Martingale Schrödinger Bridges and Optimal Semistatic Portfolios

Long Zhao

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023
Abstract
Martingale Schrödinger Bridges and Optimal Semistatic Portfolios
Long Zhao

This thesis studies the problems of semistatic trading strategies in a discrete-time financial market, where stocks are traded dynamically and European options at maturity are traded statically. First, we show that pointwise limits of semistatic trading strategies are again semistatic strategies. The analysis is carried out in full generality for a two-period model, and under a probabilistic condition for multi-period, multi-stock models. Our result contrasts with a counterexample of Acciaio, Larsson and Schachermayer, and shows that their observation is due to a failure of integrability rather than instability of the semistatic form. Mathematically, our results relate to the decomposability of functions as studied in the context of Schrödinger bridges.

Second, we study the so-called martingale Schrödinger bridge $Q_*$ in a two-period financial market; that is, the minimal-entropy martingale measure among all models calibrated to option prices. This minimization is shown to be in duality with an exponential utility maximization over semistatic portfolios. Under a technical condition on the physical measure $P$, we show that an optimal portfolio exists and provides an explicit solution for $Q_*$. Specifically, we exhibit a dense subset of calibrated martingale measures with particular properties to show that the portfolio in question has a well-defined and integrable option position.
# Table of Contents

Chapter 1: Introduction ................................................. 1
  1.1 Semistatic Trading Strategies .................................. 2
  1.2 Martingale Schrödinger Bridges ............................... 4

Chapter 2: Limits of Semistatic Trading Strategies ................. 7
  2.1 Introduction .................................................. 7
  2.2 Two-Period Model ............................................ 10
  2.3 General Probabilistic Model ................................... 18
  2.4 Proof of Proposition 2.3.2 .................................. 22

Chapter 3: Martingale Schrödinger Bridges and Optimal Semistatic Portfolios ............. 37
  3.1 Introduction and Main Results ................................. 37
  3.2 Wellposedness and Duality ................................... 44
  3.3 Admissibility and Dual Attainment ............................ 51
    3.3.1 Preliminary Considerations ............................... 51
    3.3.2 Proof of Theorem 3.1.3 .................................. 52
    3.3.3 Proof of Proposition 3.3.2 ............................... 55

Bibliography ............................................................... 67
Acknowledgments

I would like to express my earnest gratitude to my advisor, Prof. Marcel Nutz, for his unwa-
vering support and invaluable vision during my Ph.D. studies. It has been my great honor to be a
doctoral student in the Department of Statistics at Columbia University, where I am very fortunate
to also collaborate with Prof. Johannes Wiesel and Prof. Kevin Webster on projects in mathemat-
ical finance. I would like to thank Prof. Richard Davis, Prof. Julien Guyon, Prof. Daniel Lacker
and Prof. Johannes Wiesel for readily accepting to serve on my dissertation defense committee.

It has been valuable experience for me to serve as a teaching assistant to Prof. Marcel Nutz and
Prof. Richard Davis on probability theory, to Prof. Kevin Webster on price impact modeling and
algorithmic trading, to Prof. Julien Guyon and Prof. Bryan (Jianfeng) Liang on nonlinear option
pricing, and to Prof. Harvey Stein on risk management and regulation, all of which has greatly
helped me integrate my theoretical rigor with practical wisdom. I also thank all of the students for
their thought-provoking questions and constructive feedback.

During my journey in New York City, I have benefited from numerous discussions with col-
leagues within and outside of Columbia University and truly enjoyed my time living here. Last but
not the least, I am deeply indebted to my family and friends for their unbounded love and support.
I am now looking forward to continuing my wonderful journey in Singapore.
This page intentionally left blank.
以梦为马，以汗为泉，
不忘初心，不负韶华。
This page intentionally left blank.
Chapter 1: Introduction

Starting with the seminal work of Fischer Black, Myron Scholes and Robert Merton [10, 41] in the early seventies, the pricing and hedging of financial derivatives have become the cornerstone of modern mathematical finance. Their key idea is to replicate option payoffs by dynamically trading the underlying assets to eliminate market risk, which is now commonly known as delta hedging.

The principle of no arbitrage crystallizes this idea, stating that the financial market in equilibrium does not admit opportunities of possible profits that come with no downside risk. Gradually developed since the late seventies, a rather general framework (see [16, 17, 32, 33, 38, 55], among others), collectively known as the Fundamental Theorem of Asset Pricing, reveals the intimate relation between the no-arbitrage principle on the one hand, and the martingale theory on the other hand. Behind the Fundamental Theorem lie the concepts of dynamic trading in self-financing portfolios and the collection of equivalent martingale measures.

In the present thesis, we consider semistatic trading strategies which, in addition to a dynamic position in some financial asset, include a static (buy-and-hold) position in options written on this asset. We connect the semistatic portfolios with the set of calibrated martingale measures, which are pricing models that can correctly reproduce option prices observed in the market. To this end, we study the so-called martingale Schrödinger bridge — the minimal-entropy martingale measure among all models calibrated to option prices — and establish a duality result with an exponential utility maximization problem over semistatic portfolios.

This thesis is organized as follows. Chapter 1 recalls some widely used terminology in mathematical finance and introduces the problems that we are about to address. Chapter 2 focuses on semistatic trading strategies and shows that the set of such strategies is closed under pointwise convergence. Chapter 3 studies the problem of martingale Schrödinger bridges and presents the duality result. Chapter 2 and 3 are self-contained and can be read independently.
1.1 Semistatic Trading Strategies

We motivate our problem from the well-known concept of dynamic portfolios in discrete time. For simplicity of notation, we use $\mathbb{R}$-valued stochastic processes and model only one financial asset. Our introduction can be easily extended to $\mathbb{R}^d$-valued processes for $d$ assets.

Given a fixed investment horizon $T \geq 1$ and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, P)$, the prices of a financial asset in discrete time can be modeled by an adapted stochastic process $X = (X_t)_{0 \leq t \leq T}$. A dynamic trading strategy is a predictable process $H = (H_t)_{1 \leq t \leq T}$ and at time $T$ gives rise to the final investment outcome 

$$(H \cdot X)_T := \sum_{t=1}^{T} H_t (X_t - X_{t-1}).$$

For a measurable function $g : \mathbb{R} \rightarrow \mathbb{R}$ with suitable integrability conditions, we call $g(X_T)$ a European option written on $X$ with maturity $T$. Common examples include call options $g(X_T) = \max\{X_T - K, 0\}$ and put options $g(X_T) = \max\{K - X_T, 0\}$ for a range of strikes $K$.

In the past five decades, significant progress has been made to develop a general framework of no-arbitrage principle and options pricing theory using the dynamic trading strategies. With rising popularity in options trading, it seems equally natural to develop a theory centering on the semistatic trading strategies $(H, g)$, which combines a dynamic position in $X$ and a static (buy-and-hold) position in $g(X_T)$ with the final outcome

$$V = (H \cdot X)_T + g(X_T)$$

at time $T$, where $g$ can take an arbitrary functional form as long as it satisfies some integrability conditions. The static nature of the option position reflects the increased transaction cost relative to directly trading $X$ itself. Such formulation has been extensively studied in literature before; see [1, 5, 21, 26] and the references therein.

Closedness properties of trading strategies have played a pivotal role in mathematical finance.
They are at the heart of the separation arguments underlying many fundamental results. However, it has been observed by Acciaio, Larsson and Schachermayer [1, Theorem 1.1] that the space of trading outcomes from semistatic strategies lacks good closedness properties for a theory to be developed along the usual approach. More specifically, they give an example of a two-period model on a countable sample space $\Omega$ and exhibit a sequence of semistatic strategies

$$0 \leq V^{(n)} = (H^{(n)} \cdot X)_T + g^{(n)}(X_T) \to V$$

that converges almost surely and in $L^p$ for every $p \in [1, \infty)$ to $V$ as $n \to \infty$, but the limiting random variable $V$ cannot be written as, or even dominated by, $(H \cdot X)_T + g(X_T)$ for some semistatic strategy $(H, g)$ such that $(H \cdot X)_T$ and $g(X_T)$ are separately integrable.

Despite the aforementioned example, in Chapter 2 we aim to provide some positive closedness results on semistatic trading strategies in the discrete-time setting. The first main result focuses on a two-period model and, in a non-probabilistic setting, shows that the limit of pointwise-convergent semistatic strategies is again a semistatic strategy. Returning to the example in [1], our result indicates that the main failure relates to the integrability of the limiting option position rather than the semistatic functional form. Of course, one may hope that specific limiting portfolios nevertheless enjoy good integrability properties; see Chapter 3 for detailed discussion.

In our second main result of Chapter 2, we exhibit a reasonably weak financial/probabilistic condition that enables a general closedness result by a more generic analysis. Indeed, our result covers the standard setting with any (finite) number of stocks and periods, and options on all individual stocks, and with modest efforts it could be adapted for options at intermediate dates, options only on some of the stocks, or similar scenarios. The proposed financial condition is that the reference measure $P$ of the model be equivalent to (i.e., have the same nullsets as) a product measure. Intuitively, this means that none of the (a priori possible) future stock prices become completely impossible given an intermediate state of the prices.
1.2 Martingale Schrödinger Bridges

The Schrödinger bridge problem has a long history in physics [52]; see [22, 39, 42] for surveys. In its static form, the classical problem studies minimization of relative entropy over the set of probability couplings $\Pi(\mu, \nu)$ on a product space $X \times Y$ with given marginals $\mu, \nu$ with respect to a reference measure $R$:

$$\inf_{\pi \in \Pi(\mu, \nu)} H(\pi | R),$$

where the relative entropy (or Kullback–Leibler divergence) is defined as

$$H(\pi | R) := \begin{cases} E^\pi \left[ \log \frac{d\pi}{dR} \right], & \pi \ll R \\ \infty, & \pi \not\ll R. \end{cases}$$

The martingale Schrödinger bridge was introduced by Henry-Labordère [34] as a pricing model achieving perfect calibration to all Vanilla options while retaining stylized facts of a reference model. With increasing activities in exchange-traded, standardized call and put options, it has become straightforward for market participants to observe and react to option prices written on individual securities or indexes with various strikes and maturities. By the Breeden–Litzenberger formula [13], the risk-neutral distribution $\nu$ of $X_T$ can be inferred from the observed option prices. More specifically, when the interest rate is assumed to be zero, the probability density $f$ of $X_T$ is

$$f(K) = \frac{\partial^2 C(K, T)}{\partial K^2},$$

where $C(K, T)$ denotes the market price of a call option with strike $K$ and maturity $T$. It is natural to require a pricing model correctly reproduce the prices of actively traded instruments, and the arbitrage-free price of a general option $g(X_T)$ should be given by the integral $E^\nu[g]$. Numerous pricing models, including the Heston model, the SABR model and the Bergomi model, have adopted a stochastic volatility in the dynamics of $X$, and they can generate implied
volatility smiles or skews that are consistent with market observations; see [8] for a comprehensive list of examples. As these models typically depend on a finite set of parameters, they are unable to perfectly reproduce all option prices observed in the market.

Starting from a reference pricing model that cannot be perfectly calibrated, the martingale Schrödinger bridge is constructed as the calibrated measure that is closest to the reference model in the sense of relative entropy. In contrast to the classical Schrödinger bridge in Léonard [39] and Avellaneda et al. [3, 4], this problem features an additional martingale constraint to generate an arbitrage-free model. A similar approach is used by Guyon [30, 31] in a two-period setting to solve the longstanding joint S&P 500/VIX smile calibration puzzle; here entropy minimization is utilized to construct a model that is jointly calibrated to the S&P 500, VIX futures and VIX options.

In Chapter 3, we prove strong duality between the martingale Schrödinger bridge problem and an exponential utility maximization problem over semistatic portfolios. On the one hand, this duality, as well as the existence of the martingale Schrödinger bridge itself (primal attainment), is obtained along the lines of classical entropy minimization and Schrödinger bridge theory [14, 22]. On the other hand, we prove (under a technical condition) that the dual problem is attained in a natural space of admissible portfolios, and that this dual solution yields the log-density of the martingale Schrödinger bridge.

We thus derive from first principles the type of implicit condition assumed on the optimal log-density, e.g., in Guyon [31, Theorem 16], and overcome the non-closedness issue discovered in Acciaio et al. [1]. To wit, while in general a convergent sequence of semistatic portfolios may have an undesirable limit with unclear financial interpretation, the specific limit of a utility-maximizing sequence in our problem is shown to be an admissible portfolio. Since we have not been able to use the classical arguments of mathematical finance, we hope that our results can open the door to developing other aspects of mathematical finance in the semistatic setting.
This page intentionally left blank.
Chapter 2: Limits of Semistatic Trading Strategies

The chapter is based on the article [45] of the same title, authored by Marcel Nutz, Johannes Wiesel and Long Zhao. It is published in Mathematical Finance.

2.1 Introduction

Closedness properties of trading strategies play a pivotal role in mathematical finance. They are at the heart of the separation arguments underlying the Fundamental Theorem of Asset Pricing, the superhedging duality, the existence of optimal portfolios for utility maximization problems, and other key results (see, e.g., [18, 24] and the references therein). In the most classical setting, strategies refer to dynamic trading in a stock or other liquidly traded securities. Closedness refers to the limit of a sequence of final outcomes from self-financing trading again being an outcome of some admissible trading strategy, or at least being dominated by one.

In addition to dynamic trading in a stock \((X_t)\), a semistatic strategy allows for buy-and-hold trading in options written on \(X\), usually European options maturing at the time horizon \(T\) of the model. The static nature of this position reflects the increased trading cost relative to the stock. By combining options such as calls with different strikes, the trader can approximate the payoff \(g(X_T)\) for an arbitrary function \(g\). A linear pricing rule for all such payoffs is equivalent to fixing the risk-neutral distribution of \(X_T\) [13]. Taking this distribution as a primitive, a large body of literature starting with [35] developed the theory of model-free or robust finance, where the distribution of \(X\) is taken to be unknown or partially unknown, respectively, up to a no-arbitrage condition (see [5, 12, 26, 36], among many others).

In the classical setting where \(X\) is modeled on a given probability space, semistatic trading is equally natural but the mathematical foundations of the theory are not well developed. We might
expect this setting to lie between the usual one (with a fixed reference measure but without options) and the model-free one (with options but without reference measure). Maybe surprisingly, this is not the case: it had been observed for some time that the standard arguments for closure and separation do not apply in a straightforward way. The explanation is provided by [1] whose main result consists of two counterexamples, one in discrete and one in continuous time, stating that the space of (final outcomes from) semistatic trading strategies is not closed in the sense defined there. This clearly creates an obstacle to developing the theory along the usual lines. In this paper, we focus on the discrete-time setting and provide positive closedness results. In particular, we illuminate what goes wrong in the example of [1]. Our results open the door to developing other aspects of mathematical finance in the semistatic setting. One such aspect, the existence of an optimal portfolio for the exponential utility maximization problem, is treated in a companion paper [46] which takes our result as its starting point. A related question comes up in [30, 31], where trading also involves the volatility index VIX and it is postulated that a certain optimal log-density has the form of a semistatic portfolio. The arguments of the present work should be useful to deduce this from first principles.

The aforementioned discrete-time example of [1] considers a two-period model of dynamic trading in a stock \((X_t)_{t=0,1,2}\) and static trading in options \(g(X_2)\), where \(g\) is integrable under the law \(\mu\) of \(X_2\). The physical measure \(P\) of the model is taken to be a risk-neutral measure, thus excluding issues related to arbitrage. The authors exhibit a sequence of bounded strategies with nonnegative outcomes converging in \(L^p\) for any finite \(p\) and prove that the limit of the outcomes is not the outcome of an admissible strategy (not even dominated by one). The proof is based on a clever contradiction argument which circumvents studying the limiting random variable in detail.

Our first main result below focuses on a two-period model and a sequence of semistatic strategies converging pointwise. In a non-probabilistic setting, it shows that the limiting outcome is again a semistatic strategy. This result implies stability under almost-sure convergence in a probabilistic setting by application to a set of measure one. As there are no restrictions on the probability measure, the result holds regardless of arbitrage considerations or other probabilistic assumptions.
Returning to the example of [1], our result indicates that the main failure relates to the integrability of the limiting option position rather than the semistatic functional form: the limit is still a sum of dynamic trading and an option; however, the integrability of the option can fail. (This is by no means a negligible issue—it removes the obvious way of assigning a price to the option.) Of course, one may hope that specific limiting portfolios nevertheless enjoy good integrability properties. In the companion paper [46], we prove this for the optimal portfolio of exponential utility maximization.

The proof of our two-period, single stock result involves analyzing some finer algebraic structures of the set where convergence takes place. While our experience suggests that the result may extend to more general models, the complexity of our analysis grows rather quickly, possibly to the extent of becoming infeasible. In our second main result, we aim to exhibit a reasonably weak financial/probabilistic condition that eliminates some of the subtleties and enables a general closedness result by a more generic analysis. Indeed, our result covers the standard setting with any (finite) number of stocks and periods, and options on all individual stocks. The analysis immediately extends to trading constraints such as no-shorting, and with modest efforts it could be adapted for options at intermediate dates, options only on some of the stocks, or similar scenarios. The proposed financial condition is that the reference measure $P$ of the model be equivalent to (i.e., have the same nullsets as) a product measure; namely, the product of all the marginals of the individual stocks at the individual dates. Intuitively, this means that none of the (a priori possible) future stock prices becomes completely impossible given an intermediate state of the prices. We remark that this interpretation is similar to the “conditional full support” assumption known in the theory of transaction costs; see [28] and the literature thereafter.

Remarkably, we have not been able to use the classical arguments of mathematical finance, nor the techniques known from the dual problem of martingale optimal transport [7] in this work, despite the setting lying between those two. Instead, we draw inspiration from the literature on the dual of the Schrödinger bridge problem, especially [11, 50]. See also [39, 42] for general introductions. We shall see that in the present context, the analysis is significantly more involved
as the increments of the stock create an interaction between the variables. Nevertheless, notions such as the connectedness defined in [11] are crucially helpful. The observation that semistatic trading is related to a martingale version of the Schrödinger bridge problem goes back to [34] which discusses a continuous-time problem.

The remainder of this paper is organized as follows. Section 2.2 contains a complete analysis of the two-period, single-stock model. Building on its insights, Section 2.3 proposes a probabilistic condition enabling the analysis of a multi-period, multi-stock model: we show that the problem of closedness reduces to a problem of linear algebra. The latter consists in establishing that a certain matrix has full rank, which is the content of Section 2.4.

2.2 Two-Period Model

In this section, we provide a pointwise convergence analysis for a two-period model with a single stock. In our formulation, the variable \( x \) represents the stock price at date 1 and \( y \) the price at date 2. Equivalently, the stock price process is given by the canonical process \((X, Y)\) on \( \mathbb{R}^2 \). Throughout the section, we fix \( E \subseteq \mathbb{R}^2 \), interpreted as the set of possible states for \((x, y)\).

**Definition 2.2.1.** A function \( v : E \rightarrow \mathbb{R} \) is a semistatic strategy if

\[
v(x, y) = h(x)(y - x) + g(y), \quad (x, y) \in E
\]

for some functions \( h, g : \mathbb{R} \rightarrow \mathbb{R} \). We refer to \( h \) as a stock position and \( g \) as an option position.

In general, \( h \) and \( g \) are not uniquely determined by \( v \), hence one should think of \( v \) as a function \( E \rightarrow \mathbb{R} \) rather than consisting of the pair \( h, g \). We note that trading between the initial date 0 and date 1 is not represented explicitly. This entails no loss of generality: if the initial state is deterministic, an expression of the form \( h_0(x - x_0) + h(x)(y - x) + g(y) \) can always be rewritten as a semistatic strategy in the above sense.

**Theorem 2.2.2.** The set of semistatic strategies is sequentially closed under pointwise convergence.
While the theorem states that a limit of semistatic strategies $v_n$ is again a semistatic strategy, we emphasize that the corresponding stock and option positions do not converge in general. Clearly the theorem entails its probabilistic counterparts: if convergence holds $P$-a.s. under some measure $P$, we can apply the above with $E$ being the set of full measure where convergence holds. Even if $P$ satisfies a no-arbitrage condition, the structure of $E$ can be quite complicated in this context, hence the importance of leaving $E$ general in the theorem.

Theorem 2.2.2 is trivial if $E$ is $\mathbb{R}^2$ or a finite set. However it is much less obvious for generic sets $E$. We state its proof at the end of Section 2.2, after an analysis which also provides detailed insight about what happens with the stock and option positions in the passage to the limit.

**Analysis of the Two-Period Model**

We divide the given set $E \subseteq \mathbb{R}^2$ into its set $E_d := \{(x, y) \in E : x = y\}$ of diagonal points and the complement $E_o := E \setminus E_d$ of off-diagonal points. Consider a semistatic strategy $v(x, y) = h(x)(y - x) + g(y)$ at a point $(x, y) \in E_o$, then as $y - x \neq 0$, the value $h(x)$ uniquely determines $g(y)$, and vice versa. Similarly, for a sequence $v_n(x, y) = h_n(x)(y - x) + g_n(y)$ where $v_n(x, y)$ converges, the value $h_n(x)$ converges if and only if $g_n(y)$ does. For $(x, y) \in E_d$, the situation is quite different: $v(x, y) = h(x)(y - x) + g(y) = g(y)$, so that the option position is determined (or convergent, respectively), whereas $h(x)(y - x) = 0$ irrespectively of the stock position $h(x)$ at $x$. This does not mean that $h(x)$ can be ignored—if $(x, y') \in E$ for some $y' \neq y$, the value $h(x)$ is nevertheless relevant for the strategy.

We first focus our analysis on $E_o$. The following notion was introduced by [11] in a different context.

**Definition 2.2.3.** Two points $(x, y), (x', y') \in E_o$ are connected, denoted $(x, y) \sim (x', y')$, if there exist $k \in \mathbb{N}$ and $(x_i, y_i)_{i=1}^k \in E_o^k$ with $(x_1, y_1) := (x, y)$ such that the points

$$\begin{align*}
(x_1, y_1), (x_2, y_1), (x_2, y_2), (x_3, y_2), \ldots, (x_k, y_k), (x', y_k), (x', y')
\end{align*}$$

(2.1)
all belong to $E_o$. In that case, $(x_i, y_i)_{i=1}^k$ is called a path (in $E_o$) from $(x, y)$ to $(x', y')$. A set $C \subseteq E_o$ is connected (in $E_o$) if any two points in $C$ are connected.

In particular, $(x, y)$ is connected to itself with $k = 1$. For the list (2.1), the crucial property is that only one coordinate is changed in each step. In our notation, the first coordinate changes first, but because a point can be repeated in the list, this entails no loss of generality. We observe that $\sim$ is an equivalence relation on $E_o$. The corresponding equivalence classes $C = \{C_\gamma : \gamma \in \Gamma\}$ are called the connected components of $E_o$.

Given a subset $S \subseteq \mathbb{R}^2$, we denote by $S^x$ and $S^y$ its projections onto the first and second coordinate, respectively. We record some properties of $C_\gamma$ in the following remark.

**Remark 2.2.4.** For each $\gamma \in \Gamma$,

$$C_\gamma = (C^x_\gamma \times C^y_\gamma) \cap E_o \subseteq (C^x_\gamma \times C^y_\gamma) \cap E;$$

the last inclusion can be strict as $C^x_\gamma \times C^y_\gamma$ can contain points from the diagonal $E_d$. Conversely, some points from the diagonal may not pertain to $C^x_\gamma \times C^y_\gamma$ for any $\gamma \in \Gamma$; these points form the set

$$N := E \setminus \bigcup_{\gamma \in \Gamma} (C^x_\gamma \times C^y_\gamma) \subseteq E_d.$$

We say that *uniqueness of portfolio positions* holds at $(x, y) \in E$ if for any semistatic strategy $v : E \to \mathbb{R}$, the stock and option positions $h(x)$ and $g(y)$ are uniquely determined at $(x, y)$. In that situation, *convergence of portfolio positions* holds at $(x, y) \in E$ if for any semistatic strategies $v_n$ converging pointwise, the positions $h_n(x)$ and $g_n(y)$ are also convergent.

**Lemma 2.2.5.** Let $C \subseteq E_o$ be connected. If uniqueness (convergence) of portfolio positions holds at some point of $C$, then uniqueness (convergence) of portfolio positions holds at all points of $C$.

**Proof.** Let uniqueness of portfolio positions hold at $(x, y)$ and let $(x, y) \sim (x', y')$. Consider a path as in (2.1), then as discussed at the beginning of Section 2.2, uniqueness of the option position at $y$
implies uniqueness of the stock position at \( x_2 \) which in turn implies uniqueness of the option position at \( y_2 \), and so on, leading to uniqueness at \((x', y')\). Similarly for the convergence.

\[ \square \]

**Definition 2.2.6.** A path \((x_i, y_i)_{i=1}^k\) from \((x, y) \in E_o\) to itself is called a cycle. The cycle is identifying if

\[ \prod_{i=1}^k (y_i - x_i) - \prod_{i=1}^k (y_i - x_{i+1}) \neq 0, \tag{2.2} \]

where we use the cyclical convention \( x_{k+1} := x_1 \).

The terminology is explained by the subsequent lemma: such a cycle uniquely “identifies” the portfolio positions along its points. We note that \( k \geq 2 \) must hold for any identifying cycle. If \(((x_1, y_1), (x_2, y_2))\) is a cycle, then it is identifying if and only if \( x_1 \neq x_2 \) and \( y_1 \neq y_2 \), because (2.2) reduces to \((x_1 - x_2)(y_1 - y_2) \neq 0\). In particular, the cycle can be envisioned as a nondegenerate rectangle \( \{x_1, x_2\} \times \{y_1, y_2\} \subseteq E_o\). This simple characterization does not extend to larger cycles.

**Lemma 2.2.7.** Let \((x_i, y_i)_{i=1}^k\) be an identifying cycle. Then uniqueness and convergence of portfolio positions hold at \((x_1, y_1)\).

**Proof.** Consider a semistatic strategy \( v(x, y) = h(x)(y - x) + g(y) \) along the points (2.1),

\[ v(x, y) = h(x)(y - x) + g(y) \quad \text{for} \quad (x, y) \in \{(x_i, y_i), (x_{i+1}, y_i) : 1 \leq i \leq k\}. \]

This can be cast as the \(2k \times 2k\) linear system

\[
\begin{bmatrix}
y_1 - x_1 & 1 \\
1 & y_1 - x_2 \\
y_2 - x_2 & 1 \\
1 & y_2 - x_3 \\
\vdots & \ddots \\
y_k - x_1 & 1 \\
1 & y_k - x_k
\end{bmatrix}
\begin{bmatrix}
h(x_1) \\
g(y_1) \\
h(x_2) \\
g(y_2) \\
\vdots \\
h(x_k) \\
g(y_k)
\end{bmatrix}
= 
\begin{bmatrix}
v(x_1, y_1) \\
v(x_2, y_1) \\
v(x_2, y_2) \\
v(x_3, y_2) \\
\vdots \\
v(x_k, y_k) \\
v(x_1, y_k)
\end{bmatrix}
\]

13
where omitted matrix entries are zero. Using Laplace expansion along the first column and the convention $x_{k+1} := x_1$, we see that the determinant of the matrix is

$$
\begin{align*}
(y_1 - x_1) [(y_2 - x_2) \cdots (y_k - x_k)] \\
+ (-1)^{2k+1} (y_k - x_1) [(y_1 - x_2) \cdots (y_{k-1} - x_k)] \\
= \prod_{i=1}^{k} (y_i - x_i) - \prod_{i=1}^{k} (y_k - x_{i+1}).
\end{align*}
$$

As the cycle is identifying, it follows that the matrix is invertible, and the inverse map is continuous as a finite-dimensional linear map. In summary, the numbers $(h(x_i), g(y_i))_{1 \leq i \leq k}$ are uniquely determined by a continuous function of the numbers $(v(x_i, y_i), v(x_{i+1}, y_i))_{1 \leq i \leq k}$, showing the result.

Combining Lemma 2.2.5 and Lemma 2.2.7, we have the following.

**Corollary 2.2.8.** Let $C \subseteq E_o$ be connected. If $C$ contains an identifying cycle, then uniqueness and convergence of portfolio positions hold on $C$.

Next, we study what happens in the absence of identifying cycles. Given a subset $S \subseteq \mathbb{R}^2$, we denote by $S^x$ and $S^y$ its projections onto the first and second coordinate, respectively.

**Proposition 2.2.9.** Let $C$ be a connected component of $E_o$ containing no identifying cycles.

(a) Uniqueness of portfolio positions fails at each point of $C$. For any semistatic strategy, the set of portfolio positions is a one-parameter family.

(b) Closedness of semistatic strategies holds on $C$. More precisely, let $v_n(x, y) = h_n(x)(y - x) + g_n(y)$ be semistatic strategies converging pointwise on $C$ to $v : C \rightarrow \mathbb{R}$. Then there exist $h'_n, g'_n$ such that

$$
v_n(x, y) = h'_n(x)(y - x) + g'_n(y), \quad (x, y) \in C
$$
and their pointwise limits exist,

\[ h' := \lim h'_n \text{ on } C^x, \quad \quad g' := \lim g'_n \text{ on } C^y. \]

In particular, \( v(x, y) = h'(x)(y - x) + g'(y) \) on \( C \).

The positions \( h'_n, g'_n \) can be constructed as follows. Fix an arbitrary point \( (x_0, y_0) \in C \). Then there exist unique functions \( a, b : \mathbb{R} \to \mathbb{R} \setminus \{0\} \) such that \( a(x)b(y) = y - x \) on \( C \) and \( a(x_0) = 1 \). We can choose

\[ h'_n(x) := h_n(x) - \frac{h_n(x_0)}{a(x)}, \quad \quad g'_n(y) := g_n(y) + h_n(x_0)b(y). \]

Before stating the proof, we recall the Borwein–Lewis characterization for the decomposability of a function of two variables into a product of single-variable functions. More generally, this result applies to group-valued functions on arbitrary sets; see [11, Theorem 3.3].

**Lemma 2.2.10.** Let \( S \subseteq \mathbb{R} \times \mathbb{R} \) and \( c : S \to \mathbb{R} \setminus \{0\} \). The following are equivalent:

1. There exist \( a, b : \mathbb{R} \to \mathbb{R} \setminus \{0\} \) such that

\[ c(x, y) = a(x)b(y), \quad (x, y) \in S. \]

2. For any cycle \( (x_i, y_i)_{i=1}^k \) in \( S \),

\[ \prod_{j=1}^k c(x_j, y_j) = \prod_{j=1}^k c(x_{j+1}, y_j), \quad (2.3) \]

where \( x_{k+1} := x_1 \).

In that case, on each connected component of \( S \), the functions \( a \) and \( b \) are unique up on a scalar multiple.
Proof of Proposition 2.2.9. The absence of identifying cycles means that (2.3) holds for the function $c(x, y) := y - x$ on $C$. Note that $c$ is valued in $\mathbb{R} \setminus \{0\}$ due to $C \subseteq E_0$. Hence Lemma 2.2.10 implies that there exist $a, b : \mathbb{R} \rightarrow \mathbb{R} \setminus \{0\}$ such that $a(x)b(y) = y - x$ on $C$, and these functions are uniquely determined by the normalization that $a(x_0) = 1$ for some fixed $x_0 \in C^\times$.

Consider a semistatic strategy $v(x, y) = h(x)(y - x) + g(y)$ on $C$. Given $\alpha \in \mathbb{R}$, let $h_\alpha(x) := h(x) + \alpha/a(x)$ and $g_\alpha(y) := g(y) - \alpha b(y)$. Then

$$h_\alpha(x)(y - x) + g_\alpha(y) = [h(x) + \alpha/a(x)]a(x)b(y) + g(y) - \alpha b(y) = h(x)a(x)b(y) + g(y) = v(x, y),$$

showing that the portfolio positions inducing $v$ include the one-parameter family $(h_\alpha, g_\alpha)_{\alpha \in \mathbb{R}}$. Conversely, by connectedness, we know that portfolio positions are uniquely determined as soon as the option position is determined at one point $y_0$. Because $\alpha \mapsto g_\alpha(y_0) = g(y_0) - \alpha b(y_0)$ is surjective onto $\mathbb{R}$, this shows that $(h_\alpha, g_\alpha)_{\alpha \in \mathbb{R}}$ exhausts all portfolio positions inducing $v$.

Turning to the convergence, note that

$$\bar{v}_n(x, y) := \frac{v_n(x, y)}{b(y)} = \frac{h_n(x)a(x) + g_n(y)}{b(y)},$$

on $C$. Define

$$\bar{h}_n(x) := h_n(x)a(x) - h_n(x_0)a(x_0), \quad \bar{g}_n(y) := \frac{g_n(y)}{b(y)} + h_n(x_0)a(x_0),$$

so that

$$\bar{v}_n(x, y) = \bar{h}_n(x) + \bar{g}_n(y). \quad (2.4)$$

Clearly $\bar{v}_n(x, y)$ is convergent for all $(x, y) \in C$. Moreover, $\bar{h}_n(x_0) = 0$ for all $n$, so that $\bar{h}_n(x_0)$ is convergent. As $C$ is connected, the additive decomposition (2.4) implies as in the proof of Lemma 2.2.5 (or [11]) that the separate limits $\bar{g}(y) := \lim_n \bar{g}_n(y)$ and $\bar{h}(x) := \lim_n \bar{h}_n(x)$ exist for
all \((x, y) \in C\). It follows that

\[
h'_n(x) = \frac{\bar{h}_n(x)}{a(x)} \rightarrow \frac{\bar{h}(x)}{a(x)} = h'(x),
\]
\[
g'_n(y) = \frac{\bar{g}_n(y)}{b(y)} \rightarrow \frac{\bar{g}(y)}{b(y)} = g'(y),
\]

completing the proof.

We can now prove the main result.

**Proof of Theorem 2.2.2.** Let \(v_n(x, y) = h_n(x)(y - x) + g_n(y)\) be semistatic strategies converging pointwise on \(E\) to \(v : E \rightarrow \mathbb{R}\). We shall construct \(h : E^x \rightarrow \mathbb{R}\) and \(g : E^y \rightarrow \mathbb{R}\) such that \(v(x, y) = h(x)(y - x) + g(y)\) on \(E\).

Recall the partition \((C_{\gamma})_{\gamma \in \Gamma}\) of \(E_o\) into connected components and Remark 2.2.4. Let \((D_j)_{j \in J}\) be the collection consisting of all singletons \(\{(x, y)\}\) with \((x, y) \in N\) as well as the sets \((C^x_{\gamma} \times C^y_{\gamma}) \cap E, \gamma \in \Gamma\). Note that two points in \(E_o\) sharing one coordinate are necessarily connected, and two points in \(E_d\) sharing one coordinate must coincide. This implies that \((D_j)_{j \in J}\) is a partition of \(E\) with the following property: If \((x, y), (x, y') \in E\), then \((x, y)\) and \((x, y')\) belong to the same component \(D_j\). Similarly, if \((x, y), (x', y) \in E\), then \((x, y)\) and \((x', y)\) belong to the same component \(D_j\). As a consequence, we may construct the positions \(h, g\) separately on each \(D_j\) without danger of creating any inconsistencies. (This is not true for \(\{C_{\gamma}, E_d\}\), whence the need for yet another collection.)

1. Let \(D_j = (C^x_{\gamma} \times C^y_{\gamma}) \cap E\) for some \(\gamma \in \Gamma\), where \(C_{\gamma}\) contains an identifying cycle. Then we can choose \((h, g) := \lim(h_n, g_n)\) on \((C^x_{\gamma}, C^y_{\gamma})\) according to Corollary 2.2.8.

2. Let \(D_j = (C^x_{\gamma} \times C^y_{\gamma}) \cap E\) for some \(\gamma \in \Gamma\), where \(C_{\gamma}\) does not contain an identifying cycle.

Then we can choose \((h, g)\) on \((C^x_{\gamma}, C^y_{\gamma})\) according to Proposition 2.2.9.

3. Let \(D_j = \{(x, y)\}\) for some \((x, y) \in N \subseteq E_d\). Then we can define \(h(x) = 0\) and \(g(y) = v(x, y)\). In fact, as \(y - x = 0\), any choice for \(h(x)\) will do.
The preceding analysis also quantifies the non-uniqueness for portfolio positions; that is, the exact number of degrees of freedom in choosing the portfolio positions for any given semistatic strategy.

**Remark 2.2.11.** Recalling Proposition 2.2.9(a) and the proof above, for any semistatic strategy \( v \) on \( E \), the set of all portfolio positions \( (h, g) \) with \( v(x, y) = h(x)(y - x) + g(y) \) for all \((x, y) \in E\) is a \( k \)-parameter family, where

\[
k = \text{card} \left( N \cup \{ \gamma \in \Gamma : C_{\gamma} \text{ contains no identifying cycle}\} \right).
\]

A sufficient condition for measurability of portfolio positions \((h, g)\) is thus \( k \leq \text{card}(\mathbb{N}) \). Establishing a necessary condition is not obvious and left for future research.

**Remark 2.2.12.** Our arguments in Proposition 2.2.9 and Theorem 2.2.2 remain true if the sequence of strategies \((v_n)_n\) is replaced by a net \((v_a)_a\).

### 2.3 General Probabilistic Model

Given integers \( d \geq 1 \) and \( T \geq 2 \), we denote by \( X = (X_t)_{t=1}^T \) the canonical process on \((\mathbb{R}^d)^T\), where \( X_t = (X_{t,j})_{j=1}^d \) are interpreted as the prices of \( d \) stocks at date \( t \). We also fix a probability measure \( P \) on \((\mathbb{R}^d)^T\); only the nullsets of \( P \) will matter for our results. For the purposes of this section, a semistatic strategy is a random variable \( V \) satisfying

\[
V = \sum_{t=1}^{T-1} \sum_{j=1}^d \hat{h}_{t,j}(X_1, \ldots, X_t)(X_{t+1,j} - X_{t,j}) + \sum_{j=1}^d \hat{g}_j(X_{T,j})
\]  

(2.5)

\( P \)-a.s. for some real-valued measurable functions \( \hat{h}_{t,j} \) and \( \hat{g}_j \). It will be notationally convenient to work instead with the random variables

\[
h_{t,j} := \hat{h}_{t,j}(X_1, \ldots, X_t), \quad g_j := \hat{g}_j(X_{T,j}).
\]
We call \( h = (h_{t,j}) \) the stock position and \( g = (g_j) \) the option position, respectively. Together, they form the portfolio position \((h, g)\) of \( V \). The portfolio position is not uniquely determined by \( V \) in general: it is clearly possible to add a constant to \( g_1 \) and subtract the same from \( g_j \) for any \( j \neq 1 \), without affecting \( V \). These \( d - 1 \) degrees of freedom are easily removed by fixing an “anchor” point \( x^0 \in (\mathbb{R}^d)^T \) and normalizing

\[
\hat{g}_j(x^0) = 0, \quad j = 2, \ldots, d. \tag{2.6}
\]

In the context of Theorem 2.3.1 below, it will be shown that the anchor point can be chosen arbitrarily outside a certain nullset.

In this probabilistic setting, we say that uniqueness of portfolio positions holds if for any semistatic strategy \( V \), after a normalization of the form (2.6), the portfolio position \((h, g)\) is uniquely determined \( P \)-a.s. Convergence of portfolio positions holds if, after a normalization of the form (2.6), for any semistatic strategies \( V^{(n)} \) converging \( P \)-a.s., the corresponding positions \((h^{(n)}, g^{(n)})\) also converge \( P \)-a.s.

The aim of this section is to exhibit a probabilistic condition circumventing some of the complications highlighted in Section 2.2. To that end, let \( \mu_{t,j} \) be the law of \( X_{t,j} \), or equivalently, the one-dimensional marginal law of \( P \) on the component \((t, j)\). We assume throughout that \( \mu_{t,j} \) is not a Dirac measure for any \( t, j \). (This serves to simplify the exposition; while the results and arguments could be generalized, the degenerate case is not relevant financially and hence omitted.) The key condition for our result is that \( P \) be measure-theoretically equivalent to the product of its marginals.

**Theorem 2.3.1.** Suppose that \( P \sim \bigotimes_{t=1}^T \bigotimes_{j=1}^d \mu_{t,j} \). Then the set of semistatic strategies is closed under \( P \)-a.s. convergence. Moreover, uniqueness and convergence of portfolio positions hold.

As mentioned in the Introduction, \( P \sim \bigotimes_{t=1}^T \bigotimes_{j=1}^d \mu_{t,j} \) intuitively means that none of the (a priori possible) future stock prices becomes completely impossible given an intermediate state of the prices, similarly as in the condition of conditional full support [28]. Technically, we shall see
that for such \( P \), any set of full measure contains an abundance of cuboids that will play the role of identifying cycles (cf. Section 2.2). Indeed, fix \( x^0, x^1 \in (\mathbb{R}^d)^T \) such that \( x^0_{t,j} \neq x^1_{t,j} \) for all \( t, j \) and consider the cuboid \( Q \) generated by their components,

\[
Q := \prod_{t=1}^T \prod_{j=1}^d \{x^0_{t,j}, x^1_{t,j}\} \subseteq (\mathbb{R}^d)^T. \tag{2.7}
\]

That is, each point in \( Q \) is a matrix \((x^0_{t,j})_{t,j}\) where \( \epsilon_{t,j} \in \{0, 1\} \).

**Proposition 2.3.2.** Recall \( Q \) defined in (2.7). Let \( V \) and \( V^{(n)} \) be semistatic strategies with portfolio positions \((h, g)\) and \((h^{(n)}, g^{(n)})\), respectively, such that (2.5) and (2.6) hold on \( Q \). Then \((h, g)\) are uniquely determined on \( Q \). Moreover, \( V^{(n)} \to V \) pointwise on \( Q \) implies that \((h^{(n)}, g^{(n)}) \to (h, g)\) pointwise on \( Q \).

The proof is lengthy and deferred to Section 2.4. In a nutshell, we view (2.5) as linear system where the values of \( h \) and \( g \) at points in \( Q \) are the variables; each equation of the system corresponds to evaluating \( V \) at a point in \( Q \). We prove that the (finite-dimensional) linear map associated with the system is injective, hence admits a continuous inverse. As a result, the portfolio position \((h, g)\) is a continuous function of the strategy \( V \).

The next lemma is a general measure-theoretic fact; it formalizes the claim that there is an abundance of cuboids in any set of full \( P \)-measure.

**Lemma 2.3.3.** Consider probability spaces \((\Omega_i, \mathcal{F}_i, \mu_i)_{i=1}^n\) and their product \((\Omega, \mathcal{F}, \mu)\) given by \( \Omega = \prod_{i=1}^n \Omega_i, \mathcal{F} = \otimes_{i=1}^n \mathcal{F}_i \) and \( \mu = \otimes_{i=1}^n \mu_i. \) If \( A \in \mathcal{F} \) satisfies \( \mu(A) = 1, \) then \( \mu \)-a.e. \( x^0 = (x^0_1, x^0_2, \ldots, x^0_n) \in A \) satisfies

\[
\mu \left\{ x \in A : \prod_{i=1}^n \{x^0_i, x_i\} \subseteq A \right\} = 1.
\]

**Proof.** Let \( \Omega^0, \Omega^1 \) be two copies of \( \Omega \) with components denoted \( \Omega_i^{0,1} \). Consider the product space \( \hat{\Omega} = \prod_{i=1}^n (\Omega_i^0 \times \Omega_i^1) \) endowed with the product \( \sigma \)-field \( \hat{\mathcal{F}} \) and the product measure \( \hat{\mu} = \otimes_{i=1}^n (\mu_i \otimes \mu_i) \). For each multi-index \( J = (j_1, j_2, \ldots, j_n) \in \{0, 1\}^n \), define the projection \( \pi_J : \hat{\Omega} \to \prod_{i=1}^n \Omega_i^{0,1} \).
by

\[(x_0^0, x_1^1, x_2^0, x_2^1, \ldots, x_n^0, x_n^1) \mapsto (x_1^0, x_2^1, \ldots, x_n^0).\]

Clearly the law of \(\pi_J\) under \(\hat{\mu}\) is \(\mu\), so that 
\(\hat{\mu}(\pi_J \in A) = \mu(A) = 1\). Defining \(S := \bigcap\{\pi_J \in A\}\) as the intersection over all \(J \in \{0, 1\}^n\), it follows that 
\(\hat{\mu}(S) = 1\). Denote by \(S_{x^0}\) the section of \(S\) at \(x^0 \in \Omega^0\). In view of \(\hat{\mu}(S) = 1\), Fubini’s theorem implies

\[\mu \{x^0 \in \Omega^0 : \mu(S_{x^0}) = 1\} = 1.\]

The desired result follows once we observe that 
\((x_1^0, x_1, x_2^0, x_2, \ldots, x_n^0, x_n) \in S\) if and only if

\[\prod_{i=1}^n \{x_i^0, x_i\} \subseteq A.\]

We can now deduce the main result.

**Proof of Theorem 2.3.1.** Suppose that \(V^{(n)}\) is of the form (2.5) with portfolio position \((h^{(n)}, g^{(n)})\) and that \(V^{(n)} \to V\) on \(A \subseteq (\mathbb{R}^d)^T\) with \(P(A) = 1\). In view of \(P \sim \bigotimes_{t=1}^T \bigotimes_{j=1}^d \mu_{t,j}\), Lemma 2.3.3 implies that there exists \(x^0 \in A\) such that

\[B := \left\{ x \in A : \prod_{t=1}^T \prod_{j=1}^d \{x_{t,j}^0, x_{t,j}\} \subseteq A \right\}
\]

satisfies \(P(B) = 1\). (In fact, \(P\)-almost any \(x^0 \in A\) will do.) We use this point \(x^0\) for the normalization (2.6).

We claim that the limit \((h(x), g(x)) := \lim_n (h^{(n)}(x), g^{(n)}(x))\) exists for all \(x \in B\). To prove this, we first consider \(x^1 \in B\) that satisfies \(x_{t,j}^1 \neq x_{t,j}^0\) for all \(t, j\) and construct the cuboid \(Q\) determined by \(x^0\) and \(x^1\); cf. (2.7). Then we see from Proposition 2.3.2 that \((h^{(n)}, g^{(n)})\) converges to some \((h, g)\) on \(Q\), and that \((h, g)\) is uniquely determined on \(Q\). In particular, the limit exists at \(x := x^0\) and at \(x := x^1\). Next, for an arbitrary \(x \in B\), as \(\mu_{t,j}\) is not a Dirac measure, we can find \(x^1 \in B\) such that \(x_{t,j}^1 = x_{t,j}\) if \(x_{t,j} \neq x_{t,j}^0\) and \(x_{t,j}^1 \neq x_{t,j}^0\) for all \(t, j\). Applying the same reasoning as above to \(x^0\) and \(x^1\), we see that the limit exists at \(x\). The same argument also establishes the uniqueness and convergence of portfolio positions.

\[\square\]
2.4 Proof of Proposition 2.3.2

Recall Q defined in (2.7). Suppose that \( V : Q \rightarrow \mathbb{R} \) is of the form (2.5) with portfolio position \((h, g)\), where \( g \) satisfies (2.6) at \( x^0 \). This induces a linear system with the values of \( h \) and \( g \) at the points in \( Q \) as variables and the price increments \( \Delta_{t+1,j} = X_{t+1,j} - X_{t,j} \) as coefficients. To start with the simplest example, consider \( T = 2 \) and \( d = 1 \), so that \( V = h_{1,1}(X_{2,1} - X_{1,1}) + g_1 \) and \( Q = \{(x_{1,1}^0, x_{2,1}^0), (x_{1,1}^0, x_{2,1}^1), (x_{1,1}^1, x_{2,1}^0), (x_{1,1}^1, x_{2,1}^1)\} \). This corresponds to 4 equations and can be cast as a \( 4 \times 4 \) linear system

\[
\begin{bmatrix}
x_{2,1}^0 - x_{1,1}^0 & 1 & \hat{h}_{1,1}(x_{1,1}^0) & V(x_{1,1}^0, x_{2,1}^0) \\
x_{2,1}^1 - x_{1,1}^0 & 1 & \hat{h}_{1,1}(x_{1,1}^1) & V(x_{1,1}^0, x_{2,1}^1) \\
x_{2,1}^1 - x_{1,1}^1 & 1 & \hat{g}_1(x_{2,1}^0) & V(x_{1,1}^1, x_{2,1}^0) \\
x_{2,1}^1 - x_{1,1}^1 & 1 & \hat{g}_1(x_{2,1}^1) & V(x_{1,1}^1, x_{2,1}^1)
\end{bmatrix}
\]

(2.8)

In this example, the condition (2.6) is vacuous as \( d = 1 \).

In the general case, (2.5) on \( Q \) can be viewed as a \( N_r \times N_c \) linear system that we describe next. For \( 1 \leq t \leq T \), consider the binary vector \( \varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, \ldots, \varepsilon_{t,d}) \in \{0,1\}^d \). We view

\[\varepsilon_t := (\varepsilon_s)_{s=1}^t := (\varepsilon_{1,1}, \ldots, \varepsilon_{1,d}, \varepsilon_{2,1}, \ldots, \varepsilon_{2,d}, \ldots, \varepsilon_{t,1}, \ldots, \varepsilon_{t,d}) \in \{0,1\}^{dt}\]

as a multi-index of length \( dt \) and denote the collection of all such \( \varepsilon_t \) by \( \mathcal{E}_t \). There is a one-to-one correspondence between \( \mathcal{E}_T \) and \( Q \) via \( \varepsilon_T \mapsto x^{\varepsilon_T} := (x_{t,j}^{\varepsilon_{t,j}})_{t,j} \). More generally, for \( 1 \leq t \leq T \), the set \( \mathcal{E}_t \) corresponds to the set \( \prod_{s=1}^t \prod_{j=1}^d \{x_{s,j}^0, x_{s,j}^1\} \) via \( \varepsilon_t \mapsto x^{\varepsilon_t} := (x_{s,j}^{\varepsilon_{s,j}})_{s,j} \), where \( 1 \leq s \leq t \) and \( 1 \leq j \leq d \).

Every point \( x^{\varepsilon_T} \in Q \) gives rise to an equation as we evaluate \( V \) at \( x^{\varepsilon_T} \). Hence the number of rows in our system is \( N_r = 2^{dT} \), the cardinality of \( \mathcal{E}_T \). On the other hand, for each \( 1 \leq t \leq T - 1 \) and each \( 1 \leq j \leq d \), \( \hat{h}_{t,j} \) gives rise to \( 2^{dt} \) variables, namely \( \hat{h}_{t,j}(x^{\varepsilon_t}) \) for \( \varepsilon_t \in \mathcal{E}_t \). In addition, recalling (2.6), \( \hat{g} \) gives rise to \( d + 1 \) variables, namely \( \hat{g}_1(x_{T,1}^{\varepsilon_{T,1}}) \) and \( \hat{g}_j(x_{T,j}^{\varepsilon_{T,j}}) \) for \( 1 \leq j \leq d \). As a
result, the total number of variables (hence columns in the system) is

\[ N_c = d \sum_{t=1}^{T-1} 2^{dt} + (d + 1) = d \frac{2^{dT} - 1}{2^d - 1} + 1. \]

As in (2.8), the coefficients of the matrix are given by stock price increments \((x_{t+1,j}^\varepsilon - x_{t,j}^\varepsilon)_{t,j}\) and binary entries related to the options. To unambiguously determine the matrix, we need to specify an order for the rows and columns; in fact, we tailor our order to facilitate the exposition below. For \(1 \leq t \leq T\), consider the natural lexicographic order on \(E_t\) and equip \((x_{t}^\varepsilon)_{\varepsilon \in E_t}\) with the induced order; that is, \(x_{t}^\varepsilon \leq x_{t}^{\tilde{\eta}}\) if and only if \(\varepsilon \leq \tilde{\eta}\). In particular, the order on \(E_T\) induces an order on \(Q = (x_{T}^\varepsilon)_{\varepsilon \in E_T}\). The row ordering is set by evaluating (2.5) on \(Q\) in that order. The column ordering is given by the following (top-to-bottom) hierarchy:

1. The variables from \(\hat{h}\) appear before those from \(\hat{g}\).

2. The \(d \sum_{t=1}^{T-1} 2^{dt}\) variables \(\hat{h}_{t,j}(x_{t}^\varepsilon)\) are sorted
   
   (a) in **descending** order of \(t \in \{T - 1, T - 2, \ldots, 1\}\),
   
   (b) then in ascending order of \(\varepsilon_t \in E_t\), and
   
   (c) lastly in ascending order of \(j \in \{1, 2, \ldots, d\}\).

3. The \(d + 1\) variables from \(\hat{g}\) are ordered as

   \[ \hat{g}_1(x_{T,1}^0), \hat{g}_1(x_{T,1}^1), \hat{g}_2(x_{T,2}^1), \hat{g}_3(x_{T,3}^1), \ldots, \hat{g}_d(x_{T,d}^1). \]

The reader can verify that the matrix in (2.8) follows the desired ordering; a more advanced example can be found in (2.13) below. For the general case \(T \geq 2\) and \(d \geq 1\), the above convention uniquely determines a matrix \(L_T\) which will be shown to have the following property.

**Lemma 2.4.1.** The matrix \(L_T\) has full column rank.
**Remark 2.4.2.** In the case $d = 1$ of a single stock, $L_T$ is a square matrix and
\[
\det L_T = (-1)^{T-1} \prod_{t=1}^{T-1} \left( x_{t,1}^0 - x_{t,1}^1 \right)^{2^{t-1}} \left( x_{T,1}^0 - x_{T,1}^1 \right)^{2^{T-1}-1} \neq 0.
\]

The proof uses arguments similar to the proof of Lemma 2.4.1 below.

Once the lemma is established, Proposition 2.3.2 is a direct consequence:

**Proof of Proposition 2.3.2.** By Lemma 2.4.1, the linear map associated with $L_T$ is injective. Its inverse is continuous as a linear finite-dimensional map, showing that the portfolio position is a continuous function of the strategy.

\[\square\]

**Proof of Lemma 2.4.1**

Next, we introduce some additional notation for the proof of Lemma 2.4.1. Given $1 \leq t \leq T-1$ and $\vec{\varepsilon}_t \in \mathcal{E}_t$, we define $Q^{\vec{\varepsilon}_t} \subseteq Q$ by
\[
Q^{\vec{\varepsilon}_t} := \{ x^{\vec{\varepsilon}_t} \} \times \left( \prod_{s=t+1}^{T} \prod_{j=1}^{d} \{ x^0_{s,j}, x^1_{s,j} \} \right).
\]

That is, $Q^{\vec{\varepsilon}_t}$ consists of the $2^{d(T-t)}$ points in $Q$ that share the vector $x^{\vec{\varepsilon}_t}$ in their first $dt$ coordinates.

Note that for each $1 \leq t \leq T-1$, $\{ Q^{\vec{\varepsilon}_t} : \vec{\varepsilon}_t \in \mathcal{E}_t \}$ forms a partition of $Q$.

For each $1 \leq t \leq T-1$, $1 \leq j \leq d$ and $(k, l) \in \{0, 1\}^2$, we introduce a shorthand for the corresponding stock price increment
\[
\Delta_{t+1,j}^{k,l} := x_{t+1,j}^k - x_{t,j}^l,
\]

as well as the stock price difference
\[
\delta_{t,j} := x_{t,j}^0 - x_{t,j}^1
\]
which is nonzero by assumption. We record two identities for later use,

\[
\Delta_{t+1,j}^0 - \Delta_{t+1,j}^1 = \delta_{t+1,j} \quad \text{and} \quad \Delta_{t+1,j}^0 - \Delta_{t+1,j}^1 = -\delta_{t,j},
\]

(2.12)

where the right-hand sides do not depend on \( l \in \{0, 1\} \).

As the proof of Lemma 2.4.1 is somewhat involved, we first state an example to illustrate some of the arguments.

\[\text{Example 2.4.3 (} T = 2 \text{ and } d = 2 \text{). Note that (2.5) reads}\]

\[
V = h_{1,1} (X_{2,1} - X_{1,1}) + h_{1,2} (X_{2,2} - X_{1,2}) + g_1 + g_2.
\]

The matrix \( L_2 \) has 16 rows and 11 columns; see Figure 2.1(a). Since \( \mathcal{E}_1 = \{(0, 0), (0, 1), (1, 0), (1, 1)\} \),
the column ordering corresponds to the vector of 11 variables

\[
\left[ \hat{h}_{1,1}(x^{\vec{\varepsilon}_1}), \hat{h}_{1,2}(x^{\vec{\varepsilon}_1}) \right. \text{ for } \vec{\varepsilon}_1 \in \mathcal{E}_1, \left. \hat{g}_1(x^0_{2,1}), \hat{g}_1(x^1_{2,1}), \hat{g}_2(x^1_{2,2}) \right].
\] (2.13)

To show that \(L_2\) has full column rank, we proceed in four steps.

**Step 1.** Our goal is to divide \(L_2\) into 4 submatrices \((L_{\vec{\varepsilon}_1}^\varepsilon)\), each of which has 4 rows and 11 columns. We will identify a linearly dependent row in each \(L_{\vec{\varepsilon}_1}^\varepsilon\).

Consider (2.9) with \(T = 2\). For each \(\vec{\varepsilon}_1 \in \mathcal{E}_1\), evaluating \(V\) on \(Q_{\vec{\varepsilon}_1}^\varepsilon\) gives rise to a \(4 \times 11\) submatrix \(L_{\vec{\varepsilon}_1}^\varepsilon\) of \(L_2\). It has 4 rows as \(|Q_{\vec{\varepsilon}_1}^\varepsilon| = 4\); there are in total 4 submatrices as \(|\mathcal{E}_1| = 4\).

The coefficients across these submatrices follow a similar pattern. For concreteness, consider \(\vec{\varepsilon}_1 = (0, 0) \in \mathcal{E}_1\). Recalling the definition of price increments from (2.10), evaluating \(V\) on \(Q_{\vec{\varepsilon}_1}^\varepsilon\) gives

\[
\begin{bmatrix}
\Delta^0_{2,1} & \Delta^0_{2,2} & \cdots & 1 \\
\Delta^0_{2,1} & \Delta^1_{2,2} & \cdots & 1 \\
\Delta^1_{2,1} & \Delta^0_{2,2} & \cdots & 1 \\
\Delta^1_{2,1} & \Delta^1_{2,2} & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
\hat{h}_{1,1}(x^{(0,0)}) \\
\hat{h}_{1,2}(x^{(0,0)}) \\
\hat{g}_1(x^0_{2,1}) \\
\hat{g}_1(x^1_{2,1})
\end{bmatrix}
= \begin{bmatrix}
v(x^{(0,0)}, x^0_{2,1}, x^0_{2,2}) \\
v(x^{(0,0)}, x^0_{2,1}, x^1_{2,2}) \\
v(x^{(0,0)}, x^1_{2,1}, x^0_{2,2}) \\
v(x^{(0,0)}, x^1_{2,1}, x^1_{2,2})
\end{bmatrix},
\] (2.14)

where the omitted entries (including six entire columns) are zero. Denote the matrix in (2.14) by \(L_{\vec{\varepsilon}_1}^\varepsilon\) and let \(R_i\) stand for its Row \(i\). Note that \(R_1 - R_2 = R_3 - R_4\). Since \(R_4\) is a linear combination of the other rows, dropping it from \(L_{\vec{\varepsilon}_1}^\varepsilon\) does not alter the matrix rank. This argument applies to all \(\vec{\varepsilon}_1 \in \mathcal{E}_1\). Denote the remaining matrix by \((L_{\vec{\varepsilon}_1}^\varepsilon)'\). We form the matrix \(L'_2\) by vertically stacking \((L_{\vec{\varepsilon}_1}^\varepsilon)'\) for all \(\vec{\varepsilon}_1 \in \mathcal{E}_1\); see Figure 2.1(b). It follows that \(L'_2\) has the same rank as \(L_2\).

**Step 2.** Recall that elementary row and column operations preserve the matrix rank. In this step,
we aim to bring $L'_2$ to a block lower triangular matrix

$$
\begin{bmatrix}
L''_2 \\
B & C
\end{bmatrix}
$$

where $C = \begin{bmatrix} 1 & 1 \\ 1 & \end{bmatrix}$ \hspace{1cm} (2.15)

by suitable row operations. Indeed, let $\mathbf{1} = (1, 1) \in \mathcal{E}_1$ and $\vec{v}_1 \neq \mathbf{1} \in \mathcal{E}_1$. After subtracting $(L'_2)_{\vec{v}_1}$ from $(L'_2)^\top_{\vec{v}_1}$, we denote the resultant matrix by $(L'_{\vec{v}_1})''$. By applying this procedure to all $\vec{v}_1 \neq \mathbf{1}$, we bring $L'_2$ into the desired form, where

$$
L''_2 = \begin{bmatrix}
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{0,1} \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{1,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{1,1} \\
\Delta_{2,1}^{1,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{1,1} & -\Delta_{2,2}^{0,1} \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{1,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{1,1} \\
\Delta_{2,1}^{1,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{1,1} & -\Delta_{2,2}^{0,1} \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{1,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{1,1} \\
\Delta_{2,1}^{1,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{1,1} & -\Delta_{2,2}^{0,1} \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{1,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{1,1} \\
\Delta_{2,1}^{1,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{1,1} & -\Delta_{2,2}^{0,1}
\end{bmatrix}
$$

Clearly $C$ has full rank, and it follows that $L_2$ has full column rank if and only if $L''_2$ does.

**Step 3.** Recall the definition of stock price differences (2.11). Let $\vec{v}_1 \neq \mathbf{1}$. In $(L'_2)^\top_{\vec{v}_1}$, subtracting its Row 1 from all other rows leaves each of them with precisely two nonzero entries of the same
magnitude, $\delta_{2,j}$, with the opposite signs; cf. (2.12). Applying this procedure to all $\bar{\varepsilon}_1 \neq 1$ yields

$$
\begin{bmatrix}
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{0,1} & -\delta_{2,2} \\
-\delta_{2,1} & \delta_{2,1} & \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{0,1} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{0,1} & -\delta_{2,2} \\
-\delta_{2,1} & \delta_{2,1} & \\
\Delta_{2,1}^{0,1} & \Delta_{2,2}^{0,0} & -\Delta_{2,1}^{0,1} & -\Delta_{2,2}^{0,1} & -\delta_{2,2} \\
-\delta_{2,1} & \delta_{2,1} &
\end{bmatrix}
$$

(2.17)

Recall that the odd and even numbered columns correspond to the variables $\hat{h}_{1,1}(x^{\bar{\varepsilon}_1})$ and $\hat{h}_{1,2}(x^{\bar{\varepsilon}_1})$ for $\bar{\varepsilon}_1 \in \mathcal{E}_1$, respectively. In particular, Column 7 corresponds to $\hat{h}_{1,1}(x^{\bar{\varepsilon}_1})$ and Column 8 corresponds to $\hat{h}_{1,2}(x^{\bar{\varepsilon}_1})$. We add Columns 1, 3, 5 to Column 7, and Columns 2, 4, 6 to Column 8:

$$
\begin{bmatrix}
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{0,0} & -\delta_{1,1} & -\delta_{1,2} \\
-\delta_{2,2} & \\
\Delta_{2,1}^{0,0} & \Delta_{2,2}^{0,1} & -\delta_{1,1} & \\
-\delta_{2,2} & \\
\Delta_{2,1}^{0,1} & \Delta_{2,2}^{0,0} & -\delta_{1,2} & \\
-\delta_{2,2} & \\
\end{bmatrix}
$$

(2.18)

All rows except Rows 1, 4, 7 have exactly one nonzero entry, and none of these entries share a common column. (We remark that Rows 1, 4, 7 correspond to the multi-indices $(\bar{\varepsilon}_1, 0, 0)$ for $\bar{\varepsilon}_1 \neq 1$
in \( \mathcal{E}_1 \).) Dropping these rows and columns, we denote the remaining \( 3 \times 2 \) submatrix by \( M_2 \):

\[
M_2 = \begin{bmatrix}
-\delta_{1,1} & -\delta_{1,2} \\
-\delta_{1,1} & \\
-\delta_{1,2}
\end{bmatrix}.
\]

(2.19)

It follows that \( L_2 \) has full column rank if and only if \( M_2 \) does.

**Step 4.** By inspection, \( M_2 \) indeed has full column rank. The example is complete.

**Remark 2.4.4.** The example can be generalized to \( T = 2 \) and \( d \geq 1 \) without much effort; cf. Steps 1 to 3 in the proof of Lemma 2.4.1 below. In the presence of \( d \) stocks, (2.15) holds with a \((d + 1) \times (d + 1)\) matrix \( C \) of binary coefficients

\[
C = \begin{bmatrix}
1 \\
1 & 1 \\
\vdots & \ddots \\
1 & 1 \\
1
\end{bmatrix}.
\]

(2.20)

Evidently, \( C \) has full rank. In this case, \( M_2 \) has \( (2^d - 1) \) rows and \( d \) columns; cf. (2.19). For \( 1 \leq j \leq d \), the coefficients in Column \( j \) alternate among

\[
( -\delta_{1,j}, \ldots, -\delta_{1,j}, 0, \ldots, 0 )_{\times 2^d-j}. \]

It is straightforward to verify that \( M_2 \) has full column rank; cf. (2.23) below. (In fact, the rows in \( M_2 \) are linear combinations of the rows in \( C'' \).)

**Proof of Lemma 2.4.1.** We proceed in four steps. In Step 1, we show that keeping only a selected number of rows in \( L_T \) does not decrease the matrix rank. In Steps 2 and 3, we use row and column operations to reduce the problem to showing that a certain matrix \( M_T \) has full column rank. In
Step 1. We will divide $L_T$ into $2^d(T-1)$ submatrices $L_{T}^{\vec{\varepsilon}_{T-1}}$ for $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$, each of which has $2^d$ rows and $N_c$ columns. Then, we will identify and remove linearly dependent rows in each submatrix. Thus, the remaining matrix, denoted by $L_T'$, will have the same rank as $L_T$.

Recall the definition (2.9). For each $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$, evaluating $V$ on $Q^{\vec{\varepsilon}_{T-1}}$ gives rise to a $2^d \times N_c$ submatrix $L_{T}^{\vec{\varepsilon}_{T-1}}$ of $L_T$. It has $2^d$ rows as $|Q^{\vec{\varepsilon}_{T-1}}| = 2^d$; there are in total $2^d(T-1)$ submatrices as $|\mathcal{E}_{T-1}| = 2^d(T-1)$.

Fix $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$. Consider two points in $Q^{\vec{\varepsilon}_{T-1}}$ that share all but one entry. Formally, they can be denoted as $x^{\vec{\eta}_T}$ and $x^{\vec{\rho}_T}$ for $\vec{\eta}_T, \vec{\rho}_T \in \mathcal{E}_T$ such that for some fixed $1 \leq k \leq d$, we have

$$\eta_{t,j} = \rho_{t,j} = \varepsilon_{t,j} \quad \text{for all} \quad 1 \leq t \leq T-1 \quad \text{and} \quad 1 \leq j \leq d, \quad \eta_{T,j} = \rho_{T,j} \quad \text{for all} \quad j \neq k, \quad \text{and} \quad \eta_{T,k} = 0 \quad \text{and} \quad \rho_{T,k} = 1.$$  

Then, it follows from (2.5) that

$$V(x^{\vec{\eta}_T}) - V(x^{\vec{\rho}_T}) = \hat{h}_{T-1,k}(x^{\vec{\varepsilon}_{T-1}})'(x_{T,k}^0 - x_{T,k}^1) + \hat{g}_k(x_{T,k}^0) - \hat{g}_k(x_{T,k}^1), \quad (2.21)$$

where the right-hand side does not depend on any $x_{T,j}^{\vec{\eta}_T} = x_{T,j}^{\vec{\rho}_T}$ for $j \neq k$ (i.e., the shared stock prices at date $T$).

Let 0 and $e_j$, $1 \leq j \leq d$, be the zero vector and the unit vectors in $\{0,1\}^d$. Note that $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$ is still fixed, and that $(\vec{\varepsilon}_{T-1},0)$ and $(\vec{\varepsilon}_{T-1},e_j)$, $1 \leq j \leq d$, are $d+1$ elements in $\mathcal{E}_T$. Let $(L_{T}^{\vec{\varepsilon}_{T-1}})'$ be the submatrix of $L_{T}^{\vec{\varepsilon}_{T-1}}$ whose rows are generated by these elements. By repeated applications of (2.21), it is not hard to see that all rows in $L_{T}^{\vec{\varepsilon}_{T-1}}$ are linear combinations of the rows in $(L_{T}^{\vec{\varepsilon}_{T-1}})'$. This argument applies to all $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$. We form the matrix $L_T'$ by vertically stacking $(L_{T}^{\vec{\varepsilon}_{T-1}})'$ for all $\vec{\varepsilon}_{T-1} \in \mathcal{E}_{T-1}$. It follows that $L_T'$ has the same rank as $L_T$.

As a remark for later use, we note that for each $1 \leq t \leq T-1$, coefficients of $L_T$ (and $L_T'$) in the columns corresponding to $\hat{h}_{t,j}(x^{\vec{\varepsilon}_{t}})$ for $1 \leq j \leq d$ are nonzero if and only if they belong to the rows that arise from evaluating $V$ on $Q^{\vec{\varepsilon}_{t}}$. In other words, for each $1 \leq t \leq T-1$, the $d \cdot 2^{d_t}$ columns of $L_T$ that correspond to $\hat{h}_{t,j}(x^{\vec{\varepsilon}_{t}})$ for $1 \leq j \leq d$ and $\vec{\varepsilon}_t \in \mathcal{E}_t$ form a block diagonal
matrix, where each block has $2^{d(T-t)}$ rows and $d$ columns.

**Step 2.** After some block-by-block row operations, we will bring $L_2'$ to a block lower triangular matrix

$$
\begin{bmatrix}
L''_T \\
B \\
C
\end{bmatrix}, \quad \text{with } C \text{ as in (2.20)}.
$$

For $1 \leq t \leq T - 1$, we denote $(1, \ldots, 1) \in \mathcal{E}_t$ by $1_t$. Fix $\bar{\varepsilon}_{T-1} \neq 1_{T-1}$ in $\mathcal{E}_{T-1}$. After subtracting $(L_1^T)^' \varepsilon_{T-1}$ from $(L_1^T)^' \varepsilon_{T-1}$, we denote the resultant matrix by $(L_1^T)^'' \varepsilon_{T-1}$. Applying this procedure to all $\bar{\varepsilon}_{T-1} \neq 1_{T-1}$ brings $L_T'$ into the form above, where the submatrix $[B, C]$ is precisely $(L_1^T)^'.

Our choice of $d + 1$ rows in Step 1 guarantees that $C$ is the matrix (2.20) and has full rank. Thus, it follows that $L_T$ has full column rank if and only if $L_T''$ does.

To understand the structure of $L_T''$, we need to recall the remark in Step 1 and the stock price difference (2.11). Note that the columns of $L_T''$ that correspond to $\hat{h}_{t,j}(x_t)$ for $1 \leq t \leq T - 1$, $1 \leq j \leq d$ and $\bar{\varepsilon}_t \neq 1_t$ in $\mathcal{E}_t$ remain the same as in $L_T'$, since these columns are identically zero in $(L_1^T)^'$. On the other hand, the nonzero coefficients of $(L_T^T)'$ in the columns that correspond to $\hat{h}_{t,j}(x_t)$ are subtracted from the coefficients of $L_T'$ in these columns. We state two particular instances for later use:

1. The columns in $L_T''$ that correspond to $\hat{h}_{T-1,j}(x_{T-1}^\varepsilon)$ for $1 \leq j \leq d$ repeat the negative values of those columns in $(L_1^T)^';$ cf. (2.16).

2. For any $\varepsilon_{T-1} \in \{0, 1\}^d$ and $1 \leq j \leq d$, the coefficient of $L_T''$ in the row corresponding to $(1_{T-2}, \varepsilon_{T-1}, 0)$ and the column corresponding to $\hat{h}_{T-2,j}(x_{T-2}^\varepsilon)$ is

$$
(x_{T-1,j}^{\varepsilon_{T-1}} - x_{T-2,j}^1) - (x_{T-1,j}^1 - x_{T-2,j}^1) = \begin{cases}
\delta_{T-1,j} & \text{if } \varepsilon_{T-1,j} = 0 \\
0 & \text{if } \varepsilon_{T-1,j} = 1.
\end{cases}
$$

Analogous patterns hold in the columns corresponding to $\hat{h}_{t,j}(x_t)$ for $1 \leq t \leq T - 3.$
Step 3. Note that $L''_T$ has $(d + 1)(2^{(T-1)} - 1)$ rows and $d \sum_{t=0}^{T-1} 2^d t$ columns. Fix $\vec{\varepsilon}_{T-1} \neq \mathbb{1}_{T-1}$. In $(L''_T)^\prime\prime$, we subtract its Row 1 from all other rows. In view of (i) in Step 2, a variant of (2.21) implies that for $1 \leq j \leq d$, Row $(d + 2 - j)$ has two nonzero entries of the same magnitude, $\delta_{T,j}$, with the opposite signs. More precisely, $-\delta_{T,j}$ is in the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$ and $\delta_{T,j}$ is in the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$; cf. (2.17). For $1 \leq j \leq d$, we add the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$ to the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$. As a result, for $1 \leq j \leq d$, Row $(d + 2 - j)$ now has precisely one nonzero entry, $-\delta_{T,j}$, and this entry is located in the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$; cf. (2.18). Moreover, for $1 \leq j \leq d$, the first coefficient of $(L''_T)^\prime\prime$ in the column corresponding to $\hat{h}_{T-1,j}(x^{\vec{\varepsilon}_{T-1}})$ is

$$
\begin{cases}
-\delta_{T-1,j} & \text{if } \varepsilon_{T-1,j} = 0 \\
0 & \text{if } \varepsilon_{T-1,j} = 1 
\end{cases}
$$

(2.22)

and all other coefficients are zero. We apply this procedure to all $\vec{\varepsilon}_{T-1} \neq \mathbb{1}_{T-1}$. As a result, $d(2^{d(T-1)} - 1)$ rows have only one nonzero entry each, and these entries respectively belong to the first $d(2^{d(T-1)} - 1)$ columns of $L''_T$. Dropping these rows and columns, we denote the remaining matrix by $M_T$. Thus, we have shown that $L_T$ has full column rank if and only if $M_T$ does.

Step 4. We now show that $M_T$ indeed has full column rank. We argue by induction on $T$; the base case $T = 2$ was established in Remark 2.4.4. Note that $M_T$ has $(2^{d(T-1)} - 1)$ rows and $d \sum_{t=0}^{T-2} 2^d t$ columns, where

- the rows correspond to $(\vec{\varepsilon}_{T-1}, 0)$ for all $\vec{\varepsilon}_{T-1} \neq \mathbb{1}_{T-1}$ in $E_{T-1}$;

- the first $d$ columns correspond to $\hat{h}_{T-1,j}(x^{1_{T-1}})$ for $1 \leq j \leq d$; and

- the rest of the columns correspond to $\hat{h}_{t,j}(x^{\vec{\varepsilon}_t})$ for $t \in \{T - 2, \ldots, 1\}$, $\vec{\varepsilon}_t \in E_t$ and $j \in \{1, \ldots, d\}$. 

32
The matrix in Figure 2.3(a) may serve as an illustration. As

\[
\{(\tilde{\epsilon}_{T-1}, 0) : \tilde{\epsilon}_{T-1} \in \mathcal{E}_{T-1}\} = \{(\tilde{\epsilon}_{T-2}, \epsilon_{T-1}, 0) : \tilde{\epsilon}_{T-2} \in \mathcal{E}_{T-2}, \epsilon_{T-1} \in \{0, 1\}^{d}\}
\]

we can divide \(M_T\) into submatrices (\(M_T^{\tilde{\epsilon}_{T-2}}\))\(_{\tilde{\epsilon}_{T-2} \in \mathcal{E}_{T-2}}\) such that the rows of \(M_T^{\tilde{\epsilon}_{T-2}}\) correspond to \((\tilde{\epsilon}_{T-2}, \epsilon_{T-1}, 0)\) for all \(\epsilon_{T-1} \in \{0, 1\}^{d}\). Each \(M_T^{\tilde{\epsilon}_{T-2}}\) has \(2^{d}\) rows, with the exception that \(M_T^{\tilde{\epsilon}_{T-2}}\) has \(2^{d} - 1\) rows (since \(M_T\) does not contain the row corresponding to \(1_{T-1}\)). For this reason, we need to treat the cases \(d \geq 2\) and \(d = 1\) separately.

(a) Case \(d \geq 2\): Similarly to Step 1, the rows in \(M_T\) are linearly dependent, and we can reduce \(M_T\) to a matrix \(M'_T\) of \((d + 1)2^{d(T-2)}\) rows without decreasing its rank. That is, \(M'_T\) consists of \(2^{d(T-2)}\) submatrices \((M_T^{\tilde{\epsilon}_{T-2}})'\) of \(d + 1\) rows for \(\tilde{\epsilon}_{T-2} \in \mathcal{E}_{T-2}\). As a consequence of (2.22), the first \(d\) columns of \((M_T^{\tilde{\epsilon}_{T-2}})'\) make up a \((d + 1) \times d\) matrix

\[
C'' := \begin{bmatrix}
-\delta_{T-1,1} & -\delta_{T-1,2} & \ldots & -\delta_{T-1,d-1} & -\delta_{T-1,d} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \ddots & -\delta_{T-1,d-1} & \vdots \\
-\delta_{T-1,1} & 0 & \ldots & \vdots & \vdots \\
0 & -\delta_{T-1,2} & \ldots & -\delta_{T-1,d-1} & -\delta_{T-1,d}
\end{bmatrix}
\]

which does not depend on \(\tilde{\epsilon}_{T-2}\) and has full column rank. Thus, the first \(d\) columns of \(M'_T\) are formed by vertically stacking \(C'' \ 2^{d(T-2)}\) times.

Recalling (ii) from Step 2, we note that the last \(d+1\) rows of \(M'_T\) correspond to \((1_{T-2}, 0, 0)\) and \((1_{T-2}, e_j, 0)\) for \(1 \leq j \leq d\). As a result, the submatrix of \(M'_T\) formed by taking the last \(d + 1\) rows
and the columns corresponding to \( h_{T-2,j}(x^{1_{T-2}}) \) for \( 1 \leq j \leq d \) is precisely \(-C''\). Now, adding the columns corresponding to \( \hat{h}_{T-1,j}(x^{1_{T-1}}) \) to the columns corresponding to \( \hat{h}_{T-2,j}(x^{1_{T-2}}) \) for \( 1 \leq j \leq d \) can bring \( M'_T \) to the block triangular form

\[
\begin{bmatrix}
A'' & M''_T \\
C''
\end{bmatrix}.
\]

(b) Case \( d = 1 \): There is no need to remove any rows from \( M_T \). We set \( M'_T = M_T \) and note that it has \( 2^{T-1} - 1 \) rows. A subtle difference to (a) is that \( M'_T \) does not contain the row corresponding to \((1_{T-1}, e_1, 0)\). (We remark that \( e_1 = 1 \) and \( 0 = 0 \) in the case \( d = 1 \).) The coefficients of \( M'_T \) in the first column alternate between \(-\delta_{T-1,1}\) and 0. Also, the last row of \( M'_T \) has only one nonzero entry, \( \delta_{T-1,1} \), in the column corresponding to \( h_{T-2,1}(x^{1_{T-2}}) \). Adding the column corresponding to \( \hat{h}_{T-1,1}(x^{1_{T-1}}) \) to the column corresponding to \( \hat{h}_{T-2,1}(x^{1_{T-2}}) \) brings \( M'_T \) to the form above with the \( 1 \times 1 \) matrix \( C'' = [-\delta_{T-1,1}] \).

Therefore, in either case, \( M''_T \) is a matrix with \( (d+1)(2^d(T-2) - 1) \) rows and \( d \sum_{t=1}^{T-2} 2^{d_t} \) columns.

It remains to show that it has full column rank. Applying the same procedure as in Step 3 to \( M''_T \), we see that \( d(2^d(T-2) - 1) \) rows have only one nonzero entry each, and none of these entries share a common column. We then drop these rows and columns. One can check that the remaining matrix, of \( (2^d(T-2) - 1) \) rows and \( d \sum_{t=0}^{T-3} 2^{d_t} \) columns, is precisely \( M_{T-1} \); we omit further details in the interest of brevity. The inductive hypothesis applies.
Figure 2.2: Illustration of $L_T$ in the proof of Lemma 2.4.1 for $T = 3$ and $d = 2$. Columns 1–32 correspond to the variables $\hat{h}_{2,j}(x^{\tilde{e}_2})$ for $1 \leq j \leq 2$ and $\tilde{e}_2 \in \mathcal{E}_2$, Columns 33–40 correspond to the variables $\hat{h}_{1,j}(x^{\tilde{e}_1})$ for $1 \leq j \leq 2$ and $\tilde{e}_1 \in \mathcal{E}_1$, and lastly Columns 41–43 correspond to the variables $\hat{g}_1(x^{0}_{3,1}), \hat{g}_1(x^{1}_{3,1}), \hat{g}_2(x^{1}_{3,2})$. 

35
Figure 2.3: Illustration of Step 4 in the proof of Lemma 2.4.1 for $T = 3$ and $d = 2$. The row and column numbers in black are inherited from $L_3$, and the new row and column numbers are printed in red. The first two columns correspond to $\hat{h}_{2,j}(x^{12})$ for $j = 1, 2$; the rest of the columns correspond to $\hat{h}_{1,j}(x^{\vec{\varepsilon}_1})$ for $j = 1, 2$ and $\vec{\varepsilon}_1 \in \mathcal{E}_1$. Coefficients in each light gold block are the negatives of those in the dark gold block.
Chapter 3: Martingale Schrödinger Bridges and Optimal Semistatic Portfolios

The chapter is based on the article [46] of the same title, authored by Marcel Nutz, Johannes Wiesel and Long Zhao. It is published in Finance and Stochastics.

3.1 Introduction and Main Results

The martingale Schrödinger bridge was introduced by Henry-Labordère [34] as a pricing model achieving perfect calibration to all Vanilla options while retaining stylized facts of a reference model. Starting from a reference stochastic volatility model (SVM) which typically cannot be calibrated perfectly, the martingale Schrödinger bridge is constructed as the calibrated measure which is closest to the SVM in the sense of relative entropy. In contrast to the classical Schrödinger bridge in Léonard [39] and Avellaneda et al. [3, 4], this problem features an additional martingale constraint to generate an arbitrage-free model. A similar approach is used by Guyon [30, 31] in a two-period setting to solve the longstanding joint S&P 500/VIX smile calibration puzzle; here entropy minimization is utilized to construct a model that is jointly calibrated to the S&P 500, VIX futures and VIX options.

The aforementioned works rest on (sometimes implicit) mathematical assumptions of strong duality and attainment. These are plausible as natural extensions of standard results in markets without option trading (see Delbaen et al. [19], Frittelli [25], Schachermayer [51], Zariphopoulou [56], among others). However, Acciaio et al. [1] exhibited a surprising obstacle to obtaining such extensions: the space of semistatic portfolios of stocks and options is not closed (both in a two-period model and in continuous time). In classical mathematical finance, closedness results are at the very heart of the separation arguments underlying the Fundamental Theorem of Asset Pricing.
and the existence of optimal portfolios for utility maximization. As a consequence, it is not obvious how to formulate and prove the desired results.

The purpose of the present paper is to provide such results, at least in one setting. On the one hand, we prove strong duality between the martingale Schrödinger bridge problem and an exponential utility maximization problem over semistatic portfolios. This duality, as well as the existence of the martingale Schrödinger bridge itself (primal attainment), is obtained along the lines of classical entropy minimization and Schrödinger bridge theory. On the other hand, we prove (under a technical condition) that the dual problem is attained in a natural space of admissible portfolios, and that this dual solution yields the log-density of the martingale Schrödinger bridge. We thus derive from first principles the type of implicit condition assumed on the optimal log-density, e.g., in Guyon [31, Theorem 16], and overcome the non-closedness issue discovered in Acciaio et al. [1]. To wit, while in general a convergent sequence of semistatic portfolios may have an undesirable limit with unclear financial interpretation, the specific limit of a utility-maximizing sequence in our problem is shown to be an admissible portfolio.

We consider a two-period model where the price of a stock is modeled by the canonical process \((X, Y)\) on \(\mathbb{R}^2\) under a (physical) reference probability \(P\). Here \(X\) is the stock price at date \(t = 1\) and \(Y\) is the price at the terminal date \(t = 2\). In addition, European options \(g(Y)\) are liquidly traded at time zero. By the Breeden–Litzenberger formula [13], the risk-neutral distribution \(\nu\) of \(Y\) can be derived from the prices of call options with arbitrary strikes, where \(\nu\) is assumed to have a finite first moment. Then the arbitrage-free price of a general option \(g(Y)\) is given by the integral \(E^\nu[g]\). The martingale Schrödinger bridge problem can now be formalized as

\[
\inf_{Q \in \mathcal{M}(\nu)} H(Q \| P),
\]
where $H$ is the relative entropy (or Kullback–Leibler divergence)

$$H(Q|P) := \begin{cases} 
E^Q \left[ \log \frac{dQ}{dP} \right] , & Q \ll P \\
\infty , & Q \not\ll P 
\end{cases}$$

and $\mathcal{M}(\nu)$ is the set of calibrated equivalent martingale measures,

$$\mathcal{M}(\nu) := \{ Q \in \mathcal{P}(\mathbb{R}^2) : Q \sim P, Q^2 = \nu, E^Q[Y|X] = X \}. \quad (3.2)$$

Here $\mathcal{P}(\mathbb{R}^2)$ is the set of probability measures on $\mathbb{R}^2$ and $Q^2$ denotes the second marginal of $Q \in \mathcal{P}(\mathbb{R}^2)$, or equivalently, the distribution of the price $Y$ under $Q$.

We remark in passing that (3.1) relates to the classical (static) Schrödinger bridge problem $\inf_{Q \in \Pi(\mu, \nu)} H(Q|P)$ over the set $\Pi(\mu, \nu)$ of couplings of two measures $\mu, \nu$; see Föllmer [22], Léonard [39] and Nutz [42] for surveys. In this problem, there is no martingale constraint. On the other hand, (3.1) relates to the martingale optimal transport problem $\inf_{Q \in \mathcal{M}(\mu, \nu)} E^Q[c]$ which minimizes an integrated cost over the set $\mathcal{M}(\mu, \nu)$ of martingale couplings; see Beiglböck et al. [5], Galichon et al. [26], Hobson [35] and the literature thereafter. In that problem, there is no reference measure. The resulting value yields model-independent bounds for the price of the exotic option $c$ and as a consequence of the linear structure, solutions tend to be degenerate. By contrast, solutions of (3.1) tend to preserve features of the reference model $P$, as emphasized in Henry-Labordère [34].

The classical Schrödinger bridge problem arises from the classical optimal transport problem by entropic regularization as used in the context of Sinkhorn’s algorithm by Cuturi and Peyré [15, 47]. Similarly, entropic regularization of martingale optimal transport leads to the martingale Schrödinger bridge, and this was used in De March and Henry-Labordère [40] to develop a version of Sinkhorn’s algorithm for martingale optimal transport. See also Guo and Oblój [29] for a related algorithm using a different relaxation.
Returning to our problem (3.1)—for it to be meaningful, we must assume that

\[ \mathcal{M}_{\text{fin}}(\nu) := \{ Q \in \mathcal{M}(\nu) : H(Q|P) < \infty \} \neq \emptyset; \]  

(3.3)

that is, there exists a calibrated martingale measure with finite relative entropy. This condition implies the absence of arbitrage in semistatic trading strategies. It implies the usual no-arbitrage condition on the stock alone, but also depends on the interplay of \( P \) and \( \nu \). A precise characterization of (3.3), or even just \( \mathcal{M}(\nu) \neq \emptyset \), in terms of trading strategies along the lines of a fundamental theorem of asset pricing in Dalang et al. [16], is an interesting open problem. (Like the question studied in the present paper, the answer is not obvious due to the failure of closedness in Acciaio et al. [1].) We can now state the basic wellposedness result.

**Proposition 3.1.1.** The problem (3.1) admits a unique minimizer \( Q_* \in \mathcal{M}(\nu) \), called the martingale Schrödinger bridge.

This will essentially follow from standard entropy minimization theory in Csiszár [14] and properties of \( \mathcal{M}(\nu) \) which are variations of results found, e.g., in Beiglböck et al. [5]. Proposition 3.1.1 lacks a more specific description: we expect by (formal) duality that the log-density of \( Q_* \) corresponds to a semistatic portfolio with certain admissibility criteria, and those criteria are crucial for any further analysis of the martingale Schrödinger bridge and its computation (as seen, e.g., in Guyon [31]). Specifically, trading in our market gives rise to a semistatic outcome of the form

\[ V = h(X)(Y - X) + g(Y), \]

where \( h(X) \) is the number of stocks held over the second period. Stock trading in the first period, starting from a deterministic initial stock price \( X_0 \), corresponds to a term \( h_0(X_0)(X - X_0) \) which can be absorbed into the functions \( h, g \) above and hence will not be represented explicitly. We write

\[ \mathcal{V} = \{ V \text{ measurable: } V = h(X)(Y - X) + g(Y) \text{ for some } h, g : \mathbb{R} \to \mathbb{R} \}. \]  

(3.4)
In order to have a well-defined option price, the function $g$ needs to be (measurable and) integrable under the pricing measure $\nu$. We thus set

$$V_1 := \{ V \in \mathcal{V} : h, g \text{ are measurable}, g \in L^1(\nu), E^\nu[g] = 0 \}$$  \hspace{1cm} (3.5)

for those outcomes whose option is available from zero initial capital. Finally, we want $h(X)(Y - X)$ to have suitable martingale properties. There is some flexibility here regarding the definition; one natural choice is to require the martingale property under all $Q \in \mathcal{M}_{\text{fin}}(\nu)$ (see also Remark 3.2.7 for another possible choice). For $V \in V_1$, this is equivalent to $V \in L^1(Q)$ for all $Q \in \mathcal{M}_{\text{fin}}(\nu)$. In summary, our set of admissible portfolios (for zero initial capital) is

$$\mathcal{V}_{\text{adm}} = \left\{ V \in \mathcal{V} : h, g : \mathbb{R} \to \mathbb{R} \text{ are measurable}, E^\nu[g] = 0, E^Q[h(X)(Y - X)] = 0 \text{ for all } Q \in \mathcal{M}_{\text{fin}}(\nu) \right\}.$$  

We then have the following strong duality between the martingale Schrödinger bridge (primal) problem and the dual problem of exponential utility maximization over semistatic portfolios.

**Proposition 3.1.2.** Let $u(x) = -e^{-\gamma x}/\gamma$ for some $\gamma > 0$. Then

$$\frac{1}{\gamma} \inf_{Q \in \mathcal{M}(\nu)} H(Q|P) = \sup_{V \in \mathcal{V}_{\text{adm}}} u^{-1} \left( E^P[u(V)] \right).$$  \hspace{1cm} (3.6)

The duality will be obtained by showing that the log-density of $Q_*$ can be approximated by semistatic portfolios with good integrability properties; cf. Proposition 3.2.4. That proposition, in turn, is inspired by seminal results in the theory of (classical) Schrödinger bridges, especially Föllmer’s construction of Schrödinger potentials [22]. Our argument does not require dual attainment and thus avoids discussing delicate properties of the portfolios: the supremum in (3.6) would be the same if taken, say, over portfolios $h(X)(Y - X) + g(Y)$ with bounded continuous functions $h, g$. But of course, this space would not allow for attainment in general.

Turning to the delicate part, we want to show that the dual problem is attained at an admissible
portfolio \( V_* \) and that this maximizer yields the log-density of \( Q_* \). We denote by \( P = P^1 \otimes P^* \) the disintegration of \( P \); that is, \( P^1 \) is the law of \( X \) under \( P \) and \( P^*(x,dy) \) is the conditional law of \( Y \) given \( X = x \).

**Theorem 3.1.3.** Suppose that \( dP^*/d\nu \) is \( P^1 \)-a.s. uniformly bounded from above and away from zero. Then the minimizer \( Q_* \) of (3.1) is given by the density

\[
Z_* := \frac{dQ_*}{dP} = e^{H(Q_*|P)+V_*},
\]

(3.7)

where \( V_* \in V_{\text{adm}} \) is the unique solution of the dual problem,

\[
V_* = \arg \max_{V \in V_{\text{adm}}} E^P[u(V)].
\]

In particular, \( V_* = h(X)(Y - X) + g(Y) \), where \( h, g \) are measurable functions with \( g \in L^1(\nu) \) and \( E^\nu[g] = 0 \) as well as \( h(X)(Y - X) \in L^1(Q) \) and \( E^Q[h(X)(Y - X)] = 0 \) for all \( Q \in M_{\text{fin}}(\nu) \).

The boundedness condition in Theorem 3.1.3 can be weakened to an integrability condition; see Remark 3.3.4. In contrast to the other results, this theorem does not seem to follow from classical arguments. If the space of admissible portfolios were closed, the theorem would follow from the approximation result in Proposition 3.2.4, broadly as in the classical framework of mathematical finance without options. To overcome the failure of closedness (specifically, of \( V_1 \) and \( V_{\text{adm}} \), as shown in Acciaio et al. [1]), we first leverage a result from our companion paper [45], where it is shown that the functional form of semistatic portfolios is stable under pointwise limits. As a consequence, the approximation result still implies that \( V_* \) is of the general form \( V_* = h(X)(Y - X) + g(Y) \) for some measurable functions \( h, g \).

On the flip side, another insight from [45] is that the key failure in the counterexample of [1] is the integrability of the option \( g \) which is in turn crucial to associate a price. Hence, it is not surprising that establishing this integrability occupies the lion’s share of the proof of Theorem 3.1.3; it uses novel arguments and seems to be the first result in this direction. Our line of attack is to con-
struct a measure $\tilde{Q}$ in (a relaxation of) $\mathcal{M}_{\text{fin}}(\nu)$ such that $h(X)(Y - X)$ is $\tilde{Q}$-integrable; once that is achieved, soft arguments imply that $g \in L^1(\nu)$. In fact, we establish that such measures are dense: in Proposition 3.3.2 we show that any $Q \in \mathcal{M}_{\text{fin}}(\nu)$ is the limit of calibrated (absolutely continuous) martingale measures $\tilde{Q}_n$ under which the dynamic trading strategy $h$ is uniformly bounded a.s. The proof is intricate and develops, among other things, explicit stability properties of the convex order, building on ideas from martingale optimal transport in Beiglböck and Juillet [6]. See also Section 3.3.3 for further comments.

We do not know how far the technical condition on $P$ in Theorem 3.1.3 can be relaxed. However, analogy with the classical Schrödinger bridge problem suggests that some condition may be necessary. Indeed, the corresponding question in that setting—without martingale constraint but with two marginal constraints—is to show that the log-density of the Schrödinger bridge is of the form $f(x) + g(y)$ and establish the measurability and integrability properties of those “Schrödinger potentials” $(f, g)$. This problem has a long history (e.g., Beurling [9]). A series of results revealed that the additive form $f(x) + g(y)$ always holds, but also that the measurability of $(f, g)$ fails without additional conditions; moreover, even when measurability holds, integrability fails without further conditions (see Borwein and Lewis [11], Csiszár [14], Föllmer and Gantert [23], Rüschendorf and Thomsen [49, 50]). The study of Schrödinger potentials remains an area of active study (see for instance Altshuler et al. [2], Deligiannidis et al. [20], Gigli and Tamanini [27], Nutz and Wiesel [43, 44]) that we have benefited from, especially for our companion paper [45]. For the present work, we have not been able to transfer as many of those techniques.

Regarding potential future work, it seems likely that our line of argument can be extended to show the existence of optimal portfolios for more general utility functions. Generalizations in the structure of the market, for instance also adding options with maturity $t = 1$, are relatively straightforward in the general parts whereas replacing our argument for the integrability of the option is nontrivial. Similarly, an extension to multiple periods through backward induction is not obvious, as our proof crucially relies on the technical condition (3.13). In a different direction, one may remember how Rogers [48] used existence for exponential utility to show the Fundamental
Theorem of Asset Pricing. Of course, that is not immediately applicable here, as we have used (3.3) in our proof of existence.

The remainder of the paper has a simple structure: Section 3.2 derives the wellposedness and duality results (Propositions 3.1.1 and 3.1.2), and Section 3.3 provides the proof of dual attainment (Theorem 3.1.3).

3.2 Wellposedness and Duality

In this section, we first prove the wellposedness of the martingale Schrödinger bridge \( Q_* \) (Proposition 3.1.1). Then, we prove the duality with exponential utility maximization (Proposition 3.1.2) through an approximation of \( Q_* \) (Proposition 3.2.4).

We start by recalling a general result on entropy minimization.

**Lemma 3.2.1.** Consider a measurable space \((\Omega, \mathcal{F})\) and denote by \(\mathcal{P}(\Omega)\) its collection of probability measures. Fix \(R \in \mathcal{P}(\Omega)\), let \(Q \subseteq \mathcal{P}(\Omega)\) be convex and closed in variation, and suppose that \(Q_{\text{fin}} := \{Q \in Q : H(Q|R) < \infty\} \neq \emptyset\). Then there exists a unique \(Q_* \in Q\) such that

\[
H(Q_* | R) = \inf_{Q \in Q} H(Q | R) \in [0, \infty).
\]

Moreover, \(Q_* \gg Q\) for any \(Q \in Q_{\text{fin}}\). In particular, if there exists \(Q \in Q_{\text{fin}}\) with \(Q \sim R\), then \(Q_* \sim R\). Furthermore,

\[
\log \frac{dQ_*}{dR} \in L^1(Q) \quad \text{for all} \quad Q \in Q_{\text{fin}}.
\]  

(3.8)

**Proof.** In the stated form, the result can be found in Nutz [42, Theorem 1.10 and Corollary 1.13]. Its main part is very classical; cf. Csiszár [14]. The integrability (3.8) is less known but can also be deduced from [14]. \(\square\)

Lemma 3.2.1 is not directly applicable to the set \(Q = \mathcal{M}(\nu)\) of martingale measures defined in (3.2) as this set is not closed due to the equivalence constraint. Writing

\[
\Pi(\nu) = \{Q \in \mathcal{P}(\mathbb{R}^2) : Q^2 = \nu\},
\]
we consider instead the following relaxations defined with absolute continuity,

\[ \widetilde{M}(\nu) := \{ Q \in \Pi(\nu) : Q \ll P, E^Q[Y|X] = X \} \supseteq \mathcal{M}(\nu), \]

\[ \widetilde{M}_{\text{fin}}(\nu) := \{ Q \in \widetilde{M}(\nu) : H(Q|P) < \infty \} \supseteq \mathcal{M}_{\text{fin}}(\nu) \]  

(3.9)

and argue that \( \widetilde{M}(\nu) \) satisfies the hypotheses of Lemma 3.2.1. To this end, we first give an extension of Beiglböck et al. [5, Lemma 2.2 and Theorem 2.4]. Recall that two measures \( \mu, \nu \in \mathcal{P}(\mathbb{R}) \) are in convex order (e.g. in Shaked and Shanthikumar [53]), denoted \( \mu \preceq_c \nu \), if they have finite first moments and \( E^\mu[f] \leq E^\nu[f] \) holds for all convex functions \( f : \mathbb{R} \to \mathbb{R} \). As an example, \( E^Q[Y|X] = X \) implies \( Q^1 \preceq_c Q^2 \) by Jensen’s inequality.

**Lemma 3.2.2.** The set \( \{ Q \in \Pi(\nu) : E^Q[Y|X] = X \} \) is weakly closed.

**Proof.** Let \( (Q_n)_{n \geq 1} \) be a sequence of measures converging weakly to some limit \( Q \), then \( Q \in \Pi(\nu) \) by the continuity of the projection \( Y \). To see that \( E^{Q_n}[Y|X] = X \) implies \( E^Q[Y|X] = X \), we show that \( |X| + |Y| \) is \( (Q_n) \)-uniformly integrable. Indeed, as \( \{ \nu \} \) is uniformly integrable, the de la Vallée–Poussin theorem yields a convex function \( f : \mathbb{R} \to \mathbb{R}_+ \) of superlinear growth with \( \int f \, d\nu < \infty \). Thus

\[
\sup_{\mu : \mu \preceq_c \nu} \int f \, d\mu \leq \int f \, d\nu < \infty
\]

by the definition of the convex order, showing that \( \{ \mu : \mu \preceq_c \nu \} \) is uniformly integrable. As a result, \( |X| + |Y| \) is \( \{ Q \in \Pi(\nu) : Q^1 \preceq_c \nu \} \)-uniformly integrable and in particular \( (Q_n) \)-uniformly integrable.

We can now show the wellposedness of the martingale Schrödinger bridge \( Q_* \).

**Proof of Proposition 3.1.1.** Using Lemma 3.2.2, we readily verify that \( \widetilde{M}(\nu) \) is convex and closed in variation. Since \( \widetilde{M}_{\text{fin}}(\nu) \supseteq \mathcal{M}_{\text{fin}}(\nu) \neq \emptyset \) by our assumption (3.3), applying Lemma 3.2.1 with
$Q = \tilde{M}(\nu)$ yields existence and uniqueness of

$$Q_* = \arg\min_{Q \in \tilde{M}(\nu)} H(Q|P)$$

(3.10)

as well as $Q_* \sim P$; that is, $Q_* \in \mathcal{M}_{\text{fin}}(\nu)$. It now follows that $Q_*$ is also the unique minimizer of

$$\inf_{Q \in \tilde{M}(\nu)} H(Q|P).$$

We record the following observation for use in Section 3.3.

**Remark 3.2.3.** For any $Q \in \mathcal{M}_{\text{fin}}(\nu)$, a straightforward calculation shows that the density $Z := dQ/dP$ can be written as $Z = e^{H(Q|P)+V}$ for some $V \in L^1(Q)$ with $E^Q[V] = 0$. For the density

$$Z_* := \frac{dQ_*}{dP} = e^{H(Q_*|P)+V_*},$$

(3.11)

of the minimizer, we have not only that $E^{Q_*}[V_*] = 0$ but also that $V_* \in L^1(\tilde{Q})$ for all $\tilde{Q} \in \tilde{M}_{\text{fin}}(\nu)$. This follows from (3.8) by way of (3.10).

The next result characterizes the minimizer $Q_*$ through certain approximating sequences of semistatic portfolios and will serve as the basis to prove the duality (Proposition 3.1.2). We write $\mathcal{V}_b$ for the set of portfolios $V = h(X)(Y-X) + g(Y)$ where $h, g : \mathbb{R} \to \mathbb{R}$ are bounded measurable and $E^\nu[g] = 0$; clearly $\mathcal{V}_b \subseteq \mathcal{V}_{\text{adm}}$.

**Proposition 3.2.4.** Given $Q_* \in \mathcal{M}_{\text{fin}}(\nu)$ with density (3.11), the following statements are equivalent:

(i) $Q_*$ is the minimizer of (3.1).

(ii) There exist probability measures $(Q_n)_{n \geq 1}$ with densities

$$Z_n := dQ_n/dP = e^{H(Q_n|P)+V_n} \quad \text{with} \quad V_n \in \mathcal{V}_b$$

46
\[ H(Q_n|P) \to H(Q_\ast|P) \quad \text{and} \quad V_n \to V_\ast \quad \text{in} \quad L^1(Q_\ast) \quad \text{as} \quad n \to \infty. \]

(iii) There exist \((V_n)_{n \geq 1} \subseteq \mathcal{V}_b\) such that
\[ E^P[e^{V_n}] \to E^P[e^{V_\ast}] \quad \text{as} \quad n \to \infty. \]

(iv) There exist \((V_n)_{n \geq 1} \subseteq \mathcal{V}_b\) such that
\[ E^{Q_\ast}[e^{V_n - V_\ast - 1}] \to 0 \quad \text{as} \quad n \to \infty. \]

This result is inspired by a characterization of (classical) Schrödinger bridges in Föllmer and Gantert [23, Proposition 3.6], see also Csiszár [14], and Föllmer’s construction of Schrödinger potentials [22]. The key feature is that the approximating random variables \(V_n\) are portfolios and have good integrability properties (whereas the properties of \(V_\ast\) are unclear at this stage).

**Remark 3.2.5.** The assertion of Proposition 3.2.4 remains valid if \(\mathcal{V}_b\) is replaced by \(\mathcal{V}_{adm}\) or by \(\mathcal{V}_1\). This will be clear from the proof.

**Proof of Proposition 3.2.4.** (i) \(\Rightarrow\) (ii): By separability of \(L^1(\mathbb{R})\) we can write
\[ \mathcal{M}(\nu) = \{Q \sim P : E^Q[h_i(X)(Y - X)] = 0, E^Q[g_i(Y)] = 0, \ i = 1, 2, \ldots \} \]
for a countable collection of bounded measurable functions \(h_i, g_i : \mathbb{R} \to \mathbb{R}\). Denote the set of measures for which only the first \(n\) constraints are enforced by
\[ \mathcal{M}_n(\nu) = \{Q \sim P : E^Q[h_i(X)(Y - X)] = 0, E^Q[g_i(Y)] = 0, \ i = 1, 2, \ldots, n\}. \]
Clearly \(\mathcal{M}(\nu) \subseteq \mathcal{M}_n(\nu)\) and \(\mathcal{M}_n(\nu)\) is convex and closed in variation. Consider the problem
\[ \inf_{Q \in \mathcal{M}_{n}(\nu)} H(Q|P). \]  This minimization problem over measures with finitely many linear constraints is well known to be in duality with exponential utility maximization over (static) trading in the finitely many assets \( h_i(X)(Y - X), g_i(Y) \) defining the constraints. Specifically, by Föllmer and Schied [24, Section 3, esp. Corollary 3.25], the minimizer \( Q_n \) of \( \inf_{Q \in \mathcal{M}_{n}(\nu)} H(Q|P) \) is of the form

\[
Z_n := \frac{dQ_n}{dP} = \exp \left( c_n + \tilde{h}_n(X)(Y - X) + \tilde{g}_n(Y) \right)
\]

for some \( c_n \in \mathbb{R} \), where \( \tilde{h}_n(X) = \sum_{i=1}^{n} a_{i,n} h_i(X) \) and \( \tilde{g}_n(Y) = \sum_{i=1}^{n} b_{i,n} g_i(Y) \) for some \( a_{i,n}, b_{i,n} \in \mathbb{R} \). As \( Q_n \in \mathcal{M}_{n}(\nu) \) we have \( c_n = H(Q_n|P) \). That is, \( \log Z_n \) is of the form

\[
H(Q_n|P) + V_n \quad \text{for some} \quad V_n \in \mathcal{V}_b.
\]

Applying Nutz [42, Theorem 1.17] to the sets \( \mathcal{Q}_n := \mathcal{M}_n(\nu) \) and \( Q := \mathcal{M}(\nu) \) satisfying \( \cap_n \mathcal{Q}_n = Q \), and recalling that \( \mathcal{M}_{\text{fin}}(\nu) \neq \emptyset \) by our assumption (3.3), we conclude that

\[
H(Q_*|Q_n) \to 0, \quad H(Q_n|P) \to H(Q_*|P) \quad \text{and} \quad \log Z_n \to \log Z_* \text{ in } L^1(Q_*).
\]

In particular, \( V_n \to V_* \) in \( L^1(Q_*) \) follows.

1. \( (ii) \Rightarrow (iii) \): Since \( Z_n, Z \) are probability densities, we have \( e^{-H(Q_n|P)} = E^P[e^{V_n}] \) and \( e^{-H(Q_*|P)} = E^P[e^{V_*}] \). Thus \( H(Q_n|P) \to H(Q_*|P) \) is equivalent to \( E^P[e^{V_n}] \to E^P[e^{V_*}] \).

2. \( (iii) \Leftrightarrow (iv) \) : By a change of measure, (iii) is equivalent to

\[
E^{Q_*}[e^{V_n - V_*}] = e^{H(Q_*|P)} E^P[e^{V_n}] \to e^{H(Q_*|P)} E^P[e^{V_*}] = 1,
\]

and now Scheffé’s lemma yields the equivalence with (iv).

1. \( (iii) \Rightarrow (i) \): Without loss of generality, we assume \( E^P[e^{V_n}] < \infty \) for all \( n \). Define probability measures \( Q_n \) by

\[
Z_n := \frac{dQ_n}{dP} = e^{H(Q_n|P) + V_n}
\]

and recall that (iii) is equivalent to \( H(Q_n|P) \to H(Q_*|P) \). Take any \( Q \in \mathcal{M}_{\text{fin}}(\nu) \). Using the
definition of \( H(\cdot|P) \) and Lemma 3.2.6 below,

\[
H(Q|P) - H(Q|Q_n) = E^Q[\log Z_n] = H(Q_n|P) + E^Q[V_n] = H(Q_n|P).
\]

As \( H(Q|Q_n) \geq 0 \), it follows that

\[
H(Q|P) \geq \lim_{n \to \infty} H(Q_n|P) = H(Q^*|P).
\]

Since \( Q \in \mathcal{M}_{\text{fin}}(\nu) \) was arbitrary, we conclude that \( Q^* \) is the minimizer of (3.1).

The following technical result was used in the preceding proof.

**Lemma 3.2.6.** Let \( V \in \mathcal{V}_1 \) satisfy \( E^P[e^V] < \infty \). Then \( V \in L^1(Q) \) and \( E^Q[V] = 0 \) for all \( Q \in \mathcal{M}_{\text{fin}}(\nu) \).

**Proof.** Define an auxiliary probability measure \( Q' \) via

\[
Z' := \frac{dQ'}{dP} = e^{c+V},
\]

where \( c \in \mathbb{R} \) is the normalization constant. Moreover, let \( Q \in \mathcal{M}_{\text{fin}}(\nu) \) and denote by \( Z \) its density. Applying the inequality \( \log x \leq x - 1 \) to \( x = z'/z > 0 \) yields \( \log z' \leq \log z + z'/z - 1 \) and hence

\[
\log Z' \leq \log Z + Z'/Z - 1 \quad \text{on} \quad \{Z > 0\},
\]

where \( \log 0 := -\infty \). In view of \( H(Q|P) < \infty \), we have \( \log Z + Z'/Z - 1 \in L^1(Q) \) and conclude that \( (\log Z')^+ \in L^1(Q) \). By the definition of \( \mathcal{V}_1 \),

\[
\log Z' = c + V = h(X)(Y - X) + g(Y)
\]

for some \( g \in L^1(\nu) \) with \( E^\nu[g] = 0 \), and hence \( (\log Z')^+ \in L^1(Q) \) translates to the positive part of the martingale transform \( h(X)(Y - X) \) being \( Q \)-integrable. As \( Q \) is a martingale measure,
this already implies (see Jacod and Shiryaev [37, Theorem 2b]) that $h(X)(Y - X) \in L^1(Q)$ and $E^Q[h(X)(Y - X)] = 0$. The claim follows.

We are now in a position to prove the duality result.

**Proof of Proposition 3.1.2.** Let $Q_*$ be the minimizer from Proposition 3.1.1 and recall from (3.11) the notation $Z_* = dQ_*/dP = e^{H(Q_*|P)+V_*}$ where $E^{Q_*}[V_*] = 0$. Let $u(x) = -e^{-\gamma x}/\gamma$ for some $\gamma > 0$. A change of measure and Jensen’s inequality yield that for any $V \in \mathcal{V}_{adm}$,

$$E^P[u(V)] = E^{Q_*} [Z_*^{-1}u(V)] = -\frac{1}{\gamma} E^{Q_*} [e^{-H(Q_*|P)+V_*-\gamma V}]$$

$$\leq -\frac{1}{\gamma} e^{-H(Q_*|P)+E^{Q_*}[V_*]-\gamma E^{Q_*}[V]} = -\frac{1}{\gamma} e^{-H(Q_*|P)},$$

where the last equality used that $E^{Q_*}[V] = 0$ due to $Q_* \in \mathcal{M}_{fin}(\nu)$ and $V \in \mathcal{V}_{adm}$.

On the other hand, Proposition 3.2.4 (iv) shows that there exist $(V_n) \subseteq \mathcal{V}_b$ such that

$$E^{Q_*} [e^{-V_*-\gamma V_n}] \to 1$$

and consequently $-E^{Q_*} [e^{-H(Q_*|P)-V_*-\gamma V_n}] \to -e^{-H(Q_*|P)}$. In view of $\mathcal{V}_b \subseteq \mathcal{V}_{adm}$, this yields

$$\sup_{V \in \mathcal{V}_{adm}} E^P[u(V)] \geq -\frac{1}{\gamma} e^{H(Q_*|P)}.$$

Lastly,

$$\inf_{Q \in \mathcal{M}(\nu)} u\left(\frac{1}{\gamma} H(Q|P)\right) = u\left(\frac{1}{\gamma} H(Q_*|P)\right) = -\frac{1}{\gamma} e^{-H(Q_*|P)},$$

so that combining the two inequalities yields

$$\sup_{V \in \mathcal{V}_{adm}} E^P[u(V)] = \inf_{Q \in \mathcal{M}(\nu)} u\left(\frac{1}{\gamma} H(Q|P)\right)$$

as claimed. □
Remark 3.2.7. By the proof, the duality (3.6) still holds if the supremum is taken over the larger set $\mathcal{V}_1 \cap L^1(Q_\ast) \supseteq \mathcal{V}_{\text{adm}}$, providing an alternative definition of admissibility.

3.3 Admissibility and Dual Attainment

3.3.1 Preliminary Considerations

Let $Q_\ast$ be the minimizer from Proposition 3.1.1 and recall from (3.11) the notation $Z_\ast = dQ_\ast/dP = e^{H(Q_\ast|P)+V_\ast}$. With a view towards the duality relation, note that

$$E^P \left[ u \left( \frac{1}{\gamma} V_\ast \right) \right] = -E^P \left[ \frac{1}{\gamma} e^{V_\ast} \right] = -E^P \left[ \frac{1}{\gamma} e^{-H(Q_\ast|P)} Z_\ast \right] = -\frac{1}{\gamma} e^{-H(Q_\ast|P)}.$$

It is thus tempting to conclude that $-V_\ast/\gamma$ “attains” the supremum in (3.6). However, it far from obvious whether $V_\ast$ belongs to the dual domain $\mathcal{V}_{\text{adm}}$ (or is a portfolio in any sense). At this stage, we know that $E^{Q_\ast}[V_\ast] = 0$ and that $V_\ast$ is the limit of certain portfolios $(V_n) \subseteq \mathcal{V}_b \subseteq \mathcal{V}_{\text{adm}} \subseteq \mathcal{V}_1$; cf. Proposition 3.2.4. The missing conclusion would be obvious if any of these spaces had a good closure property. However, as mentioned in the Introduction, Acciaio et al. [1] have shown that this is not the case: specifically, the authors exhibit a two-period model and an $L^p$-convergent sequence $(V_n) \subseteq \mathcal{V}_b$ whose limit is outside $\mathcal{V}_1$. The proof uses a clever contradiction argument avoiding a detailed study of the limiting random variable, so it may not be clear what exactly goes wrong in the limit.

The first possible issue is whether the limit still has the functional form $h(X)(Y-X) + g(Y)$ for some functions $h, g$. A second issue is whether (these functions are measurable and) $g$ is integrable as required by the definition of $\mathcal{V}_1$. The first issue is analyzed in our companion paper [45] which shows that the functional form is stable even under pointwise limits. Under the mild condition that $P \sim P^1 \otimes P^2$ (which is implied by the condition in Theorem 3.1.3), we can also guarantee that $h, g$ remain measurable.

Lemma 3.3.1. We have $V_\ast \in \mathcal{V}$; that is, $V_\ast = h(X)(Y-X) + g(Y)$ for some functions $h, g : \mathbb{R} \to \mathbb{R}$. If $P \sim P^1 \otimes P^2$, the functions $h, g$ are a.s. uniquely determined and measurable.
Proof. By Proposition 3.2.4 we can find \((V_n) \subseteq \mathcal{V}_b\) with \(V_n \to V_* P\text{-a.s.}\). The two claims then follow from Nutz et al. [45, Theorem 2.2 and 3.1], respectively.

This stability of the functional form indicates that the key failure in the counterexample of Acciaio et al. [1] is the integrability of the option. It is then clear that some original arguments will be required to obtain that the option position in our specific limit \(V_*\) is nevertheless integrable—which motivates the rest of this section.

3.3.2 Proof of Theorem 3.1.3

Recall that the disintegration of a probability measure \(R \in \mathcal{P}(\mathbb{R}^2)\) is denoted \(R = R^1 \otimes R^\bullet\), where \(R^1\) is the first marginal (distribution of \(X\)) and \(R^\bullet : \mathbb{R} \to \mathcal{P}(\mathbb{R})\) is a stochastic kernel (conditional distribution of \(Y \) given \(X\)). We interchangeably use \(R^\bullet(x), R^\bullet(x, \cdot)\) or \(R^\bullet(x, dy)\) to denote the conditional distribution given \(X = x\).

Our basic line of attack is simple (yet seems to be novel): recalling Remark 3.2.3,

\[
V_* = h(X)(Y - X) + g(Y) \in L^1(Q) \quad \text{for all} \quad Q \in \tilde{\mathcal{M}}_{\text{fin}}(\nu),
\]

(3.12)

where \(\tilde{\mathcal{M}}_{\text{fin}}(\nu)\) was defined in (3.9). We shall construct \(\tilde{Q} \in \tilde{\mathcal{M}}_{\text{fin}}(\nu)\) such that \(h\) is uniformly bounded \(\tilde{Q}^1\)-a.s. Then clearly \(h(X)(Y - X) \in L^1(\tilde{Q})\) and now (3.12) yields \(g(Y) \in L^1(\tilde{Q})\), or equivalently \(g \in L^1(\nu)\), as desired.

On the other hand, the construction of \(\tilde{Q}\) is somewhat intricate. It is based on an approximation of \(Q_*\) by a sequence of probability measures \((\tilde{Q}_n)\), which simultaneously retain the martingale property and satisfy \(h\) is \(\tilde{Q}_n^1\)-a.s. uniformly bounded for all \(n \in \mathbb{N}\). Next, we state a general version of this approximation result, applicable to any measurable function \(h : \mathbb{R} \to \mathbb{R}\) and any \(Q \in \mathcal{M}_{\text{fin}}(\nu)\) satisfying the technical condition (3.13) below. In the proof of Theorem 3.1.3, the result will be applied to the specific function \(h\) in \(V_* = h(X)(Y - X) + g(Y)\) and \(Q = Q_*\).

**Proposition 3.3.2.** Let \(h : \mathbb{R} \to \mathbb{R}\) be measurable and \(Q = Q^1 \otimes Q^\bullet \in \mathcal{M}_{\text{fin}}(\nu)\). Suppose that
there exists a $Q^1$-integrable function $I : \mathbb{R} \to [0, \infty)$ such that

$$H(Q^*(x')|P^*(x)) \leq I(x') \quad \text{for } (Q^1 \otimes Q^1)\text{-a.a. } (x, x').$$  \hfill (3.13)

Then there exist measures $\tilde{Q}_n = \tilde{Q}_n^1 \otimes \tilde{Q}_n^* \in \tilde{\mathcal{M}}_\text{fin}(\nu)$ such that

1. $h$ is $\tilde{Q}_n^1$-a.s. uniformly bounded for each $n \in \mathbb{N}$,

2. $\tilde{Q}_n \to Q$ in variation,

3. $H(\tilde{Q}_n|P) \to H(Q|P)$.

In particular there exists $\tilde{Q} \in \tilde{\mathcal{M}}_\text{fin}(\nu)$ such that $h$ is uniformly bounded $\tilde{Q}^1$-a.s.

The proof is lengthy and deferred to Section 3.3.3. For ease of reference, we record some standard facts in the next lemma.

**Lemma 3.3.3.** Given probability measures $Q = Q^1 \otimes Q^*$ and $R = R^1 \otimes R^*$ on $\mathbb{R}^2$,

1. $Q \ll R$ if and only if $Q^1 \ll R^1$ and $Q^* \ll R^* Q^1$-a.s.,

2. if $Q \ll R$, then

$$\frac{dQ}{dR} = \frac{dQ^1}{dR^1} \frac{dQ^*}{dR^*} \quad R\text{-a.s.},$$

3. $H(Q|R) = H(Q^1|R^1) + E^{Q^1}[H(Q^*|R^*)]$.

We are now ready to detail the proof of Theorem 3.1.3.

**Proof of Theorem 3.1.3.** Set $\mu := Q^1_\ast$. We first construct a function $I$ satisfying (3.13) with $Q = Q_\ast$. By our assumption, there are constants $0 < l < L < \infty$ such that

$$l \leq \frac{dP^*}{d\nu}(y) \leq L \quad (\mu \otimes \nu)\text{-a.a. } (x, y).$$  \hfill (3.14)
Using Lemma 3.3.3(i), \( Q_* \sim P \) implies that \( Q_*^* \sim P^* \sim \nu \)-a.s. Note also

\[
\frac{dQ_*^*(x')}{dP^*(x)}(y) = \frac{dQ_*^*(x')}{dP^*(x)}(y) \frac{dP^*(x')}{dP^*(x)}(y)
= \frac{dQ_*^*(x')}{dP^*(x)}(y) \frac{dP^*(x')}{d\nu}(y) \frac{d\nu}{dP^*(x)}(y), \quad (\mu \otimes \mu \otimes \nu)-a.a. \ (x, x', y).
\]

(3.15)

Combining (3.14) and (3.15) we obtain

\[
\log \frac{dQ_*^*(x')}{dP^*(x)}(y) \leq \log \frac{dQ_*^*(x')}{dP^*(x)}(y) + \log(L/l)
\]

and now integrating against \( Q_*^*(x') \) yields

\[
H(Q_*^*(x')|P^*(x)) \leq H(Q_*^*(x')|P^*(x')) + \log(L/l) =: I(x'), \quad (\mu \otimes \mu)-a.a. \ (x, x').
\]

Lemma 3.3.3 (iii) together with \( H(Q_*|P) < \infty \) then implies \( I \in L^1(\mu) \).

Noting that \( Q_* \sim \mu \otimes \nu \), Lemma 3.3.1 yields that

\[
V_* = h(X)(Y - X) + g(Y)
\]

for some measurable functions \( h, g \). Next, we verify that \( g \) is \( \nu \)-integrable with \( E^\nu[g] = 0 \). Indeed, the function \( I \) satisfies (3.13) with \( Q = Q_* \), hence Proposition 3.3.2 provides \( \widetilde{Q} \in \tilde{\mathcal{M}}_{\text{fin}}(\nu) \) such that \( h \) is \( \widetilde{Q}^1 \)-a.s. uniformly bounded. This clearly implies \( h(X)(Y - X) \in L^1(\widetilde{Q}) \). Recalling from Remark 3.2.3 that \( V_* \) is \( \widetilde{Q} \)-integrable, we can deduce that \( g(Y) \in L^1(\widetilde{Q}) \); that is, \( g \in L^1(\widetilde{Q}^2) = L^1(\nu) \). We can now conclude from \( E^{\widetilde{Q}_*}[V_*] = 0 \) that \( E^\nu[g] = E^{\widetilde{Q}_*}[g(Y)] = 0 \), completing the proof that \( V_* \in \mathcal{V}_1 \).

Recall from Remark 3.2.3 that \( V_* \in L^1(Q) \) for all \( Q \in \tilde{\mathcal{M}}_{\text{fin}}(\nu) \). Having established \( g \in L^1(\nu) \), this implies \( h(X)(Y - X) \in L^1(Q) \) and then \( E^Q[h(X)(Y - X)] = 0 \) by the martingale property. As \( \tilde{\mathcal{M}}_{\text{fin}}(\nu) \supset M_{\text{fin}}(\nu) \), this shows that \( V_* \in \mathcal{V}_{\text{adm}} \).

\begin{remark}
As seen in the proof, the boundedness condition in Theorem 3.1.3 can be weakened
\end{remark}
to the following integrability condition:

1. \( P \sim P^1 \otimes \nu, \)

2. there exists a \( Q^1 \) integrable function \( I : \mathbb{R} \to [0, \infty) \) such that

\[
E^{Q^1(x')} \left[ \log \frac{dP^\ast(x')}{dP^\ast(x)} \right] \leq I(x') \quad \text{for } (P^1 \otimes P^1) \text{-a.a. } (x, x').
\]

### 3.3.3 Proof of Proposition 3.3.2

The program for this proof can be sketched as follows. First, we shall identify a sequence \( (\mu_n) \) of sub-probability measures \( \mu_n \ll Q^1 \) such that \( h \) is uniformly bounded \( \mu_n \)-a.e. and, when renormalized, \( \mu - \mu_n \) dominates \( \mu_n \) in convex order. Strassen’s theorem then guarantees the existence of martingale measures \( M_n \) with first marginal \( \mu_n \) and second marginal \( \mu - \mu_n \). The desired measures \( \widetilde{Q}_n \) have marginals \( \mu_n/\mu_n(\mathbb{R}) \) and \( \nu \): they will be built by embedding mass \( \mu_n(\mathbb{R}) \) according to \( Q^\ast \) and mass \( 1 - \mu_n(\mathbb{R}) \) according to the composition of \( M_n^\ast \) with \( Q^\ast \).

Let us first recall that the convex order of two probability measures \( \mu, \nu \) can be characterized via their quantile functions \( F_{\mu}^{-1}, F_{\nu}^{-1} \). Indeed \( \mu \preceq_c \nu \) if and only if

\[
\int_u^1 F_{\mu}^{-1}(p) \, dp \leq \int_u^1 F_{\nu}^{-1}(p) \, dp \quad \text{(3.16)}
\]

for all \( u \in [0, 1] \), with equality for \( u = 0 \), see Shaked and Shanthikumar [53, Theorem 3.A.5]. If \( \mu, \nu \) are finite measures with the same total mass, then \( \mu \preceq_c \nu \) if and only if \( \mu/\mu(\mathbb{R}) \preceq_c \nu/\nu(\mathbb{R}) \).

In particular, we can apply the characterization (3.16) to these normalized measures. To simplify notation, we omit the normalizing constant and write \( F_{\mu}^{-1} \) instead of \( F_{\mu/\mu(\mathbb{R})}^{-1} \) in this case.

As a preparation for the proof of Proposition 3.3.2, we first establish two lemmas. Lemma 3.3.5 (i) has the same assertion as Beiglböck and Juillet [6, Example 2.4] but is obtained with a different, more quantitative argument which is then used in Lemma 3.3.5 (ii) to elaborate on finer properties. Those properties are instrumental for the proof of Lemma 3.3.6 which describes a stability property of the convex order that will be applied in the proof of Proposition 3.3.2.
Lemma 3.3.5. Let \( A = [a, b] \subseteq \mathbb{R} \). Suppose that \( \mu_A \) and \( \mu_B \) are finite measures with the same mass and zero barycenter such that \( \mu_A \) is concentrated on \( A \) and \( \mu_B \) is concentrated on \( B := \mathbb{R} \setminus (a, b) \).

1. We have \( \mu_A \preceq_c \mu_B \).

2. Define
\[
E := \left\{ u \in (0, 1) : \int_u^1 F^{-1}_{\mu_A}(p) \, dp = \int_u^1 F^{-1}_{\mu_B}(p) \, dp \right\}.
\]
Then \( E \) is of the form \((0, \alpha] \cup [\beta, 1)\) for some \( 0 \leq \alpha \leq \beta \leq 1 \). Furthermore,

(a) \( E = (0, 1) \) if and only if \( \mu_A = \mu_B \), in which case both measures are concentrated on \( \{a, b\} \),

(b) \( \mu_B((-\infty, a)) = 0 \) and \( \mu_B(a) > \mu_A(a) = \alpha \), whenever \( \alpha > 0 \) and \( E \neq (0, 1) \),

(c) \( \mu_B((b, \infty)) = 0 \) and \( \mu_B(b) > \mu_A(b) = 1 - \beta \), whenever \( \beta < 1 \) and \( E \neq (0, 1) \).

Proof. We may assume that \( \mu_A \) and \( \mu_B \) are probability measures. We first show (i) by verifying (3.16) for \( u \in (0, 1) \). Indeed, define
\[
u^* := \sup \left\{ p \in (0, 1) : F^{-1}_{\mu_B}(p) \leq a \right\}.
\]

Note that \( F^{-1}_{\mu_B}(p) \geq b \) for all \( p \in (u^*, 1) \) and \( F^{-1}_{\mu_A}(p) \in [a, b] \) for all \( p \in (0, 1) \). Hence
\[
\int_u^1 F^{-1}_{\mu_A}(p) \, dp \leq \int_u^1 F^{-1}_{\mu_B}(p) \, dp \tag{3.17}
\]
for all \( u \in [u^*, 1) \). Suppose, towards a contradiction, that there exists \( \hat{u} \in (0, u^*) \) such that (3.17) holds with the reverse, strict inequality at \( u = \hat{u} \). As \( F^{-1}_{\mu_A}(p) \geq a \) for all \( p \in (0, 1) \) and \( F^{-1}_{\mu_B}(p) \leq a \) for all \( p \in (0, u^*) \), we deduce that
\[
\int_0^1 F^{-1}_{\mu_A}(p) \, dp > \int_0^1 F^{-1}_{\mu_B}(p) \, dp, \tag{3.18}
\]

\[1\]The conventions \((0, 0] := \emptyset\) and \([1, 1) := \emptyset\) are used.
contradicting that \( \mu_A \) and \( \mu_B \) have the same barycenter. This shows (i).

Turning to (ii), the proof of (a) is immediate. We can thus assume that there exists \( \tilde{u} \in (0, 1) \) such that (3.17) holds with strict inequality at \( u = \tilde{u} \). If there exists no \( \hat{u} \in (0, \tilde{u}) \) such that

\[
\int_{\tilde{u}}^{1} F_{\mu_A}(p) \, dp = \int_{\tilde{u}}^{1} F_{\mu_B}(p) \, dp ,
\]

then \( \alpha = 0 \). Whereas if such \( \hat{u} \) exists, then necessarily \( F_{\mu_B}(p) \leq a \) for all \( p \in (0, \hat{u}] \), for otherwise \( F_{\mu_B}(\hat{u}) \geq b \) and the equality in (3.19) cannot hold. It follows that

\[
\int_{u}^{1} F_{\mu_A}(p) \, dp \geq \int_{u}^{1} F_{\mu_B}(p) \, dp \tag{3.20}
\]

for all \( u \in (0, \hat{u}] \). Since we have shown the reverse inequality in (i), we conclude that (3.20) holds with equality for all \( u \in (0, \hat{u}] \). That is, \( E \) contains an interval of the form \((0, \alpha] \) for some \( \alpha \geq 0 \).

Changing the integral bounds from \((u, 1)\) to \((0, u)\) by subtracting the barycenter on both sides of the above equations, an analogous argument shows that \( E \) contains an interval of the form \([\beta, 1) \) for some \( \beta \in [0, 1] \). In conclusion, \( E \) is of the form \((0, \alpha] \cup [\beta, 1) \) for \( 0 \leq \alpha \leq \beta \leq 1 \).

To show (b), suppose that \( E \neq (0, 1) \) and \( \alpha > 0 \). The assumption implies that

\[
\int_{0}^{\alpha} F_{\mu_A}(p) \, dp = \int_{0}^{\alpha} F_{\mu_B}(p) \, dp .
\]

Consequently, \( F_{\mu_A}^{-1}(p) = a = F_{\mu_B}^{-1}(p) \) for all \( p \in (0, \alpha] \). That is, \( \mu_B((-\infty, a)) = 0 \), along with \( \mu_A(a) \geq \alpha \) and \( \mu_B(a) \geq \alpha \). Since

\[
\int_{0}^{u} F_{\mu_A}^{-1}(p) \, dp > \int_{0}^{u} F_{\mu_B}^{-1}(p) \, dp \quad \text{for} \quad u \in (\alpha, \beta) ,
\]

it is necessary that \( F_{\mu_A}^{-1}(p) \in (a, b] \) for \( p \in (\alpha, 1) \) and that \( F_{\mu_B}^{-1}(p) = a \) for all \( p > \alpha \) that are sufficiently close to \( \alpha \). We conclude that \( \mu_B(a) > \mu_A(a) = \alpha \), showing (b). Part (c) is proved analogously. \( \Box \)
Lemma 3.3.6. In the setting of Lemma 3.3.5, suppose that $\mu_A \neq \mu_B$ are probability measures. Let $(\mu^n_A), (\mu^n_B)$ be sequences of probability measures with barycenter zero such that $\mu^n_A \ll \mu_A$ for all $n$ as well as $d_{TV}(\mu^n_A, \mu_A) \to 0$, $d_{TV}(\mu^n_B, \mu_B) \to 0$ and $W_1(\mu^n_B, \mu_B) \to 0$ for $n \to \infty$. Then $\mu^n_A \preceq_c \mu^n_B$ for all $n$ sufficiently large.

Proof. Consider the set $E$ in Lemma 3.3.5 (ii). As $\mu_A \neq \mu_B$, we have $E \neq (0,1)$ and $\alpha < \beta$.

Let us consider the cases $\alpha > 0$ and $\alpha = 0$ separately. If $\alpha > 0$, Lemma 3.3.5 (ii) (b) states that $\bar{\alpha} := \mu_B(a) > \mu_A(a) = \alpha$. Let $\bar{\alpha}_n := \mu^n_B(a)$ and $\alpha_n := \mu^n_A(a)$. Since $d_{TV}(\mu^n_A, \mu_A) \to 0$ and $d_{TV}(\mu^n_B, \mu_B) \to 0$ as $n \to \infty$, it follows that $\bar{\alpha}_n \to \bar{\alpha}$ and $\alpha_n \to \alpha$. In view of $\mu^n_A \ll \mu_A$, we conclude that $F_{\mu^n_A}^{-1}(p) \in [a,b]$ for $p \in (0,1)$. Fix $\varepsilon < \bar{\alpha} - \alpha$. Then $\bar{\alpha}_n > \alpha + \varepsilon$ when $n$ is sufficiently large, and

$$\int_0^u F_{\mu^n_A}^{-1}(p) \, dp \geq \int_0^u F_{\mu^n_B}^{-1}(p) \, dp \quad \text{for} \quad u \in (0, \alpha + \varepsilon). \tag{3.21}$$

Whereas in the case $\alpha = 0$, we define $\bar{\alpha} := \mu_B((-\infty, a]) > 0$ and $\bar{\alpha}_n := \mu^n_B((-\infty, a])$. Again, for a fixed $\varepsilon < \bar{\alpha} - \alpha = \bar{\alpha}$, we have $\bar{\alpha}_n > \alpha + \varepsilon$ when $n$ is sufficiently large, and (3.21) holds.

Similarly, we consider the cases $\beta < 1$ and $\beta = 1$ and define $\bar{\beta}$ accordingly. In either case we can fix $\varepsilon < \beta - \bar{\beta}$ and find $n$ sufficiently large so that $\bar{\beta}_n < \beta - \varepsilon$ and

$$\int_u^1 F_{\mu^n_A}^{-1}(p) \, dp \leq \int_u^1 F_{\mu^n_B}^{-1}(p) \, dp \quad \text{for} \quad u \in [\beta - \varepsilon, 1). \tag{3.22}$$

To complete the proof, it remains to show that the inequality in (3.22) holds for $u \in O := (\alpha + \varepsilon, \beta - \varepsilon)$, where $\varepsilon < \min\{\bar{\alpha} - \alpha, \beta - \bar{\beta}\}$ is fixed. Note that $(0,1) \setminus E = (\alpha, \beta)$ and $O \subseteq (\alpha, \beta)$. As the integrals below are continuous functions of $u$, there exists $\gamma > 0$ such that

$$\int_u^1 F_{\mu^n_A}^{-1}(p) \, dp + \gamma \leq \int_u^1 F_{\mu^n_B}^{-1}(p) \, dp \quad \text{for all} \quad u \in O, \tag{3.23}$$

thanks to the definition of $E$. In view of $d_{TV}(\mu^n_A, \mu_A) \to 0$ and $d_{TV}(\mu^n_B, \mu_B) \to 0$, the quantile
functions converge pointwise. Moreover, we recall that the 1-Wasserstein distance satisfies

\[ W_1(\mu_B^n, \mu_B) = \int_0^1 \left| F_{\mu_B^n}^{-1}(p) - F_{\mu_B}^{-1}(p) \right| dp. \]

Dominated convergence and \( W_1(\mu_B^n, \mu_B) \to 0 \) thus imply that

\[ \lim_{n \to \infty} \int_u^1 F_{\mu_B^n}^{-1}(p) dp = \int_u^1 F_{\mu_B}^{-1}(p) dp, \quad \lim_{n \to \infty} \int_u^1 F_{\mu_B}^{-1}(p) dp = \int_u^1 F_{\mu_B}^{-1}(p) dp \]

uniformly in \( u \in O \). It now follows from (3.23) that

\[ \int_u^1 F_{\mu_B^n}^{-1}(p) dp \leq \int_u^1 F_{\mu_B}^{-1}(p) dp + \frac{\gamma}{2} \leq \int_u^1 F_{\mu_B}^{-1}(p) dp - \frac{\gamma}{2} \leq \int_u^1 F_{\mu_B}^{-1}(p) dp \]

for all \( u \in O \) and \( n \in \mathbb{N} \) large enough. This completes the proof.

Given measures \( \lambda, \mu \) on \( \mathbb{R} \), we write \( \lambda \leq \mu \) if \( \lambda(A) \leq \mu(A) \) for all \( A \in \mathcal{B}(\mathbb{R}) \). The total variation distance between \( \lambda \) and \( \mu \) is defined as

\[ d_{TV}(\lambda, \mu) = \sup \left\{ |\lambda(A) - \mu(A)| : A \in \mathcal{B}(\mathbb{R}) \right\}. \]

If \( \lambda \leq \mu \), it is clear that \( d_{TV}(\lambda, \mu) = (\mu - \lambda)(\mathbb{R}) \). For ease of reference, we record the following consequence.

**Lemma 3.3.7.** Let \( 0 \neq \lambda \leq \mu \) be finite measures on \( \mathbb{R} \). Then the probability measures \( \bar{\lambda} := \lambda/\lambda(\mathbb{R}) \) and \( \bar{\mu} := \mu/\mu(\mathbb{R}) \) satisfy \( \bar{\lambda} \ll \bar{\mu} \) and

\[ d_{TV}(\bar{\lambda}, \bar{\mu}) \leq \frac{(\mu - \lambda)(\mathbb{R})}{\mu(\mathbb{R})}. \] (3.24)
Proof. We have $\bar{\lambda} \ll \bar{\mu}$ and $\lambda(\mathbb{R}) \leq \mu(\mathbb{R})$, so that

$$d_{TV}(\bar{\lambda}, \bar{\mu}) = E^\mu \left[ (1 - \frac{d\bar{\lambda}}{d\bar{\mu}})^+ \right] = \frac{1}{\mu(\mathbb{R})} E^\mu \left[ (1 - \frac{\mu(\mathbb{R})}{\lambda(\mathbb{R})} \frac{d\lambda}{d\mu})^+ \right] \leq \frac{1}{\mu(\mathbb{R})} E^\mu \left[ (1 - \frac{d\lambda}{d\mu})^+ \right] = \frac{1}{\mu(\mathbb{R})} (\mu - \lambda)(\mathbb{R}).$$

We are now ready to give the proof of Proposition 3.3.2.

Proof of Proposition 3.3.2. To simplify notation we write $\mu := Q^1$ and assume without loss of generality that $\mu, \nu$ have zero barycenter. We may also assume that $\mu \neq \delta_0$; otherwise the claim is trivial. The proof proceeds in six steps.

**Step 1.** Given $\delta > 0$ sufficiently small we shall construct sub-probability measures $\mu_A, \mu_B$ (depending on $\delta$) that satisfy the hypotheses of Lemma 3.3.5, and hence $\mu_A \preceq_c \mu_B$.

As $\mu$ has zero barycenter and is not a Dirac measure, there exists $c > 0$ such that $\tilde{\delta} := \mu((-, -c]) \land \mu([c, \infty)) > 0$. Then, for all $0 < \delta < \tilde{\delta}$, there exist an interval $A := [a, b]$ and a measure $\lambda_A$ such that

$$\mu((-, a] \cup [b, \infty)) \leq \delta, \quad \mu((a, -c]) \geq \delta, \quad \mu([c, b]) \geq \delta$$

and

- $\lambda_A \leq \mu|_A$,
- $\lambda_A$ has zero barycenter,
- $\mu - \lambda_A$ is concentrated on $\mathbb{R} \setminus (a, b)$ and nonzero.

In particular $\mu|_{(a,b)} \leq \lambda_A \leq \mu|_A$. In fact, if $\mu$ has no atoms, we can set $\lambda_A = \mu|_A$. In the presence of atoms at $a$ or $b$, we may have to remove part of that mass so that $\lambda_A$ has zero barycenter and
\( \lambda_A \neq \mu \). Define
\[
\mu_A := \left( \frac{1}{\lambda_A(\mathbb{R})} - 1 \right) \lambda_A \quad \text{and} \quad \mu_B := \mu - \lambda_A.
\]
Evidently, the hypotheses of Lemma 3.3.5 are satisfied, and hence \( \mu_A \preceq_c \mu_B \).

**Step 2.** Fix \( \varepsilon > 0 \). As \( h \) takes values in \( \mathbb{R} \) there exists a set \( A^\varepsilon \subseteq A \) such that \( h \) is uniformly bounded on \( A^\varepsilon \) and
\[
\mu(A \setminus A^\varepsilon) \leq \frac{\delta \wedge \varepsilon}{2} \left( 1 \wedge \frac{c}{|a| \vee |b|} \right).
\]
(3.25)
The choice of the upper bound in (3.25) ensures that
\[
\mu((a, -c] \cap A^\varepsilon) \geq \frac{\delta}{2}, \quad \mu([c, b) \cap A^\varepsilon) \geq \frac{\delta}{2}
\]
and, writing \( X \) for the identity function on \( \mathbb{R} \) in an abuse of notation,
\[
|E^{\lambda_A} [1_{A^\varepsilon} X]| \leq |E^{\lambda_A} [1_A X]| + |E^{\lambda_A} [1_{A \setminus A^\varepsilon} X]|
\leq 0 + (|a| \vee |b|) \frac{(\delta \wedge \varepsilon)c}{2(|a| \vee |b|)} = \frac{(\delta \wedge \varepsilon)c}{2}.
\]
By restricting \( \lambda_A \) to \( A^\varepsilon \) and possibly removing some mass on \( (a, -c] \) or \([c, b)\), we can construct a measure \( \lambda_A^\varepsilon \) that is concentrated on \( A^\varepsilon \), satisfies \( \lambda_A^\varepsilon \leq \lambda_A \leq \mu \), has zero barycenter and
\[
d_{\text{TV}}(\lambda_A^\varepsilon, \lambda_A) = (\lambda_A - \lambda_A^\varepsilon)(\mathbb{R}) \leq \mu(A \setminus A^\varepsilon) + \frac{1}{c} |E^{\lambda_A} [1_{A^\varepsilon} X]| \leq \delta \wedge \varepsilon.
\]
In consequence,
\[
\lambda_A^\varepsilon(\mathbb{R}) = \lambda_A(\mathbb{R}) - (\lambda_A - \lambda_A^\varepsilon)(\mathbb{R}) \geq 1 - \delta - \delta \wedge \varepsilon.
\]
(3.26)
Step 3. Let $\delta \leq \bar{\delta}$ be fixed and consider a sequence $(\varepsilon_n)_{n \in \mathbb{N}} \downarrow 0$. For each $n \in \mathbb{N}$ we apply Step 3.3.3 to obtain

$$\mu_A^n := \left( \frac{1}{\lambda_A^n(\mathbb{R})} - 1 \right) \lambda_A^n \quad \text{and} \quad \mu_B^n := \mu - \lambda_A^n,$$

where $\lambda_A^n := \lambda_A^{\varepsilon_n}$. Both measures have the same total mass and zero barycenter for all $n \in \mathbb{N}$. Moreover we have $\lambda_A^n \leq \lambda_A$ and $d_{TV}(\lambda_A, \lambda_A^n) = (\lambda_A - \lambda_A^n)(\mathbb{R}) \leq \varepsilon_n \downarrow 0$.

In order to apply Lemma 3.3.6, we first need to scale the measures $\mu_A, \mu_B, \mu_A^n$ and $\mu_B^n$ so that they are probability measures. Set

$$\bar{\mu}_A = \frac{\lambda_A}{\lambda_A(\mathbb{R})} \quad \text{and} \quad \bar{\mu}_B = \frac{\mu - \lambda_A}{1 - \lambda_A(\mathbb{R})}$$

and define $\bar{\mu}_A^n$ and $\bar{\mu}_B^n$ analogously. Observe that $\bar{\mu}_A^n \ll \bar{\mu}_A$ and

$$d_{TV}(\bar{\mu}_A^n, \bar{\mu}_A) \leq \frac{(\lambda_A - \lambda_A^n)(\mathbb{R})}{\lambda_A(\mathbb{R})} \leq \frac{\varepsilon_n}{\lambda_A(\mathbb{R})} \downarrow 0$$

by (3.24). Similarly we have $d_{TV}(\bar{\mu}_B^n, \bar{\mu}_B) \to 0$. In particular it suffices to show that $E^{\bar{\mu}_B^n}[|X|] \to E^{\bar{\mu}_B}[|X|]$ in order to verify $W_1(\bar{\mu}_B^n, \bar{\mu}_B) \to 0$. In light of the definition of $\bar{\mu}_B$ in (3.28) and $d_{TV}(\lambda_A^n, \lambda_A) \to 0$ this readily follows from $E^{\lambda_A^n}[|X|] \to E^{\lambda_A}[|X|]$.

Now we are in a position to apply Lemma 3.3.6 to $\bar{\mu}_A, \bar{\mu}_B, \bar{\mu}_A^n$ and $\bar{\mu}_B^n$, which yields $n_0 \in \mathbb{N}$ such that $\bar{\mu}_A^n \preceq_c \bar{\mu}_B^n$. Since the convex order is invariant under scaling, it follows that

$$\mu_A^* := \mu_A^{n_0} \preceq_c \mu_B^{n_0} =: \mu_B^*.$$  

(3.29)

We recall that $h$ is uniformly bounded on $A^* := A^{\varepsilon_{n_0}}$ by construction and define

$$\lambda_A^* := \lambda_A^{n_0}$$

(3.30)

in preparation for Step 3.3.3 below.
Step 4. By Strassen [54, Theorem 8] the relation \( \mu_A^* \preceq \mu_B^* \) implies the existence of a mean-preserving probability kernel \( M^* \) : \( \mathbb{R} \to \mathcal{P}(\mathbb{R}) \) sending \( \mu_A^* \) to \( \mu_B^* \); that is, \( M_\delta := \mu_A^* \otimes M_\delta^* \in \Pi(\mu_A^*, \mu_B^*) \) and \( M_\delta^*(x) \) has barycenter \( x \) for all \( x \in \mathbb{R} \).

Recall that \( Q = \mu \otimes Q^* \in \mathcal{M}_{\text{fin}}(\nu) \) and denote by \( Q_\delta^* \) the composition of \( M_\delta^* \) with \( Q^* \),

\[
Q_\delta^*(x, C) := E^{M_\delta^*(x)}[Q^*(\cdot, C)] \quad \text{for} \quad C \in \mathcal{B}(\mathbb{R}).
\]

(3.31)

Note that \( Q_\delta^* \) is again mean-preserving.

Step 5. For \( \delta \leq \tilde{\delta} \), we set

\[
\tilde{Q}_\delta := \lambda_A^* \otimes Q^* + \mu_A^* \otimes Q_\delta^*,
\]

(3.32)

where \( \mu_A^* \), \( \lambda_A^* \) and \( Q_\delta^* \) were defined in (3.29), (3.30) and (3.31), respectively (the set \( A \) and the measures \( \lambda_A^*, \mu_A^* \) depend on \( \delta \)). In view of (3.27), the first marginal of \( \tilde{Q}_\delta \) is the probability measure \( \lambda_A^* / \lambda_A(\mathbb{R}) \).

We claim that \( \tilde{Q}_\delta \in \widetilde{\mathcal{M}}(\nu) \). Indeed, recall \( \mu \otimes Q^* = Q \sim P \) and observe that \( \mu_A^* \sim \lambda_A^* \ll \mu \), \( Q^* \sim \nu \mu_A^* \) in a.s. and \( Q_\delta^* \ll \nu \mu_A^* \)-a.s. Lemma 3.3.3 (i) then yields \( \tilde{Q}_\delta \ll P \). Moreover, the martingale property \( E^{\tilde{Q}_\delta}[Y|X] = X \) follows from the fact that \( Q^* \) and \( Q_\delta^* \) are mean-preserving. Finally, recall that \( Q_\delta^* \) is the composition of \( M_\delta^* \) with \( Q^* \), and \( \mu_A^* \otimes M_\delta^* \in \Pi(\mu_A^*, \mu_B^*) \). Thus \( \mu = \lambda_A^* + \mu_B^* \) and \( \mu \otimes Q^* \in \Pi(\mu, \nu) \) imply \( \tilde{Q}_\delta \in \Pi(\tilde{\mu}_A^*, \nu) \) and in particular \( \tilde{Q}_\delta^2 = \nu \). In summary, \( \tilde{Q}_\delta \in \widetilde{\mathcal{M}}(\nu) \) as desired.

From Step 2 and (3.26) we see that \( \lambda_A^* \leq \mu_\delta \) and \( \lambda_A^*(\mathbb{R}) \geq 1 - 2\delta \), hence \( \lambda_A^* \to \mu \) and \( \mu_A^* \to 0 \) in variation as \( \delta \to 0 \). It is now clear from the definition (3.32) that \( \tilde{Q}_\delta \to \mu \otimes Q^* = Q \) in variation.

Moreover, as \( \lambda_A^* \) is concentrated on \( A^* \) (cf. Step 3.3.3) and \( h \) is uniformly bounded on \( A^* \), we see that \( h \) is \( \tilde{\mu}_A^* \)-a.s. uniformly bounded for every \( \delta \leq \tilde{\delta} \). This proves Proposition 3.3.2 (i),(ii) after choosing \( \delta = \delta(n) \) small enough, modulo showing that \( H(\tilde{Q}_\delta|P) < \infty \) for small \( \delta \) (which will follow from the next step).
Step 6. As \( \liminf_{\delta \to 0} H(\tilde{Q}_\delta | P) \geq H(Q | P) \) due to the lower semicontinuity of \( H(\cdot | P) \) and the convergence \( \tilde{Q}_\delta \to Q \), it remains to show

\[
\limsup_{\delta \to 0} H(\tilde{Q}_\delta | P) \leq H(Q | P).
\] (3.33)

Note that \( \tilde{Q}_\delta \) is a convex combination of two probability measures:

\[
\tilde{Q}_\delta = \lambda^*_A(\mathbb{R}) \mu_A^* \otimes Q^* + [1 - \lambda^*_A(\mathbb{R})] \bar{\mu} \otimes Q^*_\delta.
\]

As \( H(\cdot | P) \) is convex, it follows that

\[
H(\tilde{Q}_\delta | P) \leq \lambda^*_A(\mathbb{R}) H(\bar{\mu} \otimes Q^* | P) + [1 - \lambda^*_A(\mathbb{R})] H(\bar{\mu} \otimes Q^* \delta | P).
\] (3.34)

We show that the first term converges to \( H(Q | P) \) and the second converges to zero. Indeed, Lemma 3.3.3 (iii) yields

\[
H(\bar{\mu} \otimes Q^* | P) = H(\bar{\mu}^*_A | P^1) + E^{\bar{\mu}^*_A} \left[ H(Q^* | P^*) \right],
\]

where

\[
H(\bar{\mu}^*_A | P^1) = \frac{1}{\lambda^*_A(\mathbb{R})} E^{\lambda^*_A} \left[ \log \left( \frac{d\lambda^*_A}{dP^1} \right) \right] - \log \lambda^*_A(\mathbb{R})
\]

\[
\to E^\mu \left[ \log \left( \frac{d\mu}{dP^1} \right) \right] = H(\mu | P^1)
\] (3.35)

by dominated convergence and \( \lambda^*_A(\mathbb{R}) \geq 1 - 2\delta \). Similarly, \( E^{\bar{\mu}^*_A} \left[ H(Q^* | P^*) \right] \to E^\mu \left[ H(Q^* | P^*) \right] \), so that the first term in (3.34) satisfies

\[
\lambda^*_A(\mathbb{R}) H(\bar{\mu}^*_A \otimes Q^* | P) \to H(\mu | P^1) + E^\mu \left[ H(Q^* | P^*) \right] = H(Q | R).
\]

It remains to show that the second term in (3.34) converges to zero,

\[
[1 - \lambda^*_A(\mathbb{R})] H(\bar{\mu}^*_A \otimes Q^* \delta | P) \to 0.
\]
Using again Lemma 3.3.3 (iii),

\[ H(\bar{\mu}_A^* \otimes Q_\delta^* | P) = H(\bar{\mu}_A^* | P^1) + E^{\bar{\mu}_A^*} [H(Q_\delta^* | P^*)]. \tag{3.36} \]

In view of (3.35) and $\lambda_A^*(\mathbb{R}) \to 1$, it follows that $[1 - \lambda_A^*(\mathbb{R})] H(\bar{\mu}_A^* | P^1) \to 0$. For the second term in (3.36), we use the definitions of $\bar{\mu}_A^*$ and $\mu_A^*$ to see that

\[ [1 - \lambda_A^*(\mathbb{R})] E^{\bar{\mu}_A^*} [H(Q_\delta^* | P^*)] = \frac{1 - \lambda_A^*(\mathbb{R})}{\lambda_A^*(\mathbb{R})} E^{\lambda_A^*} [H(Q_\delta^* | P^*)] = E^{\mu_A^*} [H(Q_\delta^* | P^*)]. \]

In view of (3.31), Jensen’s inequality and convexity of $H$ imply

\[ E^{\mu_A^*} [H(Q_\delta^* | P^*)] = E^{\mu_A^*} \left[ H(EM_\delta^*(X) | Q^*(X)) \right] \leq E^{\mu_A^*} \left[ EM_\delta^*(X) [H(Q_\delta^* | P^*(X))] \right]. \]

Lastly, the assumptions that $H(Q^*(x') | P^*(x)) \leq I(x')$ for $(\mu \otimes \mu)$-a.a. $(x, x')$ and $I \in L^1(\mu)$ together with the facts that $\mu_A^* \otimes M_\delta^* \in \Pi(\mu_A^*, \mu_B^*)$ and $\mu_B^* \to 0$ yield

\[ E^{\mu_A^*} \left[ EM_\delta^*(X) [H(Q_\delta^* | P^*(X))] \right] \leq E^{\mu_A^*} [EM_\delta^*(X) [I]] = E^{\mu_B^*} [I] \to 0 \]

by the dominated convergence theorem. This shows (3.33) and hence Proposition 3.3.2 (iii), completing the proof. \qed
This page intentionally left blank.
Bibliography

[1] B. Acciaio, M. Larsson, and W. Schachermayer. “The space of outcomes of semi-static trading strategies need not be closed”. In: Finance Stoch. 21.3 (2017), pp. 741–751.

[2] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme. “Asymptotics for semidiscrete entropic optimal transport”. In: SIAM J. Math. Anal. 54.2 (2022), pp. 1718–1741.

[3] M. Avellaneda. “Minimum-relative-entropy calibration of asset-pricing models”. In: Int. J. Theor. Appl. Finance 1.4 (1998), pp. 447–472.

[4] M. Avellaneda et al. “Weighted Monte Carlo: a new technique for calibrating asset-pricing models”. In: Int. J. Theor. Appl. Finance 4.1 (2001), pp. 91–119.

[5] M. Beiglböck, P. Henry-Labordère, and F. Penkner. “Model-independent bounds for option prices: a mass transport approach”. In: Finance Stoch. 17.3 (2013), pp. 477–501.

[6] M. Beiglböck and N. Juillet. “On a problem of optimal transport under marginal martingale constraints”. In: Ann. Probab. 44.1 (2016), pp. 42–106.

[7] M. Beiglböck, M. Nutz, and N. Touzi. “Complete duality for martingale optimal transport on the line”. In: Ann. Probab. 45.5 (2017), pp. 3038–3074.

[8] L. Bergomi. Stochastic volatility modeling. Chapman & Hall. CRC Press, Boca Raton, FL, 2016.

[9] A. Beurling. “An automorphism of product measures”. In: Ann. of Math. 72 (1960), pp. 189–200.

[10] F. Black and M. Scholes. “The pricing of options and corporate liabilities”. In: J. Polit. Econ. 81.3 (1973), pp. 637–654.

[11] J. M. Borwein and A. S. Lewis. “Decomposition of multivariate functions”. In: Canad. J. Math. 44.3 (1992), pp. 463–482.

[12] B. Bouchard and M. Nutz. “Weak dynamic programming for generalized state constraints”. In: SIAM J. Control Optim. 50.6 (2012), pp. 3344–3373.

[13] D. T. Breeden and R. H. Litzenberger. “Prices of state-contingent claims implicit in option prices”. In: J. Bus. 51.4 (1978), pp. 621–651.
[14] I. Csiszár. “$I$-divergence geometry of probability distributions and minimization problems”. In: *Ann. Probability* 3 (1975), pp. 146–158.

[15] M. Cuturi. “Sinkhorn Distances: Lightspeed Computation of Optimal Transport”. In: *Advances in Neural Information Processing Systems* 26. 2013, pp. 2292–2300.

[16] R. C. Dalang, A. Morton, and W. Willinger. “Equivalent martingale measures and no-arbitrage in stochastic securities market models”. In: *Stochastics and Stochastics Rep.* 29.2 (1990), pp. 185–201.

[17] F. Delbaen and W. Schachermayer. “A general version of the fundamental theorem of asset pricing”. In: *Math. Ann.* 300 (1994), pp. 463–520.

[18] F. Delbaen and W. Schachermayer. *The Mathematics of Arbitrage*. Berlin: Springer, 2006.

[19] F. Delbaen et al. “Exponential hedging and entropic penalties”. In: *Math. Finance.* 12 (2002), pp. 99–123.

[20] G. Deligiannidis, V. De Bortoli, and A. Doucet. “Quantitative Uniform Stability of the Iterative Proportional Fitting Procedure”. In: *Preprint arXiv:2108.08129v1* (2021).

[21] Y. Dolinsky and H. M. Soner. “Martingale Optimal Transport and Robust Hedging in Continuous Time”. In: *Probab. Theory Related Fields* 160.1–2 (2014), pp. 391–427.

[22] H. Föllmer. “Random fields and diffusion processes”. In: *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Vol. 1362. Lecture Notes in Math. Springer, Berlin, 1988, pp. 101–203.

[23] H. Föllmer and N. Gantert. “Entropy minimization and Schrödinger processes in infinite dimensions”. In: *Ann. Probab.* 25.2 (1997), pp. 901–926.

[24] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. 3rd. Berlin: W. de Gruyter, 2011.

[25] M. Frittelli. “The minimal entropy martingale measure and the valuation problem in incomplete markets”. In: *Math. Finance* 10.1 (2000), pp. 39–52.

[26] A. Galichon, P. Henry-Labordère, and N. Touzi. “A stochastic control approach to no-arbitrage bounds given marginals, with an application to Lookback options”. In: *Ann. Appl. Probab.* 24.1 (2014), pp. 312–336.

[27] N. Gigli and L. Tamanini. “Second order differentiation formula on $\text{RCD}^*(K,N)$ spaces”. In: *J. Eur. Math. Soc. (JEMS)* 23.5 (2021), pp. 1727–1795.
[28] P. Guasoni, M. Rásonyi, and W. Schachermayer. “Consistent price systems and face-lifting pricing under transaction costs”. In: Ann. Appl. Probab. 18.2 (2008), pp. 491–520.

[29] G. Guo and J. Obłój. “Computational methods for martingale optimal transport problems”. In: Ann. Appl. Probab. 29.6 (2019), pp. 3311–3347.

[30] J. Guyon. “The joint S&P 500/VIX smile calibration puzzle solved”. In: Risk (2020).

[31] J. Guyon. “Dispersion-Constrained Martingale Schrödinger Problems and the Exact Joint S&P 500/VIX Smile Calibration Puzzle”. In: Preprint SSRN:3853237 (2021).

[32] J. M. Harrison and D. M. Kreps. “Martingales and arbitrage in multiperiod securities markets”. In: J. Econom. Theory 20.3 (1979), pp. 381–408.

[33] J. M. Harrison and S. R. Pliska. “Martingales and stochastic integrals in the theory of continuous trading”. In: Stochastic Process. Appl. 11.3 (1981), pp. 215–260.

[34] P. Henry-Labordère. “From (Martingale) Schrödinger Bridges to a New Class of Stochastic Volatility Model”. In: Preprint SSRN:3353270 (2019).

[35] D. Hobson. “Robust hedging of the lookback option”. In: Finance Stoch. 2.4 (1998), pp. 329–347.

[36] D. Hobson. “The Skorokhod embedding problem and model-independent bounds for option prices”. In: Paris-Princeton Lectures on Mathematical Finance 2010. Vol. 2003. Lecture Notes in Math. Berlin: Springer, 2011, pp. 267–318.

[37] J. Jacod and A. N. Shiryaev. “Local martingales and the fundamental asset pricing theorems in the discrete-time case”. In: Finance Stoch. 2.3 (1998), pp. 259–273.

[38] D. M. Kreps. “Arbitrage and equilibrium in economies with infinitely many commodities”. In: J. Math. Econom. 8.1 (1981), pp. 15–35.

[39] C. Léonard. “A survey of the Schrödinger problem and some of its connections with optimal transport”. In: Discrete Contin. Dyn. Syst. 34.4 (2014), pp. 1533–1574.

[40] H. De March and P. Henry-Labordère. “Building Arbitrage-Free Implied Volatility: Sinkhorn’s Algorithm and Variants”. In: Preprint SSRN:3326486 (2019).

[41] R. C. Merton. “Theory of rational option pricing”. In: Bell J. Econom. and Management Sci. 4 (1973), pp. 141–183.

[42] M. Nutz. Introduction to Entropic Optimal Transport. https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf. Lecture notes, Columbia University, 2021.
[43] M. Nutz and J. Wiesel. “Entropic optimal transport: convergence of potentials”. In: Probab. Theory Related Fields 184.1-2 (2022), pp. 401–424.

[44] M. Nutz and J. Wiesel. “Stability of Schrödinger potentials and convergence of Sinkhorn’s algorithm”. In: Ann. Probab. 51.2 (2023), pp. 699–722.

[45] M. Nutz, J. Wiesel, and L. Zhao. “Limits of semistatic trading strategies”. In: Math. Finance 33.1 (2023), pp. 185–205.

[46] M. Nutz, J. Wiesel, and L. Zhao. “Martingale Schrödinger bridges and optimal semistatic portfolios”. In: Finance Stoch. 27.1 (2023), pp. 233–254.

[47] G. Peyré and M. Cuturi. “Computational Optimal Transport: With Applications to Data Science”. In: Foundations and Trends in Machine Learning 11.5-6 (2019), pp. 355–607.

[48] L. C. G. Rogers. “Equivalent martingale measures and no-arbitrage”. In: Stochastics and Stochastics Rep. 51.1-2 (1994), pp. 41–49.

[49] L. Rüschendorf and W. Thomsen. “Note on the Schrödinger equation and J-projections”. In: Statist. Probab. Lett. 17.5 (1993), pp. 369–375.

[50] L. Rüschendorf and W. Thomsen. “Closedness of sum spaces and the generalized Schrödinger problem”. In: Teor. Veroyatnost. i Primenen. 42.3 (1997), pp. 576–590.

[51] W. Schachermayer. “Optimal Investment in Incomplete Markets when Wealth may Become Negative”. In: Ann. Appl. Probab. 11.3 (2001), pp. 694–734.

[52] E Schrödinger. “Über die Umkehrung der Naturgesetze. Sitzungsberichte Preuss. Akad. Wiss.” In: Akad. Wiss., Berlin. Phys. Math. 144 (1931), pp. 144–153.

[53] M. Shaked and J. G. Shanthikumar. Stochastic orders. Springer Series in Statistics. Springer, New York, 2007.

[54] V. Strassen. “The existence of probability measures with given marginals”. In: Ann. Math. Statist. 36 (1965), pp. 423–439.

[55] J. A. Yan. “Caractérisation d’une classe d’ensembles convexes de $L^1$ ou $H^1$”. In: Seminar on Probability, XIV. Vol. 784. Lecture Notes in Math. Springer, Berlin, 1980, pp. 220–222.

[56] T. Zariphopoulou. “A solution approach to valuation with unhedgeable risks”. In: Finance Stoch. 5.1 (2001), pp. 61–82.