Schrödinger Bridges with Multimarginal Constraints

Master Thesis

Abdulwahab Mohamed

Supervisors:
Dr. Alberto Chiarini
Dr. Oliver Tse

May 7, 2021

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Contents

Common notations

a.s.	almost surely (a property holds except on a set of probability measure zero)
a.e.	almost everywehre (a property holds except on a set of measure zero)
$\mathcal{B}(S),\ \mathcal{B}$	the Borel σ-algebra on S; if the space is clear from context $\mathcal{B}$ is used
$\mathfrak{B}(S)$	Borel measurable functions on S with values in $\mathbb{R}, \mathbb{R}^d$ or $\mathbb{R}^{d\times d}$
$\mathfrak{B}_b(S)$	bounded Borel measurable functions on S with values in $\mathbb{R}, \mathbb{R}^d$ or $\mathbb{R}^{d\times d}$
$C(\Omega; S)$	the set of continuous functions $f : \Omega \to S$
$C^k(\Omega; S)$	the set of k-times continuously differentiable functions $f : \Omega \to S$
$\Delta,\ \Delta_x$	the Laplacian operator working on the spatial variable x
$\mathbb{E}$	expectation operator, e.g. $\mathbb{E}_{\mathsf{P}}[X] = \int X \, \mathrm{d}\mathsf{P}$
$\mathcal{E}(L)$	stochastic exponential, i.e. for a local martingale L, $\mathcal{E}(L) = \exp(L - \frac{1}{2}[L])$
$\mathcal{F}$	σ-algebra
$\mathcal{F}_t$	filtration generated by the canonical process $(X_t)_{t\in[0,1]}$, i.e. $\mathcal{F}_t := \sigma(X_s : \ s \leq t)$
$H(\mathsf{P}\vert\mathsf{R})$	relative entropy of P with respect to R
$\mathrm{MP}(L)$	generic martingale problem associated to an operator L
$\mathrm{MP}(L, \eta)$	martingale problem associated to an operator L with initial distribution η
$\mathbb{N}$	the set of natural numbers
$\nabla,\ \nabla_x$	gradient operator working on the spatial variable x
$\mathcal{P}(\Omega)$	the set of probability measures on Ω
$\mathbb{p}$, P	probability measure
P^x	regular conditional probability given a random variable X, i.e. $\mathsf{P}^x = \mathsf{P}(\cdot \mid X = x)$
$\mathbb{q}$, Q	(probability) measure
$\mathbb{r}$, R	(probability) measure (usually used as a reference measure)
$X_{\#}\mathsf{P}$	the law of X under P

Chapter 1

Introduction

You have probably heard of Schrödinger's cat experiment. Good, because this thesis is not about that experiment, but another good thought experiment of Schrödinger that you may not have heard about. It is called the *lazy gas experiment.*

Imagine we have a large amount of particles scattered in a box. These particles evolve in time independently of each other and we are able to observe the distribution of these particles at the initial time $t = 0$ and a later time, say $t = 1$. Until now there is no reason why this thought experiment would cause troubles in one's head. However, the trouble starts as soon as one observes a distribution at time $t = 1$ that is different from the distribution expected from the law of large numbers.

That is what Schrödinger considered in [41] in 1931. In his paper he assumed that each particle evolves like a Brownian motion in $\mathbb{R}$ independent of the other particles. More precisely, he assumed that the distribution density $w(\cdot, t)$ of each particle at time $t \geq 0$ solves the heat equation

$$\partial_t w = \frac{1}{2}\partial_{xx} w, \quad w(x, 0) = w_0(x).$$

If the particles have distribution density $w_1(\cdot)$ at time $t = 1$, then we are also given the conditions

$$w(x, 0) = w_0(x), \quad w(x, 1) = w_1(x).$$

As Schrödinger remarks such w cannot be a solution to the heat equation for an arbitrary w_1, since the solution to the heat equation is uniquely determined by the initial value. Nevertheless, we can still ask what is the most probable evolution from w_0 to w_1. This is in fact a problem understood rigorously in terms of large deviations, where one studies events that deviate from the law of large numbers.

To make that precise in the context of the experiment, suppose R is the distribution of Brownian motion which is what the particles are supposed to follow, and let P_n be the observed distribution of the n particles. A well-known result from large deviations, namely Sanov's theorem, roughly states that for n large

$$\mathbb{P}(\mathsf{P}_n \in A) \approx \exp\left(-n \inf_{\mathsf{P}\in A} H(\mathsf{P}|\mathsf{R})\right),$$

where $H(\mathsf{P}|\mathsf{R})$ is the relative entropy of P with respect to R, which formally is

$$H(\mathsf{P}|\mathsf{R}) = \int \log\left(\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}\right) \mathrm{d}\mathsf{P}.$$

In our case we are interested in the set A of all probability measures that have density w_0 and w_1 at time $t = 0$ and $t = 1$ respectively. The experiment and the induced problem that Schrödinger started with turned out to boil down to minimizing a relative entropy.

However, at that time Schrödinger did not have Sanov's theorem in his disposal (which is formalized later by Sanov [39]). Schrödinger could still make sense of the notion of most probable distribution of the evolution of particles via statistical methods and concludes that the most probable distribution density w is uniquely determined by functions f and g solving

$$f(x_0)\int_{\mathbb{R}} p_1(x_0, x_1)g(x_1)\,\mathrm{d}x_1 = w_0(x_0), \quad g(x_1)\int_{\mathbb{R}} p_1(x_1, x_0)f(x_0)\,\mathrm{d}x_0 = w_1(x_1), \tag{1.1}$$

where $p_t(x, \cdot)$ is the probability density of Brownian motion with initial value $x \in \mathbb{R}$ at time $t \geq 0$. The most probable distribution density w is then given by

$$w(x, t) = \left(\int_{\mathbb{R}} f(x_0)p_t(x_0, x)\,\mathrm{d}x_0\right)\left(\int_{\mathbb{R}} g(x_1)p_{1-t}(x, x_1)\,\mathrm{d}x_1\right).$$

Although Equation (1.1) is non-linear and seemingly non-trivial, Schrödinger was certain that a solution (f, g) has to exist for nice enough w_0 and w_1. The reason for the conjecture is that the way he derives the equations to arrive at Equation (1.1) is statistically very logical. We refer to (1.1) as the *Schrödinger system*.

As previously explained, the function $w(\cdot, t)$ describes the most probable evolution of the particles at time t. In fact, it corresponds to a process which can be seen as a generalized Brownian bridge. Recall a Brownian bridge is a Brownian motion that starts in $a \in \mathbb{R}$ and is conditioned to be in $b \in \mathbb{R}$ at time $t = 1$. The process that we observe, however, is a Brownian motion starting with a distribution density w_0 and is conditioned to have density w_1 at time $t = 1$.

Of course, the description in terms of generalized Brownian bridge may not be obvious immediately from the density w. It is best understood when we look at it from the relative entropy perspective. Recall that we vaguely explained that Schrödinger's concern is equivalent to minimizing $H(\mathsf{P}|\mathsf{R})$ over all probability measures P with marginal densities w_0 and w_1 at time $t = 0$ and $t = 1$ respectively. The minimizer P turns out to be of the form

$$\mathsf{P} = \int_{\mathbb{R}\times\mathbb{R}} \mathsf{R}^{x_0,x_1}\,\mu(\mathrm{d}x_0\,\mathrm{d}x_1),$$

with R^{x_0,x_1} being the law of a Brownian bridge starting at $x_0 \in \mathbb{R}$ and ending at $x_1 \in \mathbb{R}$, and μ being the minimizer of the so-called "static Schrödinger problem". Obviously, if $\mu = \delta_{x_0} \otimes \delta_{x_1}$ then we get P to be the law of a classical Brownian bridge. In our case μ must have marginal densities w_0 and w_1. To understand where this μ comes from, define a new measure ν on $\mathbb{R} \times \mathbb{R}$ by

$$\nu(\mathrm{d}x_0\,\mathrm{d}x_1) := p_1(x_0, x_1)\,\mathrm{d}x_0\,\mathrm{d}x_1,$$

This ν is the joint law of a Brownian motion at time $t = 0$ and $t = 1$ with initial distribution being the Lebesgue measure. Then we can formulate the static Schrödinger problem for this particular case, and it is

Static Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mu|\nu), \\ \text{subject to} \quad & \mu_0(\mathrm{d}x_0) = w_0(x_0)\,\mathrm{d}x_0, \\ & \mu_1(\mathrm{d}x_1) = w_1(x_1)\,\mathrm{d}x_1. \end{aligned} \tag{SSP}$$

We note that (SSP) is about finding the optimal coupling μ of two probability densities w_0 and w_1 such that the relative entropy $H(\mu|\nu)$ is minimized.

Moreover, in Chapter 4 we will see that the solutions (f, g) to the Schrödinger system (1.1) are related to the minimizer μ of (SSP) via the Radon-Nikodym derivative of μ with respect to ν as follows:

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x_0, x_1) = f(x_0)g(x_1). \tag{1.2}$$

We will also see that the existence of a minimizer to (SSP) implies the existence of a solution to (1.1).

For the particles it means that they move from the density w_0 to the density w_1 such that minimal energy is used where energy is measured in terms of the relative entropy. That is the reason the experiment is called the *lazy gas experiment* for trying to lose as little energy as possible moving from the distribution density w_0 to w_1.

We have assumed that we could observe those particles at times $t = 0$ and $t = 1$, but what if we could also observe the particles at time $t = 1/2$? One may want to generalize this to $k \in \mathbb{N}$ observation points or even infinitely many observation points. What is then the most probable evolution of the particles?

As we have explained above this question boils down to a problem in large deviations which then via Sanov's theorem leads to a problem of minimizing some relative entropy. We let R be the law of a stochastic process denoted by $X = (X_t)_{t\in[0,1]}$ e.g. Brownian motion or an Itô-diffusion. Assume that we can observe the distribution of the particles at times $\mathcal{T} \subset [0, 1]$ and these are given by $(\mu_t)_{t\in\mathcal{T}}$. In the case of Schrödinger experiment we have $\mathcal{T} = \{0, 1\}$. The problem is then reduced to the multimarginal Schrödinger problem which is the following:

Multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mathsf{P}|\mathsf{R}), \\ \text{subject to} \quad & (X_t)_\#\mathsf{P} = \mu_t, \qquad \text{for all } t \in \mathcal{T}. \end{aligned} \tag{MSP}$$

We make an attempt in providing an overview about what is known about existence and characterization of the solutions. We repeat many of the already-existing proofs to understand them better. Furthermore, we provide some proofs of results that are believed to be true, but are not worked out explicitly in the literature.

We begin by considering the minimization of relative entropy in a general framework in **Chapter 2**. We state the existence and uniqueness of solutions to problems of the form (MSP) in a more abstract setting. We will state crucial parts of Csiszar's work in [10] and we repeat some of the proofs with a slower pace.

We then apply the results in the same chapter on product spaces which corresponds to the static Schrödinger problem (SSP) with finitely many marginal constraints, i.e. $|\mathcal{T}| = k < \infty$. We prove that the minimizer μ of (SSP) has a form as given in (1.2) but with k functions. We follow the proof of Föllmer and Gantert for $|\mathcal{T}| = 2$ [16]. Under additional assumptions we extend the result for general $|\mathcal{T}| = k \in \mathbb{N}$ via induction.

As soon as the foundations are settled, we consider the minimization problem in the context of Itô-diffusions in **Chapter 3**. We assume that the reference measure R is the law of the solution to

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t,$$

where B is a Brownian motion. This means that the particles are expected to follow the evolution of the SDE described by the measure R.

We prove the Markov property for the minimizer P of (MSP). We also state a version of Girsanov's theorem stating that the minimizer P is the law of the following SDE:

$$\mathrm{d}X_t = (b(X_t, t) + \sigma\sigma^\top(X_t, t)u(X_t, t))\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t, \tag{1.3}$$

for some measurable function u satisfying a finite energy condition. This implies that the solution is also an Itô-diffusion. The extra drift term u in (1.3) can be seen as a *control* that is trying to minimize the relative entropy while keeping the right marginal constraints.

Using the Markov property and Girsanov's theorem we are able to deduce a formula in terms of a limit that describes the Itô-diffusion in terms of the SDE generalizing the result in [15] for the case of Brownian motion. Furthermore, we give some words on how finitely many marginal constraints problems could be used to approximate infinitely many constraints. This is related to what Föllmer implicitly used in [14] for $\mathcal{T} = [0, 1]$. We show that the minimizer P_n of the approximate minimization problem converges weakly to the solution P of the full problem and $H(\mathsf{P}_n|\mathsf{R}) \to H(\mathsf{P}|\mathsf{R})$.

We end **Chapter 3** with some observations about time-reversibility which can also be found in [15] for the Brownian motion.

After that we come back to the Schrödinger gas experiment where one is able to observe the distribution of the particles at finitely many points given by $\mathcal{T} = \{t_1, ..., t_k\}$ for some $k \in \mathbb{N}$. In **Chapter 4** we show that the corresponding multimarginal Schrödinger problem reduces to the following seemingly simpler problem:

Static multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mu|\nu), \\ \text{subject to} \quad & \frac{\mathrm{d}\mu_i}{\mathrm{d}\nu_i} = \rho_i, \quad \text{for all } i \in \{1, ..., k\}, \end{aligned} \tag{SMSP}$$

where μ_i and ν_i are the i-th marginal of μ and ν respectively; and $(\rho_i)_{1\le i\le k}$ is given fixed densities with respect to $(\nu_i)_{1\le i\le k}$. This is a straightforward generalization of (SSP) which corresponds to a particular case with $k = 2$.

We show that (MSP) is equivalent to (SMSP) as done in [14] for $k = 2$. Furthermore, we give a result concerning the Radon-Nikodym derivative of the minimizer of (SMSP); more precisely, the following factorization holds for the minimizer:

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x_1, ..., x_k) = \prod_{i=1}^{k} f_i(x_i), \tag{1.4}$$

for some measurable functions $(f_i)_{1\le i\le k}$. The assumptions we have are slightly different from the ones given in [44] for $k = 2$ and in [3] for general $k \in \mathbb{N}$.

The factorization of the minimizer leads to solutions of a generalized version of the Schrödinger system (1.1) for general $k \in \mathbb{N}$. Moreover, we can prove that a solution to the generalized Schrödinger system implies the existence of a unique minimizer to (SMSP) under some assumptions. This equivalence is well-known for $k = 2$ (see e.g. [16]) and a slight adaptation leads to the result for general $k \in \mathbb{N}$.

After that we adapt the methods used in [44] to obtain regularity results for the functions $(f_i)_{1\le i\le k}$ in (1.4), which apply for example to reversible overdamped Langevin diffusions, i.e. we assume R is the law of the SDE

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

with initial distribution $e^{-U}\,\mathrm{d}x$. We are able to give criteria for the functions $(f_i)_{1\le i\le k}$ to be smooth or in L^p depending on the regularity of the coefficient U and the prescribed marginal densities $(\rho_i)_{1\le i\le k}$. The regularity is obtained via heat kernel lower bounds (log-Harnack inequality), the analyticity of the semigroup of the Langevin diffusion, and L^p-elliptic regularity.

The regularity of the functions $(f_i)_{1\le i\le k}$ allows us through Itô's formula to derive the Hamilton-Jacobi equations in [24] which tells us that the control u in (1.3) solves

$$(\partial_t + \mathcal{L}_t)u + \frac{1}{2}|\sigma^\top \nabla u|^2 = 0, \tag{1.5}$$

where $\mathcal{L}_t$ is the generator corresponding to the reference SDE.

We illustrate the theory above for a fairly simple example with Ornstein-Uhlenbeck process. We take $\mathcal{T} = \{0, 1\}$ and the simplicity allows us to solve for the corresponding functions $(f_i)_{1\le i\le k}$ in (1.4) explicitly. We are also able to compare the regularity that would be obtained via different methods.

A natural next step would be to treat the case with $|\mathcal{T}| = \infty$. However, $\mathcal{T}$ in general can be quite pathological, so we will focus mainly on the case $\mathcal{T} = [0, 1]$ in **Chapter 5** where we can derive better characterization of the minimizer.

We approach the problem with $\mathcal{T} = [0, 1]$ along the same lines as done in [7]. We will prove existence and characterization results of the minimizer of (MSP). We have a slightly different set of assumptions than given in [7], which is less general but still broadly applicable. The reason for the detour is that we find our way a little bit clearer while still being on the same lines.

We find a remarkable result under the following setting (see Proposition 5.12). We consider prescribed marginals $(\mu_t)_{t\in[0,1]}$ that is the marginal law of an SDE of the form

$$\mathrm{d}X_t = \sigma\sigma^\top(X_t, t)\left(\nabla V(X_t, t) + \gamma(X_t, t)\right)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t,$$

for some Brownian motion $B = (B_t)_{t\in[0,1]}$; and the reference measure R is the law of

$$\mathrm{d}X_t = \sigma\sigma^\top(X_t, t)\left(\nabla U(X_t, t) + \beta(X_t, t)\right)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t,$$

where both functions γ and β are divergence free in a generalized sense

$$\text{“}\operatorname{div}(\sigma\sigma^\top\beta\mu_t) = \operatorname{div}(\sigma\sigma^\top\gamma\mu_t) = 0\text{”}.$$

Then the minimizer of (MSP) for this particular setting is the law of

$$\mathrm{d}X_t = \sigma\sigma^\top(X_t, t)\left(\nabla V(X_t, t) + \beta(X_t, t)\right)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t.$$

As one may guess from the classical Fokker-Planck equation, the divergence free condition on γ makes the gradient ∇V the only responsible term ensuring the correct marginal laws. The function β is in some sense the only term getting the minimizer “close” to the reference measure. Surprisingly, this fact is independent of the initial laws.

Furthermore, the results obtained in the last chapter will have consequences for the Schrödinger problem (MSP) for general $\mathcal{T}$. Finally, we conjecture that the control u from (1.3) is always a limit of functions satisfying the Hamilton-Jacobi equation (1.5).

Literature review

Short after Schrödinger published [41], Fortet became interested in the problem that Schrödinger ended up with, especially about solving the Schrödinger system (1.1) which Schrödinger left open. In [17], Fortet tackles the problem with fixed-point iteration methods. A little bit later around 1960 Beurling in [4] works on the same problem and actually uses an argument involving an entropy functional in his proof without direct reference to large deviations.

In the same spirit as Fortet and Beurling, also Jamison worked on Schrödinger system in 1974 [21]. However, Jamison related the problem more to stochastic processes and in particular Markov and reciprocal processes.

Around 1988 Föllmer noted in [14] that the problem that Schrödinger posed in [41] can be seen as a problem in large deviations. It was 50 years later that it was broadly realized that Schrödinger's experiment could be understood in terms of large deviations and Sanov's theorem provides the most probable distribution that the particles followed. Sanov wrote the theorem in [39] around 1960 which is approximately 30 years later after Schrödinger came up with the gas experiment.

Moreover, Föllmer shows that Schrödinger's instinct in arriving to the Schrödinger system (1.1) is justified by results obtained by Csiszar in [10]. Csiszar's paper was mostly to study minimization of relative entropy more abstractly. Its motivation came from Sanov's theorem and did not show any particular interest in solving Schrödinger's problem. However, the results obtained there were still of great significance in understanding Schrödinger's system (1.1). It helped Föllmer and Gantert to show existence of solutions to (1.1) in [16].

In the works of Csiszar, Föllmer and Gantert counterexamples to existence of solution to (1.1) are provided. It became clear that it is not always true that solutions to Schrödinger's system (1.1) exists.

Existence results of (1.1) based on Csiszar's paper are also given in [38, 37] by Rüschendorf and Thomsen. They even generalized the problem to the case where one observes the distribution at $k \in \mathbb{N}$ different times.

Föllmer also added a twist to Schrödinger's experiment and assumed that one is able to observe the distribution at each time $0 \leq t \leq 1$, and in [14] he solves that problem via approximation argument using Csiszar's results. Föllmer worked with Brownian motion most of the time. However, the results hold more generally as proved by Leonard and Cattiaux in [7, 8].

The works of Leonard and Cattiaux is based on ideas and results by, among others, Carlen [6] and Mikami [29]. Obviously all this is related, but Carlen and Mikami were interested more in Nelson's stochastic mechanics which is motivated by the other Schrödinger's equations [31].

Mikami and Leonard worked on the problem to various extent. For instance, Mikami did a large contribution to stochastic optimal control which can be seen as generalization to the Schrödinger problem, see for instance [28]. The recent paper of Leonard and Baradat is very illuminating for it covers very general results where the underlying process can be Markov or reciprocal [3]. They give a characterization of the minimizer of the relative entropy for which the factorization result in (1.4) is a special case.

Also Tamanini in his PhD thesis [44] worked on the Schrödinger problem. A notable contribution is the local and global boundedness of the solution to Schrödinger system. To the best of my knowledge, the methods used to show boundedness is not used somewhere else before.

Finally, there are a lot of books and surveys for this topics. We mention Aebi's book [1] and Nagasawa's book [30] for a good reference on the developments around the 90s. Furthermore, Leonards's survey [24] puts the Schrödinger problem in a historical perspective. Another one with a similar spirit and more humorous narrative is [9]. Leonard's survey [24] also mentions the applications and the growth caused by the Schrödinger problem in other fields of science and in particular mathematics. Examples are Euclidean quantum mechanics, the theory of reciprocal processes, (weak) Fokker-Planck equations, stochastic optimal control, regularization of optimal transport [24] and data-assimilation [34].

Chapter 2

Preliminary results on the minimization of relative entropy

We provide the foundations of minimizing the relative entropy in an abstract setting. We will state the main results from [10] and give some consequences in Section 2.1. The abstract setting allows us to prove general results using as few notations as possible.

We consider a Polish space $\mathcal{X}$, i.e. a separable complete metrizable topological space. We endow $\mathcal{X}$ with its Borel σ-algebra $\mathcal{B}(\mathcal{X})$. Let $\mathcal{P}(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$. We endow $\mathcal{P}(\mathcal{X})$ with the topology of weak convergence. Fix a probability measure $\mathbb{r} \in \mathcal{P}(\mathcal{X})$. Recall the definition of the relative entropy of $\mathbb{p}$ with respect to $\mathbb{r}$ denoted by $H(\mathbb{p}|\mathbb{r})$ and defined as

$$H(\mathbb{p}|\mathbb{r}) := \begin{cases} \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right) \mathrm{d}\mathbb{p}, & \text{if } \mathbb{p} \ll \mathbb{r}, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ be a set of probability measures which will serve as a set of constraints. The results that we will mention require $\mathcal{C}$ to be convex and/or closed. Note that $\mathcal{C}$ being closed means closed in $\mathcal{P}(\mathcal{X})$ with respect to the topology of weak convergence. We consider the following minimization problem:

Minimization of relative entropy.

$$\begin{aligned} \min \quad & H(\mathbb{p}|\mathbb{r}), \\ \text{subject to} \quad & \mathbb{p} \in \mathcal{C}. \end{aligned} \tag{2.2}$$

Any probability measure $\mathbb{p} \in \mathcal{C}$ is called *feasible*. Note that for any $\mathbb{p} \in \mathcal{C}$ such that $\mathbb{p} \not\ll \mathbb{r}$ gives an infinite relative entropy. So, only the case where we have $\mathbb{p} \ll \mathbb{r}$ giving finite entropy is interesting. Any probability measure $\mathbb{p} \in \mathcal{C}$ such that $H(\mathbb{p}|\mathbb{r}) < \infty$ is called a *competitor*. The set of all competitors deserves a notation, namely we define

$$\mathcal{C}_H := \{\mathbb{p} \in \mathcal{C} : \ H(\mathbb{p}|\mathbb{r}) < \infty\}.$$

It happens so that the existence of one competitor leads to the existence and uniqueness of a minimizer to the problem (2.2).

Theorem 2.1 (Existence of minimizer). *Let $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ be convex and closed. Assume there exists a competitor $\mathbb{q} \in \mathcal{C}_H$, then there exists a unique solution for the minimization problem* (2.2).

Proof. The proof is based on three facts about relative entropy, namely that the map $\mathbb{p} \mapsto H(\mathbb{p}|\mathbb{r})$ has compact sublevel sets, is lower semi-continuous and on the set of probability measures yielding finite relative entropy it is strictly convex (Proposition B.3 in Appendix B).

Let us proceed and define

$$\mathcal{C}' := \{\mathbb{p} \in \mathcal{C} : \ H(\mathbb{p}|\mathbb{r}) \leq H(\mathbb{q}|\mathbb{r})\}.$$

Note that we only have to look for the minimizer inside the intersection $\mathcal{C} \cap \mathcal{C}'$, because we already know that the minimizer $\mathbb{p} \in \mathcal{C}$ satisfies $H(\mathbb{p}|\mathbb{r}) \leq H(\mathbb{q}|\mathbb{r})$ by definition. By assumption $\mathcal{C}$ is closed and $\mathcal{C}'$ is compact since the relative entropy has compact sublevel sets, that implies that the intersection $\mathcal{C} \cap \mathcal{C}' \subset \mathcal{P}(\mathcal{X})$ is compact.

Keeping these facts in mind, take any sequence $(\mathbb{p}_n)_n \subset \mathcal{C}$ for which

$$H(\mathbb{p}_n|\mathbb{r}) \to \inf_{\mathbb{p}' \in \mathcal{C}} H(\mathbb{p}'|\mathbb{r}), \quad \text{ as } n \to \infty.$$

By compactness, there exists a subsequence $(\mathbb{p}_{n_k})_k$ that converges to some $\mathbb{p} \in \mathcal{C}$. But then by lower semi-continuity

$$H(\mathbb{p}|\mathbb{r}) \leq \liminf_{k \to \infty} H(\mathbb{p}_{n_k}|\mathbb{r}) = \inf_{\mathbb{p}' \in \mathcal{C}} H(\mathbb{p}'|\mathbb{r}).$$

Hence

$$H(\mathbb{p}|\mathbb{r}) = \inf_{\mathbb{p}' \in \mathcal{C}} H(\mathbb{p}'|\mathbb{r}).$$

Why is it unique? Well, assume there exist two feasible probability measures $\mathbb{p}_1, \mathbb{p}_2 \in \mathcal{C}$ minimizing $H(\cdot|\mathbb{r})$. In such case we must have $H(\mathbb{p}_i|\mathbb{r}) \leq H(\mathbb{q}|\mathbb{r}) < \infty$ for each $i = 1, 2$. If $\mathbb{p}_1 \neq \mathbb{p}_2$ then by strict convexity

$$H\left(\frac{1}{2}\mathbb{p}_1 + \frac{1}{2}\mathbb{p}_2|\mathbb{r}\right) < \frac{1}{2}H(\mathbb{p}_1|\mathbb{r}) + \frac{1}{2}H(\mathbb{p}_2|\mathbb{r}) = H(\mathbb{p}|\mathbb{r}) \leq H(\mathbb{q}|\mathbb{r}).$$

But since $\frac{1}{2}\mathbb{p}_1 + \frac{1}{2}\mathbb{p}_2 \in \mathcal{C} \cap \mathcal{C}'$ for $\mathcal{C}$ being convex and the previous inequality, we found that our minimizer was wrong which contradicts what we have started with. $\square$

We do not get into details about finding competitors for the moment. In the next sections, e.g. Chapter 4 and Chapter 5, where the constraint set $\mathcal{C}$ takes the form of marginal constraints, then we can easily find a candidate that might be a competitor.

There is a useful result about the minimizer, namely a kind of a triangle inequality. It allows us to do approximations and to verify whether a feasible probability measure $\mathbb{p} \in \mathcal{C}$ is a minimizer. Before stating the result, we introduce the notion of an *algebraic inner point* of $\mathcal{C}$ which is any probability measure $\mathbb{p} \in \mathcal{C}$ such that

$$\mathbb{p} = \lambda \mathbb{p}_1 + (1 - \lambda)\mathbb{p}_2,$$

for some $\lambda \in (0, 1)$ and $\mathbb{p}_1, \mathbb{p}_2 \in \mathcal{C} \setminus \{\mathbb{p}\}$. Now we can state Theorem 2.2 in [10] which has the following formulation:

Theorem 2.2. *Let $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ be convex. A probability measure $\mathbb{p} \in \mathcal{C}_H$ is the unique solution to* (2.2) *if and only if*

$$H(\mathbb{q}|\mathbb{r}) \geq H(\mathbb{q}|\mathbb{p}) + H(\mathbb{p}|\mathbb{r}) \quad \text{for all } \mathbb{q} \in \mathcal{C}. \tag{2.3}$$

Moreover, if $\mathbb{p}$ is an algebraic inner point of $\mathcal{C}$ then $\mathcal{C} \subset \mathcal{C}_H$ and we have an equality in (2.3).

This theorem allows us to obtain results that we will use very often throughout the whole thesis. The first one is a slightly different formulated version of Theorem 3.1 in [10] which is about a special type of constraint set.

Theorem 2.3. *Let $\mathcal{C}$ be given by*

$$\mathcal{C} := \{\mathbb{q} \in \mathcal{P}(\mathcal{X}) : \ \mathbb{E}_{\mathbb{q}}[\xi_i] = a_i \ \text{ for all } i \in \mathbb{N}\},$$

for some sequence of numbers $(a_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ and $(\xi_i)_{i\in\mathbb{N}}$ a family of bounded measurable functions on $\mathcal{X}$ taking values in $\mathbb{R}$. If a unique minimizer $\mathbb{p} \in \mathcal{C}_H$ exists, then $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ is in the $L^1(\mathbb{p})$-closure of the linear span of the constant function 1 and the functions $(\xi_i)_{i\in\mathbb{N}}$, i.e.

$$\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} \in \overline{\operatorname{span}\{1, \xi_i : \ i \in \mathbb{N}\}}^{L^1(\mathbb{p})}.$$

Proof. Obviously $\mathcal{C}$ is convex so Theorem 2.2 is applicable. We use that and a little bit of functional analysis. Define the set of functions

$$M := \operatorname{span}\{1, \xi_i : \ i \in \mathbb{N}\}.$$

Note that $M \subset L^1(\mathbb{p})$ for all the functions in M are bounded. Since the dual space of $L^1(\mathbb{p})$ is $(L^1(\mathbb{p}))' = L^\infty(\mathbb{p})$, the annihilator of M is

$$M^\perp := \{\zeta \in L^\infty(\mathbb{p}) : \ \mathbb{E}_{\mathbb{p}}[\zeta\xi] = 0 \ \text{ for all } \xi \in M\}.$$

If $M^\perp = \{0\}$, then M is dense in $L^1(\mathbb{p})$. But by assuming $\mathbb{p} \in \mathcal{C}_H$ we are assuming $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} \in L^1(\mathbb{p})$ and the claim follows. Otherwise, if $M^\perp \neq \{0\}$, then there exists a function $\zeta \neq 0$ with $|\zeta| \leq 1$ $\mathbb{p}$-a.s. and as such the measure $\tilde{\mathbb{p}}$ defined by

$$\frac{\mathrm{d}\tilde{\mathbb{p}}}{\mathrm{d}\mathbb{p}} = 1 + \zeta,$$

is a feasible probability measure. Indeed, because for any $i \in \mathbb{N}$

$$\mathbb{E}_{\tilde{\mathbb{p}}}[\xi_i] = \mathbb{E}_{\mathbb{p}}\left[(1+\zeta)\xi_i\right] = \underbrace{\mathbb{E}_{\mathbb{p}}[\xi_i]}_{=a_i} + \underbrace{\mathbb{E}_{\mathbb{p}}[\zeta\xi_i]}_{=0} = a_i.$$

This all implies that the set

$$\mathcal{C}' := \left\{\mathbb{q} \in \mathcal{C} : \ \frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}} \leq 2\right\}$$

contains at least three elements, namely: $\mathbb{p}$, $\tilde{\mathbb{p}}$ and $2\mathbb{p} - \tilde{\mathbb{p}}$. Moreover

$$\mathbb{p} = \frac{1}{2}\tilde{\mathbb{p}} + \frac{1}{2}(2\mathbb{p} - \tilde{\mathbb{p}})$$

which makes $\mathbb{p}$ an algebraic inner point in $\mathcal{C}'$. It is easy to see that $\mathcal{C}'$ is convex so that we are allowed to apply Theorem 2.2 on $\mathcal{C}'$. It gives $\mathcal{C}' \subset \mathcal{C}_H$ and the equality

$$H(\mathbb{q}|\mathbb{r}) = H(\mathbb{q}|\mathbb{p}) + H(\mathbb{p}|\mathbb{r}) \quad \text{for all } \mathbb{q} \in \mathcal{C}'. \tag{2.4}$$

So for any $\mathbb{q} \in \mathcal{C}'$ we have

$$\begin{aligned} H(\mathbb{p}|\mathbb{r}) &= H(\mathbb{q}|\mathbb{r}) - H(\mathbb{q}|\mathbb{p}) \\ &= \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\right) \mathrm{d}\mathbb{q} - \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}}\right) \mathrm{d}\mathbb{q} \\ &= \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}}\right)^{-1}\right) \mathrm{d}\mathbb{q}. \end{aligned}$$

Each element is well-defined and $\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}} > 0$ $\mathbb{q}$-a.s. which means that we can write

$$H(\mathbb{p}|\mathbb{r}) = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}\right) \mathrm{d}\mathbb{q} = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right) \mathrm{d}\mathbb{q} = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right) \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}} \,\mathrm{d}\mathbb{p}.$$

Hence by putting $H(\mathbb{p}|\mathbb{r})$ to the other side we can write (2.4) as follows:

$$\int_{\mathcal{X}} \left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}} - 1\right) \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right) \mathrm{d}\mathbb{p} = 0, \qquad \text{for all } \mathbb{q} \in \mathcal{C}'. \tag{2.5}$$

Recall that for any $\zeta \in M^{\perp}$ satisfying $\|\zeta\|_{L^\infty(\mathbb{p})} \leq 1$ we can define a probability measure $\mathbb{q} \in \mathcal{C}'$ through

$$\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}} = 1 + \zeta.$$

Using this $\mathbb{q} \in \mathcal{C}'$ in (2.5) yields

$$\int_{\mathcal{X}} \zeta \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right) \mathrm{d}\mathbb{p} = 0.$$

This holds for any $\zeta \in M^{\perp}$ by scaling. Therefore

$$\log\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} \in \{\xi \in L^1(\mathbb{p}) : \ \mathbb{E}_{\mathbb{p}}[\zeta\xi] = 0 \quad \text{for all } \zeta \in M^{\perp}\} = \overline{M}^{L^1(\mathbb{p})},$$

by a standard property of annihilators (Theorem 4.7 in [36]). This is the claim we were after. $\square$

There is actually a somehow converse statement to the previous theorem stating that whenever $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ is in the subspace spanned by 1 and $(\xi_i)_{i\in\mathbb{N}}$ without closure, i.e.

$$\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} \in \operatorname{span}\{1, \xi_i : \ i \in \mathbb{N}\},$$

then $\mathbb{p}$ is the unique minimizer of (2.2) (see Theorem 3.1 in [10]). We do not need that result in that particular form, but we will state and prove a more general result in Theorem 2.6 for a specific type of minimization problem to be considered next.

2.1 Minimization of relative entropy given marginal constraints on product spaces

In this section we consider a special set of constraints, namely marginal constraints. The underlying space now is assumed to be of the form $\mathcal{X} = \bigtimes_{i=1}^k \mathcal{X}_i$ where each of the $\mathcal{X}_i$ is a Polish space. Endowing $\mathcal{X}$ with the product topology makes $\mathcal{X}$ also a Polish space. Let $\pi_i : \mathcal{X} \to \mathcal{X}_i$ be the projection to the i-th component, i.e. for any $x = (x_1, ..., x_k) \in \mathcal{X}$ we have

$$\pi_i(x) = x_i.$$

For each $1 \leq i \leq k$ consider a given probability measure μ_i on $\mathcal{X}_i$. We consider the constraints where we let the i-th marginal of $\mathbb{p}$ be equal to μ_i. More precisely, define the set of constraints

$$\mathcal{C}_\mu := \{\mathbb{p} \in \mathcal{P}(\mathcal{X}) : \ (\pi_i)_\# \mathbb{p} = \mu_i \ \text{ for all } 1 \leq i \leq k\}.$$

Similarly define $\mathcal{C}_{\mu,H}$ as the set of probability measures $\mathbb{p} \in \mathcal{C}_\mu$ for which $H(\mathbb{p}|\mathbb{r}) < \infty$. We show that we can write $\mathcal{C}_\mu$ in the form of the constraint set that is given in Theorem 2.3.

First, since each $\mathcal{X}_i$ is a Polish space, we can find a countable π-system Π_i generating the Borel σ-algebra $\mathcal{B}(\mathcal{X}_i)$. In particular Π_i can be written as

$$\Pi_i = \{A_{i,j} \in \mathcal{B}(\mathcal{X}_i) : \ j \in \mathbb{N}\}.$$

By uniqueness of measure theorem (Lemma 1.17 in [22]) we have

$$\mathcal{C}_\mu = \{\mathbb{p} \in \mathcal{P}(\mathcal{X}) : \ \mathbb{E}_\mathbb{p}[\mathbb{1}_{A_{i,j}}(\pi_i)] = \mu_i(A_{i,j}) \quad \text{for all } (i,j) \in \{1, ..., k\} \times \mathbb{N}\}.$$

Since $\{1, ..., k\} \times \mathbb{N}$ is countable we can write the sequence $(\mathbb{1}_{A_{i,j}}(\pi_i))_{1\leq i\leq k, j\in\mathbb{N}}$ as a single sequence of functions $(\xi_i)_{i\in\mathbb{N}}$. Similarly we can gather the sequence $(\mu_i(A_{i,j}))_{1\leq i\leq k, j\in\mathbb{N}}$ into a single sequence $(a_i)_{i\in\mathbb{N}}$. This allows us to write $\mathcal{C}_\mu$ in the form given in Theorem 2.3. For this particular choice of constraints we get that for the minimizer $\mathbb{p}$ the function $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ is a limit of sums. More precisely,

$$\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}(x) = \lim_{n\to\infty} \sum_{i=1}^k \varphi_{i,n}(x_i) \qquad \text{in } L^1(\mathbb{p}). \tag{2.6}$$

for some measurable functions $(\varphi_{i,n})$. We may ask ourselves if the limit is also of such form, i.e. a sum of functions of each variable separately. That turns out to be indeed the case. However, before stating the main result, we first state the following technical lemma:

Lemma 2.4 (Converging sum converges to sum)**.** *Let $\mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2$ be a product of two Polish spaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ endowed with a probability measure $\mathbb{q}$. Consider a measurable function $c : \mathcal{Y} \to \mathbb{R}$. Assume there exists a measurable set $A \subset \mathcal{Y}$ such that $\mathbb{q}(A) = 1$ and*

$$c(y_1, y_2) = \lim_{n\to\infty} (a_n(y_1) + b_n(y_2)) \qquad \textit{for all } (y_1, y_2) \in A,$$

for some sequences of measurable functions $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$. Then there exist two measurable sets $B_1 \subset \mathcal{Y}_1$ and $B_2 \subset \mathcal{Y}_2$ for which the following hold:

(i) there exists a function $T : \mathcal{Y}_1 \to \mathcal{Y}_1$ such that $a_n(y_1) - a_n(T(y_1))$ converges for all $y_1 \in B_1$ and $b_n(y_2) + a_n(T(y_1))$ converges for all $y_2 \in B_2$ such that $(y_1, y_2) \in A \cap (B_1 \times B_2)$;

(ii) there exist two functions $a : \mathcal{Y}_1 \to \mathbb{R}$ and $b : \mathcal{Y}_2 \to \mathbb{R}$ such that

$$c(y_1, y_2) = a(y_1) + b(y_2) \qquad \textit{for all } (y_1, y_2) \in A \cap (B_1 \times B_2);$$

(iii) $q_1(B_1) = 1$ and $q_2(B_2) = 1$ with q_1 and q_2 being the marginals of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively and in particular $\mathbb{q}(A \cap (B_1 \times B_2)) = 1$;

(iv) if in addition $\mathbb{q} \ll q_1 \otimes q_2$ or $q_1 \otimes q_2 \ll \mathbb{q}$, then we can take the function $T : \mathcal{Y}_1 \to \mathcal{Y}_1$ to be measurable. In particular, the functions a and b can be chosen to be measurable as well.

Proof. The proof is based on the proof of Proposition 3.19 in [16]. We have a similar result except that we state what we get during the proof explicitly in the statement of the theorem.

The result seems to be very intuitive. We will use a "slicing" argument where we slice A into a product of two sets having full measure. This allows us to have a set of elements $B_1 \subset \mathcal{Y}_1$ for which $a_n(y_1)$ minus a "correction" term converges. We achieve this by introducing carefully chosen elements of equivalence classes of these good elements in B_1. We do something similar for the second coordinate. So the proof is split in determining the good guys, showing existence of the limit for these good guys and finally showing that we can choose the function T to be measurable under the additional condition $\mathbb{q} \ll q_1 \otimes q_2$ or $q_1 \otimes q_2 \ll \mathbb{q}$.

"Finding the good guys". Let us start making this rigorous by introducing the sections A_{y_1} for each $y_1 \in \mathcal{Y}_1$

$$A_{y_1} := \{y_2 \in \mathcal{Y}_2 : \ (y_1, y_2) \in A\}.$$

The set A_{y_1} is measurable for any $y_1 \in \mathcal{Y}_1$. The following claim is crucial:

Claim. For any $y_1, y_1' \in \mathcal{Y}_1$ we either have

$$A_{y_1} = A_{y_1'} \quad \text{or} \quad A_{y_1} \cap A_{y_1'} = \varnothing. \tag{2.7}$$

If $A_{y_1} \cap A_{y_1'} = \varnothing$, then we are done. Otherwise, there is $z \in A_{y_1} \cap A_{y_1'}$ and take any $y_2 \in A_{y_1}$ and let us show that $y_2 \in A_{y_1'}$ which is equivalent to the existence of the limit as $n \to \infty$ of $a_n(y_1') + b_n(y_2)$. Note that

$$a_n(y_1') + b_n(y_2) = (a_n(y_1') + b_n(z)) - (a_n(y_1) + b_n(z)) + (a_n(y_1) + b_n(y_2)),$$

converges for being a sum of three converging sequences. This proves the claim.

Let π_1 and π_2 be the projections on $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively. Let q_1 and q_2 be the laws of π_1 and π_2 under $\mathbb{q}$. We set $\mathbb{q}(\cdot|y_1)$ as the regular conditional probability measure given $\pi_1 = y_1$. Define

$$B_1 := \{y_1 \in \mathcal{Y}_1 : \ \mathbb{q}(A|y_1) = 1\}.$$

Note that $q_1(B_1) = 1$ which can be obtained by the tower property for conditional expectations

$$1 = \mathbb{q}(A) = \mathbb{E}_{\mathbb{q}}[\mathbb{q}(A \mid \pi_1)].$$

The equality we have obtained and the fact $\mathbb{q}(A \mid \pi_1 \leq 1$ $\mathbb{q}$-a.s. yields $\mathbb{q}(A \mid y_1) = 1$ q_1-a.s., which is $q_1(B_1) = 1$. Furthermore, for every $y_1 \in B_1$ we have $A_{y_1} \neq \varnothing$, because

$$1 = \mathbb{q}(A \mid y_1) = \mathbb{q}(\{y_1\} \times A_{y_1} \mid y_1).$$

Similarly, we can define the sections

$$A^{y_2} := \{y_1 \in B_1 : \ (y_1, y_2) \in A\}.$$

Note that we are sectioning A and taking the intersection with B_1 in the definition of A^{y_2}. Through a similar procedure we get a set $B_2 \subset \mathcal{Y}_2$ such that $q_2(B_2) = 1$ and $A^{y_2} \neq \varnothing$ for all $y_2 \in B_2$.

These sets B_1 and B_2 contain the good guys. However, we need to filter them out via equivalence classes. Due to (2.7) we can define an equivalence relation on B_1 as follows:

$$y_1 \sim y_1' \quad \text{if and only if} \quad A_{y_1} = A_{y_1'}.$$

Let $E(y_1) \subset B_1$ be the equivalence class of $y_1 \in B_1$. For any $y_2 \in A_{y_1}$, it follows from the definition of A_{y_1} and $E(y_1)$ that

$$E(y_1) = \{y_1' \in B_1 : \ (y_1', y_2) \in A\},$$

and therefore $E(y_1) \subset B_1$ is measurable. Using the Axiom of Choice we take a representative of each equivalence class $E(y_1)$ and call it $T(y_1)$ (see Definition 1.1 in [19]). We take an arbitrary y_1 from the representatives and extend T by that element on $\mathcal{Y}_1 \setminus B_1$ so this gives us the map $T : \mathcal{Y}_1 \to \mathcal{Y}_1$. Let us summarize the last part. Each of these good guys in B_1 which are similar in the sense that they have the same sets of y_2s for which the limit exists be put into one group and get assigned one as the boss which is our representative.

"Showing the limit exists for the good guys". With these representatives, i.e. the mapping $T : \mathcal{Y}_1 \to \mathcal{Y}_1$, we can show the limit as $n \to \infty$ of this newly defined sequence

$$\tilde{a}_n(y_1) := a_n(y_1) - a_n(T(y_1)),$$

exists for all $y_1 \in B_1$. To that end, observe that for any $y_1 \in B_1$ and $y_2 \in A_{y_1}$, we have $y_2 \in A_{T(y_1)}$, so that

$$\tilde{a}_n(y_1) = a_n(y_1) + b_n(y_2) - (a_n(T(y_1)) + b_n(y_2))$$

converges for being a sum of two convergent sequences. Hence there exists a finite limit for any $y_1 \in B_1$

$$a(y_1) := \lim_{n\to\infty} \tilde{a}_n(y_1).$$

We want to assign for each $y_2 \in B_2$ the good guy y_1 in B_1 for which the limit $a_n(y_1)+b_n(y_2)$ exists. But we know that there exists exactly one equivalence class $E(y_1)$ for which the limit exists due to (2.7). The assigning y_2 to the representative y_1 is done by a mapping $S : B_2 \to B_1$ such that $y_2 \in A_{S(y_2)}$. Now define for each $y_2 \in B_2$ the function sequence

$$\tilde{b}_n(y_2) := b_n(y_2) + a_n(S(y_2)).$$

Then by construction we get for each $y_2 \in B_2$ the existence of the limit:

$$b(y_2) := \lim_{n\to\infty} \tilde{b}_n(y_2).$$

For any $(y_1, y_2) \in A \cap (B_1 \times B_2)$ we have

$$\underbrace{a_n(y_1) + b_n(y_2)}_{\text{converges}} = (a_n(y_1) - a_n(T(y_1))) + (b_n(y_2) + a_n(T(y_1))) = \underbrace{\tilde{a}_n(y_1)}_{\text{converges}} + (b_n(y_2) + a_n(T(y_1))).$$

We know that $y_2 \in A_{T(y_1)}$, since the left-hand side and the first term on the right-hand side converge, so $T(y_1) = S(y_2)$. That implies that the last term on the right-hand side is equal to $\tilde{b}_n(y_2)$ which converges too. This all yields for any $(y_1, y_2) \in A \cap (B_1 \times B_2)$

$$\lim_{n\to\infty} (a_n(y_1) + b_n(y_2)) = \lim_{n\to\infty} \tilde{a}_n(y_1) + \lim_{n\to\infty} \tilde{b}_n(y_2) = a(y_1) + b(y_2).$$

At this point we have shown everything except $\mathbb{q}(A \cap (B_1 \times B_2)) = 1$. We know that $\mathbb{q}(A) = 1$ which means that it is enough to show $\mathbb{q}(B_1 \times B_2) = 1$. We write $B_1 \times B_2$ as follows:

$$B_1 \times B_2 = (B_1 \times \mathcal{Y}_2) \cap (\mathcal{Y}_1 \times B_2).$$

We have constructed B_1 and B_2 to satisfy

$$\mathbb{q}(B_1 \times \mathcal{Y}_2) = q_1(B_1) = 1, \quad \text{and} \quad \mathbb{q}(\mathcal{Y}_1 \times B_2) = q_2(B_2) = 1,$$

from which we immediately obtain $\mathbb{q}(A \cap (B_1 \times B_2)) = 1$, since a finite intersection of events happening almost surely happens almost surely. This shows the points (i)-(iii).

"Measurability of the function T under the condition $\mathbb{q} \ll q_1 \otimes q_2$ or $q_1 \otimes q_2 \ll \mathbb{q}$". Now we show the last point (iv). First we assume $\mathbb{q} \ll q_1 \otimes q_2$. What happens with the extra assumption is the possibility to consider a new measurable set $\tilde{B}_1 \subset \mathcal{Y}_1$ instead of B_1, namely

$$\tilde{B}_1 := \{y_1 \in B_1 : \; q_2(A_{y_1}) > 0\}.$$

This set is measurable by e.g. Tonelli's theorem.

We will show that $q_1(\tilde{B}_1) = 1$. To that end, we denote by $q_2(\cdot \mid y_1)$ the regular conditional law of π_2 given $\pi_1 = y_1$, i.e. we set

$$q_2(\cdot \mid y_1) = \mathbb{q}(\mathcal{Y}_1 \times \cdot \mid \pi_1 = y_1).$$

Note that

$$q_2(A_{y_1} \mid y_1) = \mathbb{q}(\mathcal{Y}_1 \times A_{y_1} \mid \pi_1 = y_1) = \mathbb{q}(\{y_1\} \times A_{y_1} \mid \pi_1 = y_1) = \mathbb{q}(A \mid \pi_1 = y_1).$$

By definition of B_1 we have $q_2(A_{y_1} \mid y_1) = 1$ for all $y_1 \in B_1$. The assumption $\mathbb{q} \ll q_1 \otimes q_2$ implies $q_2(\cdot|y_1) \ll q_2$ for q_1-a.s. y_1 by Proposition B.1 in [3]. Therefore, for q_1-a.s. $y_1 \in B_1$ we have $q_2(A_{y_1}) > 0$ for otherwise we would have $q_2(A_{y_1} \mid y_1) = 0$ which contradicts $y_1 \in B_1$. Hence $q_1(\tilde{B}_1) = 1$. We can similarly define a measurable set $\tilde{B}_2 \subset B_2$ such that $q_2(\tilde{B}_2) = 1$.

Recall the equivalence classes from above. We can do it for $\tilde{B}_1$ now, except that now we only get at most countably many equivalent classes. To see the reason, note that we can write $\tilde{B}_1$ as follows:

$$\tilde{B}_1 = \bigcup_{j=0}^{\infty} \left\{y_1 \in B_1 : \; q_2(A_{y_1}) \in (2^{-j-1}, 2^{-j}]\right\}.$$

For each $j \geq 0$ we have that $q_2(A_{y_1}) \in (2^{-j-1}, 2^{-j}]$ can only hold for a finitely many disjoint sets A_{y_1}. If there would be infinitely many sets, then they are all disjoint by (2.7) allowing them to destroy the boundedness of q_2 via the countable additivity property of the probability measure. Therefore, we basically look at a union of finite collections of A_{y_1} which is again at most countable.

The Axiom of Choice now gives us representatives of each class in $\tilde{B}_1$ denote them by $(z_j)_{j \in J}$ with J at most countable index set. Again, we get a map from $\tilde{B}_1$ to the representatives in $\tilde{B}_1$. We denote it by $\tilde{T}$ (which may be different from the previously defined T). So this function $\tilde{T}$ is only defined on $\tilde{B}_1$ but we extend it by z_1 on $\mathcal{Y}_1 \setminus \tilde{B}_1$. We get a map $\tilde{T} : \mathcal{Y}_1 \to \mathcal{Y}_1$ which can be written as

$$\tilde{T}(y_1) = z_1 \mathbb{1}_{\mathcal{Y}_1 \setminus \tilde{B}_1}(y_1) + \sum_{j \in J} z_j \mathbb{1}_{E(z_j)}(y_1).$$

This shows that $\tilde{T}$ is a (at most countable) sum of indicators over mutually disjoint measurable sets, so it is measurable. Just as before we can show that this $\tilde{T}$ allows for the same convergence for a_n and b_n as T. We note that we can also get a measurable map $\tilde{S}$ instead of S as well for the second coordinate $\mathcal{Y}$. We know that

$$\tilde{a}_n = a_n - a_n \circ \tilde{T}, \quad \tilde{b}_n = b_n + a_n \circ \tilde{S},$$

converge on $\tilde{B}_1$ and $\tilde{B}_2$ respectively. We set

$$a(y_1) := \limsup_{n\to\infty} a_n(y_1) - a_n(T(y_1)), \quad b(y_2) := \limsup_{n\to\infty} b_n(y_2) + a_n(S(y_2)).$$

Then a and b are both measurable. We can show that $c(y_1, y_2) = a(y_1) + b(y_2)$ $\mathbb{q}$-a.s. as done before. This concludes the proof under the condition $\mathbb{q} \ll q_1 \otimes q_2$

Finally, we want to show the measurability of T under the condition $q_1 \otimes q_2 \ll \mathbb{q}$. The same argument also works if $q_1 \otimes q_2 \ll \mathbb{q}$ by taking the same $\tilde{B}_1$, because then we have the limit converges $q_1 \otimes q_2$-a.s. We have the same conclusion, for instance $q_1(\tilde{B}_1) = 1$ is what mattered most and in this case we have it, simply because

$$1 = \int_{\mathcal{Y}_1\times\mathcal{Y}_2} \mathbb{1}_A \mathrm{d}q_1 \otimes q_2 = \int_{\mathcal{Y}_1} q_2(A_{y_1})\, q_1(\mathrm{d}y_1).$$

Therefore, we must have $q_2(A_{y_1}) = 1$ y_1-a.s. and hence $q_1(\tilde{B}_1) = 1$. In fact, this implies there is only one equivalence class which is kind of intuitive. That implies that there is one $y_1^* \in \tilde{B}_1$ such that $a_n(y_1) - a_n(y_1^*)$ converges for all $y_1 \in \tilde{B}_1$. Now we can continue as before to finish the proof. □

This lemma generalizes to a sum of finitely many terms through induction with slightly more assumptions. We want to apply it to $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ where $\mathbb{p}$ is the solution to problem (2.2) with constraints $\mathcal{C}_\mu$. This lemma, as useful as it may sound, is not so practical unless the functions a and b are measurable. Even though the functions a_n and b_n are measurable, the limiting functions a and b can still be unmeasurable. In fact, there are cases where no measurable decomposition exists as one finds in Section 5 of [16].

The reason why the measurability might break down in Lemma 2.4 is that we have

$$a = \lim_{n\to\infty} a_n - a_n \circ T, \quad b = \lim_{n\to\infty} b_n + a_n \circ T,$$

and in our construction the function T without the assumption in (iv) might be unmeasurable. It might be impossible to prove measurability for we rely on the Axiom of Choice. After all, after what happened to the Vitali set, it is not the first time the Axiom of Choice is leaving us with something unmeasurable.

As mentioned before, we want to apply the previous lemma as an inductive step to get the following result about the minimizer of (2.2):

Theorem 2.5. *Consider the minimization problem* (2.2) *with constraint set $\mathcal{C}_\mu$. Assume that the minimization problem has a unique minimizer $\mathbb{p} \in \mathcal{C}_{\mu,H}$. Then there is a sequence of measurable functions $\varphi_{i,n} : \mathcal{X}_i \to R$ indexed by $n \in \mathbb{N}$ and $i \in \{1, ..., k\}$ such that*

$$\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}(x) = \lim_{n\to\infty} \sum_{i=1}^{k} \varphi_{i,n}(x_i), \quad \mathbb{p}\text{-a.s.}$$

If in addition

$$\bigotimes_{i=1}^{k} \mu_i \sim \mathbb{p},$$

then there are measurable functions $\varphi_i : \mathcal{X}_i \to \mathbb{R}$ indexed by $i \in \{1, ..., k\}$, such that

$$\log \frac{d\mathbb{p}}{d\mathbb{r}}(x) = \sum_{i=1}^{k} \varphi_i(x_i), \quad \mathbb{p}\text{-}a.s.$$

Moreover, the decomposition is $\mathbb{p}$-almost surely unique up to additive constants, i.e. we can only change $(\varphi_i)_{1\le i\le k}$ to $(\varphi_i + c_i)_{1\le i\le k}$ for some $(c_i)_{1\le i\le k}$ such that $\sum_{i=1}^{k} c_i = 0$.

Proof. We split the proof in three parts. First we show the approximation in terms of sequence of sums. Secondly, we show the existence of a measurable decomposition with induction. Finally we show the uniqueness statement.

"Approximation of minimizer via a sum of functions". First of all Theorem 2.3 tells us that $\log \frac{d\mathbb{p}}{d\mathbb{r}}$ is in the $L^1(\mathbb{p})$-closure of the space spanned by the indicators as already remarked in (2.6). We know that $L^1(\mathbb{p})$-convergence implies $\mathbb{p}$-a.s.-convergence along a subsequence. Therefore by taking a subsequence we conclude that there are measurable functions $(\varphi_{i,n})$ such that

$$\log \frac{d\mathbb{p}}{d\mathbb{r}}(x) = \lim_{n\to\infty} \sum_{i=1}^{k} \varphi_{i,n}(x_i) \quad \mathbb{p}\text{-a.s.}$$

"The existence of a measurable sum decomposition". Under the additional assumption $\otimes_{i=1}^{k}\mu_i \sim \mathbb{p}$, it is enough to prove the decomposition by induction over k for any measurable function $c : \mathcal{X} \to \mathbb{R}$ such that

$$c(x) = \lim_{n\to\infty} \sum_{i=1}^{k} \varphi_{i,n}(x_i) \quad \mathbb{p}\text{-a.s.}$$

"Base case $k = 1$". For this case, there is nothing much to prove.

"Inductive step". Assume the claim is true for some k. We will show it for $k + 1$. To that end, we use the previous Lemma 2.4. We need to write everything as in the setting of Lemma 2.4. We introduce the space $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$ through

$$\mathcal{Y}_1 = \mathcal{X}_1, \quad \mathcal{Y}_2 = \bigtimes_{i=2}^{k+1} \mathcal{X}_i.$$

Since $(\mathcal{X}_i)_{1\le i\le k}$ is a family of Polish spaces, the spaces $\mathcal{Y}_1$, $\mathcal{Y}_2$ and also $\mathcal{Y}$ are Polish spaces. We need a probability measure. Since $\mathcal{Y} = \mathcal{X}$ we can take $\mathbb{q} = \mathbb{p}$, but the marginals of $\mathbb{q}$, namely q_1 and q_2 should be understood as

$$q_1 = \mathbb{q}(\tilde{\pi}_1 \in \cdot), \quad q_2 = \mathbb{q}(\tilde{\pi}_2 \in \cdot),$$

with $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are the projections from $\mathcal{Y}$ to $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively. We need to check either $q_1 \otimes q_2 \ll \mathbb{q}$ or $\mathbb{q} \ll q_1 \otimes q_2$ to ensure the existence of measurable functions. In this case both are true, because denoting by $p_{2,...,k+1}$ the marginal on $\bigtimes_{i=2}^{k+1} \mathcal{Y}_i$ under $\mathbb{p}$ we get

$$p_{2,...,k+1} \sim \otimes_{i=2}^{k+1} p_i,$$

which is a consequence of the fact that $\mu_i = p_i$, $\mathbb{p} \sim \otimes_{i=1}^{k+1}\mu_i$ and that absolute continuity of measures implies the absolute continuity of the marginals. Using that we get

$$q_1 \otimes q_2 = p_1 \otimes p_{2,\dots,k+1} \sim p_1 \otimes \left(\otimes_{i=2}^{k+1} p_i\right) \sim \mathbb{p} = \mathbb{q}.$$

We use the variables $y_1 = x_1 \in \mathcal{Y}_1$ and $y_2 = (x_2, \dots, x_{k+1}) \in \mathcal{Y}_2$ to define the functions a_n and b_n as follows:

$$a_n(y_1) = \varphi_{1,n}(x_1), \quad b_n(y_2) = \sum_{i=2}^{k+1} \varphi_{i,n}(x_i).$$

We finally need to specify the measurable set $A \subset \mathcal{Y}$ which we set it to be

$$A := \left\{ (y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2 : \lim_{n\to\infty} (a_n(y_1) + b_n(y_2)) \in \mathbb{R} \text{ exists} \right\}.$$

Then $\mathbb{q}(A) = 1$ by construction.

We can apply Lemma 2.4 to the setting we have just constructed, to get two measurable sets $B_1 \subset \mathcal{Y}_1$ and $B_2 \subset \mathcal{Y}_2$, two functions $a : \mathcal{Y}_1 \to \mathbb{R}$ and $b : \mathcal{Y}_2 \to \mathbb{R}$ such that $\mathbb{q}(A \cap (B_1 \times B_2)) = 1$ and

$$c(y_1, y_2) = a(y_1) + b(y_2) \quad \text{for all } (y_1, y_2) \in A \cap (B_1 \times B_2).$$

There also exists one element $y_1^* \in B_1$ such that the sequence defined by

$$\tilde{a}_n(y_1) = a_n(y_1) - a_n(y_1^*),$$

converges for all $y_1 \in B_1$ to $a(y_1)$ (due to having one equivalence class, see the argument in the last part of Lemma 2.4). We also have

$$\tilde{b}_n(y_2) = b_n(y_2) + a_n(y_1^*)$$

converges for all $y_2 \in B_2$ to $b(y_2)$. Now we can apply the induction hypothesis on $\mathcal{Y}_2 = \bigtimes_{i=2}^{k+1} \mathcal{X}_i$ which is a product of k Polish spaces and that $\tilde{b}_n(y_2)$ is a sum of k terms converging on B_2 which has the property $q_2(B_2) = 1$. Hence we get that the existence of a measurable set $\tilde{B}_2 \subset \mathcal{Y}_2$ such that $q_2(\tilde{B}_2) = 1$

$$b(y_2) = \sum_{i=2}^{k+1} \varphi_i(x_i) \quad \text{for all } y_2 \in \tilde{B}_2.$$

Hence for any $(y_1, y_2) \in A \cap (B_1 \times \tilde{B}_2)$ we have

$$\begin{aligned} c(y_1, y_2) &= \lim_{n\to\infty} a_n(y_1) - a_n(y_1^*) + b_n(y_2) + a_n(y_1^*) \\ &= \lim_{n\to\infty} \tilde{a}_n(y_1) + \tilde{b}_n(y_2) \\ &= a(y_1) + \sum_{i=2}^{k+1} \varphi_i(x_i). \end{aligned}$$

Finally, by Lemma 2.4 we also get $\mathbb{q}(A \cap (B_1 \times \tilde{B}_2)) = 1$ and by setting $a(y_1) = a(x_1) =: \varphi_1(x_1)$ we get the claim for $k+1$.

"Uniqueness of the decomposition". For the uniqueness, we do not need an induction argument. It follows from Fubini's theorem. Assume that there exists a collection of functions such that $(\tilde{\varphi}_i)_{1 \le i \le k}$

$$\sum_{i=1}^{k} \varphi_i(x_i) = \sum_{i=1}^{k} \tilde{\varphi}_i(x_i), \quad \mathbb{p}\text{-a.s.}$$

Define the set where we have equality

$$U := \left\{ x \in \mathcal{X} : \sum_{i=1}^{k} \varphi_i(x_i) = \sum_{i=1}^{k} \tilde{\varphi}_i(x_i) \right\}.$$

By assumption $\mathbb{p}(U) = 1$ and therefore $\otimes_{i=1}^k \mu_i(U) = 1$.

Now take any $j \in \{1, ..., k\}$ and note that by Fubini's theorem

$$1 = \otimes_{i=1}^k \mu_i(U) = \int_{\times_{i=1}^k \mathcal{X}_i} \mathbb{1}_U(x) \, \mathrm{d} \otimes_{i=1}^k \mu_i = \int_{\times_{i \neq j} \mathcal{X}_i} \int_{X_j} \mathbb{1}_U(x) \, \mathrm{d}\mu_j \, \mathrm{d} \otimes_{i \neq j} \mu_i.$$

But the inner integral is in $[0,1]$ so we must have

$$\int_{X_j} \mathbb{1}_U(x) \, \mathrm{d}\mu_j = 1, \quad \otimes_{i \neq j} \mu_i\text{-a.s.}$$

This tells us that there exists $(\hat{x}_i)_{i \neq j} \in \times_{i \neq j} \mathcal{X}_i$ such that

$$\mathbb{1}_U(\hat{x}_1, ..., \hat{x}_{j-1}, x_j, \hat{x}_{j+1}, ..., \hat{x}_k) = 1, \quad \mu_j\text{-a.s. } x_j \in \mathcal{X}_j.$$

By definition of U this means that

$$\varphi_j(x_j) + \sum_{i \neq j} \varphi_i(\hat{x}_i) = \tilde{\varphi}_j(x_j) + \sum_{i \neq j} \tilde{\varphi}_i(\hat{x}_i), \quad \mu_j\text{-a.s. } x_j \in \mathcal{X}_j.$$

By setting

$$c_j := \sum_{i \neq j} \tilde{\varphi}_i(\hat{x}_i) - \sum_{i \neq j} \varphi_i(\hat{x}_i),$$

we get

$$\varphi_j(x_j) = \tilde{\varphi}_j(x_j) + c_j, \quad \mu_j\text{-a.s. } x_j \in \mathcal{X}_j.$$

Remember that $j \in \{1, ..., k\}$ was arbitrary which means we can write

$$\sum_{i=1}^{k} \tilde{\varphi}_i(x_i) = \sum_{i=1}^{k} \varphi_i(x_i) = \sum_{i=1}^{k} \left(\tilde{\varphi}_i(x_i) + c_i \right), \quad \otimes_{i=1}^k \mu_i\text{-a.s.}$$

But since c_i is constant for all $i \in \{1, ..., k\}$ we get the asserted claim

$$\sum_{i=1}^{k} c_i = 0.$$

$\square$

To show the existence of the decomposition we use that $\mathbb{p} \sim \otimes_{i=1}^{k} \mu_i$ in the previous theorem. In [37] the authors claim that there always exists a decomposition which may not be measurable even without the assumption $\mathbb{p} \sim \otimes_{i=1}^{k} \mu_i$. However, they state the claim without a proof as an easy generalization of the case $k = 2$. Even though the proof that we have for Lemma 2.4 is similar to theirs in [38, 37], we could not generalize the idea straightforwardly. Fortunately, the assumptions that we impose ensure measurability of the decomposition in return. That makes the result more practical than the existence of a decomposition which may not be measurable.

One could ask whether the converse statement holds. We have the following giving us a sufficient condition for a probability measure $\mathbb{p} \in \mathcal{C}_\mu$ to be the unique minimizer:

Theorem 2.6. *Consider the problem* (2.2) *with constraint set* $\mathcal{C}_\mu$. *Assume there is a feasible* $\mathbb{p} \in \mathcal{C}_{\mu,H}$ *for which the Radon-Nikodym derivative* $\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ *can be written as*

$$\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}(x) = \prod_{i=1}^{k} f_i(x_i), \quad \mathbb{r}\text{-}a.s., \qquad \text{with } \log f_i \in L^1(\mathcal{X}_i, \mu_i),\ i \in \{1, ..., k\}.$$

Then $\mathbb{p}$ *is the unique minimizer.*

Proof. In our proof we will use a similar idea as in the proof of Theorem 3.1 in [10] which is related, in fact what we want to prove is part of Corollary 3.1 in the same paper. The argument is based on Theorem 2.2.

Let us elaborate the argument for the sake of completeness. To that end, we start by taking an arbitrary feasible $\mathbb{q} \in \mathcal{C}_\mu$. We want to show

$$H(\mathbb{q}|\mathbb{r}) \geq H(\mathbb{q}|\mathbb{p}) + H(\mathbb{p}|\mathbb{r}),$$

so that the claim follows from Theorem 2.2. If $H(\mathbb{q}|\mathbb{r}) = \infty$ we do not have to do anything. Let us assume $H(\mathbb{q}|\mathbb{r}) < \infty$. First we calculate

$$H(\mathbb{p}|\mathbb{r}) = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}(x)\right) \mathbb{p}(\mathrm{d}x) = \int_{\mathcal{X}} \sum_{i=1}^{k} \log f_i(x_i)\, \mathbb{p}(\mathrm{d}x) = \sum_{i=1}^{k} \int_{\mathcal{X}_i} \log f_i(x_i)\, \mu_i(\mathrm{d}x_i),$$

where we have used that $x \mapsto \log f_i(x_i)$ is in $L^1(\mathbb{p})$ via Tonelli's theorem which allows us to exchange the sum and the integral. Now recall that $\mathbb{q}$ has the same marginals as $\mathbb{p}$ so that $x \mapsto \log f_i(x_i)$ is in $L^1(\mathbb{q})$ as well and for each $i \in \{1, ..., k\}$

$$\int_{\mathcal{X}_i} \log f_i(x_i)\, \mu_i(\mathrm{d}x_i) = \int_{\mathcal{X}} \log f_i(x_i)\, \mathbb{q}(\mathrm{d}x).$$

Therefore, we can do the previous calculation in reverse to get

$$H(\mathbb{p}|\mathbb{r}) = \sum_{i=1}^{k} \int_{\mathcal{X}} \log f_i(x_i)\,\mathbb{q}(\mathrm{d}x) = \int_{\mathcal{X}} \sum_{i=1}^{k} \log f_i(x_i)\,\mathbb{q}(\mathrm{d}x) = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}(x)\right)\mathbb{q}(\mathrm{d}x),$$

where we have used the expression for $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}$ in terms of $(f_i)_{1\le i\le k}$ holds $\mathbb{r}$-a.s., and in particular $\mathbb{q}$-a.s. by the assumption $H(\mathbb{q}|\mathbb{r}) < \infty$. The previous calculation also ensures that $\log \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} \in L^1(\mathbb{q})$ from which we get that $\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}} > 0$ $\mathbb{q}$-a.s. for free. We want to use that observation to argue that $\mathbb{q} \ll \mathbb{p}$. Indeed, for any $A \in \mathcal{B}(\mathcal{X})$ we have

$$\mathbb{E}_{\mathbb{q}}[\mathbb{1}_A] = \mathbb{E}_{\mathbb{q}}\left[\mathbb{1}_A \mathbb{1}_{\{\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}>0\}} \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)^{-1}\right].$$

Moreover the assumption $H(\mathbb{q}|\mathbb{r}) < \infty$ gives $\mathbb{q} \ll \mathbb{r}$ so we can continue writing

$$\mathbb{E}_{\mathbb{q}}[\mathbb{1}_A] = \mathbb{E}_{\mathbb{r}}\left[\mathbb{1}_A \mathbb{1}_{\{\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}>0\}} \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)^{-1}\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\right] = \mathbb{E}_{\mathbb{p}}\left[\mathbb{1}_A \mathbb{1}_{\{\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}>0\}} \left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)^{-1}\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\right].$$

This allows us to conclude via Radon-Nikodym's theorem that $\mathbb{q} \ll \mathbb{p}$ and

$$\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}>0\}}\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)^{-1}\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}} = \left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)^{-1}\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}} \quad \mathbb{q}\text{-a.s..}$$

We are almost done, because all the calculations above finally lead to

$$H(\mathbb{q}|\mathbb{p}) = \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{p}}\right)\mathrm{d}\mathbb{q} = \int_{\mathcal{X}}\left[\log\left(\frac{\mathrm{d}\mathbb{q}}{\mathrm{d}\mathbb{r}}\right) - \log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{r}}\right)\right]\mathrm{d}\mathbb{q} = H(\mathbb{q}|\mathbb{r}) - H(\mathbb{p}|\mathbb{r}).$$

This proves the inequality we were after which in our case is an equality. Hence by Theorem 2.2 $\mathbb{p}$ is the unique minimizer of (2.2). $\square$

Remark 2.7. Note that we do need $\log f_i \in L^1(\mathcal{X}_i, \mu_i)$ to conclude

$$\int_{\mathcal{X}} \sum_{i=1}^{k} \log f_i(x_i)\,\mathbb{p}(\mathrm{d}x) = \int_{\mathcal{X}} \sum_{i=1}^{k} \log f_i(x_i)\,\mathbb{q}(\mathrm{d}x),$$

although both $\mathbb{p}$ and $\mathbb{q}$ from the previous proof have the same marginals. It is indeed tempting to say that the expectation of a sum of functions is determined by the marginal laws of each function. That is true for $k = 2$, but surprisingly enough it is not necessarily true for $k > 2$ as proved in [42].

The fact that the integrability of $\log f_i$ is not needed for $k = 2$ explains why Corollary 3.15 in [16] works; it is similar to our Theorem 2.6 with the additional assumption $\mathbb{p} \sim \mathbb{r}$ instead of $\log f_i \in L^1(\mathcal{X}_i, \mu_i)$ as we have.

In the next chapter we will apply the results obtained in this section to minimization problems in the form of (2.2) where the reference measure is the law of an Itô-diffusion.

Chapter 3

Itô-diffusions and multimarginal Schrödinger problem

We consider a path space $\Omega = C([0,1];\mathbb{R}^d)$ endowed with the completion of the Borel σ-algebra $\mathcal{F}$ induced by the topology of supremum norm. Let $X = (X_t)_{t\in[0,1]}$ be the canonical process defined by $X_t(\omega) := \omega(t)$. We consider the filtration $(\mathcal{F}_t)_{t\in[0,1]}$ generated by X, i.e. $\mathcal{F}_t := \sigma(X_s : \ s \le t)$. In such case we actually have $\mathcal{F} = \mathcal{F}_1$ by Lemma 16.1 in [22]. Furthermore, for any probability measure $\mathsf{P} \in \mathcal{P}(\Omega)$ we use the notation $(X_t)_{\#}\mathsf{P}$ for the law of X_t under P.

We let $\mathsf{R} \in \mathcal{P}(\Omega)$ be the reference measure for which the canonical process X is an Itô-diffusion with initial law $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$. To make that precise, consider the (pre)generator $(\mathcal{L}_t)_{t\in[0,1]}$ defined for all $f \in C_c^\infty(\mathbb{R}^d)$ by

$$\mathcal{L}_t f(x) := \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x,t)\partial_{ij} f(x) + \sum_{i=1}^{d} b_i(x,t)\partial_i f(x), \tag{3.1}$$

where $a = \sigma\sigma^\top$ for some real $d \times d'$ matrix-valued locally bounded Borel measurable function $\sigma : \mathbb{R}^d \times [0,1] \to \mathbb{R}^{d\times d'}$ and a vector-valued locally bounded Borel measurable function $b : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$. We assume that R is a Markov probability measure solving the martingale problem corresponding to $\mathcal{L}_t$ with initial law ν_0 meaning that for any $f \in C_c^\infty(\mathbb{R}^d)$

$$N_t^f := f(X_t) - f(X_0) - \int_0^t \mathcal{L}_s f(X_s)\,\mathrm{d}s,$$

is a $(\mathsf{R}, \mathcal{F}_t)$-martingale; and we use the notation $\mathsf{R} \in \mathrm{MP}(\mathcal{L}_t, \nu_0)$ and the notation $\mathrm{MP}(\mathcal{L}_t)$ to consider a general martingale problem with generator $\mathcal{L}_t$ without specifying the initial law.

In such case, we know that X is a R-semimartingale and has the semimartingale decomposition

$$X_t = X_0 + M_t + A_t,$$

with the finite variation part $A = (A_t)_{t\in[0,1]}$ and the local martingale part $M = (M_t)_{t\in[0,1]}$

being

$$A_t = \int_0^t b(X_s, s)\,\mathrm{d}s, \qquad M_t = X_t - X_0 - \int_0^t b(X_s, s)\,\mathrm{d}s.$$

Moreover, the quadratic variation of M is in matrix notation

$$[M]_t = \int_0^t a(X_s, s)\,\mathrm{d}s,$$

which more precisely means

$$[M_t^{(i)}, M_t^{(j)}]_t = \int_0^t a_{ij}(X_s, s)\,\mathrm{d}s, \quad 1 \le i, j \le d.$$

The semimartingale decomposition above (with $X_0 \sim \nu_0$) is equivalent to $\mathsf{R} \in \mathrm{MP}(\mathcal{L}_t, \nu_0)$, and that is equivalent to saying that X under R has the same law as that of the weak solution to the SDE

$$\mathrm{d}Y_t = b(Y_t, t)\,\mathrm{d}t + \sigma(Y_t, t)\,\mathrm{d}B_t,$$

for some Brownian motion B. However, in general we may not be able to write the semimartingale decomposition of X in terms of the SDE that Y solves on the same probability space Ω, but an enlargement of the space might be required (see Theorem 21.7 in [22] for full details).

Assume a measurable set $\mathcal{T} \subset [0, 1]$ and a family of probability measures $(\mu_t)_{t\in\mathcal{T}}$ on $\mathbb{R}^d$ are given. We want to minimize the relative entropy $H(\mathsf{P}|\mathsf{R})$ under the constraint that the marginals of X_t under P is equal to μ_t for all $t \in \mathcal{T}$. More precisely, we consider the following minimization problem:

Multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mathsf{P}|\mathsf{R}), \\ \text{subject to} \quad & (X_t)_\#\mathsf{P} = \mu_t, \qquad \text{for all } t \in \mathcal{T}. \end{aligned} \tag{MSP}$$

This corresponds to the minimization problem (2.2) with the constraint set being

$$\mathcal{C}^{\mathrm{MSP}} := \{\mathsf{P} \in \mathcal{P}(\Omega) : \ (X_t)_\#\mathsf{P} = \mu_t \ \text{ for all } t \in \mathcal{T}\}.$$

Just as before we define the set of competitors as well:

$$\mathcal{C}_H^{\mathrm{MSP}} := \{\mathsf{P} \in \mathcal{C}^{\mathrm{MSP}} : \ H(\mathsf{P}|\mathsf{R}) < \infty\}.$$

Remark 3.1. Since the relative entropy is infinite when $\mathsf{P} \not\ll \mathsf{R}$, it makes sense to only consider the case where $\mu_t \ll (X_t)_\#\mathsf{R}$ for all $t \in \mathcal{T}$. That is due to the fact that $\mathsf{P} \ll \mathsf{R}$ implies $(X_t)_\#\mathsf{P} \ll (X_t)_\#\mathsf{R}$ so as soon as $\mu_t \not\ll (X_t)_\#\mathsf{R}$ we cannot have $\mathsf{P} \ll \mathsf{R}$.

The constraint set $\mathcal{C}^{\mathrm{MSP}}$ being closed and convex leads to the same existence result as given in Theorem 2.1. For overview we state the existence and uniqueness of minimizer for this setting.

Theorem 3.2 (Existence of minimizer of Schrödinger problem)**.** *Assume there exists a competitor* $\mathsf{Q} \in \mathcal{C}_H^{\mathrm{MSP}}$*, then there exists a unique solution for the minimization problem* (MSP)*.*

Proof. We only have to check whether $\mathcal{C}^{\mathrm{MSP}}$ is closed and convex. We show that $\mathcal{C}^{\mathrm{MSP}}$ is a closed set. To that end, consider a sequence $(\mathsf{P}_n)_n \subset \mathcal{C}^{\mathrm{MSP}}$ that converges weakly to some P. We need to show P is in $\mathcal{C}^{\mathrm{MSP}}$, i.e. $(X_t)_\# \mathsf{P} = \mu_t$ for all $t \in \mathcal{T}$. Let $f \in C_b(\mathbb{R}^d)$, then

$$\int_{\mathbb{R}^d} f \,\mathrm{d}(X_t)_\# \mathsf{P} = \int_\Omega f(X_t)\,\mathrm{d}\mathsf{P}.$$

But note that $f(X_t) \in C_b(\Omega)$, indeed boundedness is clear and we have

$$\lim_{\omega' \to \omega} f(X_t(\omega')) = \lim_{\omega' \to \omega} f(\omega'(t)) = f(\omega(t)) = f(X_t(\omega)),$$

since uniform convergence implies pointwise convergence. Therefore

$$\int_{\mathbb{R}^d} f \,\mathrm{d}(X_t)_\# \mathsf{P} = \int_\Omega f(X_t)\,\mathrm{d}\mathsf{P} = \lim_{n\to\infty} \int_\Omega f(X_t)\,\mathrm{d}\mathsf{P}_n = \int_{\mathbb{R}^d} f\,\mathrm{d}\mu_t,$$

yielding $(X_t)_\# \mathsf{P} = \mu_t$, since bounded continuous functions separate probability measures on a Polish space (Theorem 13.11 in [23]). This shows that $\mathsf{P} \in \mathcal{C}^{\mathrm{MSP}}$ and hence $\mathcal{C}^{\mathrm{MSP}}$ is closed in $\mathcal{P}(\Omega)$ under the topology of weak convergence.

For convexity, take any $\lambda \in [0,1]$ and two probability measures $\mathsf{P}_1, \mathsf{P}_2 \in \mathcal{C}^{\mathrm{MSP}}$, then for any $t \in \mathcal{T}$ we have

$$(X_t)_\#(\lambda \mathsf{P}_1 + (1-\lambda)\mathsf{P}_2) = \lambda \underbrace{(X_t)_\# \mathsf{P}_1}_{=\mu_t} + (1-\lambda) \underbrace{(X_t)_\# \mathsf{P}_2}_{=\mu_t} = \lambda \mu_t + (1-\lambda)\mu_t = \mu_t$$

implying $\lambda \mathsf{P}_1 + (1-\lambda)\mathsf{P}_2 \in \mathcal{C}^{\mathrm{MSP}}$. That gives the convexity of $\mathcal{C}^{\mathrm{MSP}}$.

Now that we know $\mathcal{C}^{\mathrm{MSP}}$ is closed and convex, the result follows from Theorem 2.1. □

3.1 Markovian nature of the solution

The minimizer P of (MSP) (if it exists) inherits some properties of the reference measure R. Recall the definition of Markov process/measure.

Definition 3.3 (Markov process/measure)**.** A probability measure $\mathsf{Q} \in \mathcal{P}(\Omega)$ is called a *Markov measure* or *Markovian* if for any $t \in [0,1]$ and $A \in \sigma(X_s : \ s \geq t)$ we have

$$\mathsf{Q}(A \mid \mathcal{F}_t) = \mathsf{Q}(A \mid X_t) \quad \mathsf{Q}\text{-a.s.}.$$

In such case we say that the canonical process X is a *Markov process* under Q.

This definition is equivalent to saying that for any $t \in [0,1]$, $A \in \sigma(X_s : \ s \leq t)$ and $B \in \sigma(X_s : \ s \geq t)$ we have

$$\mathsf{Q}(A \cap B \mid X_t) = \mathsf{Q}(A \mid X_t)\mathsf{Q}(B \mid X_t) \quad \mathsf{Q}\text{-a.s.}. \tag{3.2}$$

Proposition 3.4 (Markovian minimizer)**.** *Assume a unique minimizer* P *of* (MSP) *exists with finite entropy. If the reference measure* R *is Markovian, then the minimizer* P *is also Markovian.*

Proof. We follow the same argument given in the proof of Theorem 4.5 in [3]. We want to show the Markov property (3.2) for P. The idea is to construct a probability measure $\tilde{\mathsf{P}}$ from P that satisfies the Markov property (3.2) for each $s \in [0,1]$. We then show it is also a minimizer which makes it equal to the minimizer P by uniqueness.

We need some notations first. We start by fixing $s \in [0,1]$. For any probability measure $\mathsf{Q} \in \mathcal{P}(\Omega)$ we denote the regular conditional probability given X_s by $\mathsf{Q}^{X_s=x} := \mathsf{Q}(\cdot \mid X_s = x)$. We denote the space of continuous functions on $[0,s]$ and $[s,1]$ by $\Omega_{[0,s]}$ and $\Omega_{[s,1]}$ respectively. We let $X^{[0,s]}$ and $X^{[s,1]}$ be the canonical processes on $\Omega_{[0,s]}$ and $\Omega_{[s,1]}$. We endow the Polish spaces $\Omega_{[0,s]}$ and $\Omega_{[s,1]}$ with the σ-algebras $\mathcal{F}_{[0,s]} = \sigma(X_r^{[0,s]} : \ r \in [0,s])$ and $\mathcal{F}_{[s,1]} = \sigma(X_r^{[s,1]} : r \in [s,1])$ respectively. On each sample space, we define a probability measure, namely

$$\mathsf{Q}_{\leftarrow s} := (X_{[0,s]})_{\#}\mathsf{Q}, \quad \text{and} \quad \mathsf{Q}_{s\rightarrow} := (X_{[s,1]})_{\#}\mathsf{Q} \quad \text{respectively.}$$

We know that $\mathcal{F} = \mathcal{F}_1$ and $\mathcal{F}_1$ is generated by cylinder sets, i.e. sets of the form

$$\bigcap_{i=1}^{n} \{X_{t_i} \in C_i\},$$

for some $n \in \mathbb{N}$, $0 \le t_1 < t_2 < ... < t_n \le 1$ and $C_i \in \mathcal{B}(\mathbb{R}^d)$, which implies that for any $E \in \mathcal{F}$ there exists two sets $E_1 \in \mathcal{F}_{[0,s]}$ and $E_2 \in \mathcal{F}_{[s,1]}$ such that

$$E = \{\omega \in \Omega : \ \omega|_{[0,s]} \in E_1 \text{ and } \omega|_{[s,1]} \in E_2\}.$$

Keeping these notations in mind, for any $E \in \mathcal{F}$ we define

$$\tilde{\mathsf{Q}}^{(s)}(E) := \int_{\mathbb{R}^d} [\mathsf{Q}_{\leftarrow s}^{X_s=x} \otimes \mathsf{Q}_{s\rightarrow}^{X_s=x}](E_1 \times E_2)\,(X_s)_{\#}\mathsf{Q}(\mathrm{d}x).$$

One can check that $\tilde{\mathsf{Q}}^{(s)}$ is a well-defined probability measure on Ω. Loosely speaking, the reason why we construct the probability measure $\tilde{\mathsf{Q}}^{(s)}$ is that it satisfies the Markov property at time s.

We proceed by showing the following:

(i) $\tilde{\mathsf{P}}^{(s)}$ preserves marginals, in particular $\tilde{\mathsf{P}}^{(s)}$ is feasible for the problem (MSP).

(ii) $\tilde{\mathsf{P}}^{(s)}$ satisfies the Markov property at time s; more precisely, for any $A \in \sigma(X_r : \ r \in [0,s])$ and $B \in \sigma(X_r : \ r \in [s,1])$ we have

$$\tilde{\mathsf{P}}^{(s)}(A \cap B \mid X_s) = \tilde{\mathsf{P}}^{(s)}(A \mid X_s)\tilde{\mathsf{P}}^{(s)}(B \mid X_s), \quad \tilde{\mathsf{P}}^{(s)}\text{-a.s.}$$

(iii) $H(\tilde{\mathsf{P}}^{(s)} \mid \mathsf{R}) \le H(\mathsf{P} \mid \mathsf{R})$.

To show (i), we first fix $t \in [0,1]$ and take an arbitrary measurable set $C \in \mathcal{B}(\mathbb{R}^d)$. We either have $t \leq s$ or $t > s$. Assume $t \leq s$, because the argument is similar to the other case anyways. We can write

$$\{X_t \in C\} = \{\omega \in \Omega : \ \omega|_{[0,s]} \in \{X_t^{[0,s]} \in C\} \quad \text{and} \quad \omega|_{[s,1]} \in \Omega_{[s,1]}\}.$$

By definition of $\mathsf{P}^{(s)}$ we get

$$\mathsf{P}^{(s)}(X_t \in C) = \int_{\mathbb{R}^d} \mathsf{P}_{\leftarrow s}^{X_s=x}(X_t^{[0,s]} \in C)\mathsf{P}_{s\rightarrow}^{X_s=x}(\Omega_{[s,1]})\,(X_s)_\#\mathsf{P}(\mathrm{d}x).$$

We have $\mathsf{P}_{s\rightarrow}^{X_s=x}(\Omega_{[s,1]}) = 1$ which gives us

$$\mathsf{P}^{(s)}(X_t \in C) = \int_{\mathbb{R}^d} \mathsf{P}_{\leftarrow s}^{X_s=x}(X_t^{[0,s]} \in C)\,(X_s)_\#\mathsf{P}(\mathrm{d}x) = \int_{\mathbb{R}^d} \mathsf{P}^{X_s=x}(X_t^{[0,s]} \in C)\,(X_s)_\#\mathsf{P}(\mathrm{d}x),$$

where we have used the obvious fact

$$\mathsf{P}_{\leftarrow s}^{X_s=x}(X_t^{[0,s]} \in C) = \mathsf{P}^{X_s=x}(X_t \in C).$$

Law of the unconscious statistician and the tower property gives

$$\mathsf{P}^{(s)}(X_t \in C) = \mathbb{E}_{\mathsf{P}}\left[\mathsf{P}^{X_s}(X_t \in C)\right] = \mathbb{E}_{\mathsf{P}}[\mathbb{E}_{\mathsf{P}}[\mathbb{1}_C(X_t) \mid X_s]] = \mathbb{E}_{\mathsf{P}}[\mathbb{1}_C(X_t)] = \mathsf{P}(X_t \in C).$$

We can do this for all $t \in [0,1]$ and in particular we get

$$(X_t)_\#\tilde{\mathsf{P}}^{(s)} = (X_t)_\#\mathsf{P} = \mu_t, \quad \text{for all } t \in \mathcal{T},$$

which makes $\tilde{\mathsf{P}}^{(s)}$ feasible.

Now we show the second point (ii). For any $A \in \sigma(X_r : \ r \in [0,s])$ and $B \in \sigma(X_r : \ r \in [s,1])$ there exists two corresponding sets $E_1 \in \mathcal{F}_{[0,s]}$ and $E_2 \in \mathcal{F}_{[s,1]}$ such that

$$A = \{\omega \in \Omega : \ \omega|_{[0,s]} \in E_1, \ \ \omega|_{[s,1]} \in \Omega_{[s,1]}\},$$

and

$$B = \{\omega \in \Omega : \ \omega|_{[0,s]} \in \Omega_{[0,s]}, \ \ \omega|_{[s,1]} \in E_2\}.$$

This characterization of the sets A and B can be proved using the cylinder sets. It is clear that

$$A \cap B = \{\omega \in \Omega : \ \omega|_{[0,s]} \in E_1, \ \ \omega|_{[s,1]} \in E_2\}.$$

Now the fact that $\tilde{\mathsf{P}}^{(s)}$ preserves marginals implies

$$\begin{aligned}
\mathbb{E}_{\tilde{\mathsf{P}}^{(s)}}[\mathbb{1}_C(X_s)\mathbb{1}_{A\cap B}] &= \int_C \left[\mathsf{P}_{\leftarrow s}^{X_s=x}(E_1)\mathsf{P}_{s\rightarrow}^{X_s=x}(E_2)\right](X_s)_\#\mathsf{P}(\mathrm{d}x) \\
&= \int_C \left[\mathsf{P}_{\leftarrow s}^{X_s=x}(E_1)\mathsf{P}_{s\rightarrow}^{X_s=x}(E_2)\right](X_s)_\#\tilde{\mathsf{P}}^{(s)}(\mathrm{d}x) \\
&= \mathbb{E}_{\tilde{\mathsf{P}}^{(s)}}[\mathbb{1}_C(X_t)\mathsf{P}_{\leftarrow s}^{X_s}(E_1)\mathsf{P}_{s\rightarrow}^{X_s}(E_2)].
\end{aligned}$$

Therefore by definition and uniqueness of conditional expectation

$$\tilde{\mathsf{P}}^{(s)}(A \cap B \mid X_s) = \mathsf{P}^{X_s}_{\leftarrow s}(E_1)\mathsf{P}^{X_s}_{s\rightarrow}(E_2), \quad \tilde{\mathsf{P}}^{(s)}\text{-a.s.}$$

It is not what we exactly want yet, but the same argument shows that

$$\tilde{\mathsf{P}}^{(s)}(A \mid X_s) = \mathsf{P}^{X_s}_{\leftarrow s}(E_1) = \mathsf{P}^{X_s}(A), \quad \tilde{\mathsf{P}}^{(s)}\text{-a.s.}$$

Similarly, we obtain

$$\tilde{\mathsf{P}}^{(s)}(B \mid X_s) = \mathsf{P}^{X_s}_{s\rightarrow}(E_2) = \mathsf{P}^{X_s}(B), \quad \tilde{\mathsf{P}}^{(s)}\text{-a.s.}$$

Putting the previous three equations together yields

$$\tilde{\mathsf{P}}^{(s)}(A \cap B \mid X_s) = \tilde{\mathsf{P}}^{(s)}(A \mid X_s)\tilde{\mathsf{P}}^{(s)}(B \mid X_s), \quad \tilde{\mathsf{P}}^{(s)}\text{-a.s..}$$

This is the assertion in (ii).

Finally, we prove the third point (iii). By conditioning on $\{X_s = x\}$ (Lemma B.4 in Appendix B)

$$H(\mathsf{P}|\mathsf{R}) = H((X_s)_\#\mathsf{P}|(X_s)_\#\mathsf{R}) + \int_{\mathbb{R}^d} H(\mathsf{P}^{X_s=x} \mid \mathsf{R}^{X_s=x})\,(X_s)_\#\mathsf{P}(\mathrm{d}x).$$

By the assumption that R is Markovian we can write $\mathsf{R}^{X_s=x} = \mathsf{R}^{X_s=x}_{\leftarrow s} \otimes \mathsf{R}^{X_s}_{s\rightarrow}$ so that by applying Lemma B.5 in Appendix B we get the following inequality

$$H(\mathsf{P}|\mathsf{R}) \geq H((X_s)_\#\mathsf{P}|(X_s)_\#\mathsf{R}) + \int_{\mathbb{R}^d} H(\mathsf{P}^{X_s=x}_{\leftarrow s} \otimes \mathsf{P}^{X_s=x}_{s\rightarrow} \mid \mathsf{R}^{X_s=x}_{\leftarrow s} \otimes \mathsf{R}^{X_s=x}_{s\rightarrow})\,(X_s)_\#\mathsf{P}(\mathrm{d}x).$$

The fact that R is Markovian, $\tilde{\mathsf{P}}^{(s)}(\cdot \mid X_s = x) = \mathsf{P}^{X_s=x}_{\leftarrow s} \otimes \mathsf{P}^{X_s=x}_{s\rightarrow}$ and that $\tilde{\mathsf{P}}^{(s)}$ preserves the marginals allows us to write the inequality as follows:

$$H(\mathsf{P}|\mathsf{R}) \geq H((X_s)_\#\tilde{\mathsf{P}}^{(s)}|(X_s)_\#\mathsf{R}) + \int_{\mathbb{R}^d} H((\tilde{\mathsf{P}}^{(s)})^{X_s=x} \mid \mathsf{R}^{X_s=x})\,(X_s)_\#\tilde{\mathsf{P}}^{(s)}(\mathrm{d}x).$$

We apply Lemma B.4 from Appendix B in reverse to get

$$H(\mathsf{P}|\mathsf{R}) \geq H(\tilde{\mathsf{P}}^{(s)} \mid \mathsf{R}),$$

which is exactly (ii).

We are almost done. We have shown that $\tilde{\mathsf{P}}^{(s)}$ is feasible and $H(\tilde{\mathsf{P}}^{(s)}|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R})$. So $\tilde{\mathsf{P}}^{(s)}$ is minimizer, but the minimizer is unique which means $\tilde{\mathsf{P}}^{(s)} = \mathsf{P}$. But we know $s \in [0,1]$ was arbitrary, so for any $s \in [0,1]$ and $A \in \sigma(X_r : r \leq s)$ and $B \in \sigma(X_r : r \geq s)$ we have by point (ii)

$$\mathsf{P}(A \cap B \mid X_s) = \tilde{\mathsf{P}}^{(s)}(A \cap B \mid X_s) = \tilde{\mathsf{P}}^{(s)}(A \mid X_s)\tilde{\mathsf{P}}^{(s)}(B \mid X_s) = \mathsf{P}(A \mid X_s)\mathsf{P}(B \mid X_s), \quad \mathsf{P}\text{-a.s.}$$

Hence P is Markovian. □

Knowing that P is Markovian is not just a beautiful fact for its own sake, but it will provide usefulness in characterizing $\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R}$ as becomes clear later, for instance in Theorem 3.14 below.

Another property is that if R is the law of a semimartingale, then so is P as one would expect from Girsanov's theorem. In addition, Leonard gives an expression for the relative entropy in terms of the "drift" in [25].

Before we state the theorem, we first need to define the notion of semi-uniqueness:

Definition 3.5 (Semi-uniqueness of martingale problem)**.** We say $\mathrm{MP}(\mathcal{L}_t)$ has the *semi-uniqueness* condition whenever two probability measures $\mathsf{Q}, \mathsf{Q}' \in \mathrm{MP}(\mathcal{L}_t)$ satisfying $(X_0)_\#\mathsf{Q}' = (X_0)_\#\mathsf{Q}$ and $\mathsf{Q}' \ll \mathsf{Q}$ we must have $\mathsf{Q}' = \mathsf{Q}$.

The previous condition is the same uniqueness condition that is called "condition (U)" in [25]. We found the name semi-uniqueness more describing, because in some sense it means unique among "half" the probability measures namely $\mathsf{Q}' \ll \mathsf{Q}$.

Of course, if the martingale problem corresponding to (3.1) has a unique solution, then the semi-uniqueness holds. Conditions to ensure that can be found in e.g. Chapter 9 from [35]. We state Theorem 2.1 and Theorem 2.3 from [25] in the following theorem:

Theorem 3.6 (Girsanov under finite entropy)**.** *Let $\mathsf{P} \in \mathcal{P}(\Omega)$ be any probability measure such that $H(\mathsf{P}|\mathsf{R}) < \infty$. Then there exists a progressively measurable process $(\beta_t)_{t\in[0,1]}$ such that*

$$\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s,s)^\top \beta_s|^2 \,\mathrm{d}s\right] < \infty$$

and the canonical process under P has the semimartingale decomposition $X = X_0 + M^{\mathsf{P}} + A^{\mathsf{P}}$ with

$$A_t^{\mathsf{P}} = \int_0^t b(X_s,s) + a(X_s,s)\beta_s \,\mathrm{d}s, \quad M_t^{\mathsf{P}} = M_t - \int_0^t a(X_s,s)\beta_s \,\mathrm{d}s,$$

and $[M^{\mathsf{P}}] = [M]$. Moreover, the relative entropy satisfies the inequality

$$H((X_0)_\#\mathsf{P}|(X_0)_\#\mathsf{R}) + \frac{1}{2}\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s,s)^\top \beta_s|^2 \,\mathrm{d}s\right] \le H(\mathsf{P}|\mathsf{R}).$$

If in addition the martingale problem $\mathrm{MP}(\mathcal{L}_t)$ satisfies the semi-uniqueness condition, then we have equality in the previous inequality and $\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R}$ can be written as

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}} \frac{\mathrm{d}(X_0)_\#\mathsf{P}}{\mathrm{d}(X_0)_\#\mathsf{R}}(X_0) \exp\left(\int_0^1 \beta_s \,\mathrm{d}M_s - \frac{1}{2}\int_0^1 |\sigma(X_s,s)^\top \beta_s|^2 \,\mathrm{d}s\right). \tag{3.3}$$

Finally, we refer to β as the control.

Remark 3.7. It is important to note that the control β is only defined P-a.s., and only defined R-a.s. on $\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} > 0\}$ as follows from the construction given in [25]. The fact that

$$\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s,s)^\top \beta_s|^2 \,\mathrm{d}s\right] < \infty,$$

implies that the stochastic integral $\int_0^1 \beta_s \,\mathrm{d}M_s$ is well-defined P-a.s. and R-a.s. on $\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} > 0\}$.

Remark 3.8. Note that when $\mathsf{P} \sim \mathsf{R}$ classical Girsanov's theorem would also give us a similar expression in terms of stochastic exponential, say $\mathcal{E}(L) = \exp(L - [L]/2)$ with L some local martingale. In such case we only know that L is a local martingale, and a priori we do not know if we can write L in terms of a stochastic integral w.r.t. M. We need some kind of martingale representation theorem to obtain the latter. However, in [25] the construction is due to a Riesz' representation argument which is applicable due to the finiteness of the relative entropy.

Note that for Theorem 3.6 we did not assume that P is Markovian. In our purpose, for the minimization problem (MSP) we consider the minimizer P which is Markovian due to Proposition 3.4. So in such case, it is tempting to claim that the control β_s is a function of (X_s, s) since X is Markovian under P. With assuming a little bit more we get the following:

Proposition 3.9. *Assume $\mathsf{P} \in \mathcal{P}(\Omega)$ is Markovian and satisfies $H(\mathsf{P}|\mathsf{R}) < \infty$. Assume that σ is uniformly bounded from above, i.e. there exists $C > 0$ such that*

$$|\sigma(x,t)| \leq C \quad \text{for all } (x,t) \in \mathbb{R}^d \times [0,1].$$

Let $(\beta_s)_{s\in[0,1]}$ be the control from Theorem 3.6. Then for almost every $t \in [0,1]$

$$\lim_{h\to 0^+} \frac{1}{h} \mathbb{E}_{\mathsf{P}} \left[X_{t+h} - X_t - \int_t^{t+h} b(X_s, s) \, \mathrm{d}s \mid \mathcal{F}_t \right] = a(X_t, t)\beta_t \quad \text{in } L^2(\Omega, \mathsf{P}).$$

In particular, there exists a Borel measurable function $F_t : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$a(X_t, t)\beta_t = F_t(X_t), \quad \mathsf{P}\text{-a.s.}$$

Proof. The following proof is along the same lines as the proof of Proposition 2.5 in [15] which is about the case when X is a BM under R. Our statement above includes such case by taking $\sigma \equiv I$.

Let us take any $t \in [0,1]$ and $h > 0$. We use the semimartingale decomposition of X under P to get

$$X_{t+h} - X_t - \int_t^{t+h} b(X_s, s) \, \mathrm{d}s - \int_t^{t+h} a(X_s, s)\beta_s \, \mathrm{d}s = M^{\mathsf{P}}_{t+h} - M^{\mathsf{P}}_t.$$

The local martingale M^{P} on the right-hand side has quadratic variation

$$[M^{\mathsf{P}}]_t = [M]_t = \int_0^t a(X_s, s) \, \mathrm{d}s,$$

and it is bounded since $|a| = |\sigma\sigma^\top| \leq C^2$. Therefore, the local martingale M^{P} is actually a martingale by e.g. Corollary 17.8 in [22]. That means that we are allowed to take conditional expectation with respect to $\mathcal{F}_t$, and moreover the conditional expectation of the martingale M^{P} on the right-hand side vanishes, i.e.

$$\mathbb{E}_{\mathsf{P}} \left[X_{t+h} - X_t - \int_t^{t+h} b(X_s, s) \, \mathrm{d}s - \int_t^{t+h} a(X_s, s)\beta_s \, \mathrm{d}s \mid \mathcal{F}_t \right] = 0. \tag{3.4}$$

We want to divide by $h > 0$ and take $h \to 0^+$. Everything seems to be as in the statement except the part $\int_t^{t+h} a(X_s, s)\beta_s \, \mathrm{d}s$ which requires a special treatment.

To that end, define for $h \in \mathbb{Q}_+$ the function $K_h : [0,1] \times \Omega \to \mathbb{R}^d$

$$K_h(t, \omega) := \frac{1}{h} \int_t^{t+h} a(X_s(\omega), s)\beta_s(\omega) \, \mathrm{d}s.$$

This function K_h is product measurable since the integrand is progressively measurable (and we extend it by zero outside $[0,1]$). Let us assume integrability of $K_h(t, \omega)$ with respect to P for the moment and divide by $h > 0$ in (3.4) to get

$$\frac{1}{h}\mathbb{E}_{\mathsf{P}}\left[X_{t+h} - X_t - \int_t^{t+h} b(X_s, s) \, \mathrm{d}s \mid \mathcal{F}_t\right] = \mathbb{E}_{\mathsf{P}}[K_h(t, \cdot) \mid \mathcal{F}_t]. \tag{3.5}$$

By conditional Jensen's inequality and tower property we get

$$\begin{aligned} \mathbb{E}_{\mathsf{P}}\left[|\mathbb{E}_{\mathsf{P}}[K_h(t,\cdot) \mid \mathcal{F}_t] - a(X_t, t)\beta_t|^2\right] &= \mathbb{E}_{\mathsf{P}}\left[|\mathbb{E}_{\mathsf{P}}[K_h(t,\cdot) - a(X_t, t)\beta_t \mid \mathcal{F}_t]|^2\right] \\ &\leq \mathbb{E}_{\mathsf{P}}\left[|K_h(t,\cdot) - a(X_t, t)\beta_t|^2\right]. \end{aligned} \tag{3.6}$$

By (3.5) and the last inequality (3.6) we see that we establish our claim once we prove that for a.e. $t \in [0,1]$ we have

$$\lim_{h \to 0^+} K_h(t, \cdot) = a(X_t, t)\beta_t, \quad \text{in } L^2(\Omega, \mathsf{P}).$$

Although we have ignored integrability issues to arrive at (3.6), we will see that the way to arrive at the inequality (3.6) addresses these issues naturally.

Let us get to business and note that the boundedness of σ and the fact that $a = \sigma\sigma^\top$ allow us to estimate

$$\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |a(X_s, s)\beta_s|^2 \, \mathrm{d}s\right] \leq C^2 \mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s, s)^\top \beta_s|^2 \, \mathrm{d}s\right] \leq C^2 H(\mathsf{P}|\mathsf{R}) < \infty. \tag{3.7}$$

This implies the existence of a measurable set $\Omega' \in \mathcal{F}$ such that $\mathsf{P}(\Omega') = 1$ and for each $\omega \in \Omega'$

$$\int_0^1 |a(X_s(\omega), s)\beta_s(\omega)|^2 \, \mathrm{d}s < \infty,$$

that is

$$[s \mapsto a(X_s(\omega), s)\beta_s(\omega)] \in L^2((0,1), \mathrm{Leb}).$$

Since $L^2((0,1), \mathrm{Leb})$ is a probability space we get $L^2((0,1), \mathrm{Leb}) \subset L^1((0,1), \mathrm{Leb})$ which in particular tells us through Lebesgue differentiation theorem that for any $\omega \in \Omega'$ we have

$$\lim_{h \to 0^+} K_h(t, \omega) = a(X_t(\omega), t)\beta_t(\omega) \quad \text{ a.e. } t \in (0,1). \tag{3.8}$$

Note that the set

$$E := \left\{ (t,\omega) \in [0,1] \in \Omega : \ \exists \lim_{h\to 0^+} K_h(t,\omega) = a(X_t(\omega),t)\beta_t(\omega) \right\},$$

is $\mathcal{B}([0,1]) \otimes \mathcal{F}$-measurable. Moreover, previously in (3.8) we have showed that

$$\int_0^1 \mathbb{1}_E(t,\omega)\,\mathrm{d}t = 1, \qquad \text{for all } \omega \in \Omega'.$$

We can apply Fubini to get that

$$(\mathrm{Leb} \otimes \mathsf{P})(E) = \int_0^1 \mathbb{E}_{\mathsf{P}}[\mathbb{1}_E]\,\mathrm{d}t = \mathbb{E}_{\mathsf{P}}\left[\int_0^1 \mathbb{1}_E\,\mathrm{d}t\right] = 1.$$

In particular since $0 \le \mathbb{1}_E \le 1$ and we are dealing with probability measures

$$\mathbb{E}_{\mathsf{P}}[\mathbb{1}_E(t,\cdot)] = 1 \quad \text{a.e. } t \in [0,1].$$

In other words, there exists $I_0 \in \mathcal{B}([0,1])$ such that $\mathrm{Leb}(I_0) = 1$ and for each $t \in I_0$ we have

$$\mathbb{E}_{\mathsf{P}}[\mathbb{1}_E(t,\cdot)] = 1,$$

which by definition of E means that

$$\lim_{h\to 0^+} K_h(t,\omega) = a(X_t(\omega),t)\beta_t(\omega), \quad \mathsf{P}\text{-a.s. } \omega \in \Omega.$$

Differently formulated, we have for each $t \in I_0$

$$\lim_{h\to 0^+} |K_h(t,\cdot) - a(X_t,t)\beta_t|^2 = 0 \quad \mathsf{P}\text{-a.s.}$$

At this point, we hope to improve the P-a.s. to $L^1(\Omega,\mathsf{P})$-convergence through Dominated convergence theorem which is equivalent to the $L^2(\Omega,\mathsf{P})$-convergence of $K_h(t,\cdot)$.

Note that with Jensen's inequality

$$\begin{aligned}
|K_h(t,\cdot) - a(X_t,t)\beta_t|^2 &= \left| \frac{1}{h}\int_t^{t+h} (a(X_s,s)\beta_s - a(X_t,t)\beta_t)\,\mathrm{d}s \right|^2 \\
&\le \frac{1}{h}\int_t^{t+h} |a(X_s,s)\beta_s - a(X_t,t)\beta_t|^2\,\mathrm{d}s.
\end{aligned}$$

Now we use the inequality $|x+y|^2 \le 2(|x|^2+|y|^2)$ to get

$$|K_h(t,\cdot) - a(X_t,t)\beta_t|^2 \le \frac{2}{h}\int_t^{t+h} |a(X_s,s)\beta_s|^2\,\mathrm{d}s + 2|a(X_t,t)\beta_t|^2.$$

The right-hand side serves as a dominating function as soon as we show the right-hand side converges in $L^1(\Omega,\mathsf{P})$.

To establish that, we note that (3.7) implies via Tonelli that there exists a measurable $I_1 \subset [0,1]$ with $\mathrm{Leb}(I_1) = 1$ such that for all $t \in I_1$ we have

$$\mathbb{E}_{\mathsf{P}}[|a(X_t,t)\beta_t|^2] < \infty,$$

and

$$\int_0^1 \mathbb{E}_{\mathsf{P}}\left[|a(X_s,s)\beta_s|^2\right] \,\mathrm{d}s \leq C^2 H(\mathsf{P}|\mathsf{R}).$$

Lebesgue differentiation theorem applied on the left-hand side yields the existence of a measurable $I_2 \subset [0,1]$ satisfying $\mathrm{Leb}(I_2)$ such that for all $t \in I_2$

$$\lim_{h\to 0^+} \frac{1}{h}\int_t^{t+h} \mathbb{E}_{\mathsf{P}}\left[|a(X_s,s)\beta_s|^2\right] \mathrm{d}s = \mathbb{E}_{\mathsf{P}}\left[|a(X_t,t)\beta_t|^2\right].$$

We know the same argument applied to $K_h(t,\omega)$ can be applied to $\frac{1}{h}\int_t^{t+h} |a(X_s,s)\beta_s|^2 \,\mathrm{d}s$ to get another set $I_3 \subset [0,1]$ satisfying $\mathrm{Leb}(I_3) = 1$ such that for all $t \in I_3$

$$\lim_{h\to 0^+} \int_t^{t+h} |a(X_s,s)\beta_s|^2 \,\mathrm{d}s = |a(X_t,t)\beta_t|^2, \quad \mathsf{P}\text{-a.s.}$$

Now let $I := \cap_{i=1}^4 I_i$ and of course the intersection of these sets satisfies $\mathrm{Leb}(I) = 1$. We can apply Sheffe's lemma (see Proposition 4.12 in [22]) to get that for all $t \in I$

$$\frac{1}{h}\int_t^{t+h} |a(X_s,s)\beta_s|^2 \,\mathrm{d}s \to |a(X_s,s)\beta_s|^2 \;\text{ in } L^1(\Omega,\mathsf{P}) \quad \text{as } h \to 0^+.$$

We have showed that our dominating function converges for all $t \in I$, so this tells us via generalized dominated convergence that for all $t \in I$

$$\lim_{h\to 0^+} \mathbb{E}_{\mathsf{P}}[|K_h(t,\cdot) - a(X_t,t)\beta_t|^2] = 0.$$

In particular, by (3.6) we get that for all $t \in I$

$$\lim_{h\to 0^+} \mathbb{E}_{\mathsf{P}}\left[\left|\frac{1}{h}\mathbb{E}_{\mathsf{P}}\left[X_{t+h} - X_t - \int_t^{t+h} b(X_s,s)\,\mathrm{d}s \mid \mathcal{F}_t\right] - a(X_t,t)\beta_t\right|^2\right] = 0.$$

This shows the first assertion since $\mathrm{Leb}(I) = 1$.

For the second assertion we note that X is a Markov process which allows us to write

$$\frac{1}{h}\mathbb{E}_{\mathsf{P}}\left[X_{t+h} - X_t - \int_t^{t+h} b(X_s,s)\,\mathrm{d}s \mid \mathcal{F}_t\right] = \frac{1}{h}\mathbb{E}_{\mathsf{P}}\left[X_{t+h} - X_t - \int_t^{t+h} b(X_s,s)\,\mathrm{d}s \mid X_t\right].$$

On the other hand that random variable converges in $L^2(\Omega,\mathsf{P})$ to $a(X_t,t)\beta_t$. Which means that there is a measurable function $F_t : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$a(X_t,t)\beta_t = F_t(X_t) \quad \mathsf{P}\text{-a.s.}$$

$\square$

The previous result gives that $a(X_t,t)\beta_t = F_t(X_t)$ with a way to find $a(X_t,t)\beta_t$ through a limit. Actually, more is true without any more assumptions than the assumptions imposed at the beginning of the chapter.

To that end, we need to introduce new spaces. Let $\mathsf{Q} \in \mathcal{P}(\Omega)$ and set $\gamma_t := (X_t)_{\#}\mathsf{Q}$ for all $t \in [0,1]$. We can define the following bilinear forms for any measurable functions $f, g : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$

$$(f,g)_{L^2(a,\gamma)} := \int_0^1 \int_{\mathbb{R}^d} f(x,t)^\top a(x,t) g(x,t)\, \gamma_t(\mathrm{d}x)\, \mathrm{d}t,$$

which is well-defined for functions in the space

$$\mathfrak{L}^2(\mathbb{R}^d \times [0,1], a, \gamma \otimes \mathrm{Leb}) := \{f : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d : \ f \text{ measurable and } (f,f)_{L^2(a,\gamma)} < \infty\}.$$

Note that $a = \sigma\sigma^\top$ which means that $f^\top a f = |\sigma^\top f|^2 \geq 0$ for all functions f. This implies that $(f,f)^{1/2}_{L^2(a,\mu)}$ is a semi-norm. Let us also define for any progressively measurable processes $\zeta, \xi : \Omega \times [0,1] \to \mathbb{R}^d$ the bilinear form

$$(\zeta,\xi)_{L^2(a,\mathsf{Q})} := \int_0^1 \mathbb{E}_{\mathsf{Q}}\left[\zeta_t^\top a(X_t,t)\xi_t\right] \mathrm{d}t,$$

and the space

$$\mathfrak{L}^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb}) = \{\zeta : \Omega \times \mathbb{R}^d \to \mathbb{R}^d : \ \zeta \text{ prog. measurable}, (\zeta,\zeta)_{L^2(a,\mathsf{Q})} < \infty\}.$$

These spaces lead to the definition of spaces using equivalence classes. We define $\sim^*$ on $\mathfrak{L}^2(\mathbb{R}^d \times [0,1], a, \gamma \otimes \mathrm{Leb})$ by $f \sim^* g \iff (f-g, f-g)_{L^2(a,\gamma)} = 0$. Now we can define the space

$$L^2(\mathbb{R}^d \times [0,1], a, \gamma \otimes \mathrm{Leb}) := \mathfrak{L}^2(\mathbb{R}^d \times [0,1], a, \gamma \otimes \mathrm{Leb})/\sim^* . \tag{3.9}$$

Similarly we can define $\sim^{**}$ on $\mathfrak{L}^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb})$ and set

$$L^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb}) := \mathfrak{L}^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb})/\sim^{**} \tag{3.10}$$

These spaces are Hilbert spaces. We need a small lemma first:

Lemma 3.10. *Let $\mathsf{Q} \in \mathcal{P}(\Omega)$. Assume there exists a progressively measurable $\mathbb{R}^d$-valued process $(\xi_t)_{t\in[0,1]}$ such that for a.e. $t \in [0,1]$*

$$\xi_t \in \mathrm{Range}(a(X_t,t)), \quad \mathsf{Q}\text{-}a.s.$$

Then there exists a progressively measurable process $(\zeta_t)_{t\in[0,1]}$ such that for a.e. $t \in [0,1]$

$$a(X_t,t)\zeta_t = \xi_t, \quad \mathsf{Q}\text{-}a.s.$$

Moreover, $\zeta_t = a(X_t,t)^+\xi_t$ where $a(X_t,t)^+$ is the Moore-Penrose inverse of $a(X_t,t)$ (see Section 20.1 in [18]). In addition, if $\xi_t = G(X_t,t)$ for some Borel measurable function $G : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, then one can choose $\zeta_t = F(X_t,t)$ for some Borel measurable function $F : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$.

Proof. We are basically after solving the equation $a(X_t,t)y = \xi_t$ for $y \in \mathbb{R}^d$ in a measurable way. Note that the matrix $a(x,t)$ may not be invertible so we cannot take the inverse. Fortunately, the Moore-Penrose inverse $a(x,t)^+$ works like an inverse for vectors in the range of $a(x,t)$. More precisely, for any $z \in \mathrm{Range}(a(x,t))$ we have $a(x,t)^+ a(x,t)z = z$ (see Corollary 20.4.3 in [18]). Therefore, by defining $\zeta_t := a(X_t,t)^+\xi_t$ we almost get the claim except the progressive measurability of ζ. We use Theorem 20.7.1 in [18] which states that the Moore-Penrose inverse can be expressed in terms of a limit

$$a(x,t)^+ = \lim_{\delta\to 0^+}(a(x,t)^2 + \delta I_{d\times d})^{-1}a(x,t).$$

Clearly this is Borel measurable for being a limit of Borel measurable functions which makes ζ progressively measurable.

Finally, for the case $\xi_t = G(X_t,t)$ we have that $\zeta_t = a(X_t,t)^+G(X_t,t) = F(X_t,t)$ where $F(x,t) := a(x,t)^+G(x,t)$ is a Borel measurable function. □

Proposition 3.11. *The spaces* $L^2(\mathbb{R}^d \times [0,1], a, \gamma\otimes\mathrm{Leb})$ *and* $L^2(\Omega\times[0,1], a, \mathsf{Q}\otimes\mathrm{Leb})$ *are Hilbert spaces with inner products* $(\cdot,\cdot)_{L^2(a,\gamma)}$ *and* $(\cdot,\cdot)_{L^2(a,\mathsf{Q})}$ *respectively.*

Proof. Showing these spaces are innerproduct spaces is standard in functional analysis so we omit those details. However, showing completeness might be tricky, because the matrix a is allowed to be singular. We show completeness of $L^2(\Omega\times[0,1], a, \mathsf{Q}\otimes\mathrm{Leb})$, because a similar argument works for $L^2(\mathbb{R}^d \times [0,1], a, \gamma\otimes\mathrm{Leb})$. Note that the induced norm is

$$\|\zeta\|^2_{L^2(a,\mathsf{Q})} = (\zeta,\zeta)_{L^2(a,\mathsf{Q})} = \int_0^1 \mathbb{E}_{\mathsf{Q}}\left[\zeta_t^\top a(X_t,t)\zeta_t\right]\mathrm{d}t = \int_0^1 \mathbb{E}_{\mathsf{Q}}\left[|\sigma(X_t,t)^\top\zeta_t|^2\right]\mathrm{d}t.$$

Now consider any Cauchy sequence $(\zeta^{(n)})_{n\in\mathbb{N}}$ in $L^2(\Omega\times[0,1], a, \mathsf{Q}\otimes\mathrm{Leb})$, then $(\sigma(X_\cdot,\cdot)^\top\zeta^{(n)})_{n\in\mathbb{N}}$ is a Cauchy sequence in the standard $L^2(\Omega\times[0,1],\mathsf{Q}\otimes\mathrm{Leb})$ which we know is a Hilbert space. Therefore, there exists $\psi \in L^2(\Omega\times[0,1],\mathsf{Q}\otimes\mathrm{Leb})$ such that $\sigma(X_\cdot,\cdot)^\top\zeta^{(n)} \to \psi$ in $L^2(\Omega\times[0,1],\mathsf{Q}\otimes\mathrm{Leb})$. We are done as soon as we show that there exists $\zeta \in L^2(\Omega\times[0,1], a, \mathsf{Q}\otimes\mathrm{Leb})$ such that $\sigma(X_\cdot,\cdot)^\top\zeta = \psi$.

To that end, note that there exists a subsequence such that $\sigma(X_\cdot,\cdot)^\top\zeta^{(n_k)} \to \psi$ $\mathsf{Q}\otimes$Leb-a.s. Moreover, $\mathrm{Range}(\sigma(X_t(\omega),t)^\top)$ is a linear subspace of the finite-dimensional space $\mathbb{R}^{d'}$ which makes $\mathrm{Range}(\sigma(X_t(\omega),t)^\top)$ closed. Therefore, $\psi_t(\omega) \in \mathrm{Range}(\sigma(X_t(\omega),t)^\top)$ for a.s. (ω,t). In particular,

$$\min_{x\in\mathbb{R}^d}|\sigma(X_t(\omega),t)^\top x - \psi_t(\omega)|^2 = 0, \quad \mathsf{Q}\otimes\text{Leb-a.s. } (\omega,t). \tag{3.11}$$

Such least squares problem is minimized by solutions $x \in \mathbb{R}^d$ of

$$\sigma(X_t(\omega),t)\sigma(X_t(\omega),t)^\top x = \sigma(X_t(\omega),t)\psi_t(\omega), \tag{3.12}$$

by Theorem 20.6.1 in [18]. Note that $a = \sigma\sigma^\top$ which makes $\xi_t(\omega) := \sigma(X_t(\omega),t)\psi_t(\omega) \in \mathrm{Range}(a(X_t(\omega),t))$ a.s. (ω,t). Moreover $(\xi)_{t\in[0,1]}$ is progressively measurable for being a limit of progressively measurable process. Hence, we can apply the previous Lemma 3.10 to get the existence of progressively measurable process ζ such that $a(X_t(\omega), t)\zeta_t(\omega) = \xi_t(\omega)$ $\mathsf{Q} \otimes \mathrm{Leb}$-a.s (ω, t). But as explained above a solution to (3.12) is a solution to (3.11). Hence, we have $\sigma(X_t, t)^\top \zeta_t = \psi_t$ $\mathsf{Q} \otimes \mathrm{Leb}$-a.s. proving completeness of $L^2(\Omega \times [0, 1], a, \mathsf{Q} \otimes \mathrm{Leb})$. □

The following proposition tells us that stochastic integration of $(\zeta)_{t\in[0,1]}$ with respect to M^{Q} with $[M^{\mathsf{Q}}] = \int_0^\cdot a(X_s, s)\,\mathrm{d}s$ is completely determined by the values of $(a(X_t, t)\zeta_t)_{t\in[0,1]}$.

Proposition 3.12. *Let $\mathsf{Q} \in \mathcal{P}(\Omega)$ and let M^{Q} be a Q-local martingale such that $[M^{\mathsf{Q}}] = \int_0^\cdot a(X_s, s)\,\mathrm{d}s$. Consider two progressively measurable processes ζ and ξ for which $\int_0^\cdot \zeta_s\,\mathrm{d}M_s^{\mathsf{Q}}$ and $\int_0^\cdot \xi_s\,\mathrm{d}M_s^{\mathsf{Q}}$ are well-defined Q-local martingales. Then we have for a.e. $t \in [0, 1]$*

$$a(X_t, t)\zeta_t = a(X_t, t)\xi_t, \quad \mathsf{Q}\text{-}a.s.,$$

if and only if

$$\left(\int_0^t \zeta_s\,\mathrm{d}M_s^{\mathsf{Q}}\right)_{t\in[0,1]} = \left(\int_0^t \xi_s\,\mathrm{d}M_s^{\mathsf{Q}}\right)_{t\in[0,1]}, \quad \mathsf{Q}\text{-}a.s.$$

In particular, if $\zeta, \xi \in L^2(\Omega \times [0, 1], a, \mathsf{Q} \otimes \mathrm{Leb})$, then the previous assertions are equivalent to $\zeta = \xi$ in $L^2(\Omega \times [0, 1], a, \mathsf{Q} \otimes \mathrm{Leb})$.

Proof. Let $(T_n^\zeta)_{n\in\mathbb{N}}$ and $(T_n^\xi)_{n\in\mathbb{N}}$ be the localizing sequences for the local martingales $\int_0^\cdot \zeta_s\,\mathrm{d}M_s^{\mathsf{Q}}$ and $\int_0^\cdot \xi_s\,\mathrm{d}M_s^{\mathsf{Q}}$ respectively. Let $T_n := T_n^\zeta \wedge T_n^\xi$ and note that by BDG-inequality there exists two universal constants $c, C > 0$ such that (see Theorem 18.17 in [40])

$$\begin{aligned} c\mathbb{E}_{\mathsf{Q}}\left[\int_0^{t\wedge T_n} (\zeta_s - \xi_s)^\top a(X_s, s)(\zeta_s - \xi_s)\,\mathrm{d}s\right] &\le \mathbb{E}_{\mathsf{Q}}\left[\sup_{t\in[0,1]}\left|\int_0^{t\wedge T_n} (\zeta_s - \xi_s)\,\mathrm{d}M_s^{\mathsf{Q}}\right|^2\right] \\ &\le C\mathbb{E}_{\mathsf{Q}}\left[\int_0^{t\wedge T_n} (\zeta_s - \xi_s)^\top a(X_s, s)(\zeta_s - \xi_s)\,\mathrm{d}s\right]. \end{aligned}$$

Hence

$$\int_0^{t\wedge T_n} (\zeta_s - \xi_s)\,\mathrm{d}M_s^{\mathsf{Q}} = 0, \quad \mathsf{Q}\text{-a.s.},$$

if and only if

$$a(X_t, t)\zeta_t = a(X_t, t)\xi_t, \quad \mathsf{Q}\text{-a.s.}$$

By taking $n \to \infty$ with the fact that $T_n \to \infty$ Q-a.s. we get that $\int_0^\cdot \alpha_s\,\mathrm{d}M_s^{\mathsf{Q}}$ is a modification of $\int_0^\cdot \beta_s\,\mathrm{d}M_s^{\mathsf{Q}}$ if and only if $a(X_t, t)\zeta_t = a(X_t, t)\xi_t$ a.s. But by continuity of the stochastic integrals we get that those stochastic integrals are indistinguishable which is the intended claim.

The additional claim under the assumption $\zeta, \xi \in L^2(\Omega \times [0, 1], a, \mathsf{Q} \otimes \mathrm{Leb})$ follows by Monotone convergence theorem. □

Remark 3.13. We note that $\zeta = \xi$ in $L^2(\Omega \times [0,1], a, \mathsf{Q})$ does not mean $\zeta = \xi$ almost surely. The reason is that a may be singular which may result in the null space of $\|\cdot\|_{L^2(a,\mathsf{Q})}$ containing nonzero processes. However, the previous proposition is telling us that in such case the the stochastic integrals of ζ and ξ with respect to M^{Q} are the same anyways.

However, if $a(X_t, t)$ is non-singular $\mathsf{Q} \otimes$ Leb-a.s., then $\zeta = \xi$ in $L^2(\Omega \times [0,1], a, \mathsf{Q})$ if and only if $\zeta = \xi$ $\mathsf{Q} \otimes$ Leb a.s.

We have finally gathered all the pieces to state and prove the following theorem which tells us that the control β from Theorem 3.6 belonging to the minimizer of (MSP) can always be written in the form $\beta_s = u(X_s, s)$ for some Borel measurable function u.

Theorem 3.14 (Markovian drift). *Let $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ be a Markovian probability measure and β be the control obtained by Theorem 3.6. Then there exists a function $u \in L^2(\mathbb{R}^d \times [0,1], a, (X_t)_\# \mathsf{P} \otimes \mathrm{Leb})$ such that*

$$\beta = u(X_\cdot, \cdot), \qquad \text{in } L^2(\Omega \times [0,1], a, \mathsf{P} \otimes \mathrm{Leb}).$$

Proof. The existence of u is based on Riesz representation theorem. Let $\gamma_t^{\mathsf{P}} := (X_t)_\# \mathsf{P}$ and define the functional $\Upsilon : L^2(\mathbb{R}^d \times [0,1], a, \gamma^{\mathsf{P}})$ by

$$\Upsilon(\varphi) := \mathbb{E}_{\mathsf{P}} \left[\int_0^1 \beta_s^\top a(X_s, s) \varphi(X_s, s) \, \mathrm{d}s \right]. \tag{3.13}$$

Then we can check via Cauchy-Schwarz (and $|\sigma^\top \beta|^2 = \beta^\top a \beta$) that

$$|\Upsilon(\varphi)|^2 \leq \mathbb{E}_{\mathsf{P}} \left[\int_0^1 \beta_s^\top a(X_s, s) \beta_s \, \mathrm{d}s \right] \|\varphi\|^2_{L^2(a,\gamma^{\mathsf{P}})} = H(\mathsf{P}|\mathsf{R}) \|\varphi\|^2_{L^2(a,\gamma^{\mathsf{P}})}.$$

So by Riesz' representation theorem there exists $u \in L^2(\mathbb{R}^d \times [0,1], a, \gamma^{\mathsf{P}})$ such that for all $\varphi \in L^2(\mathbb{R}^d \times [0,1], a, \gamma^{\mathsf{P}})$

$$\Upsilon(\varphi) = (u, \varphi)_{L^2(a,\gamma^{\mathsf{P}})}. \tag{3.14}$$

Now that we have this $u \in L^2(a, \gamma^{\mathsf{P}})$ we show that P solves the martingale problem $\mathrm{MP}(\mathcal{L}_t + u^\top a \nabla, \mu_0)$. To that end, take an arbitrary function $f \in C_c^\infty(\mathbb{R}^d)$. By Itô formula and Theorem 3.6 we have

$$N_t^{\mathsf{P}} := \int_0^t \nabla f(X_s) \, \mathrm{d}M_s^{\mathsf{P}} = f(X_t) - f(X_0) - \int_0^t \left[\mathcal{L}_s + \beta_s^\top a(X_s, s) \nabla \right] f(X_s) \, \mathrm{d}s,$$

is a P-local martingale. The quadratic variation of N^{P} is

$$[N^{\mathsf{P}}]_1 = \int_0^1 \nabla f(X_s)^\top a(X_s, s) \nabla f(X_s) \, \mathrm{d}s.$$

Since a is locally bounded and ∇f is continuous with compact support $\nabla f^\top a \nabla f$ is uniformly bounded, i.e.

$$[N^{\mathsf{P}}]_1 \leq C.$$

By Corollary 17.8 in [22] gives that $(N^{\mathsf{P}})_{t\in[0,1]}$ is a P-martingale.

This is still not enough, because we have β_s instead of $u(X_s,s)$. Nevertheless, let us use the martingale property which implies that for any bounded measurable function $g:\mathbb{R}^d\to\mathbb{R}$ and $s\le t$

$$\mathbb{E}_{\mathsf{P}}[(N_t^{\mathsf{P}}-N_s^{\mathsf{P}})g(X_s)]=0,$$

which is

$$\mathbb{E}_{\mathsf{P}}\left[\left(f(X_t)-f(X_s)-\int_s^t\left[\mathcal{L}_r+\beta_r^\top a(X_r,r)\nabla\right]f(X_r)\,\mathrm{d}r\right)g(X_s)\right]=0.$$

One can check that each term separately is integrable with respect to P so that

$$\mathbb{E}_{\mathsf{P}}\left[\left(f(X_t)-f(X_s)-\int_s^t\mathcal{L}_rf(X_r)\,\mathrm{d}r\right)g(X_s)\right]=\mathbb{E}_{\mathsf{P}}\left[g(X_s)\int_s^t\beta_r^\top a(X_r,r)\nabla f(X_r)\,\mathrm{d}r\right].$$

For example, the integrability of the right-hand side is obtained via Cauchy-Schwarz and the finiteness of the relative entropy. We will focus on the right-hand side, because we want to get rid of β_r which we do by first applying Fubini to get

$$\mathbb{E}_{\mathsf{P}}\left[g(X_s)\int_s^t\beta_r^\top a(X_r,r)\nabla f(X_r)\,\mathrm{d}r\right]=\int_s^t\mathbb{E}_{\mathsf{P}}[g(X_s)\beta_r^\top a(X_r,r)\nabla f(X_r)]\,\mathrm{d}r.$$

Now we that know this we can apply a similar argument as in Proposition 3.9 to conclude that

$$\lim_{h\to0^+}\frac{1}{h}\mathbb{E}_{\mathsf{P}}\left[f(X_{t+h})-f(X_t)-\int_t^{t+h}b(X_s,s)\,\mathrm{d}s\mid X_t\right]=\beta_t^\top a(X_t,t)\nabla f(X_t),\qquad\text{in }L^2(\Omega,\mathsf{P}).$$

In particular there exists a Borel measurable function $G_r:\mathbb{R}^d\to\mathbb{R}$ such that for a.e. $r\in[0,1]$

$$\beta_r^\top a(X_r,r)\nabla f(X_r)=G_r(X_r),\quad\mathsf{P}\text{-a.s.}$$

In particular for a.e. $r\in[s,t]$ we have that $\beta_r^\top a(X_r,r)\nabla f(X_r)$ is $\sigma(X_r)$-measurable which together with the definition of conditional expectation gives

$$\mathbb{E}_{\mathsf{P}}[g(X_s)\beta_r^\top a(X_r,r)\nabla f(X_r)]=\mathbb{E}_{\mathsf{P}}[\mathbb{E}_{\mathsf{P}}[g(X_s)\mid X_r]\beta_r^\top a(X_r,r)\nabla f(X_r)].$$

There exists a measurable function $v:\mathbb{R}^d\to\mathbb{R}$ such that $\mathbb{E}_{\mathsf{P}}[g(X_s)\mid X_r]=v(X_r)$ so that after applying Fubini again we get

$$\mathbb{E}_{\mathsf{P}}\left[\left(f(X_t)-f(X_s)-\int_s^t\mathcal{L}_rf(X_r)\,\mathrm{d}r\right)g(X_s)\right]=\mathbb{E}_{\mathsf{P}}\left[\int_s^t v(X_r)\beta_r^\top a(X_r,r)\nabla f(X_r)\,\mathrm{d}r\right].$$

The definition of $\Upsilon(\varphi)$ in (3.13) and the representation in (3.14) for $\varphi(x,r)=v(x)\nabla f(x)\mathbb{1}_{[s,t]}(r)$ yields

$$\mathbb{E}_{\mathsf{P}}\left[g(X_s)\int_s^t\beta_r^\top a(X_r,r)\nabla f(X_r)\,\mathrm{d}r\right]=\mathbb{E}_{\mathsf{P}}\left[\int_s^t v(X_r)u(X_r,r)^\top a(X_r,r)\nabla f(X_r)\,\mathrm{d}r\right].$$

The same argument as before allows us to recover g and get

$$\mathbb{E}_{\mathsf{P}}\left[\left(f(X_t) - f(X_s) - \int_s^t \left[\mathcal{L}_r + u(X_r, r)^\top a(X_r, r)\nabla\right] f(X_r)\,\mathrm{d}r\right) g(X_s)\right] = 0.$$

This implies that the new process $\hat{N}^{\mathsf{P}}$ defined as follows:

$$\hat{N}_t^{\mathsf{P}} := f(X_t) - \int_0^t \left[\mathcal{L}_r + u(X_r, r)^\top a(X_r, r)\nabla\right] f(X_r)\,\mathrm{d}r,$$

satisfies

$$\mathbb{E}_{\mathsf{P}}\left[(\hat{N}_t^{\mathsf{P}} - \hat{N}_s^{\mathsf{P}})g(X_s)\right] = 0.$$

Integrability of $\hat{N}^{\mathsf{P}}$ can be shown with Cauchy-Schwarz, local boundedness of the coefficients of $\mathcal{L}_t$ and the fact that f and all its derivatives are continuous and have compact support. This shows by the definition of conditional expectation that

$$\mathbb{E}_{\mathsf{P}}\left[\hat{N}_t^{\mathsf{P}} - \hat{N}_s^{\mathsf{P}} \mid X_s\right] = 0.$$

Therefore since $\hat{N}_t^{\mathsf{P}} - \hat{N}_s^{\mathsf{P}}$ is $\sigma(X_r :\ r \in [s, 1])$-measurable we get by the Markov property that

$$\mathbb{E}_{\mathsf{P}}\left[\hat{N}_t^{\mathsf{P}} - \hat{N}_s^{\mathsf{P}} \mid \mathcal{F}_s\right] = \mathbb{E}_{\mathsf{P}}\left[\hat{N}_t^{\mathsf{P}} - \hat{N}_s^{\mathsf{P}} \mid X_s\right] = 0.$$

Hence $\hat{N}_t^{\mathsf{P}}$ is a martingale.

We conclude that P solves the martingale problem $\mathrm{MP}(\mathcal{L}_t + u^\top a\nabla, \mu_0)$ and the uniqueness of semimartingale decomposition together with Lebesgue differentiation shows that for a.e. $t \in [0, 1]$ $a(X_t, t)\beta_t = a(X_t, t)u(X_t, t)$ P-a.s. By Proposition 3.12 we get the assertion that $\beta = u(X_\cdot, \cdot)$ in $L^2(\Omega \times [0, 1], a, \gamma^{\mathsf{P}} \otimes \mathrm{Leb})$ which is what we were after. □

The previous theorem tells us that the canonical process X under the minimizer P has the semimartingale decomposition $X = X_0 + A^{\mathsf{P}} + M^{\mathsf{P}}$ with

$$A_t^{\mathsf{P}} = \int_0^t b(X_s, s) + a(X_s, s)u(X_s, s)\,\mathrm{d}s, \quad M_t^{\mathsf{P}} = M_t - \int_0^t a(X_s, s)u(X_s, s)\,\mathrm{d}s. \tag{3.15}$$

Let us set $\gamma_t^{\mathsf{P}} = (X_t)_\# \mathsf{P}$ as in the previous theorem. The multimarginal Schrödinger problem (MSP) seems to boil down to minimizing

$$(\mu_0, u) \mapsto H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\gamma^{\mathsf{P}})}, \tag{3.16}$$

over pairs $(\mu_0, u) \in \mathcal{P}(\mathbb{R}^d) \times L^2(\Omega \times [0, 1], a, \gamma^{\mathsf{P}} \otimes \mathrm{Leb})$ such that X under P has the semimartingale decomposition given in (3.15) with the constraint $\gamma_t^{\mathsf{P}} = \mu_t$ for all $t \in \mathcal{T}$.

However, it is not completely clear whether this in fact leads to an equivalent problem. For example, it is not clear that X under P having the semimartingale decomposition given in (3.15) implies that $H(\mathsf{P}|\mathsf{R}) < \infty$. Fortunately, it turns out to be the case as follows from the same techniques used in Chapter 5. We do not get into such details as we believe that this is more straightforward than what we will prove in Chapter 5.

3.2 Approximation of the Schrödinger problem

It can be sometimes useful to reduce the constraints and consider "smaller" problems. For instance, when we have $|\mathcal{T}| = \infty$, then we may ask if we can approximate such problem with a Schrödinger problem with finitely many constraints. That is possible, and in fact, there are many ways to do so. However, we illustrate one particular way to do it which is also used in [14] to treat the case $\mathcal{T} = [0,1]$ and the reference measure being the law of a Brownian motion.

To that end, we start by arguing that we can always assume $\mathcal{T} \subset [0,1]$ is countable. The marginal law $t \mapsto (X_t)_\# \mathsf{P}$ is continuous under the topology of weak convergence. Indeed, by dominated convergence theorem we have for every bounded continuous $f \in C_b(\mathbb{R}^d)$

$$\lim_{s\to t} \int_{\mathbb{R}^d} f \,\mathrm{d}(X_s)_\# \mathsf{P} = \lim_{s\to t} \mathbb{E}_{\mathsf{P}}[f(X_s)] = \mathbb{E}_{\mathsf{P}}[f(X_t)] = \int_{\mathbb{R}^d} f \,\mathrm{d}(X_t)_\# \mathsf{P}.$$

Therefore, for the problem to have a solution, we must have that the mapping $t \mapsto \mu_t$ with domain $\mathcal{T}$ is continuously extendable on $\overline{\mathcal{T}}$. Once that is established, we see that μ_t is uniquely defined by a dense subset of $\mathcal{T}$ due to continuity. But $[0,1]$ is a separable metric space and so is every subset $\mathcal{T} \subset [0,1]$ which means we can find a countable dense subset of $\mathcal{T}' \subset \mathcal{T}$ (see Proposition 3.25 in [5]).

That leads us to define the following set of probability measures:

$$\mathcal{C}^{(n)} := \{\mathsf{P} \in \mathcal{P}(\Omega) : \ \mathbb{E}_{\mathsf{P}}[\mathbb{1}_{A_i}(X_t)] = \mu_t(A_i), \quad \text{for all } i \in \{1,...,n\},\ t \in \mathcal{T}_n\}, \tag{3.17}$$

for some closed sets $(A_i)_{i\in\mathbb{N}} \subset \mathcal{B}(\mathbb{R}^d)$ generating $\mathcal{B}(\mathbb{R}^d)$; and a collection of increasing finite sets $(\mathcal{T}_n)_{n\in\mathbb{N}}$ such that

$$\mathcal{T}' = \bigcup_{n\in\mathbb{N}} \mathcal{T}_n.$$

We obviously have $\mathcal{C}^{\mathrm{MSP}} \subset \mathcal{C}^{(n+1)} \subset \mathcal{C}^{(n)}$ for all $n \in \mathbb{N}$ and by uniqueness of measure we get

$$\mathcal{C}^{\mathrm{MSP}} = \bigcap_{n\in\mathbb{N}} \mathcal{C}^{(n)}.$$

We let $\mathcal{C}_H^{(n)} \subset \mathcal{C}^{(n)}$ to be the subset of measures with finite entropy with respect to R.

Now we can define the approximate multimarginal Schrödinger problem:

Approximate multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mathsf{P}|\mathsf{R}), \\ \text{subject to} \quad & \mathsf{P} \in \mathcal{C}^{(n)}. \end{aligned} \tag{AMSP^n}$$

Theorem 3.15 (Convergence of the approximation)**.** *Assume* (MSP) *has a unique minimizer* $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$. *Then for all* $n \in \mathbb{N}$ (AMSP^n) *has a unique minimizer* $\mathsf{P}_n \in \mathcal{C}_H^{(n)}$ *such that*

$$\mathsf{P}_n \to \mathsf{P} \qquad \textit{weakly as } n \to \infty,$$

and

$$\lim_{n\to\infty} H(\mathsf{P}_n|\mathsf{R}) = H(\mathsf{P}|\mathsf{R}).$$

Proof. We know that the sublevel set

$$\mathcal{C} = \{\mathsf{Q} \in \mathcal{P}(\Omega) : H(\mathsf{Q}|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R})\}$$

is a compact subset of $\mathcal{P}(\Omega)$ under the narrow topology (i.e. the topology of weak convergence) (Proposition B.3 in Appendix B). The fact that $\mathsf{P} \in \mathcal{C}^{\mathrm{MSP}} \subset \mathcal{C}^{(n)}$ is feasible for (AMSP^n) implies there exists a unique minimizer $\mathsf{P}_n \in \mathcal{C}$ for (AMSP^n) by Theorem 2.1. Since $\mathsf{P}_n \in \mathcal{C}$ we have that every subsequence (n_k) of P_n has a convergent subsequence (n_{k_j}), say $\mathsf{P}_{n_{k_j}} \rightharpoonup \tilde{\mathsf{P}}$. By lower semi-continuity of the relative entropy (Proposition B.3 in Appendix B), we have

$$H(\tilde{\mathsf{P}} \mid \mathsf{R}) \leq \liminf_{j\to\infty} H(\mathsf{P}_{n_{k_j}}|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R}).$$

As soon as we show that $\tilde{\mathsf{P}} \in \mathcal{C}^{\mathrm{MSP}}$ we get the claim by uniqueness of the minimizer P. Fix $i \in \mathbb{N}$ and $t \in \mathcal{T}'$. By Lemma 13.10 in [23] we get the existence of a continuous function $f_m \in C_b(\mathbb{R}^d)$ such that $0 \leq f_m \leq 1$ and $f_m \to \mathbb{1}_{A_i}$ everywhere, as $m \to \infty$. Take an arbitrary $\varepsilon > 0$; and note that there exists $m^* \in \mathbb{N}$ such that for all $m \geq m^*$, together with dominated convergence, we have

$$\int_\Omega \mathbb{1}_{A_i}(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_j}} - \varepsilon \leq \int_\Omega f_m(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_j}} \leq \int_\Omega \mathbb{1}_{A_i}(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_j}} + \varepsilon.$$

There exists j^* such that $i \leq n_{k_{j^*}}$ and $t \in \mathcal{T}_{n_{k_{j^*}}}$ which means that for all $j \geq j^*$ we have

$$\int_\Omega \mathbb{1}_{A_i}(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_j}} = \int_\Omega \mathbb{1}_{A_i}(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_{j^*}}} = \mu_t(A_i),$$

and in particular

$$\mu_t(A_i) - \varepsilon \leq \int_\Omega f_m(X_t)\,\mathrm{d}\mathsf{P}_{n_{k_j}} \leq \mu_t(A_i) + \varepsilon.$$

Take $j \to \infty$ to conclude by the fact that $\mathsf{P}_{n_{k_j}} \rightharpoonup \tilde{\mathsf{P}}$

$$\mu_t(A_i) - \varepsilon \leq \int_\Omega f_m(X_t)\,\mathrm{d}\tilde{\mathsf{P}} \leq \mu_t(A_i) + \varepsilon.$$

Now take $m \to \infty$ to obtain

$$\mu_t(A_i) - \varepsilon \leq \int_\Omega \mathbb{1}_{A_i}(X_t)\,\mathrm{d}\tilde{\mathsf{P}} \leq \mu_t(A_i) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary we conclude that

$$\mathbb{E}_{\tilde{\mathsf{P}}}[\mathbb{1}_{A_i}(X_t)] = \mu_t(A_i).$$

This holds for arbitrary $i \in \mathbb{N}$ and $t \in \mathcal{T}'$ which yields $\tilde{\mathsf{P}} \in \mathcal{C}^{\mathrm{MSP}}$. Hence $\tilde{\mathsf{P}} = \mathsf{P}$ by the uniqueness of P.

We have established that any subsequence of P_n has a subsequence converging to P so that the assertion $\mathsf{P}_n \to \mathsf{P}$ weakly follows from a standard result in analysis. Finally, the last part is mostly done already, lower semi-continuity and $H(\mathsf{P}_n|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R})$ implies

$$H(\mathsf{P}|\mathsf{R}) \leq \liminf_{n\to\infty} H(\mathsf{P}_n|\mathsf{R}) \leq \limsup_{n\to\infty} H(\mathsf{P}_n|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R}).$$

□

We can prove a stronger statement for which we first need the following result concerning the uniqueness of the control from Theorem 3.6:

Proposition 3.16 (Uniqueness of control)**.** *Assume that* $\mathsf{Q} \in \mathcal{P}(\Omega)$ *satisfies* $H(\mathsf{Q}|\mathsf{R}) < \infty$. *If*

$$\frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}}>0\}} \frac{\mathrm{d}(X_0)_\#\mathsf{Q}}{\mathrm{d}(X_0)_\#\mathsf{R}}(X_0) \exp\left(\int_0^1 \zeta_s \,\mathrm{d}M_s - \frac{1}{2}\int_0^1 |\sigma(X_s,s)^\top \zeta_s|^2 \,\mathrm{d}s\right), \quad \mathsf{R}\text{-}a.s.,$$

for some progressively measurable process ζ *such that the semimartingale* $\int_0^t \zeta_s \,\mathrm{d}M_s$ *is well-defined* Q*-a.s. Then* $\zeta \in L^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb})$ *and* $\zeta = \beta^{\mathsf{Q}}$ *in* $L^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb})$ *where* β^{Q} *is the control from Theorem 3.6. In particular*

$$H(\mathsf{Q}|\mathsf{R}) \geq H((X_0)_\#\mathsf{Q}|(X_0)_\#\mathsf{R}) + \frac{1}{2}\mathbb{E}_{\mathsf{Q}}\left[\int_0^1 |\sigma(X_s,s)^\top \zeta_s|^2 \,\mathrm{d}s\right],$$

with equality if $\mathrm{MP}(\mathcal{L}_t)$ *satisfies the semi-uniqueness condition.*

Proof. Define the stopping time

$$T_n := \inf\left\{t \in [0,1] : \int_0^t |\sigma(X_s,s)^\top \zeta_s|^2 \,\mathrm{d}s \geq n\right\} \wedge 1,$$

and consider $\mathrm{d}\mathsf{Q}/\mathrm{d}\mathsf{R}$ on $\mathcal{F}_{t\wedge T_n}$. It is not difficult to check that

$$\frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}}\Big|_{\mathcal{F}_{t\wedge T_n}} = \frac{\mathrm{d}(X_0)_\#\mathsf{Q}}{\mathrm{d}(X_0)_\#\mathsf{R}}(X_0) \exp\left(\int_0^{t\wedge T_n} \zeta_s \,\mathrm{d}M_s - \frac{1}{2}\int_0^{t\wedge T_n} |\sigma(X_s,s)^\top \zeta_s|^2 \,\mathrm{d}s\right), \quad \mathsf{R}\text{-a.s.}$$

Then we apply classical Girsanov's theorem and take $n \to \infty$ (and the fact that the semimartingale $\int \zeta \,\mathrm{d}M$ is well-defined implies that $T_n \to \infty$ Q-a.s.) to get that the semimartingale decomposition of X under Q is $X = X_0 + A^{\mathsf{Q}} + M^{\mathsf{Q}}$ with finite variation part

$$A_t^{\mathsf{Q}} = \int_0^t (b(X_s,s) + a(X_s,s)\zeta_s) \,\mathrm{d}s.$$

By the uniqueness of semimartingale decomposition and Theorem 3.6 we get

$$\int_0^t (b(X_s,s) + a(X_s,s)\zeta_s) \,\mathrm{d}s = \int_0^t \left(b(X_s,s) + a(X_s,s)\beta_s^{\mathsf{Q}}\right) \mathrm{d}s.$$

By Lebesgue differentiation we get that for a.e. $t \in [0,1]$

$$a(X_t,t)\zeta_t = a(X_t,t)\beta_t^{\mathsf{Q}}, \quad \mathsf{Q}\text{-a.s.},$$

which proves the result after invoking Theorem 3.6 and Proposition 3.12. □

We can now prove the following statement:

Theorem 3.17. *Assume in addition to the assumptions in Theorem 3.15, that* $\mathrm{MP}(\mathcal{L}_t)$ *satisfies the semi-uniqueness condition. Let* $\beta^{(n)}$ *and* β *be the control from Theorem 3.6 corresponding to* P_n *and* P *respectively. Then* $\beta^{(n)} \to \beta$ *in* $L^2(\Omega \times [0,1], a, \mathsf{P} \otimes \mathrm{Leb})$*, i.e.*

$$\lim_{n\to\infty} \mathbb{E}_{\mathsf{P}} \left[\int_0^1 |\sigma(X_s, s)^\top (\beta_s - \beta_s^{(n)})|^2 \, \mathrm{d}s \right] = 0.$$

Proof. We know under the semi-uniqueness of $\mathrm{MP}(\mathcal{L}_t)$, there exists processes β and $\beta^{(n)}$ by Theorem 3.6 such that

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}} \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0) \exp\left(\int_0^1 \beta_s \, \mathrm{d}M_s - \frac{1}{2} \int_0^1 |\sigma(X_s, s)^\top \beta_s|^2 \, \mathrm{d}s \right),$$

and

$$\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}}>0\}} \frac{\mathrm{d}\mu_0^{(n)}}{\mathrm{d}\nu_0}(X_0) \exp\left(\int_0^1 \beta_s^{(n)} \, \mathrm{d}M_s - \frac{1}{2} \int_0^1 |\sigma(X_s, s)^\top \beta_s^{(n)}|^2 \, \mathrm{d}s \right).$$

Note that P is feasible for (AMSP^n) so that by Theorem 2.2 the following inequality holds:

$$H(\mathsf{P}|\mathsf{R}) \geq H(\mathsf{P}|\mathsf{P}_n) + H(\mathsf{P}_n|\mathsf{R}). \tag{3.18}$$

In particular $\mathsf{P} \ll \mathsf{P}_n$ (because $H(\mathsf{P}|\mathsf{P}_n) < \infty$) and we can write the Radon-Nikodym derivative as follows:

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_n} = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} \left(\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}} \right)^{-1} \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}}>0\}}, \quad \mathsf{P}_n\text{-a.s.}$$

Writing everything in terms of the stochastic exponential $\mathcal{E}(\cdot)$ (i.e. $\mathcal{E}(L) := \exp(L - [L]/2)$ for any local martingale L) yields

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_n} = \frac{\frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0)}{\frac{\mathrm{d}\mu_0^{(n)}}{\mathrm{d}\nu_0}(X_0)} \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\} \cap \{\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}}>0\}} \mathcal{E}\left(\int^{\cdot} (\beta_s - \beta_s^{(n)}) \, \mathrm{d}M_s \right)_1.$$

Note that $\mathrm{d}\mathsf{P}_n/\mathrm{d}\mathsf{R}$ is finite R-a.s., because it is integrable with respect to R. Due to $\mathsf{P}_n \ll \mathsf{R}$ it is finite P_n-a.s. too. These facts combined we see that

$$\mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\} \cap \{\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}}>0\}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_n}>0\}} \quad \mathsf{P}_n\text{-a.s.}$$

Therefore we can write

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_n} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{P}_n}>0\}} \frac{\frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0)}{\frac{\mathrm{d}\mu_0^{(n)}}{\mathrm{d}\nu_0}(X_0)} \mathcal{E}\left(\int^{\cdot} (\beta_s - \beta_s^{(n)}) \, \mathrm{d}M_s \right)_1.$$

By uniqueness of the process $(\beta - \beta^{(n)})$ due to Proposition 3.16 applied to $\mathrm{d}\mathsf{P}_n/\mathrm{d}\mathsf{P}$ yields that

$$H(\mathsf{P}|\mathsf{P}_n) \geq H(\mu_0|\mu_0^{(n)}) + \frac{1}{2} \mathbb{E}_{\mathsf{P}} \left[\int_0^1 |\sigma(X_s, s)^\top (\beta_s - \beta_s^{(n)})|^2 \, \mathrm{d}s \right].$$

Inequality (3.18) implies

$$H(\mathsf{P}|\mathsf{R}) - H(\mathsf{P}_n|\mathsf{R}) \geq H(\mu_0|\mu_0^{(n)}) + \frac{1}{2}\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s, s)^\top(\beta_s - \beta_s^{(n)})|^2 \, \mathrm{d}s\right].$$

Since $H(\mathsf{P}_n|\mathsf{R}) \to H(\mathsf{P}|\mathsf{R})$ by Theorem 3.15 we get after taking $\limsup$

$$\limsup_{n\to\infty} H(\mu_0|\mu_0^{(n)}) + \frac{1}{2}\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_s, s)^\top(\beta_s - \beta_s^{(n)})|^2 \, \mathrm{d}s\right] = 0,$$

which gives the desired claim since the relative entropy is non-negative. □

3.3 Reversible processes

We give a short discussion on how sometimes reversing the time in our problem (MSP) might be of help for reversible processes. We first recall the definition of a reversible Markov process, namely

Definition 3.18 (Reversible Markov process)**.** A Markov process $(X_t)_{t\in[0,1]}$ is said to be *reversible* under $\mathsf{Q} \in \mathcal{P}(\Omega)$ if for every $t \in [0,1]$

$$(X_s)_{s\in[0,t]} \sim (X_{t-s})_{s\in[0,t]} \quad \text{under } \mathsf{Q}.$$

Moreover, we say a Markovian measure $\mathsf{Q} \in \mathcal{P}(\Omega)$ is *reversible*, if the canonical process X is a reversible Markov process under Q.

We assume that the reference measure $\mathsf{R} \in \mathcal{P}(\Omega)$ is reversible. Let us define the reversing map $\Phi : \Omega \to \Omega$ by

$$\Phi(\omega)(t) := \omega(1-t).$$

The measurability of the map Φ is immediate, and that makes $\mathsf{P} \circ \Phi$ defined by

$$\mathsf{P} \circ \Phi(\cdot) = \mathsf{P}(\Phi(\cdot)),$$

a well-defined probability measure for any $\mathsf{P} \in \mathcal{P}(\Omega)$. Moreover, Φ is bijective which implies

$$H(\mathsf{P} \circ \Phi|\mathsf{R} \circ \Phi) = H(\mathsf{P}|\mathsf{R}),$$

for any probability measures $\mathsf{P}, \mathsf{R} \in \mathcal{P}(\Omega)$ by (Corollary B.7 in Appendix B). We also know that $\mathsf{R} = \mathsf{R} \circ \Phi$ due to reversibility. Therefore, for this special case we can consider the time-reversed equivalent version of problem (MSP), namely

Time-reversed multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mathsf{P}|\mathsf{R}), \\ \text{subject to} \quad & (X_t)_\# \mathsf{P} \circ \Phi = \mu_t, \quad \text{for all } t \in \mathcal{T}. \end{aligned} \tag{TRMSP}$$

We define $\mathcal{C}^{\mathrm{TRMSP}} \subset \mathcal{P}(\Omega)$ for the set of all probability measures such that $\mathsf{P} \circ \Phi \in \mathcal{C}^{\mathrm{MSP}}$, i.e.

$$\mathcal{C}^{\mathrm{TRMSP}} := \{\mathsf{P} \in \mathcal{P}(\Omega) : \ (X_t)_{\#}\mathsf{P} \circ \Phi = \mu_t, \quad \text{for all } t \in \mathcal{T}\},$$

and we similarly define $\mathcal{C}_H^{\mathrm{TRMSP}} \subset \mathcal{C}^{\mathrm{TRMSP}}$ for the competitors in $\mathcal{C}^{\mathrm{TRMSP}}$. In general $\mathcal{C}^{\mathrm{MSP}} \neq \mathcal{C}^{\mathrm{TRMSP}}$, but if $\mathcal{C}^{\mathrm{MSP}}$ is symmetric in the sense that $t \in \mathcal{T}$ if and only if $1-t \in \mathcal{T}$ and $\mu_t = \mu_{1-t}$, then we have $\mathcal{C}^{\mathrm{MSP}} = \mathcal{C}^{\mathrm{TRMSP}}$.

This seemingly trivial reformulation of the problem can be used as a consistency criterion in the following sense:

Lemma 3.19 (Consistency under reversibility)**.** *Assume the reference probability measure* $\mathsf{R} \in \mathcal{P}(\Omega)$ *is reversible. A probability measure* $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ *is the minimizer of* (MSP) *if and only if* $\mathsf{P} \circ \Phi \in \mathcal{C}_H^{\mathrm{TRMSP}}$ *is the unique minimizer of* (TRMSP)*. In particular, if* $\mathcal{C}^{\mathrm{MSP}} = \mathcal{C}^{\mathrm{TRMSP}}$*, then the minimizers of each problems* (MSP) *and* (TRMSP) *coincide, i.e.* $\mathsf{P} = \mathsf{P} \circ \Phi$*.*

Proof. It follows from the fact that and $\Phi : \Omega \to \Omega$ is a bijection and (Corollary B.7 in Appendix B) which gives for any probability measure $\mathsf{Q} \in \mathcal{P}(\Omega)$

$$H(\mathsf{Q}|\mathsf{R}) = H(\mathsf{Q} \circ \Phi|\mathsf{R} \circ \Phi) = H(\mathsf{Q} \circ \Phi|\mathsf{R}),$$

where we have used the reversibility of R in the last equality. To prove the direction "$\Longrightarrow$", assume $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ is the minimizer of (MSP). Then for any $\mathsf{Q} \in \mathcal{C}_H^{\mathrm{TRMSP}}$ we have $\mathsf{Q} \circ \Phi \in \mathcal{C}_H^{\mathrm{MSP}}$ and

$$H(\mathsf{Q}|\mathsf{R}) = H(\mathsf{Q} \circ \Phi|\mathsf{R}) \geq H(\mathsf{P}|\mathsf{R}) = H(\mathsf{P} \circ \Phi|\mathsf{R}).$$

The reverse direction is proved similarly. Finally, the same reasoning yields $\mathsf{P} = \mathsf{P} \circ \Phi$ for the case $\mathcal{C}^{\mathrm{MSP}} = \mathcal{C}^{\mathrm{TRMSP}}$. $\square$

Time-reversibility can be useful to prove results by a "symmetry argument" as we will see in Section 4.2 and Section 4.4. As such when the reference measure R is reversible, the Schrödinger problem is easier to handle.

Finally, we end this chapter by noting that most results in this chapter extend to more general diffusion processes. While Proposition 3.4 about the minimizer P being Markovian only the continuity of the canonical process X was used, for Girsanov's theorem and the characterization of the control β we have used the semimartingale decomposition of X. Most obvious extensions of these theorems include general continuous Markovian semimartingales as it is actually proved in [25]. In such case Proposition 3.9 will take a different form where the division of h is replaced by the measure induced by the quadratic variation of the local martingale part of X.

We will see a lot of occasions where R being the law of an Itô-diffusion is not needed. It is not difficult to distinguish these results from the rest and we will omit such discussions after this chapter.

Chapter 4

Finitely many marginal constraints

We study the multimarginal Schrödinger problem (MSP) for the case $\mathcal{T}$ is finite. We relate the problem to a static version, show existence of solutions, and characterize the solutions. Furthermore, we give regularity results with applications to reversible Langevin diffusions. At the end of this chapter, in Section 4.4, we apply the obtained results for Ornstein-Uhlenbeck process.

As we will see below the problem (MSP) is equivalent to a seemingly smaller problem. More precisely, let $\mathcal{T} = \{t_1, ..., t_k\}$ and define μ and ν as the joint law of $(X_t)_{t\in\mathcal{T}}$ under P and R respectively, i.e.

$$\mu := (X_{t_1}, ..., X_{t_k})_{\#}\mathsf{P}, \quad \nu := (X_{t_1}, ..., X_{t_k})_{\#}\mathsf{R}.$$

Let π_{t_j} the projection to the t_j-th component. The law of the t_j-th component under μ is $(\pi_{t_j})_{\#}\mu$. Now consider a new minimization problem, namely

Static multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & H(\gamma|\nu), \\ \text{subject to} \quad & (\pi_t)_{\#}\gamma = \mu_t, \qquad \text{for all } t \in \mathcal{T}. \end{aligned} \tag{SMSP}$$

We call this problem the static version of (MSP). The name makes sense, because while in the original problem (MSP) we look at the whole path measure, in (SMSP) we only look at the different points where marginals are given.

We use the notations $\mathcal{C}^{\text{SMSP}}$ and $\mathcal{C}_H^{\text{SMSP}}$ for the set of probability measures in $\mathcal{P}((\mathbb{R}^d)^k)$ that are feasible for (SMSP) and competitors respectively. Using a similar argument as used in the proof Theorem 3.2 we can show that $\mathcal{C}^{\text{SMSP}} \subset \mathcal{P}((\mathbb{R}^d)^k)$ is convex and closed so that by Theorem 2.1 we get the following:

Theorem 4.1 (Existence of minimizer to (SMSP))**.** *Assume there exists a competitor $\gamma \in \mathcal{C}_H^{\text{SMSP}}$, then there exists a unique solution for the minimization problem* (SMSP).

The following lemma tells us that problems (MSP) and (SMSP) are equivalent, so it seems that we only have to solve a seemingly smaller problem.

Lemma 4.2 (Equivalence of (MSP) to (SMSP))**.** *A unique solution* $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ *to* (MSP) *exists if and only if a unique solution* $\mu^* \in \mathcal{C}_H^{\mathrm{SMSP}}$ *to* (SMSP) *exists. In either case, the minimizer* P *of* (MSP) *is characterized by*

$$\mathbb{E}_{\mathsf{P}}[Y] = \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[Y]\, \mu^*(\mathrm{d}x), \qquad \textit{for all non-negative random variables } Y,$$

with R^x *being the regular conditional probability given* $(X_t)_{t\in\mathcal{T}}$*. Moreover, the objective values of each problem is equal, i.e.*

$$H(\mathsf{P}|\mathsf{R}) = H(\mu^*|\nu).$$

Proof. We start with the "if"-part, so assume a solution $\mu^* \in \mathcal{C}_H^{\mathrm{SMSP}}$ to (SMSP) exists. We prove the assertion by defining a measure $\mathsf{Q} \in \mathcal{P}(\Omega)$ that is feasible for (MSP) and satisfies $H(\mu^*|\nu) = H(\mathsf{Q}|\mathsf{R})$. We define Q, by the characterization mentioned for the minimizer, namely for any non-negative random variable Y we set

$$\mathbb{E}_{\mathsf{Q}}[Y] := \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[Y]\, \mu^*(\mathrm{d}x). \tag{4.1}$$

This is well-defined, since Ω is a Polish space we have the existence of regular conditional probability R^x given $X_\mathcal{T} := (X_t)_{t\in\mathcal{T}} = x \in (\mathbb{R}^d)^k$.

For the feasibility of Q, we calculate for any $B \in \mathcal{B}((\mathbb{R}^d)^k)$

$$(X_\mathcal{T})_\#\mathsf{Q}(B) = \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[\mathbb{1}_B(X_\mathcal{T})]\, \mu^*(\mathrm{d}x) = \int_{(\mathbb{R}^d)^k} \mathbb{1}_B(x)\, \mu^*(\mathrm{d}x) = \mu^*(B).$$

Hence Q satisfies the marginal constraints, because μ^* is the solution to (SMSP). It only remains to show that $H(\mathsf{Q}|\mathsf{R}) = H(\mu^*|\nu)$. To that end, note that for any set $A \in \mathcal{F}$ we have using the definition of Q and the Radon-Nikodym theorem

$$\mathbb{E}_{\mathsf{Q}}[\mathbb{1}_A] = \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[\mathbb{1}_A]\, \mu^*(\mathrm{d}x) = \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[\mathbb{1}_A] \frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(x)\, \nu(\mathrm{d}x)$$

Using the definition of regular conditional probability and conditional expectation we get

$$\mathbb{E}_{\mathsf{Q}}[\mathbb{1}_A] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{E}_{\mathsf{R}}[\mathbb{1}_A \mid X_\mathcal{T}] \frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(X_\mathcal{T})\right] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(X_\mathcal{T})\right].$$

By uniqueness of the Radon-Nikodym derivative

$$\frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(X_\mathcal{T}), \quad \mathsf{R}\text{-a.s. and } \mathsf{Q}\text{-a.s.}$$

We are almost done for we can write

$$H(\mathsf{Q}|\mathsf{R}) = \int_\Omega \log\left(\frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(X_\mathcal{T})\right) \mathrm{d}\mathsf{Q} = \int_{(\mathbb{R}^d)^k} \log\left(\frac{\mathrm{d}\mu^*}{\mathrm{d}\nu}(x)\right) \mu^*(\mathrm{d}x) = H(\mu^*|\nu) < \infty.$$

Now $\mathsf{Q} \in \mathcal{C}_H^{\mathrm{MSP}}$ so a solution $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ to (MSP) exists by Theorem 3.2 and $H(\mathsf{P}|\mathsf{R}) \leq H(\mathsf{Q}|\mathsf{R})$. This establishes, the "if"-part.

For the "only-if"-part, we take the minimizer $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ of (MSP) and consider the measure $\mu = (X_{\mathcal{T}})_{\#}\mathsf{P}$. This measure is obviously feasible for (SMSP), i.e. $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$. Let P^x denote the regular conditional probability $\mathsf{P}(\cdot \mid X_{\mathcal{T}} = x)$. Then by conditioning (Lemma B.4 in Appendix B) we have

$$H(\mathsf{P}|\mathsf{R}) = H(\mu|\nu) + \int_{(\mathbb{R}^d)^k} H(\mathsf{P}^x|\mathsf{R}^x)\,\mu(\mathrm{d}x).$$

We also know that $H(\mathsf{P}^x|\mathsf{R}^x) \geq 0$ which gives

$$H(\mathsf{P}|\mathsf{R}) \geq H(\mu|\nu).$$

In particular $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$ and by the same argument as in Theorem 3.2 and Theorem 2.1 we get the existence of a solution $\mu^* \in \mathcal{C}_H^{\mathrm{SMSP}}$ to (SMSP). We can now define the measure $\mathsf{Q} \in \mathcal{C}_H^{\mathrm{MSP}}$ from the previous part using μ^* as done in (4.1) to get

$$H(\mathsf{P}|\mathsf{R}) \geq H(\mathsf{Q}|\mathsf{R}).$$

But we know that P is the unique minimizer of (MSP) so $\mathsf{P} = \mathsf{Q}$. By definition of Q we have

$$\mathbb{E}_{\mathsf{P}}[Y] = \int_{(\mathbb{R}^d)^k} \mathbb{E}_{\mathsf{R}^x}[Y]\,\mu^*(\mathrm{d}x), \quad \text{for all non-negative random variables Y,}$$

and

$$H(\mathsf{P}|\mathsf{R}) = H(\mu^*|\nu),$$

which is what we were after. $\square$

By the previous lemma, to solve (MSP), it is enough to find the minimizer μ of $H(\mu|\nu)$ with the specified marginals $(\mu_t)_{t\in\mathcal{T}}$. Since $(\mathbb{R}^d)^k$ is a product space, we can apply the existence result obtained in Section 2.1. We summarize it in the following theorem:

Theorem 4.3 (Existence of minimizer for (SMSP))**.** *Assume that*

$$H\left(\otimes_{i=1}^k \mu_{t_i}|\nu\right) < \infty,$$

and

$$\nu \ll \otimes_{i=1}^k \nu_{t_i}.$$

Then there exists a unique minimizer $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$ and there exists a collection of measurable functions $f_i : \mathbb{R}^d \to \mathbb{R}$ indexed by $i \in \{1, ..., k\}$ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \prod_{i=1}^k f_i(x_i), \quad \nu\text{-a.s.} \tag{4.2}$$

Moreover, the factorization is unique up to transformations $(f_i)_{1\leq i\leq k} \mapsto (c_i f_i)_{1\leq i\leq k}$ for some $(c_i)_{1\leq i\leq k} \subset (0,\infty)$ satisfying $\prod_{i=1}^k c_i = 1$. Additionally, if $\otimes_{i=1}^k \mu_{t_i} \sim \nu$, then $\mu \sim \nu$. Finally, we call the functions $(f_i)_{1\leq i\leq k}$ the Schrödinger factorization.

Proof. Note that $\otimes_{i=1}^k \mu_{t_i}$ is feasible and the fact that it has finite relative entropy implies it is a competitor. So a unique minimizer $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$ exists by Theorem 4.1.

We now want to apply Theorem 2.5, but first we need to show $\mu \sim \otimes_{i=1}^k \mu_{t_i}$. The fact that $\otimes_{i=1}^k \mu_{t_i}$ is a competitor implies $H(\otimes_{i=1}^k \mu_{t_i}|\mu) < \infty$ by Theorem 2.2, and in particular $\otimes_{i=1}^k \mu_{t_i} \ll \mu$. It only remains to show $\mu \ll \otimes_{i=1}^k \mu_{t_i}$. It happens that the condition $\nu \ll \otimes_{i=1}^k \nu_{t_i}$ implies $\mu \ll \otimes_{i=1}^k \mu_{t_i}$. It is based on the fact that the set

$$B := \left\{ x \in (\mathbb{R}^d)^k : \prod_{i=1}^k \frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}}(x) > 0 \right\},$$

has full measure under μ, because

$$\mu(B^c) = \mu\left(\prod_{i=1}^k \frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}} = 0\right) \le \sum_{i=1}^k \mu\left(\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}}(\pi_i) = 0\right) = \sum_{i=1}^k \mu_{t_i}\left(\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}} = 0\right) = 0.$$

Now let us take $A \in \mathcal{B}((\mathbb{R}^d)^k)$ such that $\otimes_{i=1}^k \mu_{t_i}(A) = 0$. Since $\otimes_{i=1}^k \mu_{t_i} \ll \otimes_{i=1}^k \nu_{t_i}$ (because $\mu_{t_i} \ll \nu_{t_i}$ otherwise $\mu \notin \mathcal{C}_H^{\mathrm{SMSP}}$) we have

$$0 = \int_A \prod_{i=1}^k \frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}} \,\mathrm{d}\nu_{t_1} \ldots \mathrm{d}\nu_{t_k},$$

implying $\otimes_{i=1}^k \nu_{t_i}(A \cap B) = 0$. We also know $\mu \ll \nu \ll \otimes_{i=1}^k \nu_{t_i}$ yielding $\mu(A \cap B) = 0$. The fact that $\mu(B)$ allows us to write $\mu(A) = \mu(A \cap B) = 0$, which by the choice of A gives $\mu \ll \otimes_{i=1}^k \mu_{t_i}$.

We showed $\otimes_{i=1}^k \mu_{t_i} \sim \mu$, then Theorem 2.5 gives us (after taking the exponential function) that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \prod_{i=1}^k f_i(x_i), \quad \mu\text{-a.s.},$$

with the corresponding uniqueness statement. We want the equality to hold ν-a.s. For that we need to adjust f_i a little bit. Define $\tilde{f}_i := f_i \mathbb{1}_{U_i}$, with $U_i := \{\frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}} > 0\}$.

By definition $B = \times_{i=1}^k U_i$, and since $\mu(B) = 1$, we have

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \mathbb{1}_B(x) \prod_{i=1}^k f_i(x_i) = \prod_{i=1}^k \mathbb{1}_{U_i}(x_i) f_i(x_i) = \prod_{i=1}^k \tilde{f}_i(x_i), \quad \mu\text{-a.s.},$$

which only holds ν-a.s. on $\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu} > 0\}$. However, once we show that $\mathbb{1}_B = \mathbb{1}_{B \cap \{\frac{\mathrm{d}\mu}{\mathrm{d}\nu} > 0\}}$ ν-a.s., we get that everything holding μ-a.s. on B also holds ν-a.s. on B.

Let us proceed; for any non-negative random variable $Y : (\mathbb{R}^d)^k \to \mathbb{R}_+$ satisfying $Y = 0$ on B^c we have

$$\int_{(\mathbb{R}^d)^k} Y \,\mathrm{d}\nu = \int_{(\mathbb{R}^d)^k} Y \mathbb{1}_B \,\mathrm{d}\nu = \int_{(\mathbb{R}^d)^k} \frac{\mathrm{d}\nu}{\mathrm{d} \otimes_i \nu_{t_i}} Y \mathbb{1}_B \,\mathrm{d} \otimes_i \nu_{t_i},$$

by the fact that $\nu \ll \otimes_i \nu_{t_i}$. We also have $\mu \sim \otimes_i \mu_{t_i} \sim \otimes_i \nu_{t_i}$ on B, which then allows us to write

$$\begin{aligned}\int_{(\mathbb{R}^d)^k} Y \,\mathrm{d}\nu &= \int_{(\mathbb{R}^d)^k} \frac{\mathrm{d}\nu}{\mathrm{d}\otimes_i \nu_{t_i}} \left(\frac{\mathrm{d}\otimes_i \mu_{t_i}}{\mathrm{d}\otimes_i \nu_{t_i}}\right)^{-1} Y \mathbb{1}_B \,\mathrm{d}\otimes_i \mu_{t_i} \\ &= \int_{(\mathbb{R}^d)^k} \frac{\mathrm{d}\nu}{\mathrm{d}\otimes_i \nu_{t_i}} \left(\frac{\mathrm{d}\otimes_i \mu_{t_i}}{\mathrm{d}\otimes_i \nu_{t_i}}\right)^{-1} \frac{\mathrm{d}\otimes_i \mu_{t_i}}{\mathrm{d}\mu} Y \mathbb{1}_B \mathbb{1}_{\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}>0\}} \,\mathrm{d}\mu,\end{aligned}$$

where we have added the indicator of $\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu} > 0\}$ on the last line for the event has full measure under μ. By uniqueness of Radon-Nikodym derivative and the fact that $\nu \sim \mu$ on $\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu} > 0\}$ yields

$$\int_{(\mathbb{R}^d)^k} Y \,\mathrm{d}\nu = \int_{(\mathbb{R}^d)^k} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)^{-1} Y \mathbb{1}_{\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}>0\}} \,\mathrm{d}\mu = \int_{(\mathbb{R}^d)^k} Y \mathbb{1}_{\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}>0\}} \,\mathrm{d}\nu.$$

Therefore, for any set $A \in \mathcal{B}((\mathbb{R}^d)^k)$ we have

$$\mu(A) = \int_A \mathbb{1}_{B\cap\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}>0\}}(x)\,\mu(\mathrm{d}x) = \int_A \mathbb{1}_{\{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}>0\}}(x)\mathbb{1}_B(x)\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\,\nu(\mathrm{d}x) = \int_A \prod_{i=1}^k \tilde{f}_i(x_i)\,\nu(\mathrm{d}x),$$

implying

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \prod_{i=1}^k \tilde{f}_i(x_i), \quad \nu\text{-a.s.},$$

due to uniqueness of Radon-Nikodym derivative.

Finally, under the additional assumption $\otimes_{i=1}^k \mu_{t_i} \sim \nu$, the assertion $\mu \sim \nu$ follows from $\nu \ll \otimes_{i=1}^k \mu_{t_i} \ll \mu \ll \nu$. □

Remark 4.4. The assumption we have in the previous Theorem 4.3 is fairly practical, since we can verify the existence of a unique minimizer through checking whether the trivial candidate $\otimes_{i=1}^k \mu_{t_i}$ is a competitor. However, it may suggest that it is the only way to find a competitor which is not true. In fact, once one knows the existence of a unique minimizer $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$, then one can also get the Schrödinger factorization via other arguments.

For instance, assuming a unique minimizer $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$ exists, Theorem 5.1 in [3] gives the Schrödinger factorization under irreducibility of the reference measure R, i.e.

$$(X_s, X_t)_\# \mathsf{R} \sim (X_s)_\# \mathsf{R} \otimes (X_t)_\# \mathsf{R}, \quad \text{for all } s, t \in [0, 1].$$

Such condition implies $\otimes_{i=1}^k \nu_{t_i} \sim \nu$ as obtained in Lemma 3.8 in the same paper [3].

Comparing our result with [3], we see that we both (implicitly) assume $\nu \ll \otimes_{i=1}^k \nu_{t_i}$, while we have $H(\otimes_{i=1}^k \mu_{t_i}|\nu) < \infty$ and they have $\otimes_{i=1}^k \nu_{t_i} \ll \nu$ instead.

Whenever a Schrödinger factorization exists for which (4.2) holds ν-a.s., we also obtain

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \prod_{i=1}^k f_i(X_{t_i}), \quad \mathsf{R}\text{-a.s.} \tag{4.3}$$

Indeed, using Lemma 4.2 we have for any $A \in \mathcal{F}$ by change of variables

$$\mathsf{P}(A) = \int_{\mathbb{R}^d} \mathbb{E}_{\mathsf{R}^x}[\mathbb{1}_A]\,\mu(\mathrm{d}x) = \int_{\mathbb{R}^d} \mathbb{E}_{\mathsf{R}^x}[\mathbb{1}_A] \prod_{i=1}^{k} f_i(x_i)\,\nu(\mathrm{d}x) = \mathbb{E}_{\mathsf{R}}\left[\mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \mid X_{\mathcal{T}}\right] \prod_{i=1}^{k} f_i(X_{t_i})\right].$$

The pull-out and tower property for conditional expectation gives

$$\mathsf{P}(A) = \mathbb{E}_{\mathsf{R}}\left[\mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \prod_{i=1}^{k} f_i(X_{t_i}) \mid X_{\mathcal{T}}\right]\right] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \prod_{i=1}^{k} f_i(X_{t_i})\right],$$

and the claim follows by the uniqueness of the Radon-Nikodym derivative.

By means of (4.3), one finds compatibility conditions for the Schrödinger factorization. Recall that in Chapter 1 we have seen a similar conditions in terms of the Schrödinger system (1.1). These conditions come from the marginal constraints. To see that, take any $j \in \{1, ..., k\}$ and use the definition of conditional expectation for non-negative random variables

$$\mu_{t_j}(A) = \mathbb{E}_{\mathsf{P}}[\mathbb{1}_A(X_{t_j})] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A(X_{t_j})\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}\right] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A(X_{t_j})\mathbb{E}_{\mathsf{R}}\left[\prod_{i=1}^{k} f_i(X_{t_i}) \mid X_{t_j}\right]\right].$$

We can take out $f_i(X_{t_i})$ for it is $\sigma(X_{t_i})$-measurable to deduce

$$\mu_{t_j}(A) = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A(X_{t_j}) f_j(X_{t_j})\mathbb{E}_{\mathsf{R}}\left[\prod_{i \neq j} f_i(X_{t_i}) \mid X_{t_j}\right]\right].$$

Hence due to uniqueness of the Radon-Nikodym derivative we get

$$\frac{\mathrm{d}\mu_{t_j}}{\mathrm{d}\nu_{t_j}}(x_j) = f_j(x_j)\mathbb{E}_{\mathsf{R}}\left[\prod_{i \neq j} f_i(X_{t_i}) \mid X_{t_j} = x_j\right], \quad j \in \{1, ..., k\}. \tag{4.4}$$

Alternatively, due to the Markov property one can write

$$\frac{\mathrm{d}\mu_{t_j}}{\mathrm{d}\nu_{t_j}}(x_j) = f_j(x_j)\mathbb{E}_{\mathsf{R}}\left[\prod_{i=1}^{j-1} f_i(X_{t_i}) \mid X_{t_j} = x_j\right]\mathbb{E}_{\mathsf{R}}\left[\prod_{i=j+1}^{k} f_i(X_{t_i}) \mid X_{t_j} = x_j\right], \quad j \in \{1, ..., k\}. \tag{4.5}$$

We know that for each $j \in \{1, ..., k\}$ the probability measure μ_{t_j} is fixed by the constraints given in (SMSP) and ν_{t_j} is fixed for it is a marginal of the reference measure. Therefore $\mathrm{d}\mu_{t_j}/\mathrm{d}\nu_{t_j}$ is fixed which then puts constraints on the functions $(f_j)_{1 \leq j \leq k}$ by (4.5).

Equation (4.5) indexed by $j \in \{1, ..., k\}$ is just the Schrödinger system (1.1) from the introduction in a new jacket.

4.1 Regularity of the Schrödinger factorization

In this section we study the regularity of the Schrödinger factorization. It makes sense to assume throughout this whole chapter that the minimizer $\mu \in \mathcal{C}_H^{\mathrm{SMSP}}$ exists and admits a Schrödinger factorization, namely

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) = \prod_{i=1}^{k} f_i(x_i), \quad \nu\text{-a.s.},$$

for some measurable functions $(f_i)_{1\leq i\leq k}$. Until now, we have not seen any results concerning the regularity of the functions $(f_i)_i$. We only know that the functions are non-negative almost surely and that the product is ν-integrable. Regularity of the Schrödinger factorization is studied in [44], for instance in Proposition 4.1.5 we find conditions to ensure the functions $(f_i)_{1\leq i\leq k}$ are bounded almost surely for the case $k=2$. The result given in [44] works well whenever the marginal densities $\mathrm{d}\mu_{t_i}/\mathrm{d}\nu_{t_i}$ have bounded support. Fortunately, we were able to use the same techniques to get a generalization of the result allowing for $\mathrm{d}\mu_{t_i}/\mathrm{d}\nu_{t_i}$ to have unbounded support while decaying fast enough. It might be too restrictive to aim for boundedness of the functions $(f_i)_{1\leq i\leq k}$; so we give conditions for these functions to be in particular L^p-spaces.

Before doing so, let us introduce a notation for $\mathrm{d}\mu_{t_i}/\mathrm{d}\nu_{t_i}$, $i \in \{1, ..., k\}$, namely

$$\rho_{t_i} := \frac{\mathrm{d}\mu_{t_i}}{\mathrm{d}\nu_{t_i}}.$$

The result we are after is the following:

Theorem 4.5 (L^p-regularity for Schrödinger factorization)**.** *Assume the following:*

(i) there exists a function $c : (\mathbb{R}^d)^k \to [0,\infty)$ of the form $c(x) = \prod_{i=1}^k c_i(x_i)$ with $c_i : \mathbb{R}^d \to [0,\infty)$ such that $\otimes_{i=1}^k c_i \nu_{t_i} \leq \nu$ in the sense that

$$\prod_{i=1}^{k} \int_{A_i} c_i(x)\, \nu_{t_i}(\mathrm{d}x_i) \leq \nu(A) \quad \textit{for all } A = \bigtimes_{i=1}^{k} A_i \in \mathcal{B}((\mathbb{R}^d)^k);$$

(ii) there exists a collection of measurable sets $(U_i)_{1\leq i\leq k} \subset \mathcal{B}(\mathbb{R}^d)$ such that for all $i \in \{1, ..., k\}$

$$\nu_{t_i}(U_i \cap \{\rho_{t_i} > 0\}) > 0, \quad \textit{and} \quad \nu_{t_i}(U_i \cap \{c_i = 0\}) = 0;$$

(iii) there exists a collection of measures $(\eta_i)_{1\leq i\leq k}$ such that $\eta_i \ll \nu_{t_i}$ for all $i \in \{1, ..., k\}$; and there exists a collection of numbers $(p_i)_{1\leq i\leq k} \subset [1,\infty]$ such that

$$\left.\frac{\rho_{t_i}}{c_i}\right|_{U_i} \in L^{p_i}(U_i, \eta_i), \quad \textit{for all } i \in \{1, ..., k\}.$$

Then $f_i \in L^{p_i}(U_i, \eta_i)$ *for all* $i \in \{1, ..., k\}$ *and*

$$\|f_i\|_{L^{p_i}(U_i,\eta_i)} \le \left(\prod_{j\neq i} \|c_j f_j\|_{L^1(\mathbb{R}^d,\nu_{t_j})} \right)^{-1} \left\| \frac{\rho_{t_i}}{c_i} \right\|_{L^{p_i}(U_i,\eta_i)}.$$

Proof. Fix $i \in \{1, ..., k\}$ and note that the first assumption with monotone class argument (Theorem 2.2 from Chapter 0 in [35]) implies that for any measurable set $A_i \in U_i$

$$\int_{A_i} \left(\prod_{j\neq i} \int_{\mathbb{R}^d} c_j f_j \, \mathrm{d}\nu_{t_j} \right) c_i f_i \, \mathrm{d}\nu_{t_i} \le \int_{(\mathbb{R}^d)^k} f(x_i) \mathbb{1}_A(x_i) \prod_{j\neq i} f_j(x_j) \, \nu(\mathrm{d}x).$$

The right-hand side with change of variables and conditioning can be written as

$$\begin{aligned} \int_{(\mathbb{R}^d)^k} f(x_i) \mathbb{1}_A(x_i) \prod_{j\neq i} f_j(x_j) \, \nu(\mathrm{d}x) &= \mathbb{E}_{\mathsf{R}} \left[f_i(X_{t_i}) \mathbb{1}_{A_i}(X_{t_i}) \prod_{j\neq i} f_j(X_{t_j}) \right] \\ &= \int_{A_i} \mathbb{E}_{\mathsf{R}} \left[\prod_{j\neq i} f_j(X_{t_j}) \mid X_{t_i} = x_i \right] f_i(x_i) \, \nu_{t_i}(\mathrm{d}x_i) \end{aligned}$$

Recall (4.4) which tells us that

$$\rho_{t_i}(x_i) = f(x_i) \mathbb{E}_{\mathsf{R}} \left[\prod_{j\neq i} f_j(X_{t_j}) \mid X_{t_i} = x_i \right], \quad \nu_{t_i}\text{-a.s.},$$

so that combining what we have obtained above yields the inequality

$$\int_{A_i} \left(\prod_{j\neq i} \int_{\mathbb{R}^d} c_j f_j \, \mathrm{d}\nu_{t_j} \right) c_i f_i \, \mathrm{d}\nu_{t_i} \le \int_{A_i} \rho_{t_i}(x_i) \, \nu_{t_i}(\mathrm{d}x_i).$$

Since $A_i \subset U_i$ was arbitrary, by the fact that inequality of integrals implies inequality of integrands, we conclude

$$c_i(x_i) f_i(x_i) \prod_{j\neq i} \|c_j f_j\|_{L^1(\mathbb{R}^d,\nu_{t_j})} \le \rho_{t_i}(x_i), \qquad \nu_{t_i}\text{-a.s. } x_i \in U_i.$$

For any $j \in \{1, ..., k\}$ we have $f_j > 0$ on $\{\rho_{t_j} > 0\}$ by (4.4) and $c_j > 0$ on U_j which means $c_j f_j > 0$ on $U_j \cap \{\rho_{t_j} > 0\}$. By the second point $\nu_{t_j}(U_j \cap \{\rho_{t_j} > 0\}) > 0$ which together with the previous observation implies that $\|c_j f_j\|_{L^1(\mathbb{R}^d,\nu_{t_j})} > 0$.

Hence we can divide by c_i and the products of $\|c_j f_j\|_{L^1(\mathbb{R}^d,\nu_{t_j})}$ to end up with

$$f_i(x_i) \le \left(\prod_{j\neq i} \|c_j f_j\|_{L^1(\mathbb{R}^d,\nu_{t_j})} \right)^{-1} \frac{\rho_{t_i}(x_i)}{c_i(x_i)}, \qquad \nu_{t_i}\text{-a.s. } x_i \in U_i.$$

We have $\eta_i \ll \nu_{t_i}$ for all $i \in \{1, ..., k\}$ by the last assumption which implies that the previous inequality holds η_i-a.e. The function f_i is positive and dominated by a function in $L^{p_i}(U_i, \eta_i)$ by the last assumption too. Hence, $f_i \in L^{p_i}(U_i, \eta_i)$ and the estimate of its norm follows from the same inequality. □

Remark 4.6. There are two remarks:

- By taking $U_i = \mathbb{R}^d$, we see that we must have $c_i > 0$ ν_{t_i}-a.e. and the result is a global L^p-integrability for the Schrödinger functions.
- By allowing c_i to be zero in the previous theorem, it is enough to have $\otimes_{i=1}^k c_i \nu_{t_i} \le \nu$ on $\bigtimes_{i=1}^k U_i$.

The previous remark suggests that we can recover the results obtained in Proposition 4.1.5 in [44], but one should be a little bit cautious. It is tempting to say that we may take $U_i = \{\rho_{t_i} > 0\}$ and a function c that is a positive constant on $\bigtimes_{i=1}^k U_i$ and 0 outside, and recover integrability on $\mathbb{R}^d$. It is not true, because we only get $f_i \in L^{p_i}(\{\rho_{t_i} > 0\}, \eta_i)$ unless $f_i = 0$ ν_{t_i}-a.s. on $\{\rho_{t_i} = 0\}$. Luckily, there is a condition that ensures the latter which is also in the statement of Proposition 4.1.5 in [44]. We have seen the condition that we refer to in Theorem 4.3 previously, we state it once more for its relevance:

Corollary 4.7. *Assume either $\mu \ll \otimes_{i=1}^k \mu_{t_i}$ or $\nu \ll \otimes_{i=1}^k \nu_{t_i}$. If the assumptions of Theorem 4.5 holds for $U_i = \{\rho_{t_i} > 0\}$ for each $i \in \{1, ..., k\}$, then $f_i \in L^{p_i}(\mathbb{R}^d, \eta_i)$ for all $i \in \{1, ..., k\}$.*

Proof. As seen in the proof of Theorem 4.3 the assumptions $\nu \ll \otimes_{i=1}^k \nu_{t_i}$ or $\mu \ll \otimes_{i=1}^k \mu_{t_i}$ allow us to take each function f_i in the Schrödinger factorization to have support inside $\{\rho_{t_i} > 0\}$. Therefore it is enough to study the integrability of f_i on $U_i = \{\rho_{t_i} > 0\}$. □

4.2 Regularity of the factorization for Langevin diffusions

We illustrate the theory of the previous section for a class of Itô-diffusions, namely Langevin diffusions. Consider a convex function $U \in C^2(\mathbb{R}^d)$ such that $e^{-U} \in L^1(\mathbb{R}^d, \mathrm{d}x)$. We let the reference measure $\mathsf{R} \in \mathcal{P}(\Omega)$ to be the law of a Langevin diffusion

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \tag{4.6}$$

with the initial probability law being

$$\nu_0(\mathrm{d}x) = \eta(\mathrm{d}x) := \left(\int_{\mathbb{R}^d} e^{-U(y)}\,\mathrm{d}y\right)^{-1} e^{-U(x)}\,\mathrm{d}x.$$

Recall the definition of reversible Markov process in Definition 3.18. With some additional assumptions to ensure existence and regularity, we will see that the canonical process X is stationary and reversible under R.

Recall that the transition kernel of the Markov process denoted by $p_t(x, \mathrm{d}y)$ induces a semigroup $(P_t)_{t\in[0,1]}$ on suitably chosen function spaces (see Section 1.2.1 in [2]). We write some conditions on the function U such that the semigroup $(P_t)_{t\in[0,1]}$ has smoothing effect so that the functions $(f_i)_{1\le i\le k}$ in the Schrödinger factorization turn out to be smooth. Throughout this section we consider the following assumption on U:

Assumption 4.8. (i) $U \in C^2(\mathbb{R}^d)$; (ii) for every $\varepsilon > 0$ there exists $c_\varepsilon > 0$ such that $|D^2U| \leq \varepsilon|\nabla U|^2 + c_\varepsilon$; (iii) $D^2U \geq \theta I$ for some $\theta > 0$; and (iv) $e^{-U} \in L^1(\mathbb{R}^d, \eta)$.

Remark 4.9 (Existence of transition density). Assumption (ii) above implies that $\Delta U - \frac{1}{2}|\nabla U|^2 \leq M$ for some constant $M > 0$ allowing us to apply the argument around Equation (7.4.4) and Proposition 7.4.9 in [2] to get the existence of a transition density $p_t(x, y)$ with respect to the invariant measure. Moreover the existence of a transition density with respect to the invariant measure implies $\nu \ll \otimes_{i=1}^k \eta$ which allows us to use Corollary 4.7.

Recall the definition of analytic semigroup which basically says the semigroup $(P_t)_{t\geq 0}$ has analytic extension on some parts of the complex plane (see Definition 5.1 in [32]). It happens so that under Assumption 4.8 we get that the semigroup $(P_t)_{t\in[0,1]}$ is analytic on $L^p(\mathbb{R}^d, \eta)$ with $p \in (1, \infty)$ and is symmetric for $p = 2$ (e.g. see [27]). The generator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ of the semigroup $(P_t)_{t\geq 0}$ is

$$\mathcal{L} = \Delta - \nabla U \cdot \nabla, \quad \mathcal{D}(\mathcal{L}) = W^{2,p}(\mathbb{R}^d, \eta), \tag{4.7}$$

where $W^{m,p}(\mathbb{R}^d, \eta)$ is the space of m-times weakly differentiable functions for which all the weak derivatives up to order m are in $L^p(\mathbb{R}^d, \eta)$.

Note that the reference probability measure R solves the martingale problem associated with the generator $\mathcal{L}$. It is well-know that the L^2-symmetry of the semigroup corresponds to the reversibility of the process as given in Definition 3.18 (see Section 1.6.1 in [2]).

The special property that comes from $(P_t)_{t\in[0,1]}$ being analytic is that for any function $f \in L^p(\mathbb{R}^d, \eta)$ we can differentiate $t \mapsto P_t f$ for $t \in (0, 1]$ infinitely many times as a $L^p(\mathbb{R}^d, \eta)$-valued function and (see Lemma 4.2 in [32]):

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n} P_t f = \mathcal{L}^n P_t f.$$

In particular $t \mapsto P_t f$ is in $C^\infty((0, 1]; L^p(\mathbb{R}^d, \eta))$. Furthermore, if $U \in C^\infty(\mathbb{R}^d)$ we get a result concerning parabolic regularity.

Theorem 4.10 (Parabolic regularity for the Langevin generator)**.** *Assume $U \in C^\infty(\mathbb{R}^d)$. Then for any function $f \in L^p(\mathbb{R}^d, \eta)$ the function $u : \mathbb{R}^d \times (0, 1] \to \mathbb{R}$ defined by $u(x, t) := P_t f(x)$ is in $C^\infty(\mathbb{R}^d \times (0, 1])$ and for any fixed $t \in (0, 1]$ we have $u(\cdot, t) \in W^{2,p}(\mathbb{R}^d, \eta)$.*

Proof. See Appendix A.1. ☐

We want to apply these results to the functions occurring in the Schrödinger factorization. We define a collection of functions $(h_i^{\rightarrow})_{1\leq i\leq k}$ with $h_i^{\rightarrow} : \mathbb{R}^d \times [t_{i-1}, t_i) \to \mathbb{R}$ and

$$h_i^{\rightarrow}(y, t) := \mathbb{E}_{\mathsf{R}} \left[\prod_{j=i}^{k} f_j(X_{t_j}) \mid X_t = y \right]. \tag{4.8}$$

Similarly we define $(h_i^{\leftarrow})_{1\leq i\leq k}$ with $h_i^{\leftarrow} : \mathbb{R}^d \times (t_i, t_{i+1}] \to \mathbb{R}$ and

$$h_i^{\leftarrow}(y, t) := \mathbb{E}_{\mathsf{R}} \left[\prod_{j=1}^{i} f_j(X_{t_j}) \mid X_t = y \right]. \tag{4.9}$$

With these notations and using Equation (4.5) we can rewrite ρ_{t_i} as follows for any $i \in \{2, ..., k-1\}$:

$$\rho_{t_i}(x_i) = h^{\leftarrow}_{i-1}(x_i, t_i) f_i(x_i) h^{\rightarrow}_{i+1}(x_i, t_i).$$

For $i = 1$ and $i = k$ we have

$$\rho_{t_1}(x_1) = f_1(x_1) h^{\rightarrow}_2(x_1, t_1), \quad \rho_{t_k}(x_k) = h^{\leftarrow}_{k-1}(x_k, t_k) f_k(x_k).$$

In calculations, we may also use the convention $h^{\leftarrow}_0 \equiv 1$ and $h^{\rightarrow}_{k+1} \equiv 1$ to avoid different cases. Also, it is clear that we only need $(h^{\rightarrow}_i(\cdot, t_{i-1}))_{2\leq i\leq k}$ and $(h^{\leftarrow}_i(\cdot, t_{i+1}))_{1\leq i\leq k-1}$ in the expression for $(\rho_{t_i})_{1\leq i\leq k}$ which makes them the only relevant functions for study.

A consequence of the smoothing property stated in Theorem 4.10 is the following:

Proposition 4.11. *Assume the functions $(f_i)_{1\leq i\leq k}$ are in $L^p(\mathbb{R}^d, \eta)$ for some $p > k-1$. Then the functions $(h^{\rightarrow}_i(\cdot, t_{i-1}))_{2\leq i\leq k}$ and $(h^{\leftarrow}_i(\cdot, t_{i+1}))_{1\leq i\leq k-1}$ are in $W^{2,p/(k-1)}(\mathbb{R}^d, \eta)$. If in addition $U \in C^\infty(\mathbb{R}^d)$, then $(h^{\rightarrow}_i(\cdot, t_{i-1}))_{2\leq i\leq k}$ and $(h^{\leftarrow}_i(\cdot, t_{i+1}))_{1\leq i\leq k-1}$ are even in $C^\infty(\mathbb{R}^d)$.*

Proof. We prove the (slightly stronger) assertion that $h^{\rightarrow}_i(\cdot, t_{i-1}) \in W^{2,p/(k-i+1)}(\mathbb{R}^d, \eta)$ for $i \in \{2, ..., k\}$ starting with k. The fact that $h^{\rightarrow}_k(\cdot, t_{k-1}) = P_{t_k - t_{k-1}} f_k$ and $t_k - t_{k-1} > 0$ makes Theorem 4.10 applicable so that $h^{\rightarrow}_k(\cdot, t_{k-1}) \in W^{2,p}(\mathbb{R}^d, \eta)$. This establishes the claim for k.

Let us proceed with the inductive step now and assume the claim is true for i, i.e. $h^{\rightarrow}_i(\cdot, t_{i-1}) \in W^{2,p/(k-i+1)}(\mathbb{R}^d, \eta)$. We want to show the claim for $i-1$; for that note that

$$h^{\rightarrow}_{i-1}(X_{t_{i-2}}, t_{i-2}) = \mathbb{E}_{\mathsf{R}}\left[\prod_{j=i-1}^{k} f_j(X_{t_j}) \mid X_{t_{i-2}}\right].$$

We use the Markov property to write this as

$$h^{\rightarrow}_{i-1}(X_{t_{i-2}}, t_{i-2}) = \mathbb{E}_{\mathsf{R}}\left[\prod_{j=i-1}^{k} f_j(X_{t_j}) \mid \mathcal{F}_{t_{i-2}}\right].$$

We use the properties "smallest σ-algebra wins" on $\mathcal{F}_{t_{i-2}} \subset \mathcal{F}_{t_{i-1}}$ and "pull-out what is known" to get

$$\begin{aligned} h^{\rightarrow}_{i-1}(X_{t_{i-2}}, t_{i-2}) &= \mathbb{E}_{\mathsf{R}}\left[\mathbb{E}_{\mathsf{R}}\left[\prod_{j=i-1}^{k} f_j(X_{t_j}) \mid \mathcal{F}_{t_{i-1}}\right] \mid \mathcal{F}_{t_{i-2}}\right] \\ &= \mathbb{E}_{\mathsf{R}}\left[f_{i-1}(X_{t_{i-1}}) \mathbb{E}_{\mathsf{R}}\left[\prod_{j=i}^{k} f_j(X_{t_j}) \mid \mathcal{F}_{t_{i-1}}\right] \mid \mathcal{F}_{t_{i-2}}\right]. \end{aligned}$$

Now the Markov property yields

$$h^{\rightarrow}_{i-1}(X_{t_{i-2}}, t_{i-2}) = \mathbb{E}_{\mathsf{R}}\left[f_{i-1}(X_{t_{i-1}}) \mathbb{E}_{\mathsf{R}}\left[\prod_{j=i}^{k} f_j(X_{t_j}) \mid X_{t_{i-1}}\right] \mid \mathcal{F}_{t_{i-2}}\right].$$

We recognize the random variable inside the outer conditional expectation which simplifies the expression to

$$h^{\rightarrow}_{i-1}(X_{t_{i-2}}, t_{i-2}) = \mathbb{E}_{\mathsf{R}}\left[f_{i-1}(X_{t_{i-1}})h^{\rightarrow}_{i}(X_{t_{i-1}}, t_{i-1}) \mid \mathcal{F}_{t_{i-2}}\right] = P_{t_{i-1}-t_{i-2}}[f_{i-1}h^{\rightarrow}_{i}(\cdot, t_{i-1})](X_{t_{i-2}}).$$

So now we have

$$h^{\rightarrow}_{i-1}(x_{i-2}, t_{i-2}) = P_{t_{i-1}-t_{i-2}}[f_{i-1}h^{\rightarrow}_{i}(\cdot, t_{i-1})](x_{i-2}).$$

Note that $f_{i-1} \in L^p(\mathbb{R}^d, \eta)$ and $h^{\rightarrow}_{i}(\cdot, t_{i-1}) \in L^{p/(k-i+1)}(\mathbb{R}^d, \eta)$ by the induction hypothesis so that

$$f_{i-1}h^{\rightarrow}_{i}(\cdot, t_{i-1}) \in L^{p/(k-i+2)}(\mathbb{R}^d, \eta),$$

by Hölder's inequality. We know that $i \geq 2$ and $p > k-1$ so that $p/(k-i+2) > 1$. Therefore Theorem 4.10 gives $h^{\rightarrow}_{i-1}(\cdot, t_{i-2}) \in W^{2,p/(k-i+2)}(\mathbb{R}^d, \eta)$ as required. This finishes the induction.

We have actually obtained a stronger statement, and the precise statement in the theorem can be obtained by noting that for $2 \leq i \leq k$

$$\frac{p}{k-i+1} \geq \frac{p}{k-1},$$

so that for each $i \in \{2, ..., k\}$ we get $h^{\rightarrow}_{i}(\cdot, t_{i-1}) \in W^{2,p/(k-1)}(\mathbb{R}^d, \eta)$, since η is a probability measure.

The statement for $(h^{\leftarrow}_{i}(\cdot, t_{i+1}))_{1 \leq i \leq k-1}$ can be similarly obtained by working backwards in time which yields the same result since the process X is reversible under R.

Finally, if $U \in C^\infty(\mathbb{R}^d)$ the application of Theorem 4.10 actually gives us functions in $C^\infty(\mathbb{R}^d)$. □

Remark 4.12. In the proof of the previous Proposition 4.11, we have applied Hölder's inequality using merely that for instance $h^{\rightarrow}_{i}(\cdot, t_{i-1}) \in L^{p/(k-i+1)}(\mathbb{R}^d, \eta)$ while actually we know more, namely $h^{\rightarrow}_{i}(\cdot, t_{i-1}) \in W^{2,p/(k-i+1)}(\mathbb{R}^d, \eta)$. That leads straightforwardly to the question whether we can improve the integrability in the sense that we may have a Sobolev embedding of the form $W^{2,p/(k-i+1)}(\mathbb{R}^d, \eta) \subset L^{\bar{p}}(\mathbb{R}^d, \eta)$ for some $\bar{p} > p/(k-i+1)$. If such Sobolev embedding exists, then we can reduce the integrability assumption on $(f_i)_{1 \leq i \leq k}$.

Corollary 4.13. *Consider the same setting as Proposition 4.11 with $U \in C^\infty(\mathbb{R}^d)$. If $\rho_{t_i} \in C^m(\mathbb{R}^d)$ for some $m \in \mathbb{N}$, then $f_i \in C^m(\mathbb{R}^d)$.*

Proof. This is a straightforward application of the previously found result. Fix $i \in \{1, ..., k\}$ and note that

$$\rho_{t_i}(x_i) = h^{\leftarrow}_{i-1}(x_i, t_i)f_i(x_i)h^{\rightarrow}_{i+1}(x_i, t_i).$$

Note that both $h^{\leftarrow}_{i-1}(\cdot, t_i)$ and $h^{\rightarrow}_{i+1}(\cdot, t_i)$ are in $C^\infty(\mathbb{R}^d)$ by Proposition 4.11. As soon as we show that $h^{\leftarrow}_{i-1}(\cdot, t_i)$ and $h^{\rightarrow}_{i+1}(\cdot, t_i)$ are positive we can divide by them to get $f_i \in C^m(\mathbb{R}^d)$.

The positivity of $h^{\rightarrow}_{i+1}$ can be obtained as follows. We start with $h^{\rightarrow}_{k}$ and argue by induction. First of all the fact that the semigroup P_t has a transition density p_t with respect η implies for any $A \in \mathcal{B}((\mathbb{R}^d)^k)$

$$\nu(A) = \int_A \prod_{i=1}^{k-1} p_{t_{i+1}-t_i}(x_i, x_{i+1})\, \eta(\mathrm{d}x_1) \cdots \eta(\mathrm{d}x_k), \quad \nu\text{-a.s.},$$

so that

$$\frac{\mathrm{d}\nu}{\mathrm{d} \otimes_{i=1}^k \eta}(x) = \prod_{i=1}^{k-1} p_{t_{i+1}-t_i}(x_i, x_{i+1}).$$

By Lemma 4.15 below we get $\otimes_{i=1}^k \eta \ll \nu$ which implies that $\mathrm{d}\nu/\mathrm{d} \otimes_{i=1}^k \eta > 0$ ν-a.s., which then implies

$$p_{t_k - t_{k-1}}(x_{k-1}, x_k) > 0, \quad \eta \otimes \eta\text{-a.s. } (x_{k-1}, x_k) \in \mathbb{R}^d \times \mathbb{R}^d.$$

Therefore by Fubini we have for η-a.s. $x_{k-1} \in \mathbb{R}^d$

$$p_{t_k - t_{k-1}}(x_{k-1}, x_k) > 0, \quad \eta\text{-a.s. } x_k.$$

In particular,

$$h^{\rightarrow}_k(x_{k-1}, t_{k-1}) = P_{t_k - t_{k-1}} f_k(x_{k-1}) = \int_{\mathbb{R}^d} f_k(y) p_{t_k - t_{k-1}}(x_{k-1}, y)\, \eta(\mathrm{d}y) > 0, \quad \eta\text{-a.s. } x_{k-1}.$$

We know by Theorem 4.10 that $h^{\rightarrow}_k(\cdot, t_{k-1})$ is in $C^\infty(\mathbb{R}^d)$ and together with the fact that η has a strictly positive Lebesgue density, implies that $h^{\rightarrow}_k(\cdot, t_{k-1})$ has to be positive everywhere on $\mathbb{R}^d$.

We argue via induction and reversibility as in the proof of Proposition 4.11 to obtain that for all relevant i we have that $h^{\leftarrow}_{i-1}$ and $h^{\rightarrow}_{i+1}$ are strictly positive. This concludes the proof. □

Remark 4.14. The condition $U \in C^\infty(\mathbb{R}^d)$ is mostly for elliptic regularity which was needed in Theorem 4.10. But since we can at most get f_i as many times continuously differentiable as ρ_{t_i} we do not need to have $h^{\rightarrow}_i$ and $h^{\leftarrow}_i$ to be $C^\infty(\mathbb{R}^d)$. So we can relax $U \in C^\infty(\mathbb{R}^d)$ slightly. One can argue via Morrey's inequality (Theorem 6, Section 5.6 in [13]) that $U \in C^{\bar{m}}(\mathbb{R}^d)$ with $\bar{m} > m + \lceil \frac{d}{p} \rceil$ does the job.

We want to apply these results for the Schrödinger factorization in case of Langevin diffusion, but the results that we have seen above require us to know that the functions $(f_i)_{1 \le i \le k}$ are in some suitable L^p-space already. We can establish that through Theorem 4.5. To apply these results we need lower bounds of the form $\otimes_{i=1}^k c_i \eta \ll \nu$. Fortunately, the following lemma gives us an infinite amount of possibilities to choose the function c in Theorem 4.5 for the case of Langevin diffusions.

Lemma 4.15. *Let $\varepsilon = (\varepsilon_i)_{2\le i\le k} \subset (0,\infty)$ be an arbitrary collection of real numbers. Then $\otimes_{i=1}^k c_i\eta \le \nu$ with the functions $c_i = c_i(\cdot,\varepsilon)$ being*

$$c_i(x_i) = \exp\left(-\theta\beta_i(\varepsilon)|x_i|^2\right),$$

with

$$\beta_i(\varepsilon) = \frac{1+\varepsilon_i}{2(e^{\theta(t_i-t_{i-1})}-1)} + \frac{1+\varepsilon_{i+1}^{-1}}{2(e^{\theta(t_{i+1}-t_i)}-1)}, \quad i \in \{2,...,k-1\},$$

and for $i = 1, k$ we have

$$\beta_1(\varepsilon) = \frac{1+\varepsilon_2^{-1}}{2(e^{\theta(t_2-t_1)}-1)}, \quad \text{and} \quad \beta_k(\varepsilon) = \frac{1+\varepsilon_k}{2(e^{\theta(t_k-t_{k-1})}-1)}.$$

Proof. Recall that the semigroup has a transition density p_t (see Remark 4.9). The strict convexity of U (Assumption 4.8 (iii)) yields a Gaussian lower bound for the transition density p_t. It is derived in Remark 5.6.2 in [2] using log-Harnack inequality that

$$p_t(x,y) \ge \exp\left(-\frac{\theta|x-y|^2}{2(e^{\theta t}-1)}\right) \qquad \text{for all } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \text{ and } t > 0.$$

Now take any measurable $A = \times_{i=1}^k A_i \in \mathcal{B}((\mathbb{R}^d)^k)$ and see that

$$\nu(A) = \int_{A_1}\int_{A_2}\cdots\int_{A_k}\prod_{i=2}^k p_{t_i-t_{i-1}}(x_{i-1},x_i)\,\eta(\mathrm{d}x_1)\,\ldots\,\eta(\mathrm{d}x_k).$$

Using the inequality we stated, we can bound the integrand as follows:

$$\prod_{i=2}^k p_{t_i-t_{i-1}}(x_{i-1},x_i) \ge \exp\left(-\theta\sum_{i=2}^k \frac{|x_i-x_{i-1}|^2}{2(e^{\theta(t_i-t_{i-1})}-1)}\right).$$

Take any $(\varepsilon_i)_{2\le i\le k} \subset (0,\infty)$ to lower bound the quadratic term inside the exponential function with Young's inequality

$$-|x_i-x_{i-1}|^2 = -\left(|x_i|^2 + 2(x_i,x_{i-1})_{\mathbb{R}^d} + |x_{i-1}|^2\right) \ge -(1+\varepsilon_i)|x_i|^2 - (1+\varepsilon_i^{-1})|x_{i-1}|^2.$$

We obtain

$$\prod_{i=2}^k p_{t_i-t_{i-1}}(x_{i-1},x_i) \ge \exp\left(-\theta\sum_{i=2}^k \frac{(1+\varepsilon_i)|x_i|^2 + (1+\varepsilon_i^{-1})|x_{i-1}|^2}{2(e^{\theta(t_i-t_{i-1})}-1)}\right).$$

If we look at the coefficients of $|x_i|^2$ closely, we get the asserted claim

$$\nu(A) \ge \prod_{i=1}^k \int_{A_i} \exp\left(-\theta\beta_i(\varepsilon)|x_i|^2\right)\eta(\mathrm{d}x_i),$$

with $\beta_i(\varepsilon)$ as stated above. $\square$

One thing that we can learn from the previous lemma is that we have a lot of freedom choosing our functions $(c_i)_i$. In fact, for a fixed $i \in \{1, ..., k\}$ we can choose c_i to be as "optimal" as possible to show the function f_i is in a particular L^p-space. One can repeat the same procedure independently for each $i \in \{1, ..., k\}$. The following theorem illustrates this procedure. Unfortunately, the result in the following theorem is still not optimal, because we use rough estimations, but it may be useful for its quiet simple presentation.

Theorem 4.16. *Assume that there exist constants $M > 0$ and $\delta \in \mathbb{R}$ such that*

$$|\rho_{t_i}(x)| \leq M e^{\delta|x|^2} \qquad \text{for all } x \in \mathbb{R}^d \text{ and } i \in \{1, ..., k\}.$$

Define

$$\kappa(\theta) := \max_{2 \leq i \leq k} \left(\frac{\theta}{e^{\theta(t_i - t_{i-1})} - 1} \right). \tag{4.10}$$

Then $f_i \in L^p(\mathbb{R}^d, \eta)$ for all $p \in [1, \overline{p_i})$ with $\overline{p_i}$

$$\overline{p_i} = \begin{cases} \dfrac{\theta}{2(\delta + \kappa(\theta))} & \text{if } \ i \notin \{1, k\} \ \text{ and } \ \delta > -\kappa(\theta), \\ \dfrac{\theta}{2\delta + \kappa(\theta)} & \text{if } \ i \in \{1, k\} \ \text{ and } \ 2\delta > -\kappa(\theta), \\ \infty & \text{otherwise.} \end{cases}$$

If $\delta < -\kappa(\theta)$, then $f_i \in L^\infty(\mathbb{R}^d, \eta)$ for all $i \in \{1, ..., k\}$, and $2\delta < -\kappa(\theta)$ is enough to yield the same conclusion for $i \in \{1, k\}$.

Proof. We only give the proof for $i \notin \{1, k\}$ and $\delta > -\kappa(\theta)$. With the functions c_i from the Lemma 4.15, we see that for ρ_{t_i}/c_i to be in $L^p(\mathbb{R}^d, \eta)$ we are basically seeking the Lebesgue-integrability of

$$|\rho_{t_i}|^p \exp\left(p\theta\beta_i(\varepsilon)|x_i|^2 - U(x_i)\right).$$

Note that if we take $\gamma_i \in (0, 1/2)$, then

$$U_{\gamma_i} := \max_{y \in \mathbb{R}^d} \left(\gamma_i \theta |y|^2 - U(y)\right) < \infty,$$

due to uniform convexity of U and standard result on optimizing uniformly convex functions. Now we can bound as follows:

$$|\rho_{t_i}|^p \exp\left(p\theta\beta_i(\varepsilon)|x_i|^2 - U(x_i)\right) \leq \exp\left((p\delta + p\theta\beta_i(\varepsilon) - \gamma_i\theta)|x_i|^2 + U_{\gamma_i}\right).$$

As soon as we show that we can make the coefficient of $|x_i|^2$ negative we are done. Note for $i \in \{2, ..., k-1\}$ we have the coefficient of $|x_i|^2$ being equal to

$$p\delta + p\theta\beta_i(\varepsilon) - \gamma_i\theta = p\delta + p\theta\left[\frac{1 + \varepsilon_i}{2(e^{\theta(t_i - t_{i-1})} - 1)} + \frac{1 + \varepsilon_{i+1}^{-1}}{2(e^{\theta(t_{i+1} - t_i)} - 1)}\right] - \gamma_i\theta.$$

Assuring

$$p\delta + p\theta \left[\frac{1}{2(e^{\theta(t_i - t_{i-1})} - 1)} + \frac{1}{2(e^{\theta(t_{i+1} - t_i)} - 1)} \right] - \frac{1}{2}\theta < 0$$

allows us to deduce that the previous quantity can be chosen to be negative by varying ε and γ. The negativity of the last quantity is assured if

$$p\delta + p\kappa(\theta) - \frac{1}{2}\theta < 0.$$

Therefore for any $p < \frac{\theta}{2(\delta+\kappa(\theta))}$ we get

$$p\delta + p\kappa(\theta) - \frac{1}{2}\theta < \frac{1}{2}\theta - \frac{1}{2}\theta = 0.$$

This ensures $\rho_{t_i}/c_i \in L^p(\mathbb{R}^d, \eta)$ and we are done by Theorem 4.5. $\square$

Remark 4.17. The case where $\rho_{t_i} \in C_c(\mathbb{R}^d)$ is a case where one automatically gets $f_i \in L^\infty(\mathbb{R}^d, \eta)$ by taking $\delta < 0$ negatively large and $M > 0$ large enough. We also do not need to seek for different c_i from the ones given in Lemma 4.15, because they are positive on compact sets.

In the previous Theorem 4.16 the constant $\kappa(\theta)$ behaves like

$$\kappa(\theta) \approx \frac{1}{\min_i (t_i - t_{i-1})}.$$

Therefore, if if for some $i \in \{2, ..., k\}$ the term $t_i - t_{i-1}$ is small the $\kappa(\theta)$ is big and therefore it ruins the regularity for all the other functions in the Schrödinger factorization. To have a more optimal result on the regularity one is advised to study each i separately.

The next theorem shows that if we can ensure that $f_i \in L^p(\mathbb{R}^d, \eta)$ with $p > k-1$ we can even get higher regularity in the sense of classical differentiability. Indeed, this is an obvious consequence of Corollary 4.13. We state the result with a slight warning for the reader that it is not the most general statement.

Theorem 4.18. *Assume that $U \in C^\infty(\mathbb{R}^d)$ and $\rho_{t_i} \in C^m(\mathbb{R}^d)$ for all $i \in \{1, ..., k\}$ for some $m \in \mathbb{N}$. In addition, assume that there exist constants $M > 0$ and $\delta < \frac{1}{2(k-1)}\theta - \kappa(\theta)$, with $\kappa(\theta)$ given by* (4.10), *such that*

$$|\rho_{t_i}(x)| \leq M e^{\delta |x|^2}, \qquad \textit{for all } x \in \mathbb{R}^d.$$

Then $f_i \in L^p(\mathbb{R}^d, \eta) \cap C^m(\mathbb{R}^d)$ for all $p \in [1, k-1]$. For the case $k = 2$, it is enough to require $\delta < \frac{1}{2}(\theta - \kappa(\theta))$.

Proof. The result is a combination of Theorem 4.16 and Corollary 4.13. Indeed, the condition on δ ensures $f_i \in L^p(\mathbb{R}^d, \eta)$ for some $p > k-1$ (and in particular for all $p \in [1, k-1]$ for η is a probability measure). By Corollary 4.13 we get $f_i \in C^m(\mathbb{R}^d)$, since $\otimes_{i=1}^k c_i \eta \leq \nu$ and $c_i > 0$ everywhere implies $\otimes_{i=1}^k \eta \ll \nu$. The calculations are all basic algebra and therefore we omit the details. $\square$

Needless to say, reversible Langevin diffusions are not the only processes that could be treated with the methods in this section. A straightforward generalization applies to the case where X is a weak solution to

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t,$$

with initial distribution being the corresponding stationary measure. The results above that required analyticity, elliptic regularity and some type of log-Harnack inequality are actually applicable in the generalization above (see e.g. [27] and [2]). The only thing that requires attention is that the norms in the exponential in e.g. Lemma 4.15 must be changed to the Riemannian metric induced by $a = \sigma\sigma^\top$.

4.3 Characterization of the drift of the minimizer

The minimizer $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ (if it exists) of the multimarginal Schrödinger problem (MSP) for $\mathcal{T} = \{t_1, ..., t_k\}$ can be characterized via the Schrödinger factorization. We also know by Theorem 3.6 that $\mathsf{P} \in \mathrm{MP}(\mathcal{L}_t + \beta_t^\top a\nabla, \mu_0)$ for some progressively measurable process $\beta = (\beta_t)_{t\in[0,1]}$ which we have called the control. We will study the control β in the sequel.

Via classical Girsanov's theorem we know that the control β is related to a local martingale defined through the Doob's martingale $(\mathbb{E}_{\mathsf{R}}\left[\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R} \mid \mathcal{F}_t\right])_{t\in[0,1]}$ (see Theorem 1.7 from Chapter 8 in [35]). We also know that $\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R}$ is given by the Schrödinger factorization which, obviously, makes them related to the control β. We will assume that a unique minimizer $\mathsf{P} \in \mathcal{C}_H^{\mathrm{MSP}}$ exists that admits a Schrödinger factorization

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \prod_{i=1}^k f_i(X_{t_i}), \quad \mathsf{R}\text{-a.s.},$$

for some measurable functions $(f_i)_{1\le i\le k}$. We can define a martingale $Z = (Z_t)_{t\in[0,1]}$ by

$$Z_t := \mathbb{E}_{\mathsf{R}}\left[\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} \mid \mathcal{F}_t\right].$$

We can write this Z_t for $j \in \{1, ..., k\}$ and $t \in [t_{j-1}, t_j)$ via the Markov property as follows:

$$Z_t = \mathbb{E}_{\mathsf{R}}\left[\prod_{i=j}^k f_i(X_{t_i}) \mid \mathcal{F}_t\right]\prod_{i=1}^{j-1} f_i(X_{t_i}) = \mathbb{E}_{\mathsf{R}}\left[\prod_{i=j}^k f_i(X_{t_i}) \mid X_t\right]\prod_{i=1}^{j-1} f_i(X_{t_i}).$$

Using the definition of $(h_j^{\rightarrow})_{1\le j\le k}$ in (4.8), we see that

$$Z_t = h_j(X_t, t)\prod_{i=1}^{j-1} f_i(X_{t_i}), \tag{4.11}$$

where we have dropped the arrow "$\rightarrow$", because there will be no confusion about the direction for what we will use these functions for. We define a function $h : \mathbb{R}^d \times [0, t_k] \to \mathbb{R}$ through gluing the functions $(h_j)_{1\le j\le k}$ together

$$h(x,t) := \sum_{j=1}^{k} h_j(x,t)\mathbb{1}_{[t_{j-1},t_j)}(t). \tag{4.12}$$

Under regularity assumptions we see that P is the h-transform with this particular h defined above. Recall the usual notations $\nu_0 = (X_0)_\#\mathsf{R}$ and $\mu_0 = (X_0)_\#\mathsf{P}$.

Theorem 4.19 (Characterization of the drift)**.** *Assume that $h_j \in C^{2,1}(\mathbb{R}^d \times (t_{j-1}, t_j)) \cap C(\mathbb{R}^d \times [t_{j-1}, t_j))$ for any $j \in \{1, ..., k\}$. Then X under P has the semimartingale decomposition $X = X_0 + A^{\mathsf{P}} + M^{\mathsf{P}}$ with*

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s,s) + a(X_s,s)\nabla_x \log h(X_s,s))\,\mathrm{d}s, \quad M_t^{\mathsf{P}} = M_t - \int_0^t a(X_s,s)\nabla_x \log h(X_s,s)\,\mathrm{d}s.$$

If in addition $\mathrm{MP}(\mathcal{L}_t)$ *satisfies the semi-uniqueness condition (see Definition 3.5), then*

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\mathbb{E}_{\mathsf{P}}\left[\int_0^{t_k} |\sigma(X_s,s)^\top \nabla_x \log h(X_s,s)|^2\,\mathrm{d}s\right].$$

Proof. We will start with the first interval, namely $[0, t_1)$, except if $t_1 = 0$, then we take $[t_1, t_2)$ instead. The argument stays the same, so we assume $t_1 > 0$. In such case, for $t \in [0, t_1]$ we can write $Z_t = h_1(X_t, t)$ which in particular implies

$$h_1(X_0, 0) = \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0).$$

Note that conditional expectation preserves non-negativity, therefore $h_1(X_t, t) \ge 0$. We want to take the logarithm of $h_1(X_t, t)$, but that may not be well-defined. So, we will give ourselves a little bit of extra space in terms of a small $\varepsilon > 0$. Let us apply Itô's formula on $\log(h_1(X_t, t) + \varepsilon)$ to get

$$\log(h_1(X_t,t)+\varepsilon) = \log(h_1(X_0,0)+\varepsilon) + \int_0^t \nabla_x \log(h_1(X_s,s)+\varepsilon)\,\mathrm{d}M_s + \alpha_t^\varepsilon \tag{4.13}$$

where α_t^ε is the finite variation part, to be more precise

$$\alpha_t^\varepsilon = \int_0^t (\partial_t + \mathcal{L}_s)\log(h_1(X_s,s)+\varepsilon)\,\mathrm{d}s.$$

Note that $(Z_t + \varepsilon)_{t\in[0,1]}$ is a strictly positive martingale, so there exists a local martingale L^ε (Proposition 1.6 from Chapter 8 in [35]) such that

$$Z_t + \varepsilon = \mathcal{E}(L^\varepsilon)_t,$$

where $\mathcal{E}(\cdot)$ is the stochastic exponential ($\mathcal{E}(L)_t = \exp(L_t - [L]_t/2)$ for any local martingale L). For $t \in [0, t_1)$ we have

$$Z_t + \varepsilon = \exp\left(\log(h_1(X_t, t) + \varepsilon)\right).$$

Due to uniqueness of semimartingale decomposition and the application of Itô's formula as done in (4.13) we conclude that

$$L_t^\varepsilon = \log(h_1(X_0, 0) + \varepsilon) + \int_0^t \nabla_x \log(h_1(X_s, s) + \varepsilon)\, \mathrm{d}M_s \quad t \in [0, t_1)$$

and together with $-\frac{1}{2}[L^\varepsilon]_t = \alpha_t^\varepsilon$ we get for all $t \in [0, t_1)$

$$\int_0^t (\partial_t + \mathcal{L}_s) \log(h_1(X_s, s) + \varepsilon)\, \mathrm{d}s = -\frac{1}{2} \int_0^t |\sigma(X_s, s)^\top \nabla_x \log(h_1(X_s, s) + \varepsilon)|^2\, \mathrm{d}s. \tag{4.14}$$

Applying a similar argument on all the intervals $[t_{j-1}, t_j)$ with an induction argument yields

$$Z_t + \varepsilon = \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0) + \varepsilon\right) \mathcal{E}\left(\int_0^\cdot \nabla_x \log(h(X_s, s) + \varepsilon)\, \mathrm{d}M_s\right)_t \qquad \text{for all } t \in [0, 1]. \tag{4.15}$$

Now taking $t = 1$ yields

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} + \varepsilon = \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0) + \varepsilon\right) \mathcal{E}\left(\int_0^\cdot \nabla_x \log(h(X_s, s) + \varepsilon)\, \mathrm{d}M_s\right)_1, \quad \mathsf{R}\text{-a.s.} \tag{4.16}$$

Consider the new measure Q^ε defined by

$$\frac{\mathrm{d}\mathsf{Q}^\varepsilon}{\mathrm{d}\mathsf{R}} := \frac{1}{1+\varepsilon}\left(\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} + \varepsilon\right).$$

Equivalently, we can write

$$\mathsf{Q}^\varepsilon = \frac{1}{1+\varepsilon}\mathsf{P} + \left(1 - \frac{1}{1+\varepsilon}\right)\mathsf{R}$$

to see that the convexity of the relative entropy gives

$$H(\mathsf{Q}^\varepsilon|\mathsf{R}) \le \frac{1}{1+\varepsilon} H(\mathsf{P}|\mathsf{R}) + \left(1 - \frac{1}{1+\varepsilon}\right) \underbrace{H(\mathsf{R}|\mathsf{R})}_{=0} \le H(\mathsf{P}|\mathsf{R}) < \infty.$$

By (4.16) we see that

$$\frac{\mathrm{d}\mathsf{Q}^\varepsilon}{\mathrm{d}\mathsf{R}} = \frac{\frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0) + \varepsilon}{1+\varepsilon} \mathcal{E}\left(\int_0^\cdot \nabla_x \log(h(X_s, s) + \varepsilon)\, \mathrm{d}M_s\right)_1, \quad \mathsf{R}\text{-a.s.},$$

which together with $H(\mathsf{Q}^\varepsilon|\mathsf{R}) < \infty$ allow us to apply Proposition 3.16 to obtain

$$H(\mathsf{Q}^\varepsilon|\mathsf{R}) \geq H((X_0)_\#\mathsf{Q}^\varepsilon|(X_0)_\#\mathsf{R}) + \frac{1}{2}\mathbb{E}_{\mathsf{Q}^\varepsilon}\left[\int_0^1 |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right].$$

In particular, since $H((X_0)_\#\mathsf{Q}^\varepsilon|(X_0)_\#\mathsf{R}) \geq 0$ and $H(\mathsf{Q}^\varepsilon|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R})$, we get

$$\mathbb{E}_{\mathsf{Q}^\varepsilon}\left[\int_0^1 |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

By putting the indicator $\mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}$ inside the expectation, the inequality is preserved

$$\mathbb{E}_{\mathsf{Q}^\varepsilon}\left[\mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\int_0^1 |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

We apply Fatou's lemma

$$\mathbb{E}_{\mathsf{R}}\left[\int_0^1 \liminf_{\varepsilon\to 0^+} \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\frac{\mathrm{d}\mathsf{Q}^\varepsilon}{\mathrm{d}\mathsf{R}}|\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

Note that

$$\lim_{\varepsilon\to 0^+} \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\frac{\mathrm{d}\mathsf{Q}^\varepsilon}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}, \quad \mathsf{R}\text{-a.s.},$$

which means, since it is positive, that it does not affect the lim inf so that we can pull it out and be left with

$$\mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}\int_0^1 \liminf_{\varepsilon\to 0^+} |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

We change the measure to P and note that $\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} > 0\}$ has full measure under P so that

$$\mathbb{E}_{\mathsf{P}}\left[\int_0^1 \liminf_{\varepsilon\to 0^+} |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 \,\mathrm{d}s\right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

Note that $h(X_s,s) > 0$ for $s < t_k$ by the definition of h (4.12) and the relation to Z_t (4.11) which is positive P-a.s., since $Z_t = \mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R}$ on $\mathcal{F}_t$. Therefore, for $s < t_k$ we can take the limit as follows:

$$\begin{aligned}
\liminf_{\varepsilon\to 0^+} |\sigma(X_s,s)^\top \nabla_x \log(h(X_s,s)+\varepsilon)|^2 &= \liminf_{\varepsilon\to 0^+} \left|\sigma(X_s,s)^\top \frac{\nabla_x h(X_s,s)}{h(X_s,s)+\varepsilon}\right|^2 \\
&= \left|\sigma(X_s,s)^\top \frac{\nabla_x h(X_s,s)}{h(X_s,s)}\right|^2. \\
&= \left|\sigma(X_s,s)^\top \nabla_x \log h(X_s,s)\right|^2.
\end{aligned}$$

For $s \geq t_k$ $h(\cdot, s) = 0$ which implies $\nabla_x \log(h(\cdot, s) + \varepsilon) = \nabla_x \log h(\cdot, s) = 0$ and we see that the limit exists for this case as well. Hence

$$\mathbb{E}_{\mathsf{P}} \left[\int_0^1 |\sigma(X_s, s)^\top \nabla_x \log h(X_s, s)|^2 \, \mathrm{d}s \right] \leq 2H(\mathsf{P}|\mathsf{R}).$$

This is very useful since we want to apply stochastic dominated convergence theorem (Theorem 2.12 from Chapter 4 in [35]). First note from what we have previously seen

$$\lim_{\varepsilon \to 0^+} \sigma(X_s, s)^\top \nabla_x \log(h(X_s, s) + \varepsilon) = \sigma(X_s, s)^\top \nabla_x \log h(X_s, s) \quad \mathsf{P}\text{-a.s.}$$

Moreover by working out the gradient we also see that the limiting random variable is the dominating process. Therefore since M is a semimartingale under P we get by stochastic dominated convergence

$$\sup_{t \in [0,1]} \left| \int_0^t \nabla_x \log(h(X_s, s) + \varepsilon) - \nabla_x \log h(X_s, s) \, \mathrm{d}M_s \right| \xrightarrow{\mathsf{P}} 0.$$

There exists a subsequence $\varepsilon_n \to 0$ for which the convergence holds P-almost surely, i.e.

$$\lim_{n \to \infty} \sup_{t \in [0,1]} \left| \int_0^t \nabla_x \log(h(X_s, s) + \varepsilon_n) - \nabla_x \log h(X_s, s) \, \mathrm{d}M_s \right| = 0, \quad \mathsf{P}\text{-a.s.}$$

In particular

$$\lim_{n \to \infty} \int_0^1 \nabla_x \log(h(X_s, s) + \varepsilon_n) \, \mathrm{d}M_s = \int_0^1 \nabla_x \log h(X_s, s) \, \mathrm{d}M_s, \quad \mathsf{P}\text{-a.s.}$$

There is also the finite variation term, but for that term we apply monotone convergence theorem

$$\lim_{\varepsilon \to 0} \int_0^1 |\sigma(X_s, s)^\top \nabla_x \log(h(X_s, s) + \varepsilon)|^2 \, \mathrm{d}s = \int_0^1 |\sigma(X_s, s)^\top \nabla_x \log h(X_s, s)|^2 \, \mathrm{d}s \quad \mathsf{P}\text{-a.s.}$$

Since P-a.s. convergence holds R-a.s. on the event $\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} > 0\}$, gathering these two results on convergence that we have obtained allows us to conclude (after taking $\varepsilon \to 0^+$ possibly through a subsequence)

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} > 0\}} \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0) \mathcal{E} \left(\int_0^\cdot \nabla_x \log h(X_s, s) \, \mathrm{d}M_s \right)_1, \quad \mathsf{R}\text{-a.s.}$$

Now the uniqueness of the control mentioned in Proposition 3.16 together with Proposition 3.12 and Theorem 3.6 yield the claim. □

Remark 4.20. There are cases where the regularity assumptions on $(h_j)_{1 \leq j \leq k}$ are satisfied. For example, if the reference measure follows a reversible Langevin diffusion as in the Section 4.2 and under the assumptions of Theorem 4.18, we can apply Theorem 4.19 above. Indeed, because then we even get $h_j \in C^\infty(\mathbb{R}^d \times [t_{j-1}, t_j))$ by Theorem 4.10.

Formal connection to Hamilton-Jacobi equations

Let us reflect back on the proof of Theorem 4.19. In all what we have done so far we did not look at the finite variation term α_t^ε. The equality between two integrals in (4.14) deserves special attention. Firstly it implies

$$(\partial_t + \mathcal{L}_t)\log(h_1(X_t,t)+\varepsilon) = -\frac{1}{2}|\sigma(X_t,t)^\top \nabla_x \log(h_1(X_t,t)+\varepsilon)|^2 \quad \mathsf{R}\text{-a.s.} \quad \text{for a.e. } t \in [0,t_1),$$

due to the fact that equality of definite integrals implies equality of integrands. Let us ignore the almost sure issues and just take $\varepsilon \to 0^+$ to obtain

$$(\partial_t + \mathcal{L}_t)\log h_1(x,t) = -\frac{1}{2}|\sigma(x,t)^\top \nabla_x \log h_1(x,t)|^2.$$

It turns out that $u_1 = \log h_1$ solves

$$(\partial_t + \mathcal{L}_t)u_1 = -\frac{1}{2}|\sigma^\top \nabla_x u_1|^2. \tag{4.17}$$

If $h_1 > 0$ we can show that

$$(\partial_t + \mathcal{L}_t)h_1 = 0, \tag{4.18}$$

which is not so surprising, because it is nothing else but the backward equation for the semigroup.

We see from the original definition of h_1 (4.8) that we have the boundary condition at $t = t_1$

$$h_1(x,t_1) = f_1(x)h_2(x,t_2),$$

while for the PDE in (4.17) (by setting $u_2 = \log h_2$)

$$u_1(x,t_1) = \log f_1(x) + u_2(x,t).$$

Of course, this all can be done for h_j for $j = 2,...,k$. We skipped the details before and argued via "induction", but by a similar argument one finds for $j = 1,...,k$

$$\begin{cases} (\partial_t + \mathcal{L}_t)h_j(x,t) = 0, & \text{on } (x,t) \in \mathbb{R}^d \times [t_{j-1},t_j), \\ \qquad h_j(x,t_j) = f_j(x)h_{j+1}(x,t_j), & \text{on } (x,t) \in \mathbb{R}^d \times \{t_j\}, \end{cases} \tag{4.19}$$

Similarly for u_j , $j = 1,...,k$ one finds

$$\begin{cases} (\partial_t + \mathcal{L}_t)u_j(x,t) = -\frac{1}{2}|\sigma^\top \nabla_x u_j|^2, & \text{on } (x,t) \in \mathbb{R}^d \times [t_{j-1},t_j), \\ \qquad u_j(x,t_j) = \log f_j(x) + u_{j+1}(x,t_j), & \text{on } (x,t) \in \mathbb{R}^d \times \{t_j\}. \end{cases} \tag{4.20}$$

At this point it is not clear whether solving the PDEs above will be equivalent to solving the original minimization problem (SMSP). Assuming regularity on the (pre)generator $\mathcal{L}_t$ and the functions $(f_j)_{1\le j\le k}$ we expect that smooth enough solutions exist and they are equivalent to solving the original problem. It turns out that the control is approximated by functions u_j that solve the PDE in a more generalized sense as stated in Section 5.2 below.

4.4 Example: Ornstein–Uhlenbeck process

This final section is devoted to illustrate the theory for a fairly simple example of the multimarginal Schrödinger problem (MSP). The advantage of the simplicity is that we are able to do explicit computations and find the solution explicitly.

We consider the Ornstein-Uhlenbeck process on $\mathbb{R}$. We assume that the reference measure R is the law of

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

with initial distribution being $\eta(\mathrm{d}x) \propto e^{-x^2/2}\,\mathrm{d}x$. We let $\mathcal{T} = \{0,1\}$ and the marginal constraints given by

$$\mu_0 = \mu_1 \sim \mathcal{N}(0,\sigma^2),$$

for some $\sigma > 0$.

Existence of solution to the Schrödinger problem

We want to check the conditions of Theorem 4.3. For that we need to find the measure $\nu := (X_0, X_1)_{\#}\mathsf{R}$. We know the Lebesgue transition density, namely

$$p_t^{\mathrm{Leb}}(x,y) \sim \mathcal{N}\left(xe^{-t}, 1-e^{-2t}\right).$$

Of course, we emphasize the fact that it is a transition density with respect to the Lebesgue measure, because in all the theory above we have used the invariant measure. We can also do that, but it does not really matter, since the transition densities are equivalent: $p_t(x,y) = e^{y^2/2}p_t^{\mathrm{Leb}}(x,y)$. Now we can also verify that the process is reversible using the transition density, but we omit those details.

We have all the ingredients to find ν, we find for any $A \in \mathcal{B}(\mathbb{R}^2)$ that

$$\nu(A) = \int_A \frac{1}{2\pi\sqrt{1-e^{-2}}} \exp\left(-\frac{x^2}{2} - \frac{(y-xe^{-1})^2}{2(1-e^{-2})}\right)\mathrm{d}x\,\mathrm{d}y.$$

It is also straightforward to find

$$\frac{\mathrm{d}\mu_0 \otimes \mu_1}{\mathrm{d}\nu}(x,y) = \sqrt{1-e^{-2}}\exp\left(-\frac{y^2}{2} + \frac{(y-xe^{-1})^2}{2(1-e^{-2})}\right).$$

Therefore

$$H(\mu_0 \otimes \mu_1|\nu) = \int_{\mathbb{R}^2} \log\left(\sqrt{1-e^{-2}}\exp\left(-\frac{y^2}{2} + \frac{(y-xe^{-1})^2}{2(1-e^{-2})}\right)\right)\frac{e^{-(x^2+y^2)/(2\sigma^2)}}{2\pi\sigma^2}\,\mathrm{d}x\,\mathrm{d}y.$$

We see before blinking that the latter integral is finite for being a polynomial integrated with a Gaussian measure which is finite. Hence

$$H(\mu_0 \otimes \mu_1|\nu) < \infty,$$

which together with Theorem 4.3 gives us a unique minimizer μ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x,y) = f(x)g(y), \quad \nu\text{-a.s.},$$

for some measurable functions $f : \mathbb{R} \to [0,\infty)$ and $g : \mathbb{R} \to [0,\infty)$. We immediately wrote ν-a.s., since $\mu_0 \otimes \mu_1 \sim \nu$ and therefore $\mu \sim \nu$.

The reference measure R is reversible so that through reversing the problem and applying Lemma 3.19 we get that since $\mu_0 = \mu_1$, $\mathsf{P} = \mathsf{P} \circ \Phi$ where $\Phi : \Omega \to \Omega$ is time-reversing operator defined by

$$\Phi(\omega)(t) := \omega(1-t).$$

By Theorem B.6 in Appendix B, and the fact that R is reversible we get

$$f(X_0)g(X_1) = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mathsf{P}\circ\Phi}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mathsf{P}\circ\Phi}{\mathrm{d}\mathsf{R}\circ\Phi} = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} \circ \Phi = f(X_1)g(X_0).$$

But we also know that f and g are unique up to transformation of the form $(f,g) \mapsto (cf, g/c)$ for some $c > 0$ by Theorem 4.3. We conclude that $f = cg$ for some c, so that we can write (by possible redefining f)

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x,y) = f(x)f(y), \quad \nu\text{-a.s.} \tag{4.21}$$

Regularity of the Schrödinger factorization

Next we study the regularity of the Schrödinger factorization. Note that we are in the category of reversible Langevin diffusion with $U(x) = \frac{x^2}{2}$. We have $U'(x) = x$ and $U''(x) = 1$. Also for every $\varepsilon > 0$ we take $c_\varepsilon = 1$ so that we get

$$|U''(x)| = 1 \leq \varepsilon|x|^2 + 1 = \varepsilon|U'(x)|^2 + c_\varepsilon.$$

Moreover, U is uniformly convex, since $U''(x) = 1$ so that $U''(x) \geq \theta$ with $\theta = 1$. This all shows Assumption 4.8 is satisfied which gives us that the semigroup $(P_t)_{t\in[0,1]}$ corresponding to the Markov process (X, R) is analytic on $L^p(\mathbb{R}, \eta)$ for any $p \in (1,\infty)$. In fact, we could have verified this ourselves manually using the transition density.

Obviously, we can check whether $f \in L^p(\mathbb{R}, \eta)$ for some $p \in (1,\infty)$ using the theorems above and derive some conditions on σ. For that we need to study the growth of $\rho_0 = \frac{\mathrm{d}\mu_0}{\mathrm{d}\eta}$ (it is enough, because $\rho_0 = \rho_1$). We have

$$\rho_0(x) = \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(x) = \frac{e^{-x^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}}{e^{-x^2/2}/\sqrt{2\pi}} = \frac{1}{\sqrt{\sigma^2}} \exp\left(\frac{\sigma^2 - 1}{2\sigma^2} x^2\right). \tag{4.22}$$

Recall also the function $\kappa(\theta)$ from Theorem 4.16 which in our case is

$$\kappa(\theta) = \kappa(1) = \frac{1}{e-1}.$$

We have $k = 2$ and the critical value $\overline{p}_{\text{log-Harnack}}$ according to Theorem 4.16 is

$$\overline{p}_{\text{log-Harnack}} = \frac{1}{2\left(\frac{\sigma^2-1}{2\sigma^2}\right) + \frac{1}{e-1}} = \frac{\sigma^2(e-1)}{(\sigma^2-1)(e-1)+\sigma^2}, \tag{4.23}$$

for $\sigma^2 \geq e^{-1}(e-1)$. In such case $f \in L^p(\mathbb{R}, \eta)$ for all $p < \overline{p}_{\text{log-Harnack}}$. Otherwise if $\sigma^2 < e^{-1}(e-1)$ we have that $f \in L^\infty(\mathbb{R}, \eta)$.

We may also wonder whether f is $C^\infty(\mathbb{R})$, because $U \in C^\infty(\mathbb{R})$ which allows us to apply Theorem 4.18. For that we want to check whether

$$\frac{\sigma^2-1}{2\sigma^2} < \frac{1}{2}\left(1 - \frac{1}{e-1}\right),$$

which is satisfied if and only if

$$\sigma^2 < e - 1 \approx 1.72.$$

Therefore, if $\sigma^2 < e-1$, we get $f \in L^1(\mathbb{R}, \eta) \cap C^\infty(\mathbb{R})$ by Theorem 4.18. Actually in such case we have a little more than L^1 integrability meaning that we can also apply Theorem 4.19 (see also Remark 4.20) to get that the SDE that X solves under the minimizer P is

$$\mathrm{d}X_t = (-X_t + 2\nabla_x \log h(X_t, t))\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

with h being $h(x,t) = P_{1-t}f(x)$.

Verification by solving the problem explicitly

It follows from (4.21) that we need to seek for a function f such that

$$\rho_0(x) = f(x)\int_{\mathbb{R}} f(y) p_1^{\text{Leb}}(x,y)\,\mathrm{d}y. \tag{4.24}$$

Recall that $\rho_0(x)$ is a Gaussian function, i.e. $\propto e^{cx^2}$ for some $c \in \mathbb{R}$, and so is $p_1^{\text{Leb}}(x,y)$ in both variables. If we take a Gaussian function in both variables (x,y) and integrate with respect to one variable, say y, then by completing the squares we still get a Gaussian function. If we take f to also be Gaussian then we get that

$$x \mapsto f(x)\int_{\mathbb{R}} f(y) p_1^{\text{Leb}}(x,y)\,\mathrm{d}y$$

is Gaussian.

That leads us to try an ansatz of the form $f(x) = \sqrt{C}e^{\theta x^2}$ for some $\theta \in \mathbb{R}$

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x,y) = Ce^{\theta x^2 + \theta y^2}. \tag{4.25}$$

For this to be the right solution, we must have by Equation (4.24)

$$\rho_0(x) = \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(x) = Ce^{\theta x^2}\int_{\mathbb{R}} e^{\theta y^2} p_1^{\text{Leb}}(x,y)\,\mathrm{d}y, \tag{4.26}$$

by (4.4). Let us see how the latter integral looks like

$$\int_{\mathbb{R}} e^{\theta y^2} p_1^{\mathrm{Leb}}(x,y)\,\mathrm{d}y = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-e^{-2})}} \exp\left(\theta y^2 - \frac{(y-xe^{-1})^2}{2(1-e^{-2})}\right)\,\mathrm{d}y.$$

This is a Gaussian integral and by expanding the squares one finds that it converges if and only if

$$\theta < \frac{1}{2(1-e^{-2})}.$$

Moreover, completing the squares allows us to find the integral explicitly:

$$\int_{\mathbb{R}} e^{\theta y^2} p_1^{\mathrm{Leb}}(x,y)\,\mathrm{d}y = \exp\left(\left(\frac{\theta}{e^2(1-2\theta)+\theta}\right)x^2\right)\sqrt{\frac{2\pi(e^2-1)}{e^2+2\theta-2e^2\theta}}. \tag{4.27}$$

Combining (4.22), (4.26) and (4.27) we see that the following equality should hold:

$$\frac{1}{\sqrt{\sigma^2}}\exp\left(\left(\frac{1}{2}-\frac{1}{2\sigma^2}\right)x^2\right) = C\exp\left(\theta x^2 + \left(\frac{\theta}{e^2+2\theta(1-e^2)}\right)x^2\right)\sqrt{\frac{2\pi(e^2-1)}{e^2+2\theta-2e^2\theta}}.$$

We see that the constant should be

$$C = \sqrt{\frac{e^2+2\theta-2e^2\theta}{2\pi\sigma^2(e^2-1)}},$$

and θ must satisfy

$$\frac{1}{2}-\frac{1}{2\sigma^2} = \theta + \frac{\theta}{e^2+2\theta(1-e^2)}.$$

Solving for θ gives us two solutions, but the integrability condition $\theta < \frac{1}{2(1-e^{-2})}$ shows that the only solution is

$$\theta = \frac{1+e^2(2\sigma^2-1)-\sqrt{1+e^4+2e^2(2\sigma^4-1)}}{4(e^2-1)\sigma^2}.$$

By construction this θ makes μ defined by (4.25) feasible. It is not difficult to see that $\log f \in L^1(\mathbb{R},\mu_0)$ so that by Theorem 2.6 we get that μ is the unique minimizer.

Now we can check the regularity and integrability of f. Basic algebra shows that

$$\theta(\sigma) < \frac{e}{2+2e} \approx 0.366, \qquad \text{for all } \sigma > 0,$$

which shows that we definitely have $f \in L^1(\mathbb{R},\eta)$ for any $\sigma > 0$. And obviously $f \in C^\infty(\mathbb{R})$. So the last assertion that we have seen in the previous part is not only true if $\sigma^2 < e-1$, but it is always true.

We also see that $\theta(\sigma)$ goes to the upper bound as $\sigma \to \infty$ which means that we cannot have $f \in L^p(\mathbb{R}, \eta)$ for all $p \in (1, \infty)$. There is dependence on σ which can be found by studying the expression for $\theta(\sigma)$ carefully. One sees that $f \in L^p(\mathbb{R}, \eta)$ for $p \in (1, \infty)$ if

$$p\theta(\sigma) < \frac{1}{2},$$

which gives us the exact critical value $\overline{p}_{\text{exact}}$

$$\overline{p}_{\text{exact}} = \frac{2(e^2 - 1)\sigma^2}{1 + e^2(2\sigma^2 - 1) - \sqrt{1 + e^4 + 2e^2(2\sigma^4 - 1)}}, \tag{4.28}$$

for $\sigma^2 \geq 1$ and $\overline{p}_{\text{exact}} = \infty$ otherwise.

Comparison of integrability via different methods

We want to compare the L^p-integrability of f obtained via Theorem 4.16 with the exact calculation. We have remarked several times that the estimations used in Theorem 4.16 are rough. For instance Lemma 4.15 depends on log-Harnack inequality which may not be optimal. Note that we know the transition density p_t^{Leb} exactly, so the following question arises: could we get a better integrability by deriving a “better” lower bound of the form $c_0\eta \otimes c_1\eta \leq \nu$ ourselves? The answer is positive, we can do that.

To that end, recall that for any $A = A_1 \times A_1 \in \mathcal{B}(\mathbb{R}^2)$

$$\nu(A) = \int_A \frac{1}{2\pi\sqrt{1 - e^{-2}}} \exp\left(-\frac{x^2}{2} - \frac{(y - xe^{-1})^2}{2(1 - e^{-2})}\right) \mathrm{d}x \, \mathrm{d}y.$$

We want to estimate the argument of the exponential through Young’s inequality with some $\varepsilon > 0$

$$-\frac{x^2}{2} - \frac{(y - xe^{-1})^2}{2(1 - e^{-2})} = -\frac{x^2}{2} - \frac{y^2 - 2xye^{-1} + xe^{-2}}{2(1 - e^{-2})} \geq -\frac{x^2}{2} - \frac{y^2 + \varepsilon x^2 e^{-2} + \varepsilon^{-1} y^2 + xe^{-2}}{2(1 - e^{-2})}.$$

This leads to

$$\exp\left(-\frac{x^2}{2} - \frac{(y - xe^{-1})^2}{2(1 - e^{-2})}\right) \geq \exp\left(-\frac{x^2}{2} - \frac{y^2}{2}\right) c(x, y, \varepsilon),$$

with $c(x, y, \varepsilon) = c_0(x, \varepsilon)c_1(x, \varepsilon)$ defined by

$$c_0(x, \varepsilon) = \exp\left(-\frac{1 + \varepsilon}{2(e^2 - 1)} x^2\right), \quad c_1(x, \varepsilon) = \exp\left(-\frac{1 + \varepsilon^{-1} e^2}{2(e^2 - 1)} y^2\right).$$

Now that we have these functions c_0 and c_1 we derive in a similar way as we have done in Theorem 4.16 that the critical value $\overline{p}_{\text{manual}}$ is

$$\overline{p}_{\text{manual}} = \frac{\sigma^2(e^2 - 1)}{(\sigma^2 - 1)(e^2 - 1) + \sigma^2}, \tag{4.29}$$

if $\sigma^2 \geq e^{-2}(e^2-1)$; and $\overline{p}_{\text{manual}} = \infty$ otherwise.

Finally, now we can compare the different critical values of $\overline{p}$: the one obtained via Theorem 4.16 (4.23) which is based on log-Harnack inequality, the exact value (4.28) and the one obtained manually (4.29). The result can be seen below in Figure 4.1. Everything is as we would expect, for example the exact is obviously the best, and the manually obtained is better than the one based on log-Harnack. We also remark that the dashed line is the line $\sigma^2 = 1$ where f becomes the function 1 so that is why the exact is tending to infinity there, because f is becoming $L^\infty(\mathbb{R}, \eta)$.

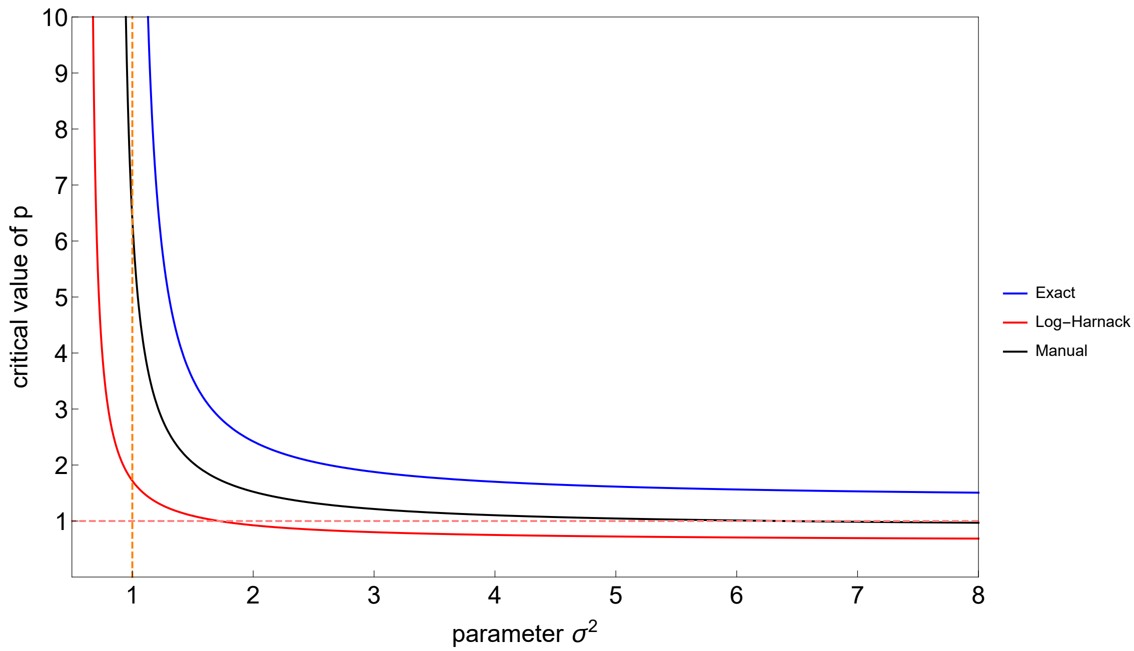

Figure 4.1: Comparison of $\overline{p}$ obtained by log-Harnack and manual calculation with the exact value as a function of σ^2

Chapter 5

Fully prescribed marginal constraints

This chapter is devoted to the study of the multimarginal Schrödinger problem (MSP) with fully prescribed marginals, i.e. $\mathcal{T} = [0,1]$. We first give some heuristics, and then we introduce an equivalent weak formulation of the problem and write its consequences. At the end of the chapter we prove that the weak problem is equivalent to the original problem.

Let us roughly recall the setting which we will formalize soon. As we had in Chapter 3, the canonical process X under the reference measure R is a weak solution to (or more precisely the law of)

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t^{\mathsf{R}}, \tag{5.1}$$

where B^{R} is a BM in a possibly enlarged space (see Theorem 21.7 in [22] for the enlargement of the space). We assume that the probability measures $(\mu_t)_{t\in[0,1]}$ are the marginals of some stochastic process, i.e. $\mu_t = (X_t)_\# \mathsf{Q}$, where Q is such that X is a weak solution to

$$\mathrm{d}X_t = c(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t^{\mathsf{Q}}, \tag{5.2}$$

where B^{Q} is a BM in a possibly enlarged space.

Although, these assumptions are not precise yet, we may start thinking about ways to solve (MSP) with $\mathcal{T} = [0,1]$ and given marginals $(\mu_t)_{t\in[0,1]}$. We give some heuristics which turn out to be accurate.

Some heuristics

We assume that Q itself is a competitor for the Schrödinger problem (MSP) so that by Theorem 3.6 there exists a Borel measurable function $v : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ such that

$$\mathrm{d}X_t = (b(X_t, t) + a(X_t, t)v(X_t, t))\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t^{\mathsf{Q}}.$$

It is not clear why such function v exists yet, but one may argue because the form that the SDE X "solves" under Q in (5.2).

On the other hand by Theorem 3.14 we know that we are after minimizing over probability measures P such that X is a weak solution to

$$\mathrm{d}X_t = (b(X_t,t) + a(X_t,t)u(X_t,t))\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}B_t^{\mathsf{P}}, \tag{5.3}$$

for some Borel measurable function $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$. In such case we have by Tonelli's theorem

$$\begin{aligned} H(\mathsf{P}|\mathsf{R}) &= H(\mu_0|\nu_0) + \frac{1}{2}\mathbb{E}_{\mathsf{P}}\left[\int_0^1 |\sigma(X_t,t)^\top u(X_t,t)|^2 \mathrm{d}t\right] \\ &= H(\mu_0|\nu_0) + \frac{1}{2}\int_0^1\int_{\mathbb{R}^d} |\sigma(x,t)^\top u(x,t)|^2\,\mu_t(\mathrm{d}x)\,\mathrm{d}t. \end{aligned}$$

Apparently, it is enough to minimize over functions u. However, it is not clear how we ensure that weak solutions to (5.3) have the prescribed marginals. We want to figure out ways to distinguish those functions u for which solutions to (5.3) yield the right marginals from the ones that do not.

To that end, we use what we already know, namely the weak forward equation also known as the Fokker-Planck equation (see Proposition 21.6 in [40]) corresponding to the SDE that X solves under Q which is

$$\partial_t \mu_t = -\operatorname{div}((b + av)\mu_t) + \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}\mu_t).$$

Similarly, the weak forward equation corresponding to P is

$$\partial_t \mu_t = -\operatorname{div}((b + au)\mu_t) + \frac{1}{2}\sum_{ij}\partial_{ij}(a_{ij}\mu_t).$$

Subtracting the previous two equations yields

$$\operatorname{div}(a(u-v)\mu_t) = 0,$$

which is understood as

$$\int_0^1\int_{\mathbb{R}^d} \nabla_x\varphi\cdot a(u-v)\,\mu_t(\mathrm{d}x)\,\mathrm{d}t = 0, \qquad \text{for all } \varphi \in C_c^\infty(\mathbb{R}^d\times\mathbb{R}). \tag{5.4}$$

We see, at least on a heuristic level, that we must have $\operatorname{div}(a(u-v)\mu_t) = 0$ for P given as the law of (5.3) to satisfy the marginal constraints. This turns out to be remarkably close to the condition that we will state below.

5.1 Fully prescribed Schrödinger problem and the weak formulation

The setting of the fully prescribed Schrödinger problem seems slightly different from the previous settings, although there is no much difference in practice (in fact it is equivalent).

The main difference is that we take the time-coordinate as a process. We will work with the canonical process $(X_t, r_t)_{t\geq 0}$ on the space $\Omega := C([0,1]; \mathbb{R}^d \times \mathbb{R})$ endowed with Borel σ-algebra induced by the supremum norm. Again, this is different from how we have defined Ω before. We set $\mathcal{F}_t = \sigma((X_t, r_t) : \ s \in [0,t])$. Let us define the (pre)generator

$$L := \partial_t + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x,t)\partial_{ij} + \sum_{i=1}^{d} b_i(x,t)\partial_i.$$

The reason why we took $(X_t, r_t)_{t\geq 0}$ instead of $(X_t)_{t\geq 0}$ is that $(X_t, r_t)_{t\geq 0}$ is time-homogeneous Markov processes while $(X_t)_{t\geq 0}$ alone can be time-in-homogeneous. It is more convenient to work with time-homogeneous Markov processes, because we want to use semigroups later on which will be neater when dealing with time-homogeneous Markov processes.

We impose the following assumptions on the generator:

Assumption 5.1.

(i) $a = \sigma\sigma^\top$ for some $d \times d'$ matrix-valued Borel measurable function $\sigma : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^{d\times d'}$;

(ii) b is Borel measurable;

(iii) there exists $C > 0$ such that $|a(x,t)| \leq C(1 + |x|^2)$, and $|b(x,t)| \leq C(1 + |x|)$ for all $(x,t) \in \mathbb{R}^d \times \mathbb{R}_+$;

(iv) $\mathrm{MP}(L)$ satisfies the semi-uniqueness condition (see Definition 3.5).

We impose the following assumptions on the reference measure:

Assumption 5.2. We assume that we have a Markovian family of probability measures $(\mathsf{R}^{x,r})_{x\in\mathbb{R}^d, t\geq 0}$ such that

(i) $\mathsf{R}^{x,r}(X_0 = x,\ r_0 = r) = 1$ for all $(x,r) \in \mathbb{R}^d \times \mathbb{R}_+$;

(ii) $\mathsf{R}^{x,r}$ is an extremal solution to $\mathrm{MP}(L, \delta_x \otimes \delta_r)$.

Remark 5.3. One can easily see that the set of solutions to a martingale problem is a convex set. An extremal solution $\mathsf{R}^{x,r}$ is then a solution that cannot be written as $\mathsf{R}^{x,r} = \alpha\mathsf{R}_1^{x,r} + (1-\alpha)\mathsf{R}_2^{x,r}$ for some $\alpha \in (0,1)$ and $\mathsf{R}_1^{x,r} \neq \mathsf{R}_2^{x,r}$. It is just the usual definition of extremal points of a convex set. In particular, if the martingale problem has a unique solution, then extremality is satisfied. For instance, it is true if σ and b are locally Lipschitz in the spatial variable x and uniformly in time as given in Theorem 19.26 in [40].

With these notations in mind the usual reference measure becomes $\mathsf{R} = \mathsf{R}^{\nu_0\otimes\delta_0}$, and under R we have that X is a semimartingale with the decomposition $X = X_0 + M + A$ where M is a local martingale and

$$A_t = \int_0^t b(X_s, s)\,\mathrm{d}s, \quad [M]_t = \int_0^t a(X_s, s)\,\mathrm{d}s. \tag{5.5}$$

The multimarginal Schrödinger problem comes with probability measures $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}(\mathbb{R}^d)$. Assume that $\mu_t = (X_t)_\#\mathsf{Q}$ for some probability measure $\mathsf{Q} \in \mathcal{P}(\Omega)$ such that X has the semimartingale decomposition $X = X_0 + M^{\mathsf{Q}} + A^{\mathsf{Q}}$ with

$$A_t^{\mathsf{Q}} = \int_0^t c(X_s, s)\,\mathrm{d}s, \quad [M^{\mathsf{Q}}]_t = \int_0^t a(X_s, s)\,\mathrm{d}s. \tag{5.6}$$

Note that this is in line with (5.2).

We can formulate the fully prescribed Schrödinger problem for the sake of completeness:

Fully prescribed Schrödinger problem.

$$\begin{aligned} \min \quad & H(\mathsf{P}|\mathsf{R}), \\ \text{subject to} \quad & (X_t)_\#\mathsf{P} = \mu_t, \qquad \text{for all } t \in [0,1]. \end{aligned} \tag{FPSP}$$

As usual we denote by $\mathcal{C}^{\mathrm{FPSP}}$ the set of feasible probability measures and $\mathcal{C}_H^{\mathrm{FPSP}}$ the set of competitors of (FPSP).

When Q itself is a competitor, then we have the existence of some process β^{Q} by Theorem 3.6 such that $X = X_0 + \tilde{A}^{\mathsf{Q}} + \tilde{M}^{\mathsf{Q}}$ with

$$\tilde{A}_t^{\mathsf{Q}} = \int_0^t (b(X_s, s) + a(X_s, s)\beta_s^{\mathsf{Q}})\,\mathrm{d}s,$$

and

$$H(\mathsf{Q}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\int_0^1 \mathbb{E}_{\mathsf{Q}}\left[|\sigma(X_s, s)^\top \beta_s|^2\right]\mathrm{d}s.$$

We want to argue that β_t^{Q} can be written in terms of (X_s, s). We can do that by using the Moore-Penrose inverse of the matrix $a(x,t)$. Firstly, by the uniqueness of the semimartingale decomposition and Lebesgue differentiation we get for a.e. $t \in [0,1]$

$$b(X_t, t) + a(X_t, t)\beta_t^{\mathsf{Q}} = c(X_t, t), \quad \mathsf{Q}\text{-a.s.}$$

Note that we must necessarily have $c(X_t, t) - b(X_t, t) \in \mathrm{Range}(a(X_t, t))$ Q-a.s. which allows us to apply Lemma 3.10 to get a Borel measurable function $v : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ explicitly given by

$$v(x,t) := a(x,t)^+(c(x,t) - b(x,t)), \tag{5.7}$$

where $a(x,t)^+$ is the Moore-Penrose inverse of the matrix $a(x,t)$ satisfying $a(X_t,t)v(X_t,t) = a(X_t,t)\beta_t^{\mathsf{Q}}$. Now we apply Proposition 3.12 to get $v(X_\cdot, \cdot) = \beta^{\mathsf{Q}}$ in $L^2(\Omega \times [0,1], a, \mathsf{Q} \otimes \mathrm{Leb})$ (see the notation in (3.9)).

As one expects from Theorem 3.14 we have to look for solutions P such that the semimartingale $X = X_0 + A^{\mathsf{P}} + M^{\mathsf{P}}$ has finite variation part A^{P} and local martingale M^{P} given by

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s, s) + a(X_s, s)u(X_s, s))\,\mathrm{d}s, \quad M_t^{\mathsf{P}} = M_t - \int_0^t a(X_s, s)u(X_s, s)\,\mathrm{d}s. \tag{5.8}$$

In such case we have

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\int_0^1 \int_{\mathbb{R}^d} |\sigma^\top(x,t)u(x,t)|^2 \, \mu_t(\mathrm{d}x)\,\mathrm{d}t.$$

Recall the space $L^2(\mathbb{R}^d \times [0,1], a, \mu \otimes \mathrm{Leb})$ defined in (3.9) which we abbreviate as $L^2(a,\mu)$ since we will use it a lot in the sequel. Basically we are after minimizing

$$u \mapsto \int_0^1 \int_{\mathbb{R}^d} |\sigma^\top u|^2 \, \mu_t(\mathrm{d}x)\,\mathrm{d}t = \|u\|^2_{L^2(a,\mu)}.$$

What we wrote here is nothing else but the rigorous version of things already said in the heuristics above.

We want to describe the set of functions u such that P characterized by (5.8) is feasible for (FPSP). To describe that set, we first need to define a generalized notion of a gradient not to be confused with distributional gradient.

We use the following notation for Borel measurable functions

$$\mathfrak{B}(\mathbb{R}^d \times \mathbb{R}) := \{f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R} \text{ with } f \text{ Borel measurable }\},$$

and

$$\mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R}) := \{f \in \mathfrak{B}(\mathbb{R}^d \times \mathbb{R}) : \ f \text{ bounded }\}.$$

Definition 5.4 (Extended domain of the generator)**.** A function $f \in \mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ is said to be in the extended domain of L and denoted by $f \in \mathcal{D}_e$ if there exists a function $g \in \mathfrak{B}(\mathbb{R}^d \times \mathbb{R})$ such that

$$\int_0^1 |g(X_s, r_s)|\,\mathrm{d}s < \infty, \quad \mathsf{R}^{x,r}\text{-a.s., for all } (x,r) \in \mathbb{R}^d \times \mathbb{R}_+,$$

and

$$N_t = f(X_t, r_t) - f(X_0, r_0) - \int_0^t g(X_s, r_s)\,\mathrm{d}s,$$

is a $\mathsf{R}^{x,r}$-local martingale for all $(x,r) \in \mathbb{R}^d \times \mathbb{R}_+$. Moreover, if $f \in \mathcal{D}_e$ we set $Lf := g$.

This definition leads immediately to the following proposition that states that for any $f \in \mathcal{D}_e$ we recover a formula similar to Itô's formula:

Proposition 5.5. *For any $f \in \mathcal{D}_e$, there exists a Borel measurable function $\overline{\nabla} f \in \mathfrak{B}(\mathbb{R}^d \times \mathbb{R})$ such that the $\mathsf{R}^{x,r}$-local martingale N given in the previous definition can be written as*

$$N_t = \int_0^t \overline{\nabla} f(X_s, r_s)\,\mathrm{d}M_s.$$

In particular we have

$$f(X_t, r_t) - f(X_0, r_0) = \int_0^t \overline{\nabla} f(X_s, r_s)\,\mathrm{d}M_s + \int_0^t Lf(X_s, r_s)\,\mathrm{d}s.$$

Proof. We give most of the proof except some few details that can be easily filled in. We follow the same arguments of Exercise 2.25 in Chapter 10 in [35]. We use Theorem 2 in [20] (or Theorem 2.7 in [43]) which states that the extremality of $\mathsf{R}^{x,r}$ implies that every $\mathsf{R}^{x,r}$-local martingale can be written as a stochastic integral with respect to M. In particular, for the local martingale N it means that there exists a progressively measurable process $(h_s)_{s\geq 0}$ such that

$$N_t = \int_0^t h_s \,\mathrm{d}M_s = \sum_{i=1}^d \int_0^t h_s^{(i)} \,\mathrm{d}M_s^{(i)}.$$

Secondly, we argue that for any $j \in \{1, ..., d\}$ the bracket process $[N, M^{(j)}]$ is an additive functional as defined in Definition 1.1 of Chapter 10 in [35]. Define the canonical shift $\Theta_t : \Omega \to \Omega$ by $\Theta_t(\omega) := \omega(t + \cdot)$. In our case, the additive functional property is

$$[N, M^{(j)}]_{t+\varepsilon} = [N, M^{(j)}]_t + [N, M^{(j)}]_\varepsilon \circ \Theta_t. \tag{5.9}$$

We can see the equality holds by the fact that (see Theorem 1.8 from Chapter 4 in [35])

$$[N, M^{(j)}]_t = \lim_{n\to\infty} \sum_{i=1}^{k_n} (N_{t_{k+1}^n} - N_{t_k^n})(M^{(j)}_{t_{k+1}^n} - M^{(j)}_{t_k^n}),$$

for some partition of $[0, t]$ denoted by $\Gamma_n(t) = \{t_1^n < t_2^n < .... < t_{k_n}^n\}$ and the limit is uniformly for $t \in [0, 1]$ P-almost surely. Then for a partition of $[0, t + \varepsilon]$ and $t \leq t_k$ we can write

$$\begin{aligned} M^{(j)}_{t_{k+1}} - M^{(j)}_{t_k} &= X_{t_{k+1}} - X_{t_k} - \int_{t_k}^{t_{k+1}} b(X_s, r_s) \,\mathrm{d}s \\ &= \left(X_{t_{k+1}-t} - X_{t_k-t} - \int_{t_k-t}^{t_{k+1}-t} b(X_s, r_s) \,\mathrm{d}s \right) \circ \Theta_t. \end{aligned}$$

Then $\{t_k - t\}_k$ for $t_k > t$ is a partition of $[0, \varepsilon]$ and the previous shift holds; and the remaining $\{t_k\}_k$ for $t_k \leq t$ is partition of $[0, t]$. We can do the same for N_t since it can also be written in terms of $(X_t, r_t)_{t\in[0,1]}$ as done for $M_t^{(j)}$. This is all that is needed to be convinced that (5.9) holds.

Using the notation a_j for the j-th row we can write

$$[N, M^{(j)}]_{t+\varepsilon} = \int_0^{t+\varepsilon} a_j(X_s, r_s)^\top h_s \,\mathrm{d}s = \int_0^t a_j(X_s, r_s)^\top h_s \,\mathrm{d}s + \int_t^{t+\varepsilon} a_j(X_s, r_s)^\top h_s \,\mathrm{d}s.$$

But by (5.9) we know that

$$[N, M^{(j)}]_{t+\varepsilon} = \int_0^t a_j(X_s, r_s)^\top h_s \,\mathrm{d}s + \int_0^\varepsilon \left[a_j(X_s, r_s)^\top h_s\right] \circ \Theta_t \,\mathrm{d}s.$$

Hence equating the previous two equations and dividing by $\varepsilon > 0$ yields

$$\frac{1}{\varepsilon} \int_t^{t+\varepsilon} a_j(X_s, r_s)^\top h_s \,\mathrm{d}s = \frac{1}{\varepsilon} \int_0^\varepsilon \left[a_j(X_s, r_s)^\top h_s\right] \circ \Theta_t \,\mathrm{d}s.$$

Taking $\varepsilon \to 0^+$ yields by Lebesgue differentiation yields

$$a_j(X_t, r_t)^\top h_t = \left[a_j^\top(X_0, r_0)h_0\right] \circ \Theta_t.$$

We know that the left-hand side is adapted for h being progressively measurable, which makes us want to take conditional expectation with respect to $\mathcal{F}_t$ as soon as we have integrability. A trick to avoid integrability issues is to apply a bounded invertible continuous function $\varphi : \mathbb{R} \to \mathbb{R}$ (e.g. $\varphi = \arctan$) as follows:

$$\varphi\left(a_j(X_t, r_t)^\top h_t\right) = \mathbb{E}_{\mathsf{R}}\left[\varphi\left(\left[a_j(X_0, r_0)^\top h_0\right] \circ \Theta_t\right) \mid \mathcal{F}_t\right].$$

Obviously we have

$$\varphi\left(\left[a_j(X_0, r_0)^\top h_0\right] \circ \Theta_t\right) = \left[\varphi\left(a_j(X_0, r_0)^\top h_0\right)\right] \circ \Theta_t.$$

We have $\varphi\left(a_j(X_0, r_0)^\top h_0\right)$ is a bounded measurable function on $(\Omega, \mathcal{F})$ and thus by the Markov property we have (see Proposition from Chapter 3 in [35])

$$\varphi\left(a_j(X_t, r_t)^\top h_t\right) = \mathbb{E}_{\mathsf{R}^{X_t, r_t}}\left[\varphi\left(a_j(X_0, r_0)^\top h_0\right)\right].$$

By inverting φ we get

$$a_j(X_t, r_t)^\top h_t = \varphi^{-1}\left(\mathbb{E}_{\mathsf{R}^{X_t, r_t}}\left[\varphi\left(a_j(X_0, r_0)^\top h_0\right)\right]\right) =: G_j(X_t, r_t).$$

for some measurable function $G_j : \Omega \to \mathbb{R}$. We can do this for all $j \in \{1, ..., d\}$ so that by the symmetry of a we get

$$a(X_t, r_t)h_t = G(X_t, r_t),$$

with $G = (G_1, ..., G_d)$. Therefore $G \in \mathrm{Range}(a(X_t, r_t))$ meaning that we can apply Lemma 3.10 to get a Borel measurable function $F : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ satisfying

$$a(X_t, t)F(X_t, t) = G(X_t, t) = a(X_t, t)h_t.$$

But by Proposition 3.12 we see that the stochastic integrals $\int h_s \, \mathrm{d}M_s = \int F(X_s, r_s) \, \mathrm{d}M_s$ coincide. We set

$$F(X_t, r_t) = a(X_t, r_t)^+ G(X_t, r_t) =: \overline{\nabla} f(X_t, r_t).$$

The rest of the assertion can be obtained by the definition of $f \in \mathcal{D}_e$. $\square$

Remark 5.6. One must always remember that $\overline{\nabla} f$ is just a notation and may in general not be a classical gradient. Moreover, $a\overline{\nabla} f$ is unique $(X_t)_\# \mathsf{R} \otimes \mathrm{Leb}$-a.e. as follows from a similar way of differentiating the quadratic variation. It turns out that such uniqueness is the most that we could ask for due to Proposition 3.12. We abuse notation and say that $\overline{\nabla} f = \hat{\nabla} f$ for any $\hat{\nabla} f$ satisfying the statement in Proposition 5.5.

Itô's formula implies that $C_c^\infty(\mathbb{R}^d \times \mathbb{R}) \subset \mathcal{D}_e$ and $\overline{\nabla} f = \nabla_x f$ as one can easily verify with Itô's formula.

There is a subset of $\mathcal{D}_e$ which we call the μ-good functions which is the subset of functions that we can work with later on.

Definition 5.7 (μ-good function)**.** We say a function $f \in \mathcal{D}_e$ is μ-good denoted by $f \in \mathcal{D}_{e,\mu}$ if

(i) $\int_0^1 \int_{\mathbb{R}^d} |Lf(x,t)| \, \mu_t(\mathrm{d}x) \, \mathrm{d}t < \infty$;

(ii) $\overline{\nabla} f \in L^2(a, \mu)$.

It is not difficult to see that $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}) \subset \mathcal{D}_{e,\mu}$, since the coefficients of L are locally bounded.

We can define the space of "generalized divergence free" functions with respect to μ_t using the new notion of gradients $\overline{\nabla}$, namely

$$L^2_{\mathrm{div}}(a, \mu) := \{f \in L^2(a, \mu) : \; (\overline{\nabla}\varphi, f)_{L^2(a,\mu)} = 0, \quad \text{for all } \varphi \in \mathcal{D}_{e,\mu}\}.$$

The name div as a subscript makes sense, because if μ_t would have a Lebesgue density $p(x,t)$, then $L^2_{\mathrm{div}}(a, \mu)$ would consists of functions for which the following holds:

$$\mathrm{div}_x(afp) = 0 \quad \text{in the weak sense.}$$

It is a little bit more generalized than that, because we integrate against all $\overline{\nabla}\varphi$ which may include more than weak gradients.

Remember also that our heuristics gave us a condition in (5.4) which was almost right except that we must take $\varphi \in \mathcal{D}_{e,\mu}$, or in other words $u - v \in L^2_{\mathrm{div}}(a, \mu)$. Thus, we can finally formulate our newly obtained minimization problem:

Weak multimarginal Schrödinger problem.

$$\begin{aligned} \min \quad & \|u\|^2_{L^2(a,\mu)}, \\ \text{subject to} \quad & u - v \in L^2_{\mathrm{div}}(a, \mu). \end{aligned} \tag{WMSP}$$

Since $u - v \in L^2_{\mathrm{div}}(a, \mu)$ may give some troubles if $v \notin L^2(a, \mu)$, we impose the following assumption:

Assumption 5.8 (Q is a competitor)**.** We assume that $H(\mu_0|\nu_0) < \infty$ and $v \in L^2(a, \mu)$, i.e. Q is a competitor.

Of course one may wonder whether the weak problem (WMSP) is equivalent to the original problem (MSP). It is indeed the case with all assumptions imposed in this section.

Theorem 5.9. *The original formulation* (MSP) *is equivalent to the weak formulation* (WMSP)*. Moreover, the finite variation part can be written in terms of the minimizer $u \in L^2(a, \mu)$ of* (WMSP)

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s, s) + a(X_s, s)u(X_s, s)) \, \mathrm{d}s,$$

and

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)}.$$

The problem (WMSP) is fairly easily solved once one establishes the following result:

Lemma 5.10. *The subset $L^2_{\text{div}}(a,\mu)$ is a closed linear subspace of $L^2(a,\mu)$. In particular, we can write*

$$L^2(a,\mu) = L^2_{\text{div}}(a,\mu) \oplus L^2_{\text{div}}(a,\mu)^{\perp}.$$

Proof. Linearity is clear. To prove closedness, take $f_n \in L^2_{\text{div}}(a,\mu)$ converging to f in $L^2(a,\mu)$. Take $\varphi \in \mathcal{D}_{e,\mu}$, then by definition $\overline{\nabla}\varphi \in L^2(a,\mu)$. Due to the fact that strong convergence implies weak convergence we get

$$(\overline{\nabla}\varphi, f)_{L^2(a,\mu)} = \lim_{n\to\infty} (\overline{\nabla}\varphi, f_n)_{L^2(a,\mu)} = 0.$$

The decomposition of $L^2(a,\mu)$ follows from standard theory on Hilbert spaces (see Remark 5 in Chapter 5 in [5]). $\square$

We denote the projection of $u \in L^2(a,\mu)$ on $L^2_{\text{div}}(a,\mu)$ and $L^2_{\text{div}}(a,\mu)^{\perp}$ by u_{div} and u_{pot} respectively. Moreover we have

$$\|u\|^2_{L^2(a,\mu)} = \|u_{\text{div}}\|^2_{L^2(a,\mu)} + \|u_{\text{pot}}\|^2_{L^2(a,\mu)}.$$

The subscript pot can be understood as functions in $L^2(a,\mu)^{\perp}$ are the closure of $\{\overline{\nabla} f : \ f \in \mathcal{D}_{e,\mu}\}$ in $L^2(a,\mu)$ which are kind of "potentials".

Let us get back to the functional we want to minimize. By simply projection we get the minimizer of $J(u)$ as appears from the following proposition:

Proposition 5.11 (Solution of (WMSP))**.** *There exists a unique minimizer $u^* \in L^2(a,\mu)$ of the problem* (WMSP) *and it is explicitly given by $u^* = v_{\text{pot}}$. The corresponding objective value is $J(u^*) = \|v_{\text{pot}}\|^2_{L^2(a,\mu)}$.*

Proof. We start by expanding the norm as follows:

$$J(u) = \|u\|^2_{L^2(a,\mu)} = \|(u-v)+v\|^2_{L^2(a,\mu)} = \|u-v\|^2_{L^2(a,\mu)} + 2(u-v,v)_{L^2(a,\mu)} + \|v\|^2_{L^2(a,\mu)}.$$

Since $u - v \in L^2_{\text{div}}(a,\mu)$ we have

$$(u-v,v)_{L^2(a,\mu)} = (u-v, v_{\text{div}})_{L^2(a,\mu)}.$$

Moreover we have

$$2(u-v,v_{\text{div}})_{L^2(a,\mu)} \geq -\|u-v\|^2_{L^2(a,\mu)} - \|v_{\text{div}}\|^2_{L^2(a,\mu)}.$$

This all shows

$$J(u) \geq -\|v_{\text{div}}\|^2_{L^2(a,\mu)} + \|v\|^2_{L^2(a,\mu)} = \|v_{\text{pot}}\|^2_{L^2(a,\mu)}.$$

This lower bound is attained by $u^* = v_{\text{pot}}$ which gives the asserted objective value. Of course $u^* - v \in L^2_{\text{div}}(a,\mu)$, because $u^* - v = v_{\text{pot}} - v = v_{\text{div}} \in L^2_{\text{div}}(a,\mu)$. So we found our minimizer. Uniqueness follows by strict convexity of J which can be easily obtained. $\square$

A straightforward consequence of the previous theorem is that when the drifts are decomposed in terms of $L^2_{\text{div}}(a, \mu) \otimes L^2(a, \mu)^\perp$, then we get

Proposition 5.12 (Minimizer for decomposed drifts)**.** *Assume the functions b and c in* (5.5) *and* (5.6) *can be written as follows:*

$$b(x, t) = a(x, t) (\nabla U(x, t) + \beta(x, t)), \qquad c(x, t) = a(x, t) (\nabla V(x, t) + \gamma(x, t)),$$

for some $U, V \in C^{2,1}(\mathbb{R}^d \times [0, 1])$ for which $\nabla U, \nabla V \in L^2(a, \mu)$ and $\beta, \gamma \in L^2_{\text{div}}(a, \mu)$. Then the canonical process X under the unique solution P of (MSP) *has finite variation part*

$$A^{\mathsf{P}}_t = \int_0^t a(X_s, s) (\nabla V(X_s, s) + \beta(X_s, s)) \, \mathrm{d}s,$$

and

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2} \int_0^1 \int_{\mathbb{R}^d} \left|\sigma(x, t)^\top (\nabla U(x, t) - \nabla V(x, t))\right|^2 \mu_t(\mathrm{d}x) \, \mathrm{d}t.$$

Proof. We note that v in this case can be written as

$$v = a^{-1}(c - b) = \nabla V + \gamma - \nabla U - \beta = \nabla(V - U) + \gamma - \beta.$$

Then $\nabla(U - V) \in L^2_{\text{div}}(a, \mu)^\perp$ since $\overline{\nabla} = \nabla_x$ for $C^{2,1}$ functions. Therefore $v_{\text{pot}} = \nabla(V - U)$ and the rest follows by Theorem 5.9 and Proposition 5.11. $\square$

It is remarkable that the coefficients of the SDE of the minimizer are independent of the initial distributions. One way to visualize the previous result is as follows: the gradients are responsible for the right marginals and the divergence free parts do not contribute to the marginals directly. However, the minimizer P has to be close to the reference measure so it inherits the divergence free part of the reference measure.

5.2 Consequences of the weak formulation

Assume everything about the reference measure R as in Assumption 5.1 and Assumption 5.2. We consider a general Schrödinger problem (MSP) where $\mathcal{T}$ is not necessarily $[0, 1]$. Assume that this (MSP) has a unique minimizer $\mathsf{P} \in \mathcal{C}^{\text{MSP}}_H$. We extend μ_t by $(X_t)_\#\mathsf{P}$ on $[0, 1] \setminus \mathcal{T}$ and consider the space $L^2(a, \mu)$.

We have the remarkable fact that the control u is a limit of "gradients", i.e.

Theorem 5.13. *Let u be the function obtained in Theorem 3.14 for the minimizer P to* (MSP)*. Then $u \in L^2_{\text{pot}}(a, \mu)$.*

Proof. We know that by Theorem 3.14 a function $\tilde{u} \in L^2(a, \mu)$ exists. We consider a new fully prescribed problem (FPSP) with the obtained $(\mu_t)_{t \in [0,1]}$ and $v = \tilde{u}$. Then P is a competitor for (FPSP) so it has a minimizer P^{FPSP} with $H(\mathsf{P}^{\text{FPSP}}|\mathsf{R}) \leq H(\mathsf{P}|\mathsf{R})$. But since the fully prescribed problem has more constraints we have $H(\mathsf{P}|\mathsf{R}) \leq H(\mathsf{P}^{\text{FPSP}}|\mathsf{R})$. Hence $\mathsf{P} = \mathsf{P}^{\text{FPSP}}$ by uniqueness of the minimizer to (FPSP) (see Theorem 3.2). Now the assertion that $u = \tilde{u}_{\text{pot}} \in L^2_{\text{pot}}(a, \mu)$ follows by Theorem 5.9 and Proposition 5.11. $\square$

The previous theorem tells us that there exists a sequence $(g_n)_{n\in\mathbb{N}} \subset \mathcal{D}_{e,\mu}$ such that $\overline{\nabla} g_n \to u$ in $L^2(a,\mu)$. We can actually choose this sequence $(g_n)_{n\in\mathbb{N}}$ to be a solution to a generalized version of Hamilton-Jacobi similar to (4.20). We have the following:

Conjecture. There exists $g_n \in \mathcal{D}_{e,\mu}$ such that $\overline{\nabla} g_n \to u$ in $L^2(a,\mu)$ and

$$Lg_n = -\frac{1}{2}|\sigma^\top \overline{\nabla} g_n|^2, \quad \mu_t \otimes \text{Leb-a.e.}$$

Sketch of proof. We use an approximate Schrödinger problems as in (3.17). Note that $\mathcal{C}^{(n)}$ is written in the form of the constraint set in Theorem 2.3 with finitely many function. The fact that the span of finitely many functions is closed we conclude that $\mathsf{P}_n \in \mathcal{C}_H^{(n)}$ the solution of the approximation satisfies

$$\frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\mathsf{R}} = \prod_{i=1}^{k(n)} f_{i,n}(X_{t_i}),$$

with $f_{i,n}$ all bounded measurable functions.

Hence now we can repeat the argument in Theorem 4.19 where the application of Itô's formula is understood in the sense of Proposition 5.5 with extended domain of generator and $\overline{\nabla}$. Every step works just as before and perhaps even easier since $\mathsf{P}_n \sim \mathsf{R}$ and the conclusion follows the same way as the calculations after Theorem 4.19. $\square$

Remark 5.14. If the coefficients for L allow for the application of Feyman-Kac's formula or Itô's formula, then we even have

$$Lg_n = -\frac{1}{2}|\sigma^\top \nabla g_n|^2, \quad \text{weakly,}$$

which is less vague than the generalized gradients $\overline{\nabla} g_n$. Therefore in such case $\nabla g_n \to u$ for some weak gradient ∇g_n.

5.3 Proof of the equivalence between the original and the weak formulation

The proof of Theorem 5.9 is technical and that is why it is put in one section on its own. The easy part is that the original formulation (FPSP) implies the weak formulation (WMSP) which we do first. After that we give the proof of the other direction.

Original implies weak formulation

Assume we have a minimizer $\mathsf{P} \in \mathcal{C}_H^{\text{FPSP}}$ of (FPSP). There exists a function $u \in L^2(a,\mu)$ by Theorem 3.6 and Theorem 3.14 such that $\mathsf{P} \in \text{MP}(L + u^\top a\nabla, \mu_0 \otimes \delta_0)$, and

$$H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)} = H(\mathsf{P}|\mathsf{R}) < \infty.$$

Now take any $f \in \mathcal{D}_{e,\mu}$. By Definition 5.4 we have that

$$N_t^{\mathsf{R}} = f(X_t, t) - f(X_0, 0) - \int_0^t Lf(X_s, s)\,\mathrm{d}s$$

is a R-local martingale. By change of measure we have that

$$N_t^{\mathsf{P}} = f(X_t, t) - f(X_0, 0) - \int_0^t (L + u^\top a\overline{\nabla})f(X_s, s)\,\mathrm{d}s,$$

is a P-local martingale. We can do the same for Q to get that

$$N_t^{\mathsf{Q}} = f(X_t, t) - f(X_0, 0) - \int_0^t (L + v^\top a\overline{\nabla})f(X_s, s)\,\mathrm{d}s,$$

is a Q-local martingale.

The way we defined $f \in \mathcal{D}_{e,\mu}$ makes sure that each term in N_t^{P} and N_t^{Q} has integrable supremum over $[0, 1]$. We check it for P, note that by triangle inequality, Definition 5.7 and Tonelli's theorem

$$\mathbb{E}_{\mathsf{P}}\left[\sup_{t\in[0,1]}\left|\int_0^t Lf(X_s, s)\,\mathrm{d}s\right|\right] \le \int_0^1 \int_{\mathbb{R}^d} |Lf(x, s)|\,\mu_s(\mathrm{d}x)\,\mathrm{d}s < \infty,$$

and by Cauchy-Schwarz

$$\mathbb{E}_{\mathsf{P}}\left[\sup_{t\in[0,1]}\left|\int_0^t u^\top a\overline{\nabla}f(X_s, s)\,\mathrm{d}s\right|\right] \le \|u\|_{L^2(a,\mu)}\|\overline{\nabla}f\|_{L^2(a,\mu)} < \infty.$$

Together with f being bounded by definition of $\mathcal{D}_e$ we get that

$$\mathbb{E}_{\mathsf{P}}\left[\sup_{t\in[0,1]}|N_t^{\mathsf{P}}|\right] < \infty,$$

which then implies that N^{P} is a martingale by Corollary 17.8 in [22]. A similar reasoning works for N^{Q}.

Now that we have proved the N^{P} and N^{Q} are martingales, we can take expectation of the martingale which is zero to get that

$$\mathbb{E}_{\mathsf{P}}\left[f(X_1, 1) - f(X_0, 0) - \int_0^1 (L + u^\top a\overline{\nabla})f(X_s, s)\,\mathrm{d}s\right] = 0,$$

and

$$\mathbb{E}_{\mathsf{Q}}\left[f(X_1, 1) - f(X_0, 0) - \int_0^1 (L + v^\top a\overline{\nabla})f(X_s, s)\,\mathrm{d}s\right] = 0.$$

Since P and Q have the same marginals we can equate both to get

$$\mathbb{E}_{\mathsf{P}}\left[\int_0^1 u^\top a\overline{\nabla} f(X_s,s)\,\mathrm{d}s\right] = \mathbb{E}_{\mathsf{P}}\left[\int_0^1 v^\top a\overline{\nabla} f(X_s,s)\,\mathrm{d}s\right].$$

That is exactly $u-v \in L^2_{\mathrm{div}}(a,\mu)$ so it is feasible to (WMSP). We conclude that the minimizer of P leads to a feasible solution to (WMSP).

In the second part below we see that the solution to the weak problem (WMSP) also gives a feasible solution to (FPSP).

Weak implies original formulation

We take any $u \in L^2(a,\mu)$ such that $u - v \in L^2_{\mathrm{div}}(a,\mu)$ and we want to show that we can construct a measure $\mathsf{P} \in \mathcal{C}_H^{\mathrm{FPSP}}$ with this u. We could be naive and just define

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}} \exp\left(\int_0^1 u(X_s,s)\,\mathrm{d}M_s - \frac{1}{2}\int_0^1 |\sigma(X_s,s)^\top u(X_s,s)|^2\,\mathrm{d}s\right).$$

However, we cannot just do this. First of all we do not know whether we can define the stochastic integral $\int_0^1 u(X_s,s)\,\mathrm{d}M_s$ neither P or R-a.s. One could think that the condition $u \in L^2(a,\mu)$ helps us to make sense out of it. That is a priori not true, because we do not know whether P has marginal laws $(\mu_t)_{t\in[0,1]}$.

Fortunately, it turns out that we can define P like that, but it is not straightforward. We proceed by using the stopping times

$$T_n := \inf\left\{t \geq 0 : \int_0^t |\sigma(X_s,s)^\top u(X_s,s)|^2\,\mathrm{d}s \geq n\right\} \wedge 1.$$

We consider the measures $\tilde{\mathsf{P}}_n$ defined as P above until time T_n. This measure $\tilde{\mathsf{P}}_n$, unlike P, is a well-defined probability measure. We work through a long procedure to show that $\tilde{\mathsf{P}}_n$ tends to P in some sense and conclude that P is well-defined as well. The inequality that will help us is the following

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}[f(X_t,t)\mathbb{1}_{\{t\leq T_n\}}] \leq \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x), \quad \text{for a large class of functions } f. \tag{5.10}$$

We call this the key inequality, because in some sense if we are allowed to take $f(x,t) = \mathbb{1}_A(x)$, then it tells us that $\tilde{\mathsf{P}}_n(A) \lessapprox \mu_t(A)$ so that taking A^c yields $\tilde{\mathsf{P}}_n(A) \approx \mu_t(A)$. The reason why we have $\approx$ is that we have the indicator $\mathbb{1}_{\{t\leq T_n\}}$ which we want to get rid of. As one gets the feeling already, it is going to be a technical proof. We follow the paper by Cattiaux and Leonard [7] which is mostly devoted to proof the claim that we are after. We split the proof in the following parts:

(i) Approximation of the function u by bounded and compactly supported functions.

(ii) Localization argument.

(iii) Derivation of the key inequality (5.10).

(iv) Construction of the probability measure P.

(i) Approximation of the function u by bounded and compactly supported functions.

We once again work with the Markov process $(X_t, r_t)_{t\geq 0}$ instead of $(X_t)_{t\geq 0}$. Let us define a bounded compactly supported approximation of u, namely

$$u_k(x,t) := u(x,t)\mathbb{1}_{\{|u|\leq k\}}(x,t)\theta_k(x),$$

with $\theta_k \in C_c^\infty(\mathbb{R}^d)$ such that $\theta(x) = 1$ on $\{|x| \leq k\}$ and $\theta_k(x) = 0$ on $\{|x| \geq k+1\}$ with smooth transition on $\{k \leq |x| \leq k+1\}$.

Define a probability measure $\mathsf{P}_k^{x,r}$ via

$$\frac{\mathrm{d}\mathsf{P}_k^{x,r}}{\mathrm{d}\mathsf{R}^{x,r}} := \exp\left(\int_0^1 u_k(X_s, r_s)\,\mathrm{d}M_s - \frac{1}{2}\int_0^1 |\sigma(X_s,r_s)^\top u_k(X_s,r_s)|^2\,\mathrm{d}s\right).$$

Using Novikov's criterion we get

Proposition 5.15. *The measure $\mathsf{P}_k^{x,r}$ with the Radon-Nikodym derivative $\frac{\mathrm{d}\mathsf{P}_k^{x,r}}{\mathrm{d}\mathsf{R}^{x,r}}$ written above is a well-defined probability measure such that $\mathsf{P}_k^{x,r} \sim \mathsf{R}^{x,r}$.*

Proof. Note that u_k has compact support, is bounded and a is locally bounded, which makes $|\sigma^\top u_k|^2 = (u_k, au_k)$ bounded. Therefore

$$\int_0^1 |\sigma(X_s,r_s)^\top u_k(X_s,r_s)|^2\,\mathrm{d}s \leq \|\sigma^\top u_k\|_{\sup}^2 < \infty.$$

We can define the local martingale $N_t = \int_0^t u_k(X_s,r_s)\,\mathrm{d}M_s$ and since $[M]_1 \leq \|\sigma^\top u_k\|_{\sup}^2$ we get by Novikov's criterion that $\mathsf{P}_k^{x,r}$ is a probability measure (see Theorem 18.23 in [22]). Since the local martingale M is defined $\mathsf{R}^{x,r}$ almost surely, we get that $\frac{\mathrm{d}\mathsf{P}_k^{x,r}}{\mathrm{d}\mathsf{R}^{x,r}} > 0$ $\mathsf{R}^{x,r}$-a.s. leading to $\mathsf{P}_k^{x,r} \sim \mathsf{R}^{x,r}$. □

We set $\mathsf{P}_k := \mathsf{P}_k^{\mu_0\otimes\delta_0}$, more precisely for any $A \in \mathcal{F}$

$$\mathsf{P}_k(A) = \int_{\mathbb{R}^d} \mathsf{P}^{x,0}(A)\,\mu_0(\mathrm{d}x).$$

We have that $\mathsf{P}_k^{x,r}$ is Markovian as follows from

Proposition 5.16. *The probability measure $\mathsf{P}_k^{x,r}$ is Markovian.*

Proof. This result is well-known, we define the $\mathsf{R}^{x,r}$-martingale

$$Z_t^{(k,x,r)} := \mathbb{E}_{\mathsf{R}}^{x,r}\left[\frac{\mathrm{d}\mathsf{P}_k^{x,r}}{\mathrm{d}\mathsf{R}^{x,r}} \mid \mathcal{F}_t\right].$$

The main reason is that for $t \geq s$

$$Z_t^{(k,x,r)} = Z_s^{(k,x,r)}\exp\left(\int_s^t u_k(X_s,r_s)\,\mathrm{d}M_s - \frac{1}{2}\int_s^t |\sigma(X_s,r_s)^\top u_k(X_s,r_s)|^2\,\mathrm{d}s\right).$$

The second term on the right-hand side is $\sigma(X_\tau\ :\ \ \tau \in [s,t])$-measurable. To be fully convinced one must recall that the stochastic integral inside the exponential can be written as a limit of Riemann sums with respect to M which is defined in terms of X (see Theorem 2.13 from Chapter 4 in [35]). Then we have for any $t \geq s$ and $A \in \mathcal{F}_s$ and bounded Borel measurable function $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}_{\mathsf{P}_k^{x,r}}[\mathbb{1}_A f(X_t, r_t)] = \mathbb{E}_{\mathsf{R}^{x,r}}[Z_t^{(k,x,r)} \mathbb{1}_A f(X_t, r_t)] = \mathbb{E}_{\mathsf{R}^{x,r}}[Z_s^{(k,x,r)} \mathbb{1}_A g(X_s, r_s)],$$

for some function g by the Markov property of $\mathsf{R}^{x,r}$. This gives for $t \geq s$

$$\mathbb{E}_{\mathsf{P}_k^{x,r}}[f(X_t, r_t) \mid \mathcal{F}_s] = \mathbb{E}_{\mathsf{P}^{x,r}}[\mathbb{1}_A g(X_s, r_s)] = g(X_s, r_s),$$

which is the asserted Markov property of $\mathsf{P}_k^{x,r}$. $\square$

Knowing that $\mathsf{P}_k^{x,r}$ is Markovian we can define the associated semigroup $P_t^{(k)}$ on $\mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ by

$$P_t^{(k)} f(x,r) := \mathbb{E}_{\mathsf{P}_k^{x,r}}[f(X_t, r_t)].$$

Define the space

$$E_k := \left\{ f \in \mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R}) :\ \|P_t^{(k)} f - f\|_{\sup} \to 0 \ \text{ as } \ t \to 0^+ \right\}.$$

The semigroup $P_t^{(k)}$ is strongly continuous Markov semigroup on the space $(E_k, \|\cdot\|_{\sup})$ (see Section 1.2.1 in [2] and see Definition 2.1 in [32]). Moreover, a simple calculation shows that E_k is a closed subspace of $\mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ with respect to the sup-norm which makes it a Banach space since $(\mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R}), \|\cdot\|_{\sup})$ is. We denote the generator of $P_t^{(k)}$ by $L^{(k)}$ with domain (see page 1 in [32])

$$\mathcal{D}(L^{(k)}) := \left\{ f \in E_k :\ \lim_{t \to 0^+} \frac{P_t^{(k)} f - f}{t} \ \text{exists w.r.t. } \|\cdot\|_{\sup}\text{-norm} \right\},$$

and for any $f \in \mathcal{D}(L^{(k)})$

$$L^{(k)} f := \lim_{t \to 0+} \frac{1}{t} \left(P_t^{(k)} f - f \right).$$

By Itô's formula we get $C_c^\infty(\mathbb{R}^d \times \mathbb{R}) \subset E_k$. Indeed, for any $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$

$$N_t^{\mathsf{R}} = f(X_t, r_t) - f(X_0, r) - \int_0^t Lf(X_s, r_s)\,\mathrm{d}s$$

is a bounded $\mathsf{R}^{x,r}$-martingale. Therefore by Girsanov's theorem

$$N_t^{\mathsf{P}_k} = f(X_t, r_t) - f(X_0, r) - \int_0^t [L + u_k^\top a \nabla] f(X_s, r_s)\,\mathrm{d}s$$

is a $\mathsf{P}_k^{x,r}$-martingale since the extra term $u_k^\top a \nabla f$ is bounded for u_k and a being locally bounded and ∇f having compact support. This leads to $\mathbb{E}_{\mathsf{P}_k^{x,r}}[N_t^{\mathsf{P}_k}] = \mathbb{E}_{\mathsf{P}_k^{x,r}}[N_0^{\mathsf{P}_k}] = 0$ so that

$$P_t^{(k)} f(x,r) - f(x,r) = \mathbb{E}_{\mathsf{P}_k^{x,r}}[f(X_t,r_t) - f(X_0,r_0)] = \mathbb{E}_{\mathsf{P}_k^{x,r}}\left[\int_0^t [L + u_k^\top a \nabla] f(X_s,r_s)\,\mathrm{d}s\right].$$

Hence taking the supremum with the local boundedness of a and b implies that $\|(L + u_k^\top a\nabla) f\|_{\sup} < \infty$ which then readily gives

$$\|P_t^{(k)} f - f\|_{\sup} \leq \int_0^t \|(L + u_k^\top a \nabla_x) f\|_{\sup}\,\mathrm{d}s = t\|(L + u_k^\top a\nabla_x)f\|_{\sup} \to 0,$$

as $t \to 0^+$, yielding $f \in E_k$.

Now take any non-negative $f \in \mathcal{D}(L^{(k)})$ and $t \in [0,1]$ and consider the function

$$F_s(x,r) := P_{t-s}^{(k)} f(x,r).$$

Well-known results from semigroup theory (see Theorem 2.4 in [32]) tells us that $F_s \in C^1([0,t],\mathcal{D}(L^{(k)}))$ and satisfies the backward equation

$$\left(\partial_s + L^{(k)}\right) F_s(x,r) = 0, \quad (s,x,r) \in [0,t] \times \mathbb{R}^d \times \mathbb{R}, \tag{5.11}$$

with the values at $s = t$ and $s = 0$ being

$$F_t(x,r) = f(x,r), \quad F_0(x,r) = \mathbb{E}_{\mathsf{P}_k^{x,r}}[f(X_t,r_t)]. \tag{5.12}$$

This function F is difficult to work with, because we do not know if it maps to $\mathcal{D}_{e,\mu}$. So, we must do a localization procedure which is multiplying F with a smooth compactly supported function to get some grip on F. That leads to the next step.

(ii) Localization argument

We need a sequence of cut-off functions with a specific purpose. The existence of such functions is ensured by the following lemma:

Lemma 5.17. *There exists a family of functions $(\zeta_\lambda)_{\lambda>0} \subset C_c^\infty(\mathbb{R}^d)$ such that the following hold:*

(1) $\zeta_\lambda \geq 0$, and $\zeta_\lambda \to 1$ everywhere as $\lambda \to \infty$;

(2) $\int_0^1 \int_{\mathbb{R}^d} |L\zeta_\lambda(x,t)|\,\mu_t(\mathrm{d}x)\,\mathrm{d}t \to 0$ as $\lambda \to \infty$;

(3) $\|\nabla\zeta_\lambda\|_{L^2(a,\mu)} \to 0$ as $\lambda \to \infty$.

All the above hold for ζ_λ^2 as well.

Proof. See Appendix A.2. □

As we have noted, we want to consider $\zeta_\lambda F$, it turns out that $[s \mapsto \zeta_\lambda(x) F_s(x,s)] \in \mathcal{D}_{e,\mu}$. The step towards showing that is the following lemma:

Lemma 5.18. *Let $\varphi \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ and $g \in \mathcal{D}(L^{(k)})$, then $\varphi g \in \mathcal{D}_{e,\mu}$.*

Proof. Roughly speaking, it is all based on integration by parts for stochastic integrals and Itô's formula. First note that Itô's formula applied to φ yields

$$\varphi(X_t, r_t) - \varphi(X_0, r) = \int_0^t \nabla_x \varphi(X_s, r_s) \, \mathrm{d}M_s + \int_0^t L\varphi(X_s, r_s) \, \mathrm{d}s.$$

By a similar argument as in Proposition 1.6 from Chapter 7 in [35] (where it is remarked to hold in more generality, see also Corollary 4.8 from [12]) we get that

$$N_t^{\mathsf{P}_k} = g(X_t, r_t) - g(X_0, r) - \int_0^t L^{(k)} g(X_s, r_s) \, \mathrm{d}s \tag{5.13}$$

is a $\mathsf{P}_k^{x,r}$-martingale. By Proposition 5.15 we know that $\mathsf{P}_k^{x,r} \sim \mathsf{R}^{x,r}$ which together with the assumption that $\mathsf{R}^{x,r}$ is an extremal solution to $\mathrm{MP}(L, \delta_x \otimes \delta_r)$ and Girsanov's theorem yields that $\mathsf{P}^{x,r}$ is an extremal solution to $\mathrm{MP}(L - u_k^\top a \nabla, \delta_x \otimes \delta_r)$. So the arguments in Proposition 5.5 imply that there exists a Borel measurable function $\overline{\nabla}^{\mathsf{P}_k} g : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ such that

$$N_t^{\mathsf{P}_k} = \int_0^t \overline{\nabla}^{\mathsf{P}_k} g(X_s, r_s) \, \mathrm{d}M_s^{\mathsf{P}_k}, \tag{5.14}$$

where M^{P_k} is local martingale part of X under $\mathsf{P}_k^{x,r}$. By Girsanov's theorem we get that

$$N_t^{\mathsf{R}} = \int_0^t \overline{\nabla} g(X_s, r_s) \, \mathrm{d}M_s,$$

is a $\mathsf{R}^{x,r}$-local martingale and it satisfies

$$N_t^{\mathsf{R}} = g(X_t, r_t) - g(X_0, r) - \int_0^t (L^{(k)} - u^\top a \overline{\nabla}^{\mathsf{P}_k}) g(X_s, r_s) \, \mathrm{d}s.$$

This implies by definition of $\mathcal{D}_e$ that $g \in \mathcal{D}_e$ and $Lg = (L^{(k)} - u_k^\top a \overline{\nabla}^{\mathsf{P}_k})g$. Now we get by Proposition 5.5 the existence of $\overline{\nabla} g$ which can be chosen to be equal to $\overline{\nabla}^{\mathsf{P}_k} g$ by the uniqueness mentioned in Remark 5.6 and the fact that $\mathsf{P}_k^{x,r} \sim \mathsf{R}^{x,r}$. Now integration by parts for semimartingale gives

$$\begin{aligned}(\varphi g)(X_t, r_t) - (\varphi g)(X_0, r_0) &= \int_0^t \varphi(X_s, r_s) \, \mathrm{d}g(X_s, r_s) + \int_0^t g(X_s, r_s) \, \mathrm{d}\varphi(X_s, r_s) \\ &\quad + \int_0^t \nabla \varphi^\top a \overline{\nabla} g(X_s, r_s) \, \mathrm{d}s.\end{aligned}$$

Note that by the expression for g we get

$$\int_0^t \varphi(X_s, r_s)\, \mathrm{d}g(X_s, r_s) = \int_0^t \varphi\overline{\nabla} g(X_s, s)\, \mathrm{d}M_s + \int_0^t \varphi L g(X_s, r_s)\, \mathrm{d}s,$$

and by the expression for φ we get

$$\int_0^t g(X_s, r_s)\, \mathrm{d}\varphi(X_s, r_s) = \int_0^t g\nabla\varphi(X_s, r_s)\, \mathrm{d}M_s + \int_0^t gL\varphi(X_s, r_s)\, \mathrm{d}s.$$

This means that

$$\begin{aligned}(\varphi g)(X_r, r_t) - (\varphi g)(X_0, r_0) &= \int_0^t \left[\varphi\overline{\nabla} g + g\nabla\varphi\right](X_s, r_s)\, \mathrm{d}M_s \\ &\quad + \int_0^t \left[\varphi L g + gL\varphi + \nabla\varphi^\top a\overline{\nabla} g\right](X_s, r_s)\, \mathrm{d}s.\end{aligned}$$

This implies that $\varphi g \in \mathcal{D}_e$ and by uniqueness of $\overline{\nabla}(\varphi g)$ we have

$$\overline{\nabla}(\varphi g) = \varphi\overline{\nabla} g + g\nabla\varphi,$$

and

$$L(\varphi g) = \varphi L g + gL\varphi + \nabla\varphi^\top a\overline{\nabla} g.$$

Now we need to show that $(\varphi g) \in \mathcal{D}_{e,\mu}$. Looking at the expressions for $\overline{\nabla}(\varphi g)$ and $L(\varphi g)$ we see that we are dealing with bounded functions except $\overline{\nabla} g$. Once we show that $\overline{\nabla} g \in L^2(a, \mu)$, then it is immediate that $\overline{\nabla}(\varphi g) \in L^2(a, \mu)$, but we also get through Cauchy-Schwarz that

$$\int_0^1 \int_{\mathbb{R}^d} |L(\varphi g)(x, s)|\, \mu_s(\mathrm{d}x)\, \mathrm{d}s < \infty.$$

Let us show $\overline{\nabla} g \in L^2(a, \mu)$. By the fact that the marginals $(\mu_t)_{t\in[0,1]}$ are coming from the measure Q, we have that

$$\|\overline{\nabla} g\|^2_{L^2(a,\mu)} = \mathbb{E}_{\mathsf{Q}}\left[\int_0^1 \overline{\nabla} g^\top a\overline{\nabla} g(X_s, s)\, \mathrm{d}s\right].$$

This leaves us wondering how we can relate this to our setting, since until now we were working with the measure P_k. The answer to such questions can usually be obtained by grabbing a cup of coffee and defining the stopping time

$$\tau_n := \inf\left\{t \geq 0 : \int_0^t |\sigma^\top v(X_s, s)|^2 \geq n\right\} \wedge 1.$$

Stopping local martingales gives one time to think before the local martingale misbehaves. We must note that although we were working with $\mathsf{R}^{x,r}$- and $\mathsf{P}_k^{x,r}$-local martingales above, but these remain R- and P_k-local martingales by integrating over the initial values.

Let us also define the stopped version of Q, namely

$$\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0)\exp\left(\int_0^{\tau_n} v(X_s,s)\,\mathrm{d}M_s - \frac{1}{2}\int_0^{\tau_n} |\sigma^\top v(X_s,s)|^2\,\mathrm{d}s\right).$$

Moreover, since $\tau_n \to 1$ Q-a.s., we see that due to continuity of the semi-martingale inside the exponential that

$$\lim_{n\to\infty}\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}}, \quad \mathsf{Q}\text{-a.s. and } \mathsf{R}\text{-a.s. on } \left\{\frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\mathsf{R}} > 0\right\}.$$

This convergence of the Radon-Nikodym derivative allows us to get via Fatou's lemma that

$$\|\overline{\nabla} g\|^2_{L^2(a,\mu)} = \mathbb{E}_{\mathsf{Q}}\left[\int_0^1 \overline{\nabla} g^\top a \overline{\nabla} g(X_s,s)\,\mathrm{d}s\right] \le \liminf_{n\to\infty}\mathbb{E}_{\mathsf{Q}_n}\left[\int_0^1 \overline{\nabla} g^\top a \overline{\nabla} g(X_s,s)\,\mathrm{d}s\right]. \quad (5.15)$$

It is still not clear where this leads to, until we figure out that $\mathsf{Q}_n \sim \mathsf{P}_k$ and one can easily check that

$$\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k} = \exp\left(-\int_0^1 (u_k - v\mathbb{1}_{\{s\le\tau_n\}})(X_s,s)\,\mathrm{d}M_s^{\mathsf{P}_k} - \frac{1}{2}\int_0^1 |\sigma^\top(u_k - v\mathbb{1}_{\{s\le\tau_n\}})(X_s,s)|^2\,\mathrm{d}s\right).$$

This brings us one step closer. We can also write by the integral representation of N^{P_k} in (5.14)

$$\mathbb{E}_{\mathsf{Q}_n}\left[\int_0^1 \overline{\nabla} g^\top a \overline{\nabla} g(X_s,s)\,\mathrm{d}s\right] = \mathbb{E}_{\mathsf{P}_k}\left[\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k}[N^{\mathsf{P}_k}]_1\right]. \quad (5.16)$$

Recall Fenchel's inequality stating that for any $w \ge 0$ and $z \in \mathbb{R}$

$$wz \le w\log w - w + e^z.$$

Now we introduce $\alpha > 0$ and apply Fenchel's inequality on the right-hand side of (5.16)

$$\begin{aligned}
\mathbb{E}_{\mathsf{P}_k}\left[\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k}[N^{\mathsf{P}_k}]_1\right] &= \alpha\mathbb{E}_{\mathsf{P}_k}\left[\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k}\frac{[N^{\mathsf{P}_k}]_1}{\alpha}\right] \\
&\le \alpha\mathbb{E}_{\mathsf{P}_k}\left[\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k}\log\left(\frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k}\right) - \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\mathsf{P}_k} + 1 + \exp\left([N^{\mathsf{P}_k}]_1\right) - 1\right] \quad (5.17) \\
&\le \alpha H(\mathsf{Q}_n|\mathsf{P}_k) + \alpha\mathbb{E}_{\mathsf{P}_k}\left[\exp\left(\frac{[N^{\mathsf{P}_k}]_1}{\alpha}\right)\right].
\end{aligned}$$

Therefore by (5.15) and (5.16) it is enough to show that the right-hand side of (5.17) is uniformly bounded in n. Looking at (5.17) we conclude that we must show that the relative entropy is bounded in n and that the expectation of the exponential of the quadratic variation is finite.

Let us show that the relative entropy is uniformly bounded in n. By Girsanov's theorem and since $(u_k - v\mathbb{1}_{s\leq\tau_n})(X_s, s)$ is uniformly bounded, we observe that

$$H(\mathsf{Q}_n|\mathsf{P}_k) = \frac{1}{2}\mathbb{E}_{\mathsf{Q}_n}\left[\int_0^1 |\sigma^\top(u_k - v\mathbb{1}_{\{s\leq\tau_n\}})(X_s, s)|^2\,\mathrm{d}s\right].$$

Using

$$|x-y|^2 \leq 2(|x|^2 + |y|^2)$$

yields

$$H(\mathsf{Q}_n|\mathsf{P}_k) \leq \mathbb{E}_{\mathsf{Q}_n}\left[\int_0^1 |\sigma^\top u_k(X_s, s)|^2\,\mathrm{d}s\right] + \mathbb{E}_{\mathsf{Q}_n}\left[\int_0^{\tau_n} |\sigma^\top v(X_s, s)|^2\,\mathrm{d}s\right].$$

Notice that u_k has compact support so that $|\sigma^\top u|^2 = (u, au)_{\mathbb{R}^d}$ is bounded for a being locally bounded. We get

$$H(\mathsf{Q}_n|\mathsf{P}_k) \leq \|\sigma^\top u_k\|_{\sup}^2 + \mathbb{E}_{\mathsf{Q}}\left[\int_0^{\tau_n} |\sigma^\top v(X_s, s)|^2\,\mathrm{d}s\right],$$

where we have used that $\mathsf{Q} = \mathsf{Q}_n$ on $\mathcal{F}_{\tau_n}$. We also know $\tau_n \leq 1$ yielding

$$H(\mathsf{Q}_n|\mathsf{P}_k) \leq \|\sigma^\top u_k\|_{\sup}^2 + \|v\|_{L^2(a,\mu)}^2 < \infty.$$

We have obtained a bound independent of n for the relative entropy.

Let us do the other term in (5.17) which by the definition of the exponential function and Tonelli can be written as

$$\mathbb{E}_{\mathsf{P}_k}\left[\exp\left(\frac{[N^{\mathsf{P}_k}]_1}{\alpha}\right)\right] = \mathbb{E}_{\mathsf{P}_k}\left[\sum_{p=0}^\infty \frac{[N^{\mathsf{P}_k}]_1^p}{\alpha^p p!}\right] = \sum_{p=0}^\infty \frac{1}{\alpha^p p!}\mathbb{E}_{\mathsf{P}_k}\left[[N^{\mathsf{P}_k}]_1^p\right].$$

We have a p-th moment of a quadratic variation so BDG-inequality comes in handy, especially since N^{P_k} is bounded, say by $C > 0$. Indeed, the boundedness of N^{P_k} follows since in its definition (5.13), we have $g \in \mathcal{D}(L^{(k)})$ and integral of $L^{(k)}g$ which both are uniformly bounded. Therefore BDG-inequality (Theorem 18.17 in [40]) implies that all those moments of $[N]_1$ are finite and in particular for any $p \geq 2$ we have

$$\mathbb{E}_{\mathsf{P}_k}\left[[N^{\mathsf{P}_k}]_1^p\right] \leq (4p)^p \mathbb{E}_{\mathsf{P}_k}\left[\sup_{s\in[0,1]} \left|N_s^{\mathsf{P}_k}\right|^{2p}\right] \leq (4p)^p C^{2p}.$$

We end up with

$$\mathbb{E}_{\mathsf{P}_k}\left[\exp\left(\frac{[N^{\mathsf{P}_k}]_1}{\alpha}\right)\right] \leq \sum_{p=0}^\infty \frac{(4p)^p C^{2p}}{\alpha^p p!}.$$

Ratio test for series shows that for $C^2/\alpha < (4/e)$ the series converge. Since $\alpha > 0$ was not specified, we will specify it to become large enough such that the convergence of the series is satisfied.

Wrapping everything up, we conclude that we can bound the right-hand side of (5.15) uniformly in n so that

$$\|\overline{\nabla} g\|_{L^2(a,\mu)} < \infty,$$

which was the remaining condition to ensure $\varphi g \in \mathcal{D}_{e,\mu}$. $\square$

(iii) Derivation of the key inequality

Using the previous two lemmas we can show that for $\varphi \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$, the function φF satisfies the weak forward equation corresponding to the Markov process $(X_t, r_t)_{t\geq 0}$ under Q as defined in Definition 4.45 in [7]. The assumption $u - v \in L^2_{\text{div}}(a, \mu)$ allows us to translate everything to u as follows from:

Lemma 5.19. *For any function $f \in \mathcal{D}_{e,\mu}$ we have*

$$\int_0^t \int_{\mathbb{R}^d} (u-v)^\top a \overline{\nabla} f(x,s)\, \mu_s(\mathrm{d}x)\, \mathrm{d}s = 0.$$

Proof. Let $\xi_\varepsilon \in C_b^\infty(\mathbb{R}^d \times \mathbb{R})$ be a function $\xi_\varepsilon(x,s) = \xi_\varepsilon(s)$ uniformly bounded and converging to $\mathbb{1}_{[0,t]}(s)$. Then one can easily see that $\xi_\varepsilon f \in \mathcal{D}_{e,\mu}$ and $\overline{\nabla}(\xi_\varepsilon f) = \xi_\varepsilon \overline{\nabla} f$. By Cauchy-Schwarz and dominated convergence one gets

$$\int_0^t \int_{\mathbb{R}^d} (u-v)^\top a \overline{\nabla} f(x,s)\, \mu_s(\mathrm{d}x)\, \mathrm{d}s = \lim_{\varepsilon \to 0^+} \int_0^1 \int_{\mathbb{R}^d} (u-v)^\top a \overline{\nabla}(\xi_\varepsilon f)(x,s)\, \mu_s(\mathrm{d}x)\, \mathrm{d}s = 0.$$

□

Lemma 5.20 (Weak forward equation)**.** *For any $\varphi \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ and $t \in [0,1]$ we have*

(1)

$$\int_{\mathbb{R}^d \times \mathbb{R}} (\varphi F_t)(x,t)\, \mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d \times \mathbb{R}} (\varphi F_0)(x,0)\, \mu_0(\mathrm{d}x) =$$
$$\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}} \left[\partial_r + L_{x,s} + u^\top a \overline{\nabla}\right] (\varphi F_r)(x,s)\, \mu_r(\mathrm{d}x)\, \delta_r(\mathrm{d}s)\, \mathrm{d}r.$$

(2)

$$\int_{\mathbb{R}^d \times \mathbb{R}} (\varphi F_t^2)(x,t)\, \mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d \times \mathbb{R}} (\varphi F_0^2)(x,0)\, \mu_0(\mathrm{d}x) =$$
$$\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}} \left[\partial_r + L_{x,s} + u^\top a \overline{\nabla}\right] (\varphi F_r^2)(x,s)\, \mu_r(\mathrm{d}x)\, \delta_r(\mathrm{d}s)\, \mathrm{d}r.$$

Proof. We use the starting point that by a similar argument used in the proof of Lemma 5.18 we have that

$$F_t(X_t,t) - F_0(X_0,0) - \int_0^t \int_{\mathbb{R}} \left[\partial_r + L_{x,s}^{(k)}\right] F_r(X_s,s)\, \delta_r(\mathrm{d}s)\, \mathrm{d}r$$

is a P_k-martingale, and by Girsanov's theorem

$$F_t(X_t,t) - F_0(X_0,0) - \int_0^t \int_{\mathbb{R}} \left[\partial_r + L_{x,s}\right] F_r(X_s,s)\, \delta_r(\mathrm{d}s)\, \mathrm{d}r$$

is a R-local martingale. This shows $[(x,s)\mapsto F_s(x,s)]\in\mathcal{D}(L^{(k)})$ so that by Lemma 5.18 we get $\varphi F\in\mathcal{D}_{e,\mu}$ and

$$N_t^{\mathsf{R}}=(\varphi F_t)(X_t,t)-(\varphi F_0)(X_0,0)-\int_0^t\int_{\mathbb{R}}\left[\partial_r+L_{x,s}\right](\varphi F_r)(X_s,s)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r$$

is a R-local martingale. Now we change the measure to Q to get that

$$N_t^{\mathsf{Q}}=(\varphi F_t)(X_t,t)-(\varphi F_0)(X_0,0)-\int_0^t\int_{\mathbb{R}}\left[\partial_r+L_{x,s}+v^\top a\overline{\nabla}\right](\varphi F_r)(X_s,s)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r$$

is a Q-local martingale.

Note that by definition F is bounded so φF is bounded. Also $\partial_r(\varphi F_r)$ is bounded by the fact that φ is bounded and independent of r and the derivative $\partial_r F_r=-P_{t-r}Lf$ is again bounded (see Theorem 2.4 in [32]). We also know that

$$\begin{aligned}\mathbb{E}_{\mathsf{Q}}\left[\int_0^1\left|\left[L_{x,s}+v^\top a\overline{\nabla}\right](\varphi F_s)(X_s,s)\right|\,\mathrm{d}s\right]\le&\int_0^1\int_{\mathbb{R}^d}|L(\varphi F_s)(x,s)|\,\mu_s(\mathrm{d}x)\,\mathrm{d}s\\&+\int_0^1\int_{\mathbb{R}^d}|v^\top a\overline{\nabla}(\varphi F_s)(x,s)|\,\mu_t(\mathrm{d}x)\,\mathrm{d}t.\end{aligned}$$

By Cauchy on the latter integral we see that everything is fine due to $\varphi F\in\mathcal{D}_{e,\mu}$ and $v\in L^2(a,\mu)$. We know that the supremum of the Q-local martingale N_t^{Q} can be treated as follows:

$$\begin{aligned}\mathbb{E}_{\mathsf{Q}}\left[\sup_{s\in[0,1]}|N_s^{\mathsf{Q}}|\right]\le&\,2\|\varphi F\|_{\sup}+\|\varphi\partial_r F\|_{\sup}+\int_0^1\int_{\mathbb{R}^d}|L(\varphi F_s)(x,s)|\,\mu_s(\mathrm{d}x)\,\mathrm{d}s\\&+\|v\|_{L^2(a,\mu)}\|\overline{\nabla}(\varphi F)\|_{L^2(a,\mu)}<\infty,\end{aligned}$$

where we see F as the mapping $(x,s)\mapsto F_s(x,s)$ when we take the $L^2(a,\mu)$-norm. Then by Corollary 17.8 in [22] we get that N^{Q} is a true Q-martingale.

So N^{Q} is a martingale which implies it has constant expectation $\mathbb{E}_{\mathsf{Q}}[N_t^{\mathsf{Q}}]=\mathbb{E}_{\mathsf{Q}}[N_0^{\mathsf{Q}}]=0$. This gives

$$\mathbb{E}_{\mathsf{Q}}\left[(\varphi F_t)(X_t,t)-(\varphi F_0)(X_0,0)-\int_0^t\int_{\mathbb{R}}\left[\partial_r+L_{x,s}+v^\top a\overline{\nabla}\right](\varphi F_r)(X_s,s)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r\right]=0.$$

Since everything inside is separately integrable we get

$$\int_{\mathbb{R}^d\times\mathbb{R}}(\varphi F_t)(x,t)\,\mu_t(\mathrm{d}x)-\int_{\mathbb{R}^d\times\mathbb{R}}(\varphi F_0)(x,0)\,\mu_0(\mathrm{d}x)=$$

$$\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}}\left[\partial_r+L_{x,s}+v^\top a\overline{\nabla}\right](\varphi F_r)(x,s)\,\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r.$$

But the fact that $u-v \in L^2_{\mathrm{div}}(a,\mu)$ allows us via $v=(u-v)+u$ and Lemma 5.19 to write

$$\int_{\mathbb{R}^d\times\mathbb{R}} (\varphi F_t)(x,t)\,\mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d\times\mathbb{R}} (\varphi F_0)(x,0)\,\mu_0(\mathrm{d}x) =$$

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} \left[\partial_r + L_{x,s} + u^\top a\overline{\nabla}\right](\varphi F_r)(x,s)\,\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r.$$

That is the first assertion.

To get the second assertion we need one extra step of integration by parts to get rid of F^2. But everything goes fine since F itself is bounded. □

Remark 5.21. The use of the measure δ_r in the previous lemma is to not confuse the reader that ∂_r is being taken with respect to the variable r in $F_r(x,s)$.

Recall the function $(\zeta_\lambda)_{\lambda\geq 0}$ form Lemma 5.17. From now we set $\zeta=\zeta_\lambda$ and apply the theorems above with $\varphi=\zeta$. We know that $\zeta(x,t)$ is not compactly support in t, but we can modify ζ such that it is 1 for $t\in[-2,2]$ and vanishing smoothly outside. Since we are interested in $t\in[0,1]$, there won't be any issue.

Using Lemma 5.20 on ζF and (5.11) we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \zeta^2 F_t(x,t)\,\mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d} \zeta^2 F_0(x,0)\,\mu_0(\mathrm{d}x) = \\
&\int_0^t \int_{\mathbb{R}^d} (u-u_k)^\top a\overline{\nabla}(\zeta^2 F_s)(x,s)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s \\
&+\int_0^t \int_{\mathbb{R}^d} F_s\left[L\zeta^2 + u_k^\top a\nabla\zeta^2 - 2\nabla\zeta\cdot a\nabla\zeta\right](x,s)\,\mu_r(\mathrm{d}x)\,\mathrm{d}s \\
&+2\int_0^t \int_{\mathbb{R}^d} \nabla\zeta\cdot a\overline{\nabla}(\zeta F_s)(x,s)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s.
\end{aligned} \tag{5.18}$$

We have by Cauchy-Schwarz

$$\left|\int_0^t \int_{\mathbb{R}^d} (u-u_k)^\top a\overline{\nabla}(\zeta^2 F_s)(x,s)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s\right| \leq \|u-u_k\|_{L^2(a,\mu)}\|\overline{\nabla}(\zeta^2 F)\|_{L^2(a,\mu)}, \tag{5.19}$$

where we consider $(s,x)\mapsto\overline{\nabla}(\zeta^2 F_s)(x,s)$ when we take the $L^2(a,\mu)$-norm. We also have

$$\begin{aligned}
&\left|\int_0^t \int_{\mathbb{R}^d} F_s\left[L\zeta^2 + u_k^\top a\nabla\zeta^2 - 2\nabla\zeta\cdot a\nabla\zeta\right](x,s)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s\right| \\
&\leq \|f\|_{\sup}\left(\int_0^1 \int_{\mathbb{R}^d} |L\zeta^2(x,s)|\,\mu_s(\mathrm{d}x)\,\mathrm{d}s + \|u_k\|_{L^2(a,\mu)}\|\nabla\zeta^2\|_{L^2(a,\mu)} + 2\|\nabla\zeta\|^2_{L^2(a,\mu)}\right).
\end{aligned} \tag{5.20}$$

Finally

$$\left|\int_0^t \int_{\mathbb{R}^d} \nabla\zeta\cdot a\overline{\nabla}(\zeta F_s)(x,s)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s\right| \leq \|\nabla\zeta\|_{L^2(a,\mu)}\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)}. \tag{5.21}$$

The term in (5.20) is the easiest to manage. Note that $\|u_k\|_{L^2(a,\mu)} \leq 2\|u\|_{L^2(a,\mu)}$ for k large enough. The rest of the terms tend to 0 as $\lambda \to \infty$ by Lemma 5.17. The rest, namely (5.19) and (5.21), also tend to zero if $k \to \infty$ and $\lambda \to \infty$ if we can show that $\|\overline{\nabla}(\zeta^2 F)\|_{L^2(a,\mu)}$ and $\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)}$ are uniformly bounded in $\lambda > 0$. That is what we show in

Lemma 5.22.

$$\liminf_{\lambda\to\infty} \|\overline{\nabla}(\zeta_\lambda F)\|_{L^2(a,\mu)} \leq 2\sqrt{2}\|f\|_{\sup}, \quad \textit{and} \quad \liminf_{\lambda\to\infty} \|\overline{\nabla}(\zeta_\lambda^2 F)\|_{L^2(a,\mu)} \leq 4\|f\|_{\sup}.$$

Proof. We start with applying the weak forward equation Lemma 5.20 on $(\zeta F)^2$

$$\int_{\mathbb{R}^d} (\zeta F_t)^2(x,t)\,\mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d} (\zeta F_0)^2(x,0)\,\mu_0(\mathrm{d}x) = \\ \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} \left[\partial_r + L_{x,s} + u^\top a\overline{\nabla}\right] (\zeta F_r)^2(x,s)\,\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r.$$

We will make a couple of calculations to write the latter integral in a favourable way. We will apply the product rule for L and $\overline{\nabla}$ a couple of times. First we see that

$$\left[\partial_r + L_{x,s} + u^\top a\overline{\nabla}\right] (\zeta F_r)^2 = 2\zeta F_r \left[\partial_r + L_{x,s} + u^\top a\overline{\nabla}\right] (\zeta F_r) + \overline{\nabla}(\zeta F_r)^\top a\overline{\nabla}(\zeta F_r).$$

Integrating the right-hand side gives the $L^2(a,\mu)$ norm of $\overline{\nabla}(\zeta F)$. With little manipulation using the product rules and integration by parts we end up with (we take the norm $\|\dot{\|}_{L^2(a,\mu)}$ but it is actually until time t)

$$\begin{aligned}
\|\overline{\nabla}(\zeta F)\|^2_{L^2(a,\mu)} &= \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} \overline{\nabla}(\zeta F_r)^\top a\overline{\nabla}(\zeta F_r)(x,s)\,\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r \\
&= \int_{\mathbb{R}^d} (\zeta F_t)^2(x,t)\,\mu_t(\mathrm{d}x) - \int_{\mathbb{R}^d} (\zeta F_0)^2(x,0)\,\mu_0(\mathrm{d}x) \\
&\quad - \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} 2\left[\zeta F_r(u-u_k)^\top a\overline{\nabla}(\zeta F_r)\right](x,s)\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r \\
&\quad - \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} 2\left[\zeta F_r^2\left[L_{x,s} + u_k^\top a\nabla\right]\zeta\right](x,s)\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r \\
&\quad - \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}} \left[2F_r\nabla\zeta^\top a\overline{\nabla}(\zeta F_r) - 2F_r^2\nabla\zeta^\top a\nabla\zeta\right](x,s)\mu_r(\mathrm{d}x)\,\delta_r(\mathrm{d}s)\,\mathrm{d}r.
\end{aligned} \tag{5.22}$$

Bounding each term with a mixture of supremum-norm and Cauchy-Schwarz gives

$$\begin{aligned}
\|\overline{\nabla}(\zeta F)\|^2_{L^2(a,\mu)} \leq \|f\|^2_{\sup} &+ 2\|f\|_{\sup}\|u-u_k\|_{L^2(a,\mu)}\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)} \\
&+ 2\|f\|^2_{\sup}\left(\|L\zeta\|_{\sup} + \|u_k\|_{L^2(a,\mu)}\|\nabla\zeta\|_{L^2(a,\mu)}\right) \\
&+ 2\|f\|_{\sup}\|\nabla\zeta\|_{L^2(a,\mu)}\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)} + 2\|f\|^2_{\sup}\|\nabla\zeta\|^2_{L^2(a,\mu)}.
\end{aligned}$$

Note that by taking k large enough we can get $\|u - u_k\|_{L^2(a,\mu)} \leq 1$ and $\|u_k\|_{L^2(a,\mu)} \leq 2\|u\|_{L^2(a,\mu)}$. So this gives with a little bit of rearranging

$$\begin{aligned}\|\overline{\nabla}(\zeta F)\|^2_{L^2(a,\mu)} \leq \|f\|^2_{\sup}\left(1 + 2\left(\|L\zeta\|_{\sup} + 2\|u\|_{L^2(a,\mu)}\|\nabla\zeta\|_{L^2(a,\mu)} + \|\nabla\zeta\|^2_{L^2(a,\mu)}\right)\right)\\ + 2\|f\|_{\sup}\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)}\left(1 + \|\nabla\zeta\|_{L^2(a,\mu)}\right).\end{aligned}$$

We use $z^2 \leq \alpha^2(1+\gamma) + 2\alpha\beta z$ with $\gamma \geq 0$ implies $|z| \leq 2|\alpha|(1+\gamma)^{1/2}(1+\beta^2)^{1/2}$ so that

$$\begin{aligned}\|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)} \leq 2\|f\|_{\sup}\left(1 + 2\left(\|L\zeta\|_{\sup} + 2\|u\|_{L^2(a,\mu)}\|\nabla\zeta\|_{L^2(a,\mu)} + \|\nabla\zeta\|^2_{L^2(a,\mu)}\right)\right)^{1/2}\\ \times\left(1 + \left(1 + \|\nabla\zeta\|_{L^2(a,\mu)}\right)^2\right)^{1/2}.\end{aligned}$$

By taking $\lambda \to \infty$ we get that

$$\liminf_{\lambda\to\infty} \|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)} \leq 2\sqrt{2}\|f\|_{\sup}.$$

The fact that

$$\overline{\nabla}(\zeta^2 F) = \zeta\overline{\nabla}(\zeta F) + \zeta F\nabla\zeta,$$

with the inequality $|x+y|^2 \leq 2|x|^2 + 2|y|^2$ allows us to bound

$$\|\overline{\nabla}(\zeta^2 F)\|^2_{L^2(a,\mu)} \leq 2\|\overline{\nabla}(\zeta F)\|^2_{L^2(a,\mu)} + 2\|f\|^2_{\sup}\|\nabla\zeta\|^2_{L^2(a,\mu)}.$$

Hence

$$\liminf_{\lambda\to\infty} \|\overline{\nabla}(\zeta^2 F)\|_{L^2(a,\mu)} \leq \sqrt{2}\liminf_{\lambda\to\infty} \|\overline{\nabla}(\zeta F)\|_{L^2(a,\mu)} \leq 4\|f\|_{\sup},$$

which is exactly what we were after. □

A straightforward corollary is

Corollary 5.23. *For any non-negative $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ we have*

$$\mathbb{E}_{\mathsf{P}_k}[f(X_t,t)] \leq \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u - u_k\|_{L^2(a,\mu)}.$$

Proof. Let $f \in \mathcal{D}(L^{(k)})$. Recall Equation (5.18) and how we bounded everything in (5.19), (5.20) and (5.21). What we then find is

$$\begin{aligned}\int_{\mathbb{R}^d} \zeta^2 F_0(x,0)\,\mu_0(\mathrm{d}x) &\leq \int_{\mathbb{R}^d} \zeta^2 F_t(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u - u_k\|_{L^2(a,\mu)} + o_\lambda(1)\\ &\leq \int_{\mathbb{R}^d} F_t(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u - u_k\|_{L^2(a,\mu)} + o_\lambda(1).\end{aligned}$$

Hence taking $\lambda \to \infty$ and applying Fatou's lemma yields, since $\zeta_\lambda \to 1$ as $\lambda \to \infty$,

$$\int_{\mathbb{R}^d} F_0(x,0)\mu_0(\mathrm{d}x) \leq \int_{\mathbb{R}^d} F_t(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u - u_k\|_{L^2(a,\mu)}.$$

Recall by the definition of F_0 and F_t in Equation (5.12), the inequality becomes

$$\mathbb{E}_{\mathsf{P}_k}[f(X_t,t)] = \int_{\mathbb{R}^d} \mathbb{E}_{\mathsf{P}_k^{x,0}}[f(X_t,t)]\,\mu_0(\mathrm{d}x) \leq \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u-u_k\|_{L^2(a,\mu)}.$$

This is the desired inequality except that we have $f \in \mathcal{D}(L^{(k)})$. We want to have it for $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$. So take any $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ such that $f \geq 0$. We have $f \in E_k$ and the function f_ε defined by

$$f_\varepsilon = \frac{1}{\varepsilon}\int_0^\varepsilon P_t^{(k)} f\,\mathrm{d}t,$$

satisfies

$$\|f_\varepsilon\|_{\sup} \leq \frac{1}{\varepsilon}\int_0^\varepsilon \|P_t^{(k)} f\|_{\sup}\,\mathrm{d}t \leq \frac{1}{\varepsilon}\int_0^\varepsilon \|f\|_{\sup}\,\mathrm{d}t = \|f\|_{\sup},$$

where we have used that $P_t^{(k)}$ is a contraction which follows from it being a Markovian semigroup (see Section 1.2.1 in [2]). It is not difficult to see that $f_\varepsilon \in \mathcal{D}(L^{(k)})$ and $f_\varepsilon \to f$ uniformly (see for instance Theorem 2.4 in [32]). This allows us to extend the inequality to all non-negative $f \in C_c^\infty(\mathbb{R}^d)$. $\square$

(iv) Construction of the probability measure P.

We can almost define our probability measure P. To that end, first define the stopping time

$$T_n := \inf\left\{t \geq 0 : \int_0^t |\sigma^\top u(X_s,s)|^2\,\mathrm{d}s \geq n\right\} \wedge 1.$$

Now we can define a new measure as follows:

$$\frac{\mathrm{d}\tilde{\mathsf{P}}_n}{\mathrm{d}\mathsf{R}} = \frac{\mathrm{d}\mu_0}{\mathrm{d}\nu_0}(X_0)\exp\left(\int_0^{T_n} u(X_s,s)\,\mathrm{d}M_s - \frac{1}{2}\int_0^{T_n} |\sigma^\top u(X_s,s)|^2\,\mathrm{d}s\right).$$

It is immediate that this is a well-defined probability measure. Finally we define

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} := \liminf_{n\to\infty}\frac{\mathrm{d}\tilde{\mathsf{P}}_n}{\mathrm{d}\mathsf{R}}.$$

We are about to celebrate soon, but it is still unclear whether that $\mathrm{d}\mathsf{P}/\mathrm{d}\mathsf{R}$ defines a probability measure.

Let us introduce the martingales

$$Z_{t\wedge T_n}^{(k)} := \mathbb{E}_{\mathsf{R}}\left[\frac{\mathrm{d}\mathsf{P}_k}{\mathrm{d}\mathsf{R}} \mid \mathcal{F}_{t\wedge T_n}\right],$$

and

$$Z_{t\wedge T_n} := \mathbb{E}_{\mathsf{R}}\left[\frac{\mathrm{d}\tilde{\mathsf{P}}_n}{\mathrm{d}\mathsf{R}} \mid \mathcal{F}_{t\wedge T_n}\right].$$

We have that both $(Z_{T_n})_{n\in\mathbb{N}}$ and $(Z_{T_n}^{(k)})_{n\in\mathbb{N}}$ are $(\mathcal{F}_{T_n},\mathsf{R})$-martingales by the optional stopping theorem (see Theorem 7.29 in [22]).

Proposition 5.24. *For any bounded non-negative measurable function $f \in \mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ we have*

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}[f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] \le \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x).$$

Proof. We have

$$\lim_{k\to\infty} \mathbb{E}_{\mathsf{R}}\left[\int_0^{T_n} |\sigma^\top (u-u_k)(X_s,s)|^2\,\mathrm{d}s\right] = 0,$$

since $u_k \to u$ and on $s \le T_n$ we have uniform boundedness of u and $u_k \le u$ so that the claim holds by dominated convergence theorem. By the use of BDG-inequality we can conclude that the local martingales in the expression of $Z^{(k)}_{t\wedge T_n}$ converges to $Z_{t\wedge T_n}$ R-a.s. as $k\to\infty$ along a subsequence (but we do not relabel it).

Now take any non-negative function $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ and note that

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}[f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] = \mathbb{E}_{\mathsf{R}}[Z_{t\wedge T_n} f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] = \mathbb{E}_{\mathsf{R}}\left[\lim_{k\to\infty} Z^{(k)}_{t\wedge T_n} f(X_t,t)\mathbb{1}_{\{t\le T_n\}}\right].$$

Applying Fatou's lemma, changing the measure to P_k and using the fact that $\mathbb{1}_{\{t\le T_n\}}$ gives

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}[f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] \le \liminf_{k\to\infty} \mathbb{E}_{\mathsf{R}}[Z^{(k)}_{t\wedge T_n} f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] \le \liminf_{k\to\infty} \mathbb{E}_{\mathsf{P}_k}[f(X_t,t)].$$

We know by Corollary 5.23 that

$$\begin{aligned}\liminf_{k\to\infty} \mathbb{E}_{\mathsf{P}_k}[f(X_t,t)] &\le \liminf_{k\to\infty}\left(\int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x) + 4\|f\|_{\sup}\|u-u_k\|_{L^2(a,\mu)}\right)\\ &= \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x).\end{aligned}$$

We conclude

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}[f(X_t,t)\mathbb{1}_{\{t\le T_n\}}] \le \int_{\mathbb{R}^d} f(x,t)\,\mu_t(\mathrm{d}x).$$

The fact that we can extend this inequality for any non-negative $f \in \mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ follows by Lusin's theorem and mollification (Lemma 1.37 in [22]). □

Note that there exists simple functions $f_m \in \mathfrak{B}_b(\mathbb{R}^d \times \mathbb{R})$ increasing to $|\sigma^\top u|^2$ as $m\to\infty$ so that

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}\left[\int_0^{T_n} f_m(X_t,t)\,\mathrm{d}t\right] = \int_0^1 \mathbb{E}_{\tilde{\mathsf{P}}_n}\left[f_m(X_t,t)\mathbb{1}_{\{t\le T_n\}}\right]\mathrm{d}t \le \int_0^1\int_{\mathbb{R}^d} f_m(x,t)\,\mu_t(\mathrm{d}x)\,\mathrm{d}t.$$

Monotone convergence yields

$$\mathbb{E}_{\tilde{\mathsf{P}}_n}\left[\int_0^{T_n} |\sigma^\top u(X_s,s)|^2\,\mathrm{d}s\right] \le \int_0^1\int_{\mathbb{R}^d} |\sigma^\top u(x,s)|^2\,\mu_s(\mathrm{d}x)\,\mathrm{d}s = \|u\|^2_{L^2(a,\mu)}.$$

This leads to the satisfying result:

Theorem 5.25. *The measure* P *is a probability measure satisfying* $H(\mathsf{P}|\mathsf{R}) < \infty$.

Proof. It is easy to see that

$$H(\tilde{\mathsf{P}}_n|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\mathbb{E}_{\tilde{\mathsf{P}}_n}\left[\int_0^{T_n} |\sigma^\top u(X_s,s)|^2 \,\mathrm{d}s\right].$$

Then

$$H(\tilde{\mathsf{P}}_n|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\mathbb{E}_{\tilde{\mathsf{P}}_n}\left[\int_0^{T_n} |\sigma^\top u(X_s,s)|^2 \,\mathrm{d}s\right] \leq H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)}.$$

This gives that

$$H(\tilde{\mathsf{P}}_n|\mathsf{R}) = \mathbb{E}_{\mathsf{R}}\left[Z_{T_n}\log(Z_{T_n})\right].$$

This implies uniform integrability of Z_{T_n}, because for any $K \geq 0$

$$\begin{aligned}\mathbb{E}_{\mathsf{R}}\left[Z_{T_n}\mathbb{1}_{\{|Z_{T_n}|\geq K\}}\right] &\leq \frac{1}{\log(K)}\mathbb{E}_{\mathsf{R}}\left[Z_{T_n}\log(Z_{T_n})\mathbb{1}_{\{|Z_{T_n}|\geq K\}}\right]\\ &\leq \frac{1}{\log(K)}\left(H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)}\right),\end{aligned}$$

and that tends to zero as $K \to \infty$ uniformly in $n \in \mathbb{N}$. We also know Z_{T_n} is a $(\mathsf{R}, \mathcal{F}_{T_n})$-martingale so that by martingale convergence theorem we get

$$\mathsf{P}(\Omega) = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_\Omega \lim_{n\to\infty} Z_{T_n}\right] = \mathbb{E}_{\mathsf{R}}\left[\lim_{n\to\infty} Z_{T_n}\right] = \lim_{n\to\infty}\mathbb{E}_{\mathsf{R}}[Z_{T_n}] = 1.$$

Finally L^1-convergence of Z_{T_n} gives that for any bounded continuous function $U : \Omega \to \mathbb{R}$ that

$$\lim_{n\to\infty}\mathbb{E}_{\tilde{\mathsf{P}}_n}[U] = \lim_{n\to\infty}\mathbb{E}_{\mathsf{R}}[Z_{T_n}U] = \mathbb{E}_{\mathsf{R}}\left[\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}U\right] = \mathbb{E}_{\mathsf{P}}[U],$$

which is saying that $\tilde{\mathsf{P}}_n \to \mathsf{P}$ weakly. Finally, the lower semi-continuity of the relative entropy yields

$$H(\mathsf{P}|\mathsf{R}) \leq \liminf_{n\to\infty} H(\tilde{\mathsf{P}}_n|\mathsf{R}) \leq H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)} < \infty.$$

$\square$

This is a very promising result. It only remains to show that P satisfies the marginal constraints and the relative entropy is given in terms of u.

Theorem 5.26. *The probability measure* P *is feasible, and*

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}} = \mathbb{1}_{\{\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}>0\}}\exp\left(\int_0^1 u(X_s,s)\,\mathrm{d}M_s - \frac{1}{2}\int_0^1 |\sigma(X_s,s)^\top u(X_s,s)|^2\,\mathrm{d}s\right).$$

In particular under P *the canonical process* X *has a finite variation part given by*

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s, s) + a(X_s, s)u(X_s, s))\,\mathrm{d}s,$$

and

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)}.$$

Proof. Note we have for any $A \in \mathcal{F}_{T_n}$

$$\mathsf{P}(A) = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}\right] = \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \liminf_{n\to\infty} \frac{\mathrm{d}\tilde{\mathsf{P}}_m}{\mathrm{d}\mathsf{R}}\right] \leq \liminf_{m\to\infty} \mathbb{E}_{\mathsf{R}}\left[\mathbb{1}_A \frac{\mathrm{d}\tilde{\mathsf{P}}_m}{\mathrm{d}\mathsf{R}}\right] = \tilde{\mathsf{P}}_n(A).$$

Hence $\mathsf{P}(A) \leq \tilde{\mathsf{P}}_n(A)$ and by taking A^c yields $\mathsf{P}(A) = \tilde{\mathsf{P}}_n(A)$. Therefore

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{R}}|_{\mathcal{F}_{T_n}} = Z_{T_n}$$

so that by Girsanov's theorem we get that

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s, s) + a(X_s, s)u(X_s, s))\,\mathrm{d}s$$

for any $t \leq T_\infty := \sup_n T_n$, since that is true for $\tilde{\mathsf{P}}_n$ and $\mathsf{P} = \tilde{\mathsf{P}}_n$ on $\mathcal{F}_{T_n}$. We also have $H(\mathsf{P}|\mathsf{R}) < \infty$ so that by Girsanov's theorem and finite entropy Theorem 3.6 we get the existence of a process $\beta = (\beta_s)_{s\in[0,1]}$ such that for all $t \in [0,1]$

$$A_t^{\mathsf{P}} = \int_0^t (b(X_s, s) + a(X_s, s)\beta_s)\,\mathrm{d}s.$$

Therefore for any $t \leq T_\infty$

$$\int_0^t a(X_s, s)u(X_s, s)\,\mathrm{d}s = \int_0^t a(X_s, s)\beta_s\,\mathrm{d}s.$$

Lebesgue differentiation gives P-a.s. $\omega \in \Omega$ for a.e. $t \leq T_\infty(\omega)$

$$a(X_t(\omega), t)u(X_t(\omega), t) = a(X_t(\omega), t)\beta_t(\omega).$$

Therefore

$$\int_0^{t\wedge T_\infty} |\sigma(X_s, s)^\top u(X_s, s)|^2\,\mathrm{d}s = \int_0^{t\wedge T_\infty} |\sigma(X_s, s)^\top \beta_s|^2\,\mathrm{d}s \leq \int_0^1 |\sigma(X_s, s)^\top \beta_s|^2\,\mathrm{d}s < \infty,$$

which implies that the left-hand side cannot explode in finite time. Therefore $T_n \to T_\infty = 1$ P-a.s. In particular one can define the P-semimartingale $(\int_0^t u(X_s, s)\,\mathrm{d}M_s)_{t\in[0,1]}$. In particular, Proposition 5.24 after taking $n \to \infty$; $\tilde{\mathsf{P}}_n = \mathsf{P}$ on $\mathcal{F}_{T_n}$, and monotone convergence gives

$$\mathbb{E}_{\mathsf{P}}[f(X_t, t)] \leq \int_{\mathbb{R}^d} f(x, t)\,\mu_t(\mathrm{d}x).$$

This holds for any bounded measurable function f. In particular for $f = \mathbb{1}_A$ and $f = \mathbb{1}_{A^c}$ where $A \in \mathcal{B}(\mathbb{R}^d)$ is arbitrary. That gives

$$\mathsf{P}(X_t \in A) = \mu_t(A),$$

which makes P feasible. □

Let us wrap everything up. In the part of the proof of the original implying the weak formulation, we have showed that the minimizer $\mathsf{P}^{\mathrm{FPSP}} \in \mathcal{C}_H^{\mathrm{FPSP}}$ to (FPSP) gives a function u^{FPSP} that is feasible to (WMSP) and

$$H(\mathsf{P}^{\mathrm{FPSP}}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\|u^{\mathrm{FPSP}}\|^2_{L^2(a,\mu)}.$$

Therefore, for the minimizer u of (WMSP) we have

$$H(\mathsf{P}^{\mathrm{FPSP}}|\mathsf{R}) \geq H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)}.$$

But since the solution u to (WMSP) yields a feasible $\mathsf{P} \in \mathcal{C}_H^{\mathrm{FPSP}}$ with

$$H(\mathsf{P}|\mathsf{R}) = H(\mu_0|\nu_0) + \frac{1}{2}\|u\|^2_{L^2(a,\mu)} \leq H(\mathsf{P}^{\mathrm{FPSP}}|\mathsf{R}),$$

we get that $\mathsf{P} = \mathsf{P}^{\mathrm{FPSP}}$ by uniqueness of $\mathsf{P}^{\mathrm{FPSP}}$ by Theorem 3.2. This concludes the equivalence of (FPSP) and (WMSP). The fact that the semimartingale decomposition of X can be written as given in the statement of Theorem 5.9 follows from Theorem 5.26.

Appendix A

Postponed proofs

A.1 Proof of Theorem 4.10

We first need a Lemma:

Lemma A.1. *Let $I \in \mathbb{R}$ be an interval and $S \subset \mathbb{R}^d$ be a bounded open set. For any function $v \in C^\infty(I, L^p(S, \rho))$ we have that any $\varphi \in C_c^\infty(S \times I)$ we have*

$$\int_I \partial_t^{(n)} v(t)\varphi(\cdot, t)\,\mathrm{d}t = (-1)^n \int_I v(t)\partial_t^{(n)}\varphi(\cdot, t)\,\mathrm{d}t.$$

Proof. A simple application of dominated convergence theorem shows that $t \mapsto v(t)\varphi(\cdot, t)$ is differentiable in $L^p(S, \rho)$ and the usual product rule holds

$$(v(t)\varphi(\cdot, t))' = v'(t)\varphi(\cdot, t) + v(t)\partial_t\varphi(\cdot, t).$$

By Theorem 2 from Section 5.9 in [13] we get that the Fundamental Theorem of Calculus holds for such Banach-valued functions and the claim follows for $n = 1$. We can do this for all $n \in \mathbb{N}$ through an induction argument. □

Now we can continue with the proof of Theorem 4.10. The semigroup is analytic for all $t \geq 0$ by Assumption 4.8 which was in force for Theorem 4.10 (see [27]). Fix $t \in (0, 1]$ and $m, n \in \mathbb{N}$. Note that for any $s \in [t, 2]$ we have that the function $s \mapsto u(\cdot, s)$ is infinitely many times differentiable as seen as a map from $[t, 2]$ to $L^p(\mathbb{R}^d, \eta)$. In particular since $[t, 2]$ is a compact set we have that $[s \mapsto u(\cdot, s)]$ and all its (time-)derivatives are bounded in $L^p(\mathbb{R}^d, \eta)$. In particular there exists a constant $C_n < \infty$ such that

$$\sup_{s\in[t,2]} \left\|\partial_s^{(n)} u(\cdot, s)\right\|_{L^p(\mathbb{R}^d,\eta)}^p \leq C_n.$$

Obviously

$$\mathcal{L}^m \partial_s^{(n)} u(\cdot, s) = \mathcal{L}^m \mathcal{L}^n u(\cdot, s) = \mathcal{L}^{n+m} u(\cdot, s) = \partial_s^{(n+m)} u(\cdot, s).$$

Therefore

$$\sup_{s\in[t,2]} \left\|\mathcal{L}^m \partial_s^{(n)} u(\cdot,s)\right\|^p_{L^p(\mathbb{R}^d,\eta)} \le C_{n+m}.$$

Consider a ball $B_r = B_r(0)$ for some $r > 0$, then since

$$-\infty < \inf_{B_r} e^{-U} \le \sup_{B_r} e^{-U} < \infty$$

we get that

$$\inf_{B_r} e^{-U} \|\cdot\|_{L^p(B_r,\mathrm{d}x)} \le \|\cdot\|_{L^p(B_r,\eta)} \le \sup_{B_r} e^{-U} \|\cdot\|_{L^p(B_r,\mathrm{d}x)}.$$

So the norms $\|\cdot\|_{L^p(B_r,\eta)}$ and $\|\cdot\|_{L^p(B_r,\mathrm{d}x)}$ are equivalent on $L^p(B_r,\eta)$. In particular, we get that

$$\sup_{s\in[t,2]} \left\|\mathcal{L}^m \partial_s^{(n)} u(\cdot,s)\right\|^p_{L^p(B_r,\mathrm{d}x)} \le C'_{n+m}$$

Elliptic regularity (Corollary 10.3.10 in [26]) tells us that for any $0 < r' < r$ (to ensure $B_{r'} \subset B_r$)

$$\left\|\partial_s^{(n)} u(\cdot,s)\right\|^p_{W^{m+2,p}(B_{r'},\mathrm{d}x)} \le \tilde{C}\left(\left\|\mathcal{L}^m \partial_s^{(n)} u(\cdot,s)\right\|^p_{L^p(B_r,\mathrm{d}x)} + \left\|\partial_s^{(n)} u(\cdot,s)\right\|^p_{L^p(B_r,\mathrm{d}x)}\right). \tag{A.1}$$

The same argument can be used by considering $\partial_s^{(n)} u(\cdot,s) - \partial_{s'}^{(n)} u(\cdot,s')$ to show that $s \mapsto \partial_s^{(n)} u(\cdot,s)$ is actually continuous in $W^{m+2,p}(B_{r'},\mathrm{d}x)$. Due to the continuity we can get a product measurable function $\partial_x^\alpha \partial_s^{(n)} u(x,s)$ for any multi-index α such that $|\alpha| \le m+2$. The product measurability can be proved with Pettis' theorem and the continuity and the fact that $W^{m+2,p}(B_{r'},\mathrm{d}x)$ is separable (See Theorem 7 from Appendix E in [13]). We can actually get rid of r' by varying r so we can keep working on B_r. Moreover, we have for any $\varphi \in C_c^\infty(B_r \times (t,2))$

$$\int_t^2 \int_{B_r} \partial_x^\alpha \partial_s^{(n)} u(x,s)\varphi(x,s)\,\mathrm{d}x\,\mathrm{d}s = (-1)^{|\alpha|} \int_t^2 \int_{B_r} \partial_s^{(n)} u(x,s)\partial_x^\alpha \varphi(x,s)\,\mathrm{d}x\,\mathrm{d}s.$$

We also know that $s \mapsto u(\cdot,s)$ being differentiable on $L^p(B_r,\mathrm{d}x)$ implies the product rule on $L^p(B_r,\mathrm{d}x)$ by Lemma A.1 above. We have to exchange the order of integration to apply Lemma A.1 which is allowed since we have $u \in L^\infty([s,t];L^p(B_r,\mathrm{d}x))$ and φ is uniformly bounded. So we get

$$\int_t^2 \int_{B_r} \partial_x^\alpha \partial_s^{(n)} u(x,s)\varphi(x,s)\,\mathrm{d}x\,\mathrm{d}s = (-1)^{|\alpha|+n} \int_t^2 \int_{B_r} u(x,s)\partial_s^{(n)}\partial_x^\alpha \varphi(x,s)\,\mathrm{d}x\,\mathrm{d}s.$$

This means that $\partial_x^\alpha \partial_s^{(n)} u(x,s)$ is the weak derivative for the function $u(x,s)$ (which is also product measurable due to strong continuity of the semigroup by the argument used for $\partial_x^\alpha \partial_s^{(n)} u$).

The integrability condition (A.1) and the fact that n and m are arbitrary we conclude that $(x,s) \mapsto u(x,s)$ is in $W^{m,p}(B_r \times (t,2), \mathrm{d}x\,\mathrm{d}s)$ for all $m \in \mathbb{N}$. Then by Morrey's inequality (Theorem 6, Section 5.6 in [13]) and that $t > 0$ was arbitrary, we get that $u \in C^\infty(\mathbb{R}^d \times (0,1])$.

The claim that for fixed $t \in (0,1]$ $u(\cdot,t) \in W^{2,p}(\mathbb{R}^d,\eta)$ follows by $u(\cdot,t) \in \mathcal{D}(\mathcal{L}) = W^{2,p}(\mathbb{R}^d,\eta)$ which holds for analytic semigroups (see Lemma 4.2 in [32]).

A.2 Proof of Lemma 5.17

Proof. The idea is to construct functions that do the purpose and then we mollify it to get smooth functions. Consider a family of functions $(\xi_\lambda)_{\lambda\geq 0}$ defined as follows:

$$\xi_\lambda(x) := \begin{cases} 1 & \text{if } |x| \leq \lambda \\ 1 - \frac{1}{\lambda}\left(\log(2+|x|) - \log(2+\lambda)\right) & \text{if } \lambda < |x| < (2+\lambda)e^\lambda - 2 \\ 0 & \text{otherwise.} \end{cases}$$

Note that this function is piecewise C^∞. It is also continuous which makes it weakly differentiable for we can integrate by parts and the boundary terms vanish due to continuity. Now mollify this function ξ_λ with the standard mollifier $\rho \in C_c^\infty(\mathbb{R}^d)$ which is

$$\rho(x) := \left(\int_{\{|y|\leq 1\}} \exp\left(-\frac{1}{1-|y|^2}\right) \mathrm{d}y\right) \exp\left(-\frac{1}{1-|x|^2}\right) \mathbb{1}_{\{|x|\leq 1\}},$$

and define

$$\zeta_\lambda := \xi_\lambda * \rho.$$

Of course we have $\zeta_\lambda \in C_c^\infty(\mathbb{R}^d)$. It only remains to be shown that the conditions are satisfied. Let us check them one by one:

(1) It is immediate that $\zeta_\lambda \geq 0$ and $\zeta_\lambda \to 1$ everywhere since for any $x \in \mathbb{R}^d$ we have that for λ large enough $\xi_\lambda(y-x) = 1$ for all $|y| \leq 1$ so that the claim follows due to $\|\rho\|_{L^1(\mathbb{R}^d)} = 1$.

(2) By the triangle inequality we are good to go if we can prove that

$$\lim_{\lambda\to\infty} \int_0^1 \int_{\mathbb{R}^d} |b_i(x,t)\partial_i\zeta_\lambda(x)|\,\mu_t(\mathrm{d}x)\,\mathrm{d}t = 0, \qquad \text{for all } i \in \{1,...,d\},$$

and

$$\lim_{\lambda\to\infty} \int_0^1 \int_{\mathbb{R}^d} |a_{ij}(x,t)\partial_{ij}\zeta_\lambda(x)|\,\mu_t(\mathrm{d}x)\,\mathrm{d}t = 0, \qquad \text{for all } i,j \in \{1,...,d\}.$$

Note that $\xi_\lambda = 0$ except on $\lambda < |x| < (2+\lambda)e^\lambda - 2$. For $\lambda < |x| < (2+\lambda)e^\lambda - 2$ we have

$$\partial_i\xi_\lambda(x) = -\frac{x_i}{\lambda|x|(2+|x|)}.$$

This gives for a.e. $x \in \mathbb{R}^d$

$$|\partial_i\xi_\lambda(x)| \leq \frac{1}{\lambda(2+|x|)}.$$

We know that ζ_λ is differentiable and $\partial_i\zeta_\lambda = \xi_\lambda * (\partial_i\rho)$, but since ξ_λ is weakly differentiable with bounded weak derivative, integration by parts gives

$$\partial_i\zeta_\lambda = (\partial_i\xi_\lambda) * \rho.$$

Using this fact and the definition of convolution we get

$$|\partial_i \zeta_\lambda(x)| \leq \operatorname{esssup}_{|y|\leq 1} |\partial_i \xi_\lambda(x+y)| \underbrace{\|\rho\|_{L^1(\mathbb{R}^d)}}_{=1} \leq \operatorname{esssup}_{|y|\leq 1} \frac{1}{\lambda(2+|x+y|)} \leq \frac{1}{\lambda(1+|x|)},$$

where we have used that $|x+y| \geq |x| - |y| \geq |x| - 1$ in the last inequality. Multiplying this by $b_i(x)$ gives for all $x \in \mathbb{R}^d$

$$|b_i(x,t)\partial_i \zeta_\lambda(x)| \leq \frac{1}{\lambda} \frac{C(1+|x|)}{(1+|x|)} \leq \frac{C}{\lambda}.$$

This leads to uniform convergence to zero. In particular, since the measure $\mu_t \otimes \mathrm{Leb}\,|_{[0,1]}$ that we are integrating against is a probability measure

$$\int_0^1 \int_{\mathbb{R}^d} |b_i(x,t)\partial_i \zeta_\lambda(x)|\, \mu_t(\mathrm{d}x)\, \mathrm{d}t \leq \frac{C}{\lambda},$$

which tends to zero as $\lambda \to 0$.

For the other one concerning a_{ij} and the second derivatives $\partial_{ij}\zeta_\lambda$, we must do a little more. Once again as before we note for $\lambda < |x| < (2+\lambda)e^\lambda - 2$ we have ξ_λ twice differentiable and for $j \neq i$

$$\partial_{ij}\xi_\lambda(x) = -\frac{x_i}{\lambda|x|^2(2+|x|)^2}\left(-2\frac{x_j}{|x|} - 2x_j\right) = \frac{2x_i x_j(1+|x|)}{\lambda|x|^3(2+|x|)^2},$$

and for $j \neq i$

$$\partial_{ii}\xi_\lambda(x) = -\frac{1}{\lambda|x|^2(2+|x|)^2}\left(|x|(2+|x|) - 2\frac{x_i^2}{|x|} - 2x_i^2\right).$$

It is not so difficult to see that we can find a constant $C_1 > 0$ such that for $\lambda < |x| < (2+\lambda)e^\lambda - 2$

$$|\partial_{ij}\xi_\lambda(x)| \leq \frac{C_1}{\lambda(1+|x|^2)}$$

for all $\lambda \geq \lambda_0$ for some λ_0.

Recall that $\operatorname{supp}\rho \subset \{|x| \leq 1\}$ which makes $\partial_{ij}\xi_\lambda$ zero outside $\{\lambda - 1 < |x| < (2+\lambda)e^\lambda - 1\}$. Moreover for $\lambda + 1 < |x| < (2+\lambda)e^\lambda - 3$ we have that $\partial_i \xi_\lambda(x-y)$ differentiable for all $|y| \leq 1$ which allows integration by parts as follows:

$$\partial_{ij}\zeta_\lambda(x) = \partial_{ij}\xi_\lambda * \rho,$$

which together with the bound obtained above gives

$$|\partial_{ij}\zeta_\lambda(x)| \leq \sup_{|y|\leq 1} \frac{C_1}{\lambda(1+|x+y|^2)} \leq \frac{C_1}{\lambda(1+(|x|-1)^2)}.$$

Multiplying this with a_{ij} keeps it bounded by the quadratic growth condition. The only tricky part is $\lambda - 1 \le |x| \le \lambda + 1$ and $(2+\lambda)e^{\lambda} - 3 \le |x| \le (2+\lambda)e^{\lambda} - 1$. For that we integrate by parts but one sees that the boundary term does not vanish. Well, the boundary term for the case $\lambda - 1 \le |x| \le \lambda + 1$ is of the form

$$\int_{\Gamma} \partial_i \xi_\lambda \rho n_j \, \mathrm{d}S,$$

with Γ some part of the boundary of the ball $\{|z| \le \lambda\}$ and n_j the normal vector on the ball in direction j. Note that on Γ we have

$$\partial_i \xi_i(x) \le \frac{1}{\lambda(2+\lambda)}$$

and ρ is bounded, because Γ can be at most of length 2. So the extra term is of order $O(\lambda^{-2})$. The non-boundary term after integration by parts as the argument before gives something of order $O(\lambda^{-3})$. Multiplying with a_{ij} which is of order λ^2 on $\{\lambda - 1 \le |x| \le \lambda + 1\}$ keeps everything bounded. A similar argument work for $(2+\lambda)e^{\lambda} - 3 \le |x| \le (2+\lambda)e^{\lambda} - 1$.

What we get is that there exists $\lambda_0 > 0$ and $C > 0$ (may be different than the C defined before) such that

$$|a_{ij}(x,t)\partial_{ij}\zeta_\lambda(x)| \le C, \qquad \text{for all } (x,t) \in \mathbb{R}^d \times [0,1] \text{ and } \lambda \ge \lambda_0.$$

Also it is not difficult to see that for all $x \in \mathbb{R}^d$ we have $\partial_{ij}\zeta_\lambda(x)$ for all λ large enough. Hence

$$\lim_{\lambda \to \infty} |a_{ij}(x,t)\partial_{ij}\zeta_\lambda(x)| = 0.$$

Now dominated convergence gives

$$\lim_{\lambda \to \infty} \int_0^1 \int_{\mathbb{R}^d} |a_{ij}(x,t)\partial_{ij}\zeta_\lambda(x)| \, \mu_t(\mathrm{d}x) \, \mathrm{d}t.$$

This shows (2).

(3) The last point follows similarly, in fact

$$|\nabla\zeta_\lambda(x)^\top a(x,t)\nabla\zeta_\lambda(x)| \le \frac{C(1+|x|^2)}{\lambda(1+|x|)^2} = \frac{C}{\lambda}, \qquad \text{for all } (x,t) \in \mathbb{R}^d \times [0,1].$$

Hence, it converges uniformly so that the integral goes to zero since we are dealing with a probability measure.

In addition, we claimed that everything holds for ζ_λ^2 as well, but that is immediate by the product rule for L and the gradient ∇ and $|\zeta_\lambda^2| \le 1$. $\square$

Appendix B

Relative entropy

In this section we define relative entropy and show some results concerning it. Let $\mathcal{X}$ be a Polish space.

Definition B.1 (Relative entropy)**.** For two probability measures $\mathbb{p}, \mathbb{q} \in \mathcal{P}(\mathcal{X})$ we define the *relative entropy* of $\mathbb{p}$ with respect to $\mathbb{q}$ by

$$H(\mathbb{p}|\mathbb{q}) := \mathbb{E}_{\mathbb{p}}\left[\log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}\right)\right],$$

whenever $\mathbb{p} \ll \mathbb{q}$ and $H(\mathbb{p}|\mathbb{q}) = \infty$ otherwise.

We see an expectation, so we may wonder whether the expectation is well-defined. The integrand may have two different signs and that leads us to wonder what happens to the positive and negative part. Note that the expectation of a random variable X is well-defined as long as at least one of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ is finite. In case only one of them is infinite, then there is no problem because we can add infinities to a finite number. In other words, that avoids $\infty - \infty$ which makes every mathematician's eye balls fly out. Let us confirm this for the relative entropy by the following:

Proposition B.2 (Lemma 1.4.1 from [11])**.** *For any $\mathbb{p}, \mathbb{q} \in \mathcal{P}(\mathcal{X})$ the relative entropy is well-defined, non-negative and $H(\mathbb{p}|\mathbb{q}) = 0$ if and only if $\mathbb{p} = \mathbb{q}$.*

The relative entropy has a lot of properties and the fundamental ones that we use are stated in the following proposition:

Proposition B.3 (Lemma 1.4.3 from [11])**.** *Let $\mathcal{X}$ be a Polish space, then the following holds:*

(i) (strict convexity) the relative entropy $\mathbb{p} \mapsto H(\mathbb{p}|\mathbb{q})$ is a strictly convex function on $\{\mathbb{p} \in \mathcal{P}(\mathcal{X}) :\ H(\mathbb{p}|\mathbb{q}) < \infty\}$;

(ii) (lower semi-continuity) the relative entropy $(\mathbb{p}, \mathbb{q}) \mapsto H(\mathbb{p}|\mathbb{q})$ is lower semi-continuous, that is for any sequence $(\mathbb{p}_n, \mathbb{q}_n) \to (\mathbb{p}, \mathbb{q})$ weakly (i.e. converges in the product topology of weak convergence), we have

$$H(\mathbb{p}|\mathbb{q}) \leq \liminf_{n\to\infty} H(\mathbb{p}_n|\mathbb{q}_n).$$

(iii) (compact level sets) the level set

$$\{\mathbb{p} \in \mathcal{P}(\mathcal{X}) : \; H(\mathbb{p}|\mathbb{q}) \le c\}$$

is a compact subset of $\mathcal{P}(\mathcal{X})$ under the topology of weak convergence.

Lemma B.4 (Conditioning and relative entropy)**.** *Let $(\mathcal{X}, \mathcal{F})$ be a measurable Polish space. Let $\mathbb{p}, \mathbb{q} \in \mathcal{P}(\mathcal{X})$ and $Y : (\mathcal{X}, \mathcal{F}) \to (S, \mathcal{S})$ be an arbitrary random variable. Let μ and ν be the law of Y under $\mathbb{p}$ and $\mathbb{q}$ respectively. Under the assumption $\mathbb{p} \ll \mathbb{q}$ we have*

$$H(\mathbb{p}|\mathbb{q}) = H(\mu|\nu) + \int H(\mathbb{p}^y|\mathbb{q}^y)\,\mu(\mathrm{d}y).$$

Proof. First we argue that $y \mapsto H(\mathbb{p}^y|\mathbb{q}^y)$ is measurable as a map from $(S, \mathcal{S})$ to the space of probability measures $(\mathcal{P}(\mathcal{X}), \mathcal{B}(\mathcal{P}(\mathcal{X}))$ where the target space is endowed with the Borel σ-algebra induced by the topology of weak convergence. We know $(\mathbb{p}, \mathbb{q}) \mapsto H(\mathbb{p}|\mathbb{q})$ is lower semi-continuous which makes it measurable. Moreover, the regular conditional probabilities $\mathbb{p}^y$ and $\mathbb{q}^y$ exist and the map $y \mapsto (\mathbb{p}^y, \mathbb{q}^y)$ is measurable (see Theorem A.5.2 in [11]). Therefore the decomposition

$$[(\mathbb{p}, \mathbb{q}) \mapsto H(\mathbb{p}|\mathbb{q})] \circ [y \mapsto (\mathbb{p}^y, \mathbb{q}^y)] = [y \mapsto H(\mathbb{p}^y|\mathbb{q}^y)],$$

is measurable. Now as soon as one is convinced that

$$\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}} = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y)\frac{\mathrm{d}\mathbb{p}^{y=Y}}{\mathrm{d}\mathbb{q}^{y=Y}}$$

holds true, we get the claim fast. Of course everything written is well-defined for we have $\mathbb{p} \ll \mathbb{q}$ implies $\mu \ll \nu$ and $\mathbb{p}^y \ll \mathbb{q}^y$ for μ-a.s. y. Also the Radon-Nikodym derivative $(y, x) \mapsto \frac{\mathrm{d}\mathbb{p}^y}{\mathrm{d}\mathbb{q}^y}(x)$ has a jointly measurable version (See Theorem A.5.7 in [11]). To be convinced, look at

$$\mathbb{E}_{\mathbb{q}}\left[\mathbb{1}_A \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y)\frac{\mathrm{d}\mathbb{p}^{y=Y}}{\mathrm{d}\mathbb{q}^{y=Y}}\right] = \mathbb{E}_{\mathbb{q}}\left[\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y)\mathbb{E}_{\mathbb{q}^{y=Y}}\left[\mathbb{1}_A \frac{\mathrm{d}\mathbb{p}^{y=Y}}{\mathrm{d}\mathbb{q}^{y=Y}}\right]\right] = \mathbb{E}_{\mathbb{q}}\left[\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y)\mathbb{E}_{\mathbb{p}^{y=Y}}[\mathbb{1}_A]\right],$$

and we know $\mathbb{E}_{\mathbb{p}^{y=Y}}[\mathbb{1}_A] = \mathbb{E}_{\mathbb{p}}[\mathbb{1}_A \mid Y]$, so that

$$\mathbb{E}_{\mathbb{q}}\left[\mathbb{1}_A \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Y)\frac{\mathrm{d}\mathbb{p}^{y=Y}}{\mathrm{d}\mathbb{q}^{y=Y}}\right] = \mathbb{E}_{\mathbb{p}}[\mathbb{E}_{\mathbb{p}}[\mathbb{1}_A \mid Y]] = \mathbb{p}(A).$$

This allows us to conclude

$$\mathbb{E}_{\mathbb{p}}\left[\log\left(\frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}\right) \mid Y = y\right] = \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(y)\right) + \log\left(\frac{\mathrm{d}\mathbb{p}^y}{\mathrm{d}\mathbb{r}^y}\right),$$

which after integrating with respect to μ, or if you prefer the tower property, yields the claim. $\square$

We will state and prove Lemma 4.1 from [3] which is a crucial lemma for proving the Markov property of the minimizer of the Schrödinger problem. The setting is as follows. Consider two probability spaces $(\mathcal{X}_1, \mathcal{F}_1, \mathbb{q}_1)$ and $(\mathcal{X}_2, \mathcal{F}_2, \mathbb{q}_2)$ with Polish state spaces. Define π_1 and π_2 on $\mathcal{X}_1 \times \mathcal{X}_2$ to be the projections on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively. For any probability measure $\mathbb{p} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ we denote $\mathbb{p}_i := \pi_{i\#}\mathbb{p}$, $i = 1, 2$, i.e. its marginal on $\mathcal{X}_i$. Under this setting we introduce

Lemma B.5. *Under the setting described above we have for any $\mathbb{p} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ such that $H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) < \infty$*

$$H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) \geq H(\mathbb{p}_1|\mathbb{q}_1) + H(\mathbb{p}_2|\mathbb{q}_2) = H(\mathbb{p}_1 \otimes \mathbb{p}_2|\mathbb{q}_1 \otimes \mathbb{q}_2).$$

Proof. By the assumption $H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) < \infty$, we obviously have $\mathbb{p} \ll \mathbb{q}_1 \otimes \mathbb{q}_2$. We apply the conditioning from Lemma B.4 to get

$$H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) = H(\mathbb{p}_1|\mathbb{q}_1) + \int_{\mathcal{X}_1} H(\mathbb{p}^{x_1} \mid \mathbb{q}^{x_1})\, \mathbb{p}_1(\mathrm{d}x_1).$$

As it is intuitively clear, conditioning on $X_1 = x_1$ leaves only X_2 as a "random variable". Indeed, for any $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ we have

$$\mathbb{p}^{x_1}(A \times B) = \mathbb{1}_A(x_1)\mathbb{p}_2^{x_1}(B), \quad \text{and} \quad \mathbb{q}^{x_1}(A \times B) = \mathbb{1}_A(x_1)\mathbb{q}_2^{x_1}(B).$$

Together with the fact that X_1 and X_2 are independent under $\mathbb{q}$ brings us to

$$H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) = H(\mathbb{p}_1|\mathbb{q}_1) + \int_{\mathcal{X}_1} H(\mathbb{p}_2^{x_1} \mid \mathbb{q}_2)\, \mathbb{p}_1(\mathrm{d}x_1).$$

Let us look at $x_1 \mapsto \mathbb{p}_2^{x_1}$ as a map to probability measures which is in fact a measurable map (see Theorem A.5.2 in [11]). Through change of variables, we can integrate over probability measures on

$$\mathbb{p}_2^{\mathcal{X}_1} := \{\mu \in \mathcal{P}(\mathcal{X}_2) : \ \mu = \mathbb{p}_2^{x_1} \text{ for some } x_1 \in \mathcal{X}_1\},$$

with respect to the push-forward

$$\rho := (x_1 \mapsto \mathbb{p}_2^{x_1})_{\#}\mathbb{p}_1.$$

Basically, that is nothing else but

$$\int_{\mathcal{X}_1} H(\mathbb{p}_2^{x_1} \mid \mathbb{q}_2)\, \mathbb{p}_1(\mathrm{d}x_1) = \int_{\mathbb{p}_2^{\mathcal{X}_1}} H(\mu|\mathbb{q}_2)\, \rho(\mathrm{d}\mu).$$

We want to apply Jensen's inequality to get the desired inequality. We need a general version of Jensen's inequality as stated in Theorem 3.10 in [33]. We need a topological locally convex vector space which we do not have. Indeed the set of probability measures is not a vector space. However, the set of regular bounded additive set functions denoted by $\mathrm{rba}(\mathcal{X}_2)$ is the dual space of $C_b(\mathcal{X}_2)$ and is a vector space. Endowing $\mathrm{rba}(\mathcal{X}_2)$ with the weak*-topology makes it locally convex (See Proposition 3.12 in [5]). Moreover $\mathcal{P}(\mathcal{X}_2) \subset \mathrm{rba}(\mathcal{X}_2)$ and $\mathcal{P}(\mathcal{X}_2)$ is actually endowed with subset weak*-topology. Therefore, the measure ρ that integrates over $\mathbb{p}_2^{\mathcal{X}_1} \subset \mathcal{P}(\mathcal{X}_2)$ can be set to be zero on $\mathrm{rba}(\mathcal{X}_2) \setminus \mathbb{p}_2^{\mathcal{X}_1}$ making it a probability measure on $\mathrm{rba}(\mathcal{X}_2)$.

We have made clear that although $\mathcal{P}(\mathcal{X}_2)$ itself does not satisfy the conditions required in [33], it is a subset of a space satisfying the conditions. So we may apply Jensen's inequality to get

$$\int_{\mathbb{p}_2^{\mathcal{X}_1}} H(\mu|\mathbb{q}_2)\,\rho(\mathrm{d}\mu) \geq H\left(\int_{\mathbb{p}_2^{\mathcal{X}_1}} \mu\,\rho(\mathrm{d}\mu) \mid \mathbb{q}_2\right),$$

where the integral $\int_{\mathbb{p}_2^{\mathcal{X}_1}} \mu\,\rho(\mathrm{d}\mu)$ is understood in the sense of Pettis integration (See the definition given in [33]). One can check by the definition of Pettis integral which states that for any $f \in (\mathcal{P}(\mathcal{X}_2))'$, i.e. f in the dual of $\mathcal{P}(\mathcal{X}_2)$, we have

$$\left\langle \int_{\mathbb{p}_2^{\mathcal{X}_1}} \mu\,\rho(\mathrm{d}\mu), f \right\rangle = \int_{\mathbb{p}_2^{\mathcal{X}_1}} \langle \mu, f\rangle\,\rho(\mathrm{d}\mu) = \int_{\mathbb{p}_2^{\mathcal{X}_1}} \left(\int_{\mathcal{X}_2} f(x_2)\,\mu(\mathrm{d}x_2)\right)\rho(\mathrm{d}\mu),$$

where $\langle \cdot, \cdot \rangle$ denotes the duality brackets. Change of variables gives

$$\left\langle \int_{\mathbb{p}_2^{\mathcal{X}_1}} \mu\,\rho(\mathrm{d}\mu), f \right\rangle = \int_{\mathcal{X}_1} \left(\int_{\mathcal{X}_2} f(x_2)\,\mathbb{p}_2^{x_1}(\mathrm{d}x_2)\right)\mathbb{p}_1(\mathrm{d}x_1) = \int_{\mathcal{X}_2} f(x_2)\,\mathbb{p}_2(\mathrm{d}x_2),$$

by the tower property. We know that $C_b(\mathcal{X}_2) \subset (\mathcal{P}(\mathcal{X}_2))'$ and $C_b(\mathcal{X}_2)$ separates probability measures (Theorem 13.11 in [23]) allowing us to conclude that

$$\int_{\mathbb{p}_2^{\mathcal{X}_1}} \mu\,\rho(\mathrm{d}\mu) = \mathbb{p}_2.$$

Gathering everything yields

$$H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) \geq H(\mathbb{p}_1|\mathbb{q}_1) + H(\mathbb{p}_2|\mathbb{q}_2).$$

It remains to prove the last equality. It follows readily from

$$\frac{\mathrm{d}\mathbb{p}_1 \otimes \mathbb{p}_2}{\mathrm{d}\mathbb{q}_1 \otimes \mathbb{q}_2} = \frac{\mathrm{d}\mathbb{p}_1}{\mathrm{d}\mathbb{q}_1} \cdot \frac{\mathrm{d}\mathbb{p}_2}{\mathrm{d}\mathbb{q}_2},$$

and by the property logarithm of product is the sum of logarithms. Keeping these facts in mind, we finish by noting that the assumption that $H(\mathbb{p}|\mathbb{q}_1 \otimes \mathbb{q}_2) < \infty$ makes both $H(\mathbb{p}_1|\mathbb{q}_1)$ and $H(\mathbb{p}_2|\mathbb{q}_2)$ finite allowing us to put them together to get

$$H(\mathbb{p}_1|\mathbb{q}_1) + H(\mathbb{p}_2|\mathbb{q}_2) = H(\mathbb{p}_1 \otimes \mathbb{p}_2|\mathbb{q}_1 \otimes \mathbb{q}_2).$$

□

Theorem B.6 (Radon-Nikodym under a bijection)**.** *Let $\varphi : (S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$ be a measurable invertible function. Assume $\mathbb{p} \ll \mathbb{q}$ are two σ-finite measures. Then $\mathbb{p} \circ \varphi \ll \mathbb{q} \circ \varphi$ and*

$$\frac{\mathrm{d}\mathbb{p} \circ \varphi}{\mathrm{d}\mathbb{q} \circ \varphi} = \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}} \circ \varphi.$$

Proof. Note that for any $A \in \mathcal{S}_1$ by change of variables/measures formula

$$\int_A \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}(\varphi(x))\mathbb{q} \circ \varphi(dx) = \int_{\varphi(A)} \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}(\varphi(\varphi^{-1}(x')))\mathbb{q}(dx') = \int_{\varphi(A)} \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}}(x')\mathbb{q}(dx') = \mathbb{p} \circ \varphi(A).$$

Hence $\mathbb{p} \circ \varphi \ll \mathbb{q} \circ \varphi$ and by uniqueness of Radon-Nikdoym derivative

$$\frac{\mathrm{d}\mathbb{p} \circ \varphi}{\mathrm{d}\mathbb{q} \circ \varphi} = \frac{\mathrm{d}\mathbb{p}}{\mathrm{d}\mathbb{q}} \circ \varphi.$$

$\square$

A straightforward corollary of this fact is

Corollary B.7 (Relative entropy under a bijection)**.** *Consider a measurable invertible function $\varphi : (S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$. Then for all probability measures $\mathbb{p}$ and $\mathbb{q}$*

$$H(\mathbb{p}|\mathbb{q}) = H(\mathbb{p} \circ \varphi|\mathbb{q} \circ \varphi).$$

References

[1] R. Aebi. *Schrodinger diffusion processes.* Basel Birkhauser, 1996.

[2] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators.* Vol. 348. Grundlehren der mathematischen Wissenschaften. Cham: Springer International Publishing, 2014.

[3] A. Baradat and C. Léonard. *Minimizing relative entropy of path measures under marginal constraints.* 2020. arXiv: `2001.10920 [math.PR]`.

[4] A. Beurling. "An automorphism of product measures". In: *Ann. Math. (2)* 72 (1960).

[5] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* 2010.

[6] E. A. Carlen. "Conservative diffusions". In: *Commun. Math. Phys.* 94 (1984).

[7] P. Cattiaux and C. Léonard. "Correction to: Minimization of the Kullback information of diffusion processes". In: *Ann. Inst. Henri Poincaré, Probab. Stat.* 31.4 (1995).

[8] P. Cattiaux and C. Léonard. "Minimization of the Kullback information for some Markov processes". In: *Séminaire de probabilités XXX.* Berlin: Springer, 1996.

[9] Y. Chen, T. T. Georgiou, and M. Pavon. *Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schroedinger bridge.* 2020. arXiv: `2005.10963 [math.OC]`.

[10] I. Csiszar. "I-Divergence Geometry of Probability Distributions and Minimization Problems". In: *The Annals of Probability* 3.1 (1975).

[11] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* John-Wiley & Sons, 1997.

[12] A. Eberle. *Markov Processes.* 2021.

[13] L. C. Evans. *Partial Differential Equations.* 2nd ed. American Mathematical Society, 2010.

[14] H. Föllmer. *Random fields and diffusion processes.* Calcul des probabilités, Éc. d'Été, Saint-Flour/Fr. 1985-87, Lect. Notes Math. 1362, 101-203. 1988.

[15] H. Föllmer. "Time reversal on Wiener space". In: Springer, Berlin, Heidelberg, 1986.

[16] H. Föllmer and N. Gantert. "Entropy minimization and Schrödinger processes in infinite dimensions". In: *Annals of Probability* 25.2 (1997).

[17] R. Fortet. "Résolution d'un système d'équations de M. Schrödinger". In: *J. Math. Pures Appl. (9)* 19 (1940).

[18] D. A. Harville. *Matrix algebra from a statistician's perspective.* New York, NY: Springer, 1997.

[19] H. Herrlich. *Axiom of choice.* Vol. 1876. Berlin: Springer, 2006.

[20] J. Jacod. *A general theorem of representation for martingales.* Probab., Proc. Symp. Pure Math., Urbana 1976, 37-53. 1977.

[21] B. Jamison. "Reciprocal processes". In: *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 30 (1974).

[22] O. Kallenberg. *Foundations of modern probability*. Second. Springer-Verlag, 2002.

[23] A. Klenke. *Probability theory : a comprehensive course*. English. Cham, 2020.

[24] C. Léonard. "A survey of the Schrödinger problem and some of its connections with optimal transport". In: *Discrete and Continuous Dynamical Systems- Series A* 34.4 (Aug. 2013). arXiv: `1308.0215`.

[25] C. Léonard. "Girsanov theory under a finite entropy condition". In: *Séminaire de probabilités XLIV*. Berlin: Springer, 2012.

[26] N. Liviu I. *Lectures On The Geometry Of Manifolds*. 2nd Ed. World Scientific, 2007.

[27] G. Metafune et al. "Lp-regularity for elliptic operators with unbounded coefficients". In: *Adv. Differ. Equ.* 10.10 (2005).

[28] T. Mikami. "Stochastic optimal transport revisited". In: *SN Partial Differential Equations and Applications* 2.1 (2021).

[29] T. Mikami. "Variational processes from the weak forward equation". In: *Commun. Math. Phys.* 135.1 (1990).

[30] M. Nagasawa. *Schrödinger equations and diffusion theory*. Vol. 86. Basel: Birkhäuser Verlag, 1993.

[31] E. Nelson. *Quantum fluctuations*. Princeton University Press, Princeton, NJ, 1985.

[32] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Vol. 44. Applied Mathematical Sciences. New York, NY: Springer New York, 1983.

[33] M. D. Perlman. "Jensen's inequality for a convex vector-valued function on an infinite-dimensional space". In: *Journal of Multivariate Analysis* 4.1 (1974).

[34] S. Reich. "Data assimilation: the Schrödinger perspective". In: *Acta Numerica* 28 (2019).

[35] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Third. Springer-Verlag, 2004.

[36] W. Rudin. *Functional analysis*. New York : McGraw-Hill, 1991.

[37] L. Rüschendorf and W. Thomsen. "Closedness of Sum Spaces and the Generalized Schrödinger Problem". In: *Theory of Probability & Its Applications* 42.3 (1998).

[38] L. Rüschendorf and W. Thomsen. "Note on the Schrödinger equation and I-projections". In: *Statistics and Probability Letters* 17.5 (1993).

[39] I. N. Sanov. "On the probability of large deviations of random variables". In: *Sel. Transl. Math. Stat. Probab.* 1 (1961).

[40] R. L. Schilling and L. Partzsch. *Brownian Motion*. De Gruyter, 2014.

[41] E. Schrödinger. "Über die Umkehrung der Naturgesetze". In: *Sitzungsber. Preuß. Akad. Wiss., Phys.-Math. Kl.* 1931 (1931).

[42] G. Simons. "An unexpected expectation". In: *Ann. Probab.* 5 (1977).

[43] D. W. Stroock and M. Yor. "On extremal solutions of martingale problems". In: *Annales scientifiques de l'École Normale Supérieure* Ser. 4, 13.1 (1980).

[44] L. Tamanini. "Analysis and Geometry of RCD spaces via the Schrödinger problem". PhD thesis. Sept. 2017.