Optimal Transport and Informed Trading

by

Lu Vy

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in the University of Michigan
2026

Doctoral Committee:

Associate Professor Ibrahim Ekren, Co-Chair
Professor Uday Rajan, Co-Chair
Professor Erhan Bayraktar
Associate Professor Asaf Cohen

Lu Vy
luvy@umich.edu
ORCID iD: 0009-0007-1055-8250

© Lu Vy 2026

TABLE OF CONTENTS

LIST OF FIGURES iv

LIST OF TABLES v

LIST OF APPENDICES vi

LIST OF SYMBOLS vii

ABSTRACT viii

CHAPTER

1 Classical Kyle-Back in Continuous Time 1

1.1 The Model 1
1.1.1 Information Asymmetry 2
1.1.2 Definitions and Assumptions 4
1.2 The Equilibrium and its Properties 6
1.2.1 The Kantorovich problem 6
1.2.2 Main Theorem 7
1.2.3 Properties of the Equilibrium 9
1.2.4 The Wealth Decomposition 13
1.2.5 Proof of Main Theorem 15
1.3 Conclusion 16

2 Dynamic Information and Activism Trading 18

2.1 The Model 18
2.1.1 Information Asymmetry 21
2.1.2 Definitions and Assumptions 21
2.2 Main Result 23
2.3 Derivation of the Equilibrium 27
2.3.1 The Wealth Decomposition 27
2.3.2 Inconspicuous Trading 28
2.3.3 The Market Maker's Filtering Problem 30
2.4 A sobering counter-example 34
2.5 Conclusion 35

3 Multidimensional Stochastic Liquidity 37

3.1 The Model 38

3.1.1 Information Asymmetry . . . 41
3.1.2 Definitions . . . 41
3.2 Equilibrium . . . 43
3.2.1 The Primal Problem . . . 43
3.2.2 The Dual Problem . . . 44
3.2.3 Main Theorem . . . 52
3.2.4 Properties of the Equilibrium . . . 53
3.2.5 The Wealth Decomposition . . . 56
3.2.6 Proof of Main Theorem . . . 60
3.3 Examples . . . 61
3.3.1 Kyle 1985 . . . 61
3.3.2 Back and Pedersen 1998, static information . . . 62
3.3.3 Back, Cocquemas, Ekren, and Lioui 2021 . . . 63
3.3.4 Dufresne and Fos 2016, deterministic growth . . . 65
3.4 Existence of an Optimizer . . . 68
3.4.1 One Dimension . . . 68
3.5 Heuristic Derivation . . . 70
3.6 Conclusion . . . 72
4 Kyle's Single Auction Model . . . 74
4.1 The Model . . . 74
4.2 Main Theorem . . . 76
4.3 Relation to Static Mean Field Games . . . 81
4.4 Transaction Costs . . . 82
4.4.1 Existence . . . 85
4.5 Conclusion . . . 89
5 Summary . . . 91
5.1 Main Contributions . . . 91
5.2 Conceptual Implications . . . 92
5.3 Future Directions . . . 93
5.4 Final Remarks . . . 94

APPENDICES . . . 95

BIBLIOGRAPHY . . . 131

LIST OF FIGURES

FIGURE

1.1 Flow chart describing the classical Kyle-Back model. 2
1.2 Information asymmetry in prices. 3
1.3 Information asymmetry in trades. 3
1.4 PnL of market participants. Note that the market maker tends to break even, and the informed trader profits at the expense of noise traders. 6
1.5 The market maker's filtering problem. Numerical simulation with $\tilde{v} \sim \mathcal{N}(5, 1)$ and $\sigma = 3/2$. 10
1.6 The inconspicuous trading property. The noise trades and the aggregate trades are both Brownian motions in their own filtrations. 11
1.7 The informed trader's profit is the area under the residual supply curve. 16

2.1 Flow chart describing the dynamic information model. 19
2.2 Prices in the dynamic setting. 20
2.3 Trades in the dynamic setting. 20
2.4 The market maker's filtering problem in the dynamic case. 31

3.1 Flow chart describing multidimensional stochastic liquidity model. 38
3.2 Price information asymmetry in the multidimensional stochastic liquidity model. 39
3.3 Trading information asymmetry in the multidimensional stochastic liquidity model. 40
3.4 Numerical simulation of market maker's filtering problem. 54
3.5 The informed trader's profit. 58

4.1 Flow chart describing Kyle's single auction model. 75

A.1 The reflection principle. 111

LIST OF TABLES

TABLE

1.1 Summary of symbols in classical Kyle-Back model. 4

2.1 Summary of symbols in the dynamic information and activism trading model. . 21

3.1 Summary of symbols in stochastic liquidity model. 43

LIST OF APPENDICES

A Probability Theory **95**

B Stochastic Optimal Control Theory **121**

C Filtering Theory **123**

LIST OF SYMBOLS

$\tilde{v}$ the (hidden) value of a risky asset

P the market maker's pricing rule (his best guess of $\tilde{v}$)

Σ variance of $\tilde{v}$

X the informed trader's position

θ the informed trader's trading rate

Z market orders from liquidity / noise traders

σ_Z volatility of noise trading

Y net market orders; the sum of X and Z

S the informed trader's signal

σ_S volatility of private signal

π the informed trader's profits

J the informed trader's optimized expected profits

λ Kyle's lambda; inverse market depth

φ the Brenier (or Kantorovich) potential from $\mathcal{L}(Y_T)$ to $\mathcal{L}(\tilde{v})$

h the Brenier map from $\mathcal{L}(Y_T)$ to $\mathcal{L}(\tilde{v})$

$\mathcal{S}^n_+$ the set of symmetric positive semi-definite matrices

$\mathcal{S}^n_{++}$ the set of symmetric positive definite matrices

$\mathcal{S}^n_-$ the set of symmetric negative positive semi-definite matrices

$\mathcal{S}^n_{--}$ the set of symmetric negative positive definite matrices

ABSTRACT

We investigate applications of optimal transport theory to Kyle–Back models. We begin by developing a general methodology that constructs Kyle–Back equilibria from an optimal transport problem between the hidden asset price and liquidity trades. By combining Monge–Kantorovich duality with backward stochastic partial differential equations (BSPDEs), we represent equilibrium prices and trading strategies in terms of Kantorovich potentials and transport maps, yielding a unified framework that encompasses nearly all known continuous-time Kyle–Back models, including those with dynamic information and activist trading.

The second part of the dissertation introduces a variational formulation of the Kyle model that accommodates stochastic and multidimensional liquidity. In this formulation, the equilibrium is reinterpreted as an optimal liquidation of information: the insider's control variable is the rate at which information is released to the market, rather than inventory itself. This perspective links market microstructure to optimal execution theory and clarifies the role of the adjoint process identified by Collin–Dufresne and Fos (2016), showing that it is the co-state variable in the stochastic maximum principle associated with the dual problem.

Finally, we revisit Kyle's single-auction model and show that its equilibrium can be cast as a fixed-point problem on the space of probability measures. In particular, we prove that Kyle's single-auction equilibrium is a static mean-field equilibrium. This perspective does not rely on Gaussian assumptions and recovers the equilibria of Foster and Viswanathan (1993), Carré, Dufresne, and Gabriel (2022), and Cho and El Karoui (2000), demonstrating that the core informational structure of Kyle's model persists far beyond the Gaussian setting.

CHAPTER 1

Classical Kyle-Back in Continuous Time

There is a deep and elegant connection between optimal transport theory and Kyle–Back models. This connection was first made explicit by [BCEL20], who formulated an optimal transport problem based on the Fréchet–Hoeffding comonotone coupling [Fré51, Hoe40]. Although their problem initially appeared unrelated to informed trading or price impact, they showed that components of the problem's solution could be used to build a Kyle–Back equilibrium. Within this framework, the market maker's pricing rule, the price sensitivity (Kyle's λ), and the insider's expected profits all admit explicit representations in terms of Brenier potentials. While the economic interpretation of the transport problem itself remains somewhat opaque, its structural correspondence with equilibrium behavior is far from coincidental. We begin this chapter with a summary of their main results.

The model presented here was originally developed in [Bac92] as a generalization of [Kyl85]. Our exposition follows the later treatment of [BCEL20] in order to maintain consistency with the optimal-transport perspective that runs throughout this thesis. The resulting formulation serves as a reference point for the extensions explored in Chapters 2 and 3.

1.1 The Model

There is a public announcement to be made at time T which will affect the price of a risky asset. Immediately after this announcement, the asset will be valued at $\tilde{v}$, possibly jumping in price. The distribution of $\tilde{v}$ is public information, but no one knows exactly what value $\tilde{v}$ will take except a single informed trader. The informed trader knows $\tilde{v}$ from time 0, when the asset is traded at P_0. She submits trade orders $\mathrm{d}X_t$ over the interval $[0, T]$ and accumulates the position X_t in the risky asset. Her trades combine with noise trades $\mathrm{d}Z_t$ which follow a Brownian motion with volatility σ. A market maker sees the combined order $\mathrm{d}Y_t = \mathrm{d}X_t + \mathrm{d}Z_t$ but cannot separate $\mathrm{d}X_t$ from $\mathrm{d}Z_t$. Based on Y_t, he assigns a price $P_t = H(t, Y_t)$ to the asset.

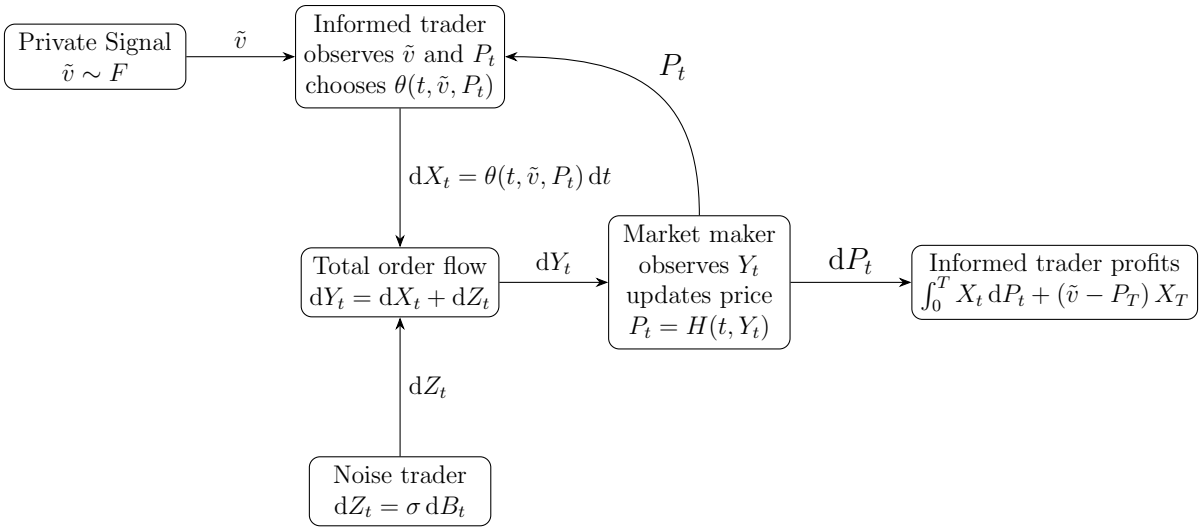


Figure 1.1: Flow chart describing the classical Kyle-Back model.

The informed trader's goal is to maximize her profit. The market maker's goal is to accurately estimate $\tilde{v}$.

1.1.1 Information Asymmetry

We model $\tilde{v}$ as a random variable whose realized value is known to the informed trader before it is announced to anyone else. Thus, information asymmetry reduces to a comparison between two filtrations, namely $\mathscr{F}_t^Y$ and $\mathscr{F}_t^Y \vee \sigma(\tilde{v})$.

The idea is that the informed trader has strictly superior information to the market maker, so the market maker's filtration is a (strict) subset of the informed trader's filtration at each $t \in [0, T]$. The market maker can only observe the aggregate trades $\{Y_t\}$, and so his filtration is $\mathscr{F}_t^Y$. On the other hand, the informed trader observes the terminal price $\tilde{v}$ in addition to the market price $P_t = H(t, Y_t)$. If we assume $H(t, \cdot)$ is invertible at each t, then the informed trader can deduce Y_t from $H(t, Y_t)$, and her information is $\mathscr{F}_t^Y \vee \sigma(\tilde{v})$.

We summarize below the relevant quantities and the extent to which they are observable by market participants.

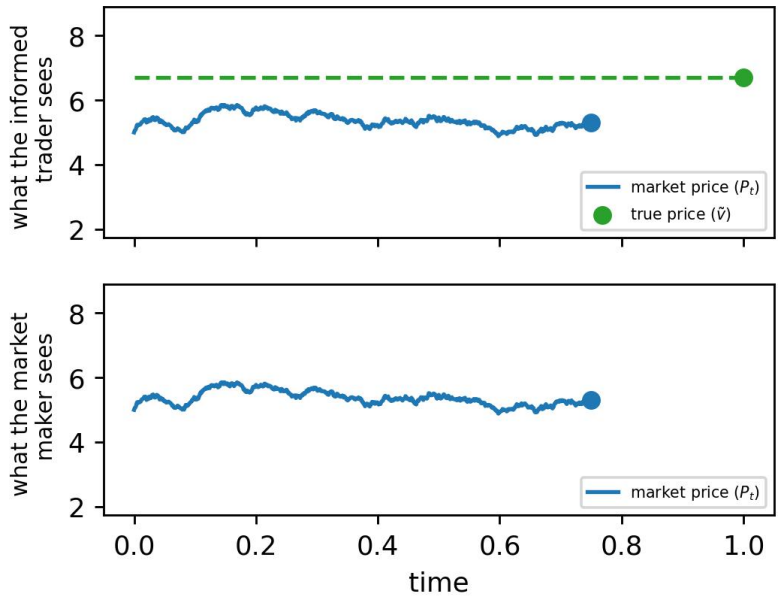


Figure 1.2: Information asymmetry in prices.

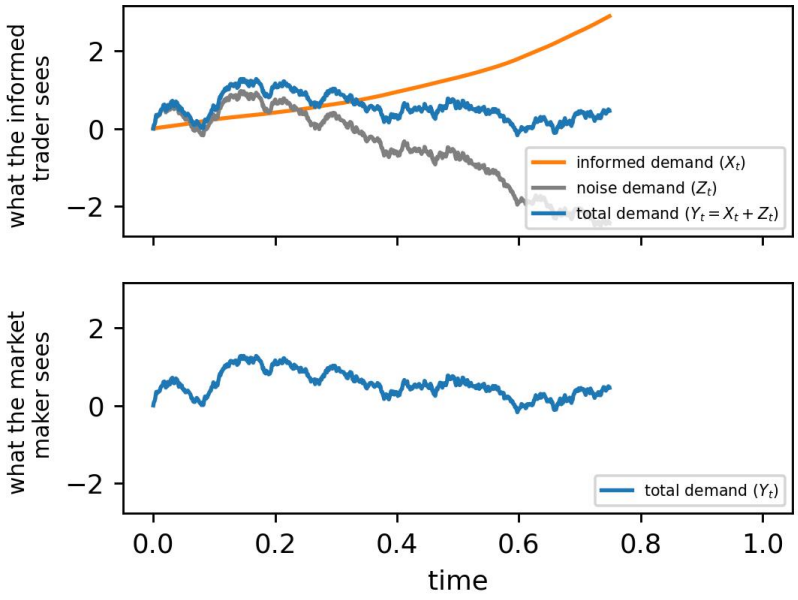


Figure 1.3: Information asymmetry in trades.

symbol	meaning	relation(s)	informed trader	market maker
$\tilde{v}$	terminal price		yes	no
$\mathcal{L}(\tilde{v})$	law of $\tilde{v}$		yes	yes
X_t	informed demand		yes	no
Z_t	noise demand		yes	no
$\mathcal{L}(Z_t)$	law of $Z_t \sim \mathcal{N}(0, \sigma^2 t)$		yes	no
Y_t	aggregate demand	$Y_t = X_t + Z_t$	yes	yes
$\theta(t, v, y)$	informed strategy	$\dot{X} = \theta$	yes	yes
$H(t, y)$	pricing rule		yes	yes
P_t	price process	$P_t = H(t, Y_t)$	yes	yes

Table 1.1: Summary of symbols in classical Kyle-Back model.

Although the market maker cannot pinpoint the informed trader's trades, we assume he knows her trading strategy.

1.1.2 Definitions and Assumptions

Definition 1.1.1. We call $H\colon [0,T] \times \mathbb{R} \mapsto \mathbb{R}$ a *pricing rule* if

(i) $H(t, \cdot)$ is strictly monotone increasing for each $t \in [0, T]$

(ii) $H \in C^{1,2}$

(iii) $\mathbb{E}\Big[\int_0^T H(t, Z_t)^2 \, \mathrm{d}t\Big] < \infty$

(iv) $\mathbb{E}\big[H(T, Z_T)^2\big] < \infty$

Item (i) ensures invertibility. Its economic interpretation is that higher demand reflects good news for the risky asset. Item (ii) is a regularity assumption required for H to satisfy the heat equation, whose discussion we defer to corollary 1.2.9. Item (iii) and item (iv) prevent doubling strategies.

Definition 1.1.2. For a given pricing rule H, we say that the $\big\{\mathscr{F}_t^Y \vee \sigma(\tilde{v})\big\}$ semi-martingale X is an admissible trading strategy if

$$\mathbb{E}\left[\int_0^T H(t, X_t + Z_t)^2 \, \mathrm{d}t\right] < \infty. \tag{1.1}$$

The constraint rules out doubling strategies.

Assumption 1.1.3. The risk free interest rate is 0.

Proposition 1.1.4. *For a fixed pricing rule H and a fixed strategy $\{X_t\}$, the informed trader profits*

$$V_T = \int_0^T (\tilde{v} - P_t)\, \mathrm{d}X_t - \langle X, P \rangle_T \tag{1.2}$$

where $P_t = H(t, X_t + Z_t)$.

Proof. Denote by $\{\beta_t\}$ the value of a riskless bond and $\{V_t\}$ the value of the informed trader's portfolio. We have

$$V_t = (V_t - X_t P_t)\, \beta_t + X_t P_t.$$

If this portfolio is self-financing, then

$$\mathrm{d}V_t = (V_t - X_t P_t)\, \mathrm{d}\beta_t + X_t\, \mathrm{d}P_t.$$

By assumption 1.1.3, $\mathrm{d}\beta_t = 0$. Integrating what remains in the above from 0 to T, we obtain

$$\lim_{t \uparrow T} V_t = V_0 + \int_0^T X_t\, \mathrm{d}P_t.$$

The limit addresses the possibility of a jump at time T. If the final market price P_T differs from $\tilde{v}$, the informed trader's wealth will jump by $(\tilde{v} - P_T)\, X_T$. Thus, at time T, the informed trader will profit

$$V_T = V_0 + (\tilde{v} - P_T)\, X_T + \int_0^T X_t\, \mathrm{d}P_t. \tag{1.3}$$

If the informed trader is risk neutral, we may take $V_0 = 0$ without loss of generality. An integration by parts re-expresses (1.3) as (1.2). □

Assumption 1.1.5. $\tilde{v}$ is such that $\mathbb{E}\left[\tilde{v}^2\right] < \infty$

Definition 1.1.6. Fix a pricing rule H. We say that a trading strategy $\{X_t^*\}$ is *optimal* for H if it maximizes the informed trader's expected profits:

$$\begin{aligned} &\mathbb{E}\left[\int_0^T \left(\tilde{v} - H(t, X_t^* + Z_t)\right) \mathrm{d}X_t^* - \langle X^*, H(\cdot, X_\cdot^* + Z_\cdot)\rangle_T\right] \\ &\qquad \geq \mathbb{E}\left[\int_0^T \left(\tilde{v} - H(t, X_t + Z_t)\right) \mathrm{d}X_t - \langle X, H(\cdot, X_\cdot + Z_\cdot)\rangle_T\right] \end{aligned} \tag{1.4}$$

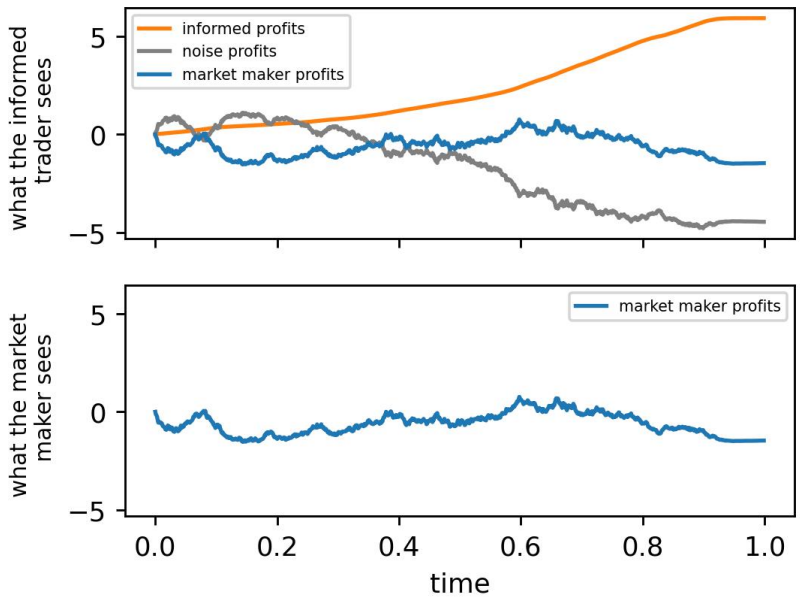


Figure 1.4: PnL of market participants. Note that the market maker tends to break even, and the informed trader profits at the expense of noise traders.

Now, fix a strategy $\{X_t\}$. We say that a pricing rule H^* is *rational* for $\{X_t\}$ if

$$H^*(t, X_t + Z_t) = \mathbb{E}\left[\tilde{v} \middle| \mathscr{F}_t^{X+Z}\right]. \tag{1.5}$$

We call a pair (H^*, X^*) an *equilibrium* if $\{X_t^*\}$ is optimal for H^* and H^* is *rational* for $\{X_t^*\}$.

Remark 1.1.7. We could have considered path dependent pricing rules of the form $H\big(t, Y_{[0,t]}\big)$, in which case the rationality condition (1.5) would read

$$H^*\big(t, Y_{[0,t]}\big) = \mathbb{E}\left[\tilde{v} \middle| \mathscr{F}_t^{Y^*}\right],$$

but [Cho03] shows that this would lead to the same equilibrium.

1.2 The Equilibrium and its Properties

1.2.1 The Kantorovich problem

Let F denote the CDF of $\tilde{v}$ and N the CDF of $\mathcal{N}\big(0, \sigma^2 T\big)$. We consider the Kantorovich problem

$$\max_{\Phi} \iint_{\mathbb{R}^2} vy \, \Phi(\mathrm{d}v, \mathrm{d}y) \tag{1.6}$$

where the maximum is over all probability measures Φ on $\mathscr{B}(\mathbb{R}^2)$ with marginals F and N, i.e.

$$\Phi(\mathrm{d}v, \mathbb{R}) = F(\mathrm{d}v) \qquad \Phi(\mathbb{R}, \mathrm{d}y) = N(\mathrm{d}y).$$

Note the abuse of notation; F and N as they appear above are Borel measures.

The Kantorovich problem admits the dual formulation

$$\min_{f,g} \left(\int_{\mathbb{R}} f(v) F(\mathrm{d}v) + \int_{\mathbb{R}} g(y) N(\mathrm{d}y) \right) \tag{1.7}$$

where the minimum is over all bounded and continuous functions f, g satisfying

$$f(v) + g(y) \geq vy.$$

See, for example, [San15, section 1.2].

The minimum of (1.7) is achieved at some Legendre dual pair $(f, g) = (\varphi^*, \varphi)$. The optimal φ is called a *Brenier potential*, and it is unique up to an additive constant [San15, theorem 1.40]. We pick in particular the φ which satisfies

$$\mathbb{E}\big[\varphi(Z_T)\big] = 0. \tag{1.8}$$

Since N is absolutely continuous, the function $h(y) = \varphi'(y)$ is defined everywhere. When F is absolutely continuous as well, h admits the representation $h = F^{-1} \circ N$ and is invertible.

1.2.2 Main Theorem

Theorem 1.2.1. *Suppose $\mathcal{L}(\tilde{v})$ is absolutely continuous and $\mathbb{E}\big[\tilde{v}^2\big] < \infty$. Let φ denote the Brenier potential from $\mathcal{L}(Z_T) \sim \mathcal{N}(0, \sigma^2 T)$ to $\mathcal{L}(\tilde{v})$. Then, the pricing rule*

$$H^*(t, y) = \mathbb{E}\big[\varphi'(Z_T) \big| Z_t = y\big] \tag{1.9}$$

and the strategy

$$X_t^* = (T - t) \int_0^t \frac{(\varphi')^{-1}(\tilde{v}) - Z_s}{(T - s)^2} \, \mathrm{d}s \tag{1.10}$$

form an equilibrium.

Remark 1.2.2. Some authors characterize the strategy (1.10) equivalently by its ODE

$$\mathrm{d}X_t^* = \frac{(\varphi')^{-1}(\tilde{v}) - Y_t^*}{T - t} \, \mathrm{d}t = \frac{(\varphi')^{-1}(\tilde{v}) - X_t^* - Z_t}{T - t} \, \mathrm{d}t$$

with initial condition $X_0^* = 0$. The pricing rule (1.9) may also be written explicitly as the integral

$$H^*(t,y) = \int_{\mathbb{R}} \varphi'(z) \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \exp\left(-\frac{(z-y)^2}{2\sigma^2(T-t)}\right) \mathrm{d}z.$$

Remark 1.2.3. φ' is called the *Brenier map* from $\mathcal{L}(Y_T)$ to $\mathcal{L}(\tilde{v})$, and it is so called because under the optimal coupling, $\tilde{v}$ and Y_T are related by $\tilde{v} = \varphi'(Y_T)$. In particular, φ' a one-to-one map which relates $\tilde{v}$ to the aggregate demand Y_T.

Example 1.2.4. If $\tilde{v} \sim \mathcal{N}(p_0, \Sigma_0)$, then $\varphi'(y) = p_0 + \frac{\sqrt{\Sigma_0}}{\sigma} y$. Taking $T = 1$ for simplicity, the pricing rule (1.9) becomes

$$H^*(t,y) = p_0 + \frac{\sqrt{\Sigma_0}}{\sigma} y$$

and the strategy (1.10) becomes

$$X_t^* = (1-t) \int_0^t \frac{\frac{\sigma}{\sqrt{\Sigma_0}} (\tilde{v} - p_0) - Z_s}{1-s} \, \mathrm{d}s. \tag{1.11}$$

This is the foundational equilibrium from [Kyl85]. The ratio

$$\lambda = \frac{\sqrt{\Sigma_0}}{\sigma}$$

is called *Kyle's lambda* and measures price sensitivity to order flow. We shall later make this notion precise with definition 1.2.11. [Kyl85] showed that when $\tilde{v}$ is Gaussian, prices move linearly in proportion to order flow.

The inverse, $1/\lambda$, is called *market depth*. Writing (1.11) in its ODE form

$$\mathrm{d}X_t^* = \frac{1}{\lambda} \frac{\tilde{v} - \overbrace{H(t, X_t^* + Z_t)}^{P_t^*}}{1-t} \, \mathrm{d}t,$$

we see that the informed trader trades in proportion to the price mismatch $\tilde{v} - P_t^*$ with the weighting $\frac{1}{\lambda(1-t)}$.

Example 1.2.5. If $\log \tilde{v} \sim \mathcal{N}(\alpha, \phi^2)$, then $\varphi'(y) = \exp\Big(\alpha + \frac{\phi}{\sigma} y\Big)$. The pricing rule (1.9) becomes

$$H^*(t,y) = \exp\left(\alpha + \frac{\phi}{\sigma} y + \frac{\phi^2}{2}(1-t)\right) \tag{1.12}$$

and the trading strategy (1.10) becomes

$$X_t^* = (1-t)\int_0^t \frac{\frac{\sigma}{\phi}(\log\tilde{v}-\alpha) - Z_s}{(1-s)^2}\,\mathrm{d}s.$$

This equilibrium was first presented by [Bac92].

We defer the proof of theorem 1.2.1 to a later section. First, we state a few more properties of the equilibrium.

1.2.3 Properties of the Equilibrium

Proposition 1.2.6. *Conditioned on Y_t^*, $(\varphi')^{-1}(\tilde{v})$ is Gaussian with mean Y_t^* and variance $\sigma^2(T-t)$.*

Proof. If we substitute $\mathrm{d}X_t^* = \frac{(\varphi')^{-1}(\tilde{v})-Y_t^*}{T-t}\,\mathrm{d}t$ into $\mathrm{d}Y_t^* = \mathrm{d}X_t^* + \mathrm{d}Z_t$, we obtain

$$\mathrm{d}Y_t^* = \frac{(\varphi')^{-1}(\tilde{v}) - Y_t^*}{T-t}\,\mathrm{d}t + \mathrm{d}Z_t, \tag{1.13}$$

which is an Ornstein Uhlenbeck process admitting the closed form solution

$$Y_t^* = (\varphi')^{-1}(\tilde{v})\frac{t}{T} + (T-t)\int_0^t \frac{1}{T-s}\,\mathrm{d}Z_s.$$

We see that the pair $(Y_t^*, (\varphi')^{-1}(\tilde{v}))$ is jointly Gaussian for each t. By the Gaussian projection theorem (see theorem A.1.1), $(\varphi')^{-1}(\tilde{v})$ conditioned on Y_t^* is Gaussian with mean

$$\mathbb{E}\big[(\varphi')^{-1}(\tilde{v})\big|Y_t^*\big] = \mathbb{E}\big[(\varphi')^{-1}(\tilde{v})\big] + \frac{\mathrm{Cov}\big(Y_t^*, (\varphi')^{-1}(\tilde{v})\big)}{\mathrm{Var}(Y_t^*)}\big(Y_t^* - \mathbb{E}[Y_t^*]\big) = Y_t^*$$

and variance

$$\mathrm{Var}\big((\varphi')^{-1}(\tilde{v})\big|Y_t^*\big) = \mathrm{Var}\big((\varphi')^{-1}(\tilde{v})\big) - \frac{\mathrm{Cov}\big(Y_t^*, (\varphi')^{-1}(\tilde{v})\big)^2}{\mathrm{Var}(Y_t^*)} = \sigma^2(T-t).$$

□

Corollary 1.2.7. *The net orders $\{Y_t^*\}$ form a Brownian bridge connecting $(0,0)$ to $(T, (\varphi')^{-1}(\tilde{v}))$. In particular, since $(\varphi')^{-1}(\tilde{v}) \sim \mathcal{N}(0,\sigma^2 T)$, $\{Y_t^*\}$ is a Brownian motion in its own filtration.*

This means that $\{Y_t^*\}$ appears as a Brownian bridge to the informed trader, but a Brownian motion to the market maker. In market microstructural terms, $\{Y_t^*\}$ looks as natural

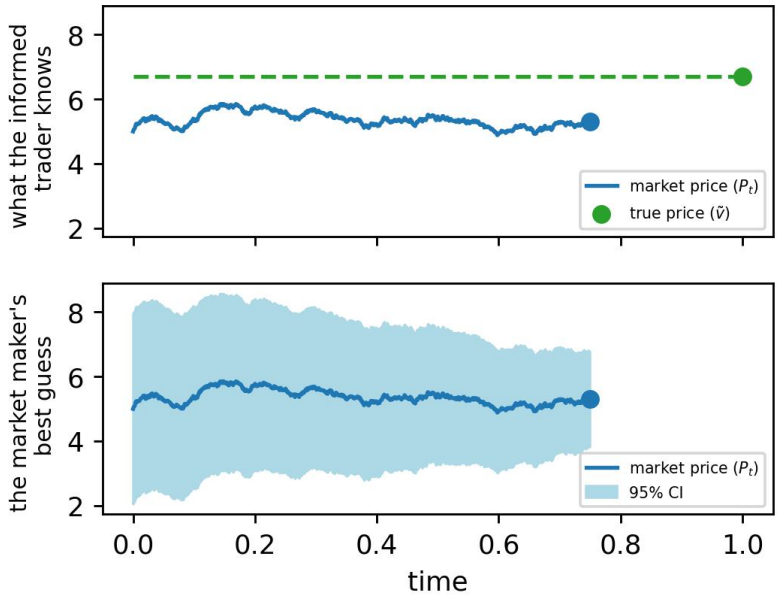


Figure 1.5: The market maker's filtering problem. Numerical simulation with $\tilde{v} \sim \mathcal{N}(5,1)$ and $\sigma = 3/2$.

to the market marker as the liquidity trades $\{Z_t\}$, so the informed trades are undetectable. This is the *inconspicuous trading property* first coined by [Cho03].

Proof. Note that (1.13) is the SDE of a Brownian bridge pinned at $(0,0)$ and $(T,(\varphi')^{-1}(\tilde{v}))$. To show that $\{Y_t^*\}$ is a Brownian motion in its own filtration, invoke [LS77, theorem 7.12] to write

$$Y_t^* = \int_0^t \mathbb{E}\left[\frac{(\varphi')^{-1}(\tilde{v}) - Y_s^*}{T-s}\middle|\mathscr{F}_s^{Y^*}\right]\mathrm{d}s + Z_t^{Y^*}$$

where $\left\{Z_t^{Y^*}\right\}$ is an $\{\mathscr{F}_t^{Y^*}\}$ Brownian motion. Next, compute

$$\mathbb{E}\left[(\varphi')^{-1}(\tilde{v})\middle|\mathscr{F}_s^{Y^*}\right] = \mathbb{E}\left[Y_T^*\middle|\mathscr{F}_s^{Y^*}\right] = \mathbb{E}[Y_T^*|Y_s^*] = \mathbb{E}\left[(\varphi')^{-1}(\tilde{v})\middle|Y_s^*\right] = Y_s^*,$$

so that the $\mathrm{d}s$ term vanishes, and we are left with $Y_t^* = Z_t^{Y^*}$. □

Corollary 1.2.8. *The pricing rule* (1.9) *is rational in the sense of* (1.5).

Proof. Since $\{Z_t\}$ and $\{Y_t^*\}$ are both Brownian motions in their natural filtrations,

$$H(t,Y_t^*) = \mathbb{E}\left[\varphi'(Z_T)\middle|Z_t = Y_t^*\right] = \mathbb{E}\left[\varphi'(Y_T^*)\middle|Y_t^*\right] = \mathbb{E}\left[\varphi'(Y_T^*)\middle|\mathscr{F}_t^Y\right]$$

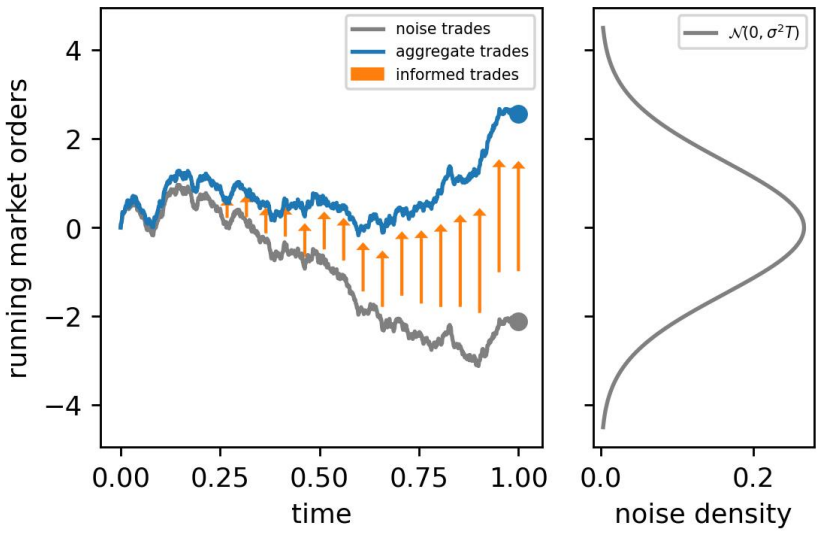


Figure 1.6: The inconspicuous trading property. The noise trades and the aggregate trades are both Brownian motions in their own filtrations.

Since $Y_T^* = (\varphi')^{-1}(\tilde{v})$ almost surely,

$$\mathbb{E}\Big[\varphi'(Y_T^*)\Big|\mathscr{F}_t^Y\Big] = \mathbb{E}\Big[\tilde{v}\Big|\mathscr{F}_t^Y\Big].$$

□

Corollary 1.2.9. *The price process* $\{P_t^*\} = \{H(t, Y_t^*)\}$ *is an* $\big\{\mathscr{F}_t^{Y^*}\big\}$ *martingale.*

This means that $\{P_t^*\}$ appears as a "fair game" to the market maker and noise traders: the public information $\big\{\mathscr{F}_t^{Y^*}\big\}$ does not reveal any directional bias. In contrast, the enlarged filtration $\big\{\mathscr{F}_t^{Y^*} \vee \sigma(\tilde{v})\big\}$ *does* reveal such a bias, because the informed trader knows in advance that $P_T^* = \tilde{v}$. Thus, $\{P_t^*\}$ is a martingale for the market maker and noise traders, but not for the informed trader.

Proof. By virtue of corollary 1.2.8, we have $P_t = H(t, Y_t^*) = \mathbb{E}\big[\tilde{v}\big|\mathscr{F}_t^{Y^*}\big]$. Fix $s < t$. Invoking the Markov property of $\{Y_t^*\}$,

$$\mathbb{E}\Big[P_t\Big|\mathscr{F}_s^{Y^*}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\tilde{v}\Big|\mathscr{F}_t^{Y^*}\Big]\Big|\mathscr{F}_s^{Y^*}\Big] = \mathbb{E}\Big[\tilde{v}\Big|\mathscr{F}_s^{Y^*}\Big] = P_s.$$

From $\mathbb{E}\big[\tilde{v}^2\big] < \infty$, we immediately obtain integrability:

$$\mathbb{E}\big[|P_t|\big] = \mathbb{E}\left[\left|\mathbb{E}\Big[\tilde{v}\Big|\mathscr{F}_t^{Y^*}\Big]\right|\right] \leq \mathbb{E}\Big[\mathbb{E}\Big[|\tilde{v}|\Big|\mathscr{F}_t^{Y^*}\Big]\Big] = \mathbb{E}\big[|\tilde{v}|\big] \leq \sqrt{\mathbb{E}[\tilde{v}^2]} < \infty.$$

□

From martingality we immediately obtain

Corollary 1.2.10. *H satisfies the heat equation*

$$\frac{\partial H}{\partial t} + \frac{\partial^2 H}{\partial y^2} = 0 \tag{1.14}$$

with terminal condition $H(T, y) = \varphi'(y)$.

Definition 1.2.11. Kyle's lambda is the function $\lambda\colon \mathbb{R}^+ \times \mathbb{R} \mapsto \mathbb{R}$ satisfying

$$\mathrm{d}P_t = \lambda(t, Y_t)\,\mathrm{d}Y_t. \tag{1.15}$$

Remark 1.2.12. Kyle's lambda quantifies price sensitivity to order flow. In example 1.2.4, we saw that $\lambda(t, y)$ is constant when $\tilde{v}$ is Gaussian. If $\tilde{v}$ is lognormal as in example 1.2.5, then

$$\lambda(t, y) = \frac{\phi}{\sigma} \exp\left(\alpha + \frac{\phi}{\sigma} y + \frac{\phi^2}{2}(1 - t)\right),$$

which we obtain from taking the y derivative of (1.12).

Proposition 1.2.13. *Let φ denote the Brenier potential from $\mathcal{L}(Z_T) = \mathcal{N}(0, \sigma^2 T)$ to $\mathcal{L}(\tilde{v})$. Then, Kyle's lambda is*

$$\lambda(t, y) = \mathbb{E}\left[\varphi''(Z_T) \middle| Z_t = y\right].$$

Proof. We apply Itô's lemma to

$$P_t = H(t, Y_t)$$

and we obtain

$$\mathrm{d}P_t = \left(\frac{\partial H}{\partial t}(t, Y_t) + \frac{\partial^2 H}{\partial y^2}(t, Y_t)\right)\mathrm{d}t + \frac{\partial H}{\partial y}(t, Y_t)\,\mathrm{d}Y_t.$$

Due to corollary 1.2.10, the $\mathrm{d}t$ term above vanishes, and we see that

$$\lambda(t, Y_t) = \frac{\partial H}{\partial y}(t, Y_t) = \frac{\partial}{\partial y}\,\mathbb{E}\left[\varphi'(Z_T) \middle| Z_t = Y_t\right] = \mathbb{E}\left[\varphi''(Z_T) \middle| Z_t = Y_t\right].$$

□

Corollary 1.2.14. *The pricing rule (1.2.1) is strictly increasing in y.*

Proof. Since φ is convex, $\varphi'' > 0$, and the inequality is strict because $\mathcal{L}(\tilde{v})$ is atomless [GS98]. Thus φ' is strictly increasing. Fix $t \in [0,T]$ and $y_1 < y_2$. We have

$$\begin{aligned} H(t,y_1) &= \mathbb{E}\big[\varphi'(Z_T)\big|Z_t = y_1\big] \\ &= \mathbb{E}\big[\varphi'(Z_T - Z_t + y_1)\big] < \mathbb{E}\big[\varphi'(Z_T - Z_t + y_2)\big] = H(t,y_2). \end{aligned}$$

□

1.2.4 The Wealth Decomposition

Theorem 1.2.15. *Let φ denote the Brenier potential from $\mathcal{L}(Z_T) = \mathcal{N}(0,\sigma^2 T)$ to $\mathcal{L}(\tilde{v})$. Normalize $\mathbb{E}\big[\varphi(Z_T)\big] = 0$. Denote*

$$\Gamma(t,y) = \mathbb{E}\big[\varphi(Z_T)\big|Z_t = y\big]. \tag{1.16}$$

Fix a trading strategy $\{X_t\}$ which is not necessarily (1.10)*. If the market maker uses the pricing rule* (1.9)*, then the informed trader's expected profits from $\{X_t\}$ admits the representation*

$$\mathbb{E}\left[\int_0^T (\tilde{v} - P_t)\,\mathrm{d}X_t - \langle X, P\rangle_T\right] = \mathbb{E}\left[\tilde{v}Y_T - \varphi(Y_T) - \frac{1}{2}\int_0^T \frac{\partial^2\Gamma}{\partial y^2}(t,Y_t)\,\mathrm{d}\langle X\rangle_t\right] \tag{1.17}$$

which we call the wealth decomposition.

Proof. First note that

$$\frac{\partial\Gamma}{\partial y}(t,Y_t) = \mathbb{E}\big[\varphi'(Z_T)\big|Z_t = Y_t\big] = \mathbb{E}\big[h(Z_T)\big|Z_t = Y_t\big] = H^*(t,Y_t) = P_t.$$

By Itô's lemma, we have

$$\mathrm{d}\Gamma(t,Y_t) = \frac{\partial\Gamma}{\partial t}(t,Y_t)\,\mathrm{d}t + \underbrace{\frac{\partial\Gamma}{\partial y}(t,Y_t)}_{P_t}\mathrm{d}Y_t + \frac{1}{2}\frac{\partial^2\Gamma}{\partial y^2}(t,Y_t)\,\mathrm{d}\langle Y\rangle_t. \tag{1.18}$$

Since $Y = X + Z$,

$$\begin{aligned} \langle Y\rangle_t = \langle X+Z\rangle_t &= \langle X\rangle_t + 2\langle X,Z\rangle_t + \langle Z\rangle_t \\ &= \langle X\rangle_t + 2\langle X, Y - X\rangle_t + \langle Z\rangle_t = \langle Z\rangle_t + 2\langle X,Y\rangle - \langle X\rangle_t. \end{aligned}$$

If $Z_t = \sigma W_t$, then $\langle Z \rangle_t = \sigma^2 t$. Putting these into (1.18), we obtain

$$\begin{aligned} \mathrm{d}\Gamma(t, Y_t) = &\left(\frac{\partial \Gamma}{\partial t}(t, Y_t) + \frac{\sigma^2}{2} \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \right) \mathrm{d}t + P_t \, \mathrm{d}Y_t \\ &+ \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X, Y \rangle_t - \frac{1}{2} \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X \rangle_t. \end{aligned} \tag{1.19}$$

Since $\Gamma(t, y)$ is defined by a conditional expectation, it solves the backwards heat equation. Thus, the $\mathrm{d}t$ term in (1.19) vanishes.

We also have from $P_t = \frac{\partial \Gamma}{\partial y}(t, Y_t)$ that

$$\mathrm{d}P_t = \frac{\partial^2 \Gamma}{\partial t \partial y}(t, Y_t) \, \mathrm{d}t + \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}Y_t + \frac{1}{2} \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle Y \rangle_t, \tag{1.20}$$

and so

$$\mathrm{d}\langle X, P \rangle_t = \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X, Y \rangle_t. \tag{1.21}$$

Updating (1.19) once again, we find

$$\mathrm{d}\Gamma(t, Y_t) = P_t \, \mathrm{d}Y_t + \mathrm{d}\langle X, P \rangle_t - \frac{1}{2} \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X \rangle_t. \tag{1.22}$$

Integration from 0 to T produces

$$\Gamma(T, Y_T) = \Gamma(0, 0) + \int_0^T P_t \, \mathrm{d}Y_t + \langle X, P \rangle_T - \frac{1}{2} \int_0^T \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X \rangle_t \tag{1.23}$$

As in [BCEL20], we normalize $\Gamma(0, 0) = 0$. Rearranging and combining with $\int_0^T \tilde{v} \, \mathrm{d}Y_t = \tilde{v} Y_T$, we find

$$\int_0^T (\tilde{v} - P_t) \, \mathrm{d}Y_t - \langle X, P \rangle_T = \tilde{v} Y_T - \Gamma(T, Y_T) - \frac{1}{2} \int_0^T \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X \rangle_t. \tag{1.24}$$

Finally, we apply expectations to obtain

$$\mathbb{E}\left[\int_0^T (\tilde{v} - P_t) \, \mathrm{d}Y_t - \langle X, P \rangle_T \right] = \mathbb{E}\left[\tilde{v} Y_T - \Gamma(T, Y_T) - \frac{1}{2} \int_0^T \frac{\partial^2 \Gamma}{\partial y^2}(t, Y_t) \, \mathrm{d}\langle X \rangle_t \right]. \tag{1.25}$$

Noting that

$$\mathbb{E}\left[\int_0^T (\tilde{v} - P_t) \, \mathrm{d}Z_t \right] = 0, \tag{1.26}$$

we see that (1.25) coincides with (1.17). □

Corollary 1.2.16. *Let φ be as in theorem 1.2.15. Any strategy is optimal which*

1. *ensures $Y_T = (\varphi')^{-1}(\tilde{v})$ almost surely*
2. *is absolutely continuous with respect to Lebesgue measure.*

The informed trader profits at most $\psi(\tilde{v})$, where ψ denotes the Legendre transform of φ.

Proof. Consider the RHS of (1.17). For fixed $\tilde{v}$, the quantity $\tilde{v}Y_T - \varphi(Y_T)$ is at most

$$\psi(\tilde{v}) = \sup_y \left\{\tilde{v}y - \varphi(y)\right\} \tag{1.27}$$

and achieves this value when y satisfies $\varphi'(y) = \tilde{v}$. The informed trader should therefore force Y_T to satisfy $\varphi'(Y_T) = \tilde{v}$ by driving the order flow. Additionally, the term $-\frac{1}{2}\int_0^T \frac{\partial^2\Gamma}{\partial y^2}(t, Y_t)\,\mathrm{d}\langle X\rangle_t$ is nonpositive but at most 0 when $\{X_t\}$ is absolutely continuous. □

Corollary 1.2.17. *On the event $\{\tilde{v} = v, Y_t = y\}$, the informed trader profits at most*

$$\psi(v) + \Gamma(t, y) - vy. \tag{1.28}$$

Proof. Integrate (1.22) over $[t, T]$ instead of $[0, T]$. Then, (1.24) becomes

$$\begin{aligned}\int_t^T (\tilde{v} - P_s)\,\mathrm{d}Y_s - \int_t^T \mathrm{d}\langle X, P\rangle_s \\ = \tilde{v}\,(Y_T - Y_t) - \varphi(Y_T) + \Gamma(t, Y_t) - \frac{1}{2}\int_t^T \frac{\partial^2\Gamma}{\partial y^2}\,(s, Y_s)\,\mathrm{d}\langle X\rangle_s.\end{aligned} \tag{1.29}$$

We apply $\mathbb{E}[\cdot|\tilde{v} = v, Y_t = y]$ and we obtain

$$\begin{aligned}&\mathbb{E}\left[\int_t^T (\tilde{v} - P_s)\,\mathrm{d}X_s - \int_t^T \mathrm{d}\langle X, P\rangle_s \middle| \tilde{v} = v, Y_t = y\right] \\ &\quad = \mathbb{E}\left[\tilde{v}\,(Y_T - Y_t) - \varphi(Y_T) + \Gamma(t, Y_t) - \frac{1}{2}\int_t^T \frac{\partial^2\Gamma}{\partial y^2}\,(s, Y_s)\,\mathrm{d}\langle X\rangle_s \middle| \tilde{v} = v, Y_t = y\right].\end{aligned} \tag{1.30}$$

If $\{X_t\}$ is as in corollary 1.2.16, the RHS is at most $\psi(v) + \Gamma(t, y) - vy$, which is (1.28). □

1.2.5 Proof of Main Theorem

Proof. Take H and $\{X_t\}$ as they appear in theorem 1.2.1.

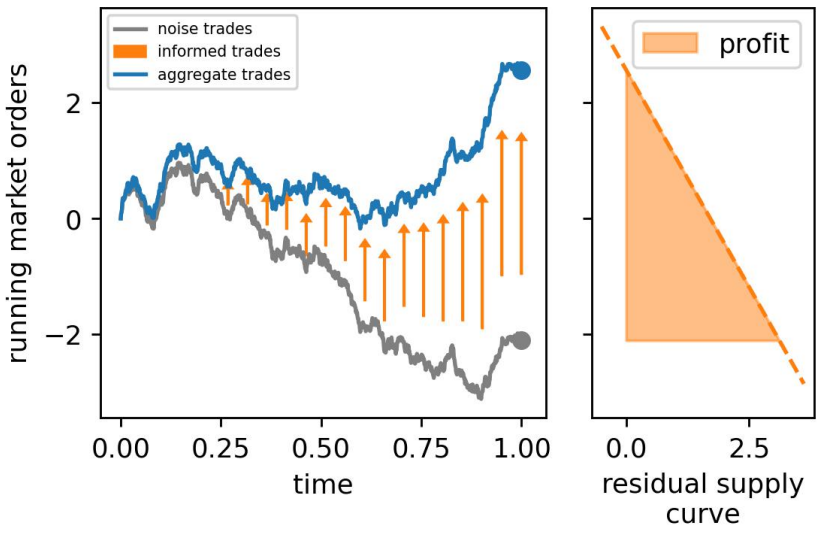


Figure 1.7: The informed trader's profit is the area under the residual supply curve.

Corollary 1.2.8 shows that H is rational. It remains to verify the properties in definition 1.1.1. Item (i) follows from corollary 1.2.14. Item (ii) follows from corollary 1.2.10. Item (iii) and item (iv) follow immediately from assumption 1.1.3.

To show that the strategy is admissible in the sense of definition 1.1.2, we just note that eq. (1.1) is precisely item (iii) due to the inconspicuous trading property discussed in corollary 1.2.7.

□

1.3 Conclusion

The equilibrium constructed in this chapter shows that the classical Kyle–Back model can be recast as an optimal transport problem between the hidden asset price and liquidity trades. The Brenier potential φ and its gradient $\nabla\varphi = \varphi'$ play a dual role: analytically, they solve the Monge–Kantorovich problem (1.6)–(1.7), while economically, they encode the pricing rule $H^*(t, y)$ and the insider's optimal strategy X^*. This perspective transforms what is usually framed as a stochastic control problem into a deterministic optimization over probability measures, revealing that Kyle's equilibrium is, at its core, a comonotone coupling.

Several key properties fall out transparently from this formulation. First, inconspicuous trading arises naturally: under the equilibrium strategy, the total order flow Y_t remains a Brownian motion in its own filtration, so informed trading is statistically indistinguishable from noise. Second, the pricing rule inherits the convexity of the Brenier potential, which

guarantees monotonicity and thus the rational-expectations property $P_t = \mathbb{E}\big[\tilde{v} \mid \mathscr{F}_t^Y\big]$. Finally, the decomposition of the informed trader's wealth in section 1.2.4 makes explicit that expected profits are given by the convex conjugate $\varphi^*(\tilde{v})$, yielding a clean geometric expression.

Conceptually, this chapter demonstrates that optimal transport offers not only an alternative proof of existence and uniqueness for the continuous-time Kyle equilibrium, but also a unifying language relating price formation, information revelation, and liquidity. The tools developed here—Brenier potentials, Kantorovich duality, and the interpretation of $\nabla^2\varphi$ as Kyle's impact parameter λ—provide the mathematical backbone for the extensions studied in the next chapters. In particular, Chapter 2 generalizes this framework to *dynamic information and activist trading*, where the insider's private signal evolves over time.

CHAPTER 2

Dynamic Information and Activism Trading

Now that we've seen optimal transport applied to the original Kyle-Back model [Bac92], we ask: will it work for extensions of the Kyle-Back model? The answer turns out to be yes, with surprising generality. By swapping the cost function in the optimal transport problem, we can accommodate nonlinear liquidation payouts, such as the ones considered in [BCDF$^+$18]. Adding a new dimension to the source measure, we can accommodate random endowments. Finally, we introduce a methodology using BSPDE's and nonlinear filtering equations to construct trading strategies, and with this we can accommodate dynamic information models such as [BP98]. The result is a general methodology of constructing Kyle-Back equilibria which encompasses all known continuous Kyle-Back extensions the literature, to our knowledge. These results appear in our arXiv paper [CENV22].

2.1 The Model

There is a public announcement to be made at time T which will affect the price of n risky assets. Immediately after this announcement, the assets will be valued at $\tilde{v}$, possibly jumping in price. The distribution of $\tilde{v}$ is public information, but no one knows exactly what value $\tilde{v}$ will take. The most knowledgeable person is an informed trader who observes in real time a private diffusive signal $\{S_t\}$ which is related to $\tilde{v}$ by

$$f(S_T) = \tilde{v} \tag{2.1}$$

for some bijective function f. The signal starts at $S_0 \sim \mathcal{N}(0, \mathrm{Var}(S_0))$ and evolves according to $\mathrm{d}S_t = \Sigma_S^{1/2}(t)\,\mathrm{d}W_t$. The informed trader starts with a random stock endowment β. Over the time interval $[0, T]$, she submits trade orders and accumulates the position X_t in the risky asset. Her trades combine with noise trades $\mathrm{d}Z_t$ which follow a Brownian motion with covariance matrix $\Sigma_Z(t)$. A market maker sees the combined order $\mathrm{d}Y_t = \mathrm{d}X_t + \mathrm{d}Z_t$ but cannot separate $\mathrm{d}X_t$ from $\mathrm{d}Z_t$. Based on Y_t, he assigns a price $P_t = H(t, Y_t)$ to the

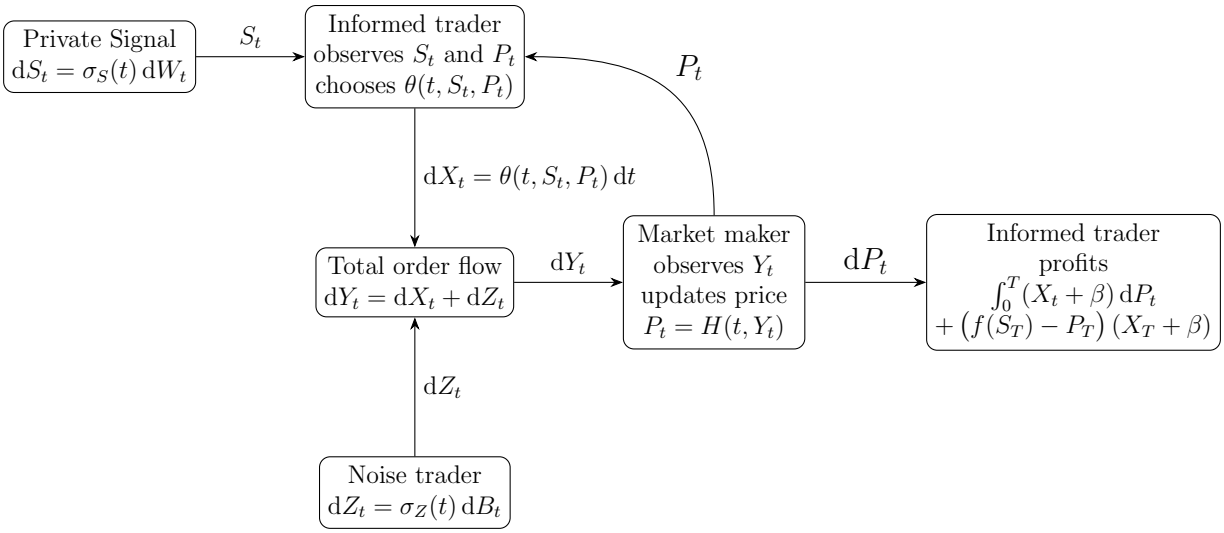


Figure 2.1: Flow chart describing the dynamic information model.

assets. At the end of the trading interval, the informed trader liquidates her position and receives the amount $V(\beta+X_T, S_T)$, where V is a possibly nonlinear function that generalizes $(\beta + X_T)\, f(S_T)$.

Remark 2.1.1. This model differs from the previous one in several aspects:

1. The informed trader does not see $\tilde{v}$ but learns about it gradually through $\{S_t\}$. This is the dynamic information component which incorporates [BP98].

2. There are n assets

3. The informed trader starts with an initial endowment β.

4. The informed trader's liquidation value is $V(\beta + X_T, S_T)$ instead of $(\beta + X_T)\, f(S_T)$. This captures *activism trading*, as discussed in [BCDF$^+$18]. The informed profits will be made precise with assumption 2.1.6.

5. The noise trades have time-varying (but deterministic) covariance matrix $\Sigma_Z(t)$:

$$\mathrm{d}Y_t = \mathrm{d}X_t + \underbrace{\Sigma_Z^{1/2}(t)\,\mathrm{d}B_t}_{\mathrm{d}Z_t}$$

We call this the *dynamic information model*, and the previous one the *static information model*.

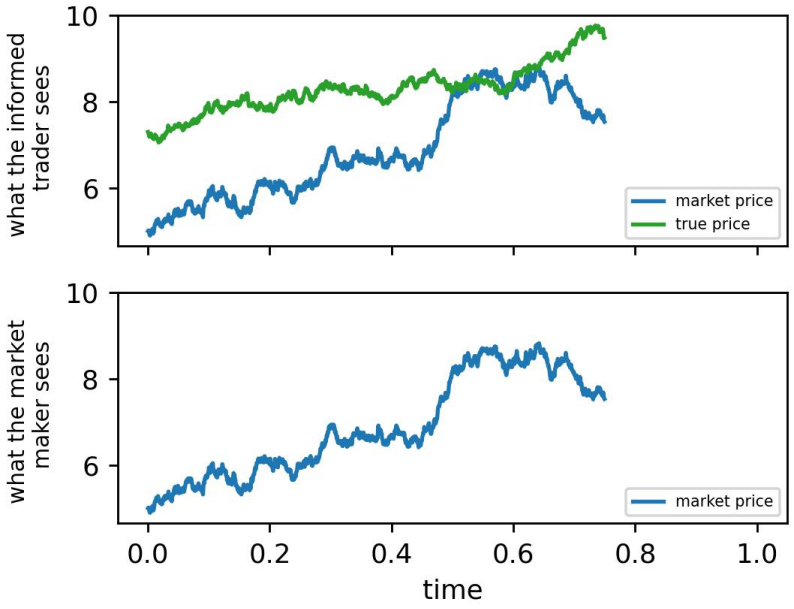


Figure 2.2: Prices in the dynamic setting.

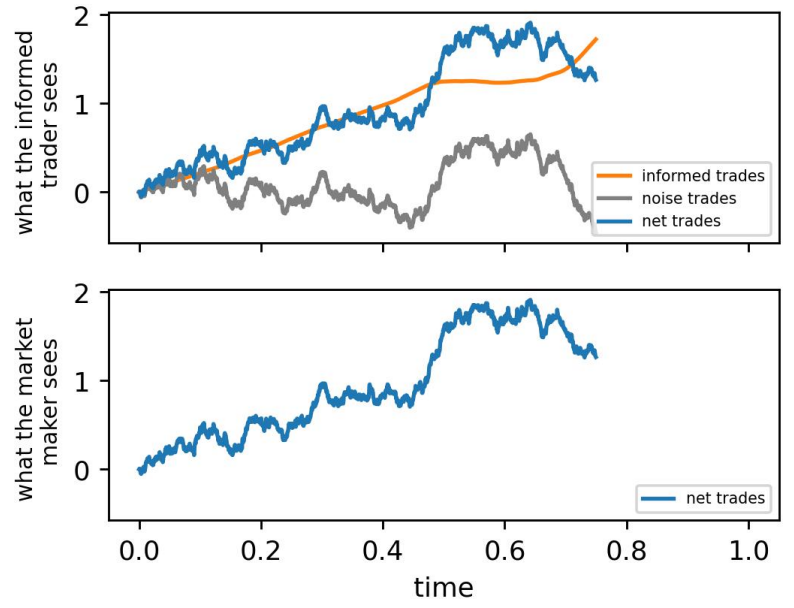


Figure 2.3: Trades in the dynamic setting.

Assumption 2.1.2. Normalizing $\{S_t\}$ as necessary,

$$\int_0^T \Sigma_Z(t)\,\mathrm{d}t = \mathrm{Var}(S_0) + \int_0^T \Sigma_S(t)\,\mathrm{d}t.$$

Assumption 2.1.3. f is strictly increasing.

2.1.1 Information Asymmetry

As before, information asymmetry is a comparison between two filtrations. The market maker's filtration is still $\mathscr{F}_t^Y$. The informed trader's filtration is now $\mathscr{F}_t^{Y,S}$.

The table from section 1.1.1 updates to

symbol	meaning	relation(s)	informed trader	market maker
$\tilde{v}$	terminal price		no	no
$\mathcal{L}(\tilde{v})$	law of $\tilde{v}$		yes	yes
S_t	private signal	$f(S_T) = \tilde{v}$	yes	no
$\mathcal{L}(S_t)$	law of S_t		yes	yes
X_t	informed demand		yes	no
$\theta(t, y, s)$	informed strategy	$\dot{X}_t = \theta(t, Y_t, S_t)$	yes	yes
Z_t	noise demand		yes	no
$\mathcal{L}(Z_t)$	law of Z_t	$Z_T \sim \mathcal{N}(0, \int_0^T \Sigma_Z(t)\,\mathrm{d}t)$	yes	yes
Y_t	aggregate demand	$Y_t = X_t + Z_t$	yes	yes
$H(t, y)$	pricing rule		yes	yes
P_t	price process	$P_t = H(t, Y_t)$	yes	yes
β	endowment		yes	no
$\mathcal{L}(\beta)$	law of β		yes	yes
$V(x, s)$	liquidation payoff		yes	yes

Table 2.1: Summary of symbols in the dynamic information and activism trading model.

2.1.2 Definitions and Assumptions

Definition 2.1.4. We call $H\colon\ [0,T] \times \mathbb{R}^n \mapsto \mathbb{R}^n$ a *pricing rule* if

(i) $H(t, \cdot)$ is strictly monotone increasing for each $t \in [0, T]$

(ii) $H \in C^{1,2}$

(iii) $\mathbb{E}\Big[\int_0^T H(t, Z_t)^2\,\mathrm{d}t\Big] < \infty$

(iv) $\mathbb{E}\big[H(T, Z_T)^2\big] < \infty$

Item (i) ensures invertibility. Its economic interpretation is that higher demand reflects good news for the risky asset. Item (ii) is a regularity assumption required for H to satisfy the heat equation, whose discussion we defer to corollary 2.3.6. Item (iii) and item (iv) prevent doubling strategies.

Definition 2.1.5. For a given pricing rule H, we say that the $\left\{\mathscr{F}_t^{Y,S}\right\}$ semi-martingale X is an admissible trading strategy if

$$\mathbb{E}\left[\int_0^T H(t, X_t + Z_t)^2 \,\mathrm{d}t\right] < \infty. \tag{2.2}$$

The constraint rules out doubling strategies.

Assumption 2.1.6. For a fixed pricing rule H and a fixed strategy $\{X_t\}$, the informed trader profits

$$V(\beta + X_T, S_T) - \int_0^T P_t \,\mathrm{d}X_t - \langle X, P\rangle_T \tag{2.3}$$

where $P_t = H(t, X_t + Z_t)$.

Remark 2.1.7. In the previous chapter, the informed profit (1.2) is

$$X_T\tilde{v} - \int_0^T P_t \,\mathrm{d}X_t - \langle X, P\rangle_T.$$

Replacing X_T with $X_T + \beta$ and $\tilde{v}$ with $f(S_T)$, we obtain

$$(X_T + \beta)\, f(S_T) - \int_0^T P_t \,\mathrm{d}X_t - \langle X, P\rangle_T. \tag{2.4}$$

To obtain (2.3), replace $(X_T + \beta)\, f(S_T)$ with $V(\beta + X_T, S_T)$.

Assumption 2.1.8. The map $V(x, s)$ is convex differentiable in x.

Assumption 2.1.9. V satisfies the growth conditions

$$\big|V(y - \tilde{z}, s)\big| \le m(y) + n(\tilde{z}, s) \tag{2.5}$$

$$\left|\frac{\partial V}{\partial x}(y - \tilde{z}, s)\right| \le \tilde{m}(y) + \tilde{n}(\tilde{z}, s), \tag{2.6}$$

where $n, \tilde{n} \in L^2$ and $m, \tilde{m}$ are at most polynomial growth.

Definition 2.1.10. Fix a pricing rule H. We say that a trading strategy $\{X_t^*\}$ is *optimal* for H if it maximizes the informed trader's expected profits:

$$\begin{aligned}\mathbb{E}\left[V(\beta + X_T^*, S_T) - \int_0^T H(t, Y_t^*)\,\mathrm{d}X_t^* - \langle X^*, H(\cdot, Y_\cdot^*)\rangle_T\right] \\ \geq \mathbb{E}\left[V(\beta + X_T, S_T) - \int_0^T H(t, Y_t)\,\mathrm{d}X_t - \langle X, H(\cdot, Y_\cdot)\rangle_T\right]\end{aligned} \tag{2.7}$$

Now, fix a strategy $\{X_t\}$. We say that a pricing rule H^* is *rational* for $\{X_t\}$ if

$$H^*(t, X_t + Z_t) = \mathbb{E}\Big[\nabla_x V(\beta + X_T, S_T)\Big|\mathscr{F}_t^{X+Z}\Big]. \tag{2.8}$$

We call a pair (H^*, X^*) an *equilibrium* if $\{X_t^*\}$ is optimal for H^* and H^* is *rational* for $\{X_t^*\}$.

Remark 2.1.11. (2.8) is a generalization of (1.5).

2.2 Main Result

We constructed the previous equilibrium in theorem 1.2.1 using the solution of a seemingly unrelated optimal transport problem. Here, we do the same.

Problem 2.2.1. Denote $\tilde{Z} = Z - \beta$. The primal problem is

$$\sup_\pi \mathbb{E}^\pi\Big[V(Y_T - \tilde{Z}, S_T)\Big] \tag{2.9}$$

where π has marginals $\mathcal{L}(\tilde{Z}, S_T)$ and $\mathcal{L}(Y_T)$. Here, $\mathcal{L}(\tilde{Z}, S_T)$ is given in the context of the informed trading problem. On the other hand, we will assume that

$$\mathcal{L}(Y_T) \sim \mathcal{N}\left(0, \int_0^T \Sigma_Z(t)\,\mathrm{d}t\right) \tag{2.10}$$

for reasons that we will specify in proposition 2.3.3 when we discuss the inconspicuous trading property. The dual problem is

$$\inf_{\varphi,\psi}\left(\mathbb{E}\big[\varphi(Y_T)\big] + \mathbb{E}\Big[\psi(\tilde{Z}, S_T)\Big]\right) \tag{2.11}$$

over all functions φ, ψ satisfying

$$\varphi(y) + \psi(\tilde{z}, s) \geq V(y - \tilde{z}, s)$$

for all $\tilde{z}, s, y \in \mathbb{R}$.

Assumption 2.2.2. V and $\mathcal{L}(\tilde{Z}, S_T)$ are such that there is no duality gap; the optimal primal value (2.9) equals the optimal dual value (2.11).

1. The dual problem (2.11) attains its infimum at some pair φ^*, ψ^* satisfying

$$\varphi^*(y) = \sup_{\tilde{z},s} \left\{ V(y - \tilde{z}, s) - \psi^*(\tilde{z}, s) \right\} \tag{2.12}$$

$$\psi^*(\tilde{z}, s) = \sup_{y} \left\{ V(y - \tilde{z}, s) - \varphi^*(y) \right\} \tag{2.13}$$

2. since $\mathcal{L}(\tilde{Z}, S_T)$ is absolutely continuous, there exists a map I for which

$$Y_T = I(\tilde{Z}, S_T) \qquad \pi^*\text{-a.s.} \tag{2.14}$$

This comes from the first order condition of (2.13), which reads

$$\nabla_x V(y - \tilde{z}, s) - \nabla \varphi^*(y) = 0.$$

The assumption (2.14) is that we may solve for y as some function of $\tilde{z}$ and s.

Theorem 2.2.3. *Let π^* denote the optimal coupling of* (2.9) *and let (φ^*, ψ^*) denote the optimal Kantorovich potentials in* (2.11)*. Suppose that the PDE*

$$\frac{\partial p}{\partial t} + \frac{1}{2}\mathrm{tr}\Big(\Sigma_Z(t)\nabla_y^2 p\Big) = \frac{1}{2}\left(\mathrm{tr}\big(\Sigma_Z(t)\nabla_{\tilde{z}}^2 p\big) + \mathrm{tr}\big(\Sigma_S(t)\nabla_s^2 p\big)\right) \tag{2.15}$$

$$p(T, \tilde{z}, s, y)\,\mathrm{d}\tilde{z}\,\mathrm{d}s = \pi^*_{Y_T}(\mathrm{d}\tilde{z}, \mathrm{d}s) \tag{2.16}$$

admits a solution $p(t, \tilde{z}, \mathbf{s}, y)$. Then, the pricing rule

$$H^*(t, y) = \mathbb{E}\big[\nabla \varphi^*(Z_T) \big| Z_t = y\big] \tag{2.17}$$

and the trading strategy

$$\mathrm{d}X_t^* = \frac{1}{p(t, Z_t - \beta, S_t, X_t^* + Z_t)} \Sigma_Z(t)\left(\nabla_y + \nabla_{\tilde{z}}\right) p(t, Z_t - \beta, S_t, X_t^* + Z_t)\,\mathrm{d}t \tag{2.18}$$

form an equilibrium.

Remark 2.2.4. p is a conditional density. It represents the market maker's filtered beliefs on the private information $\left(\tilde{Z}_t, S_t\right)$ conditioned on the order flow $\{\mathscr{F}_t^Y\}$. The PDE (2.15)

is in general ill-posed. However, with appropriate conditions on Σ_Z, Σ_S, V, and $\mathcal{L}(\tilde{Z}, S_T)$, it admits a unique solution.

Example 2.2.5. To recover the static model from chapter 1, take $\Sigma_Z(t) = \sigma^2$, $\mathrm{Var}(S_0) = \sigma^2 T$, $\Sigma_S = 0$ (so that $S_t = S_0$ is a constant random variable), $\beta = 0$, and $V(x, s) = xh^{-1}(s)$. Here, h refers to the Brenier map as defined in theorem 1.2.1, but it can also be taken as the CDF composition $F^{-1} \circ N$, with F and N defined in section 1.2.1. The primal problem in this case reads

$$\max_{\pi} \mathbb{E}^{\pi}\big[h^{-1}(S_T)Y_T\big] \tag{2.19}$$

where π has marginals $\mathcal{L}(S_T) \sim \mathcal{N}(0, \sigma^2 T)$ and $\mathcal{L}(Y_T) \sim \mathcal{N}(0, \sigma^2 T)$. With a change of variables $\tilde{v} = h^{-1}(S_T)$, the primal problem becomes the max correlation problem from section 1.2.1. The optimal coupling is the one for which $S_T = Y_T$, which is to say the map I in (2.14) is $I(s) = \mathrm{s}$. The PDE (2.15) reads

$$\frac{\partial p}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial y^2} = 0 \tag{2.20}$$

$$p(T, s, y) = \delta_y(s) \tag{2.21}$$

and the solution is

$$p(t, s, y) = \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \exp\left(-\frac{(y-s)^2}{2\sigma^2(T-t)}\right).$$

Thus, the density of S_t condition on $Y_t = y$ is Gaussian with mean y and variance $\sigma^2(T-t)$, as we had found in proposition 1.2.6. The trading strategy (2.18) is

$$\mathrm{d}X_t^* = \frac{S_t - X_t^* - Z_t}{T-t}\,\mathrm{d}t, \tag{2.22}$$

and we recover (1.13). To recover the pricing rule, we consider the dual formulation of (2.19), which reads

$$\inf_{\varphi,\psi} \Big(\mathbb{E}\big[\varphi(Y_T)\big] + \mathbb{E}\big[\psi(S_T)\big]\Big) \tag{2.23}$$

$$\varphi(y) + \psi(s) \geq yh^{-1}(s). \tag{2.24}$$

Making the change of variables $\tilde{v} = h^{-1}(S_T)$, (2.23) becomes

$$\inf_{\varphi,\psi} \Big(\mathbb{E}\big[\varphi(Y_T)\big] + \mathbb{E}\big[(\psi \circ h)\,(\tilde{v})\big] \Big) \tag{2.25}$$

$$\varphi(y) + (\psi \circ h)\,(v) \geq vy \tag{2.26}$$

Since this coincides with the max correlation formulation, φ and $\psi \circ h$ as they appear in (2.23) are Brenier potentials. In particular, φ' is the Brenier map from $\mathcal{L}(Y_T)$ to $\mathcal{L}(\tilde{v})$. Thus, the general pricing rule (2.17) agrees with the specific pricing rule (1.9).

Example 2.2.6. To recover the dynamic information model from [BP98], take $\Sigma_Z(t) = \sigma_Z(t)^2$, $\mathrm{Var}(S_0) > 0$, $\Sigma_S(t) = \sigma_S(t)^2$, $\beta = 0$, and $V(x,s) = xf(s)$ for f strictly increasing. The primal problem reads

$$\sup_{\pi} \mathbb{E}^{\pi}\big[f(S_T)Y_T\big] \tag{2.27}$$

where π has marginals $\mathcal{L}(S_T)$ and $\mathcal{L}(Y_T) \sim \mathcal{N}\Big(0, \int_0^T \sigma_Z(t)^2\,\mathrm{d}t\Big)$. It turns out that the optimal coupling π^* is the one for which $S_T = Y_T$, so again we have $I(s) = s$. The PDE (2.15) becomes

$$\frac{\partial p}{\partial t} + \frac{\sigma_Z(t)^2}{2}\frac{\partial^2 p}{\partial y^2} = \frac{\sigma_S(t)^2}{2}\frac{\partial^2 p}{\partial s^2} \tag{2.28}$$

$$p(T,s,y) = \delta_y(s) \tag{2.29}$$

This PDE in general is ill-posed, but if we assume that there exists $\varepsilon > 0$ for which

$$\int_t^T \sigma_Z(u)^2\,\mathrm{d}u > (1+\varepsilon)\int_t^T \sigma_S(u)^2\,\mathrm{d}u, \qquad \forall t \in [0,T)$$

then the PDE admits the unique solution

$$p(t,s,y) = \frac{1}{\sqrt{2\pi\Sigma(t)}}\exp\left(-\frac{(s-y)^2}{2\Sigma(t)}\right) \tag{2.30}$$

where

$$\Sigma(t) = \int_t^T \big[\sigma_Z(u)^2 - \sigma_S(u)^2\big]\,\mathrm{d}u.$$

This says that S_t conditioned on $\mathscr{F}_t^Y$ has distribution $\mathcal{N}\big(Y_t, \Sigma(t)\big)$. $\Sigma(t)$ is a measure of how much private information remains unrevealed at time t. $\Sigma(t)$ starts at $\Sigma(0) = \mathrm{Var}(S_0)$ and in equilibrium ends at $\Sigma(T) = 0$. Thus, it is optimal for the informed trader to reveal all of her private information at the end of the trading period.

Substituting p into (2.18), we obtain the trading strategy

$$\mathrm{d}X_t^* = \sigma_Z(t)^2 \frac{S_t - X_t^* - Z_t}{\Sigma(t)}\,\mathrm{d}t. \tag{2.31}$$

As for the pricing rule, consider the dual problem of (2.27), which reads

$$\inf_{\varphi,\psi} \Big(\mathbb{E}\big[\varphi(Y_T)\big] + \mathbb{E}\big[\psi(S_T)\big] \Big) \tag{2.32}$$

$$\varphi(y) + \psi(s) \geq y f(s) \tag{2.33}$$

or, with the change of variables $\tilde{v} = f(S_T)$,

$$\inf_{\varphi,\psi} \left(\mathbb{E}\big[\varphi(Y_T)\big] + \mathbb{E}\Big[\big(\psi \circ f^{-1}\big)(\tilde{v}) \Big] \right) \tag{2.34}$$

$$\varphi(y) + \big(\psi \circ f^{-1}\big)(v) \geq vy. \tag{2.35}$$

This means that φ^* and $\big(\psi \circ f^{-1}\big)^*$ are Brenier potentials between $\mathcal{L}(Y_T)$ and $\mathcal{L}(\tilde{v})$. To obtain the pricing rule (2.17), take the optimal φ^* and evaluate

$$H^*(t,y) = \mathbb{E}\Big[(\varphi^*)'(Z_T) \Big| Z_t = y \Big] \tag{2.36}$$

2.3 Derivation of the Equilibrium

2.3.1 The Wealth Decomposition

Theorem 2.3.1. *Let φ denote the Kantorovich potential from $\mathcal{L}(Y_T)$ to $\mathcal{L}(\tilde{Z}_T, S_T)$ as it appears in problem 2.2.1. Normalize $\mathbb{E}\big[\varphi(Z_T)\big] = 0$. Denote*

$$\Gamma(t,y) = \mathbb{E}\big[\varphi(Z_T)\big|Z_t = y\big]\,. \tag{2.37}$$

Fix a trading strategy $\{X_t\}$ which is not necessarily (2.18)*. If the market maker uses the pricing rule* (2.17)*, then the informed trader's expected profits from $\{X_t\}$ admits the representation*

$$\begin{aligned} &\mathbb{E}\left[V(\beta + X_T, S_T) - \int_0^T P_t\,\mathrm{d}X_t - \langle X, P\rangle_T \right] \\ &\qquad = \mathbb{E}\left[V(Y_T - \tilde{Z}_T, S_T) - \varphi(Y_T) - \frac{1}{2}\int_0^T \mathrm{tr}\Big(\nabla_y^2 \Gamma(t, Y_t)\Big)\,\mathrm{d}\langle X\rangle_t \right] \end{aligned} \tag{2.38}$$

which we call the wealth decomposition.

The proof is exactly the same as that of theorem 1.2.15.

2.3.2 Inconspicuous Trading

Definition 2.3.2. A trading strategy $\{X_t\}$ is called *inconspicuous* if $t \mapsto X_t$ has finite variation almost surely and

$$\mathbb{E}\left[\frac{\mathrm{d}X_t}{\mathrm{d}t}\middle|\mathscr{F}_t^Y\right] = 0.$$

Proposition 2.3.3. *If the informed trader follows an inconspicuous trading strategy, then the law of the aggregate demand $\{Y_t\}$ under $\{\mathscr{F}_t^Y\}$ is the same as the law of the noise demand $\{Z_t\}$ under $\{\mathscr{F}_t\}$. In particular,*

$$Y_t \sim \mathcal{N}\left(0, \int_0^t \Sigma_Z(s)\,\mathrm{d}s\right)$$

for each $t \in (0,T)$.

Proof. Write $\mathrm{d}X_t = A_t\,\mathrm{d}t$ with $\mathbb{E}\big[A_t\big|\mathscr{F}_t^Y\big] = 0$. We will show that

$$B_t^Y = \int_0^t \Sigma_Z^{-1/2}(s)\,\mathrm{d}Y_s \tag{2.39}$$

is a $\{\mathscr{F}_t^Y\}$ Brownian motion by Levy's characterization. First, note that

$$\mathrm{d}B_t^Y = \Sigma_Z^{-1/2}(t)\,(\mathrm{d}X_t + \mathrm{d}Z_t) = \Sigma_Z^{-1/2}(t)\left(A_t\,\mathrm{d}t + \Sigma_Z^{1/2}(t)\,\mathrm{d}B_t^Z\right) = \Sigma_Z^{-1/2}(t)A_t\,\mathrm{d}t + \mathrm{d}B_t^Z,$$

so that $\langle B^Y\rangle_t = t\mathbf{I}$. Next, fix $s < t$. We have

$$\begin{aligned}
\mathbb{E}\Big[B_t^Y\Big|\mathscr{F}_s^Y\Big] &= B_s^Y + \mathbb{E}\left[\int_s^t \Sigma_Z^{-1/2}(u)A_u\,\mathrm{d}u + \int_s^t \mathrm{d}B_u^Z\middle|\mathscr{F}_s^Y\right] \\
&= B_s^Y + \int_s^t \Sigma_Z^{-1/2}(u)\mathbb{E}\Big[A_u\Big|\mathscr{F}_s^Y\Big]\,\mathrm{d}u \\
&= B_s^Y + \int_s^t \Sigma_Z^{-1/2}(u)\mathbb{E}\left[\mathbb{E}\Big[A_u\Big|\mathscr{F}_u^Y\Big]\middle|\mathscr{F}_s^Y\right]\mathrm{d}u = B_s^Y.
\end{aligned}$$

Thus $\{B_t^Y\}$ is an $\{\mathscr{F}_t^Y\}$ local martingale with quadratic variation t. □

Theorem 2.3.4. *Any inconspicuous strategy is optimal which ensures $Y_T = I(\tilde{Z}_T, S_T)$ π^*- almost surely, where I is defined by* (2.14)*. The informed trader should expect to profit at*

most $\mathbb{E}\Big[\psi^*\Big(\tilde{Z}_T, S_T\Big)\Big]$, *where* ψ^* *is defined in* (2.13).

Proof. Consider the RHS of (2.38). For fixed Z_T, β, and S_T, the quantity $V(Y_T - \tilde{Z}_T, S_T) - \varphi(Y_T)$ is at most

$$\psi^*\Big(\tilde{Z}_T, S_T\Big) = \sup_y \Big\{V(y - \tilde{Z}_T, S_T) - \varphi(y)\Big\} \tag{2.40}$$

and achieves this value when $y = I(\tilde{Z}_T, S_T)$. The informed trader should therefore force Y_T to satisfy $Y_T = I(\tilde{Z}_T, S_T)$ by manipulating the order flow. In this case, her maximized profit is $\mathbb{E}\Big[\psi^*\Big(\tilde{Z}_T, S_T\Big)\Big]$. □

Corollary 2.3.5. *If the informed trader follows an optimal inconspicuous strategy, the pricing rule* (2.17) *is rational in the sense of definition 2.1.10.*

Proof. From (2.17), we have

$$H(t, Y_t) = \mathbb{E}\big[\nabla\varphi(Z_T)\big|Z_t = Y_t\big]\,.$$

If the informed trader follows an inconspicuous strategy, proposition 2.3.3 asserts that $\mathcal{L}(Y_{[0,T]})$ relative to $\{\mathscr{F}_t^Y\}$ is the same as $\mathcal{L}(Z_{[0,T]})$ relative to $\{\mathscr{F}_t\}$. Hence,

$$\mathbb{E}\big[\nabla\varphi(Z_T)\big|Z_t = Y_t\big] = \mathbb{E}\big[\nabla\varphi(Y_T)\big|Y_t\big]\,.$$

By the Markov property,

$$\mathbb{E}\big[\nabla\varphi(Y_T)\big|Y_t\big] = \mathbb{E}\Big[\nabla\varphi(Y_T)\Big|\mathscr{F}_t^Y\Big]\,.$$

Finally, note that $Y_T = I(Z_T - \beta, S_T)$ guarantees the first order condition of (2.40), which reads

$$\nabla\varphi(Y_T) = \nabla_x V(Y_T - Z_T + \beta, S_T) = \nabla_x V(X_T + \beta, S_T).$$

□

Corollary 2.3.6. *If the informed trader follows an optimal inconspicuous strategy, the price process* $\{P_t^*\} = \{H^*(t, Y_t^*)\}$ *is an* $\big\{\mathscr{F}_t^{Y^*}\big\}$ *martingale.*

Proof. Fix $s < t$. We compute

$$\begin{aligned}
\mathbb{E}\Big[P_t\Big|\mathscr{F}_s^{Y^*}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\nabla_x V(X_T+\beta,S_T)\Big|\mathscr{F}_t^{Y^*}\Big]\Big|\mathscr{F}_s^{Y^*}\Big] && \text{corollary 2.3.5}\\
&= \mathbb{E}\Big[\nabla_x V(X_T+\beta,S_T)\Big|\mathscr{F}_s^{Y^*}\Big] && \text{tower property}\\
&= P_s && \text{corollary 2.3.5}
\end{aligned}$$

From $\mathbb{E}\Big[\big\|\nabla_x V(X_T+\beta,S_T)\big\|^2\Big] < \infty$, we immediately obtain integrability:

$$\begin{aligned}
\mathbb{E}\big[\|P_t\|\big] &= \mathbb{E}\left[\left\|\mathbb{E}\Big[\nabla_x V(X_T+\beta,S_T)\Big|\mathscr{F}_t^{Y^*}\Big]\right\|\right] && \text{corollary 2.3.5}\\
&\le \mathbb{E}\left[\mathbb{E}\Big[\big\|\nabla_x V(X_T+\beta,S_T)\big\|\Big|\mathscr{F}_t^{Y^*}\Big]\right] && \text{triangle inequality}\\
&= \mathbb{E}\Big[\big\|\nabla_x V(X_T+\beta,S_T)\big\|\Big] && \text{tower property}\\
&\le \sqrt{\mathbb{E}\Big[\big\|\nabla_x V(X_T+\beta,S_T)\big\|^2\Big]} < \infty && \text{Cauchy-Schwarz}
\end{aligned}$$

□

Corollary 2.3.7. *H^* solves the backwards heat equation*

$$\frac{\partial H^*}{\partial t} + \frac{1}{2}\mathrm{tr}\big(\Sigma_Z(t)\nabla^2 H^*\big) = 0$$

with terminal condition $H^*(T,y) = \nabla\varphi(y)$.

2.3.3 The Market Maker's Filtering Problem

Proposition 2.3.8. *Suppose the informed trader uses an optimal inconspicuous strategy of the form $\mathrm{d}X_t = A_t\,\mathrm{d}t$. Then, the conditional density*

$$\rho_t(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s = \mathbb{P}\Big(\tilde{Z}_t \in \mathrm{d}\tilde{z}, S_t \in \mathrm{d}s\Big|\mathscr{F}_t^Y\Big) \tag{2.41}$$

satisfies the Kushner SPDE

$$\begin{aligned}
\mathrm{d}\rho_t(\tilde{z},s) = \frac{1}{2}\mathrm{tr}\Big(\big(\Sigma_Z(t)\nabla_{\tilde{z}}^2 + \Sigma_S(t)\nabla_s^2\big)\,\rho_t(\tilde{z},s)\Big)\,\mathrm{d}t\\
+ \big(-\nabla_{\tilde{z}}\rho_t(\tilde{z},s) + \rho_t(\tilde{z},s)\Sigma_Z^{-1}(t)A_t\big)^\top\,\mathrm{d}Y_t
\end{aligned} \tag{2.42}$$

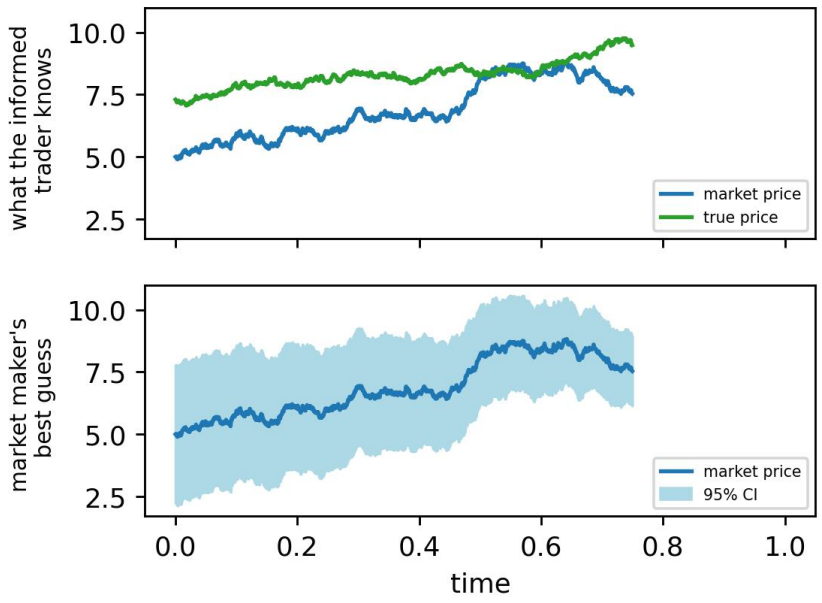


Figure 2.4: The market maker's filtering problem in the dynamic case.

with initial condition

$$\rho_0(\tilde{z}, s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s = \mathbb{P}\Big(\tilde{Z}_0 \in \mathrm{d}\tilde{z}, S_0 \in \mathrm{d}s\Big) \tag{2.43}$$

Proof. Let the test function $\varphi\colon \mathbb{R} \mapsto \mathbb{R}$ be twice differentiable. Itô's rule gives

$$\begin{aligned}\varphi(\tilde{Z}_t, S_t) = \varphi(\tilde{Z}_0, S_0) &+ \frac{1}{2}\int_0^t \mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\varphi(\tilde{Z}_u, S_u)\Big)\,\mathrm{d}u \\ &+ \int_0^t \Big(\nabla_s\varphi(\tilde{Z}_u, S_u)\Big)^\top \Sigma_S^{1/2}(u)\,\mathrm{d}W_u + \int_0^t \Big(\nabla_{\tilde{z}}\varphi(\tilde{Z}_u, S_u)\Big)^\top \Sigma_Z^{1/2}(u)\,\mathrm{d}B_u\end{aligned} \tag{2.44}$$

Together with

$$Y_t = \int_0^t A_u\,\mathrm{d}u + \int_0^t \Sigma_Z^{1/2}(u)\,\mathrm{d}B_u, \tag{2.45}$$

the pair $(\varphi(S_t), Y_t)$ forms a partially observable process. Appealing to [LS77, Theorem 8.1], we obtain the filtering equation

$$\begin{aligned}\mathbb{E}\Big[\varphi(\tilde{Z}_t, S_t)\Big|\mathscr{F}_t^Y\Big] &= \mathbb{E}\Big[\varphi(\tilde{Z}_0, S_0)\Big|\mathscr{F}_0^Y\Big] \\ &+ \frac{1}{2}\int_0^t \mathbb{E}\Big[\mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\varphi(\tilde{Z}_u, S_u)\Big)\Big|\mathscr{F}_u^Y\Big]\,\mathrm{d}u \\ &+ \int_0^t \mathbb{E}\Big[\nabla_{\tilde{z}}\varphi(\tilde{Z}_u, S_u)^\top + \varphi(\tilde{Z}_u, S_t)A_u^\top\Sigma_Z^{-1}(u)\Big|\mathscr{F}_u^Y\Big]\,\mathrm{d}Y_u.\end{aligned} \tag{2.46}$$

In integral form, this reads

$$\begin{aligned}\iint \varphi(\tilde{z},s)\rho_t(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s &= \iint \varphi(\tilde{z},s)\rho_0(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s \\ &\quad + \frac{1}{2}\int_0^t \iint \mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\varphi(\tilde{z},s)\Big)\,\rho_u(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s\,\mathrm{d}u \\ &\quad + \int_0^t \iint \big(\nabla_{\tilde{z}}\varphi(\tilde{z},s) + \varphi(\tilde{z},s)\Sigma_Z^{-1}(u)A_u\big)^\top \rho_u(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s\,\mathrm{d}Y_u\end{aligned} \tag{2.47}$$

Several applications of Fubini's theorem and integration by parts give

$$\begin{aligned}\iint \rho_t(\tilde{z},s)\varphi(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s &= \iint \rho_0(\tilde{z},s)\varphi(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s \\ &\quad + \frac{1}{2}\int_0^t \iint \mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\rho_u(\tilde{z},s)\Big)\,\varphi(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s\,\mathrm{d}u \\ &\quad + \int_0^t \iint \big(-\nabla_{\tilde{z}}\rho_u(\tilde{z},s) + \rho_u(\tilde{z},s)\Sigma_Z^{-1}(u)A_u\big)^\top \varphi(\tilde{z},s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s\,\mathrm{d}Y_u\end{aligned} \tag{2.48}$$

Since φ is arbitrary, (2.48) implies

$$\begin{aligned}\rho_t(\tilde{z},s) &= \rho_0(\tilde{z},s) + \frac{1}{2}\int_0^t \mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\rho_u(\tilde{z},s)\Big)\,\mathrm{d}u \\ &\quad + \int_0^t \big(-\nabla_{\tilde{z}}\rho_u(\tilde{z},s) + \rho_u(\tilde{z},s)\Sigma_Z^{-1}(u)A_u\big)^\top \mathrm{d}Y_u,\end{aligned} \tag{2.49}$$

which is (2.42). □

Proposition 2.3.9. *Assume that the PDE* (2.15) *admits a unique solution. If the informed trader follows the strategy* (2.18)*, then the conditional density* (2.41) *also satisfies the terminal condition*

$$\rho_T(\tilde{z},s) = \pi^*_{Y_T}(\tilde{z},s). \tag{2.50}$$

Proof. We introduce ϱ through the BSPDE

$$\varrho_t(\tilde{z},s) = \pi^*_{Y_T}(\tilde{z},s) - \frac{1}{2}\int_t^T \mathrm{tr}\Big(\big(\Sigma_Z(u)\nabla_{\tilde{z}}^2 + \Sigma_S(u)\nabla_s^2\big)\,\varrho_u(\tilde{z},s)\Big)\,\mathrm{d}u - \int_t^T R_u^\top\,\mathrm{d}B_u^Y \tag{2.51}$$

with the hope that ϱ will coincide with ρ (and indeed it does). Together with

$$Y_t = \int_0^t \Sigma_Z^{1/2}(u)\,\mathrm{d}B_u^Y, \tag{2.52}$$

the pair (Y_t, ϱ_t) forms a BSPDE. Introduce the decoupling field $g \in C^{1,2,2}$ such that

$$\varrho_t(\tilde{z}, s) = g(t, \tilde{z}, s, Y_t). \tag{2.53}$$

Applying the four-step scheme [Zha17, section 8.3.1], we obtain the PDE

$$\frac{\partial g}{\partial t} + \frac{1}{2}\mathrm{tr}\Big(\Sigma_Z(t)\nabla_y^2 g\Big) = \frac{1}{2}\left(\mathrm{tr}\big(\Sigma_Z(t)\nabla_{\tilde{z}}^2 g\big) + \mathrm{tr}\big(\Sigma_S(t)\nabla_s^2 g\big)\right) \tag{2.54}$$

$$g(T, \tilde{z}, s, y) = \frac{\partial^2}{\partial \tilde{z}\partial s}\mathbb{P}^{\pi^*}\Big(\tilde{Z}_T \le \tilde{z}, S_T \le s \Big| Y_T = y\Big) \tag{2.55}$$

by matching drift terms. But this is (2.15), which we assumed admits a unique solution. It follows that $p = g$. On the other hand, equating diffusive terms gives

$$R_t = \Sigma_Z^{1/2}(t)\nabla_y p(t, \tilde{z}, s, Y_t), \tag{2.56}$$

or, in light of (2.18),

$$R_t = -\nabla_{\tilde{z}} p(t, \tilde{z}, s, Y_t) + p(t, \tilde{z}, s, Y_t)\Sigma_Z^{-1}(t)\,\frac{\mathrm{d}X_t}{\mathrm{d}t}\,, \tag{2.57}$$

which is precisely the diffusive term of ρ in (2.42). It remains to show that ϱ satisfies the same initial condition as ρ, namely (2.43). To this end, we introduce the Fourier transform

$$\hat{\varrho}_t(u, v) = \int e^{-i\left(u^\top \tilde{z} + v^\top s\right)}\varrho_t(\tilde{z}, s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s$$

so that the BSPDE (2.51) becomes

$$\mathrm{d}\hat{\varrho}_t(u, v) = -\frac{1}{2}\left(u^\top \Sigma_Z(t)u + v^\top \Sigma_S(t)v\right)\hat{\varrho}_t(u, v)\,\mathrm{d}t + \hat{R}_t(u, v)\,\mathrm{d}Y_t \tag{2.58}$$

$$\hat{\varrho}_T(u, v) = \int e^{-i\left(u^\top \tilde{z} + v^\top s\right)}\pi^*_{Y_T}(\mathrm{d}\tilde{z}, \mathrm{d}s). \tag{2.59}$$

This is a linear BSDE whose solution admits the representation

$$\hat{\varrho}_t(u, v) = \exp\left(\frac{1}{2}\int_t^T \left(u^\top \Sigma_Z(r)u + v^\top \Sigma_S(r)v\right)\mathrm{d}r\right)\mathbb{E}\left[\int e^{-i\left(u^\top \tilde{z} + v^\top s\right)}\pi^*_{Y_T}(\mathrm{d}\tilde{z}, \mathrm{d}s)\,\middle|\,\mathscr{F}_t^Y\right].$$

At $t = 0$, this is

$$\hat{\varrho}_0(u, v) = \exp\left(\frac{1}{2}\int_0^T \left(u^\top \Sigma_Z(r)u + v^\top \Sigma_S(r)v\right)\mathrm{d}r\right)\mathbb{E}\left[\int e^{-i\left(u^\top \tilde{z} + v^\top s\right)}\pi^*_{Y_T}(\mathrm{d}\tilde{z}, \mathrm{d}s)\right] \tag{2.60}$$

By the properties of disintegration,

$$\begin{aligned}\mathbb{E}\left[\int e^{-i\left(u^\top \tilde{z}+v^\top s\right)}\pi^*_{Y_T}(\mathrm{d}\tilde{z},\mathrm{d}s)\right] &= \int e^{-i\left(u^\top \tilde{z}+v^\top s\right)}\pi^*(\mathrm{d}\tilde{z},\mathrm{d}s,\mathrm{d}y)\\ &= \mathbb{E}\left[e^{-i\left(u^\top \tilde{Z}+v^\top S_T\right)}\right]\\ &= \mathbb{E}\left[e^{-i\left(u^\top Z_T+v^\top \int_0^T \Sigma_S^{1/2}(r)\,\mathrm{d}W_r\right)}\right]\mathbb{E}\left[e^{-i\left(-u^\top\beta+v^\top S_0\right)}\right]\end{aligned}$$

and so (2.60) becomes

$$\hat{\varrho}_0(u,v) = \mathbb{E}\left[e^{-i\left(-u^\top\beta+v^\top S_0\right)}\right] = \mathbb{E}\left[e^{-i\left(u^\top \tilde{Z}_0+v^\top S_0\right)}\right].$$

Thus we see that $\hat{\varrho}_0(u,v)$ is the characteristic function of $(\tilde{Z}_0, S_0)$. This verifies the initial condition (2.43). □

2.4 A sobering counter-example

While our general approach yields equilibria for all known continuous-time Kyle models with risk-neutral agents, this subsection shows that the SPDE (2.42) may fail to have solutions when the function I lacks sufficient smoothness in $\tilde{z}$. For many choices of V, one should not expect I to be infinitely differentiable. Indeed, transport maps are typically non-smooth, and this lack of regularity obstructs solvability of the BSPDE.

Proposition 2.4.1. *Assume that the BSPDE* (2.42) *admits a solution and the law* μ_0 *of* $(-\beta, S_0)$ *admits a smooth density. Then,* $\tilde{z} \mapsto I(\tilde{z}, s)$ *is* C^∞.

Proof. For simplicity, we assume $n = 1$ and $\sigma_S(t) = 0$ and $\sigma_Z(t) = \sigma$ for all $t \in [0,T]$. For all $t \in [0,T)$, denote by $\gamma_t(\tilde{z})$ the Gaussian density of $\tilde{Z}_T - \tilde{Z}_t$ and ρ_t a solution to (2.42). Denoting $\widehat{\gamma_t *_{\tilde{z}} \rho_t}$ the Fourier transform of $\gamma_t *_{\tilde{z}} \rho_t$, where $*_{\tilde{z}}$ denotes the convolution in $\tilde{z}$, we have, by a direct computation,

$$(\widehat{\gamma_t *_{\tilde{z}} \rho_t})(u,v) = \mathbb{E}\left[\int e^{-i(u\tilde{z}+vs)}\pi^*_{Y_T}(\mathrm{d}\tilde{z},\mathrm{d}\xi)\middle|\mathscr{F}_t^Y\right]$$

where Y is a fixed σ-Brownian motion.

Inverting the Fourier transform, one finds the very simple expression

$$(\gamma_t *_{\tilde{z}} \rho_t)(\tilde{z}, s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s = \mathbb{E}\left[\pi^*_{Y_T}(\mathrm{d}\tilde{z},\mathrm{d}\xi)\middle|\mathscr{F}_t^Y\right],$$

or equivalently,

$$\begin{aligned}
&(\gamma_t *_{\tilde{z}} \rho_t)(\tilde{z}, s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s \\
&= \int \pi_y^*(\mathrm{d}\tilde{z}, \mathrm{d}s) e^{-\frac{(y-Y_t)^2}{2\sigma^2(T-t)}} \frac{\mathrm{d}y}{\sqrt{2\pi\sigma^2(T-t)}} \\
&= \int \pi^*(\mathrm{d}y, \mathrm{d}\tilde{z}, \mathrm{d}s) \sqrt{\frac{1}{1-t/T}}\ \exp\left(\frac{y^2}{2\sigma^2 T} - \frac{(y-Y_t)^2}{2\sigma^2(T-t)}\right)\ ,
\end{aligned} \tag{2.61}$$

where in the last step we used the fact that $\pi^*(\mathrm{d}y, \mathrm{d}\tilde{z}, \mathrm{d}s) = \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{-\frac{y^2}{2\sigma^2 T}}\,\mathrm{d}y\,\pi_y^*(\mathrm{d}\tilde{z}, \mathrm{d}s)$ is an optimal coupling between the law ν_T of Y_T and the law μ_T of $(\tilde{Z}_T, S_T)$.

By assumption on μ_0, μ_T admits a smooth density that we denote $f_\mu(\tilde{s}, \tilde{z})$. Let us write the following exact formula for equation (2.61)

$$\begin{aligned}
&(\gamma_t *_{\tilde{z}} \rho_t)(\tilde{z}, s)\,\mathrm{d}\tilde{z}\,\mathrm{d}s \\
&= \int \delta_{I(\tilde{z},s)}(\mathrm{d}y) f_\mu(\tilde{z}, s) \sqrt{\frac{1}{1-t/T}}\ \exp\left(\frac{y^2}{2\sigma^2 T} - \frac{(y-Y_t)^2}{2\sigma^2(T-t)}\right) \mathrm{d}\tilde{z}\,\mathrm{d}s \\
&= f_\mu(\tilde{z}, s) \sqrt{\frac{1}{1-t/T}}\ \exp\left(\frac{I^2(\tilde{z}, s)}{2\sigma^2 T} - \frac{(I(\tilde{z}, s)-Y_t)^2}{2\sigma^2(T-t)}\right) \mathrm{d}\tilde{z}\,\mathrm{d}s\ .
\end{aligned}$$

Note that $\tilde{z} \mapsto (\gamma_t *_{\tilde{z}} \rho_t)(\tilde{z}, s)$ is a convolution with a Gaussian density. Thus, we obtain that $\tilde{z} \mapsto \exp\left(\frac{I^2(\tilde{z},s)}{2\sigma^2 T} - \frac{(I(\tilde{z},s)-Y_t)^2}{2\sigma^2(T-t)}\right)$ has to be smooth for all $t \in (0, T)$ which imposes that $\tilde{z} \mapsto I(\tilde{z}, s)$ is smooth. □

2.5 Conclusion

This chapter extended the optimal-transport formulation of the classical Kyle–Back model to accommodate dynamic information and activist trading. While the coupling remains between the terminal distributions of the insider's private information and the market maker's beliefs, the cost function has been generalized to capture nonlinear liquidation payoffs faced by the informed trader. This modification corresponds to the setting of activist trading, in which the insider's objective is no longer purely informational but also strategic, reflecting her influence on the firm's value through trading and intervention. The resulting equilibrium is characterized by a backward stochastic partial differential equation (BSPDE), whose solution determines both the pricing rule and the optimal trading strategy.

Within this framework, inconspicuous trading reappears as a fixed-point condition: under the equilibrium strategy, the aggregate order flow remains a Brownian motion in its own

filtration, even as the insider's incentives become nonlinear. The market maker's filtering problem continues to mediate the relationship between private information, order flow, and prices, linking the geometric structure of optimal transport to the dynamic optimization faced by the insider.

Conceptually, this chapter shows that modifying the *cost function* in the optimal-transport problem suffices to capture the economics of activist trading and nonlinear liquidation payoffs. The Brenier potentials and their gradients now describe both informational and strategic dimensions of price impact. In the next chapter, we shift focus from informational asymmetry to the microstructure of liquidity itself, introducing a variational formulation that allows for stochastic and multidimensional liquidity and thereby connects information flow to the endogenous dynamics of market depth.

CHAPTER 3

Multidimensional Stochastic Liquidity

Having established in the previous chapters how optimal transport underlies both the classical and dynamic Kyle–Back models, we now turn to an even broader generalization. In practice, market liquidity is neither constant nor confined to a single asset. The volatility of noise trading fluctuates randomly throughout the day, and correlated order flows link multiple securities. To capture these features, we extend the optimal-transport formulation of informed trading to a setting with *stochastic and multidimensional liquidity*.

Our approach draws inspiration from the stochastic-volatility models of [CDF16] and the multidimensional extensions of [EMZ22]. We reinterpret their equilibria through a *variational problem* in which the insider optimally controls the rate at which private information is released to the market. The resulting structure connects the theory of Kyle–Back equilibria to that of optimal liquidation and execution, providing an economic meaning to the adjoint process that appears in the stochastic maximum principle of [CDF16].

Mathematically, this chapter develops a primal–dual framework that accommodates random, matrix-valued liquidity and correlated assets. The primal problem determines the optimal decay of the market maker's conditional variance, while the dual problem introduces a matrix-valued martingale representing the shadow price of information. These two problems jointly yield an equilibrium in which the sensitivity matrix—Kyle's multivariate Λ_t—is identified with the gradient of a convex potential in matrix space.

In summary, this chapter extends the optimal-transport perspective of informed trading to random, multidimensional environments, linking stochastic liquidity, covariance dynamics, and information flow within a unified variational formulation.

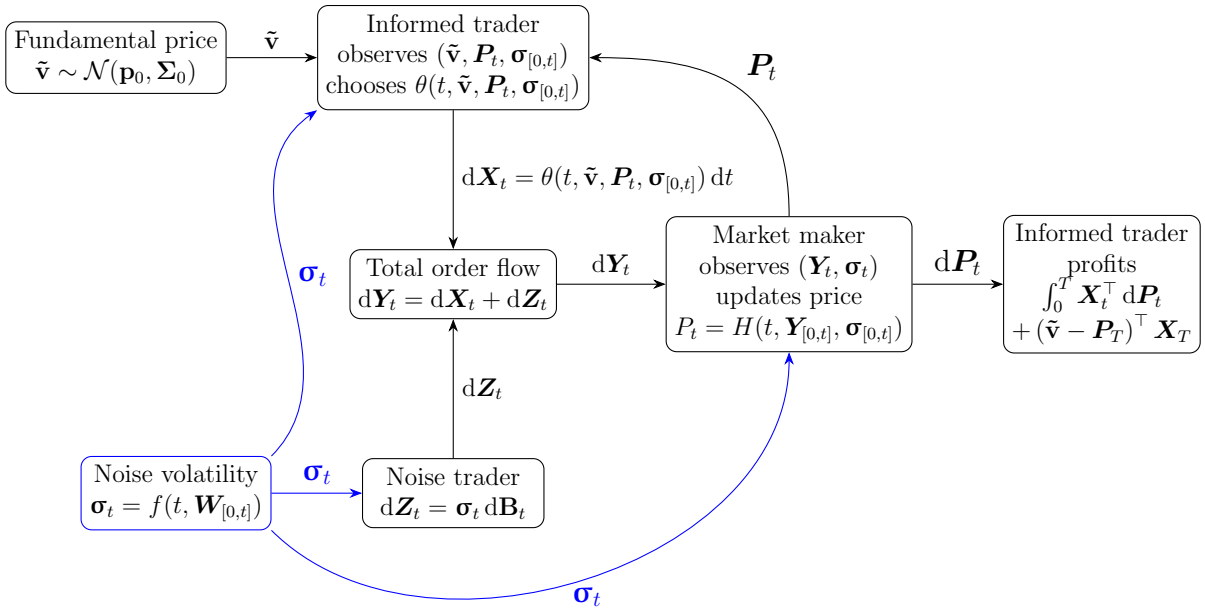


Figure 3.1: Flow chart describing multidimensional stochastic liquidity model.

3.1 The Model

There is a public announcement to be made at time T that will affect the prices of n risky assets. Immediately after this announcement, the assets are valued at

$$\tilde{\mathbf{v}} = (\tilde{v}_1, \ldots, \tilde{v}_n).$$

Before time T, the vector $\tilde{\mathbf{v}}$ is known only to a single informed trader. She knows $\tilde{\mathbf{v}}$ from time 0, when the assets are traded at prices $\boldsymbol{P}_0$.

For $t \in [0, T]$, the informed trader submits order flows $\mathrm{d}\boldsymbol{X}_t \in \mathbb{R}^n$ and accumulates a position $\boldsymbol{X}_t \in \mathbb{R}^n$ in the risky assets over $[0, t]$. Her trades are mixed with noise trades of the form

$$\mathrm{d}\boldsymbol{Z}_t = \boldsymbol{\sigma}_t\ \mathrm{d}\mathbf{B}_t,$$

where the (stochastic) volatility matrix $\boldsymbol{\sigma}_t \in \mathcal{S}^n_{++}$ is independent of the Brownian motion $\mathbf{B}_t \in \mathbb{R}^n$. At each time t, the market maker observes only the aggregate order flow

$$\mathrm{d}\boldsymbol{Y}_t = \mathrm{d}\boldsymbol{X}_t + \mathrm{d}\boldsymbol{Z}_t,$$

and cannot distinguish $\mathrm{d}\boldsymbol{X}_t$ from $\mathrm{d}\boldsymbol{Z}_t$. Based on the filtration $\mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}}$ generated by $\{\boldsymbol{Y}_s, \boldsymbol{\sigma}_s\}_{0 \le s \le t}$,

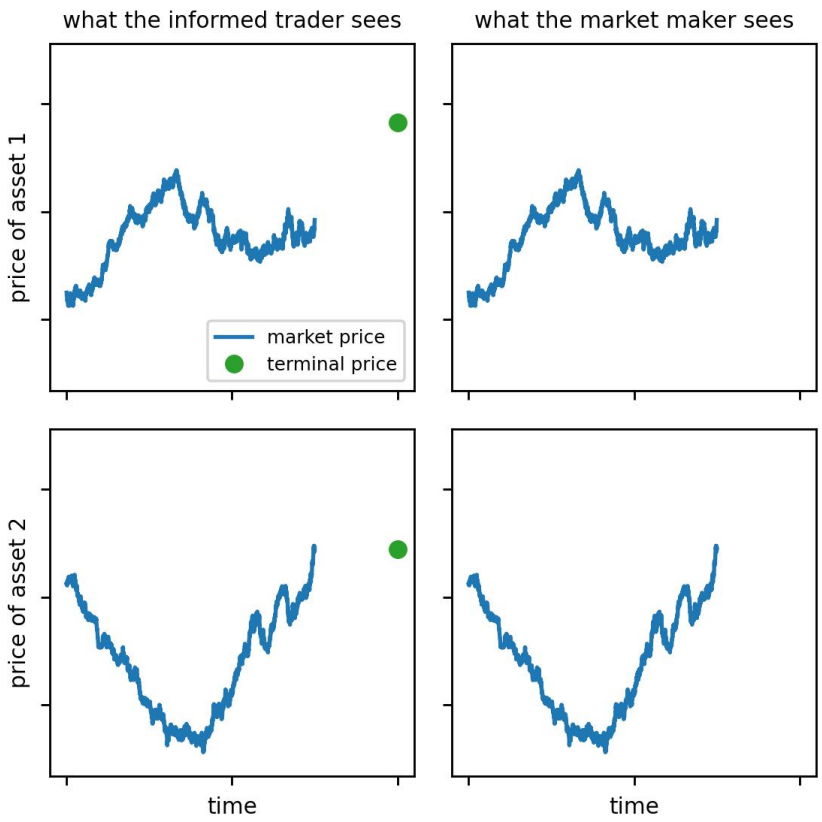


Figure 3.2: Price information asymmetry in the multidimensional stochastic liquidity model.

the market maker sets prices

$$\boldsymbol{P}_t = \mathbb{E}\left[\tilde{\mathbf{v}} \mid \mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}}\right].$$

This model is identical to that of section 1.1, but differs again in two aspects:

1. There are n traded assets instead of just one. As in 1.1, their terminal values $\tilde{\mathbf{v}} = (\tilde{v}_1, \ldots, \tilde{v}_n)$ are known to the informed trader at the start of the trading period.

2. The noise trades have diffusive stochastic volatility $\{\boldsymbol{\sigma}_t\}$. Thus, aggregate trades are

$$\mathrm{d}\boldsymbol{Y}_t = \mathrm{d}\boldsymbol{X}_t + \underbrace{\boldsymbol{\sigma}_t \, \mathrm{d}\mathbf{B}_t}_{\mathrm{d}\boldsymbol{Z}_t}.$$

We call this the *multidimensional stochastic liquidity model*.

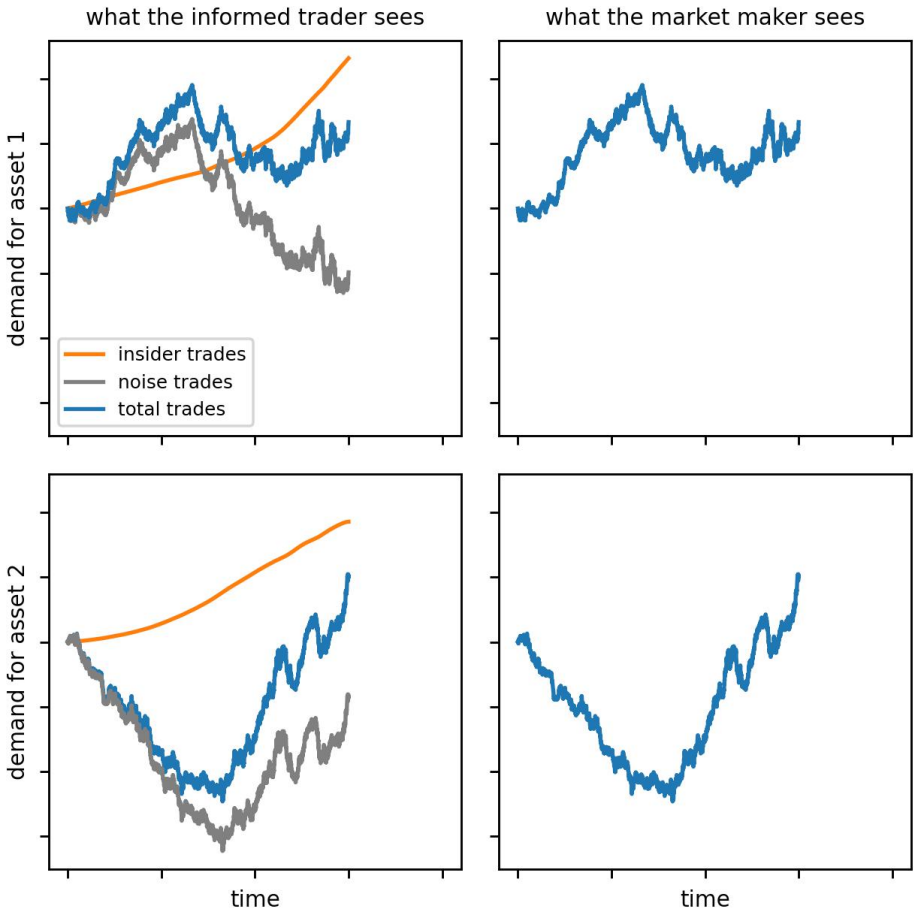


Figure 3.3: Trading information asymmetry in the multidimensional stochastic liquidity model.

3.1.1 Information Asymmetry

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a complete probability space endowed with the right continuous filtration $\{\mathscr{F}_t \colon t \in [0,T]\}$. On this filtered probability space, $(\mathbf{B}_t, \mathscr{F}_t)$ and $(\boldsymbol{W}_t, \mathscr{F}_t)$ are independent Brownian motions. We recall that $\mathbf{B}_t$ drives noise trades, while $\boldsymbol{W}_t$ drives noise volatility.

We model $\tilde{\mathbf{v}} \sim \mathcal{N}(\mathbf{p}_0, \boldsymbol{\Sigma}_0)$ as an $\mathscr{F}_T$ random variable (also independent from $\mathbf{B}_t$ and $\boldsymbol{W}_t$) whose realized value is known to the informed trader before it is announced to the public. Thus, information asymmetry reduces to a comparison between two filtrations, namely $\mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}}$ and $\mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}} \vee \sigma(\tilde{\mathbf{v}})$.

The idea is that the insider has strictly superior information to the market maker, so the market maker's filtration $\mathscr{F}_t^M$ is a (strict) subset of the insider's $\mathscr{F}_t^I$ at each $t \in [0,T]$. The market maker observes the aggregate trades $\{\boldsymbol{Y}_t\}$ and the noise volatility $\{\boldsymbol{\sigma}_t\}$. We permit him to observe noise volatility for model tractability. Thus, the market maker's information is $\mathscr{F}_t^M = \mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}}$. On the other hand, the insider observes the terminal price $\tilde{\mathbf{v}}$ in addition to everything the market maker sees. Her information is $\mathscr{F}_t^I = \mathscr{F}_t^{\boldsymbol{Y},\boldsymbol{\sigma}} \vee \sigma(\tilde{\mathbf{v}})$. Thus, we have $\mathscr{F}_t^M \subset \mathscr{F}_t^I$, as desired.

3.1.2 Definitions

Definition 3.1.1. A *pricing rule* is a pair $(\boldsymbol{H}, \boldsymbol{\Lambda})$ consisting of a function $\boldsymbol{H} \colon [0,T] \times \mathbb{R}^n \mapsto \mathbb{R}^n$ and an $\mathscr{F}_t^M$ semi-martingale $\boldsymbol{\Lambda}_t \in \mathcal{S}_{++}^n$ so that the market price is

$$\boldsymbol{P}_t = \boldsymbol{H}(t, \boldsymbol{\xi}_t), \text{ where } \boldsymbol{\xi}_t = \int_0^t \boldsymbol{\Lambda}_t \, \mathrm{d}\boldsymbol{Y}_t.$$

For a given $\boldsymbol{\Lambda}$, we require of $\boldsymbol{H} \in C^{1,2}$ that

1. there exists a convex function $\varphi(t, \boldsymbol{\xi})$ for which $\nabla_{\boldsymbol{\xi}} \varphi(t, \boldsymbol{\xi}) = \boldsymbol{H}(t, \boldsymbol{\xi})$.
2. $\mathbb{E}\left[\int_0^T \left\| \boldsymbol{H}\left(t, \int_0^t \boldsymbol{\Lambda}_s \, \mathrm{d}\boldsymbol{Z}_s\right)\right\|^2 \mathrm{d}t\right] < \infty$
3. $\mathbb{E}\left[\left\| \boldsymbol{H}\left(T, \int_0^T \boldsymbol{\Lambda}_s \, \mathrm{d}\boldsymbol{Z}_s\right)\right\|^2\right] < \infty$

In contrast to definition 1.1.6, a pricing rule in this setting is a pair. The new addition is the semi-martingale, which provides a t by t weighting of the order flow $\{\boldsymbol{Y}_{[0,t]}\}$. Yet, the conditions are mostly the same as before. Condition (1) generalizes monotonicity. Conditions (2) and (3) are related to doubling strategies and are adapted from [Bac92]. As we will see, $\int_0^t \boldsymbol{\Lambda}_s \, \mathrm{d}\boldsymbol{Z}_s$ represents the state variable in the absence of informed trades.

Definition 3.1.2. For a given pricing rule $(\boldsymbol{H}, \boldsymbol{\Lambda})$, we say that the $\mathscr{F}_t^{\boldsymbol{Z},\boldsymbol{\sigma}} \vee \sigma(\tilde{\mathbf{v}})$ semi-martingale $\boldsymbol{X}$ is an admissible trading strategy if

$$\mathbb{E}\left[\int_0^T \boldsymbol{H}(t, \boldsymbol{\xi}_t)^2 \,\mathrm{d}t \middle| \mathscr{F}_0^I\right] < \infty. \tag{3.1}$$

Here, $\boldsymbol{\xi}_t = \int_0^t \boldsymbol{\Lambda}_s \,\mathrm{d}(\boldsymbol{X}_s + \boldsymbol{Z}_s)$ denotes the *state variable*.

The point of $\boldsymbol{\xi}_t$ is to reduce the path $\boldsymbol{Y}_{[0,t]}$ to a Markovian process. The idea has been used before in [Cho03] and [EMZ22], among others. As we shall see, $\boldsymbol{\Lambda}$ plays the role of *Kyle's lambda* and its inverse has the economic interpretation of *market depth*. Kyle's lambda quantifies price sensitivity to trades, but it turns out that in equilibrium, we will have $\boldsymbol{P}_t = \mathbf{p}_0 + \int_0^t \boldsymbol{\Lambda}_s \,\mathrm{d}\boldsymbol{Y}_s$.

Definition 3.1.3. Fix a pricing rule $(\boldsymbol{H}, \boldsymbol{\Lambda})$. We say that a trading strategy $\{\boldsymbol{X}_t^*\}$ is *optimal* for $(\boldsymbol{H}, \boldsymbol{\Lambda})$ if it maximizes the informed trader's expected profits:

$$\begin{aligned} &\mathbb{E}\left[\int_0^T \left(\tilde{\mathbf{v}} - \boldsymbol{H}(t, \boldsymbol{\xi}_t^*)\right) \mathrm{d}\boldsymbol{X}_t^* - \langle \boldsymbol{X}^*, \boldsymbol{H}(\cdot, \boldsymbol{\xi}_\cdot^*)\rangle_T \middle| \mathscr{F}_0^I\right] \\ &\qquad \geq \mathbb{E}\left[\int_0^T \left(\tilde{\mathbf{v}} - \boldsymbol{H}(t, \boldsymbol{\xi}_t)\right) \mathrm{d}\boldsymbol{X}_t - \langle \boldsymbol{X}, \boldsymbol{H}(\cdot, \boldsymbol{\xi}_\cdot)\rangle_T \middle| \mathscr{F}_0^I\right]. \end{aligned} \tag{3.2}$$

Fix a strategy $\{\boldsymbol{X}_t\}$. We say that a pricing rule $(\boldsymbol{H}, \boldsymbol{\Lambda})$ is *rational* for $\{\boldsymbol{X}_t\}$ if

$$\boldsymbol{H}(t, \boldsymbol{\xi}_t) = \mathbb{E}\left[\tilde{\mathbf{v}} \middle| \mathscr{F}_t^M\right]. \tag{3.3}$$

We call $(\boldsymbol{H}, \boldsymbol{\Lambda}, \boldsymbol{X})$ an *equilibrium* if $\{\boldsymbol{X}\}$ is optimal for $(\boldsymbol{H}, \boldsymbol{\Lambda})$ and $(\boldsymbol{H}, \boldsymbol{\Lambda})$ is rational for $\{\boldsymbol{X}\}$.

We summarize below the relevant quantities and the extent to which they are observable by market participants.

symbol	meaning	relation(s)	insider	market maker
$\tilde{\mathbf{v}}$	terminal price		yes	no
$\mathcal{L}(\tilde{\mathbf{v}})$	law of $\tilde{\mathbf{v}}$	$\mathcal{N}(\mathbf{p}_0, \mathbf{\Sigma}_0)$	yes	yes
$\boldsymbol{X}_t$	informed demand		yes	no
$\boldsymbol{Z}_t$	noise demand		yes	no
$\boldsymbol{Y}_t$	aggregate demand	$\boldsymbol{Y}_t = \boldsymbol{X}_t + \boldsymbol{Z}_t$	yes	yes
$\boldsymbol{\xi}_t$	state variable	$\boldsymbol{\xi}_t = \int_0^t \mathbf{\Lambda}_s \, \mathrm{d}\boldsymbol{Y}_s$	yes	yes
$\theta(t, \mathbf{v}, \boldsymbol{\sigma})$	informed strategy		yes	yes
$\boldsymbol{H}(t, \boldsymbol{\xi})$	pricing rule		yes	yes
$\boldsymbol{P}_t$	price process	$\boldsymbol{P}_t = H(t, \boldsymbol{\xi}_t)$	yes	yes

Table 3.1: Summary of symbols in stochastic liquidity model.

3.2 Equilibrium

3.2.1 The Primal Problem

The equilibrium will be defined from the solution of the maximization problem

$$\max_{\dot{\mathbf{\Sigma}} < \mathbf{0}} \mathbb{E}\left[\int_0^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_t \dot{\mathbf{\Sigma}}_t \boldsymbol{\sigma}_t}\right) \mathrm{d}t\right] \tag{3.4}$$

subject to $\mathbf{\Sigma}(0) = \mathbf{\Sigma}_0$ and $\mathbf{\Sigma}(T) = \mathbf{0}$. In section 3.5, we provide a heuristic derivation that shows how this problem can be derived from a causal optimal transport problem.

The matrix inequality $\dot{\mathbf{\Sigma}} < \mathbf{0}$ is understood in the sense of Loewner ordering: Let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric positive semi-definite matrices. We say that $\mathbf{A} \leq \mathbf{B}$ if $\mathbf{B} - \mathbf{A}$ is symmetric positive semi-definite. The inequality is strict when $\mathbf{B} - \mathbf{A}$ is symmetric positive definite. Writing $\dot{\mathbf{\Sigma}}_t < \mathbf{0}$ as $\mathbf{0} < -\dot{\mathbf{\Sigma}}_t$, we see that $\dot{\mathbf{\Sigma}}_t$ is chosen at each t among all symmetric negative definite matrices $\mathcal{S}^n_{--}$. More precisely, we will take the admissible set to be all $\mathcal{S}^n_{--}$ valued processes $\{\dot{\mathbf{\Sigma}}_t\}$ which are continuous and $\{\mathscr{F}^{\boldsymbol{W}}_t\}$ adapted, and which satisfy, for given $\mathbf{\Sigma}_0$, the matrix identity

$$\mathbf{\Sigma}_0 + \int_0^T \dot{\mathbf{\Sigma}}_t \, \mathrm{d}t = \mathbf{0}.$$

3.2.2 The Dual Problem

In the presence of $\dot{\mathbf{\Sigma}} < \mathbf{0}$, the constraints $\mathbf{\Sigma}(0) = \mathbf{\Sigma}_0$ and $\mathbf{\Sigma}(T) \geq \mathbf{0}$ are equivalent to

$$\mathbf{\Sigma}_0 + \int_0^T \dot{\mathbf{\Sigma}}_s \, \mathrm{d}s \geq \mathbf{0}.$$

We embed this into the objective with the Lagrange multiplier $\mathbf{M}$, chosen among all $\mathcal{S}^n_{++}$ valued random matrices which are $\mathscr{F}_T^{\boldsymbol{W}}$ measurable. This suggests the Lagrangian

$$\mathbb{E}\left[\int_0^T \operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s \boldsymbol{\sigma}_s}\right) \mathrm{d}s + \operatorname{tr}\left(\frac{1}{2}\mathbf{M}\left(\mathbf{\Sigma}_0 + \int_0^T \dot{\mathbf{\Sigma}}_s \, \mathrm{d}s\right)\right)\right]$$

viewed as a map from $\dot{\mathbf{\Sigma}}_{[0,T]}$ (taken from the admissible set described above minus the terminal constraint) and $\mathbf{M} \in \mathcal{S}^n_{++}$ to $\mathbb{R}$. However, we will restrict $\mathbf{M}$ to

$$L^1(\Omega, \mathcal{S}^n_{++}, \|\|_2) = \left\{\mathbf{M} \in \mathcal{S}^n_{++} \colon \mathbb{E}\big[\|\mathbf{M}\|_2\big] < \infty\right\},$$

for then we may define the $\mathcal{S}^n_{++}$ valued martingale $\mathbf{M}_t = \mathbb{E}[\mathbf{M}|\mathscr{F}_t]$. Note by taking closure, we recover the original set, and so the infimum objective values coincide. More precisely, we define the set of admissible martingales by

$$\mathcal{M} = \left\{\mathbf{M}_{[0,T]} \in \left(\mathcal{S}^n_{++}\right)^{[0,T]} : \mathbf{M}_t = \mathbb{E}[\mathbf{M}|\mathscr{F}_t] \text{ for some } \mathbf{M} \in L^1(\Omega, \mathcal{S}^n_{++}, \|\|_2)\right\}. \tag{3.5}$$

Matrix martingality may be understood in the component-wise sense, but this interpretation does not generalize to supermartingales, which we will need later. For this reason, we borrow a standard definition from [WR25]:

Definition 3.2.1. Let $\{\mathbf{M}_t\}$ be a $\{\mathscr{F}_t^{\boldsymbol{W}}\}$ adapted matrix valued process satisfying $\mathbb{E}\big[\|\mathbf{M}_t\|_2\big] < \infty$ for each $t \geq 0$. We say that $\{\mathbf{M}_t\}$ is a *supermartingale* if, for all $s \leq t$,

$$\mathbb{E}\Big[\mathbf{M}_t \Big| \mathscr{F}_s^{\boldsymbol{W}}\Big] \leq \mathbf{M}_s$$

and a *submartingale* if

$$\mathbb{E}\Big[\mathbf{M}_t \Big| \mathscr{F}_s^{\boldsymbol{W}}\Big] \geq \mathbf{M}_s.$$

Here as before, the inequalities are understood in the sense of Loewner ordering. If $\{\mathbf{M}_t\}$ is both a supermartingale and a submartingale, we say that $\{\mathbf{M}_t\}$ is a *martingale*.

As with scalar supermartingales, we will require the Doob-Meyer decomposition:

Proposition 3.2.2 (Matrix-Valued Doob-Meyer Decomposition)**.** *Let* $\{\boldsymbol{X}_t\}$ *be a* $\mathcal{S}^n_{++}$ *valued supermartingale. Then,* $\{\boldsymbol{X}_t\}$ *admits the decomposition*

$$\boldsymbol{X}_t = \mathbf{L}_t + \mathbf{A}_t$$

for some matrix-valued local martingale $\mathbf{L}$ *and some matrix-valued non-increasing process* $\mathbf{A}$ *which are unique up to indistinguishability.*

The reason $\mathbf{L}$ is a local martingale as opposed to a martingale is because we do not assume $\boldsymbol{X}$ is of class DL [KS91].

Proof. [1] To prove uniqueness, suppose $\boldsymbol{X} = \mathbf{L}_1 + \mathbf{A}_1$ and $\boldsymbol{X} = \mathbf{L}_2 + \mathbf{A}_2$, with $\mathbf{L}_1, \mathbf{L}_2$ local martingales and $\mathbf{A}_1, \mathbf{A}_2$ non-increasing processes. Since $\mathbf{A}_1$ and $\mathbf{A}_2$ are increasing, they are finite variation processes. Then, the process $\mathbf{L}_1 - \mathbf{L}_2 = (\boldsymbol{X} - \mathbf{A}_1) - (\boldsymbol{X} - \mathbf{A}_2) = \mathbf{A}_2 - \mathbf{A}_1$ is a predictable, finite variation local martingale. Therefore, $\mathbf{L}_1 - \mathbf{L}_2 = \mathbf{A}_2 - \mathbf{A}_1$ is indistinguishable from the constant $\mathbf{0}$.

To prove existence, denote

$$\boldsymbol{X}^{\mathbf{u},\mathbf{v}} := \mathbf{u}^\top \boldsymbol{X} \mathbf{v} \quad \text{and} \quad \boldsymbol{X}^{\mathbf{u}} := \mathbf{u}^\top \boldsymbol{X} \mathbf{u}$$

for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Since $\boldsymbol{X}$ is symmetric,

$$\boldsymbol{X}^{\mathbf{u},\mathbf{v}} = \frac{1}{2}\left(\boldsymbol{X}^{\mathbf{u}+\mathbf{v}} - \boldsymbol{X}^{\mathbf{u}} - \boldsymbol{X}^{\mathbf{v}}\right).$$

By the definition of matrix-valued supermartingales, $\boldsymbol{X}^{\mathbf{u}}$ is a local supermartingale. By the scalar-valued Doob-Meyer decomposition, there exists a local martingale $\mathbf{L}^{\mathbf{u}}$ and a non-increasing predictable $\mathbf{A}^{\mathbf{u}}$ starting at zero such that $\boldsymbol{X}^{\mathbf{u}} = \mathbf{L}^{\mathbf{u}} + \mathbf{A}^{\mathbf{u}}$. Since

$$\boldsymbol{X}^{\mathbf{u},\mathbf{v}} = \frac{1}{2}\left(\mathbf{L}^{\mathbf{u}+\mathbf{v}} - \mathbf{L}^{\mathbf{u}} - \mathbf{L}^{\mathbf{v}}\right) + \frac{1}{2}\left(\mathbf{A}^{\mathbf{u}+\mathbf{v}} - \mathbf{A}^{\mathbf{u}} - \mathbf{A}\mathbf{v}\right)$$

and since $(\mathbf{u}, \mathbf{v}) \mapsto \boldsymbol{X}^{\mathbf{u},\mathbf{v}}$ is a bilinear map, it follows that

$$\mathbf{A}^{\mathbf{u},\mathbf{v}} := \frac{1}{2}\left(\mathbf{A}^{\mathbf{u}+\mathbf{v}} - \mathbf{A}^{\mathbf{u}} - \mathbf{A}^{\mathbf{v}}\right) \quad \text{and} \quad \mathbf{L}^{\mathbf{u},\mathbf{v}} := \frac{1}{2}\left(\mathbf{L}^{\mathbf{u}+\mathbf{v}} - \mathbf{L}^{\mathbf{u}} - \mathbf{L}^{\mathbf{v}}\right)$$

[1] This proof is due to Evangelos "Vaki" Nikitopoulos.

are also bilinear maps in $(\mathbf{u}, \mathbf{v})$. Indeed, if $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, then the identities

$$\boldsymbol{X}^{\mathbf{u}+\mathbf{w},\mathbf{v}} = \mathbf{L}^{\mathbf{u}+\mathbf{w},\mathbf{v}} + \mathbf{A}^{\mathbf{u}+\mathbf{w},\mathbf{v}} \tag{3.6}$$

$$\boldsymbol{X}^{\mathbf{u},\mathbf{v}} = \mathbf{L}^{\mathbf{u},\mathbf{v}} + \mathbf{A}^{\mathbf{u},\mathbf{v}} \tag{3.7}$$

$$\boldsymbol{X}^{\mathbf{w},\mathbf{v}} = \mathbf{L}^{\mathbf{w},\mathbf{v}} + \mathbf{A}^{\mathbf{w},\mathbf{v}} \tag{3.8}$$

each hold on (possibly different) full measure sets. On their intersection (which is also full measure), we have

$$\mathbf{L}^{\mathbf{u}+\mathbf{w},\mathbf{v}} - \mathbf{L}^{\mathbf{u},\mathbf{v}} - \mathbf{L}^{\mathbf{w},\mathbf{v}} = \underbrace{\left(\boldsymbol{X}^{\mathbf{u}+\mathbf{w},\mathbf{v}} - \boldsymbol{X}^{\mathbf{u},\mathbf{v}} - \boldsymbol{X}^{\mathbf{w},\mathbf{v}}\right)}_{\mathbf{0}} - \left(\mathbf{A}^{\mathbf{u}+\mathbf{w},\mathbf{v}} - \mathbf{A}^{\mathbf{u},\mathbf{v}} - \mathbf{A}^{\mathbf{w},\mathbf{v}}\right),$$

by which we see that $\mathbf{L}^{\mathbf{u}+\mathbf{w},\mathbf{v}} - \mathbf{L}^{\mathbf{u},\mathbf{v}} - \mathbf{L}^{\mathbf{w},\mathbf{v}}$ is a continuous local martingale with finite variation, and thus indistinguishable from the constant matrix $\mathbf{0}$. It follows that $\mathbf{A}^{\mathbf{u}+\mathbf{w},\mathbf{v}} - \mathbf{A}^{\mathbf{u},\mathbf{v}} - \mathbf{A}^{\mathbf{w},\mathbf{v}}$ is also indistinguishable from $\mathbf{0}$. In the same way, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then we have

$$\boldsymbol{X}^{c\mathbf{u},\mathbf{v}} = \mathbf{L}^{c\mathbf{u},\mathbf{v}} + \mathbf{A}^{c\mathbf{u},\mathbf{v}} \tag{3.9}$$

$$\boldsymbol{X}^{\mathbf{u},\mathbf{v}} = \mathbf{L}^{\mathbf{u},\mathbf{v}} + \mathbf{A}^{\mathbf{u},\mathbf{v}} \tag{3.10}$$

on (possibly different) full measure sets, and on their intersection,

$$\mathbf{L}^{c\mathbf{u},\mathbf{v}} - c\mathbf{L}^{\mathbf{u},\mathbf{v}} = \underbrace{\left(\boldsymbol{X}^{c\mathbf{u},\mathbf{v}} - c\boldsymbol{X}^{\mathbf{u},\mathbf{v}}\right)}_{\mathbf{0}} - \left(\mathbf{A}^{c\mathbf{u},\mathbf{v}} - c\mathbf{A}^{\mathbf{u},\mathbf{v}}\right)$$

from which follows $\mathbf{L}^{c\mathbf{u},\mathbf{v}} = c\mathbf{L}^{\mathbf{u},\mathbf{v}}$ and $\mathbf{A}^{c\mathbf{u},\mathbf{v}} = c\mathbf{A}^{\mathbf{u},\mathbf{v}}$. By symmetry, linearity holds in the second argument as well.

Taking the countable intersection over $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{Q}^n$ and $c \in \mathbb{Q}$ of these full measure sets, we obtain a full measure set on which is defined the bilinear maps $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{L}^{\mathbf{u},\mathbf{v}}$ and $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{A}^{\mathbf{u},\mathbf{v}}$. Since bilinear maps determine matrices, there exists matrix valued processes $\mathbf{L}$ and $\mathbf{A}$ for which

$$\mathbf{L}^{\mathbf{u},\mathbf{v}} = \mathbf{u}^\top \mathbf{L}\mathbf{v} \quad \text{and} \quad \mathbf{A}^{\mathbf{u},\mathbf{v}} = \mathbf{u}^\top \mathbf{A}\mathbf{v}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{Q}^n$. By continuity, we can extend this to $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Since $\mathbf{L}^{\mathbf{u}}$ is a local martingale, $\mathbf{L}$ is a matrix-valued local martingale. Since $\mathbf{A}^{\mathbf{u}}$ is non-increasing and starts at 0, $\mathbf{A}$ is non-increasing and starts at $\mathbf{0}$. We have proved existence. □

Definition 3.2.3. Let $\mathcal{A}$ denote all $\mathcal{S}^n_{--}$ valued processes $\{\dot{\boldsymbol{\Sigma}}_t\}$ which are continuous and

$\{\mathscr{F}_t^{\boldsymbol{W}}\}$ adapted, and let $\mathcal{M}$ be as in (3.5). The Lagrangian $L\colon \mathcal{A}\times\mathcal{M}\mapsto\mathbb{R}$ is

$$L(\dot{\boldsymbol{\Sigma}}_{[0,T]},\mathbf{M}_{[0,T]})=\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_0\mathbf{M}_0)+\mathbb{E}\left[\int_0^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t\boldsymbol{\sigma}_t}+\frac{1}{2}\mathbf{M}_t\dot{\boldsymbol{\Sigma}}_t\right)\mathrm{d}t\right]. \tag{3.11}$$

Proposition 3.2.4. *The dual function is*

$$g\big(\mathbf{M}_{[0,T]}\big):=\sup_{\dot{\boldsymbol{\Sigma}}\in\mathcal{A}} L(\dot{\boldsymbol{\Sigma}}_{[0,T]},\mathbf{M}_{[0,T]})=\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_0\mathbf{M}_0)+\frac{1}{2}\mathbb{E}\left[\int_0^T \mathrm{tr}\big(\boldsymbol{\sigma}_t^2\mathbf{M}_t^{-1}\big)\,\mathrm{d}t\right]. \tag{3.12}$$

The quantity $\int_0^T \mathrm{tr}\big(\boldsymbol{\sigma}_t^2\mathbf{M}_t^{-1}\big)\,\mathrm{d}t$ is called *slippage cost* and is discussed in [CDF16].

Proof. Let $\mathbf{M}_{[0,T]}\in\mathcal{M}$ be given. Choose

$$\dot{\boldsymbol{\Sigma}}_t^*=-\mathbf{M}_t^{-1}\boldsymbol{\sigma}_t^2\mathbf{M}_t^{-1},\qquad t\in[0,T] \tag{3.13}$$

and note that $\dot{\boldsymbol{\Sigma}}^*_{[0,T]}\in\mathcal{A}$ because $\dot{\boldsymbol{\Sigma}}_t^*\in\mathcal{S}^n_{--}$ for each t and $\dot{\boldsymbol{\Sigma}}_t^*$ is adapted to $\mathscr{F}_t^{\boldsymbol{W}}$.

Consider an admissible perturbation $\dot{\boldsymbol{\Sigma}}^*+\varepsilon\dot{\boldsymbol{\eta}}$. We will show that (3.13) satisfies the first order condition, which reads

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left[\int_0^T\left(\mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_t\left(\dot{\boldsymbol{\Sigma}}_t^*+\varepsilon\dot{\boldsymbol{\eta}}_t\right)\boldsymbol{\sigma}_t}-\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}+\frac{1}{2}\varepsilon\mathbf{M}_t\dot{\boldsymbol{\eta}}_t\right)\right)\mathrm{d}t\right]=0. \tag{3.14}$$

From [MN18], we have the first order Taylor approximation

$$\sqrt{-\boldsymbol{\sigma}_t\left(\dot{\boldsymbol{\Sigma}}_t^*+\varepsilon\dot{\boldsymbol{\eta}}_t\right)\boldsymbol{\sigma}_t}=\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}-\varepsilon\int_0^\infty e^{-s\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}}\boldsymbol{\sigma}_t\dot{\boldsymbol{\eta}}_t\boldsymbol{\sigma}_t e^{-s\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}}\,\mathrm{d}s+O(\varepsilon^2).$$

With this, the expected value in (3.14) reduces to

$$\mathbb{E}\left[\int_0^T\mathrm{tr}\left(\frac{1}{2}\mathbf{M}_t\dot{\boldsymbol{\eta}}_t-\int_0^\infty e^{-s\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}}\boldsymbol{\sigma}_t\dot{\boldsymbol{\eta}}_t\boldsymbol{\sigma}_t e^{-s\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}}\,\mathrm{d}s\right)\mathrm{d}t\right].$$

We cyclically permute the integrand within the trace to obtain

$$\mathbb{E}\left[\int_0^T \operatorname{tr}\left(\left(\frac{1}{2}\mathbf{M}_t - \boldsymbol{\sigma}_t \int_0^\infty e^{-2y\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}}\,\mathrm{d}y\boldsymbol{\sigma}_t\right)\dot{\boldsymbol{\eta}}_t\right)\mathrm{d}t\right]$$
$$= \mathbb{E}\left[\int_0^T \operatorname{tr}\left(\left(\frac{1}{2}\mathbf{M}_t - \frac{1}{2}\boldsymbol{\sigma}_t\left(\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}\right)^{-1}\boldsymbol{\sigma}_t\right)\dot{\boldsymbol{\eta}}_t\right)\mathrm{d}t\right].$$

For the integrand to vanish, it suffices that

$$\mathbf{M}_t = \boldsymbol{\sigma}_t\left(\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}_t}\right)^{-1}\boldsymbol{\sigma}_t, \qquad (t,\omega)\in[0,T]\times\Omega \quad \text{a.e.}, \tag{3.15}$$

but this is the same as (3.13). Thus (3.14) holds.

It remains to show that (3.13) is a global maximum, and not a saddle point or a local minimum. But this follows from the observation that $L(\dot{\boldsymbol{\Sigma}}_{[0,T]}, \mathbf{M}_{[0,T]})$ is concave in $\dot{\boldsymbol{\Sigma}}_{[0,T]}$ for each $\mathbf{M}_{[0,T]}\in\mathcal{M}$. To show this, take $\dot{\boldsymbol{\Sigma}}^1_{[0,T]}, \dot{\boldsymbol{\Sigma}}^2_{[0,T]}\in\mathcal{A}$ and $\lambda\in[0,1]$. Then, by the trace concavity of square root [WR25, fact 3], we have for each $t\in[0,T]$

$$\operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}_t\left(\lambda\dot{\boldsymbol{\Sigma}}^1_t + (1-\lambda)\dot{\boldsymbol{\Sigma}}^2_t\right)\boldsymbol{\sigma}_t}\right) \geq \lambda\operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}^1_t\boldsymbol{\sigma}_t}\right) + (1-\lambda)\operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}_t\dot{\boldsymbol{\Sigma}}^2_t\boldsymbol{\sigma}_t}\right). \tag{3.16}$$

Substituting this into (3.11), we see that

$$L\left(\lambda\dot{\boldsymbol{\Sigma}}^1_{[0,T]} + (1-\lambda)\dot{\boldsymbol{\Sigma}}^2_{[0,T]}, \mathbf{M}_{[0,T]}\right) \geq \lambda L\left(\dot{\boldsymbol{\Sigma}}^1_{[0,T]}, \mathbf{M}_{[0,T]}\right) + (1-\lambda)L\left(\dot{\boldsymbol{\Sigma}}^2_{[0,T]}, \mathbf{M}_{[0,T]}\right).$$

Finally, substitution of (3.15) into the Lagrangian (3.11) produces the dual function (3.12). □

Proposition 3.2.5. *The dual function g is convex.*

Proof. Let $\mathbf{M}^1_{[0,T]}, \mathbf{M}^2_{[0,T]}\in\mathcal{M}$ and $\lambda\in[0,1]$. Evaluating g at $\lambda\mathbf{M}^1_{[0,T]} + (1-\lambda)\mathbf{M}^2_{[0,T]}$ gives

$$\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}_0\left(\lambda\mathbf{M}^1_0 + (1-\lambda)\mathbf{M}^2_0\right)\right) + \frac{1}{2}\mathbb{E}\left[\int_0^T \operatorname{tr}\left(\boldsymbol{\sigma}_t^2\left[\lambda\mathbf{M}^1_t + (1-\lambda)\mathbf{M}^2_t\right]^{-1}\right)\mathrm{d}t\right]. \tag{3.17}$$

Since $\mathbf{A}\mapsto\operatorname{tr}\left(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}\right)$ is convex, we have

$$\operatorname{tr}\left(\boldsymbol{\sigma}_t\left[\lambda\mathbf{M}^1_t + (1-\lambda)\mathbf{M}^2_t\right]^{-1}\boldsymbol{\sigma}_t\right) \leq \lambda\operatorname{tr}\left(\boldsymbol{\sigma}_t\left[\mathbf{M}^1_t\right]^{-1}\boldsymbol{\sigma}_t\right) + (1-\lambda)\operatorname{tr}\left(\boldsymbol{\sigma}_t\left[\mathbf{M}^2_t\right]^{-1}\boldsymbol{\sigma}_t\right).$$

This means the expected value in (3.17) is bounded above by

$$\lambda\mathbb{E}\left[\int_0^T \operatorname{tr}\Big(\boldsymbol{\sigma}_t^2\left[\mathbf{M}_t^1\right]^{-1}\Big)\,\mathrm{d}t\right] + (1-\lambda)\mathbb{E}\left[\int_0^T \operatorname{tr}\Big(\boldsymbol{\sigma}_t^2\left[\mathbf{M}_t^2\right]^{-1}\Big)\,\mathrm{d}t\right], \tag{3.18}$$

and so

$$g\Big(\lambda\mathbf{M}^1_{[0,T]} + (1-\lambda)\mathbf{M}^2_{[0,T]}\Big) \le \lambda g\Big(\mathbf{M}^1_{[0,T]}\Big) + (1-\lambda)g\Big(\mathbf{M}^2_{[0,T]}\Big).$$

□

Theorem 3.2.6. *The maximization problem* (3.4) *admits the dual formulation*

$$\inf_{\mathbf{M}\in\mathcal{M}} g\big(\mathbf{M}_{[0,T]}\big) \tag{3.19}$$

where g is given by (3.12)*. Furthermore, if $\boldsymbol{\sigma}$ is uniformly bounded from above and below by a constant and $\Sigma_0 > 0$, there exists a unique $\{\mathscr{F}_t^{\boldsymbol{W}}\}$-local martingale optimizer $\mathbf{M}^{*,0,\Sigma_0}$ of* (3.19) *which satisfies $|\mathbf{M}_T^{*,0,\Sigma_0}| < \infty$ a.s.*

Proof. Denote the value in (3.19) as L, and note that $L \ge 0 > -\infty$.

We first prove existence. Let $C > 0$ be such that $\frac{1}{C}\mathbf{I} \le \boldsymbol{\sigma}_t \le C\mathbf{I}$ for all $t \in [0,T]$, almost surely. For the choice of the constant matrix $\mathbf{M}_t = \mathbf{I}$ we see that

$$L \le g(\mathbf{I}) \le \frac{C^2Tn + \operatorname{tr}(\boldsymbol{\Sigma}_0)}{2}.$$

For each k, let $\{\mathbf{M}_t^k\}$ be an $\mathcal{S}_{++}^n$ valued martingale satisfying

$$g(\mathbf{M}^k) \le L + \frac{1}{k} \le \frac{C^2Tn + \operatorname{tr}(\boldsymbol{\Sigma}_0)}{2} + 1.$$

The non-degeneracy of $\boldsymbol{\Sigma}_0$ and the bound on $g(\mathbf{M}^k)$ allows us to claim that $(\mathbf{M}_0^k)_{k\ge 0}$ is bounded. Thus, without loss of generality we can assume that $\mathbf{M}_0^k \to \mathbf{M}_0^\infty \in \mathcal{S}_+^n$ as $k \to \infty$.

We construct a martingale sequence $\{\tilde{\mathbf{M}}^k\}$ from $\{\mathbf{M}^k\}$ with the property that $\{\tilde{\mathbf{M}}^k\}$ converges almost surely to a component-wise limit $\mathbf{M}^\infty$. Denote $\mathbf{e}_i = (0,\dots,0,1,0,\dots,0)^\top$, with 1 being at ith position for $i = 1,\dots,n$. By the martingale representation theorem, there exists some process $\{H_t^{11,k}\}$ for which

$$\mathbf{e}_1^\top\mathbf{M}_0^k\mathbf{e}_1 + \int_0^t H_s^{11,k}\,\mathrm{d}W_s = \mathbf{e}_1^\top\mathbf{M}_t^k\mathbf{e}_1 \ge 0.$$

Thus $\int_0^t H_s^{11,k}\,\mathrm{d}W_s$ is uniformly bounded below by $-\mathbf{e}_1^\top\mathbf{M}_0^k\mathbf{e}_1$. [DS99, 1.10 Theorem D] asserts that there are finite convex combinations $\tilde{H}^{11,k} \in \operatorname{conv}\{H^{11,k}, H^{11,k+1},\dots\}$ and a

super-martingale V_t^{11} for which

$$\lim_{\substack{s\downarrow t \\ s\in\mathbb{Q}_+}} \lim_{k\to\infty} \int_0^s \tilde{H}_r^{11,k}\,\mathrm{d}W_r = V_t^{11} \tag{3.20}$$

$$V_0^{11} \leq 0. \tag{3.21}$$

Applying these same convex combinations to the matrices $\mathbf{M}^k, \mathbf{M}^{k+1}, \ldots$, we build the martingale sequence $\tilde{\mathbf{M}}^k, \tilde{\mathbf{M}}^{k+1} \ldots$. Take this new sequence and repeat the same procedure on the second diagonal term $\mathbf{e}_2^\top \mathbf{M}^k \mathbf{e}_2$ to obtain $\{V_t^{22}\}$. Notice that the limit V_t^{11} of the first diagonal does not change through this procedure.

We continue in this way along all diagonal entries, denoting again the new sequences obtained by $\{\tilde{\mathbf{M}}\}$. For the off-diagonal entries, we consider instead $\left(\mathbf{e}_i + \mathbf{e}_j\right)^\top \tilde{\mathbf{M}}^k \left(\mathbf{e}_i + \mathbf{e}_j\right)$ with $i \neq j$ and we denote these limits V_t^{ij}. Altogether, this process is repeated $n(n+1)/2$ times. Define the matrix process $\{\mathbf{M}_t^\infty\}$ by

$$[\mathbf{M}_t^\infty]_{i,j} = \begin{cases} V_t^{ii} & i = j \\ \frac{1}{2}\left(V_t^{ij} - V_t^{ii} - V_t^{jj}\right) & i \neq j \end{cases}$$

and notice that it coincides with the component-wise limit of $\{\tilde{\mathbf{M}}^k\}$.

Let $\mathbf{x} \in \mathbb{R}^n$ and $s \leq t$. For each k, $\{\tilde{\mathbf{M}}_t^k\}$ is a martingale, and so

$$\mathbb{E}\left[\mathbf{x}^\top \tilde{\mathbf{M}}_t^k \mathbf{x} \middle| \mathscr{F}_s\right] = \mathbf{x}^\top \tilde{\mathbf{M}}_s^k \mathbf{x}.$$

Taking $k \to \infty$, Fatou's lemma gives

$$\mathbb{E}\left[\liminf_k \mathbf{x}^\top \tilde{\mathbf{M}}_t^k \mathbf{x} \middle| \mathscr{F}_s\right] \leq \liminf_k \mathbf{x}^\top \tilde{\mathbf{M}}_s^k \mathbf{x}.$$

But a component-wise limit exists and is $\{\mathbf{M}_t^\infty\}$ and so

$$\mathbb{E}\left[\mathbf{x}^\top \mathbf{M}_t^\infty \mathbf{x} \middle| \mathscr{F}_s\right] \leq \mathbf{x}^\top \mathbf{M}_s^\infty \mathbf{x}. \tag{3.22}$$

Since $\mathbf{x}$ is arbitrary, this means that $\mathbb{E}[\mathbf{M}_t^\infty | \mathscr{F}_s] \leq \mathbf{M}_s^\infty$.

Thus $\{\mathbf{M}_t^\infty\}$ is a supermartingale as defined by definition 3.2.1.

But (3.12) is convex in $\mathbf{M}$, so any supermartingale minimizer may be improved by taking only its local martingale component. By proposition 3.2.2, Doob's decomposition permits the representation $\mathbf{M}_t^\infty = \mathbf{L}_t + \mathbf{A}_t$, for some local martingale $\mathbf{L}_t$ and some nonincreasing process $\mathbf{A}_t$ which starts at $\mathbf{0}$. Since $\mathbf{M}_t^\infty \leq \mathbf{L}_t$ for each t, we have also that $(\mathbf{L}_t)^{-1} \leq (\mathbf{M}_t^\infty)^{-1}$, and

so

$$g\Big(\mathbf{M}^{\infty}_{[0,T]}\Big) = \frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_0\mathbf{M}^{\infty}_0) + \frac{1}{2}\mathbb{E}\left[\int_0^T \mathrm{tr}\Big(\boldsymbol{\sigma}_t^2\left(\mathbf{M}^{\infty}_t\right)^{-1}\Big)\,\mathrm{d}t\right]$$
$$\geq \frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_0\mathbf{M}^{\infty}_0) + \frac{1}{2}\mathbb{E}\left[\int_0^T \mathrm{tr}\Big(\boldsymbol{\sigma}_t^2\left(\mathbf{L}_t\right)^{-1}\Big)\,\mathrm{d}t\right] = g\big(\mathbf{L}_{[0,T]}\big)\,.$$

We now show that $\mathbf{L}$ is an optimizer of (3.19). By convex combination we have that

$$L + \frac{1}{k} \geq g\Big(\mathbf{M}^{k}_{[0,T]}\Big) \geq g\Big(\tilde{\mathbf{M}}^{k}_{[0,T]}\Big)\,.$$

By Fatou's lemma, we have that

$$g(\mathbf{M}^{\infty}) = g(\liminf_k \tilde{\mathbf{M}}^k) \leq \liminf_k L + \frac{1}{k}.$$

Thus we obtain

$$g(\mathbf{L}) \leq g(\mathbf{M}^{\infty}) \leq L$$

which shows the optimality of $\mathbf{L}$.

We now assume that there are two optimizers denote $\mathbf{M}^{*,1}, \mathbf{M}^{*,2}$ and by abuse of notation define the convex function

$$\lambda \in [0,1] \mapsto g(\lambda) := g(\lambda\mathbf{M}^{*,2} + (1-\lambda)\mathbf{M}^{*,1})$$

which by convexity and optimality is constant along $\lambda \in [0,T]$. Thus,

$$0 = g''(0) = \mathbb{E}\left[\int_0^T \mathrm{tr}\left(\boldsymbol{\sigma}_t\left[\mathbf{M}^{*,1}_t\right]^{-1}(\mathbf{M}^{*,1}_t - \mathbf{M}^{*,2}_t)\left[\mathbf{M}^{*,1}_t\right]^{-1}(\mathbf{M}^{*,1}_t - \mathbf{M}^{*,2}_t)\left[\mathbf{M}^{*,1}_t\right]^{-1}\boldsymbol{\sigma}_t\right)\mathrm{d}t\right]$$

which implies $\mathbf{M}^{*,1}_t = \mathbf{M}^{*,2}_t$, $\mathrm{d}t \times \mathrm{d}\mathbb{P}$-a.s. due to the positivity of the matrices involved. □

Corollary 3.2.7. *Strong duality holds in the sense that*

$$\sup_{\dot{\mathbf{\Sigma}}}\inf_{\mathbf{M}} L(\dot{\mathbf{\Sigma}}_{[0,T]}, \mathbf{M}_{[0,T]}) = \inf_{\mathbf{M}}\sup_{\dot{\mathbf{\Sigma}}} L(\dot{\mathbf{\Sigma}}_{[0,T]}, \mathbf{M}_{[0,T]}). \tag{3.23}$$

Proof. We established the existence of a unique local martingale minimizer in proposition 3.2.6. Denote this minimizer $\mathbf{M}^*$. Choose

$$\dot{\mathbf{\Sigma}}^* = [\mathbf{M}^*_t]^{-1}\,\boldsymbol{\sigma}_t^2\,[\mathbf{M}^*_t]^{-1} \tag{3.24}$$

in accordance with (3.13). We will show that $\left(\dot{\mathbf{\Sigma}}^*, \mathbf{M}^*\right)$ is a saddle point of L, which will prove (3.23).

It suffices to show that

$$\nabla_{\dot{\mathbf{\Sigma}}} L\left(\dot{\mathbf{\Sigma}}^*, \mathbf{M}^*\right) = \mathbf{0} \tag{3.25}$$

$$\nabla_{\mathbf{M}} L\left(\dot{\mathbf{\Sigma}}^*, \mathbf{M}^*\right) = \mathbf{0}. \tag{3.26}$$

The identity (3.25) reduces to

$$\mathbf{M}_t^* = \boldsymbol{\sigma}_t \left(\sqrt{-\boldsymbol{\sigma}_t \dot{\mathbf{\Sigma}}_t^* \boldsymbol{\sigma}_t}\right)^{-1} \boldsymbol{\sigma}_t, \qquad (t, \omega) \in [0, T] \times \Omega \quad \text{a.e.}, \tag{3.27}$$

as was shown in the proof of proposition 3.2.4, and this is satisfied by (3.24). On the other hand, (3.26) reduces to

$$\mathbf{\Sigma}_0 + \int_0^T \dot{\mathbf{\Sigma}}_t^* \, \mathrm{d}t = \mathbf{0},$$

which, due to our choice of $\dot{\mathbf{\Sigma}}_t^*$, becomes

$$\mathbf{\Sigma}_0 - \int_0^T \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\sigma}_t^2 \left[\mathbf{M}_t^*\right]^{-1} \mathrm{d}t = \mathbf{0}. \tag{3.28}$$

This is just the statement $\nabla_{\mathbf{M}} g(\mathbf{M}^*) = \mathbf{0}$, where g is the dual function (3.12). By the convexity of g (proposition 3.2.5) and the optimality of $\mathbf{M}^*$, this holds too. □

3.2.3 Main Theorem

Theorem 3.2.8. *Suppose* $\{\boldsymbol{\sigma}_t\}$ *and* $\mathbf{\Sigma}_0$ *are such that the dual problem* (3.19) *admits a martingale minimizer. Denote this minimizer by* $\{\mathbf{M}_t^*\}$*. Then, there exists an equilibrium where the pricing rule is*

$$\mathbf{\Lambda}_t^* = \left[\mathbf{M}_t^*\right]^{-1} \tag{3.29}$$

$$\boldsymbol{H}^*(t, \boldsymbol{\xi}) = \mathbf{p}_0 + \boldsymbol{\xi} \tag{3.30}$$

and the trading strategy is

$$\boldsymbol{X}_t^* = \int_0^t \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathbf{\Sigma}_s^{-1} \left(\tilde{\mathbf{v}} - \boldsymbol{H}^*(s, \boldsymbol{\xi}_s^*)\right) \mathrm{d}s. \tag{3.31}$$

An optimum (to either problem) may or may not exist depending on the data $\{\boldsymbol{\sigma}_t\}$ and

$\boldsymbol{\Sigma}_0$. We discuss in section 3.4 a class of processes $\{\boldsymbol{\sigma}_t\}$ for which the dual problem (3.19) always has a solution.

We defer the proof of theorem 3.2.8 to section 3.2.6. Here, we state a few more properties of the equilibrium.

3.2.4 Properties of the Equilibrium

Proposition 3.2.9. *The equilibrium price process in* (3.31) *is*

$$\boldsymbol{P}_t^* = \boldsymbol{H}^*(s, \boldsymbol{\xi}_s^*) = \mathbb{E}\Big[\tilde{\mathbf{v}}\Big|\mathscr{F}_t^M\Big] = \mathbf{p}_0 + \underbrace{\int_0^t [\mathbf{M}_s^*]^{-1}\,\mathrm{d}\boldsymbol{Y}_s^*}_{\boldsymbol{\xi}_t^*} \tag{3.32}$$

and the market maker's optimal filtering error is

$$\mathrm{Var}\Big(\tilde{\mathbf{v}}\Big|\mathscr{F}_t^M\Big) = \boldsymbol{\Sigma}_t^* = \boldsymbol{\Sigma}_0 - \int_0^t [\mathbf{M}_s^*]^{-1}\,\boldsymbol{\sigma}_s^2\,[\mathbf{M}_s^*]^{-1}\,\mathrm{d}s \tag{3.33}$$

$$= \int_t^T [\mathbf{M}_s^*]^{-1}\,\boldsymbol{\sigma}_s^2\,[\mathbf{M}_s^*]^{-1}\,\mathrm{d}s. \tag{3.34}$$

Notice that the integral (3.34) is $\mathscr{F}_t^M$ measurable, even though it "goes into the future."

Proof. This is an application of the Kalman-Bucy filter [LS13, theorem 12.7]. The hidden process is the static random variable $\tilde{\mathbf{v}}$, and the observable process is the pair $(\boldsymbol{\xi}_t, \boldsymbol{\sigma}_t)$, which by the strategy (3.31) evolves according to

$$\mathrm{d}\boldsymbol{\xi}_t^* = [\mathbf{M}_t^*]^{-1}\,\boldsymbol{\sigma}_t^2\,[\mathbf{M}_t^*]^{-1}\,[\boldsymbol{\Sigma}_t^*]^{-1}\,(\tilde{\mathbf{v}} - \mathbf{p}_0 - \boldsymbol{\xi}_t^*)\,\mathrm{d}t + [\mathbf{M}_t^*]^{-1}\,\boldsymbol{\sigma}_t\,\mathrm{d}\mathbf{B}_t, \quad \boldsymbol{\xi}_0^* = \mathbf{0} \tag{3.35}$$

$$\boldsymbol{\sigma}_t \in \mathscr{F}_t^W \subset \mathscr{F}_t^M. \tag{3.36}$$

Note that

$$\int_0^T \sum_{i,j}^n \Big([\mathbf{M}_t^*]^{-1}\,\boldsymbol{\sigma}_t\Big)_{ij}^2\,\mathrm{d}t = \int_0^T \mathrm{tr}\Big([\mathbf{M}_t^*]^{-1}\,\boldsymbol{\sigma}_t^2\,[\mathbf{M}_t^*]^{-1}\Big)\,\mathrm{d}t$$

$$= -\int_0^T \mathrm{tr}\Big(\dot{\boldsymbol{\Sigma}}_t^*\Big)\,\mathrm{d}t = \mathrm{tr}(\boldsymbol{\Sigma}_0) < \infty$$

The filtering mean $\mathbf{m}_t = \mathbb{E}\big[\tilde{\mathbf{v}}\big|\mathscr{F}_t^M\big]$ satisfies the SDE

$$\mathrm{d}\mathbf{m}_t = \boldsymbol{\gamma}_t\,[\boldsymbol{\Sigma}_t^*]^{-1}\Big(\mathrm{d}\boldsymbol{\xi}_t - [\mathbf{M}_t^*]^{-1}\,\boldsymbol{\sigma}_t^2\,[\mathbf{M}_t^*]^{-1}\,[\boldsymbol{\Sigma}_t^*]^{-1}\,(\mathbf{m}_t - \mathbf{p}_0 - \boldsymbol{\xi}_t^*)\Big) \tag{3.37}$$

$$\mathbf{m}_0 = \mathbf{p}_0, \tag{3.38}$$

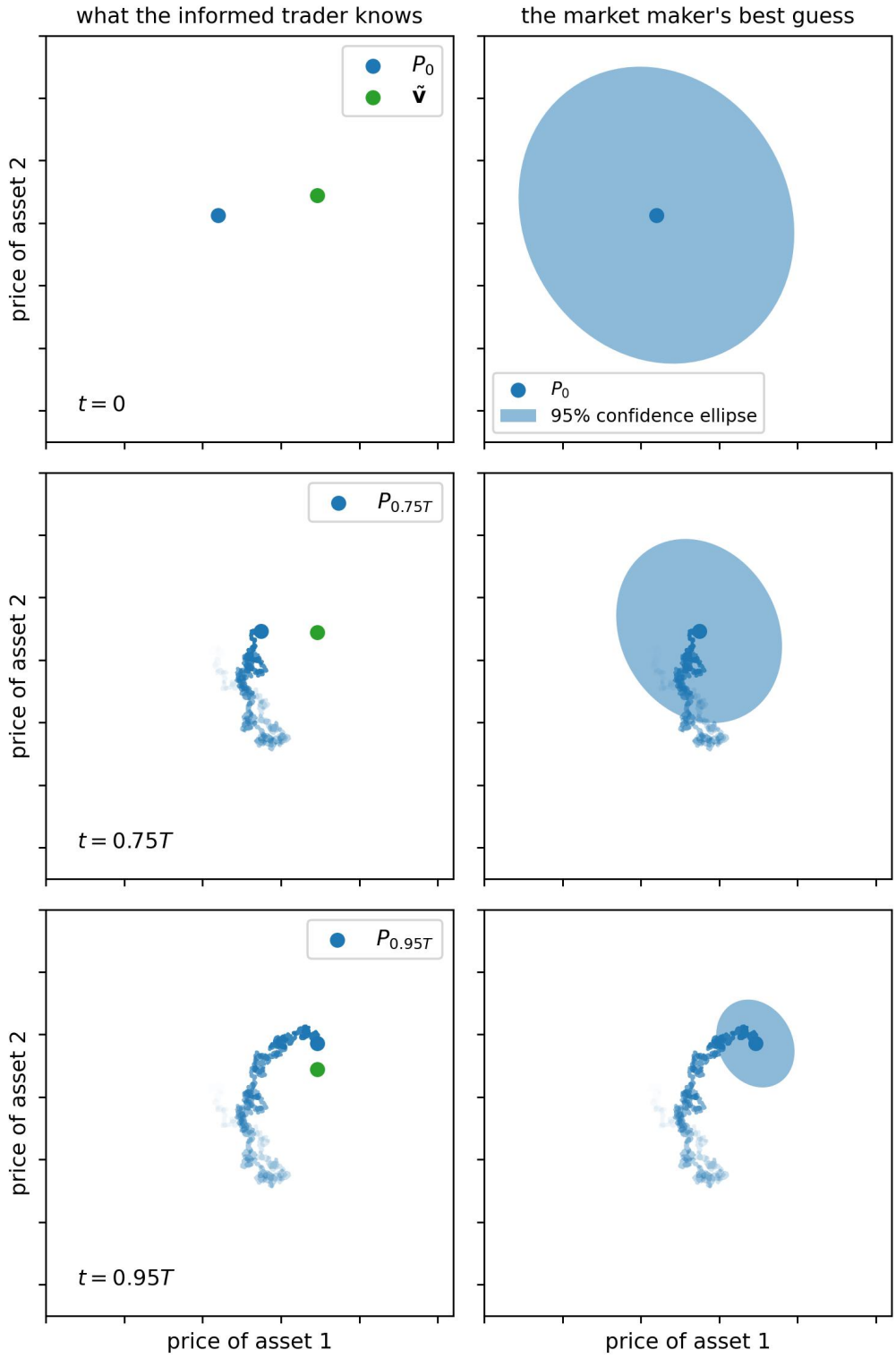


Figure 3.4: Numerical simulation of market maker's filtering problem.

and the filtering variance $\boldsymbol{\gamma}_t = \mathrm{Var}\big(\tilde{\mathbf{v}}\big|\mathscr{F}_t^M\big)$ satisfies the Riccati equation

$$\dot{\boldsymbol{\gamma}}_t = -\boldsymbol{\gamma}_t \left[\boldsymbol{\Sigma}_t^*\right]^{-1} \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\sigma}_t^2 \left[\mathbf{M}_t^*\right]^{-1} \left[\boldsymbol{\Sigma}_t^*\right]^{-1} \boldsymbol{\gamma}_t \tag{3.39}$$

$$\boldsymbol{\gamma}_0 = \boldsymbol{\Sigma}_0. \tag{3.40}$$

Now, in light of (3.13), we see that $\boldsymbol{\gamma}_t = \boldsymbol{\Sigma}_t^*$ is a solution to (3.39). This gives us (3.33). Updating (3.37) with this selection, it becomes clear that $\boldsymbol{P}_t$ as it appears in (3.32) is a solution to (3.37). □

Proposition 3.2.10. *Let* $\{\boldsymbol{\xi}_t^*\}$ *and* $\{\mathbf{M}_t^*\}$ *be defined as in* (3.29) *and* (3.32)*. Then,*

$$\hat{\mathbf{B}}_t = \int_0^t \boldsymbol{\sigma}_s^{-1} \mathbf{M}_s^* \, \mathrm{d}\boldsymbol{\xi}_s^* \tag{3.41}$$

is an n*-dimensional Brownian motion with respect to the market maker's filtration* $\{\mathscr{F}_t^M\}$*.*

Proof. We invoke Levy's characterization. First, observe that (3.41) is continuous. Next, compute

$$\begin{aligned} \mathrm{d}\langle \hat{\mathbf{B}} \rangle_t &= \boldsymbol{\sigma}_t^{-1} \mathbf{M}_t^* \, \mathrm{d}\langle \boldsymbol{\xi}^* \rangle_t \mathbf{M}_t^* \boldsymbol{\sigma}_t^{-1} \\ &= \boldsymbol{\sigma}_t^{-1} \mathbf{M}_t^* \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\sigma}_t \, \mathrm{d}\langle \mathbf{B} \rangle_t \boldsymbol{\sigma}_t \left[\mathbf{M}_t^*\right]^{-1} \mathbf{M}_t^* \boldsymbol{\sigma}_t^{-1} = \mathbf{I} \, \mathrm{d}t. \end{aligned}$$

Finally, note that the processes $\{\boldsymbol{\sigma}_t\}$, $\{\mathbf{M}_t^*\}$, and $\{\boldsymbol{\xi}_t^*\}$ are all $\{\mathscr{F}_t^M\}$ adapted. □

Proposition 3.2.11. *Conditioned on* $\mathscr{F}_t^M$*,* $\boldsymbol{\xi}_T^*$ *is Gaussian with mean* $\boldsymbol{\xi}_t^*$ *and variance*

$$\boldsymbol{\Sigma}_t^* = \boldsymbol{\Sigma}_0 - \int_0^t \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathrm{d}s = \int_t^T \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathrm{d}s.$$

Proof. Fix $\mathbf{u} \in \mathbb{R}^n$. By Itô's lemma, we have

$$\mathrm{d}e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} = ie^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} \mathbf{u}^\top \, \mathrm{d}\boldsymbol{\xi}_t^* - \frac{1}{2} e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} \mathbf{u}^\top \, \mathrm{d}\langle \boldsymbol{\xi}^* \rangle_t \mathbf{u},$$

or, in light of (3.2.10),

$$\mathrm{d}e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} = ie^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} \mathbf{u}^\top \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\sigma}_t \, \mathrm{d}\hat{\mathbf{B}}_t - \frac{1}{2} e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} \mathbf{u}^\top \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\sigma}_t^2 \left[\mathbf{M}_t^*\right]^{-1} \mathbf{u} \, \mathrm{d}t.$$

We integrate from t to T to obtain

$$e^{i\mathbf{u}^\top \boldsymbol{\xi}_T^*} - e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} = \int_t^T ie^{i\mathbf{u}^\top \boldsymbol{\xi}_s^*} \mathbf{u}^\top \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s \, \mathrm{d}\hat{\mathbf{B}}_s - \frac{1}{2} \int_t^T e^{i\mathbf{u}^\top \boldsymbol{\xi}_s^*} \mathbf{u}^\top \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathbf{u} \, \mathrm{d}s.$$

Applying $\mathscr{F}_t^M$ conditional expectation, the $\mathrm{d}\hat{\mathbf{B}}$ integral vanishes because

$$\mathbb{E}\left[\int_0^T \mathbf{u}^\top \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathbf{u}\, \mathrm{d}s\right] = \mathbf{u}^\top \boldsymbol{\Sigma}_0 \mathbf{u} < \infty,$$

and we are left with

$$\mathbb{E}\left[e^{i\mathbf{u}^\top \boldsymbol{\xi}_T^*} \middle| \mathscr{F}_t^M\right] - e^{i\mathbf{u}^\top \boldsymbol{\xi}_t^*} = -\frac{1}{2}\int_t^T \mathbb{E}\left[e^{i\mathbf{u}^\top \boldsymbol{\xi}_s^*} \middle| \mathscr{F}_t^M\right] \mathbf{u}^\top \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathbf{u}\, \mathrm{d}s.$$

This is an integral equation whose solution gives

$$\mathbb{E}\left[e^{i\mathbf{u}^\top \boldsymbol{\xi}_T^*} \middle| \mathscr{F}_t^M\right] = \exp\left(i\mathbf{u}^\top \boldsymbol{\xi}_t^* - \frac{1}{2}\mathbf{u}^\top \left(\int_t^T \left[\mathbf{M}_s^*\right]^{-1} \boldsymbol{\sigma}_s^2 \left[\mathbf{M}_s^*\right]^{-1} \mathrm{d}s\right) \mathbf{u}\right).$$

□

3.2.5 The Wealth Decomposition

Definition 3.2.12. A trading strategy $\{\boldsymbol{X}\}$ is said to be *inconspicuous* provided that it is absolutely continous and that, for all $t \in [0, T]$,

$$\mathbb{E}\left[\frac{\mathrm{d}\boldsymbol{X}_t}{\mathrm{d}t} \middle| \mathscr{F}_t^M\right] = 0.$$

Proposition 3.2.13. *The strategy* (3.31) *is inconspicuous.*

Proof. Clearly (3.31) is absolutely continuous. Next, compute

$$\begin{aligned}\mathbb{E}\left[\frac{\mathrm{d}\boldsymbol{X}_t^*}{\mathrm{d}t} \middle| \mathscr{F}_t^M\right] &= \mathbb{E}\left[\boldsymbol{\sigma}_t^2 \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\Sigma}_t^{-1} \left(\tilde{\mathbf{v}} - \boldsymbol{P}_t^*\right) \middle| \mathscr{F}_t^M\right] \\ &= \boldsymbol{\sigma}_t^2 \left[\mathbf{M}_t^*\right]^{-1} \boldsymbol{\Sigma}_t^{-1} \left(\mathbb{E}\left[\tilde{\mathbf{v}} \middle| \mathscr{F}_t^M\right] - \boldsymbol{P}_t^*\right) = 0.\end{aligned}$$

The quantity in the parenthesis vanishes due to (3.32). □

Definition 3.2.14. Define $V\colon \mathbb{R}^+ \times \mathcal{S}_{++}^n \mapsto \mathbb{R}$ the dynamic programming value of (3.4) by

$$V(t, \mathbf{C}) = \sup_{\dot{\boldsymbol{\Sigma}} \leq \mathbf{0}} \left\{\mathbb{E}\left[\int_t^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\boldsymbol{\Sigma}}_s \boldsymbol{\sigma}_s}\right) \mathrm{d}s \middle| \mathscr{F}_t^{\boldsymbol{W}}\right] : \mathbf{C} + \int_t^T \dot{\boldsymbol{\Sigma}}_s \,\mathrm{d}s = 0\right\}. \tag{3.42}$$

Lemma 3.2.15. *Suppose there exists a maximizer* $\{\dot{\mathbf{\Sigma}}_t^*\}$ *to the primal problem* (3.4)*, and denote by* $\{\mathbf{\Sigma}_t^*\}$ *its optimal trajectory. Then,* $V(t, \mathbf{\Sigma}_t^*)$ *admits the dynamics*

$$\mathrm{d}V(t, \mathbf{\Sigma}_t^*) = -\mathrm{tr}\big(\boldsymbol{\sigma}_t [\mathbf{M}_t^*]^{-1} \boldsymbol{\sigma}_t\big)\,\mathrm{d}t + \mathbf{R}(t, \mathbf{\Sigma}_t^*)^\top\,\mathrm{d}\boldsymbol{W}_t \tag{3.43}$$

for some adapted square integrable random field $\{\mathbf{R}(t, \mathbf{A})\}$*.*

Proof. If $\{\mathbf{\Sigma}_t^*\}$ denotes the optimal trajectory, then we have from (3.42) the identity

$$V(t, \mathbf{\Sigma}_t^*) = \mathbb{E}\left[\int_t^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s \,\middle|\, \mathscr{F}_t^{\boldsymbol{W}}\right].$$

We add to both sides the $\mathscr{F}_t^{\boldsymbol{W}}$ measurable quantity $\int_0^t \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s$ and thereby obtain

$$V(t, \mathbf{\Sigma}_t^*) + \int_0^t \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s = \mathbb{E}\left[\int_0^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s \,\middle|\, \mathscr{F}_t^{\boldsymbol{W}}\right]. \tag{3.44}$$

The RHS is a martingale and thus admits the representation

$$\mathbb{E}\left[\int_0^T \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s \,\middle|\, \mathscr{F}_t^{\boldsymbol{W}}\right] = \int_0^t \mathbf{R}(s, \mathbf{\Sigma}_s^*)^\top\,\mathrm{d}\boldsymbol{W}_s$$

for some adapted square integrable $\{\mathbf{R}(t, \mathbf{A})\}$. The integral which appears on the RHS may be written instead as

$$\int_0^t \mathrm{tr}\left(\sqrt{-\boldsymbol{\sigma}_s \dot{\mathbf{\Sigma}}_s^* \boldsymbol{\sigma}_s}\right)\mathrm{d}s = \int_0^t \mathrm{tr}\big(\boldsymbol{\sigma}_s [\mathbf{M}_s^*]^{-1} \boldsymbol{\sigma}_s\big)\,\mathrm{d}s$$

due to the relation (3.13). Thus, (3.43) is just the identity (3.44). □

Theorem 3.2.16. *Suppose there exists a martingale minimizer* $\{\mathbf{M}_t^*\}$ *to the dual problem* (3.19)*. Let* $\{\dot{\mathbf{\Sigma}}_t^*\}$ *be defined as in* (3.25)*. Define the random potentials* $\varphi\colon \Omega \times \mathbb{R}^+ \times \mathbb{R}^n \mapsto \mathbb{R}$ *and* $\psi\colon \Omega \times \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ *by*

$$\varphi(\omega, t, \boldsymbol{\xi}) = \frac{(\boldsymbol{\xi} + \mathbf{p}_0)^\intercal\, \mathbf{M}_t^*(\omega)\,(\boldsymbol{\xi} + \mathbf{p}_0)}{2} + \frac{1}{2} V(t, \mathbf{\Sigma}_t^*) \tag{3.45}$$

$$\psi^{\tilde{\mathbf{v}}}(\omega, t, \mathbf{p}) = \frac{(\tilde{\mathbf{v}} - \mathbf{p})^\intercal\, \mathbf{M}_t^*(\omega)\,(\tilde{\mathbf{v}} - \mathbf{p})}{2} + \frac{1}{2} V(t, \mathbf{\Sigma}_t^*). \tag{3.46}$$

If the market maker sets prices by (3.29)-(3.30)*, then the informed trader's* $\tilde{\mathbf{v}}$ *conditional*

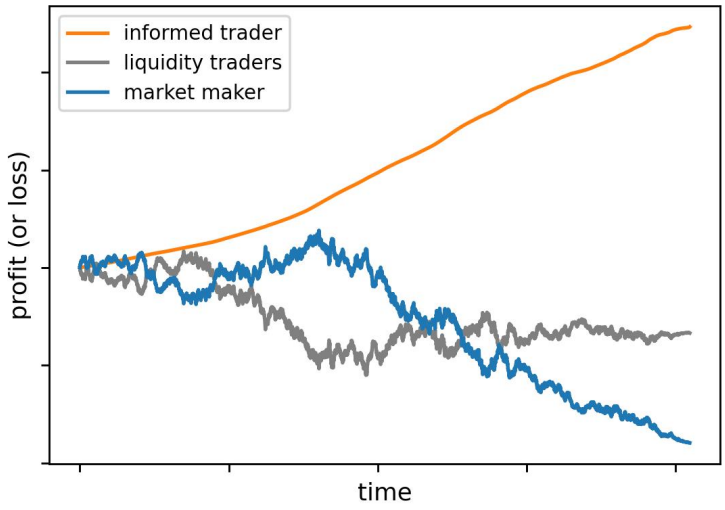


Figure 3.5: The informed trader's profit.

expected profit is at most

$$\mathbb{E}\left[\int_0^T (\tilde{\mathbf{v}} - \boldsymbol{P}_t)^\intercal \,\mathrm{d}\boldsymbol{X}_t - \langle \boldsymbol{X}, \boldsymbol{P}\rangle_T \middle| \mathscr{F}_0^I\right] \leq \mathbb{E}\left[\psi^{\tilde{\mathbf{v}}}(0, \mathbf{p}_0) \middle| \mathscr{F}_0^I\right] \tag{3.47}$$

from time 0*, and in general at most*

$$\mathbb{E}\left[\int_t^T (\tilde{\mathbf{v}} - \boldsymbol{P}_s)^\intercal \,\mathrm{d}\boldsymbol{X}_s - \langle \boldsymbol{X}, \boldsymbol{P}\rangle_T + \langle \boldsymbol{X}, \boldsymbol{P}\rangle_t \middle| \mathscr{F}_t^I\right] \leq \mathbb{E}\left[\psi^{\tilde{\mathbf{v}}}(t, \boldsymbol{P}_t) \middle| \mathscr{F}_t^I\right] \tag{3.48}$$

from time $t \in [0, T]$.

Furthermore, these upper bounds are attained when $\boldsymbol{X}_t$ *is inconspicuous and ensures* $\boldsymbol{\xi}_T = \tilde{\mathbf{v}} - \mathbf{p}_0$ *a.s.*

Proof. Let $\boldsymbol{Y}_t = \boldsymbol{X}_t + \boldsymbol{Z}_t$ denote the order flow from any admissible strategy $\boldsymbol{X}_t$. By Itô's lemma, we have

$$\begin{aligned} \mathrm{d}\varphi(t, \boldsymbol{\xi}_t) = (\boldsymbol{\xi}_t + \mathbf{p}_0)^\top \mathbf{M}_t^* \,\mathrm{d}\boldsymbol{\xi}_t + \frac{1}{2}(\boldsymbol{\xi}_t + \mathbf{p}_0)^\top \mathrm{d}\mathbf{M}_t^* (\boldsymbol{\xi}_t + \mathbf{p}_0) \\ + \frac{1}{2}\mathrm{d}\boldsymbol{\xi}_t^\top \mathbf{M}_t^* \,\mathrm{d}\boldsymbol{\xi}_t + (\boldsymbol{\xi}_t + \mathbf{p}_0)^\top \mathrm{d}\mathbf{M}_t^* \,\mathrm{d}\boldsymbol{\xi}_t + \frac{1}{2}\mathrm{d}V(t, \boldsymbol{\Sigma}_t^*). \end{aligned} \tag{3.49}$$

Observe that $\mathrm{d}\mathbf{M}_t^* \,\mathrm{d}\boldsymbol{\xi}_t = \mathbf{0}$ because $\mathrm{d}\mathbf{M}_t^*$ and $\mathrm{d}\boldsymbol{\xi}_t$ are comprised of independent Brownian differentials ($\mathrm{d}W_t$ for $\mathbf{M}^*$, $\mathrm{d}B_t$ for $\boldsymbol{\xi}$). In light of (3.29)-(3.30), we may replace $\boldsymbol{\xi}_t + \mathbf{p}_0$ by the price process $\boldsymbol{P}_t$. We have $\mathbf{M}_t^* \,\mathrm{d}\boldsymbol{\xi}_t = \mathrm{d}\boldsymbol{Y}_t$ due to (3.32). These considerations taken into

account, (3.49) updates to

$$\mathrm{d}\varphi(t,\boldsymbol{\xi}_t) = \boldsymbol{P}_t^\top \,\mathrm{d}\boldsymbol{Y}_t + \frac{1}{2}\boldsymbol{P}_t^\top \,\mathrm{d}\mathbf{M}_t^*\boldsymbol{P}_t + \frac{1}{2}\,\mathrm{d}\boldsymbol{Y}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Y}_t + \frac{1}{2}\,\mathrm{d}V(t,\boldsymbol{\Sigma}_t^*). \tag{3.50}$$

The third RHS term is

$$\begin{aligned}
\mathrm{d}\boldsymbol{Y}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Y}_t &= (\mathrm{d}\boldsymbol{X}_t + \mathrm{d}\boldsymbol{Z}_t)^\top \boldsymbol{\Lambda}_t^*\,(\mathrm{d}\boldsymbol{X}_t + \mathrm{d}\boldsymbol{Z}_t)\\
&= \mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t + 2\,\mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Z}_t + \mathrm{d}\boldsymbol{Z}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Z}_t\\
&= \mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t + 2\,\mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,(\mathrm{d}\boldsymbol{Y}_t - \mathrm{d}\boldsymbol{X}_t) + \mathrm{d}\boldsymbol{Z}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Z}_t\\
&= -\,\mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t + 2\,\mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Y}_t + \mathrm{d}\boldsymbol{Z}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{Z}_t\\
&= -\,\mathrm{d}\boldsymbol{X}_t^\top \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t + 2\,\mathrm{d}\langle \boldsymbol{X}, \boldsymbol{P}\rangle_t + \mathrm{tr}(\boldsymbol{\sigma}_t\boldsymbol{\Lambda}_t^*\boldsymbol{\sigma}_t)\,\mathrm{d}t,
\end{aligned}$$

so again (3.50) becomes, with the help of lemma (3.2.15),

$$\mathrm{d}\varphi(t,\boldsymbol{\xi}_t) = \boldsymbol{P}_t^\top \,\mathrm{d}\boldsymbol{Y}_t + \frac{1}{2}\boldsymbol{P}_t^\top \,\mathrm{d}\mathbf{M}_t^*\boldsymbol{P}_t - \frac{1}{2}\,\mathrm{d}\boldsymbol{X}_t\boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t + \mathrm{d}\langle \boldsymbol{X}, \boldsymbol{P}\rangle_t + \frac{1}{2}\mathbf{R}_t^\top \,\mathrm{d}\boldsymbol{W}_t. \tag{3.51}$$

Similar calculations with $\tilde{\mathbf{v}}^\top \mathbf{M}_t\,(\boldsymbol{\xi}_t + \mathbf{p}_0)$ produces

$$\mathrm{d}\Big(\tilde{\mathbf{v}}^\top \mathbf{M}_t\,(\boldsymbol{\xi}_t + \mathbf{p}_0)\Big) = \tilde{\mathbf{v}}^\top \,\mathrm{d}\mathbf{M}_t^*\boldsymbol{P}_t + \tilde{\mathbf{v}}^\top \,\mathrm{d}\boldsymbol{Y}_t. \tag{3.52}$$

Subtracting (3.51) from (3.52), we find

$$\begin{aligned}
(\tilde{\mathbf{v}} - \boldsymbol{P}_t)^\intercal \,\mathrm{d}\boldsymbol{Y}_t - \mathrm{d}\langle \boldsymbol{X}, \boldsymbol{P}\rangle_t = \mathrm{d}\Big(\tilde{\mathbf{v}}^\top \mathbf{M}_t\,(\boldsymbol{\xi}_t + \mathbf{p}_0) - \varphi(t,\boldsymbol{\xi}_t)\Big) + \frac{1}{2}\mathbf{R}_t^\top \,\mathrm{d}\boldsymbol{W}_t\\
+ \left(\frac{1}{2}\boldsymbol{P}_t - \tilde{\mathbf{v}}\right)^\intercal \mathrm{d}\mathbf{M}_t^*\boldsymbol{P}_t^* - \frac{1}{2}\,\mathrm{d}\boldsymbol{X}_t^\intercal \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t
\end{aligned} \tag{3.53}$$

If we integrate (3.53) over $t \in [0,T]$ and apply $\mathscr{F}_0^I$ conditional expectation, $\mathrm{d}\boldsymbol{Y}_t = \mathrm{d}\boldsymbol{X}_t + \mathrm{d}\boldsymbol{Z}_t$ reduces to $\mathrm{d}\boldsymbol{X}_t$ on the LHS while the $\mathrm{d}\mathbf{M}^*$ integral vanishes from the RHS because $\mathbf{M}^*$ is a martingale. We are left with

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T (\tilde{\mathbf{v}} - \boldsymbol{P}_t)^\intercal \,\mathrm{d}\boldsymbol{X}_t - \langle \boldsymbol{X}, \boldsymbol{P}\rangle_T \middle| \mathscr{F}_0^I\right] =\\
&\qquad \mathbb{E}\left[\tilde{\mathbf{v}}^\top \mathbf{M}_T\,(\boldsymbol{\xi}_T + \mathbf{p}_0) - \varphi(T,\boldsymbol{\xi}_T) - \tilde{\mathbf{v}}^\top \mathbf{M}_0\mathbf{p}_0 + \varphi(0,\mathbf{0}) - \int_0^T \frac{1}{2}\,\mathrm{d}\boldsymbol{X}_t^\intercal \boldsymbol{\Lambda}_t^*\,\mathrm{d}\boldsymbol{X}_t \middle| \mathscr{F}_0^I\right].
\end{aligned} \tag{3.54}$$

Since $\varphi(T,\boldsymbol{\xi}_T)$ is quadratic in $\boldsymbol{\xi}_T$, the RHS terms which involve $\boldsymbol{\xi}_T$ achieve their maximum

at $\boldsymbol{\xi}_T = \tilde{\mathbf{v}} - \mathbf{p}_0$. Furthermore, $-\int_0^T \mathrm{d}\boldsymbol{X}_t^\intercal \boldsymbol{\Lambda}_t^* \, \mathrm{d}\boldsymbol{X}_t$ is at most 0, and this occurs when $\boldsymbol{X}_t$ is absolutely continuous. Altogether, we have the upper bound

$$\mathbb{E}\left[\int_0^T (\tilde{\mathbf{v}} - \boldsymbol{P}_t)^\intercal \, \mathrm{d}\boldsymbol{X}_t - \langle \boldsymbol{X}, \boldsymbol{P} \rangle_T \middle| \mathscr{F}_0^I \right] \leq \mathbb{E}\left[\frac{(\tilde{\mathbf{v}} - \mathbf{p}_0)^\intercal \, \mathbf{M}_0^* \, (\tilde{\mathbf{v}} - \mathbf{p}_0)}{2} + V(0, \boldsymbol{\Sigma}_0) \middle| \mathscr{F}_0^I \right] \tag{3.55}$$

which is (3.47). To obtain (3.48), integrate (3.53) over $[t, T]$ instead and apply $\mathscr{F}_t^I$-conditional expectations. □

This result generalizes [EMZ22, theorem 2.1] to multiple dimensions when $\tilde{\mathbf{v}}$ is Gaussian, but it should also be compared to [CDF16, theorem 1].

3.2.6 Proof of Main Theorem

Proof. First, fix the trading strategy (3.31). By proposition 3.2.9, (3.29) and (3.30) are rational in the sense of definition 3.1.3. Furthermore, notice that $\mathbf{p}_0 + \boldsymbol{\xi} = \nabla_{\boldsymbol{\xi}} \left(\mathbf{p}_0 \top \boldsymbol{\xi} + \frac{1}{2} \boldsymbol{\xi}^\top \boldsymbol{\xi} \right)$, so (3.30) coincides with the gradient of a convex function. To check the integrability assumptions of definition 3.1.1, note that

$$\begin{aligned} \mathbb{E}\left[\left\| \boldsymbol{H}(T, \boldsymbol{\xi}_T^*) \right\|^2 \middle| \mathscr{F}_0^M \right] &= \mathbb{E}\left[\left\| \mathbf{p}_0 + \boldsymbol{\xi}_T^* \right\|^2 \middle| \mathscr{F}_0^M \right] \\ &= \left\| \mathbf{p}_0 \right\|^2 + \operatorname{tr}\left(\operatorname{Var}\left(\boldsymbol{\xi}_T^* \middle| \mathscr{F}_0^M \right) \right) = \left\| \mathbf{p}_0 \right\|^2 + \operatorname{tr}(\boldsymbol{\Sigma}_0) < \infty. \end{aligned}$$

The last equality follows from proposition 3.2.11.

Similarly,

$$\mathbb{E}\left[\int_0^T \left\| \boldsymbol{H}(t, \boldsymbol{\xi}_t^*) \right\|^2 \mathrm{d}t \middle| \mathscr{F}_0^M \right] = \mathbb{E}\left[\int_0^T \left\| \mathbf{p}_0 + \boldsymbol{\xi}_t^* \right\|^2 \mathrm{d}t \middle| \mathscr{F}_0^M \right]. \tag{3.56}$$

Appealing again to proposition 3.2.11 and Jensen's inequality,

$$\left\| \mathbf{p}_0 + \boldsymbol{\xi}_t^* \right\|^2 \leq \mathbb{E}\left[\left\| \mathbf{p}_0 + \boldsymbol{\xi}_T^* \right\|^2 \middle| \mathscr{F}_t^M \right]$$

and so (3.56) is bounded above by

$$\begin{aligned} \mathbb{E}\left[\int_0^T \mathbb{E}\left[\left\| \mathbf{p}_0 + \boldsymbol{\xi}_T^* \right\|^2 \middle| \mathscr{F}_t^M \right] \mathrm{d}t \middle| \mathscr{F}_0^M \right] &= T \mathbb{E}\left[\left\| \mathbf{p}_0 + \boldsymbol{\xi}_T^* \right\|^2 \middle| \mathscr{F}_0^M \right] \\ &= T \left(\left\| \mathbf{p}_0 \right\|^2 + \operatorname{tr}(\boldsymbol{\Sigma}_0) \right) < \infty \end{aligned}$$

Thus, we have established that for the fixed strategy (3.31), the pricing rule (3.29)-(3.30) satisfies the requirements of definition 3.1.1 and is rational in the sense of definition 3.1.3.

□

3.3 Examples

3.3.1 Kyle 1985

Fix $\sigma > 0$ constant and take $T = 1$. The primal problem (3.4) reads

$$\max_{\dot{\Sigma} \leq 0} \int_0^1 \sigma \sqrt{-\dot{\Sigma}(t)}\, \mathrm{d}t, \tag{3.57}$$

where the maximization is taken over all differentiable functions $\Sigma(t)$ satisfying $\Sigma(0) = \Sigma_0$ and $\Sigma(1) = 0$.

We turn to standard calculus of variations. The optimal Σ^* if it exists satisfies the Euler-Lagrange equation

$$\left(-\dot{\Sigma}^*(t)\right)^{-1/2} = \text{constant} \tag{3.58}$$

and the particular solution which satisfies $\Sigma^*(0) = \Sigma_0$ and $\Sigma^*(1) = 0$ is

$$\Sigma^*(t) = \Sigma_0 \left(1 - t\right). \tag{3.59}$$

By (3.27), the solution to the dual problem (3.19) is

$$M_t^* = \frac{\sigma}{\sqrt{\Sigma_0}}. \tag{3.60}$$

This is clearly a martingale. Since a martingale minimizer exists, our main theorem may be invoked. It tells us that (3.59) is the filtering variance and that (3.60) is the market depth corresponding to some equilibrium. This equilibrium has the pricing rule

$$\lambda_t^* = \frac{\sqrt{\Sigma_0}}{\sigma} \tag{3.61}$$

$$H^*(t, \xi) = p_0 + \xi \tag{3.62}$$

as given by (3.29) and (3.30). The price process (3.32) is

$$P_t^* = p_0 + \frac{\sqrt{\Sigma_0}}{\sigma} Y_t,$$

and the insider's trading strategy (3.31) is

$$\mathrm{d}X_t^* = \frac{\sigma}{\sqrt{\Sigma_0}} \frac{\tilde{v} - P_t^*}{1-t}\,\mathrm{d}t.$$

Thus, we recover the equilibrium from [Kyl85].

3.3.2 Back and Pedersen 1998, static information

Let $\sigma(t)$ be time dependent but deterministic. The primal problem reads

$$\max_{\dot{\Sigma}\leq 0} \int_0^T \sigma(t)\sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t. \tag{3.63}$$

The optimal Σ^* if it exists satisfies the Euler-Lagrange equation

$$\sigma(t)\left(-\dot{\Sigma}^*(t)\right)^{-1/2} = \text{constant}. \tag{3.64}$$

Imposing the conditions $\Sigma^*(0) = \Sigma_0$ and $\Sigma^*(T) = 0$, the solution is

$$\Sigma^*(t) = \Sigma_0 \frac{\int_t^T \sigma(s)^2\,\mathrm{d}s}{\int_0^T \sigma(s)^2\,\mathrm{d}s}. \tag{3.65}$$

Thus, the optimal $\dot{\Sigma}^*$ is

$$\dot{\Sigma}^*(t) = -\Sigma_0 \frac{\sigma(t)^2}{\int_0^T \sigma(s)^2\,\mathrm{d}s}, \tag{3.66}$$

and the optimal M^* by (3.27) is

$$M^* = \sqrt{\frac{\int_0^T \sigma(s)^2\,\mathrm{d}s}{\Sigma_0}}. \tag{3.67}$$

Since M^* is constant, and thus a martingale, theorem 3.2.8 asserts that an equilibrium exists. The equilibrium trading strategy (3.31) is

$$\mathrm{d}X_t^* = \sqrt{\frac{\int_0^T \sigma(s)^2\,\mathrm{d}s}{\Sigma_0}}\sigma(t)^2 \frac{\tilde{v} - P_t^*}{\int_t^T \sigma(s)^2\,\mathrm{d}s}\,\mathrm{d}t \tag{3.68}$$

and the equilibrium price process (3.32) is

$$P_t^* = p_0 + \sqrt{\frac{\Sigma_0}{\int_0^T \sigma(s)^2 \, \mathrm{d}s}} Y_s^*. \tag{3.69}$$

These coincide with equations (8) and (9) of [BP98] when $\sigma_S(t) = 0$ (in their notation). Thus we recover their equilibrium in the case of static information.

3.3.3 Back, Cocquemas, Ekren, and Lioui 2021

Fix $\boldsymbol{\sigma} > 0$. The primal problem (3.4) reads

$$\max_{\dot{\boldsymbol{\Sigma}} \leq 0} \int_0^T \operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}(t)\boldsymbol{\sigma}}\right) \mathrm{d}t \tag{3.70}$$

where the maximization is taken over all differentiable matrix functions satisfying $\boldsymbol{\Sigma}(0) = \boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}(T) = \mathbf{0}$.

Proposition 3.3.1. *The solution to* (3.70) *is*

$$\boldsymbol{\Sigma}^*(t) = \boldsymbol{\Sigma}_0 \left(1 - \frac{t}{T}\right). \tag{3.71}$$

Note that (3.71) is the matrix analog of (3.59). We state it as a proposition because its proof may not be trivial to the reader.

Proof. Let $\boldsymbol{\Sigma}^*$ be an as in (3.71). Fix a perturbation direction $\boldsymbol{\eta} \in C^1\left([0,T], \mathcal{S}_{++}^n\right)$ so that the perturbed optimum $\boldsymbol{\Sigma}^* + \varepsilon\boldsymbol{\eta}$ is admissible for all ε in some neighborhood of 0. In particular, this means that

$$\int_0^T \dot{\boldsymbol{\eta}}(t) \, \mathrm{d}t = \mathbf{0}. \tag{3.72}$$

Define

$$J(\boldsymbol{\Sigma}) = \int_0^T \operatorname{tr}\left(\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}(t)\boldsymbol{\sigma}}\right) \mathrm{d}t.$$

We first show that $\boldsymbol{\Sigma}^*$ satisfies the first order condition

$$0 = \lim_{\varepsilon \downarrow 0} \frac{J(\boldsymbol{\Sigma}^* + \varepsilon\boldsymbol{\eta}) - J(\boldsymbol{\Sigma}^*)}{\varepsilon}. \tag{3.73}$$

Due to [MN18, theorem 1.1], we have the Taylor expansion

$$\begin{aligned}\sqrt{-\boldsymbol{\sigma}\left(\dot{\boldsymbol{\Sigma}}^*(t)+\varepsilon\boldsymbol{\eta}(t)\right)\boldsymbol{\sigma}} =& \sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}} \\ &+\varepsilon\int_0^\infty e^{-k\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}}}\boldsymbol{\sigma}\dot{\boldsymbol{\eta}}\boldsymbol{\sigma}e^{-k\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}}}\,\mathrm{d}k+O(\varepsilon^2).\end{aligned} \tag{3.74}$$

We can write

$$\lim_{\varepsilon\downarrow 0}\frac{J(\boldsymbol{\Sigma}^*+\varepsilon\boldsymbol{\eta})-J(\boldsymbol{\Sigma}^*)}{\varepsilon}=\int_0^T\mathrm{tr}\left(\int_0^\infty e^{-k\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}}}\boldsymbol{\sigma}\dot{\boldsymbol{\eta}}\boldsymbol{\sigma}e^{-k\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}}}\,\mathrm{d}k\right)\mathrm{d}t.$$

Permuting cyclically within the trace, the RHS becomes

$$\int_0^T\mathrm{tr}\left(\left(\int_0^\infty\boldsymbol{\sigma}e^{-2k\sqrt{-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}^*(t)\boldsymbol{\sigma}}}\boldsymbol{\sigma}\,\mathrm{d}k\right)\dot{\boldsymbol{\eta}}\right)\mathrm{d}t=\int_0^T\mathrm{tr}\left(\frac{1}{2}\left(\boldsymbol{\sigma}\left(-\boldsymbol{\sigma}\dot{\boldsymbol{\Sigma}}_t^*\boldsymbol{\sigma}\right)^{-1/2}\boldsymbol{\sigma}\right)\dot{\boldsymbol{\eta}}\right)\mathrm{d}t$$

Substituting in (3.71), we get

$$\int_0^T\mathrm{tr}\left(\frac{1}{2T}\left(\boldsymbol{\sigma}\left(\boldsymbol{\sigma}\boldsymbol{\Sigma}_0\boldsymbol{\sigma}\right)^{-1/2}\boldsymbol{\sigma}\right)\dot{\boldsymbol{\eta}}\right)\mathrm{d}t=\mathrm{tr}\left(\frac{1}{2T}\left(\boldsymbol{\sigma}\left(\boldsymbol{\sigma}\boldsymbol{\Sigma}_0\boldsymbol{\sigma}\right)^{-1/2}\boldsymbol{\sigma}\right)\int_0^T\dot{\boldsymbol{\eta}}\,\mathrm{d}t\right)=0$$

where the last equality is due to (3.72). This proves (3.73). To conclude the proof, observe that $\boldsymbol{\Sigma}\mapsto J(\boldsymbol{\Sigma})$ is convex. □

According to (3.15), the optimal $\mathbf{M}^*$ is

$$\mathbf{M}^*=\boldsymbol{\sigma}\left(\sqrt{\boldsymbol{\sigma}\boldsymbol{\Sigma}_0\boldsymbol{\sigma}}\right)^{-1}\boldsymbol{\sigma}.$$

This is a constant matrix, and thus a martingale. Theorem 3.2.8 applies and we recover the equilibrium

$$\begin{aligned}H(t,\mathbf{y})&=\mathbf{p}_0+\boldsymbol{\Lambda}\mathbf{y}\\ \mathrm{d}\boldsymbol{X}_t&=\frac{\boldsymbol{\Lambda}^{-1}\left(\tilde{\mathbf{v}}-\mathbf{p}_0\right)-\boldsymbol{Y}_t}{T-t}\,\mathrm{d}t\end{aligned}$$

where $\boldsymbol{\Lambda}=\boldsymbol{\sigma}^{-1}\left(\sqrt{\boldsymbol{\sigma}\boldsymbol{\Sigma}_0\boldsymbol{\sigma}}\right)\boldsymbol{\sigma}^{-1}$.

3.3.4 Dufresne and Fos 2016, deterministic growth

Let σ_t be a one-dimensional diffusion with dynamics

$$\mathrm{d}\sigma_t = \sigma_t m(t)\,\mathrm{d}t + \sigma_t \nu(t,\sigma_t)\,\mathrm{d}W_t, \tag{3.75}$$

where $m(t)$ is deterministic.

We present two ways to recover this equilibrium, one by solving the primal problem and the other by solving the dual problem.

3.3.4.1 The Primal Problem

The primal problem reads

$$\max_{\dot{\Sigma}<0} \mathbb{E}\left[\int_0^T \sigma_t \sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t\right]. \tag{3.76}$$

First note that σ_t has the representation

$$\sigma_t = \sigma_0 \exp\left(\int_0^t m(s)\,\mathrm{d}s\right) \exp\left(-\int_0^t \frac{\nu(s,\sigma_s)^2}{2}\,\mathrm{d}s + \int_0^t \nu(s,\sigma_s)\,\mathrm{d}W_s\right). \tag{3.77}$$

This motivates the Radon Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^T \frac{\nu(s,\sigma_s)^2}{2}\,\mathrm{d}s + \int_0^T \nu(s,\sigma_s)\,\mathrm{d}W_s\right), \tag{3.78}$$

where $\mathbb{P}$ is the probability measure in the expectation (3.76). We assume $\nu(s,\sigma_s)$ satisfies the Novikov condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \nu(s,\sigma_s)^2\,\mathrm{d}s\right)\right] < \infty.$$

Once this assumption is granted, we may write (3.76) as

$$\max_{\dot{\Sigma}<0} \mathbb{E}\left[\int_0^T \sigma_t \sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t\right] = \max_{\dot{\Sigma}<0} \mathbb{E}_{\mathbb{Q}}\left[\int_0^T \sigma_0 \exp\left(\int_0^t m(s)\,\mathrm{d}s\right) \sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t\right]. \tag{3.79}$$

Note that $m(s)$ is deterministic, and so is not affected by the change of measure. The key feature of (3.79) is that the RHS contains no random data, and so we may drop the expectation. The problem (3.76) therefore may be solved using classical calculus of variations. In

fact, writing (3.79) in the form

$$\max_{\dot{\Sigma}<0} \mathbb{E}\left[\int_0^T \sigma_t \sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t\right] = \int_0^T \sigma_0 \exp\left(\int_0^t m(s)\,\mathrm{d}s\right)\sqrt{-\dot{\Sigma}(t)}\,\mathrm{d}t,$$

we see that the RHS is of the form (3.63) with $\sigma_0 \exp\left(\int_0^t m(s)\,\mathrm{d}s\right)$ in place of $\sigma(t)$. The optimal $\dot{\Sigma}^*$, according to (3.66), is

$$\dot{\Sigma}^*(t) = -\Sigma_0 \frac{\exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)}{\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t}. \tag{3.80}$$

According to (3.27), the optimal M^* is

$$M_t^* = \sigma_t \exp\left(-\int_0^t m(s)\,\mathrm{d}s\right)\sqrt{\frac{1}{\Sigma_0}\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t}. \tag{3.81}$$

3.3.4.2 The Dual Problem

The dual problem (3.19) reads

$$\min_M \left(\frac{1}{2}\Sigma_0 M_0 + \frac{1}{2}\mathbb{E}\left[\int_0^T \frac{\sigma_t^2}{M_t}\,\mathrm{d}t\right]\right) \tag{3.82}$$

where the minimization is over all positive martingales $\{M\}$. If M_t evolves according to

$$\mathrm{d}M_t = \alpha_t M_t\,\mathrm{d}W_t \tag{3.83}$$

for some $\{\mathcal{F}_t^W\}$ adapted α_t, then due to (3.75), the integrand in (3.82) may be written

$$\frac{\sigma_t^2}{M_t} = \frac{\sigma_0^2}{M_0}\exp\left(\int_0^t \left(2m(s) - \nu(s,\sigma_s)^2 + \frac{\alpha_s^2}{2}\right)\mathrm{d}s + \int_0^t \left(2\nu(s,\sigma_s) - \alpha_s\right)\mathrm{d}W_s\right).$$

We introduce the measure $\mathbb{Q}$ through the RN derivative

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\frac{1}{2}\int_0^t \left(2\nu(s,\sigma_s) - \alpha_s\right)^2\mathrm{d}s + \int_0^t \left(2\nu(s,\sigma_s) - \alpha_s\right)\mathrm{d}W_s\right),$$

and we compute

$$\mathbb{E}\left[\frac{\sigma_t^2}{M_t}\right] = \mathbb{E}_{\mathbb{Q}}\left[\frac{\sigma_0^2}{M_0}\exp\left(\int_0^t 2m(s) - \big(\alpha_s - \nu(s,\sigma_s)\big)^2\,\mathrm{d}s\right)\right]. \tag{3.84}$$

The RHS integrand is minimized pointwise when

$$\alpha_s = \nu(s,\sigma_s), \tag{3.85}$$

in which case (3.84) reduces to

$$\mathbb{E}\left[\frac{\sigma_t^2}{M_t}\right] = \frac{\sigma_0^2}{M_0}\exp\left(2\int_0^t m(s)\,\mathrm{d}s\right).$$

We substitute this into (3.82) through an invocation of the tower property. When this is done, the minimization problem reduces to

$$\min_{M_0}\left(\frac{1}{2}\Sigma_0 M_0 + \frac{1}{2}\frac{\sigma_0^2}{M_0}\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t\right),$$

and the optimal M_0 is given by

$$M_0 = \sigma_0\sqrt{\frac{1}{\Sigma_0}\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t}.$$

In light of (3.83) and (3.85), we see now that the optimal martingale $\{M_t\}$ is

$$\begin{aligned} M_t &= \sigma_0\exp\left(-\int_0^t \frac{\nu(s,\sigma_s)^2}{2}\,\mathrm{d}s + \int_0^t \nu(s,\sigma_s)\,\mathrm{d}W_s\right)\sqrt{\frac{1}{\Sigma_0}\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t} \\ &= \sigma_t\exp\left(-\int_0^t m(s)\,\mathrm{d}s\right)\sqrt{\frac{1}{\Sigma_0}\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s\right)\mathrm{d}t}. \end{aligned}$$

Since $\{M_t\}$ minimizes (3.82) and is a martingale by construction, our main theorem asserts that $\{M_t\}$ represents the market depth in some equilibrium. The price impact of this same

equilibrium is

$$\lambda_t = \frac{1}{M_t} = \frac{1}{\sigma_t} \exp\left(\int_0^t m(s)\,\mathrm{d}s \right) \sqrt{\frac{\Sigma_0}{\int_0^T \exp\left(2\int_0^t m(s)\,\mathrm{d}s \right) \mathrm{d}t}}, \tag{3.86}$$

which is precisely (25) of [CDF16].

3.4 Existence of an Optimizer

3.4.1 One Dimension

First, consider (3.19) with M_t, σ_t, and Σ_0 all one dimensional. We split the optimization over $\{M_t\}_{0\le t\le T}$ into two steps: $\{M_t\}_{0<t\le T}$ and M_0, in that order:

$$\inf_{M_0} \left(\frac{1}{2}\Sigma_0 M_0 + \frac{1}{2} \inf_{M_{(0,T]}} \mathbb{E}\left[\int_0^T \sigma_t^2 M_t^{-1}\,\mathrm{d}t \right] \right) \tag{3.87}$$

The first step renders the objective value into a coercive function of M_0.

We will assign M_t and σ_t the dynamics

$$\mathrm{d}M_t = \alpha_t M_t\,\mathrm{d}W_t \tag{3.88}$$

$$\mathrm{d}\sigma_t = \sigma_t m_t\,\mathrm{d}t + \sigma_t \nu_t\,\mathrm{d}W_t \tag{3.89}$$

with M_0 and σ_0 fixed. Here, m_t and ν_t are exogenously given processes which obey some regularity conditions. Introducing the state variable $S_t = \sigma_t^2 M_t^{-1}$, the optimization over $\{M_t\}_{0<t\le T}$ reformulates as the control problem

$$\inf_{\alpha}\ \mathbb{E}\left[\int_0^T S_t\,\mathrm{d}t \right]. \tag{3.90}$$

S_t is a controlled state variable with dynamics

$$\mathrm{d}S_t = \left(2m_t + (\alpha_t - \nu_t)^2 \right) S_t\,\mathrm{d}t + (2\nu_t - \alpha_t) S_t\,\mathrm{d}W_t \tag{3.91}$$

and $S_0 = \sigma_0^2 M_0^{-1} > 0$ is fixed.

The first adjoint equation [YZ99, equation 3.8] reads

$$\mathrm{d}G_t = -\left(\left(2m_t + (\alpha_t^* - \nu_t)^2\right) G_t + (2\nu_t - \alpha_t^*)\, U_t - 1\right) \mathrm{d}t + U_t\, \mathrm{d}W_t \tag{3.92}$$

$$G_T = 0 \tag{3.93}$$

and the second [YZ99, equation 3.9]

$$\mathrm{d}g_t = -\left(\left(4m_t + 6\nu_t^2 + 3(\alpha_t^*)^2 - 8\alpha_t^*\nu_t\right) g_t + (4\nu_t - 2\alpha_t^*)\, u_t\right) \mathrm{d}t + u_t\, \mathrm{d}W_t \tag{3.94}$$

$$g_T = 0. \tag{3.95}$$

The solution to (3.94) – (3.95) is $(g_t, u_t) = (0, 0)$. The $\mathcal{H}$ function [YZ99, equation 3.16] therefore coincides with the Hamiltonian [YZ99, equation 3.10]

$$H(t, s, \alpha, G, U) = \left(2m_t + \nu_t^2 + \alpha^2 - 2\alpha\nu_t\right) sG + (2\nu_t - \alpha)\, sU - s. \tag{3.96}$$

The stochastic maximum principle [YZ99, theorem 3.2] asserts that the optimal $\bar{\alpha}$ is a pointwise minimizer of $H\left(t, \bar{X}_t, \alpha, G_t, U_t\right)$. We find

$$\alpha_t^* = \nu_t + \frac{1}{2}\frac{U_t}{G_t}. \tag{3.97}$$

Substituting this into (3.92), we obtain

$$\mathrm{d}G_t = \left(\frac{U_t^2}{4G_t} - \nu_t U_t - 2m_t G_t + 1\right) \mathrm{d}t + U_t\, \mathrm{d}W_t \tag{3.98}$$

$$G_T = 0 \tag{3.99}$$

If we define the probability measure $\mathbb{P}_\nu$ through the Radon-Nikodym derivative

$$\frac{\mathrm{d}\mathbb{P}_\nu}{\mathrm{d}\mathbb{P}} = \exp\left(-\frac{1}{2}\int_0^T \nu_t^2\, \mathrm{d}t + \int_0^T \nu_t\, \mathrm{d}W_t\right)$$

then we may write (3.98) as

$$\mathrm{d}G_t = \left(\frac{U_t^2}{4G_t} - 2m_t G_t + 1\right) \mathrm{d}t + U_t\, \mathrm{d}W_t^\nu \tag{3.100}$$

where $W_t^\nu = W_t - \int_0^t \nu_s\, \mathrm{d}s$ is a $\mathbb{P}_\nu$ Brownian motion by Girsanov's theorem. Furthermore, we

may multiply both sides of the above by the integrating factor $\exp\left(\int_0^t m_s \,\mathrm{d}s\right)$ and change variables to

$$\tilde{G}_t = -e^{2\int_0^t m_s ds} G_t \tag{3.101}$$

$$\tilde{U}_t = -e^{2\int_0^t m_s \,\mathrm{d}s} U_t. \tag{3.102}$$

Then, the BSDE (3.99), (3.100) becomes

$$\mathrm{d}\tilde{G}_t = \left(\frac{\tilde{U}_t^2}{4\tilde{G}_t} - e^{2\int_0^t m_s \,\mathrm{d}s}\right)\mathrm{d}t + \tilde{U}_t \,\mathrm{d}W_t^\nu \tag{3.103}$$

$$\tilde{G}_T = 0, \tag{3.104}$$

which is the same as that appearing in [EMZ22, proposition 2.1]. Invoking this result, we obtain the pair (G_t, U_t). The control (3.97) is well defined, and we can substitute it into (3.91) to obtain the optimal state trajectory

$$\bar{S}_t = \sigma_0^2 M_0^{-1} \exp\left(\int_0^t \left(2m_s - \frac{\nu_s^2}{2} + \nu_s \frac{U_s}{2G_s} + \frac{U_s^2}{8G_s^2}\right)\mathrm{d}s + \int_0^t \left(\nu_s - \frac{U_s}{2G_s}\right)\mathrm{d}W_s\right). \tag{3.105}$$

If we substitute (3.105) into (3.90), we will find that the optimal value is of the form $C\sigma_0^2 M_0^{-1}$ for some $C > 0$. Thus the original objective (3.87) is coercive in M_0, and admits a minimizer. Altogether, the optimal M_t is

$$\bar{M}_t = \sigma_t^2/\bar{S}_t = M_0 \exp\left(-\frac{1}{2}\int_0^t \left(\nu_s + \frac{U_s}{2G_s}\right)^2 \mathrm{d}s + \int_0^t \left(\nu_s + \frac{U_s}{2G_s}\right)\mathrm{d}W_s\right). \tag{3.106}$$

3.5 Heuristic Derivation

Let $\tilde{\mathbf{v}} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. We start with the optimal transport type problem

$$V(t, \mathbf{\Sigma}) = \sup_{\pi} \mathbb{E}^\pi \left[\tilde{\mathbf{v}}^\intercal \int_t^T \boldsymbol{\sigma}_s \,\mathrm{d}\mathbf{B}_s \middle| \mathscr{F}_t\right] \tag{3.107}$$

where the supremum is taken over all couplings π of $(\tilde{\mathbf{v}}, \mathbf{B}, \boldsymbol{W})$ in which $\tilde{\mathbf{v}}$ and $\boldsymbol{W}$ are independent and with laws $N(0, \Sigma_0)$ for $\tilde{\mathbf{v}}$ and $\mathbf{B}$ is Browian in its filtration and coupled with $(\tilde{\mathbf{v}}, \boldsymbol{W})$ causally.

Introducing the partition $t = t_0 < \cdots < t_n = T$, we approximate (3.107) as

$$V^n(t_k, \boldsymbol{\Sigma}) = \sup_{\pi_k} \sum_{k=0}^{n-1} \mathbb{E}^\pi \left[\tilde{\mathbf{v}}^\intercal \boldsymbol{\sigma}_{t_k} \left(\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \right) \right]$$

or, written dynamically, as

$$V^n(t_k, \boldsymbol{\Sigma}) = \sup_{\pi_k} \mathbb{E}^\pi \left[\tilde{\mathbf{v}}^\intercal \boldsymbol{\sigma}_{t_k} \left(\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \right) + V^n\big(t_{k+1}, \boldsymbol{\Sigma}_{t_{k+1}}\big) \,|\mathscr{F}_t \right] \tag{3.108}$$

where

$$\boldsymbol{\Sigma}_{t_{k+1}} = \mathrm{Var}\Big(\tilde{\mathbf{v}} \Big| \mathscr{F}^{\mathbf{B}}_{t_{k+1}} \Big) . \tag{3.109}$$

To couple $\big(\tilde{\mathbf{v}}, \mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k}\big)$, let the pair be jointly Gaussian with covariance matrix

$$\mathrm{Var}\left(\begin{bmatrix} \tilde{\mathbf{v}} \\ \mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \end{bmatrix} \middle| \mathscr{F}^{\mathbf{B}}_{t_k} \right) = \begin{bmatrix} \boldsymbol{\Sigma} & \sqrt{\tilde{\boldsymbol{\Sigma}}_k}\mathbf{Q}\Delta t_k \\ \mathbf{Q}^\intercal \sqrt{\tilde{\boldsymbol{\Sigma}}_k}\Delta t_k & \mathbf{I}\Delta t_k \end{bmatrix} .$$

Note that we have polar factorized

$$\mathrm{Cov}(\tilde{\mathbf{v}})\, \mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} = \sqrt{\tilde{\boldsymbol{\Sigma}}_k}\mathbf{Q}_k \Delta t_k \tag{3.110}$$

so that $\tilde{\boldsymbol{\Sigma}}_k$ is symmetric positive semi-definite and $\mathbf{Q}_k$ is orthogonal. Invoking the Gaussian projection theorem, (3.110) and (3.109) yield

$$\boldsymbol{\Sigma}_{t_{k+1}} = \boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}_k \Delta t_k .$$

Similarly, (3.110) permits the decomposition

$$\tilde{\mathbf{v}} = \mathbf{z} + \sqrt{\tilde{\boldsymbol{\Sigma}}}\mathbf{Q}_k \left(\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \right)$$

where $\mathbf{z}$ is independent of $\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k}$. With these considerations, (3.108) updates to

$$\begin{aligned} V^n(t_k, \boldsymbol{\Sigma}) &= \sup_{\tilde{\boldsymbol{\Sigma}}_k} \sup_{\mathbf{Q}_k} \mathbb{E}^\pi \left[\left(\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \right)^\intercal \mathbf{Q}_k^\intercal \sqrt{\tilde{\boldsymbol{\Sigma}}_k} \boldsymbol{\sigma}_{t_k} \left(\mathbf{B}_{t_{k+1}} - \mathbf{B}_{t_k} \right) + V^n\Big(t_{k+1}, \boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}_k \Delta t_k\Big) \right] \\ &= \sup_{\tilde{\boldsymbol{\Sigma}}_k} \sup_{\mathbf{Q}_k} \mathbb{E}^\pi \left[\mathrm{tr}\left(\mathbf{Q}_k^\intercal \sqrt{\tilde{\boldsymbol{\Sigma}}_k} \boldsymbol{\sigma}_{t_k} \right) \Delta t_k + V^n\Big(t_{k+1}, \boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}_k \Delta t_k\Big) \right] \end{aligned}$$

To attain the supremum over $\mathbf{Q}_k$, write $\sqrt{\tilde{\boldsymbol{\Sigma}}_k}\boldsymbol{\sigma}_{t_k} = \mathbf{O}_k\boldsymbol{P}_k$ for $\mathbf{O}_k$ orthogonal and $\boldsymbol{P}_k$ positive, and choose $\mathbf{Q}_k = \mathbf{O}_k$. The trace term reduces to

$$\operatorname{tr}(\boldsymbol{P}_k) = \operatorname{tr}\left(\sqrt{\boldsymbol{\sigma}_{t_k}^\intercal \tilde{\boldsymbol{\Sigma}}_k \boldsymbol{\sigma}_{t_k}}\right),$$

and $V^n(t_k, \boldsymbol{\Sigma})$ to

$$V^n(t_k, \boldsymbol{\Sigma}) = \sup_{\tilde{\boldsymbol{\Sigma}}_k} \mathbb{E}^\pi \left[\operatorname{tr}\left(\sqrt{\boldsymbol{\sigma}_{t_k}^\intercal \tilde{\boldsymbol{\Sigma}}_k \boldsymbol{\sigma}_{t_k}}\right) \Delta t_k + V^n(t_{k+1}, \boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}_k \Delta t_k)\right]$$

where $\tilde{\boldsymbol{\Sigma}}_k$ is chosen so that $\boldsymbol{\Sigma} - \tilde{\boldsymbol{\Sigma}}_k \Delta t_k$ remains positive semi-definite. This motivates our continuous-time problem

$$V(t, \boldsymbol{\Sigma}) = \sup_{\tilde{\boldsymbol{\Sigma}}} \left\{ \mathbb{E}\left[\int_t^T \operatorname{tr}\left(\sqrt{\boldsymbol{\sigma}_s \tilde{\boldsymbol{\Sigma}}_s \boldsymbol{\sigma}_s}\right) \mathrm{d}s \middle| \mathcal{F}_t^{\boldsymbol{W}}\right] : \int_t^T \tilde{\boldsymbol{\Sigma}}_s \,\mathrm{d}s \leq \boldsymbol{\Sigma} \right\} \tag{3.111}$$

$$= \sup_{(\tilde{\boldsymbol{\Sigma}}_s, Q_s) \in \mathbb{R}^{n\times n} \times O_n} \left\{ \mathbb{E}\left[\int_t^T \operatorname{tr}\left(\tilde{\boldsymbol{\Sigma}}_s \sigma_s Q_s\right) \mathrm{d}s \middle| \mathcal{F}_t^{\boldsymbol{W}}\right] : \int_t^T \tilde{\boldsymbol{\Sigma}}_s^\top \tilde{\boldsymbol{\Sigma}}_s \,\mathrm{d}s \leq \boldsymbol{\Sigma} \right\} \tag{3.112}$$

or, written dynamically,

$$V(t, \boldsymbol{\Sigma}) = \sup_{\tilde{\boldsymbol{\Sigma}}} \left\{ \mathbb{E}\left[\int_t^\tau \operatorname{tr}\left(\sqrt{\boldsymbol{\sigma}_s \tilde{\boldsymbol{\Sigma}}_s \boldsymbol{\sigma}_s}\right) \mathrm{d}s + V\left(\tau, \boldsymbol{\Sigma} - \int_t^\tau \tilde{\boldsymbol{\Sigma}}_s \,\mathrm{d}s\right) \middle| \mathcal{F}_t^{\boldsymbol{W}}\right] : \int_t^\tau \tilde{\boldsymbol{\Sigma}}_s \,\mathrm{d}s \leq \boldsymbol{\Sigma} \right\}$$

We expect the inequality $\int_0^T \tilde{\boldsymbol{\Sigma}}_s \,\mathrm{d}s \leq \boldsymbol{\Sigma}$ to be tight if $\tilde{\boldsymbol{\Sigma}}$ is optimal.

3.6 Conclusion

In this chapter, we extended the Kyle–Back framework to a setting with stochastic and multidimensional liquidity. By reinterpreting the equilibrium construction as a variational problem over covariance paths, we recast informed trading as an *optimal liquidation of information*. The insider's control variable is no longer her inventory of shares but the rate at which private information is released into the market. This reformulation unifies the equilibrium conditions of the classical Kyle–Back model, the dynamic information model of Back and Pedersen (1998), and the stochastic-volatility formulation of Collin–Dufresne and Fos (2016), within a single convex-analytic and optimal-transport framework.

The primal problem, expressed as a maximization over decreasing covariance processes $\dot{\Sigma}_t <$

0, encodes the feasible flow of information. Its dual, formulated in terms of a matrix-valued martingale, captures the market maker's filtration and yields a natural generalization of the adjoint process identified in [CDF16]. The dual variable thus plays the role of a *shadow price of information*, quantifying the marginal value of a reduction in uncertainty about the terminal asset value. This dual interpretation also clarifies the economic content of the process G_t in Collin–Dufresne and Fos: it is the co-state variable in the stochastic maximum principle corresponding to our variational problem.

We proved that the resulting equilibrium satisfies a multidimensional analogue of the inconspicuous-trading property: both the total order flow and the noise component are Brownian motions in their respective filtrations, while the market price evolves according to the matrix process $\Lambda_t = \nabla^2_\xi \varphi(t, \xi_t)$ derived from a convex potential φ. The structure of this potential links the equilibrium directly to the Monge–Kantorovich theory of optimal transport, extending the one-dimensional results of Back, Cocquemas, Ekren, and Lioui (2021) to higher dimensions and to stochastic volatility environments.

Finally, by specializing parameters, we recovered the classical equilibria of Kyle (1985), Back (1992), Back and Pedersen (1998), and Collin–Dufresne and Fos (2016) as limiting cases of our model. This confirms that the variational and transport formulations provide a unified geometric framework for understanding information flow, price impact, and liquidity in continuous-time markets.

CHAPTER 4

Kyle's Single Auction Model

In our final project, we apply optimal transport to the one-shot model from Kyle's original 1985 paper [Kyl85]. This model differs substantially from the continuous-time ones in the previous chapters, partly because the inconspicuous trading property no longer holds in discrete time. Our application of optimal transport to this model therefore demonstrates the versatility of this tool and reinforces its deep connection to Kyle-Back models.

The equilibrium can be described through a coupling between the insider's private information $\tilde{v}$ and her optimal demand $X(\tilde{v})$. Given a pricing rule P, the trader chooses X by solving a first-order condition of the form

$$\tilde{v} = (\mathrm{id} P_\gamma)'(X(\tilde{v})),$$

where P_γ denotes the price function smoothed by noise. Conversely, given X, the market maker's rationality condition determines the pricing rule P by conditioning on total order flow. Equilibrium arises when these two relations are consistent—when the optimal strategy and the rational price rule close the loop.

This formulation highlights the variational structure underlying the discrete-time Kyle model and shows how its equilibrium can be characterized entirely through the relationship between private information, optimal trading, and the market maker's updating of prices.

4.1 The Model

There is a public announcement to be made at time 1 which will affect the price of a risky asset. Immediately after this announcement, the asset will be valued at $\tilde{v}$. The distribution of $\tilde{v}$ is public information, but no one knows exactly what value $\tilde{v}$ will take except a single informed trader. The informed trader knows $\tilde{v}$ from time 0. She submits the trade order $X(\tilde{v})$, which combines with liquidity trades Z. The distribution of Z is public information (often taken to be $\mathcal{N}(0, \sigma^2)$) and Z is independent of $\tilde{v}$, but no one knows exactly what value

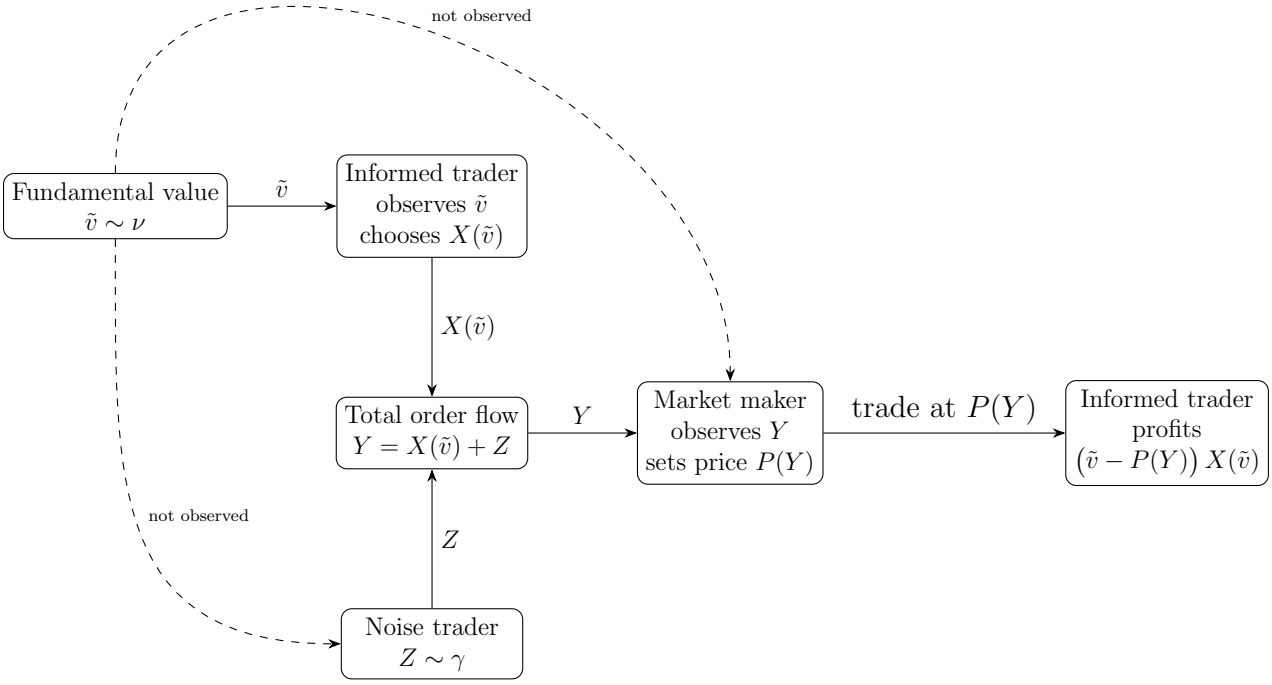


Figure 4.1: Flow chart describing Kyle's single auction model.

Z will take. The market maker observes only the total order flow $Y = X(\tilde{v}) + Z$, without being able to distinguish its components, and sets a price $P(Y)$ intended to estimate the fundamental value $\tilde{v}$. All trading occurs instantaneously at time 0, after which $\tilde{v}$ is revealed at time 1. The insider's realized profit is then

$$\pi = (\tilde{v} - P(Y))\, X(\tilde{v}).$$

This motivates the following definition.

Definition 4.1.1. A *pricing rule* P is a strictly increasing function from $\mathbb{R}$ to $\mathbb{R}$. A *trading strategy* X is a strictly increasing function from $\mathrm{supp}(\tilde{v})$ to $\mathbb{R}$. Given a pricing rule P, a trading strategy X^* is said to be *optimal* if

$$X^*(v) \in \arg\max_x \mathbb{E}\big[(v - P(Y))x\big], \qquad \forall v \in \mathrm{supp}(\tilde{v}). \tag{4.1}$$

Given a trading strategy X, a pricing rule P^* is said to be *rational* if

$$P^*(y) = \mathbb{E}\big[\tilde{v}\big|X(\tilde{v}) + Z = y\big]. \tag{4.2}$$

A pair (X^*, P^*) is said to be an equilibrium if X^* is optimal for P^* and P^* is rational for X^*.

For the informed to develop an optimal strategy, she must know the market maker's pricing rule. For the market maker to develop a rational pricing rule, he must know the insider's strategy. Finding an equilibrium amounts to solving both problems simultaneously.

4.2 Main Theorem

Lemma 4.2.1. *Fix a pricing rule P (not necessarily rational). Denote $\nu := \mathcal{L}(\tilde{v})$, $\gamma := \mathcal{L}(Z)$, and $P_\gamma(y) := \mathbb{E}\big[P(x + Z)\big]$. Let α^* be the maximizer of*

$$\alpha \mapsto \sup_{\pi \in \Pi(\alpha,\nu)} \iint vx\pi(\mathrm{d}x, \mathrm{d}v) - \int xP_\gamma(x)\alpha(\mathrm{d}x). \tag{4.3}$$

Let T^{ν,α^*} denote the Brenier map from ν to α^*. Then, $X = T^{\nu,\alpha^*}$ is optimal given P.

Proof. Denote the maximum correlation function between α and ν by

$$M(\alpha, \nu) = \sup_{\pi \in \Pi(\alpha,\nu)} \iint vx\pi(\mathrm{d}x, \mathrm{d}v). \tag{4.4}$$

At the optimum, the intrinsic derivative of $\alpha \mapsto M(\alpha, \nu) - \int x P_\gamma(x)\alpha(\mathrm{d}x)$ is 0, which gives

$$\begin{aligned} 0 &= T^{\alpha^*,\nu}(x) - xP_\gamma(x) - P'_\gamma(x) \\ &= T^{\alpha^*,\nu}(x) - (\mathrm{id}P_\gamma)'(x). \end{aligned}$$

Changing variables $v = T^{\alpha^*,\nu}(x)$, this is

$$v = (\mathrm{id}P_\gamma)'(T^{\nu,\alpha^*}(v)). \tag{4.5}$$

Thus $x = T^{\nu,\alpha^*}(v)$ solves

$$\big(x(v - P_\gamma(x))\big)' = 0,$$

and since $x(v - P_\gamma(x))$ is concave in x, this means

$$T^{\nu,\alpha^*}(v) \in \arg\max_x x(v - P_\gamma(x)) = \arg\max_x \mathbb{E}\Big[x\left(\tilde{v} - P(x + Z)\right) \,\Big|\, \tilde{v} = v\Big].$$

□

Remark 4.2.2. Lemma 4.2.1 characterizes the ideal X for a given P. In particular, one solves the variational problem (4.3) and obtains an optimal strategy. In this way, we reformulate (4.1) as a maximization over the space of measures corresponding to $\mathcal{L}(X(\tilde{v}))$ as opposed to the space of functions corresponding to $X(\tilde{v})$. The next lemma characterizes the ideal P for a given X.

Lemma 4.2.3. *Fix $X\colon \mathrm{supp}(\tilde{v}) \mapsto \mathbb{R}$. Suppose that X is strictly increasing. Denote $\alpha = \mathcal{L}(X(\tilde{v}))$ and suppose that α is absolutely continuous. Then,*

$$P^X(y) := \frac{\mathbb{E}\big[\tilde{v} f_Z(y - T^{\nu,\alpha}(\tilde{v}))\big]}{\mathbb{E}\big[f_Z(y - T^{\nu,\alpha}(\tilde{v}))\big]} \tag{4.6}$$

is efficient for X.

Proof. This proof is due to [CK00]. According to Bayes' rule,

$$\mathbb{E}\big[\tilde{v} \mid Y = y\big] = \int_{\mathbb{R}} v f_{\tilde{v}|Y}(v \mid y) dv = \frac{\int_{\mathbb{R}} v f_{Y|\tilde{v}}(y \mid v) f_{\tilde{v}}(v)\,\mathrm{d}v}{\int_{\mathbb{R}} f_{Y|\tilde{v}}(y \mid v) f_{\tilde{v}}(v)\,\mathrm{d}v},$$

where $f_{Y|\tilde{v}}$ is the conditional density function of Y given $\tilde{v}$ and $f_{\tilde{v}}$ the marginal density function of $\tilde{v}$. Note that

$$F_{Y|\tilde{v}}(y \mid v) = \mathbb{P}\big(Y \le y \mid \tilde{v} = v\big) = \mathbb{P}\big(Z \le y - X(v)\big) = F_Z(y - X(v)),$$

because $Y = X(\tilde{v}) + Z$ and $\tilde{v}$ and Z are independent. Therefore, $f_{Y|\tilde{v}}(y \mid v) = f_Z\big(y - X(v)\big)$. Finally, since $T^{\nu,\alpha}$ is the Brenier map from $\mathcal{L}(\tilde{v})$ to $\mathcal{L}(X(\tilde{v}))$ and X is strictly monotone, we have $X(v) = T^{\nu,\alpha}(v)$. □

Theorem 4.2.4. *Denote*

$$C_\alpha(x) := x \int \frac{\mathbb{E}\big[\tilde{v} f_Z(x + z - T^{\nu,\alpha}(\tilde{v}))\big]}{\mathbb{E}\big[f_Z(x + z - T^{\nu,\alpha}(\tilde{v}))\big]} f_Z(z)\,\mathrm{d}z. \tag{4.7}$$

There exists an equilibrium in Kyle's model if and only if there exists α^* which solves the fixed-point problem

$$\alpha^* = \arg\max_{\alpha \in \mathcal{P}(\mathbb{R})} \left(M(\alpha, \nu) - \int C_{\alpha^*}(x)\alpha(\mathrm{d}x) \right). \tag{4.8}$$

Proof. First suppose there exists an equilibrium (X^*, P^*). We will show that $\alpha^* = \mathcal{L}(X^*(\tilde{v}))$ maximizes

$$\alpha \mapsto M(\alpha, \nu) - \int C_{\alpha^*}(x)\alpha(\mathrm{d}x). \tag{4.9}$$

By lemma 4.2.3, P^{X^*} as defined by (4.6) is rational for X^*. But since (X^*, P^*) is an equilibrium, P^* is rational for X^* too. Thus $P^{X^*}(y)$ and $P^*(y)$ coincide and share the value $\mathbb{E}\big[\tilde{v} \mid X^*(\tilde{v}) + Z = y\big]$. Since $P^{X^*} = P^*$,

$$M(\alpha, \nu) - \int C_{\alpha^*}(x)\alpha(\mathrm{d}x) = \max_{\pi \in \Pi(\alpha,\nu)} \iint xv\pi(\mathrm{d}x, \mathrm{d}v) - \int xP^*_\gamma(x)\alpha(\mathrm{d}x) \tag{4.10}$$

$$= \max_{\pi \in \Pi(\alpha,\nu)} \iint xv\pi(\mathrm{d}x, \mathrm{d}v) - \int \mathbb{E}\big[xP^*(x + Z)\big]\,\alpha(\mathrm{d}x) \tag{4.11}$$

represents the informed trader's profit if she chooses the strategy X and generates the law $\alpha = \mathcal{L}(X(\tilde{v}))$. But since (X^*, P^*) is an equilibrium, X^* is optimal, so $\alpha^* = \mathcal{L}(X^*(\tilde{v}))$ must maximize (4.11).

Next, suppose there exists α^* satisfying (4.8). We will show that

$$X^*(v) := T^{\nu,\alpha^*}(v) \tag{4.12}$$

$$P^*(y) := \frac{\mathbb{E}\big[\tilde{v} f_Z(y - T^{\nu,\alpha^*}(\tilde{v}))\big]}{\mathbb{E}\big[f_Z(y - T^{\nu,\alpha^*}(\tilde{v}))\big]} \tag{4.13}$$

form an equilibrium. Lemma 4.2.1 asserts that given P^*, X^* is optimal. Lemma 4.2.3 asserts that given X^*, P^* is rational. □

Proposition 4.2.5. *In order for α^* to solve the fixed point problem* (4.3)*, it suffices that α^**

solves

$$\phi^{\alpha^*,\nu}(x) = C_{\alpha^*}(x) \qquad \forall x \in \mathrm{supp}(\alpha^*), \tag{4.14}$$

where $\phi^{\alpha^*,\nu}$ is the Brenier potential from α to ν with normalization $\phi^{\alpha^*,\nu}(0) = 0$.

Proof. The gradient of (4.9) in α is

$$\nabla_\alpha \left(M(\alpha, \nu) - \int C_{\alpha^*}(x)\alpha(dx) \right) = \phi^{\alpha,\nu}(x) - C_{\alpha^*}(x).$$

The Brenier potential $\phi^{\alpha,\nu}$ is unique up to a constant, so we choose the normalization $\phi^{\alpha,\nu}(0) = 0$. In this case, the first order condition coincides with (4.14). Since (4.9) is convex in α, the first order condition is sufficient for optimality. □

Example 4.2.6. The original [Kyl85] takes $\tilde{v} \sim \mathcal{N}(p_0, \Sigma_0)$ and $Z \sim \mathcal{N}(0, \sigma^2)$. In this case, the equilibrium is

$$X^*(v) = \frac{\sigma}{\sqrt{\Sigma_0}} (v - p_0) \tag{4.15}$$

$$P^*(y) = p_0 + \frac{\sqrt{\Sigma_0}}{\sigma} y. \tag{4.16}$$

We have $\alpha^* := \mathcal{L}(X^*(\tilde{v})) = \mathcal{N}(0, \sigma^2)$. To show that α^* satisfies eq. (4.14), note from optimal transport theory that

$$\phi^{\alpha^*,\nu}(x) = p_0 x + \frac{\sqrt{\Sigma_0}}{2\sigma} x^2.$$

On the other hand, a lengthy computation on (4.7) produces

$$C_{\alpha^*}(x) = p_0 x + \frac{\sqrt{\Sigma_0}}{2\sigma} x^2.$$

Thus proposition 4.2.5 applies and asserts that $\alpha^* = \mathcal{N}(0, \sigma^2)$ solves the fixed point problem (4.8) in the particular case $\tilde{v} \sim \mathcal{N}(p_0, \Sigma_0)$ and $Z \sim \mathcal{N}(0, \sigma^2)$.

Conversely, we can start with $\alpha^* = \mathcal{N}(0, \sigma^2)$ as the solution to (4.8) and then construct the equilibrium (4.15)—(4.16) from (4.12)—(4.13).

Example 4.2.7. Carré, Dufresne, and Gabriel [CCDG22] take $\tilde{v} \sim \mathcal{U}(-1, 1)$ and $Z \sim \mathcal{U}(-1, 1)$. Their equilibrium is

$$X^*(v) = v \tag{4.17}$$

$$P^*(y) = \frac{y}{2}. \tag{4.18}$$

Thus $\alpha^* := \mathcal{L}(X^*(\tilde{v})) = \mathcal{U}(-1,1)$. Optimal transport theory asserts that since $\alpha^* = \nu$,

$$\phi^{\alpha^*,\nu}(x) = \frac{1}{2}x^2 \qquad x \in [-1,1].$$

On the other hand, we can show through a lengthy calculation on (4.7) that

$$C_{\alpha^*}(x) = \frac{1}{2}x^2 \qquad x \in [-1,1].$$

Thus proposition 4.2.5 asserts that $\alpha^* = \mathcal{U}(-1,1)$ solves the fixed point problem (4.8) in the particular case $\tilde{v} \sim \mathcal{U}(-1,1)$ and $Z \sim \mathcal{U}(-1,1)$.

Conversely, we can start with $\alpha^* = \mathcal{U}(-1,1)$ as the solution to (4.8) and then construct the equilibrium (4.17)—(4.18) from (4.12)—(4.13).

Example 4.2.8. Cho and El Karoui [CK00] take

$$\tilde{v} = \begin{cases} v_1 & \text{with probability } \pi_1 \\ v_0 & \text{with probability } \pi_0 \end{cases}$$

where $v_0 < v_1$ and $\pi_0 + \pi_1 = 1$ with $Z \sim \mathcal{N}(0,\sigma^2)$. Their equilibrium is

$$X(v) = \begin{cases} \alpha_1^* & v = v_1 \\ \alpha_0^* & v = v_0 \end{cases} \tag{4.19}$$

$$P(y) = \frac{\pi_1 v_1 \mathcal{E}^{\left(\alpha_1^* + \sigma^2/\alpha_1^*\right)y - \left((\alpha_1^*)^2 - \sigma^2/(\alpha_1^*)^2\right)/2} + \pi_0 v_0}{\pi_1 \mathcal{E}^{\left(\alpha_1^* + \sigma^2/\alpha_1^*\right)y - \left((\alpha_1^*)^2 - \sigma^2/(\alpha_1^*)^2\right)/2} + \pi_0} \tag{4.20}$$

where α_1^* is the unique solution (in x) to

$$\mathbb{E}\left[\frac{1 + xZ/\sigma^2}{\pi_1 \mathcal{E}^{(x+\sigma^2/x)Z + (x+\sigma^2/x)^2/2} + \pi_0}\right] = 0 \tag{4.21}$$

and $\alpha_0^* := -\sigma^2/\alpha_1^*$. To verify that $\alpha^* = \mathcal{L}(X(\tilde{v}))$ solves (4.14), it is easier to instead work with the gradient of (4.14) in x:

$$v_1 \stackrel{?}{=} \mathbb{E}\left[\left(1 + \alpha_1^* \frac{Z}{\sigma^2}\right) P(\alpha_1^* + Z)\right] \tag{4.22}$$

$$v_0 \stackrel{?}{=} \mathbb{E}\left[\left(1 + \alpha_0^* \frac{Z}{\sigma^2}\right) P(\alpha_0^* + Z)\right]. \tag{4.23}$$

But these follow from (4.21) and $\alpha_0^* = -\sigma^2/\alpha_1^*$. In fact, (4.22)—(4.23) are exactly the first order condition eq. (6) of [CK00].

4.3 Relation to Static Mean Field Games

Theorem 4.3.1. *The fix point problem*

$$\alpha^* \in \arg\max_{\alpha} \left(M(\alpha, \nu) - \int C_{\alpha^*}(x)\, \alpha(\mathrm{d}x) \right), \tag{4.24}$$

shares the same solution as the fixed point problem

$$\alpha^* \in \arg\max_{\alpha} \int \left(\phi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x) \right) \alpha(\mathrm{d}x). \tag{4.25}$$

Proof. For brevity, fix ν and define

$$F(\alpha) := M(\alpha, \nu) - \int C_{\alpha^*}(x)\, \alpha(\mathrm{d}x), \qquad \alpha \in \mathcal{P}(\mathbb{R}).$$

By assumption F is concave on $\mathcal{P}$ and differentiable at α^*, with Gâteaux derivative

$$DF(\alpha^*)(x) = \varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x).$$

(i) From (4.24) *to* (4.25). Assume that α^* solves (4.24), i.e. $\alpha^* \in \arg\max_\alpha F(\alpha)$. Fix any $\alpha \in \mathcal{P}(\mathbb{R})$ and set

$$\alpha_t := (1-t)\alpha^* + t\alpha, \qquad t \in [0,1].$$

By optimality of α^*, we have $F(\alpha_t) \le F(\alpha^*)$ for all $t \in [0,1]$. Hence the right derivative at $t = 0$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\alpha_t)\Big|_{t=0^+} \le 0.$$

By the definition of the Gâteaux derivative,

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\alpha_t)\Big|_{t=0^+} = \int \left(\varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x) \right) (\alpha - \alpha^*)(\mathrm{d}x).$$

Thus, for every $\alpha \in \mathcal{P}$,

$$\int \left(\varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x) \right) \alpha(\mathrm{d}x) \le \int \left(\varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x) \right) \alpha^*(\mathrm{d}x),$$

which shows that α^* maximizes the linear functional in (4.25). Hence $\alpha^* \in \arg\max_\alpha \int (\varphi^{\alpha^*,\nu} - C_{\alpha^*})\, \mathrm{d}\alpha$.

(ii) From (4.25) *to* (4.24). Conversely, assume that α^* solves (4.25), i.e.

$$\int \left(\varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x)\right)(\alpha - \alpha^*)(\mathrm{d}x) \leq 0 \qquad \text{for all } \alpha \in \mathcal{P}(\mathbb{R}).$$

Fix $\alpha \in \mathcal{P}$ and again set $\alpha_t = (1-t)\alpha^* + t\alpha$. Define $g(t) := F(\alpha_t)$ for $t \in [0,1]$. Since F is concave, g is concave on $[0,1]$. For a concave differentiable function g, we have

$$g(t) - g(0) \leq t\, g'(0^+) \qquad \text{for all } t \in (0,1].$$

Using the chain rule,

$$g'(0^+) = \frac{\mathrm{d}}{\mathrm{d}t} F(\alpha_t)\bigg|_{t=0^+} = \int \left(\varphi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x)\right)(\alpha - \alpha^*)(\mathrm{d}x) \leq 0.$$

Therefore, for all $t \in (0,1]$,

$$F(\alpha_t) - F(\alpha^*) = g(t) - g(0) \leq t\, g'(0^+) \leq 0.$$

Letting $t \uparrow 1$ and using continuity of F along the segment $t \mapsto \alpha_t$, we obtain $F(\alpha) \leq F(\alpha^*)$ for every $\alpha \in \mathcal{P}$, so that $\alpha^* \in \arg\max_\alpha F(\alpha)$, i.e. α^* solves (4.24).

Combining (i) and (ii) shows that the fixed point problems (4.24) and (4.25) share the same set of solutions α^*. □

Corollary 4.3.2. *The following three are equivalent:*

1. *there exists an equilibrium in Kyle's model*
2. *there exists a solution to the fixed point problem* (4.8)
3. *there exists a mean field equilibrium to the static game with action space* $\mathbb{R}$ *and payoff* $\phi^{\alpha^*,\nu}(x) - C_{\alpha^*}(x)$.

4.4 Transaction Costs

It is not clear under which general assumptions on $\mathcal{L}(\tilde{v})$ and $\mathcal{L}(Z)$ the fixed point problem (4.3) admits a solution. However, we found that if we

1. add superlinear transaction costs

2. restrict $\tilde{v}$ to have compact support

then a solution is guaranteed to exist. To formalize these notions, we add the following definition:

Definition 4.4.1. A function $\tau\colon \mathbb{R} \mapsto \mathbb{R}$ is said to be a *transaction cost* if

1. $\tau(0) = 0$
2. τ is convex
3. τ is even: $\tau(x) = \tau(-x)$

Fix a pricing rule P and a transaction cost τ. A strategy $X\colon \operatorname{supp}(\tilde{v}) \mapsto \mathbb{R}$ is said to be *optimal for P and τ* if

$$X(v) \in \arg\max_x \mathbb{E}\big[(v - P(x+Z))x - \tau(x)\big], \qquad \forall v \in \operatorname{supp}(\tilde{v}). \tag{4.26}$$

For a fixed transaction cost τ, the pair (X^*, P^*) is said to be an equilibrium if

1. given P^*, X^* is optimal
2. given X^*, P^* is rational.

Taking $\tau := 0$ specializes to the model presented above. To characterize the equilibrium in this more general setting, we restate and re-prove lemma 4.2.3:

Lemma 4.4.2. *Fix a pricing rule P (not necessarily rational) and a transaction cost τ. Denote $\nu := \mathcal{L}(\tilde{v})$, $\gamma := \mathcal{L}(Z)$, and $P_\gamma(y) := \mathbb{E}\big[P(x+Z)\big]$. Let α^* be the maximizer of*

$$\alpha \mapsto \sup_{\pi \in \Pi(\alpha,\nu)} \iint vx\pi(\mathrm{d}x, \mathrm{d}v) - \int \big(xP_\gamma(x) + \tau(x)\big)\,\alpha(\mathrm{d}x). \tag{4.27}$$

Let T^{ν,α^*} denote the Brenier map from ν to α^*. Then, $X = T^{\nu,\alpha^*}$ is optimal given P.

Proof. Denote $M(\alpha,\nu)$ as in (4.4). At the optimum, the intrinsic derivative of $\alpha \mapsto M(\alpha,\nu) - \int \big(xP_\gamma(x) + \tau(x)\big)\,\alpha(\mathrm{d}x)$ is 0, which gives

$$\begin{aligned} 0 &= T^{\alpha^*,\nu}(x) - xP_\gamma(x) - P_\gamma'(x) - \tau'(x) \\ &= T^{\alpha^*,\nu}(x) - (\mathrm{id}P_\gamma + \tau)'(x). \end{aligned}$$

Changing variables $v = T^{\alpha^*,\nu}(x)$, this is

$$v = (\mathrm{id}P_\gamma + \tau)'(T^{\nu,\alpha^*}(v)). \tag{4.28}$$

Thus $x = T^{\nu,\alpha^*}(v)$ solves

$$\big(x(v - P_\gamma(x)) - \tau(x)\big)' = 0,$$

and since $x(v - P_\gamma(x)) - \tau(x)$ is concave in x, this means

$$T^{\nu,\alpha^*}(v) \in \arg\max_x \big(x(v - P_\gamma(x)) - \tau(x)\big) = \arg\max_x \mathbb{E}\Big[x\left(\tilde{v} - P(x + Z)\right) - \tau(x) \,\Big|\, \tilde{v} = v\Big].$$

□

Remark 4.4.3. The above lemma states that even with transaction costs, the optimal strategy for the informed trader is still the Brenier map. Lemma 4.2.3, which characterizes the ideal pricing rule, is unchanged in this new setting, so we will not restate it. We will, however, have to update theorem 4.2.4.

Theorem 4.4.4. *Define $C_\alpha(x)$ as in (4.7). There exists an equilibrium in Kyle's model if and only if there exists α^* which solves the fixed-point problem*

$$\alpha^* = \arg\max_{\alpha \in \mathcal{P}(\mathbb{R})} \left(M(\alpha, \nu) - \int \big(C_{\alpha^*}(x) + \tau(x)\big)\, \alpha(\mathrm{d}x) \right). \tag{4.29}$$

Proof. First suppose there exists an equilibrium (X^*, P^*). We will show that $\alpha^* = \mathcal{L}(X^*(\tilde{v}))$ maximizes

$$\alpha \mapsto M(\alpha, \nu) - \int \big(C_{\alpha^*}(x) + \tau(x)\big)\, \alpha(\mathrm{d}x). \tag{4.30}$$

By lemma 4.2.3, P^{X^*} as defined by (4.6) is rational for X^*. But since (X^*, P^*) is an equilibrium, P^* is rational for X^* too. Thus $P^{X^*}(y)$ and $P^*(y)$ coincide and share the value $\mathbb{E}\big[\tilde{v} \mid X^*(\tilde{v}) + Z = y\big]$. Since $P^{X^*} = P^*$,

$$\begin{aligned} &M(\alpha, \nu) - \int \big(C_{\alpha^*}(x) + \tau(x)\big)\, \alpha(\mathrm{d}x) \\ &\quad = \max_{\pi \in \Pi(\alpha,\nu)} \iint xv\pi(\mathrm{d}x, \mathrm{d}v) - \int \mathbb{E}\big[xP^*(x + Z) + \tau(x)\big]\, \alpha(\mathrm{d}x) \end{aligned}$$

represents the informed trader's profit if she chooses the strategy X and generates the law $\alpha = \mathcal{L}(X(\tilde{v}))$. But since (X^*, P^*) is an equilibrium, X^* is optimal, so $\alpha^* = \mathcal{L}(X^*(\tilde{v}))$ must maximize (4.11).

Next, suppose there exists α^* satisfying (4.8). We will show that

$$X^*(v) := T^{\nu,\alpha^*}(v) \tag{4.31}$$

$$P^*(y) := \frac{\mathbb{E}\big[\tilde{v} f_Z(y - T^{\nu,\alpha^*}(\tilde{v}))\big]}{\mathbb{E}\big[f_Z(y - T^{\nu,\alpha^*}(\tilde{v}))\big]} \tag{4.32}$$

form an equilibrium. But this follows again from lemma 4.4.2 and lemma 4.2.3. □

4.4.1 Existence

Lemma 4.4.5. *Fix $\beta \in \mathcal{P}(\mathbb{R})$. Suppose $\tilde{v}$ is compactly supported on $[-K, K]$ and suppose that $\tau(x)$ is superlinear. Then,*

$$\alpha \mapsto M(\alpha, \nu) - \int \big(C_\beta(x) + \tau(x)\big)\, \alpha(\mathrm{d}x).$$

is uniformly bounded from above.

Proof. Since $\tilde{v}$ is supported on $[-K, K]$, we have

$$\big|M(\alpha, \nu)\big| \leq K\mathbb{E}\big[|X|\big]$$

and

$$\left|\int C_\beta(x)\alpha(\mathrm{d}x)\right| \leq \int \big|C_\beta(x)\big|\, \alpha(\mathrm{d}x) = \int \big|xP(x)\big|\, \alpha(\mathrm{d}x) \leq K \int |x|\, \alpha(\mathrm{d}x),$$

whence

$$\left|M(\alpha, \nu) - \int \big(C_\beta(x) + \tau(x)\big)\, \alpha(\mathrm{d}x)\right| - \int \tau(x)\alpha(\mathrm{d}x) \leq \int \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x). \quad (4.33)$$

Since $\tau(x)$ is superlinear, we may find $x_0 \in \mathbb{R}$ for which $|x| \geq x_0$ implies $\tau(x) \geq 2K|x|$. For this x_0, we have

$$\begin{aligned}\int \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) &= \int_{[-x_0, x_0]} \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) + \int_{[-x_0, x_0]^\complement} \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) \\ &\leq \int_{[-x_0, x_0]} 2K|x|\, \alpha(dx) + 0 \leq 2K|x_0|\,.\end{aligned}$$

Thus,

$$M(\alpha, \nu) - \int \big(C_\beta(x) + \tau(x)\big)\, \alpha(\mathrm{d}x) \leq 2K|x_0|\,.$$

□

Remark 4.4.6. By virtue of the above proposition, we know that

$$\sup_\alpha \left(M(\alpha, \nu) - \int \big(C_\beta(x) + \tau(x)\big)\, \alpha(\mathrm{d}x)\right)$$

exists and is finite. Hereafter we shall denote it $\pi^*(\beta)$.

Lemma 4.4.7. *Fix $\beta \in \mathcal{P}(\mathbb{R})$. Let $\varepsilon \in \big(0, \pi^*(\beta)\big)$ and let x_0 be such that $|x| \geq x_0$ implies $\tau(x) \geq 3K|x|$. Then, each α from the set*

$$\mathcal{A}_\varepsilon := \left\{ \alpha \in \mathcal{P}(\mathbb{R}) \colon M(\alpha, \nu) - \int \big(C_\beta(x) + \tau(x)\big)\, \alpha(\mathrm{d}x) \geq \pi^*(\beta) - \varepsilon \right\} \tag{4.34}$$

satisfies the uniform first moment bound

$$\int |x|\, \alpha(\mathrm{d}x) \leq 4|x_0|\,. \tag{4.35}$$

In particular, the bound (4.35) does not depend on β or ε.

Proof. Let $\alpha \in \mathcal{A}_\varepsilon$. Choose x_0 as above. Then,

$$\begin{aligned}
&\int \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) \\
&\quad = \int_{[-x_0, x_0]} \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) + \int_{[-x_0, x_0]^{\complement}} \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) \\
&\quad \leq \int_{[-x_0, x_0]} \big(2K|x| - \tau(x)\big)\, \alpha(\mathrm{d}x) - K \int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x) \\
&\quad \leq \int_{[-x_0, x_0]} 2K|x|\, \alpha(\mathrm{d}x) - K \int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x) \\
&\quad \leq 2K|x_0| - K \int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x).
\end{aligned}$$

Since $\alpha \in \mathcal{A}_\varepsilon$, the upper bound (4.33) gives

$$\pi^*(\beta) - \epsilon \leq 2K|x_0| - K \int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x),$$

or

$$\int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x) \leq 2|x_0| - \frac{1}{K} \left(\pi^*(\beta) - \epsilon\right).$$

Thus,

$$\begin{aligned}
\int |x|\, \alpha(\mathrm{d}x) &= \int_{[-x_0, x_0]^{\complement}} |x|\, \alpha(\mathrm{d}x) + \int_{[-x_0, x_0]} |x|\, \alpha(\mathrm{d}x) && (4.36) \\
&\leq 2|x_0| - \frac{1}{K} \left(\pi^* - \epsilon\right) + \int_{[-x_0, x_0]} |x|\, \alpha(\mathrm{d}x) && (4.37) \\
&\leq 4|x_0| - \frac{1}{K} \left(\pi^* - \epsilon\right) < 4|x_0|\,. && (4.38)
\end{aligned}$$

□

Lemma 4.4.8. *Let x_0 be as in Lemma 4.4.7. Then, the set*

$$\mathcal{M} := \left\{ \alpha \in \mathcal{P}(\mathbb{R}) \colon \int |x| \, \alpha(\mathrm{d}x) \leq 4|x_0| \right\} \tag{4.39}$$

is nonempty, convex, and compact in $\mathcal{P}(\mathbb{R})$ equipped with the weak topology.

Proof. Nonemptiness is clear, e.g. $\delta_0 \in \mathcal{M}$. Convexity follows from linearity of the moment constraint: if $\alpha_1, \alpha_2 \in \mathcal{M}$ and $t \in [0,1]$, then

$$\int |x| \, \big(t\alpha_1 + (1-t)\alpha_2\big)(\mathrm{d}x) = t \int |x| \, \alpha_1(\mathrm{d}x) + (1-t) \int |x| \, \alpha_2(\mathrm{d}x) \leq 4|x_0|.$$

For compactness, note first that $\mathcal{M}$ is tight, because a uniform first–moment bound implies tightness in $\mathcal{P}(\mathbb{R})$. Hence $\mathcal{M}$ is relatively compact in the weak topology by Prokhorov's theorem. If $(\alpha_n)_n \subset \mathcal{M}$ and $\alpha_n \Rightarrow \alpha$, then by Fatou's lemma

$$\int |x| \, \alpha(\mathrm{d}x) \leq \liminf_{n \to \infty} \int |x| \, \alpha_n(\mathrm{d}x) \leq 4|x_0|,$$

so $\alpha \in \mathcal{M}$. Thus $\mathcal{M}$ is closed and relatively compact, hence compact. □

For $\alpha, \beta \in \mathcal{P}(\mathbb{R})$ define

$$F(\alpha, \beta) := M(\alpha, \nu) - \int_{\mathbb{R}} C_\beta(x) \, \alpha(\mathrm{d}x) - \int_{\mathbb{R}} \tau(x) \, \alpha(\mathrm{d}x), \tag{4.40}$$

and

$$\pi^*(\beta) := \sup_{\alpha \in \mathcal{P}(\mathbb{R})} F(\alpha, \beta).$$

Lemma 4.4.9. *For each $\beta \in \mathcal{M}$ define the best–response set*

$$\Gamma(\beta) := \arg\max_{\alpha \in \mathcal{M}} F(\alpha, \beta).$$

Then:

1. *For every $\beta \in \mathcal{M}$, the set $\Gamma(\beta)$ is nonempty, convex, and compact, and $\Gamma(\beta) \subset \mathcal{M}$.*
2. *The map $\Gamma : \mathcal{M} \mapsto 2^{\mathcal{M}}$ has a closed graph in $\mathcal{M} \times \mathcal{M}$.*

Proof. (1) Nonemptiness and compactness. Fix $\beta \in \mathcal{M}$ and, for $\varepsilon > 0$, let $\mathcal{A}_\varepsilon(\beta)$ be as in (4.34). By Lemma 4.4.7 we have $\mathcal{A}_\varepsilon(\beta) \subset \mathcal{M}$ for all $\varepsilon > 0$.

Set $\varepsilon_n := 1/n$ and choose $\alpha_n \in \mathcal{A}_{\varepsilon_n}(\beta)$ so that

$$F(\alpha_n, \beta) \geq \pi^*(\beta) - \tfrac{1}{n}.$$

By Lemma 4.4.7 each α_n belongs to $\mathcal{M}$, and by Lemma 4.4.8 the sequence $\{\alpha_n\}$ admits a weakly convergent subsequence, still denoted α_n, with

$$\alpha_n \Rightarrow \alpha^\beta \in \mathcal{M}.$$

By the standing continuity and growth assumptions on M, C_β, and τ, the map $(\alpha, \beta) \mapsto F(\alpha, \beta)$ is continuous on $\mathcal{M} \times \mathcal{M}$ for the weak topology. Hence

$$F(\alpha^\beta, \beta) = \lim_{n\to\infty} F(\alpha_n, \beta) \geq \lim_{n\to\infty} \left(\pi^*(\beta) - \tfrac{1}{n}\right) = \pi^*(\beta).$$

Since $F(\alpha^\beta, \beta) \leq \pi^*(\beta)$ by definition of $\pi^*(\beta)$, we conclude that α^β attains the supremum over $\mathcal{M}$. Thus $\Gamma(\beta)$ is nonempty and $\Gamma(\beta) \subset \mathcal{M}$.

The functional $F(\cdot, \beta)$ is affine and continuous in α on $\mathcal{M}$; therefore the maximizer set $\Gamma(\beta)$ is convex and closed in $\mathcal{M}$. By Lemma 4.4.8, $\mathcal{M}$ is compact, hence $\Gamma(\beta)$ is compact.

(2) Closed graph. Define, for each fixed $\gamma \in \mathcal{M}$,

$$K_\gamma := \big\{(\beta, \alpha) \in \mathcal{M} \times \mathcal{M} : F(\alpha, \beta) \geq F(\gamma, \beta)\big\}.$$

Then

$$\mathrm{Gr}(\Gamma) = \big\{(\beta, \alpha) \in \mathcal{M} \times \mathcal{M} : \alpha \in \Gamma(\beta)\big\} = \bigcap_{\gamma\in\mathcal{M}} K_\gamma.$$

Fix $\gamma \in \mathcal{M}$ and suppose $(\beta_n, \alpha_n) \in K_\gamma$ with $(\beta_n, \alpha_n) \to (\beta, \alpha)$ in $\mathcal{M} \times \mathcal{M}$. By definition,

$$F(\alpha_n, \beta_n) \geq F(\gamma, \beta_n), \qquad n \geq 1.$$

Using the joint continuity of F on $\mathcal{M} \times \mathcal{M}$ we obtain

$$F(\alpha, \beta) = \lim_{n\to\infty} F(\alpha_n, \beta_n) \geq \lim_{n\to\infty} F(\gamma, \beta_n) = F(\gamma, \beta),$$

so $(\beta, \alpha) \in K_\gamma$. Thus each K_γ is closed, and hence $\mathrm{Gr}(\Gamma)$ is closed as an intersection of closed sets. □

Theorem 4.4.10. *Let x_0 be as in Lemma 4.4.7. Then, the fixed point problem*

$$\alpha^* = \arg\max_{\alpha\in\mathcal{P}(\mathbb{R})} \left(M(\alpha, \nu) - \int C_{\alpha^*}(x)\alpha(\mathrm{d}x) - \int \tau(x)\alpha(\mathrm{d}x)\right) \tag{4.41}$$

admits a solution in the set

$$\mathcal{M} = \left\{ \alpha \in \mathcal{P}(\mathbb{R}) \colon \int |x| \, \alpha(\mathrm{d}x) \leq 4|x_0| \right\}.$$

Proof. By Lemma 4.4.8, $\mathcal{M}$ is a nonempty convex compact subset of $\mathcal{P}(\mathbb{R})$ with the weak topology. By Lemma 4.4.9, the correspondence $\Gamma : \mathcal{M} \mapsto 2^{\mathcal{M}}$ has nonempty convex compact values and a closed graph. Kakutani's fixed point theorem therefore yields $\alpha^* \in \mathcal{M}$ such that $\alpha^* \in \Gamma(\alpha^*)$.

Unwinding the definition of Γ, this means precisely that

$$\alpha^* = \arg\max_{\alpha \in \mathcal{M}} F(\alpha, \alpha^*).$$

Since enlarging the domain of maximization from $\mathcal{M}$ to $\mathcal{P}(\mathbb{R})$ does not change the maximizer (any maximizer must lie in $\mathcal{M}$ by Lemma 4.4.7), the same α^* solves

$$\alpha^* = \arg\max_{\alpha \in \mathcal{P}(\mathbb{R})} F(\alpha, \alpha^*),$$

and the claimed form follows from the definition (4.40) of F together with the bound $\alpha^* \in \mathcal{M}$. □

4.5 Conclusion

This chapter revisited Kyle's original one-shot model and showed that its equilibrium can be formulated as a distributional fixed-point problem. Rather than relying on Gaussian assumptions or explicit functional forms, the equilibrium emerges from the mutual consistency between the insider's optimal trading decision and the market maker's pricing rule. This viewpoint highlights the essential mechanism of the model: the insider chooses a demand schedule that is optimal given prices, and the market maker sets prices that are rational given the insider's strategy, with equilibrium occurring precisely at the point where these responses reinforce one another.

A key advantage of the fixed-point formulation is its flexibility. Because it does not require normality of the noise or the asset value, it naturally accommodates equilibria beyond the classical Gaussian setting, including those of Foster and Viswanathan [FV93], Cho and El Karoui [CK00], and Carré, Dufresne, and Gabriel [CCDG22]. More broadly, this perspective clarifies how static trading models fit into the transport-based and convex-analytic themes developed throughout the dissertation: the core properties of monotonicity,

comonotonic structure, and information–price consistency are preserved even in this discrete, single-round environment.

The fixed-point approach also suggests a constructive route for computation. By iteratively updating the insider's demand distribution and the market maker's pricing rule, one may approximate equilibria numerically in more general discrete-time or nonlinear settings. This opens the door to future extensions, including multi-period trading, non-Gaussian noise, transaction costs, and richer forms of strategic behavior, all within a conceptually simple and economically transparent framework.

CHAPTER 5

Summary

This dissertation developed a unified framework linking optimal transport theory and the models of informed trading initiated by Kyle (1985) and Back (1992). Each chapter pursued a distinct generalization of the classical model, extending its scope from static Gaussian settings to dynamic information, nonlinear liquidation, and stochastic liquidity models. The common thread across these developments is the use of convex duality and transport potentials to represent equilibrium prices, trading strategies, and the evolution of information in markets.

5.1 Main Contributions

Optimal transport formulation of the Kyle–Back model. Chapter 1 is a literature review. It examines the connection between the continuous–time Kyle–Back equilibrium and the Monge–Kantorovich problem with quadratic cost. From the Brenier potential φ we can construct everything from the market maker's pricing rule to the informed trader's profit to price sensitivity. This connection reproduces all known properties of the classical equilibrium – inconspicuous trading, price martingality, and movement along the residual supply curve – but expressed through the geometry of optimal transport. The gradient $\nabla\varphi$ provides an explicit map between the terminal value of the asset and the aggregate order flow, offering a new interpretation of market depth and price sensitivity as derivatives of a convex potential.

Dynamic information and activist trading. Chapter 2 generalized the model to allow the informed trader's private signal to evolve dynamically through time and for the informed trader to influence the firm's terminal value. The terminal liquidation value of the informed trader is represented by a nonlinear function $V(\beta + X_T, S_T)$, capturing activist or strategic intervention effects. By combining Monge–Kantorovich duality with backward stochastic partial differential equations (BSPDEs), we characterized the equilibrium

in terms of Kantorovich potentials and the solution of a backward Kushner–Zakai filtering equation. This formulation unifies the classical model of Back, Cocquemas, Ekren, and Lioui (2021) with dynamic-information models of Back and Pedersen (1998) and Campi, Çetin, and Danilova (2010) with the activist-trading model of Back, Dufresne, Fos, Li, and Ljungqvist (2018). Economically, it shows how information flow, price sensitivity, and liquidation payout can all be viewed as facets of a single transport problem with a generalized cost function. Our findings have been submitted to ArXiv as [CENV22].

Stochastic and multidimensional liquidity. Chapter 3 introduced a variational formulation of the Kyle model that accommodates both stochastic volatility of noise trading and multiple correlated assets. In this formulation, the insider optimally "liquidates" information rather than inventory: her control variable is the rate of decline of the market maker's conditional variance Σ_t. The primal problem, posed as a maximization over decreasing covariance paths $\dot{\Sigma}_t < 0$, represents the feasible release of information, thereby linking market microstructure to optimal execution theory. The dual problem introduces a matrix-valued martingale interpreted as the shadow price of information. The model admits a multidimensional inconspicuous-trading property and recovers the equilibria of Kyle, Back, Back and Pedersen, and Collin–Dufresne and Fos as specific cases.

Kyle's One-Shot Model. Chapter 4 revisited Kyle's original single-auction framework and showed that its equilibrium can be described through a fixed-point formulation. In this setting, the insider's trading decision and the market maker's pricing rule must be mutually consistent, and the equilibrium is obtained at the point where these responses reinforce one another. A key advantage of this approach is that it does not rely on Gaussian assumptions: neither the asset value nor the noise-trader distribution must be normal. The fixed-point formulation therefore accommodates a much broader class of equilibria, including those studied by Foster and Viswanathan [FV93], Cho and El Karoui [CK00], and Carré, Dufresne, and Gabriel [CCDG22]. This perspective also offers a constructive way to approximate equilibria numerically, which we discuss further in section 5.3.

5.2 Conceptual Implications

Taken together, the three chapters demonstrate that the structure of continuous-time information equilibria is inherently geometric. The convex potentials and transport maps arising from optimal transport encode the relationship between order flow, prices, and private information. The duality between the insider and the market maker mirrors the duality between

primal and conjugate transport problems, providing a common mathematical language for diverse market phenomena: price impact, inventory risk, dynamic information, and stochastic liquidity.

5.3 Future Directions

Several natural extensions of this work suggest themselves.

Discrete-time models and fixed-point methods. Although Chapters 1–3 are formulated in continuous time, Chapter 4 shows that the ideas of this dissertation also apply in a purely discrete-time, single-auction setting. A natural next step is to develop multi-period discrete-time Kyle models in which each trading round is governed by a fixed-point interaction between the insider's trading rule and the market maker's pricing rule. The one-shot formulation in Chapter 4 suggests iterative numerical schemes: starting from an initial guess for the insider's demand distribution and the pricing rule, one can update each in turn until they stabilize. Extending this approach to multi-round trading, to non-Gaussian noise, or to models with transaction costs and other frictions could provide a flexible computational framework for discrete-time market microstructure and clarify how the continuous-time limits studied in earlier chapters emerge from underlying discrete dynamics.

Multiple insiders. The optimal-transport approach also extends to settings with several informed traders, such as the model of Back, Cao, and Willard (2000). In this case, the equilibrium corresponds to a multi-marginal optimal transport problem, where the convex potential encodes the joint strategic interaction among insiders. The dual variables of this problem naturally represent each insider's marginal value of information and clarify how competition affects liquidity and price impact.

Static mean field games. The fixed-point problem developed in Chapter 4 closely resembles the equilibrium condition in static mean field games. This connection suggests a promising avenue: by drawing on the extensive literature on mean field games, one may establish the existence of Kyle's one-shot equilibrium for very general distributions, requiring only minimal assumptions on the noise or the asset value. The main challenge in pursuing this direction is that the insider's action space is not compact, which complicates the application of standard fixed-point theorems.

5.4 Final Remarks

The central message of this dissertation is that optimal transport provides a unifying mathematical framework for understanding informed trading and price formation in continuous time. Through convex duality, it reveals deep connections between information, liquidity, and volatility—concepts that are often treated separately in financial economics. By viewing equilibrium as an optimal coupling between beliefs and prices, the transport approach not only clarifies existing models but also opens new directions for the study of market microstructure, information design, and the geometry of stochastic processes.

APPENDIX A

Probability Theory

A.1 The Gaussian Projection Theorem

Theorem A.1.1. *Let $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$ be Gaussian vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$. Suppose $(\boldsymbol{X}, \boldsymbol{Y})$ is jointly Gaussian. Then, the distribution of $\boldsymbol{X}$ conditioned on $\boldsymbol{Y}$ is Gaussian with mean*

$$\mathbb{E}\big[\boldsymbol{X} \mid \boldsymbol{Y}\big] = \boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\left(\boldsymbol{Y} - \boldsymbol{\mu}_Y\right)$$

and variance

$$\mathrm{Var}\big(\boldsymbol{X} \mid \boldsymbol{Y}\big) = \boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX}$$

where

$$\boldsymbol{\Sigma}_{XY} := \begin{bmatrix} \mathrm{Cov}(X_1, Y_1) & \cdots & \mathrm{Cov}(X_1, Y_m) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, Y_1) & \cdots & \mathrm{Cov}(X_n, Y_m) \end{bmatrix}$$
$$\boldsymbol{\Sigma}_{YX} := \boldsymbol{\Sigma}_{XY}^{\top}.$$

Proof. Put $\boldsymbol{Y}^{\perp} := \boldsymbol{X} - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}$. Then, the pair $\big(\boldsymbol{Y}^{\perp}, \boldsymbol{Y}\big)$ is jointly Gaussian, and $\boldsymbol{Y}^{\perp}$ is uncorrelated with $\boldsymbol{Y}$:

$$\mathrm{Cov}\Big(\boldsymbol{Y}^{\perp}, \boldsymbol{Y}\Big) = \mathrm{Cov}\big(\boldsymbol{X} - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}, \boldsymbol{Y}\big) = \boldsymbol{0}.$$

Thus $\boldsymbol{Y}^{\perp}$ is independent from $\boldsymbol{Y}$. In addition, we compute

$$\mathbb{E}\Big[\boldsymbol{Y}^{\perp}\Big] = \boldsymbol{\mu}_X - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\mu}_Y$$
$$\mathrm{Var}\Big(\boldsymbol{Y}^{\perp}\Big) = \boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX}$$

Fix $\mathbf{t} \in \mathbb{R}^n$. We have

$$\begin{aligned}
\mathbb{E}\Big[e^{\mathbf{t}^\top \boldsymbol{X}} \,\Big|\, \boldsymbol{Y}\Big] &= \exp\Big(\mathbf{t}^\top \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}\Big)\,\mathbb{E}\Big[\exp\Big(\mathbf{t}^\top \boldsymbol{Y}^\perp\Big)\,\Big|\,\boldsymbol{Y}\Big] \\
&= \exp\Big(\mathbf{t}^\top \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{Y}\Big)\,\mathbb{E}\Big[\exp\Big(\mathbf{t}^\top \boldsymbol{Y}^\perp\Big)\Big] \\
&= \exp\Big(\mathbf{t}^\top\left(\boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\left(\boldsymbol{Y}-\boldsymbol{\mu}_Y\right)\right) + \mathbf{t}^\top\left(\boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX}\right)\mathbf{t}\Big),
\end{aligned}$$

which proves the result. □

Remark A.1.2. The above is a linear algebraic proof. A probabilistic proof using Baye's theorem is also possible.

Example A.1.3. Let $X \sim \mathcal{N}\big(\mu, \sigma^2\big)$ and $\epsilon \sim \mathcal{N}\big(0, \delta^2\big)$ be independent. If $Y = X + \epsilon$ is observed (but not X or ϵ), then

$$X \sim \mathcal{N}\left(\frac{\sigma^2 y + \delta^2 \mu}{\sigma^2+\delta^2}, \frac{\sigma^2\delta^2}{\sigma^2+\delta^2}\right) \qquad \{Y = y\}\ \mathbb{P}\text{-a.s.}$$

Example A.1.4. Let $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathcal{N}(0,1)$. If $X = \sum_{i=1}^n a_i Z_i$ and $Y = \sum_{i=1}^n b_i Z_i$ for arbitrary $\{a_i\}, \{b_i\}$, then

$$X|Y \sim \mathcal{N}\left(\frac{\sum a_i b_i}{\sum b_i^2} Y, \frac{\left(\sum a_i^2\right)\left(\sum b_i^2\right) - \left(\sum a_i b_i\right)^2}{\sum b_i^2}\right)$$

A.2 The Feynman Kac Representation

A.2.1 Examples

To explain the Feynman-Kac representation theorem, it is best to start with examples.

Example A.2.1. Consider the Cauchy problem

$$\begin{aligned}
\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x) = 0, \qquad (t,x) \in (0,T)\times\mathbb{R} \\
u(T,x) = x^2
\end{aligned}$$

Assume that a sufficiently regular solution $u(t,x)$ exists. Let $\{B_t\}$ be a Brownian motion

and consider the process $u(t, B_t)$. By Itô's formula,

$$\mathrm{d}u(t, B_t) = \underbrace{\left(\frac{\partial u}{\partial t}(t, B_t) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, B_t) \right)}_{0} \mathrm{d}t + \frac{\partial u}{\partial x}(t, B_t)\, \mathrm{d}B_t.$$

Integrating from $s \in (0, T)$ to T gives

$$\underbrace{u(T, B_T)}_{B_T^2} - u(s, B_s) = \int_s^T \frac{\partial u}{\partial x}(t, B_t)\, \mathrm{d}B_t.$$

Now apply $\mathbb{E}\big[\, \cdot \mid B_s = x\big]$ to both sides:

$$\mathbb{E}\big[B_T^2 \mid B_s = x\big] - \underbrace{\mathbb{E}\big[u(s, B_s) \mid B_s = x\big]}_{u(s,x)} = \underbrace{\mathbb{E}\left[\int_s^T \frac{\partial u}{\partial x}(t, B_t)\, \mathrm{d}B_t \,\middle|\, B_s = x \right]}_{0}. \tag{A.1}$$

Thus we obtain the Feynman–Kac representation

$$u(s, x) = \mathbb{E}\big[B_T^2 \mid B_s = x\big].$$

Since $B_T \mid \{B_s = x\}$ is Gaussian with mean x and variance $T - s$, we have

$$\mathrm{Var}\big(B_T \mid B_s = x\big) = \mathbb{E}\big[B_T^2 \mid B_s = x\big] - \underbrace{\mathbb{E}\big[B_T \mid B_s = x\big]}_{x}^2$$

so that $\mathbb{E}\big[B_T^2 \mid B_s = x\big] = x^2 + T - s$. Because $s \in [0, T)$ is arbitrary, relabeling s as t yields the solution

$$\boxed{u(t, x) = x^2 + T - t}$$

Remark A.2.2. If we replaced the terminal condition $u(T, x) = x^2$ with $u(T, x) = g(x)$, all of the steps above would be the same. We would arrive at $u(s, x) = \mathbb{E}\big[g(B_T) \mid B_s = x\big]$ instead of $u(s, x) = \mathbb{E}\big[B_T^2 \mid B_s = x\big]$.

Remark A.2.3. In order to claim that the expectation vanishes in (A.1), we technically should have checked that $\int_s^T \frac{\partial u}{\partial x}(t, B_t)\, \mathrm{d}B_t$ is integrable, which is to say

$$\mathbb{E}\left[\left| \int_s^T \frac{\partial u}{\partial x}(t, B_t)\, \mathrm{d}B_t \right| \right] < \infty.$$

However, we will not worry about such integrability issues here.

Example A.2.4. Consider the PDE

$$\begin{aligned}\frac{\partial u}{\partial t}(t,x)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x)-u(t,x)&=0\\ u(T,x)&=x^2.\end{aligned}$$

Multiply both sides of the PDE by the integrating factor e^{-t} to obtain

$$\frac{\partial}{\partial t}\big(e^{-t}u(t,x)\big)+\frac{1}{2}\frac{\partial^2}{\partial x^2}\big(e^{-t}u(t,x)\big)=0.$$

This suggests the change of variables $v(t,x):=e^{-t}u(t,x)$, whereupon the PDE reads

$$\begin{aligned}\frac{\partial v}{\partial t}(t,x)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x)&=0\\ v(T,x)&=e^{-T}x^2.\end{aligned}$$

This is precisely the form of example A.2.1 with $g(x)=e^{-T}x^2$. Thus, we know that

$$v(t,x)=\mathbb{E}\Big[e^{-T}B_T^2\,\Big|\,B_t=x\Big]=e^{-T}\left(x^2+T-t\right),$$

but $v(t,x)=e^{-t}u(t,x)$, and so

$$\boxed{u(t,x)=e^{-(T-t)}(x^2+T-t)}$$

Example A.2.5. Consider the PDE

$$\begin{aligned}\frac{\partial u}{\partial t}(t,x)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x)-xu(t,x)&=0\\ u(T,x)&=1.\end{aligned}$$

Let $\{B_t\}$ denote as before a standard Brownian motion, and $u(t,x)$ a solution to the above PDE. We compute

$$\mathrm{d}u(t,B_t)=\underbrace{\left(\frac{\partial u}{\partial t}(t,B_t)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,B_t)\right)}_{B_t u(t,B_t)}\mathrm{d}t+\frac{\partial u}{\partial x}(t,B_t)\,\mathrm{d}B_t. \tag{A.2}$$

Introducing the integrating factor

$$\exp\left(-\int_0^t B_s \,\mathrm{d}s\right),$$

eq. (A.2) becomes

$$\mathrm{d}\left(\exp\left(-\int_0^t B_s \,\mathrm{d}s\right) u(t, B_t)\right) = \exp\left(-\int_0^t B_s \,\mathrm{d}s\right) \frac{\partial u}{\partial x}(t, B_t) \,\mathrm{d}B_t.$$

Integrating again from $r \in (0, T)$ to T, we find

$$\begin{aligned}\exp\left(-\int_0^T B_s \,\mathrm{d}s\right) \underbrace{u(T, B_T)}_{1} - \exp\left(-\int_0^r B_s \,\mathrm{d}s\right) u(r, B_r) \\ = \int_r^T \exp\left(-\int_0^t B_s \,\mathrm{d}s\right) \frac{\partial u}{\partial x}(t, B_t) \,\mathrm{d}B_t.\end{aligned}$$

We apply $\mathbb{E}\big[\,\cdot \mid B_r = x\big]$ to both sides. The RHS integral vanishes, and after switching the dummy variable r with t, we are left with

$$u(t, x) = \mathbb{E}\left[\exp\left(-\int_t^T B_s \,\mathrm{d}s\right) \,\middle|\, B_t = x\right].$$

This is the MGF of the random variable $\int_s^T B_s \,\mathrm{d}s$ conditioned on $B_t = x$ and evaluated at -1. It turns out that conditioned on $\{B_t = x\}$,

$$\int_t^T B_s \,\mathrm{d}s \sim \mathcal{N}\left(x(T-t), \frac{1}{3}(T-t)^3\right), \tag{A.3}$$

and so

$$\boxed{u(t, x) = \exp\left(-x(T-t) + \frac{1}{6}(T-t)^3\right)}$$

Remark A.2.6. To show (A.3), first integrate by parts to obtain

$$\int_t^T B_s \,\mathrm{d}s = B_T T - B_t t - \int_t^T s \,\mathrm{d}B_s$$

by which we see that $\int_t^T B_s \, \mathrm{d}s$ is Gaussian. Its mean is

$$\mathbb{E}\left[\int_t^T B_s \, \mathrm{d}s \,\middle|\, B_t = x\right] = \mathbb{E}\left[B_T T - B_t t - \int_t^T s \, \mathrm{d}B_s \,\middle|\, B_t = x\right] = x(T-t).$$

To compute its variance, introduce $0 = -B_t T + B_t T$, so that we may write

$$B_T T - B_t t - \int_t^T s \, \mathrm{d}B_s = B_t(T-t) + \int_t^T (T-s) \, \mathrm{d}B_s.$$

The point is that the summands are now independent, and we may compute

$$\begin{aligned}
\mathrm{Var}\left(B_t(T-t) + \int_t^T (T-s) \, \mathrm{d}B_s \,\middle|\, B_t = x\right) &= \mathrm{Var}\left(\int_t^T (T-s) \, \mathrm{d}B_s\right) \\
&= \mathbb{E}\left[\left(\int_t^T (T-s) \, \mathrm{d}B_s\right)^2\right] \\
&= \mathbb{E}\left[\int_t^T (T-s)^2 \, \mathrm{d}s\right] \\
&= \frac{1}{3}(T-t)^3.
\end{aligned}$$

Example A.2.7 (Black-Scholes Equation)**.** Consider the PDE

$$\begin{aligned}
\frac{\partial V}{\partial t}(t,s) + rs\,\frac{\partial V}{\partial s}(t,s) + \frac{1}{2}\sigma^2 s^2\,\frac{\partial^2 V}{\partial s^2}(t,s) - rV(t,s) &= 0 \\
V(T,s) &= \max\{s-K, 0\}.
\end{aligned}$$

Introduce as in example A.2.4 the integrating factor e^{-rt}. The PDE reads

$$\frac{\partial}{\partial t}\big(e^{-rt}V(t,s)\big) + rs\,\frac{\partial}{\partial s}\big(e^{-rt}V(t,s)\big) + \frac{1}{2}\sigma^2 s^2\,\frac{\partial^2}{\partial s^2}\big(e^{-rt}V(t,s)\big) = 0$$

which suggests the change of variables $U(t,s) := e^{-rt}V(t,s)$. The PDE becomes

$$\begin{aligned}
\frac{\partial U}{\partial t}(t,s) + rs\,\frac{\partial U}{\partial s}(t,s) + \frac{1}{2}\sigma^2 s^2\,\frac{\partial^2 U}{\partial s^2}(t,s) &= 0 \\
U(T,s) &= e^{-rT}\max\{s-K, 0\}.
\end{aligned}$$

Now, let $\{S_t\}$ be a geometric Brownian motion which diffuses according to

$$\mathrm{d}S_t = rS_t \, \mathrm{d}t + \sigma S_t \, \mathrm{d}B_t.$$

If U is a solution to the PDE, then Itô's lemma gives

$$\mathrm{d}U(t,S_t) = \underbrace{\left(\frac{\partial U}{\partial t}(t,S_t) + rS_t\,\frac{\partial U}{\partial s}(t,S_t) + \frac{1}{2}\sigma^2 S_t^2\,\frac{\partial^2 U}{\partial s^2}(t,S_t)\right)}_{0}\mathrm{d}t + \sigma S_t\,\frac{\partial U}{\partial s}(t,S_t)\,\mathrm{d}B_t.$$

Integrating from $r \in (0,T)$ to T, we find

$$\underbrace{e^{-rT}\max\{S_T-K,0\}}_{U(T,S_T)} - U(r,S_r) = \int_r^T \sigma S_t\,\frac{\partial U}{\partial s}(t,S_t)\,\mathrm{d}B_t.$$

Applying $\mathbb{E}\big[\,\cdot \mid S_t = s\big]$, the stochastic integral vanishes, and are are left with

$$U(t,s) = \mathbb{E}\Big[e^{-T}\max\{S_T-K,0\}\,\Big|\,S_t=s\Big],$$

or, since $U(t,s) := e^{-rt}V(t,s)$,

$$V(t,s) = e^{-(T-t)}\mathbb{E}\big[\max\{S_T-K,0\}\,\big|\,S_t=s\big].$$

From here, we can show that

$$\boxed{V(t,s) = sN(d_1) - Ke^{-r(T-t)}N(d_2)}$$

where

$$d_1 = \frac{\log\frac{S}{K} + \left(r+\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = \frac{\log\frac{S}{K} + \left(r-\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}.$$

For details on this calculation, see proposition A.4.14.

A.2.2 Statement of Feynman-Kac Theorem

This derivation is from proposition 5.6 of Bjork [Bjo20].

Theorem A.2.8. *Consider the PDE*

$$\frac{\partial F}{\partial t}(t,x) + \mu(t,x)\frac{\partial F}{\partial x}(t,x) + \frac{1}{2}\sigma(t,x)^2\frac{\partial^2 F}{\partial x^2}(t,x) - rF(t,x) = 0 \tag{A.4}$$

subject to the terminal condition

$$F(T,x) = \Phi(x). \tag{A.5}$$

Then, the solution F may be written as the conditional expectation

$$F(t,x) = e^{-r(T-t)}\mathbb{E}\big[\Phi(X_T) \mid X_t = x\big], \tag{A.6}$$

where X_t is an Itô process with dynamics

$$\mathrm{d}X_t = \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t. \tag{A.7}$$

Proof. As with example A.2.4 and example A.2.7, we introduce the integrating factor e^{-rt}. With the change of variables

$$u(t,x) := e^{-rt}F(t,x),$$

the PDE reads

$$\frac{\partial u}{\partial t}(t,x) + \mu(t,x)\,\frac{\partial u}{\partial x}(t,x) + \frac{1}{2}\sigma(t,x)^2\,\frac{\partial^2 u}{\partial x^2}(t,x) = 0 \tag{A.8}$$

$$u(T,x) = e^{-rT}\Phi(x). \tag{A.9}$$

Assume that a sufficiently regular solution $u(t,x)$ exists. We apply Itô's lemma to the process $u(t,X_t)$, where X_t is given by (A.7):

$$\begin{aligned}\mathrm{d}u(t,X_t) = &\left(\frac{\partial u}{\partial t}(t,X_t) + \mu(t,X_t)\,\frac{\partial u}{\partial x}(t,X_t) + \frac{1}{2}\sigma(t,X_t)^2\,\frac{\partial^2 u}{\partial x^2}(t,X_t)\right)\mathrm{d}t \\ &+ \sigma(t,X_t)\,\frac{\partial u}{\partial x}(t,X_t)\,\mathrm{d}W_t.\end{aligned} \tag{A.10}$$

Due to (A.8), the drift term in (A.10) vanishes, so (A.10) is actually

$$\mathrm{d}u(t,X_t) = \sigma(t,X_t)\,\frac{\partial u}{\partial x}(t,X_t)\,\mathrm{d}W_t. \tag{A.11}$$

Integrating (A.11) from t to T, we get

$$\underbrace{u(T,X_T)}_{e^{-rT}\Phi(X_T)} - u(t,X_t) = \int_t^T \sigma(s,X_s)\,\frac{\partial u}{\partial x}(s,X_s)\,\mathrm{d}W_s. \tag{A.12}$$

Now, we apply $\mathbb{E}\big[\,\cdot \mid X_t = x\big]$ to both sides of (A.12). The stochastic integral vanishes, and

we are left with

$$u(t,x) = e^{-rT}\mathbb{E}\big[\Phi(X_T) \mid X_t = x\big]. \tag{A.13}$$

But $u(t,x) = e^{-rt}F(t,x)$, so (A.13) is in fact just (A.6). □

The converse is also true:

Theorem A.2.9. *Define F by the conditional expectation* (A.6)*. Then, F solves the PDE* (A.4) *with terminal condition* (A.5)*.*

Proof. From (A.6) it follows that

$$F(t,X_t) = e^{-r(T-t)}\mathbb{E}\Big[\Phi(X_T)\Big|\mathscr{F}_t^X\Big], \tag{A.14}$$

or

$$e^{-rt}F(t,X_t) = \mathbb{E}\Big[e^{-rT}\Phi(X_T)\Big|\mathscr{F}_t^X\Big].$$

Thus $t \mapsto e^{-rt}F(t,X_t)$ is a martingale. By the martingale representation theorem, there exists some square integrable $\big\{Z_t \colon t \in [0,T]\big\}$ for which

$$e^{-rt}F(t,X_t) = e^{-rT}\Phi(X_T) - \int_t^T Z_s \,\mathrm{d}W_s.$$

Applying Itô's lemma on both sides, we find

$$e^{-rt}\,\mathrm{d}F(t,X_t) - re^{-rt}F(t,X_t)\,\mathrm{d}t = Z_t\,\mathrm{d}W_t. \tag{A.15}$$

Since

$$\begin{aligned}\mathrm{d}F(t,X_t) = &\left(\frac{\partial F}{\partial t}(t,X_t) + \mu(t,X_t)\,\frac{\partial F}{\partial x}(t,X_t) + \frac{1}{2}\sigma(t,X_t)^2\,\frac{\partial^2 F}{\partial x^2}(t,X_t)\right)\mathrm{d}t\\ &+ \sigma(t,X_t)\,\frac{\partial F}{\partial x}(t,X_t)\,\mathrm{d}W_t,\end{aligned}$$

(A.4) follows from equating $\mathrm{d}t$ terms in (A.15). The terminal condition (A.5) follows from taking $t = T$ in (A.14). □

A.3 The Kolmogorov Equations

The Kolmogorov equations, named after Soviet mathematician Andrey Kolmogorov, are PDEs that characterize the evolution of transition densities. Our derivation follows section 5.6 of [Bjo20].

A.3.1 Kolmogorov's Backward Equation

Let X_t be the solution to

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t. \tag{A.16}$$

Fix $t > 0$ and $B \in \mathscr{B}$. For $s \leq t$, define

$$P(s, y; t, B) = \mathbb{P}(X_t \in B | X_s = y) = \mathbb{E}\big[\mathbb{1}_B(X_t)\big|X_s = y\big]\,. \tag{A.17}$$

From the Feynman Kac representation, it follows that $P(s, y; t, B)$ satisfies the PDE

$$\begin{aligned} \frac{\partial P}{\partial s}(s, y; t, B) + \mu(s, y)\,\frac{\partial P}{\partial x}(s, y; t, B) + \frac{1}{2}\sigma(s, y)^2\,\frac{\partial^2 P}{\partial x^2}(s, y; t, B) = 0 \\ P(t, y; t, B) = \mathbb{1}_B(y). \end{aligned} \tag{A.18}$$

Now, $P(s, y; t, \cdot)$ is actually a measure on $\mathscr{B}$. We can obtain its RN derivative the usual way: fix $x \in \mathbb{R}$, $h > 0$, and put

$$p(s, y; t, x) = \frac{P\big(s, y; t, (x-h, x+h)\big)}{2h} = \mathbb{E}\left[\frac{\mathbb{1}_{(x-h,x+h)}(X_t)}{2h}\middle| X_s = y\right]. \tag{A.19}$$

If $\mathbb{P}(X_t \in \cdot | X_s = y)$ is absolutely continuous with respect to Lebesgue measure, letting $h \downarrow 0$ forces $p(s, y; t, x)$ to coincide with the density $f_{X_t|X_s}(x|y)$. Further, $p(s, y; t, x)$ satisfies

$$\begin{aligned} \frac{\partial p}{\partial s}(s, y; t, x) + \mu(s, y)\,\frac{\partial p}{\partial x}(s, y; t, x) + \frac{1}{2}\sigma(s, y)^2\,\frac{\partial^2 p}{\partial x^2}(s, y; t, x) = 0 \\ p(t, y; t, x) = \delta_x(y). \end{aligned} \tag{A.20}$$

The PDE (A.20) is Kolmogorov's backward equation, so called because it fixes the terminal condition $\{X_t = x\}$ and describes the evolution of the transition density before t. If we instead fix the initial condition $\{X_s = y\}$ and describe the evolution of p after s, we attain Kolmogorov's forward equation.

A.3.2 Kolmogorov's Forward Equation

Fix $s > 0$ and $y \in \mathbb{R}$. Pick $g \in C^2$ such that g vanishes at $\pm\infty$. If $t > s$, then Itô's formula gives

$$\begin{aligned} g(X_t) = g(X_s) &+ \int_s^t \mu(\tau, X_\tau)\,\frac{\partial g}{\partial x}(\tau, X_\tau) + \frac{1}{2}\sigma(\tau, X_\tau)^2\,\frac{\partial^2 g}{\partial x^2}(\tau, X_\tau)\,\mathrm{d}\tau \\ &+ \int_s^t \sigma(\tau, X_\tau)\,\frac{\partial g}{\partial x}(\tau, X_\tau)\,\mathrm{d}W_\tau. \end{aligned} \tag{A.21}$$

Applying $\mathbb{E}[\cdot|X_s = y]$ to both sides, the stochastic integral vanishes and we are left with

$$\mathbb{E}\big[g(X_t)\big|X_s = y\big] = g(y) + \int_s^t \mathbb{E}\left[\mu(\tau, X_\tau)\,\frac{\partial g}{\partial x}(\tau, X_\tau) + \frac{1}{2}\sigma(\tau, X_\tau)^2\,\frac{\partial^2 g}{\partial x^2}(\tau, X_\tau)\middle| X_s = y\right]\mathrm{d}\tau, \tag{A.22}$$

where we are allowed to switch the order of integration due to Fubini.

Conditioning over $\{X_\tau = u\}$, the expectation in the RHS integrand can be written as

$$\int_{-\infty}^{\infty}\left[\mu(\tau, u)\,\frac{\partial g}{\partial x}\,(\tau, u) + \frac{1}{2}\sigma(\tau, u)^2\,\frac{\partial^2 g}{\partial x^2}\,(\tau, u)\right] p(s, y; \tau, u)\,\mathrm{d}u. \tag{A.23}$$

Since g vanishes at $\pm\infty$, integration by parts on (A.23) gives

$$\int_{-\infty}^{\infty} g(\tau, u)\left[-\frac{\partial}{\partial x}\big[\mu(\tau, u)p(s, y; \tau, u)\big] + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\big[\sigma(\tau, u)^2 p(s, y; \tau, u)\big]\right]\mathrm{d}u. \tag{A.24}$$

Similar reasoning gives us an expression for the LHS of (A.22):

$$\int_{-\infty}^{\infty} g(u)p(s, y; t, u)\,\mathrm{d}u. \tag{A.25}$$

Now, if we pick $g(u) = \exp\!\Big(-\frac{(u-x)^2}{2h^2}\Big)$ and let $h \downarrow 0$, (A.24) becomes

$$-\frac{\partial}{\partial x}\,\big[\mu(t, x)p(s, y; \tau, x)\big] + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\,\big[\sigma(\tau, x)^2 p(s, y; \tau, x)\big] \tag{A.26}$$

and (A.25) becomes

$$p(s, y; t, x). \tag{A.27}$$

Putting these back into (A.22),

$$p(s,y;t,x) = \delta_x(y) + \int_s^t \frac{\partial}{\partial x}\left[\mu(\tau,x)p(s,y;\tau,x)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma(\tau,x)^2 p(s,y;\tau,x)\right] \mathrm{d}\tau. \quad \text{(A.28)}$$

Differentiating with respect to t, we obtain Kolmogorov's forward equation

$$\frac{\partial p}{\partial t}(s,y;t,x) = -\frac{\partial}{\partial x}\big[\mu(t,x)p(s,y;t,x)\big] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big[\sigma(t,x)^2 p(s,y;t,x)\big]\,. \quad \text{(A.29)}$$

A.4 Girsanov Theory

Girsanov theory describes how the distribution of a stochastic process—especially Brownian motion—changes when we tilt the probability measure by a suitably integrable exponential martingale. Before reaching the general theorem, it is helpful to first examine the simplest case: a Gaussian random variable. This illustrates the key idea that "adding drift" can be undone by reweighting the underlying probability measure through a likelihood ratio.

A.4.1 Gaussian Change of Measure

We begin with a one-dimensional Gaussian identity that captures the essence of Girsanov's theorem.

Theorem A.4.1 (Gaussian Shift Identity)**.** *If $Z \sim \mathcal{N}(0,1)$ and $\mu \in \mathbb{R}$, then for any measurable $f\colon \mathbb{R} \mapsto \mathbb{R}$ for which the expectations exist,*

$$\mathbb{E}\big[f(\mu+Z)\big] = \mathbb{E}\left[f(Z)\exp\left(\mu Z - \frac{\mu^2}{2}\right)\right]. \quad \text{(A.30)}$$

Proof. Start with the integral representation

$$\mathbb{E}\big[f(\mu+Z)\big] = \int_{\mathbb{R}} f(\mu+z)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\mathrm{d}z.$$

Substituting $z' = \mu + z$, the integral becomes

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z'-\mu)^2}{2}\right)\mathrm{d}z' = \int_{\mathbb{R}} \exp\left(\mu z' - \frac{\mu^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z')^2}{2}\right)\mathrm{d}z'$$

and taking the standard normal as the reference density yields the result. □

Remark A.4.2. Equation (A.30) expresses two equivalent viewpoints:

1. We can remove μ from the argument of f in $\mathbb{E}\big[f(\mu + Z)\big]$, but we need to compensate with the multiplicative factor $e^{\mu Z - \mu^2/2}$.

2. We can remove the multiplicative factor $e^{\mu Z - \mu^2/2}$, but we need to compensate with a μ in $\mathbb{E}\big[f(\mu + Z)\big]$.

Both directions are useful depending on the situation. The factor $e^{\mu Z - \mu^2/2}$ is called a *likelihood ratio*. To see why, let p_μ and p_0 denote the $\mathcal{N}(\mu, 1)$ and $\mathcal{N}(0, 1)$ densities. Then,

$$\frac{p_\mu(Z)}{p_0(Z)} = \exp\left(\mu Z - \frac{\mu^2}{2}\right).$$

In fact, the calculation above may have been performed slightly differently:

$$\begin{aligned}\mathbb{E}\big[f(\mu + Z)\big] &= \int_{\mathbb{R}} f(\mu + z) p_0(z)\,\mathrm{d}z = \int_{\mathbb{R}} f(z) p_\mu(z)\,\mathrm{d}z \\ &= \int_{\mathbb{R}} f(z)\frac{p_\mu(z)}{p_0(z)} p_0(z)\,\mathrm{d}z = \mathbb{E}\left[f(Z)\frac{p_\mu(Z)}{p_0(Z)}\right] = \mathbb{E}\left[f(Z)\exp\left(\mu Z - \frac{\mu^2}{2}\right)\right].\end{aligned}$$

This second line of reasoning explicitly introduces $p_\mu(Z)/p_0(Z)$, and it turns out to be more generalizable.

Example A.4.3 (Gaussian MGF's)**.** Setting $f := 1$ and $\mu := t$ in eq. (A.30) gives

$$1 = \mathbb{E}\left[\exp\left(tZ - \frac{t^2}{2}\right)\right],$$

so

$$M_Z(t) := \mathbb{E}\Big[e^{tZ}\Big] = e^{t^2/2}.$$

For $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$, the same identity yields

$$M_X(t) := \mathbb{E}\Big[e^{tX}\Big] = \mathbb{E}\Big[e^{t(\mu+\sigma Z)}\Big] = e^{t\mu}\mathbb{E}\Big[e^{t\sigma Z}\Big] = e^{t\mu} M_Z(t\sigma) = e^{t\mu} e^{t^2\sigma^2/2}.$$

A.4.2 The Cameron-Martin Formula

The next step is to generalize the previous idea from a single Gaussian random variable to an entire Brownian path. The Cameron–Martin formula tells us how expectations of functionals of a drifted Brownian motion can be computed under the original Wiener measure.

Theorem A.4.4. *Let $\{B_t\}$ be a Brownian motion on $[0,T]$ and let $\mu \in L^2([0,T])$. Define*

$$X_t = \int_0^t \mu(s)\,\mathrm{d}s + B_t.$$

Then for any functional f for which both expectations exist,

$$\mathbb{E}\Big[f\big(X_{[0,T]}\big)\Big] = \mathbb{E}\left[f\big(B_{[0,T]}\big)\exp\left(\int_0^T \mu(t)\,\mathrm{d}B_t - \frac{1}{2}\int_0^T \mu(t)^2\,\mathrm{d}t\right)\right]. \tag{A.31}$$

Remark A.4.5. This is the infinite-dimensional analogue of theorem A.4.1. The exponential term plays the role of a *Radon-Nikodym derivative*:

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(\int_0^T \mu(t)\,\mathrm{d}B_t - \frac{1}{2}\int_0^T \mu(t)^2\,\mathrm{d}t\right)$$

so under the measure $\mathbb{Q}$, the process $B_t + \int_0^t \mu(s)\,\mathrm{d}s$ is a Brownian motion.

Heuristic Derivation. Fix $0 < t_1 < \cdots < t_n = T$ and consider $f\colon \mathbb{R}^n \mapsto \mathbb{R}$. Write

$$\mathbb{E}\big[f(X_{t_1},\ldots,X_{t_n})\big] := \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} f(x_1,\ldots,x_n)p_X(x_1,\ldots,x_n)\,\mathrm{d}x_n\cdots\mathrm{d}x_1 \tag{A.32}$$

where

$$\begin{aligned} p_X(x_1,\ldots,x_n) &= p_X(x_n \mid x_{n-1})\cdots p_X(x_1) = \prod_{k=0}^{n-1} p_X(x_{k+1}\mid x_k) \\ &= \prod_{k=0}^{n-1}\frac{1}{\sqrt{2\pi(t_{k+1}-t_k)}}\exp\left(-\frac{(x_{k+1}-x_k-\int_{t_k}^{t_{k+1}}\mu(s)\,\mathrm{d}s)^2}{2(t_{k+1}-t_k)}\right) \end{aligned}$$

is the joint density of $(X_{t_1},\ldots,X_{t_n})$. Define now

$$p_B(x_1,\ldots,x_n) := \prod_{k=0}^{n-1}\frac{1}{\sqrt{2\pi(t_{k+1}-t_k)}}\exp\left(-\frac{(x_{k+1}-x_k)^2}{2(t_{k+1}-t_k)}\right)$$

as the joint density of $(B_{t_1},\ldots,B_{t_n})$. We multiply the integrand of (A.32) by p_B/p_B and we obtain

$$\int_{\mathbb{R}}\cdots\int_{\mathbb{R}} f(x_1,\ldots,x_n)\frac{p_X(x_1,\ldots,x_n)}{p_B(x_1,\ldots,x_n)}p_B(x_1,\ldots,x_n)\,\mathrm{d}x_n\cdots\mathrm{d}x_1. \tag{A.33}$$

Taking $p_B(x_1, \ldots, x_n)$ as our new density, this is

$$\mathbb{E}\left[f(B_{t_1}, \ldots, B_{t_n}) \frac{p_X(B_{t_1}, \ldots, B_{t_n})}{p_B(B_{t_1}, \ldots, B_{t_n})} \right]. \tag{A.34}$$

We simplify the likelihood ratio:

$$\mathbb{E}\left[f(B_{t_1}, \ldots, B_{t_n}) \exp\left(\sum_{k=0}^{n-1} \mu(t_k)(B_{t_{k+1}-B_{t_k}}) - \frac{1}{2} \sum_{k=0}^{n-1} \mu(t_k)^2 (t_{k+1} - t_k) \right) \right]. \tag{A.35}$$

Now, this identity holds for every fixed n. Taking $n \to \infty$, we hope (A.35) converges in some sense to (A.32), where f is now a functional of the entire path $X_{[0,T]}$ as opposed to just the points $X_{t_1}, \ldots, X_{t_n}$. □

A.4.2.1 Application: Maximum of Drifted Brownian motion

To illustrate the power of the Cameron–Martin theorem, we study how classical formulas for the running maximum of Brownian motion are modified after introducing a constant drift. In this section, B_t denotes a standard Brownian motion and

$$M_t := \max_{0 \le s \le t} B_s$$

its running maximum. For a fixed $\alpha \in \mathbb{R}$, we define the drifted Brownian motion by

$$X_t := \alpha t + B_t, \qquad Y_t := \max_{0 \le s \le t} X_s.$$

The maxima M_t and Y_t are well-defined almost surely, since Brownian paths are almost surely continuous and $[0, t]$ is compact, so the extreme value theorem applies.

Distribution functions We will take for granted the following classical result concerning the maximum of standard Brownian motion:

Proposition A.4.6. *Let $M_t := \max_{0 \le s \le t} B_s$. Then M_t has CDF*

$$\mathbb{P}(M_t \le m) = 2N\left(\frac{m}{\sqrt{t}} \right), \qquad m \ge 0 \tag{A.36}$$

and density

$$f_{M_t}(m) = \frac{2}{\sqrt{2\pi t}} \exp\left(-\frac{m^2}{2t} \right), \qquad m \ge 0. \tag{A.37}$$

Furthermore, the joint density of (B_t, M_t) *is*

$$f_{B_t,M_t}(b,m) = \frac{2(2m-b)}{t\sqrt{2\pi t}} \exp\left(-\frac{(2m-b)^2}{2t}\right), \qquad m \geq 0,\, b \leq m. \tag{A.38}$$

In the presence of a constant drift α, the corresponding results for X_t and Y_t take the following form:

Proposition A.4.7. *The random variable* Y_t *has CDF*

$$\mathbb{P}(Y_t \leq y) = N\left(\frac{y-\alpha t}{\sqrt{t}}\right) - e^{2\alpha y} N\left(-\frac{y+\alpha t}{\sqrt{t}}\right), \qquad y \geq 0 \tag{A.39}$$

and density

$$f_{Y_t}(y) = \frac{2}{\sqrt{2\pi t}} \exp\left(-\frac{(y-\alpha t)^2}{2t}\right) - 2\alpha e^{2\alpha y} N\left(-\frac{y+\alpha t}{\sqrt{t}}\right), \qquad y \geq 0. \tag{A.40}$$

Furthermore, the joint density of (X_t, Y_t) *is*

$$f_{X_t,Y_t}(x,y) = \frac{2(2y-x)}{t\sqrt{2\pi t}} \exp\left(2\alpha y - \frac{(2y-x+\alpha t)^2}{2t}\right), \qquad y \geq 0,\, x \leq y. \tag{A.41}$$

Proof. Fix $y \geq 0$ and $x \leq y$. We write

$$\mathbb{P}(Y_t \geq y,\, X_t \leq x) = \mathbb{E}\big[\mathbb{I}\{Y_t \geq y,\, X_t \leq x\}\big]. \tag{A.42}$$

Comparing the right-hand side with (A.31), we choose

$$f(X_{[0,t]}) = \mathbb{I}\left\{\max_{0\leq s\leq t} X_s \geq y,\; X_t \leq x\right\}, \tag{A.43}$$

$$\mu(s) = \alpha. \tag{A.44}$$

By (A.31), this gives

$$\mathbb{E}\big[\mathbb{I}\{Y_t \geq y,\, X_t \leq x\}\big] = \mathbb{E}\Big[\mathbb{I}\{M_t \geq y,\, B_t \leq x\} \exp\big(\alpha B_t - \tfrac{1}{2}\alpha^2 t\big)\Big]. \tag{A.45}$$

Recall from the reflection principle that on the event $\{M_t \geq y\}$, the random variables B_t and $2y - B_t$ are equal in distribution (see fig. A.1). Therefore, we may replace B_t by $2y - B_t$

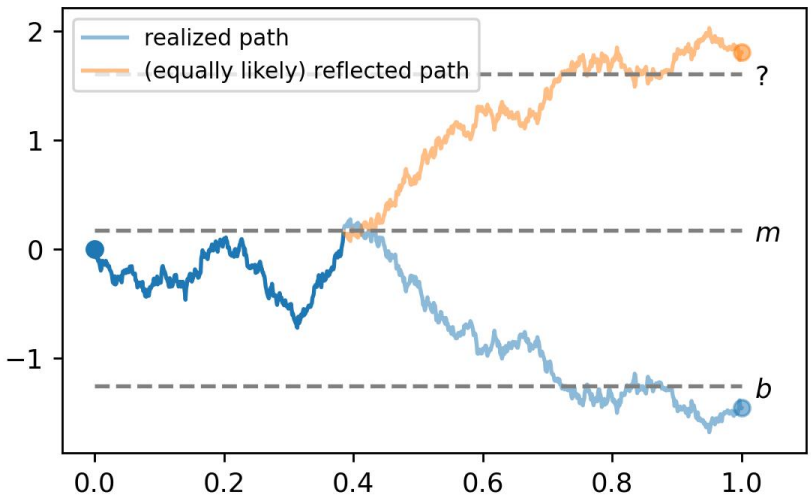


Figure A.1: The reflection principle.

inside the expectation in (A.45), which yields

$$\begin{aligned}\mathbb{E}\Big[\mathbb{I}\{M_t \geq y,\, 2y - B_t \leq x\} \exp\big(\alpha(2y - B_t) - \tfrac{1}{2}\alpha^2 t\big)\Big] \\ = e^{2\alpha y}\mathbb{E}\Big[\mathbb{I}\{B_t \geq 2y - x\} \exp\big(-\alpha B_t - \tfrac{1}{2}\alpha^2 t\big)\Big]\,.\end{aligned} \tag{A.46}$$

We now apply (A.4.4) again (this time from right to left) with

$$f\big(B_{[0,t]}\big) = \mathbb{I}\{B_t \geq 2y - x\}\,, \tag{A.47}$$

$$\mu(s) = -\alpha, \tag{A.48}$$

to obtain

$$e^{2\alpha y}\mathbb{E}\big[\mathbb{I}\{B_t - \alpha t \geq 2y - x\}\big] = e^{2\alpha y}\, N\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \tag{A.49}$$

Combining (A.42)–(A.49), we conclude that

$$\mathbb{P}(Y_t \geq y,\, X_t \leq x) = e^{2\alpha y}\, N\left(\frac{x - 2y - \alpha t}{\sqrt{t}}\right). \tag{A.50}$$

Differentiating (A.50) with respect to x and y gives the joint density (A.41). To obtain the CDF (A.39), observe that

$$\underbrace{\mathbb{P}(X_t \leq y)}_{N\left(\frac{y-\alpha t}{\sqrt{t}}\right)} = \underbrace{\mathbb{P}(Y_t \leq y,\, X_t \leq y)}_{\mathbb{P}(Y_t \leq y)} + \underbrace{\mathbb{P}(Y_t \geq y,\, X_t \leq y)}_{e^{2\alpha y} N\left(-\frac{y+\alpha t}{\sqrt{t}}\right)}.$$

This immediately yields (A.39). Finally, differentiating (A.39) with respect to y gives (A.40). □

Hitting time density Fix $L > 0$. Define the first hitting times of the level L by

$$\tau_B := \inf\{t \geq 0 : B_t = L\}, \tag{A.51}$$

$$\tau_X := \inf\{t \geq 0 : X_t = L\}, \tag{A.52}$$

corresponding to standard Brownian motion and drifted Brownian motion, respectively.

We take for granted the following classical result in the absence of drift.

Proposition A.4.8. *The random variable τ_B has density*

$$f_{\tau_B}(t) = \frac{L}{t^{3/2}\sqrt{2\pi}} \exp\left(-\frac{L^2}{2t}\right), \qquad t > 0. \tag{A.53}$$

Using this, we derive the corresponding result for the drifted process.

Proposition A.4.9. *Assume for simplicity that $\alpha > 0$. Then, the random variable τ_X has density*

$$f_{\tau_X}(t) = \frac{L}{t^{3/2}\sqrt{2\pi}} \exp\left(-\frac{(L-\alpha t)^2}{2t}\right), \qquad t > 0. \tag{A.54}$$

Proof. Fix $t \in [0, T]$ and define

$$f\big(X_{[0,T]}\big) = \mathbb{I}\{\tau_X \wedge T \leq t\} \tag{A.55}$$

$$\mu(t) = \alpha. \tag{A.56}$$

With these choices, the Cameron–Martin formula (A.31) yields

$$\mathbb{P}(\tau_X \wedge T \leq t) = \mathbb{E}\left[\mathbb{I}\{\tau_B \wedge T \leq t\} \exp\left(\alpha B_T - \frac{\alpha^2}{2}T\right)\right]. \tag{A.57}$$

By iterated expectations, the right-hand side can be written as

$$\mathbb{E}\left[\mathbb{E}\left[\mathbb{I}\{\tau_B \wedge T \leq t\} \exp\left(\alpha B_T - \frac{\alpha^2}{2}T\right) \middle| \mathscr{F}_{\tau_B \wedge T}\right]\right].$$

Since $\mathbb{I}\{\tau_B \wedge T \leq t\}$ is $\mathscr{F}_{\tau_B \wedge T}$-measurable, we may factor it out of the inner conditional

expectation. This gives

$$\mathbb{E}\left[\mathbb{I}\{\tau_B \wedge T \leq t\}\,\mathbb{E}\left[\exp\left(\alpha B_T - \frac{\alpha^2}{2}T\right)\middle|\mathscr{F}_{\tau_B\wedge T}\right]\right].$$

Now note that the process $t \mapsto \exp\Big(\alpha B_t - \frac{\alpha^2}{2}t\Big)$ is a martingale. Therefore,

$$\mathbb{E}\left[\exp\left(\alpha B_T - \frac{\alpha^2}{2}T\right)\middle|\mathscr{F}_{\tau_B\wedge T}\right] = \exp\left(\alpha B_{\tau_B\wedge T} - \frac{\alpha^2}{2}(\tau_B \wedge T)\right).$$

Substituting this back, (A.57) becomes

$$\mathbb{P}(\tau_X \wedge T \leq t) = \mathbb{E}\left[\mathbb{I}\{\tau_B \wedge T \leq t\}\exp\left(\alpha B_{\tau_B\wedge T} - \frac{\alpha^2}{2}\left(\tau_B \wedge T\right)\right)\right].$$

We now let $T \to \infty$ and apply Lebesgue's dominated convergence theorem. Note that

$$0 < \mathbb{I}\{\tau_B \wedge T \leq t\}\exp\left(\alpha B_{\tau_B\wedge T} - \frac{\alpha^2}{2}\left(\tau_B \wedge T\right)\right) < e^{\alpha L}.$$

since $B_{\tau_B} = L$ on $\{\tau_B < \infty\}$. Taking for granted that $\tau_B < \infty$ and $\tau_X < \infty$ almost surely (since $\alpha > 0$), we obtain

$$\mathbb{P}(\tau_X \leq t) = \mathbb{E}\left[\mathbb{I}\{\tau_B \leq t\}\exp\left(\alpha L - \frac{\alpha^2}{2}\tau_B\right)\right].$$

Here τ_B is the driftless hitting time. Using (A.53), its density, we can write

$$\mathbb{P}(\tau_X \leq t) = \int_0^t \exp\left(\alpha L - \frac{\alpha^2}{2}s\right)\frac{L}{s^{3/2}\sqrt{2\pi}}\exp\left(-\frac{L^2}{2s}\right)\mathrm{d}s.$$

Finally, completing the square in the exponent and differentiating with respect to t yields the density (A.54). □

Remark A.4.10. A much quicker proof follows from (A.39). Write

$$\mathbb{P}(\tau_X \leq t) = \mathbb{P}(Y_t \geq L) = 1 - N\left(\frac{y-\alpha t}{\sqrt{t}}\right) + e^{2\alpha L}N\left(-\frac{y+\alpha t}{\sqrt{t}}\right)$$

and differentiate in t and simplify to get (A.54).

Hitting time MGF We now turn to the moment generating functions of the hitting times τ_B and τ_X defined in (A.51) and (A.52). First, recall the classical driftless result.

Proposition A.4.11. *The random variable τ_B has MGF*

$$M_{\tau_B}(\lambda) = \exp\Big(-L\sqrt{2\lambda}\Big), \qquad \lambda \geq 0. \tag{A.58}$$

For a proof, see [Shr04, Theorem 3.7.1]. Our main interest here is in showing how the Cameron–Martin theorem extends proposition A.4.11 to the drifted case.

Proposition A.4.12. *Suppose $\alpha > 0$. Then the hitting time τ_X has MGF*

$$M_{\tau_X}(\lambda) := \mathbb{E}\Big[e^{-\lambda\tau_X}\Big] = \exp\bigg(-L\Big(-\alpha + \sqrt{\alpha^2 + 2\lambda}\Big)\bigg). \tag{A.59}$$

Proof. We apply (A.31) with

$$\begin{aligned} f\big(X_{[0,T]}\big) &= \exp\big(-\lambda(\tau_X \wedge T)\big), \\ \mu(t) &= \alpha, \end{aligned}$$

which gives

$$\mathbb{E}\Big[\exp\big(-\lambda(\tau_X \wedge T)\big)\Big] = \mathbb{E}\left[\exp\big(-\lambda(\tau_B \wedge T)\big)\exp\left(\alpha B_T - \frac{\alpha^2}{2}T\right)\right]. \tag{A.60}$$

Using iterated expectations together with the fact that $t \mapsto \exp\Big(\alpha B_t - \frac{\alpha^2}{2}t\Big)$ is a martingale, the right-hand side becomes

$$\mathbb{E}\left[\exp\big(-\lambda(\tau_B \wedge T)\big)\exp\left(\alpha B_{\tau_B \wedge T} - \frac{\alpha^2}{2}(\tau_B \wedge T)\right)\right].$$

Hence

$$\mathbb{E}\Big[\exp\big(-\lambda(\tau_X \wedge T)\big)\Big] = \mathbb{E}\left[\exp\left(\alpha B_{\tau_B \wedge T} - \left(\lambda + \frac{\alpha^2}{2}\right)(\tau_B \wedge T)\right)\right]. \tag{A.61}$$

We now send $T \to \infty$ and apply Lebesgue's dominated convergence theorem. Note that

$$0 < \exp\big(-\lambda(\tau_X \wedge T)\big) < 1, \qquad 0 < \exp\Big(\alpha B_{\tau_B \wedge T} - (\lambda + \tfrac{\alpha^2}{2})(\tau_B \wedge T)\Big) < e^{\alpha L},$$

since $B_{\tau_B} = L$ on the event $\{\tau_B < \infty\}$. Assuming (as we do for $\alpha > 0$) that $\tau_B < \infty$ and $\tau_X < \infty$ almost surely, letting $T \to \infty$ in (A.61) yields

$$\mathbb{E}\big[\exp(-\lambda\tau_X)\big] = \mathbb{E}\left[\exp\left(\alpha L - \left(\lambda + \frac{\alpha^2}{2}\right)\,\tau_B\right)\right].$$

Since τ_B is the driftless hitting time, we may invoke (A.58) to compute the remaining expectation. This gives exactly the expression in (A.59), completing the proof. □

Remark A.4.13. The identity (A.59) can also be obtained directly by integrating $e^{-\lambda t}$ against the density (A.54).

A.4.2.2 Application: Black–Scholes Formula

Proposition A.4.14. *Let*

$$S_T := S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right) T + \sigma B_T\right). \tag{A.62}$$

Then the arbitrage-free price of a European call option with strike K and maturity T is

$$e^{-rT}\mathbb{E}\big[(S_T - K)^+\big] = S_0\, N(d_1) - K e^{-rT} N(d_2), \tag{A.63}$$

where

$$d_1 = \frac{\log(S_0/K) + \left(r + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}, \tag{A.64}$$

$$d_2 = \frac{\log(S_0/K) + \left(r - \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}. \tag{A.65}$$

Proof. We begin by writing

$$\begin{aligned} e^{-rT}\mathbb{E}\big[(S_T - K)^+\big] &= e^{-rT}\mathbb{E}\big[(S_T - K)\,\mathbb{I}\{S_T \geq K\}\big] \\ &= e^{-rT}\mathbb{E}\big[S_T\,\mathbb{I}\{S_T \geq K\}\big] \;-\; K e^{-rT}\mathbb{P}(S_T \geq K)\,. \end{aligned}$$

The probability in the second term is

$$\mathbb{P}(S_T \geq K) = \mathbb{P}\bigg(S_0 \exp\Big((r - \tfrac{\sigma^2}{2})T + \sigma B_T \Big) \geq K \bigg) = N(d_2).$$

To evaluate the first term, observe that

$$\mathbb{E}\big[S_T \, \mathbb{I}\{S_T \geq K\} \big] = S_0 e^{rT} \mathbb{E}\left[\exp\left(\sigma B_T - \frac{\sigma^2}{2} T \right) \mathbb{I}\{S_T \geq K\} \right]. \tag{A.66}$$

We now apply the Cameron–Martin formula (A.31) with

$$\begin{aligned} f(B_{[0,T]}) &= \mathbb{I}\bigg\{ S_0 \exp\Big((r - \tfrac{\sigma^2}{2})T + \sigma B_T \Big) \geq K \bigg\}, \\ \mu(t) &= \sigma. \end{aligned}$$

Under this change of measure, the right-hand side of (A.66) becomes

$$\mathbb{P}\bigg(S_0 \exp\Big((r - \tfrac{\sigma^2}{2})T + \sigma(B_T + \sigma T) \Big) \geq K \bigg) = N(d_1).$$

Combining this with the earlier expression for $\mathbb{P}(S_T \geq K)$ yields (A.63), proving the Black–Scholes formula. □

A.4.2.3 Application: Put-Call Symmetry

Proposition A.4.15. *Let S_T be defined as in* (A.62)*, and define the call and put prices by*

$$C(S_0, K, T) := e^{-rT} \mathbb{E}\Big[(S_T - K)^+ \Big] \tag{A.67}$$

$$P(S_0, K, T) := e^{-rT} \mathbb{E}\Big[(K - S_T)^+ \Big]. \tag{A.68}$$

Then,

$$C(S_0, K, T) = P(Ke^{-rT}, S_0 e^{rT}, T). \tag{A.69}$$

Proof. We begin by writing

$$\begin{aligned} C(S_0, K, T) &= e^{-rT}\mathbb{E}\big[(S_T - K)^+\big] \\ &= e^{-rT}\mathbb{E}\left[\left(S_0 \exp\Big((r - \tfrac{\sigma^2}{2})T + \sigma B_T\Big) - K\right)^+\right] \\ &= e^{-rT}\mathbb{E}\left[\exp\Big(\sigma B_T - \tfrac{\sigma^2}{2}T\Big)\left(S_0 e^{rT} - K\exp\Big(\tfrac{\sigma^2}{2}T - \sigma B_T\Big)\right)^+\right]. \end{aligned} \tag{A.70}$$

We now apply the Cameron–Martin formula theorem A.4.4 with

$$\begin{aligned} f(B_{[0,T]}) &= \left(S_0 e^{rT} - K\exp\Big(\tfrac{\sigma^2}{2}T - \sigma B_T\Big)\right)^+, \\ \mu(t) &= \sigma. \end{aligned}$$

Under this change of measure, the right-hand side of (A.70) becomes

$$\begin{aligned} & e^{-rT}\mathbb{E}\left[\left(S_0 e^{rT} - K\exp\Big(\tfrac{\sigma^2}{2}T - \sigma(B_T + \sigma T)\Big)\right)^+\right] \\ &\quad = e^{-rT}\mathbb{E}\left[\left(S_0 e^{rT} - Ke^{-rT}\exp\Big((r - \tfrac{\sigma^2}{2})T - \sigma B_T\Big)\right)^+\right] \\ &\quad = P\Big(Ke^{-rT},\, S_0 e^{rT},\, T\Big), \end{aligned}$$

which completes the proof. □

A.4.3 Girsanov's Formula

We now introduce Girsanov's theorem. Unlike the Cameron–Martin theorem, which treats deterministic drifts, Girsanov's theorem allows the drift to be a random adapted process.

Theorem A.4.16. *Let $\{B_t\}_{0\le t\le T}$ be a Brownian motion and let $\{\theta_t\}_{0\le t\le T}$ be an adapted process satisfying the Novikov condition*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2 \,\mathrm{d}t\right)\right] < \infty.$$

Define the drifted process

$$X_t := \int_0^t \theta_s \,\mathrm{d}s + B_t.$$

Then for any functional f for which the expectations exist,

$$\mathbb{E}\Big[f\big(X_{[0,T]}\big)\Big] = \mathbb{E}\left[f\big(B_{[0,T]}\big)\exp\left(\int_0^T \theta_t\,\mathrm{d}B_t - \frac{1}{2}\int_0^T \theta_t^2\,\mathrm{d}t\right)\right]. \tag{A.71}$$

Remark A.4.17. The Cameron–Martin theorem is a special case of Girsanov's theorem. Indeed, if $\theta \in L^2([0,T])$ is deterministic, then $\int_0^T \theta_t^2\ \mathrm{d}t < \infty$, and the Novikov condition holds automatically:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta_t^2\ \mathrm{d}t\right)\right] = \exp\left(\frac{1}{2}\int_0^T \theta_t^2\ \mathrm{d}t\right) < \infty.$$

A.4.3.1 Application: Merton's portfolio optimization problem

Problem A.4.18. Merton's portfolio optimization problem is

$$\sup_{\alpha_{[0,T]}} \mathbb{E}\left[\frac{X_T^p}{p}\right], \qquad p \in (0,1),$$

where the controlled state variable satisfies

$$\mathrm{d}X_t = \big(\alpha_t\mu + (1-\alpha_t)r\big)X_t\ \mathrm{d}t + \alpha_t\sigma X_t\ \mathrm{d}B_t.$$

Proposition A.4.19. *The optimal control is*

$$\alpha_t^* = \frac{\mu - r}{\sigma^2(1-p)}.$$

Proof. Observe first that X_t has the explicit representation

$$X_t = X_0\exp\left(\int_0^T\left(\alpha_s\mu + (1-\alpha_s)r - \frac{\alpha_s^2\sigma^2}{2}\right)\mathrm{d}s + \int_0^T \alpha_s\sigma\,\mathrm{d}B_s\right).$$

Hence the objective equals

$$\mathbb{E}\left[\frac{X_T^p}{p}\right] = \frac{X_0^p}{p}\mathbb{E}\left[\exp\left(\int_0^T p\left(\alpha_s\mu + (1-\alpha_s)\,r - \frac{\alpha_s^2\sigma^2}{2}\right)\mathrm{d}s + \int_0^T p\alpha_s\sigma\,\mathrm{d}B_s\right)\right].$$

We now decompose the exponent into two factors,

$$\exp\left(\int_0^T p\alpha_s\sigma\,\mathrm{d}B_s - \frac{1}{2}\int_0^T p^2\alpha_s^2\sigma^2\,\mathrm{d}s\right)$$

and

$$\exp\left(\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right).$$

The first factor is a stochastic exponential and will serve as our Radon–Nikodym derivative. Applying Girsanov's theorem with

$$f(B_{[0,T]}) = \exp\left(\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right) \tag{A.72}$$

$$\theta(t,\omega) = p\sigma\alpha(t,\omega) \tag{A.73}$$

yields

$$\mathbb{E}\left[\frac{X_T^p}{p}\right] = \frac{X_0^p}{p}\mathbb{E}\left[\exp\left(\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right)\right].$$

To maximize this expression, note that

$$\begin{aligned}
&\sup_{\alpha_{[0,T]}}\mathbb{E}\left[\exp\left(\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right)\right] \\
&\quad\leq \mathbb{E}\left[\sup_{\alpha_{[0,T]}}\exp\left(\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right)\right] \\
&\quad= \mathbb{E}\left[\exp\left(\sup_{\alpha_{[0,T]}}\int_0^T \left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right)\right] \\
&\quad\leq \mathbb{E}\left[\exp\left(\int_0^T \sup_{\alpha_s}\left[pr + p\alpha_s\left(\mu - r\right) - \frac{\alpha_s^2}{2}p\sigma^2\left(1-p\right)\right]\mathrm{d}s\right)\right] \\
&\quad= \exp\left(\left(pr + \frac{p}{2\left(1-p\right)}\frac{\left(\mu-r\right)^2}{\sigma^2}\right)T\right).
\end{aligned}$$

The supremum is attained pointwise by maximizing the quadratic in α_s, which gives

$$\alpha_s^* = \frac{\mu - r}{\sigma^2 (1-p)}.$$

Thus $\alpha_s^* = \frac{\mu - r}{\sigma^2(1-p)}$ is the optimal control among all processes satisfying the Novikov condition, completing the proof. □

Remark A.4.20. Proposition A.4.19 requires Girsanov's theorem rather than the Cameron–Martin theorem because the drift term α_t is allowed to be random. Only after solving the problem do we discover that the optimal control α_t is in fact a deterministic constant.

APPENDIX B

Stochastic Optimal Control Theory

B.1 The HJB Equation

This derivation is from section 25.5 of Bjork [Bjo20]. We consider the maximization problem

$$\sup_u \mathbb{E}\left[\int_t^T F(s, X_s, u_s)\,\mathrm{d}s + \Phi(X_T)\middle| X_t = x\right], \tag{B.1}$$

where X_t evolves according to

$$\mathrm{d}X_t = \mu(t, X_t, u_t)\,\mathrm{d}t + \sigma(t, X_t, u_t)\,\mathrm{d}W_t. \tag{B.2}$$

The function u_t is called the control parameter. Suppose an optimal one exists. Denote it by u_t^*, and define

$$V(t,x) = \mathbb{E}\left[\int_t^T F(s, X_s, u_s^*)\,\mathrm{d}s + \Phi(X_T)\middle| X_t = x\right]. \tag{B.3}$$

Clearly, $V(t,x)$ specifies the value in (B.1). It has the interpretation of the most money one should expect to make trading optimally over $[t,T]$. The dynamic programming principle says that if we fail to adhere to u^* for the first part of $[t,T]$, say $[t,t+h)$, then the best thing for us to do at time $t+h$ is to continue with u^* over the remaining $[t+h,T]$. The wealth we expect to generate from doing so is

$$\mathbb{E}\left[\int_t^{t+h} F(s, X_s, u_s)\,\mathrm{d}s + \int_{t+h}^T F(s, X_s, u_s)\,\mathrm{d}s + \Phi(X_T)\middle| X_t = x\right], \tag{B.4}$$

where u_s denotes any strategy at all. Clearly, (B.4) cannot exceed $V(t,x)$. Conditioning on values of X_t and appealing to the tower property, (B.4) can be written as

$$\begin{aligned}\mathbb{E}\left[\int_t^{t+h} F(s,X_s,u_s)\,\mathrm{d}s + \mathbb{E}\left[\int_{t+h}^{T} F(s,X_s,u_s)\,\mathrm{d}s + \Phi(X_T)\middle| X_{x+h}\right]\middle| X_t = x\right] \\ = \mathbb{E}\left[\int_t^{t+h} F(s,X_s,u_s)\,\mathrm{d}s + V(t+h,X_{t+h})\middle| X_t = x\right],\end{aligned} \tag{B.5}$$

so from previous considerations, we have

$$V(t,x) \geq \mathbb{E}\left[\int_t^{t+h} F(s,X_s,u_s)\,\mathrm{d}s + V(t+h,X_{t+h})\middle| X_t = x\right]. \tag{B.6}$$

We would like to cancel $V(t,x)$ on both sides of the inequality, but $V(t+h,X_{t+h})$ is not quite $V(t,x)$. Itô's lemma with (B.2) tells us the difference. Momentarily suppressing arguments, we have

$$V(t+h,X_{t+h}) = V(t,X_t) + \int_t^{t+h} \frac{\partial V}{\partial t} + \mu\,\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2\,\frac{\partial^2 V}{\partial x^2}\,\mathrm{d}s + \int_t^{t+h} \sigma\,\frac{\partial V}{\partial x}\,\mathrm{d}W_s. \tag{B.7}$$

Placing (B.7) into the expectation (B.6) turns $V(t,X_t)$ into $V(t,x)$ and removes the stochastic integral. Thus (B.6) simplifies to

$$0 \geq \mathbb{E}\left[\int_t^{t+h} F + \frac{\partial V}{\partial t} + \mu\,\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2\,\frac{\partial^2 V}{\partial x^2}\,\mathrm{d}s\middle| X_t = x\right]. \tag{B.8}$$

Dividing through by h and sending $h \downarrow 0$, (B.8) becomes

$$0 \geq F(t,x,u) + \frac{\partial V}{\partial t}(t,x) + \mu(t,x)\,\frac{\partial V}{\partial x}(t,x) + \frac{1}{2}\sigma(t,x)^2\,\frac{\partial^2 V}{\partial x^2}(t,x). \tag{B.9}$$

(B.9) becomes an equality when $u = u^*$, which gives the HBJ equation

$$0 = \sup_u \left\{ F(t,x,u) + \frac{\partial V}{\partial t}(t,x) + \mu(t,x)\,\frac{\partial V}{\partial x}(t,x) + \frac{1}{2}\sigma(t,x)^2\,\frac{\partial^2 V}{\partial x^2}(t,x) \right\}. \tag{B.10}$$

APPENDIX C

Filtering Theory

C.1 Zakai's Equation

We derive Zakai's equation starting from Girsanov's theorem. We will obtain Girsanov's kernel using a heuristic limiting argument inspired by theorem 6.13 of Law et al. [LSZ10] which closely mimics the argument following remark A.4.5. As for Zakai's equation, we follow his original 1969 paper [Zak69].

C.1.1 Revisiting Girsanov's Theorem

C.1.1.1 The Static Case

Consider a measurable space $(\Omega, \mathscr{F})$ and let (X, Y) be a pair of random variables whose distribution is under dispute. There are two observers, the first with beliefs $\mathbb{P}$ and the second with beliefs $\mathbb{Q}$. They agree on the distribution of X, but the first observer believes

$$Y = \alpha(X) + \beta Z^{\mathbb{P}},$$

while the second believes

$$Y = \beta Z^{\mathbb{Q}}.$$

Here, $\alpha(\cdot)$ denotes a deterministic Borel measurable function, β a real number, and $Z^{\mathbb{P}}, Z^{\mathbb{Q}}$ standard Gaussian random variables under their respective measures. A sample is drawn and takes the value (x, y). To the first observer, the likelihood that this event occurred is

$$\mathbb{P}\big((X,Y) \in (\mathrm{d}x, \mathrm{d}y)\big) = \mathbb{P}(X \in \mathrm{d}x)\, \mathbb{P}(Y \in \mathrm{d}y | X = x) \tag{C.1}$$

$$= \mathbb{P}(X \in \mathrm{d}x) \times \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{1}{2} \left(\frac{y - \alpha(x)}{\beta} \right)^2 \right) \mathrm{d}y. \tag{C.2}$$

The second observer, believing X and Y to be independent, thinks the likelihood is

$$\mathbb{Q}\big((X,Y)\in(\mathrm{d}x,\mathrm{d}y)\big)=\mathbb{Q}(X\in\mathrm{d}x)\,\mathbb{Q}(Y\in\mathrm{d}y) \tag{C.3}$$

$$=\mathbb{Q}(X\in\mathrm{d}x)\times\frac{1}{\sqrt{2\pi\beta^2}}\exp\left(-\frac{1}{2}\left(\frac{y}{\beta}\right)^2\right)\mathrm{d}y. \tag{C.4}$$

The quantities $\mathbb{P}(X\in\mathrm{d}x)$ and $\mathbb{Q}(X\in\mathrm{d}x)$ are the same because there is no disagreement over the distribution of X. If this quantity is nonzero, then the likelihood ratio of (C.2) and (C.4) is

$$\frac{\mathbb{P}\big((X,Y)\in(\mathrm{d}x,\mathrm{d}y)\big)}{\mathbb{Q}\big((X,Y)\in(\mathrm{d}x,\mathrm{d}y)\big)}=\exp\left(\frac{\alpha(x)}{\beta^2}y-\frac{1}{2}\frac{\alpha(x)^2}{\beta^2}\right),$$

or, what is equivalent,

$$\frac{\mathbb{P}(\mathrm{d}\omega)}{\mathbb{Q}(\mathrm{d}\omega)}=\exp\left(\frac{\alpha\big(X(\omega)\big)}{\beta^2}Y(\omega)-\frac{1}{2}\frac{\alpha\big(X(\omega)\big)^2}{\beta^2}\right).$$

We have derived a Radon-Nikodym derivative which permits the change of measure from $\mathbb{Q}$ to $\mathbb{P}$. The other direction is attained by direct reciprocal. Both RN derivatives exist due to equivalence between $\mathbb{P}$ and $\mathbb{Q}$.

Intuitively, $\mathbb{P}(\mathrm{d}\omega)/\mathbb{Q}(\mathrm{d}\omega)$ is a conversion ratio that allows one observer to interpret an event from the other's point of view. Such a ratio only exists when the beliefs of the two observers are "compatible" in the sense that events which appear outlandish to one will appear outlandish to the other.

C.1.1.2 Discrete Markov Chains

We now generalize the previous example to a discrete Markov process. Endow $(\Omega,\mathscr{F})$ with the filtration $\{\mathscr{F}_n\colon 0\le n\le N\}$ and consider the process

$$\big\{(X_n,Y_n,\mathscr{F}_n):0\le n\le N\big\}.$$

There are two observers, just as before, and they agree on the distribution of X_n for each n. It is known that $\big\{(X_n,\mathscr{F}_n)\big\}$ is Markovian.

The observers also agree on the distribution of Y_0. The disagreement is over subsequent Y_n. Observer $\mathbb{P}$ believes that $\{Y_n\}$ evolves according to

$$Y_{n+1}=Y_n+\alpha_n(X_n,Y_n)+\beta_n(Y_n)Z_n^{\mathbb{P}},\qquad 0\le n\le N-1$$

(the notation is similar to before) while observer $\mathbb{Q}$ believes

$$Y_{n+1} = Y_n + \beta_n(Y_n) Z_n^{\mathbb{Q}}.$$

The event $\big\{(X_0, Y_0) = (x_1, y_1), \ldots, (X_N, Y_N) = (x_N, y_N)\big\}$ is observed. Observer $\mathbb{P}$, noting the Markov property, thinks the likelihood of this event is

$$\begin{aligned}
&\mathbb{P}\big((X_0, Y_0) \in (\mathrm{d}x_0, \mathrm{d}y_0), \ldots, (X_N, Y_N) \in (\mathrm{d}x_N, \mathrm{d}y_N)\big) \\
&= \mathbb{P}\big((X_0, Y_0) \in (\mathrm{d}x_0, \mathrm{d}y_0)\big) \prod_{n=0}^{N-1} \mathbb{P}\big((X_{n+1}, Y_{n+1}) \in (\mathrm{d}x_{n+1}, \mathrm{d}y_{n+1}) \big| (X_n, Y_n) = (x_n, y_n)\big)\,.
\end{aligned} \tag{C.5}$$

Note that by independence, each of the rightmost terms in (C.5) simplify to

$$\mathbb{P}(X_{n+1} \in \mathrm{d}x_{n+1} | X_n = x_n)\, \mathbb{P}\big(Y_{n+1} \in \mathrm{d}y_{n+1} \big| (X_n, Y_n) = (x_n, y_n)\big)\,. \tag{C.6}$$

Observer $\mathbb{Q}$ performs the same analysis, and the only difference is in his analog of (C.6), which instead takes the form

$$\mathbb{Q}(X_{n+1} \in \mathrm{d}x_{n+1} | X_n = x_n)\, \mathbb{Q}(Y_{n+1} \in \mathrm{d}y_{n+1} | Y_n = y_n)\,.$$

Because both observers agree on $X_0, \ldots, X_N, Y_0$, the likelihood ratio simplifies to

$$\begin{aligned}
&\prod_{n=0}^{N-1} \frac{\mathbb{P}\big(Y_{n+1} \in \mathrm{d}y_{n+1} \big| (X_n, Y_n) = (x_n, y_n)\big)}{\mathbb{Q}(Y_{n+1} \in \mathrm{d}y_{n+1} | Y_n = y_n)} \\
&\qquad = \exp\left(\sum_{n=0}^{N-1} \frac{\alpha_n(x_n, y_n)}{\beta_n(x_n)^2} \Delta y_n - \frac{1}{2} \sum_{n=0}^{N-1} \frac{\alpha_n(x_n, y_n)^2}{\beta_n(y_n)^2} \right)
\end{aligned} \tag{C.7}$$

where $\Delta y_n = y_{n+1} - y_n$.

C.1.1.3 The Continuous Filtering Problem

Consider the partially observable process $\big\{(X_t, Y_t, \mathscr{F}_t)\big\}$.

Let us say under $\mathbb{P}$, it solves

$$\mathrm{d}X_t = F(t, X_t)\, \mathrm{d}t + Q(t, X_t)\, \mathrm{d}W_t \tag{C.8}$$

$$\mathrm{d}Y_t = H(t, X_t, Y_t)\, \mathrm{d}t + R(t, Y_t)\, \mathrm{d}V_t^{\mathbb{P}}, \tag{C.9}$$

while under $\mathbb{Q}$,

$$\mathrm{d}X_t = F(t, X_t)\,\mathrm{d}t + Q(t, X_t)\,\mathrm{d}W_t \tag{C.10}$$

$$\mathrm{d}Y_t = R(t, Y_t)\,\mathrm{d}V_t^{\mathbb{Q}}. \tag{C.11}$$

The usual assumptions must be made: F, Q, H, and R are deterministic, Lipschitz in their space variables, and uniformly continuous in time. W_t, $V_t^{\mathbb{P}}$, and $V_t^{\mathbb{Q}}$ are Brownian motions according to their respective measures.

The natural discretizations of the systems are, respectively,

$$\begin{aligned} X_{t_{n+1}} &\approx X_{t_n} + F(t_n, X_n)\Delta t_n + Q(t_n, X_{t_n})\Delta W_{t_n} \\ Y_{t_{n+1}} &\approx Y_{t_n} + H(t_n, X_n, Y_n)\Delta t_n + R(t_n, Y_{t_n})\Delta V_{t_n} \end{aligned}$$

and

$$\begin{aligned} X_{t_{n+1}} &\approx X_{t_n} + F(t_n, X_n)\Delta t_n + Q(t_n, X_{t_n})\Delta W_{t_n} \\ Y_{t_{n+1}} &\approx Y_{t_n} + R(t_n, Y_{t_n})\Delta V_{t_n}. \end{aligned}$$

$\{X_{t_n}\}$ is Markovian in n and $\{Y_{t_n}\}$ fits the form above with the identifications

$$\begin{aligned} \alpha_n(x, y) &= H(t_n, x, y)\Delta t_n \\ \beta_n(y) &= R(t_n, y)\sqrt{\Delta t_n}. \end{aligned}$$

From (C.7), the RN derivative is

$$\exp\left(\sum_{n=0}^{N-1} \frac{H(t_n, x_{t_n}, y_{t_n})}{R(t_n, y_{t_n})^2}\Delta y_{t_n} - \frac{1}{2}\sum_{n=0}^{N-1} \frac{H(t_n, x_{t_n}, y_{t_n})^2}{R(t_n, y_{t_n})^2}\Delta t_n\right).$$

Taking $\Delta t \downarrow 0$, this converges to

$$\exp\left(\int_0^T \frac{H(t, x_t, y_t)}{R(t, y_t)^2}\,\mathrm{d}y_t - \frac{1}{2}\int_0^T \frac{H(t, x_t, y_t)^2}{R(t, y_t)^2}\,\mathrm{d}t\right).$$

C.1.2 The Zakai Equation

Let $\big\{(X_t, Y_t, \mathscr{F}_t)\big\}$, $\mathbb{P}$, and $\mathbb{Q}$ be defined as before (see equations (C.8) – (C.11)) and define the likelihood process

$$L_t = \exp\left(\int_0^t \frac{H(s, X_s, Y_s)}{R(s, Y_s)^2}\,\mathrm{d}Y_s - \frac{1}{2}\int_0^t \frac{H(s, X_s, Y_s)^2}{R(s, Y_s)^2}\,\mathrm{d}s\right),$$

noting that $L_T = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$.

Let $g\colon \mathbb{R} \mapsto \mathbb{R}$ be a function that

1. belongs to C^2 (is twice differentiable everywhere on $\mathbb{R}$)
2. belongs to $L^1\big(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda\big)$ $\big(\int_{\mathbb{R}} |g|\,\mathrm{d}x < \infty\big)$
3. vanishes at $\pm\infty$.

From Itô's lemma, we have

$$\begin{aligned} g(X_t)L_t = g(X_s)L_s &+ \int_s^t L_\tau \mathcal{L} g(X_\tau)\,\mathrm{d}\tau + \int_s^t L_\tau g(X_\tau)\frac{H(\tau, X_\tau, Y_\tau)}{R(\tau, Y_\tau)}\,\mathrm{d}Y_\tau \\ &+ \int_s^t L_\tau Q(\tau, X_\tau)\,\frac{\partial g}{\partial x}\,(X_\tau)\,\mathrm{d}W_\tau, \end{aligned} \tag{C.12}$$

where $\mathcal{L} = F(\cdot,\cdot,\cdot)\frac{\partial}{\partial x} + \frac{1}{2}Q(\cdot,\cdot)^2\frac{\partial^2}{\partial x^2}$ is Kolmogorov's backward operator.

The rest of the derivation is a walkthrough of applying

$$\mathbb{E}^{\mathbb{Q}}\big[\cdot\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\big]$$

to (C.12). For organizational purposes, the process is broken down into five steps.

1. We consider first the single LHS term

$$\mathbb{E}^{\mathbb{Q}}\big[g(X_t)L_t\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\big]. \tag{C.13}$$

Conditioning on values of X_t, the above becomes

$$\int_{\mathbb{R}} \mathbb{E}^{\mathbb{Q}}\big[g(X_t)L_t\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s, X_t = x\big]\,\mathbb{Q}\big(X_t \in \mathrm{d}x\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\big). \tag{C.14}$$

Under $\mathbb{Q}$, $\{X_t\}$ and $\{Y_t\}$ are independent, hence

$$\mathbb{Q}\big(X_t \in \mathrm{d}x\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\big) = \mathbb{Q}(X_t \in \mathrm{d}x|X_s = x_s)$$

and (C.14) becomes

$$\begin{aligned}\int_{\mathbb{R}} g(x)\mathbb{E}^{\mathbb{Q}}\big[L_t\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s,X_t=x\big]\,\mathbb{Q}(X_t\in\mathrm{d}x|X_s=x_s)\\ =\int_{\mathbb{R}} g(x)\mathbb{E}^{\mathbb{Q}}\big[L_t\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s,X_t=x\big]\,q_{X_t|X_s}(x|x_s)\,\mathrm{d}x,\end{aligned} \tag{C.15}$$

where $q_{X_t|X_s}$ is the density of X_t conditioned on X_s. Here, let us define

$$\rho(s,x_s;t,x)=\mathbb{E}^{\mathbb{Q}}\big[L_t\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s,X_t=x\big]\,q_{X_t|X_s}(x|x_s)\,. \tag{C.16}$$

It turns out that ρ is the quantity of interest – the unnormalized transition density of $\{X_t\}$ conditioned on $\{Y_t\}$ (the proof, which we omit here for the sake of brevity, is by abstract Baye's theorem, see equation 10 of [Zak69]). Substituting this into (C.15), we finally attain the expression for (C.13):

$$\mathbb{E}^{\mathbb{Q}}\big[g(X_t)L_t\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s\big]=\int_{\mathbb{R}} g(x)\rho(s,x_s;t,x)\,\mathrm{d}x. \tag{C.17}$$

2. If $t\downarrow s$ in (C.17), then $x\to x_s$ and $\rho(s,x_s;t,x)\to\delta_{x_s}(x)$. Thus, the first RHS term in (C.12) becomes
$$\mathbb{E}^{\mathbb{Q}}\big[g(X_s)L_s\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s\big]=g(x_s). \tag{C.18}$$

3. Let us move onto the second RHS term
$$\mathbb{E}^{\mathbb{Q}}\left[\int_s^t L_\tau\mathcal{L}g(X_\tau)\,\mathrm{d}\tau\,\middle|\,Y_{[0,t]}=y_{[0,t]},X_s=x_s\right]. \tag{C.19}$$
Under certain conditions (see [Zak69]), the order of integration may be changed, which gives us
$$\int_s^t \mathbb{E}^{\mathbb{Q}}\big[L_\tau\mathcal{L}g(X_\tau)\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s\big]\,\mathrm{d}\tau. \tag{C.20}$$
Consider the integrand
$$\mathbb{E}^{\mathbb{Q}}\big[L_\tau\mathcal{L}g(X_\tau)\big|Y_{[0,t]}=y_{[0,t]},X_s=x_s\big]$$

and condition on values of X_τ:

$$\int_{\mathbb{R}} \mathbb{E}^{\mathbb{Q}}\big[L_\tau \mathcal{L}g(X_\tau)\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s, X_\tau = x\big]\, \mathbb{Q}\big(X_\tau \in \mathrm{d}x\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\big)$$
$$= \int_{\mathbb{R}} \mathcal{L}g(x)\mathbb{E}^{\mathbb{Q}}\big[L_\tau\big|Y_{[0,t]} = y_{[0,t]}, X_s = x_s, X_\tau = x\big]\, q_{X_\tau|X_s}(x|x_s)\,\mathrm{d}x. \tag{C.21}$$

Since L_τ depends on the path of $Y_{[0,t]}$ only up to time τ, we are allowed to "omit" the history of $\{Y_\tau\}$ over the period $[s,t]$, and we see that (C.21) is equal to

$$\int_{\mathbb{R}} \mathcal{L}g(x)\rho(s,x_s;\tau,x)\,\mathrm{d}x,$$

or, since $g(x)$ vanishes at $\pm\infty$,

$$\int_{\mathbb{R}} g(x)\mathcal{L}^*\rho(s,x_s;\tau,x)\,\mathrm{d}x, \tag{C.22}$$

where $\mathcal{L}^* = -\frac{\partial}{\partial x}F(\cdot,\cdot,\cdot) + \frac{1}{2}\frac{\partial^2}{\partial x^2}Q(\cdot,\cdot)^2$ is Kolmogorov's forward operator (this can be verified using integration by parts). Substituting (C.22) into (C.20), we see that the second RHS term of (C.12) simplifies to

$$\int_s^t \int_{\mathbb{R}} g(x)\mathcal{L}^*\rho(s,x_s;\tau,x)\,\mathrm{d}x\,\mathrm{d}\tau. \tag{C.23}$$

4. The third RHS term of (C.12) is

$$\mathbb{E}^{\mathbb{Q}}\left[\int_s^t g(X_\tau)\frac{H(\tau,X_\tau,Y_\tau)}{R(\tau,Y_\tau)}L_\tau\,\mathrm{d}Y_\tau\middle|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\right]. \tag{C.24}$$

Again switching the order of integration (for justification, see [Zak69]), we get

$$\int_s^t \mathbb{E}^{\mathbb{Q}}\left[g(X_\tau)\frac{H(\tau,X_\tau,Y_\tau)}{R(\tau,Y_\tau)}L_\tau\middle|Y_{[0,t]} = y_{[0,t]}, X_s = x_s\right]\mathrm{d}y_\tau.$$

Conditioning on X_τ and simplifying just as before, we see that (C.24) becomes

$$\int_s^t \int_{\mathbb{R}} g(x)\frac{H(\tau,x,y_\tau)}{R(\tau,y_\tau)^2}\rho(s,x_s;\tau,x)\,\mathrm{d}x\,\mathrm{d}y_\tau. \tag{C.25}$$

5. Finally, the last RHS term in (C.12) vanishes because $\{W_t\}$ is a Brownian motion under

$\mathbb{Q}$.

Altogether, (C.12) becomes

$$\begin{aligned}\int_{\mathbb{R}} g(x)\rho(s,x_s;t,x)\,\mathrm{d}x = g(x_s) &+ \int_s^t \int_{\mathbb{R}} g(x)\mathcal{L}^*\rho(s,x_s;\tau,x)\,\mathrm{d}x\,\mathrm{d}\tau \\ &+ \int_s^t \int_{\mathbb{R}} g(x)\frac{H(\tau,x,y_\tau)}{R(\tau,y_\tau)^2}\rho(s,x_s;\tau,x)\,\mathrm{d}x\,\mathrm{d}y_\tau.\end{aligned} \tag{C.26}$$

Taking $g(u) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(u-x)^2}{2\sigma^2}\right)$ and letting $\sigma \downarrow 0$, we attain

$$\rho(s,x_s;t,x) = \delta_x(x_s) + \int_s^t \mathcal{L}^*\rho(s,x_s;\tau,x)\,\mathrm{d}\tau + \int_s^t \frac{H(\tau,x,y_\tau)}{R(\tau,y_\tau)^2}\rho(s,x_s;\tau,x)\,\mathrm{d}y_\tau,$$

which is Zakai's equation.

BIBLIOGRAPHY

[Bac92] Kerry Back. Insider trading in continuous time. *The Review of Financial Studies*, 5(3):387–409, 1992.

[BCDF+18] Kerry Back, Pierre Collin-Dufresne, Vyacheslav Fos, Tao Li, and Alexander Ljungqvist. Activism, strategic trading, and liquidity. *Econometrica*, 86(4):1431–1463, 2018.

[BCEL20] Kerry Back, Francois Cocquemas, Ibrahim Ekren, and Abraham Lioui. Optimal transport and risk aversion in kyle's model of informed trading. *arXiv e-prints*, pages arXiv–2006, 2020.

[Bjo20] Tomas Bjork. *Arbitrage Theory in Continuous Time*. Oxford University Press, 4 edition, 2020.

[BP98] Kerry Back and Hal Pedersen. Long-lived information and intraday patterns. *Journal of financial markets*, 1(3-4):385–402, 1998.

[CCDG22] Sylvain Carré, Pierre Collin-Dufresne, and Franck Gabriel. Insider trading with penalties. *Journal of Economic Theory*, 203:105461, 2022.

[CDF16] Pierre Collin-Dufresne and Vyacheslav Fos. Insider trading, stochastic liquidity, and equilibrium prices. *Econometrica*, 84(4):1441–1475, 2016.

[CENV22] Reda Chhaibi, Ibrahim Ekren, Eunjung Noh, and Lu Vy. A unified approach to informed trading via monge-kantorovich duality. *arXiv preprint*, arXiv:2210.17384, 2022. Submitted on 31 Oct 2022.

[Cho03] Kyung-Ha Cho. Continuous auctions and insider trading: uniqueness and risk aversion. *Finance and Stochastics*, 7(1):47–71, 2003.

[CK00] Kyung-Ha Cho and Nicole El Karoui. Insider trading and nonlinear equilibria: Single auction case. *Annales d'Économie et de Statistique*, 60:21–41, 2000. Microstructure des marchés financiers / Financial Market Microstructure.

[DS99] Freddy Delbaen and Walter Schachermayer. A compactness principle for bounded sequences of martingales with applications. In *Seminar on Stochastic Analysis, Random Fields and Applications: Centro Stefano Franscini, Ascona, September 1996*, pages 137–173. Springer, 1999.

[EMZ22] Ibrahim Ekren, Brad Mostowski, and Gordan Zitković. Kyle's model with stochastic liquidity. *arXiv preprint arXiv:2204.11069*, 2022.

[Fré51] Maurice Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Section A: Sciences Mathématiques et Astronomie*, 14:53–77, 1951.

[FV93] F. Douglas Foster and S. Viswanathan. The effect of public information and competition on trading volume and price volatility. *Review of Financial Studies*, 6(1):23–56, 1993. Sponsor: The Society for Financial Studies.

[GS98] Wilfrid Gangbo and Andrzej Swiech. Optimal maps for the multidimensional Monge–Kantorovich problem. *Communications on Pure and Applied Mathematics*, 51(1):23–45, 1998.

[Hoe40] Wassily Hoeffding. Maßstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5(3):179–233, 1940.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2nd edition, 1991.

[Kyl85] Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985.

[LS77] Robert Shevilevich Liptser and Al'bert Nikolaevich Shiriaev. *Statistics of random processes: General theory*, volume 394. Springer, 1977.

[LS13] Robert S Liptser and Albert N Shiryaev. *Statistics of random processes II: Applications*, volume 6. Springer Science & Business Media, 2013.

[LSZ10] Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62. Springer, 2010.

[MN18] P. Del Moral and A. Niclas. A taylor expansion of the square root matrix function. *Journal of Mathematical Analysis and Applications*, 465:259–266, 2018.

[San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[Shr04] Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance. Springer, New York, 2004.

[WR25] Hongjian Wang and Aaditya Ramdas. Positive semidefinite matrix supermartingales. *arXiv preprint arXiv:2401.15567*, 2025.

[YZ99] Jiongmin Yong and Xun Yu Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999.

[Zak69] M Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(2):230–243, 1969.

[Zha17] Jianfeng Zhang. *Backward Stochastic Differential Equations*. Springer, 2017.