On some aspects of Skorokhod Embedding Problem and its applications in Mathematical Finance

These notes originated from short lecture courses I gave at the 5^{th} European Summer School in Financial Mathematics in August 2012 in Paris, at the Chinese University of Hong Kong in March 2013 and at the RTD 1845 (Berlin–Potsdam)—IMPAN Summer School in Będlewo in October 2014.[1]

Jan Obłój

This version October 2014; first version August 2012

[1] I am grateful to all the participants in the above mentioned courses and many others who brought in valuable corrections. These notes undertake excursions into several fascinating and vibrant fields of research. They are not meant to provide a comprehensive overview and often struggle to keep up with the newest developments. In particular, some relevant contributions may not be cited, for which I apologise. *Any comments and corrections are most welcomed and should be emailed to:* `obloj@maths.ox.ac.uk`

Contents

Preliminaries

We list here some conventions and notation.

Throughout increasing and decreasing are understood in a weak sense of non-decreasing and non-increasing respectively. We sometimes use the latter descriptives to stress this convention.

For a set Δ we denote its complement by Δ^c.

For a random variable X, $\mathcal{L}(X)$ denotes its distribution and $X \sim \mu$, or $\mathcal{L}(X) = \mu$, means "X has the distribution μ." For two random variables X and Y, $X \sim Y$ and $X \stackrel{\mathcal{L}}{=} Y$ signify both that X and Y have the same distribution. The Dirac point mass at a is denoted δ_a, or $\delta_{\{a\}}$ when the former might cause confusion. The Normal (Gaussian) distribution with mean m and variance σ^2 is denoted $\mathcal{N}(m, \sigma^2)$. We sometimes write $\mathcal{N} = \mathcal{N}(0, 1)$. Finally, $\mu_n \Rightarrow \mu$ signifies weak convergence of probability measures.
For any probability measure μ, we define its distribution function $\underline{\mu}(x) := \mu((-\infty, x])$ and its tail function $\overline{\mu}(x) := \mu([x, \infty))$.

We will use $\cdot \circ \theta_\tau$ to denote "a shift by a stopping time τ". Note that this is just a notation and is different from the classical shift operator in the theory of Markov processes. Specifically if $H_\Delta(X)$ is the first time the process enters the set Δ and τ is a stopping time then $H_\Delta \circ \theta_\tau$ denotes the first time *after time* τ that the process enters the set Δ.

Chapter 1

From Classical to Robust framework for valuation and hedging

1.1 The Classical modelling framework

Contemporary Mathematical Finance gained momentum with the seminal contributions of Black and Scholes [BS73] on option pricing and Merton [Mer69, Mer71] on optimal investment. At its heart was the classical modelling setup which had its roots in Louis Bachelier's 1900 thesis and Samuelson (1965). The ideal mathematical tools it required were ready to use: the theory of (semi)martingales and stochastic integration going back to Kyioshi Itô and the Strasbourg school of Paul-André Meyer. A rapid growth of the field ensued together with beautiful mathematics: works on no-arbitrage and the Fundamental Theorem of Asset Pricing from Harrison and Kreps [HK79] to Delbaen and Schachermayer [DS94], the mean-variance theory of Schweizer [Sch92] or the optional decomposition of El Karoui and Quenez [EKQ95] and Kramkov [Kra96], the duality theory in portfolio optimisation – to mention a few among many important developments, see e.g. Karatzas and Shreve [KS98]. It also stimulated a dynamic growth of the financial industry. Its importance was underlined with two Nobel Prizes for Harry Markowitz and William Sharpe in 1990 and for Robert Merton and Myron Scholes in 1997.

The dominant "classical" modelling setup in Mathematical Finance consists of specifying a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and an adapted stochastic process (S_t), which models the dynamics of the price process of a *risky asset*. S_0 is the current market price but the future prices are random and *exoge-*

nous[1]. S is often refereed to as the *underlying* since derivative products will have payoffs contingent on paths of S. The market also features another asset S^0, often called the *money market account* or the *bond*, which is a "riskless" asset. It represents the cash value of investment confined to a savings account. In particular, as long as interest rates are non-negative, (S^0_t) is non-decreasing.

Agents are then allowed to trade. A simple trading strategy consists in rebalancing (changing) the holdings in the risky assets at finite number of times t_i, $0 = t_0 < t_1 < \ldots < t_{n-1}$. Say $\phi_{t_{i-1}}$, which is $\mathcal{F}_{t_{i-1}}$ measurable, corresponds to number of units of risky asset S held between t_{i-1} and t_i. The trading is said to be *self-financing* if the capital needed to carry out the above strategy was borrowed (or invested) from the money market account, i.e. the agents holds ϕ^0_t units of the bond at time t with

$$\phi_{t_{i-1}} S_{t_i} + \phi^0_{t_{i-1}} S^0_{t_i} = \phi_{t_i} S_{t_i} + \phi^0_{t_i} S^0_{t_i}, \quad i = 1, 2, \ldots, n-1.$$

It is convenient to work in *discounted units*, i.e. to express all prices in units of S^0: $\tilde{S}_t := S_t / S^0_t$. Assuming the agent starts with no initial capital, the final payoff from the trading, in units of S^0, is given as

$$\begin{aligned}\phi_{t_{n-1}} \tilde{S}_{t_n} + \phi^0_{t_{n-1}} &= \phi_{t_{n-1}} \left(\tilde{S}_{t_n} - \tilde{S}_{t_{n-1}}\right) + \phi_{t_{n-2}} \tilde{S}_{t_{n-1}} + \phi^0_{t_{n-2}} = \ldots \\ &= \sum_{i=1}^{n} \phi_{t_{i-1}} \left(\tilde{S}_{t_i} - \tilde{S}_{t_{i-1}}\right).\end{aligned}$$

Taking the limit to continuous trading we obtain the stochastic (Itô) integral $\int \phi_u \mathrm{d}S_u$, which hints at the intimate link between mathematical finance and stochastic calculus.

The fundamental idea of Black and Scholes [BS73] is that pricing is done through hedging. The unique fair price for a payoff ξ at some future date T is equal to v_0, the capital needed to replicate its cashflow through trading: $\tilde{\xi} = v_0 + \int_0^T \phi_u \mathrm{d}\tilde{S}_u$. To compute v_0 we look for an equivalent probability measure $\mathbb{Q}$ under which $(\tilde{S}_t)$ is a martingale, which yields the *risk-neutral pricing*: $v_0 = \mathbb{E}^{\mathbb{Q}}[\xi / S^0_T]$. It turns out that existence of $\mathbb{Q}$, possibly in some weaker incarnation, is equivalent to absence of *arbitrage* opportunities. The latter is a fundamental economic postulate: the principle of market efficiency. The mentioned equivalence, asserted by the first *Fundamental Theorem of Asset Pricing (FTAP)*, was subject of a groundbreaking stream of research, cf. Delbaen and Schachermayer [DS06]. In particular, it shows that the principle of market efficiency implies S has to be a semimartingale[2]. Again, we see a natural deep link between modern mathematical finance and stochastic analysis.

[1] This is a very important point which was rather revolutionary in 1960ties and to which we will come back later.

[2] Bichteler-Meyer theorem stats that this is necessary for the theory of stochastic (Itô) integration to be well defined, see [DM80].

The insights of Black and Scholes, Merton and others have been hugely influential and instigated a whole industry trading financial derivatives. A *derivative*, also called an *option* or a *contingent claim*, gives the holder a right to an agreed cashflow at a future date T, the maturity. The cashflow is expressed as a function of the path $(S_t : t \leq T)$ of the underlying. Ideas of Black and Scholes show that *assuming a particular stochastic model holds true* a seller can perfectly hedge the risk associated with selling a derivative. Starting from the appropriate initial capital, which the buyer pays, and following the appropriate trading (hedging) strategy the seller can perfectly replicate the pre-agreed payoff.

In [BS73], the risky asset S is assumed to follow the geometric Brownian motion model postulated by Samuelson [Sam65]:

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t, \quad t \geq 0, \tag{1.1}$$

where μ, σ are two constants and W is a standard Brownian motion. This is a *complete market model* and every $\mathcal{F}_T$-measurable non-negative payoff is replicable by an admissible trading strategy[3]. This model, even if appealing in its simplicity, is known to be inadequate for modelling purposes[4]. Many much more complex models have been introduced since. They were often *incomplete* – some risk was inherent in the dynamics and could not be hedged away by trading in the underlying. However this risk could again be well understood and strategies minimising (some measure of) the residual risk have been developed, e.g. the mean-variance hedging mentioned previously. Again, these methods yield a description of "optimal behaviour" *assuming a given model holds.*

1.2 Critique of the classical framework

Both from a theoretical and a practical standpoint, the classical modelling setup is a simplification vulnerable to important critique on at least three grounds. Firstly, it ignores the information present in the market such as the prices of liquidly traded options[5] or time series of data. Secondly, it makes very specific modelling assumptions, in particular it specifies a unique probabilistic description of dynamics of the price process (S_t). Thirdly, it is concerned with an

[3] When we consider continuous trading we have to exclude strategies requiring infinite credit line since they can produce arbitrage. This is easily done by requiring that the wealth process associated to a trading strategy remains above a certain constant $a \leq 0$.

[4] Its simplest default is probably that it prescribes a flat volatility surface while in practice market prices exhibit a implied volatility (σ) which is different for different strikes and maturities.

[5] Market prices are available not only for assets but also for many derivative products written on them and should be treated as inputs and not outputs. The industry deals with it through model calibration (reverse-engineering): tweak model parameters so as to match today's prices. This has to be then repeated on daily basis effectively changing the model and introducing theoretical inconsistency.

idealised frictionless market which can be quite different from a realistic market where participants pay transaction costs, liquidity is limited, counterparties may default, etc.

As we saw above, the powerful martingale methods were tailored suited to the classical modelling framework and in a way what was a blessing became a curse. The models could get arbitrary complex but they largely shared the underlying framework with its weaknesses: inflexibility due to a choice of a particular probabilistic setup and failure to incorporate market information in a consistent manner. This coupled with inaccuracy resulting from ignoring market frictions.

However, in practice, the classical framework was being widely used with its important limitations being overlooked. The 2008 financial crisis played out possible negative consequences in a rather spectacular fashion. It should be highlighted that there is nothing "wrong" with the classical modelling framework. As any modelling approach, it has its advantages and its limitations. The problem was that it provided appealing and relatively simple answers which were often applied without thinking about the standing assumptions made in the first place.

Naturally, to understand how violations of fundamental assumptions of the classical approach impact its answers we need to abandon the classical framework and consider a more general, alternative approach. An alternative does not mean a replacement. Indeed, the classical framework is not easily replaced for its scope of applications or for its clear outputs: unique prices and hedging strategies or explicit optimal investment policies. Industry practitioners will most likely continue to apply it. Crucially however, they need tools to understand the risks taken and a new framework to apply when the classical one fails. In the words of Steven Shreve[6] : *For banks, the only way to avoid a repetition of the current crisis is to measure and control all their risks, including the risk that their models give incorrect results.* Given both the theoretical importance of the topic and its dramatic practical illustrations, it is not surprising that it has fuelled an impressive stream of research.

We stressed above that the classical framework starts by postulating a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. As we describe it here, this is a courageous leap of faith. It was also a rather shocking departure from economics practice which tried to model the prices as arising in an equilibrium between supply and demand, see the foreword by Samuelson in [DE06]. It looks more natural if we consider the actual goal of Black, Scholes and Merton. The assumption was that the stock was liquid and its price random and given by the market. The interest was in pricing (and hedging) a warrant, a call option, on this stock. The an-

[6] *Don't Blame The Quants*, Forbes, Aug 2008; available at: http://www.forbes.com/2008/10/07/securities-quants-models-oped-cx_ss_1008shreve.html.

swers were clear and, as one would expect from a good model, quite robust with respect to departures from the specific modelling assumptions in (1.1). More specifically, if we are hedging a European option with convex payoff and the true volatility is a random process σ_t which is however close to the assumed constant value σ then our hedging error is not too large. If $\sigma_t \leq \sigma$ then we are guaranteed to superreplicate, see El Karoui, Jeanblanc and Shreve [EKJPS98] and Hobson [Hob11]. However this is not necessarily true when one considers complex derivatives or more involved stochastic dynamics for the underlying. This motivates the idea to relax the assumption of a uniquely specified probability measure and introduce the inherent *model uncertainty* or *model ambiguity*.

Knight [Kni21] was possibly the first one to distinguish *risk* and *uncertainty*. The former comes from the randomness within a given probabilistic universe (our classical model), or the *known unknown*, and is quantified in the model. The latter corresponds to the possibility that our given classical model is an inadequate description of reality, or the *unknown unknown*. To try to describe it, understand it and quantify it, we need to consider setups without a single specified probability measure. The first stream of research concentrates on *drift uncertainty* in assuming that a specific model is unknown but comes from a set of probability measures, all equivalent to a reference measure $\mathbb{P}$. For the purposes of valuation of derivatives this is largely irrelevant – different models will induce *the same* dynamics of $\tilde{S}$ under the martingale measure. In contrast, for optimal investment problems this extended setup presents new considerable difficulties. Nevertheless, such a setting for model ambiguity essentially relies on classical methods. It would be hard to give justice here to these developments; see for example Maccheroni et al. [MMR06] and Schied [Sch07]. Significantly, the economics literature on Robustness and the *Knightian uncertainty* includes a number of papers by the 2011 and 2013 Nobel Laureates Thomas Sargent and Lars Peter Hansen, e.g. [HS10].

Generalising the form of model uncertainty, researchers considered so-called non-dominated setups where the measures may be mutually singular. This allows for *uncertain volatility*, as pioneered by Lyons [Lyo95] and Avellaneda et al. [ALP95]. Notable recent developments here are linked to the G-expectation of Peng [Pen07] and the quasi-sure stochastic integration based on capacity theory in Denis and Martini [DM06] and on the aggregation method in Soner, Touzi and Zhang [STZ11]. In discrete time a corresponding generalisation of the FTAP was obtained by Bouchard and Nutz [BN]. There have been some more radical departures, e.g. Bick and Willinger [BW94], Cassese [Cas08] and Vovk [Vov09], proposing a modelling framework without a pre-specified probability measure. These contributions largely come from outside of mainstream financial mathematics. They involve non-standard tools such as analysis of finitely additive measures in [Cas08] and a game theoretical approach to probability in

[Vov09]. We will come back to these developments later in the text.

Let us now move towards the second critique of the classical framework. The motivating observation here is that the market prices of liquidly traded derivative instruments ought to be treated as *inputs* rather than *outputs* of modelling. To accommodate it within the classical framework, one could dramatically increase the dimensionality and model simultaneously the asset S and options written on it. This results in a highly dimensional (or infinite dimensional) universe of underlyings with added difficulty of singular constraints on their dynamics: some assets have to be equal to given functions of other assets at some future dates. Such modeling efforts go by name of *market models* and include e.g. Schönbucher [Sch99], Schweizer and Wissel [SW08], Jacod and Protter [JP10], Carmona and Nadtochiy [CN09]. Mathematically these are often very involved precisely due to the embedded consistency conditions and are typically specialised to a particular choice of the market input. More importantly for us however, these works do not offer any answer to the first fundamental critique of the classical framework. They fail to incorporate any model uncertainty. To the contrary: if specifying a unique probabilistic setup for asset prices was questionable it is even more so in the case of market models.

Finally, we should mention that there is a large body of literature which incorporates various market imperfections into the classical framework. Systematic efforts have centred mainly around the questions of liquidity and transaction costs. This is a fascinating and often mathematically very involved research area which we do not discuss here.

1.3 What is a model?

To built a suitable relaxation of the classical modelling framework in mathematical finance we need to first reflect on how models are built. The motivating need for models is practical: to obtain answers to concrete questions in real world applications. These could include

- prices and hedges for derivatives,
- optimal portfolios for investment,
- risk quantification for risk management.

We will refer to these as model *outputs*. They are obtained starting from model *inputs* and applying *reasoning principles*. It is convenient to distinguish three types of inputs:

- *Beliefs*: describe possible future evolution of risky assets. This could be very specific as in the classical framework or very general, e.g. believing that risky asset could follow any continuous path.

- *Information*: the subset of existing market data we trust and need our model to treat consistently as input. Our model will need to be consistent with this information. The information here is understood broadly: these could be given specific market quotes as well as bounds obtained using statistical analysis of time series of past data.
- *Rules*: who can trade what and how. In particular we need to specify which, if any, of market frictions we want to account for.

These inputs are of a very different type and together they yield our starting point. The classical framework corresponds to very strong beliefs, no use of information and simple rules. More precisely, in the classical framework beliefs are a leap of faith: we specify a fixed probability space $(\Omega, \mathcal{F}, (\mathcal{F}_{\mathrm{t}}), \mathbb{P})$ together with adapted processes S which represent the dynamics of risky assets. We usually do this by fixing a Stochastic Different Equation (SDE) which governs the dynamics of S. This allows us to use a single piece of market information: today's price S_0. Any other information is, for a moment, discarded. Rules are supposed to reflect both the actual market practice and our choice of simplifying assumptions. The idea is to capture these restrictions and frictions which, we think, have a first order effect. Others are neglected for simplicity. For example, we could assume there are no transaction costs and interest rates for landing and borrowing are the same but if there is a ban on short-selling we could decide to incorporate this as an important trading restriction.

The underlying *Reasoning Principle* which allows to deduce outputs from inputs is the principle of *market efficiency*. This translates into absence of arbitrage opportunities and hence into, e.g. pricing through the cost of replication. As highlighted above, this is made operational thanks to the Fundamental Theorem of Asset Pricing (FTAP) and the link to martingale tools it brings, see Delbaen and Schachermayer [DS06] for a detailed account.

In the classical approach, only once the outputs are specified we can come back to the question of market *Information*. The outputs depend on *Beliefs*, e.g. parameters in the SDE for dynamics of S. One can try to manipulate the parameters in such a way that outputs match, or are not far from, a given set of market prices of options we want to reproduce. This is called *calibration*. We stress that calibration is a reverse engineering procedure. It is not a consistent way of treating the information. Typically, on the next day new information is available which is inconsistent with the previously calibrated parameters and the model is re-calibrated, i.e. the model is changed, losing theoretical consistency underpinning hedging.

The works on model uncertainty mentioned above start with weaker beliefs: $(\Omega, \mathcal{F}, (\mathcal{F}_t), \{\mathbb{P}_\alpha : \alpha \in \Lambda\})$ where instead of a single probability measure we have a family of probability measures. S still represents the risky assets but it has different dynamics for different α. We could treat all $\mathbb{P}_\alpha$ equally seriously and require the outputs, e.g. hedging strategies, to work $\mathbb{P}_\alpha$-a.s. for *all* $\alpha \in \Lambda$. Or we could also have specify a prior describing our belief of likelihood of various $\mathbb{P}_\alpha$. This is the point of view often taken for optimal investment under model uncertainty, see Föllmer, Schied and Weber [FSW09].

We note that having a family of probability measures instead of a single measure may have far reaching consequences. Consider for example trading strategies. There is no ambiguity in defining (frictionless) trading at discrete times. We saw however that continuous time trading corresponds to stochastic integrals. These are well defined a.s., in particular if all $\mathbb{P}_\alpha$ are absolutely continuous with respect to one reference measure $\mathbb{P}$ we have a way to define stochastic integrals for all $\alpha \in \Lambda$ simultaneously. However in the case of uncertain volatility this is no longer true. Mathematically, this is a challenging issue, which was solved in number of different ways, see e.g. Lyons [Lyo95], Peng [Pen07], Denis and Martini [DM06], Soner, Touzi and Zhang [STZ11] or Nutz [Nut12].

The essential point we want to stress here is that all these works relax the classical framework. Necessarily, they obtain less precise results. It seems more natural to try to consider two weaknesses of the classical framework simultaneously: allow for model uncertainty and introduce a coherent use of market information. On one hand we relax strong assumptions and on the other hand we compensate by adding new features and making use of information. This is what the robust approach aims to do.

1.4 Robust framework for valuation and hedging

We develop here a general robust framework for valuation and hedging which is based on pathwise arguments, following [CO11b, DOR14]. The essential idea is to encode beliefs through a choice of space of possible paths of the risky asset. The motivation is that in reality we only see one path so we should be able to formulate our beliefs in terms of this path. As we will see below, this is actually a rather flexible framework which allows us in particular to recover the classical Black-Scholes (1.1) model.

1.4.1 Introducing the framework

For simplicity assume we have only one risky asset S which does not pay dividends. This can be easily incorporated but would make notation more laborious. We also fix (the last) maturity T. The *Beliefs* are given by a set $\mathfrak{P}$ which is a Borel subset of càdlàg functions on $[0, T]$. These are the paths we think S

may follow. We agree to disregard any paths outside of $\mathfrak{P}$. Simple examples are given by

- $\mathbb{D}([0,T],\mathbb{R})$ – all càdlàg functions on $[0,T]$;
- $C([0,T],\mathbb{R}_+)$ – all non-negative continuous functions on $[0,T]$;
- functions in $C([0,T],\mathbb{R}_+)$ which admit quadratic variation and for which $\frac{\mathrm{d}\langle S\rangle_t}{\mathrm{d}t}$ exists and is equal to $\sigma^2 S_t^2$.

The first example corresponds to "no beliefs" and the second one to mild beliefs about lack of jumps. The last example is essentially equivalent to assuming Black-Scholes model. Note that already at this stage we will use at least one piece of information – typically we will know S_0 and hence we will supplement the above specifications and assume all paths in $\mathfrak{P}$ have the same common starting point S_0. We have the natural filtration $\mathcal{F}_t = \sigma(S_u : u \leq t)$ i.e. the σ-field generated by one-dimensional projections, which we take right-continuous. This allows us to talk about stopping times. In particular hitting times of closed and open sets are stopping times, see Section 2.1.

More specifically, we also need to specify paths for S^0, the riskless asset. We could either prescribe one possible path (deterministic interest rates) or allow for a range of paths typically non-decreasing or with bounded variation. So in fact $\mathfrak{P}$ corresponds to two dimensional paths. However, for simplicity of notation, we write both $(S^0,S)\in\mathfrak{P}$ and $S\in\mathfrak{P}$ and hope this should not cause any confusion.

The market *Information* we want to consider comes in form of market prices for a set of (liquidly) traded options. More specifically, we let $\mathcal{X}$ be the set of payoffs of options traded today. $\mathcal{X}$ may be finite or infinite. $\xi\in\mathcal{X}$ is simply a measurable function from $\mathfrak{P}$ to $\mathbb{R}$. On $\mathcal{X}$ we can define a pricing operator $\mathcal{P}$ which encodes the market prices available today. We do not assume that options in $\mathcal{X}$ are traded at any future dates so any position in options is held until maturity.

It remains to specify the *Rules*. We assume a frictionless market setting. This implies that $\mathcal{P}$ is linear when defined and hence we can assume $\mathcal{X}$ is a vector space. Further, we assume that trading is discrete. Note that continuous trading would require a pathwise notion of stochastic integration. Any trading strategy X is semi-static: static in options and dynamic (discrete) in the stock. Its payoff is given by, with the notation $\tilde{S}_t := S_t/S_t^0$,

$$X_T = \sum_{i=1}^{n} a_i(\xi_i(S_t : t\leq T) - S_T^0\mathcal{P}\xi_i) + S_T^0\sum_{j=1}^{m}\phi_{\tau_{j-1}}(\tilde{S}_{\tau_j} - \tilde{S}_{\tau_{j-1}}), \quad \xi_i\in\mathcal{X}, \tag{1.2}$$

where $0\leq\tau_0<\tau_1<\ldots<\tau_{m-1}<\tau_m=T$ are stopping times, $a_i\in\mathbb{R}$ and ϕ_{τ_j} are $\mathcal{F}_{\tau_j}$ measurable. The initial cost of the above is zero, $X_0=\mathcal{P}X_T=0$.

However we also need to impose some admissibility condition to avoid arbitrage opportunities. For the moment being we simply assume that all ϕ_{τ_j} are bounded on $\mathfrak{P}$. We will come back to this issue. Finally, the above features a finite portfolio of traded options. In principle we could, and sometimes will, allow for an infinite portfolio as long as its initial price, here $\sum_{i=1}^{n} a_i$, is finite. We denote $\mathcal{A}$ the set of *admissible* trading strategies. Finally note that while we might speak of X_t which is a certain portfolio in options maturing at T, in stock and in bond, we can not speak of price of X_t as this is not necessarily well defined.

1.4.2 Market models, arbitrage and FTAP

By a classical model we mean a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{Q})$ with adapted processes S^0 and S. We are interested in pricing so $\mathbb{Q}$ plays the role of the risk neutral measure.

Definition 1.4.1. *We say that a classical model is a* $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$–market model *if* $\tilde{S}_t$ *is a* $\mathbb{Q}$*-martingale,* $((S_t^0, S_t) : t \le T) \in \mathfrak{P})$ *a.s. and* $\mathbb{E}^{\mathbb{Q}}[\xi/S_T^0] = \mathcal{P}\xi$ *for all* $\xi \in \mathcal{X}$*.*

A market model is thus a fully specified classical framework model which admits no arbitrage and agrees with our beliefs and with the given market information. Once we know a market model exists we can use it to *define*, through conditional expectation, joint arbitrage-free dynamics for all $t \le T$, of S and all $\xi \in \mathcal{X}$.

Hence existence of a market model is the equivalent of "existence of a risk neutral measure" in the classical context. A Robust FTAP should thus establish an equivalence between existence of a market model and market efficiency, i.e. absence of arbitrage opportunities. Note however that it is not clear at all how to even define the latter. The usual notions of arbitrage depend on $\mathbb{Q}$. Here we need a notion of arbitrage which only uses beliefs and market information. A natural candidate is given by

Definition 1.4.2. *We say that there is a* strong arbitrage *if there exists a portfolio* $X \in \mathcal{A}$ *with* $X_T(S_t : t \le T) > 0$ *for all* $S \in \mathfrak{P}$*.*

In the case of "no-beliefs", when $\mathfrak{P}$ corresponds to all càdlàg (or all continuous) paths, the above is known as *model-independent arbitrage*. Existence of a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$-market model implies absence of a strong arbitrage. Indeed, since ϕ_{τ_j} are assumed bounded, we have $\mathbb{E}^{\mathbb{Q}}[X_T/S_T^0] = 0$ and then $X_T \ge 0$ implies $X_T = 0$ $\mathbb{Q}$-a.s. The converse, giving a version of a Robust FTAP, may be true under suitably strong additional assumptions, see Acciaio et al. [ABPS13]. However in general, the converse is *not* true as seen from the following simple example of Davis and Hobson [DH07]. Assume zero interest rates, let $\mathfrak{P} = C([0, T], \mathbb{R})$,

$\mathcal{X} = \{(S_T - K_1)^+, (S_T - K_2)^+\}$ and $S_0 > \mathcal{P}(S_T - K_1)^+ = \mathcal{P}(S_T - K_2)^+ > 0$ where $0 < K_1 < K_2$. In words, we have a simple market where two call options trade at the same non-zero price. If we combine this with a classical model then it would seem that there is an obvious arbitrage opportunity. However a more careful look reveals that the arbitrage strategy depends on the zero sets of the model. More precisely, if we have a classical model where $S_T \leq K_2$ a.s. then we simply sell the call with strike K_2 making profit of $\mathcal{P}(S_T - K_2)^+$. Otherwise, we also buy the other call and obtain, at zero initial cost, a non-negative payoff which is strictly positive with positive probability. This situation was termed *weak arbitrage*, see [DH07] and [DOR14]. A notion of *weak free lunch with vanishing risk* was introduced in Cox and Obłój [CO11b] which turned out to be suitable for the case of infinite $\mathcal{X}$.

The fundamental observation we want to stress is that presence of arbitrage is a feature of a given *modelling setup* and may not correspond to a single trading strategy. Thinking that an arbitrage opportunity should be given be a (single) trading strategy is a legacy of the classical approach. Implicit in classical definition of an arbitrage was the fact that the arbitrage opportunity was *relative* to modelling inputs: beliefs, information and rules. Changing inputs affects arbitrage opportunities. As a well-known example, recall that considering fractional Brownian motion within the classical framework leads to an arbitrage (in the sense of WFLVR of Delbaen and Schachermayer [DS94]) but when arbitrarily small transaction costs are added the arbitrage opportunities disappear, as shown by Guasoni [Gua06]. We say that one modelling framework – a triplet of beliefs, information and rules – is a refinement of another one if it leads to sharper outputs. This may be due to stronger beliefs, more information or a richer set of admissible trading strategies. A modelling framework admits arbitrage if all of its refinements admit arbitrage.

A particularly interesting case is when information and rules are fixed but we keep strengthening beliefs. For simplicity assume $\mathcal{X}$ is finite. Then, it may be shown that absence of a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$*–market model* is essentially equivalent to existence of

$$\emptyset = \mathfrak{P}_0 \subset \mathfrak{P}_1 \subset \ldots \subset \mathfrak{P}_n = \mathfrak{P}$$

and a family of trading strategies X^i such that X^i is non-negative on $\mathfrak{P}_i$ and strictly positive on $\mathfrak{P}_i \setminus \mathfrak{P}_{i-1}$. We refer to this as *Robust (Sequential) Arbitrage*. If an agent is willing to accept $\mathfrak{P}_i$ as beliefs, but is not willing to restrict to paths in $\mathfrak{P}_{i-1}$ only, then she will see X^i as an arbitrage opportunity. And each agent who accepts $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$ as possible modelling setup will be in this position for some $i = 1, \ldots, n$. It is not hard to see that in the special case of Acciaio el al. [ABPS13] one can take $n = 1$ thanks to a dominating derivative outside of a compact set and possibility to aggregate the strategies on a compact set. In particular Robust Arbitrage then reduces to model–independent arbitrage.

In these notes we will largely avoid the above fine questions of no-arbitrage. In particular, we do not give a precise statement of the above general version of a Robust FTAP. We should highlight however that, to the best of our knowledge, the quest for the most general version of a Robust FTAP with continuous trading is still ongoing. This is an exciting and challenging topic for future research.

Finally, let us point out the difference between the above analysis and the quasi-sure setting mentioned before. In the classical framework, absence of arbitrage essentially means that if a payoff from trading, with zero initial capital, is nonegative $X_T \geq 0$ then $X_T = 0$ a.s. Introducing model-uncertainty by considering a family of measures $\Lambda = \{\mathbb{P}_\alpha : \alpha \in ...\}$ it is natural to say that absence of arbitrage opportunities means that $X_T \geq 0$ implies $X_T = 0$ $\mathbb{P}_\alpha$–a.s. for all α, often denoted Λ–q.s. This is a strong requirement and a version of FTAP then asserts that it is equivalent to existence of an equivalent martingale measure $\mathbb{Q}_\alpha$ for each $\mathbb{P}_\alpha$, see Bouchard and Nutz [BN]. Such approach has a conceptually different starting point to the one proposed here and it arrives at different notions of arbitrage and its duality in FTAP.

1.4.3 Outputs: pricing and hedging duality

The motivating financial question, or pair of questions, we want to consider is:

Given our modelling setup (Beliefs, Information and Rules), what is the range of prices for an option with payoff $O_T = O(S_t : t \leq T)$ which do not introduce strong arbitrage? And if the option trades at a price outside this range, what is the arbitrage strategy?

It is easy to see that the range of prices constitutes an interval. Indeed if there exists a strong arbitrage when $\mathcal{P}O_T = p$ then, depending if the arbitrage strategy goes long or short in O, there is also a strong arbitrage for all $\mathcal{P}O_T < p$ or all $\mathcal{P}O_T > p$ respectively. Let us denote its lower and upper bounds $LB_{(\mathfrak{P},\mathcal{X},\mathcal{P})}$ and $UB_{(\mathfrak{P},\mathcal{X},\mathcal{P})}$ respectively. Absence of strong arbitrage implies that a superhedging strategy for O_T will be at least as expensive as O_T. This implies that infimum over prices of all superhedging strategies will be greater or equal to $UB_{(\mathfrak{P},\mathcal{X},\mathcal{P})}$:

$$\begin{aligned} UB_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) &\leq \overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) \\ &:= \inf \{p : \exists X \in \mathcal{A}\ \forall S \in \mathfrak{P}\ O(S_t : t \leq T) \leq p + X_T(S_t : t \leq T)\}. \end{aligned} \tag{1.3}$$

Note that we require the superhedging property to hold on all paths but only in $\mathfrak{P}$ and recall that by definition X requires no initial capital: $\mathcal{P}X_T = 0$. On the other hand, if we have a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$–market model then we obtain a possible price for our exotic derivative and hence

$$\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) := \sup_{(\mathfrak{P},\mathcal{X},\mathcal{P})\text{-market models}} \mathbb{E}^{\mathbb{Q}}[O_T/S_T^0] \leq UB_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O). \tag{1.4}$$

Naturally, analogous statements hold for the lower bound with $\underline{LB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ denoting supremum of prices of subhedging strategies and $\overline{LB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ denoting the infimum over market model prices. One may expect that a no-duality gap result holds true and $\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$. In [DOR14] a particular setup of $\mathcal{X}$ consisting of n put options and $O_T = \lambda(S_T)$ a European option with a convex payoff is considered. Then no-duality gap is deduced from classical semi-infinite linear programming results. More recently, in a general discrete time setting, Beiglböck, Henry-Labordère and Penkner [BHLP11] established the desired equality using (and developing) optimal transport theory. Dolinsky and Soner [DS14] obtained results in continuous time, for continuous paths, and under strong continuity assumptions on the payoff. The general case is still an open problem. There are also cases when duality fails to hold, either due to singularities or to market setup, e.g. trading restrictions, see Cox, Hou and Obłój [CHO14]. Finally, we note that instead of pathwise one can consider superhedging in quasi-sure sense. Pricing-hedging duality can be then obtained in considerable generality, see Bouchard and Nutz [BN], Neufeld and Nutz [NN13] and Possamaï, Royer and Touzi[PRT13].

Example 1.4.3. *Note that sometimes it is trivial to establish no duality gap by showing that in fact* $\underline{LB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ *and hence we have a unique price. Consider for example* $\mathfrak{P}$ *given by continuous non-negative functions starting at* S_0*, no interest rates (i.e.* S^0 *constant equal to* 1*) and* $\mathcal{X} = \{(S_T - K)^+ : K \in \mathbb{K}\}$ *for some set* $\mathbb{K} \subset (0,\infty)$*. We assume the given prices* $\mathcal{P}$ *do not admit arbitrage in the sense that a* $(\mathfrak{P},\mathcal{X},\mathcal{P})$*–market model exists. Fix* $K \in \mathbb{K}$ *and consider an up and in put with strike and barrier at* K*:* $O_T = O(S_t : t \le T) = (K - S_T)^+ \mathbf{1}_{\sup_{u\le T} S_u \ge K}$.

Then the following strategy: buy call with strike K *and when* K *is reached enter a forward:*

$$X_T = (S_T - K)^+ + (K - S_T)\mathbf{1}_{\rho \le T}, \quad \rho = \inf\{t \ge 0 : S_t \ge K\},$$

where we use the assumption of continuity of paths to write $S_\rho = K$*. It is immediate that* $X_T = O_T$ *for any* $S \in \mathfrak{P}$ *and* $X_T - \mathcal{P}(S_T - K)^+ \in \mathcal{A}$*. Since we have a perfect replication strategy we see that* $\underline{LB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \mathcal{P}X_T = \mathcal{P}(S_T - K)^+$*.*

Existing literature on Robust Pricing and Hedging has been focusing on particular examples of $\mathfrak{P}, \mathcal{X}$ and O, establishing the equality $\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ and identifying the cheapest superhedge and the market model which achieves the supremum. The methodology, based of Skorokhod embeddings and pathwise inequalities, was pioneered by Hobson [Hob98a] for lookback options. It was continued in Brown et al. [BHR01] for barrier options. More recent developments include [CW13, CO11a, CO11b, CHO08]. We describe this methodology in the next section.

1.4.4 Robust Pricing and Hedging via Skorokhod embeddings

We outline now, briefly and in a schematic way, the underlying methodology which links the Skorkohod embedding problem with Robust Pricing and Hedging. Assume that the interest rates are zero. Alternatively we can think of S as the forward price as long as it makes sense to consider options written on the forward. Assume further that $\mathfrak{P} = C([0,T], \mathbb{R}_+)$ and that $\mathcal{X} = \{(S_T - K)^+ : K \geq 0\}$. Suppose that given prices $\mathcal{P}$ do not admit arbitrage themselves and there exists a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$-market model. In this model we have

$$C(K) := \mathcal{P}(S_T - K)^+ = \int_K^\infty (s - K)\mu(\mathrm{d}s), \quad \mu \sim_{\mathbb{Q}} S_T.$$

Direct arguments show that $C(K)$ is a non-increasing convex function. Differentiating we obtain the so-called Breeden-Litzenberger [BL78] formula

$$C'(K-) = -\mu([K,\infty)), \quad C'(K+) = -\mu((K,\infty)), \quad C''(\mathrm{d}K) = \mu(\mathrm{d}K). \tag{1.5}$$

In particular, the knowledge of call prices determines uniquely the risk neutral distribution of S_T. We often refer to μ as the distribution of S_T implied by the call prices.

In the market model, (S_t) is a continuous $\mathbb{Q}$-martingale and hence, by the Dambis-Dubins-Schwarz theorem 2.3.6, it is a time-changed Brownian motion: $S_t = W_{\tau_t}$. Further $(W_{s\wedge\tau_T} : s \geq 0) = (S_{C_s \wedge T} : s \geq 0)$ is a uniformly integrable martingale. We call such stopping time τ_T UI, see Section 2.2. Finally, suppose that the payoff of the exotic derivative we are interested in is invariant under time changes: $O(S_t : t \leq T) = O(W_u : u \leq \tau_T)$. An example is given by a one-touch (digital barrier) option $O(S_t : t \leq T) = \mathbf{1}_{\sup_{t\leq T} S_t \geq b}$. Then we have

$$\mathbb{E}^{\mathbb{Q}}[O(S_t : t \leq T)] = \mathbb{E}^{\mathbb{Q}}[O(W_u : u \leq \tau_T)].$$

We conclude that

$$\begin{aligned} &\inf_{\tau: W_\tau \sim \mu,\ \tau \text{ is UI}} \mathbb{E}^{\mathbb{Q}}[O(W_t : t \leq \tau)] = \underline{LB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) \leq \mathcal{P}O_T, \quad \text{and} \\ &\mathcal{P}O_T \leq \underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \sup_{\tau: W_\tau \sim \mu,\ \tau \text{ is UI}} \mathbb{E}^{\mathbb{Q}}[O(W_t : t \leq \tau)]. \end{aligned} \tag{1.6}$$

As we will see the stopping times appearing in the above display constitute solutions to the Skorokhod embedding problem of μ in W. For some exotic options previous probabilistic works give us immediate answers identifying the above bounds and the optimal stopping times. In other cases we have to construct new extremal embeddings. There is however a rich methodology at our disposal which we will study in detail in these notes.

Finally, if we find the stopping time τ^* which, say, maximises $\mathbb{E}^{\mathbb{Q}}[O(W_t : t \leq \tau)]$ then we can define a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$-market model which attains $\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$. Analysing the hedging strategy in this extremal model we hope to guess the cheapest superhedging strategy. More precisely, if we can construct a superhedge of O_T at the initial price $\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ then in one go we show that there is no duality gap and identify the strong arbitrage strategy available when O_T trades above the upper bound.

Note that if we have any superreplication strategy we can take it as our starting point. It obviously induces an upper bound on $\overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ and if we can construct a $(\mathfrak{P}, \mathcal{X}, \mathcal{P})$–market model in which this superhedge is a perfect hedge then we know it is the cheapest superhedge, we deduce no-duality gap as well as optimality properties of the embedding which we use to construct the market model. We continue this reasoning in Chapter 5.

Chapter 2

On some aspects of time changing

This chapter provides an overview of useful results concerning stopping times and time changing. We assume that a filtered probability space is given $(\Omega, \mathcal{F}, (\mathcal{F}_{\mathrm{t}}), \mathbb{P})$. Throughout, with the exception of Section 2.1 where a more general discussion is presented, we assume the usual hypothesis (the filtration is right continuous and all $\mathcal{F}_t$ and $\mathcal{F}$ are complete). All processes are assumed to be càdlàg and in particular progressively measurable.

2.1 Stopping times and their properties

We assume now that a filtered probability space is given $(\Omega, \mathcal{F}, (\mathcal{F}_{\mathrm{t}}), \mathbb{P})$. We let $\mathcal{F}_\infty := \sigma\left(\sum_{t\geq 0} \mathcal{F}_t\right)$.

Definition 2.1.1. *A measurable function $\tau : (\Omega, \mathcal{F}) \to [0, \infty]$ is called a stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$, for all $t \geq 0$. The class of sets $A \in \mathcal{F}_\infty$ such that $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$ is a σ-algebra denoted $\mathcal{F}_\tau$.*

Note that under the usual assumptions $(\mathcal{F}_t)$ is right-continuous and completed with $\mathbb{P}$ null sets and then τ is a stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$ or equivalently when the left-continuous process $\mathbf{1}_{(0,\tau]}(t)$ is adapted.

A deterministic time is a stopping time. Minimum and maximum of two stopping times are also stopping times. Finally, if $\rho \leq \tau$ are two stopping times then $\mathcal{F}_\rho \subset \mathcal{F}_\tau$. For many other basic properties of stopping times we refer to Revuz and Yor [RY01, Sec I.4].

Consider now a stochastic process X taking values in a metric space $(E, \mathcal{E})$ endowed with its Borel σ-field. We assume X is progressively measurable which ensures that the random variable $X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau$ measurable. We write $X^\tau = (X_{t\wedge\tau} : t \geq 0)$ for the stopped process. Note that progressive measurability is not a stringent assumption and is implied by X being adapted and left- or right-continuous.

We are interested in hitting times of X. Consider a Borel set Δ and let

$$H_\Delta(X) := \inf\{t \geq 0 : X_t \in \Delta\}, \quad \text{where } \inf\{\emptyset\} = \infty, \tag{2.1}$$

which is called the hitting time of, or the entry time to, Δ. We write $H_\Delta(X) = H_\Delta$ when there is no ambiguity about the process we consider. In all generality it is not true that H_Δ is a stopping time. But if paths of X and the set Δ are regular enough than H_Δ is indeed a stopping time. First, if Δ is closed and X has continuous paths then H_Δ is in fact a stopping time relative to natural filtration of X: $\mathcal{F}^X_t := \sigma(X_s : s \leq t)$. If X is only assumed to have right-continuous paths and Δ is open then H_Δ is a stopping time relative to the right-continuous version of the natural filtration: $(\mathcal{F}^X_{t+})$.

The hitting time of Δ is a particular example of a more general stopping time called the *debut* of a progressively measurable set. Let $\Gamma \subset \mathbb{R}_+ \times \Omega$ and define

$$D_\Gamma(\omega) := \inf\{t \geq 0 : (t,\omega) \in \Gamma)\}, \quad \text{where } \inf\{\emptyset\} = \infty. \tag{2.2}$$

We have the following important result

Theorem 2.1.2. *If $(\mathcal{F}_t)$ is right-continuous and complete then the debut D_Γ of a progressively measurable set Γ is a stopping time.*

Proof. The set $\Gamma_t := \Gamma \cap ([0,t) \times \Omega)$ belongs to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$ since Γ is progressively measurable. On the other hand we see that $\{D_\Gamma < t\}$ is the projection of Γ_t onto Ω. By the projection theorem in measure theory we conclude that $\{D_\Gamma < t\} \in \mathcal{F}_t$ which is sufficient since $(\mathcal{F}_t)$ is right-continuous. □

We note that the projection theorem invoked above is a difficult result and it requires that $\mathcal{F}_t$ is complete. We note also that if τ is a stopping time then $\tau = D_\Gamma$ where $\Gamma = \{(t,\omega) : T(\omega) \leq t\}$. Note that $\mathbf{1}_\Gamma(t) = \mathbf{1}_{[\tau,\infty)}(t)$ is an adapted right-continuous process and hence progressively measurable. We see therefore that the above Theorem in fact covers *all* stopping times. As an important consequence, under the usual hypothesis if X is a progressively measurable process and $\Delta \in \mathcal{E}$ then H_Δ is a stopping time. For a much more detailed discussion of fundamental properties and notions related to filtrations and stopping times we refer the reader to Dellacherie and Meyer [DM75, Chp IV].

2.2 Small stopping times

For number of reasons we are interested in stopping times which are "small". For example, the original motivation for considering embedding problems was to iterate the construction and obtain a representation of a random walk as Brownian motion stopped at a sequence of stopping times. In financial applications it is important that the stopped process on $[0,\tau]$ be a martingale. If we are interested only in the law of X_τ and if we have $\rho \leq \tau$ with $X_\rho \sim X_\tau$ then ρ seems better than τ. As an example, consider the first hitting time of one $H_{\{1\}}(X)$. We could also consider $\inf\{t \geq 1 : X_t = 1\}$ but this seems rather artificial. Likewise, if ρ is a stopping time we could apply it to paths shifted by $T = \inf\{t \geq 1 : X_t = X_0\}$ but for the purposes considered in these notes such a modification would not be desirable. We are interested in stopping times which are "naturally small" and in this section we explores ways of making this mathematically precise.

Definition 2.2.1. *We say that a stopping time τ is* minimal *for X if for any other stopping time $\rho \leq \tau$, $X_\rho \sim X_\tau$ implies $\rho = \tau$ a.s.*
We say that a stopping time τ is uniformly integrable (UI) *for X if the stopped process X^τ is uniformly integrable.*

Note that both notions are relative to a process X and τ may be minimal for one process but not for another. For example, if $X_t = X_0$ is constant then only $\tau \equiv 0$ is minimal while all stopping times are UI. However when X is fixed we simply say that τ is minimal or is UI without mentioning X explicitly.

Proposition 2.2.2 (Monroe [Mon72a])**.** *Let τ be a stopping time and X a continuous martingale with X_0 a constant. We assume that paths of X do not have intervals of constancy a.s.. If $\mathcal{L}(X_\tau)$ has finite first moment and $\mathbb{E}\, X_\tau = X_0$ then τ is minimal if and only if τ is UI. If further $\mathcal{L}(X_\tau)$ has bounded support with bounds $a < X_0 < b$ then minimality of τ is equivalent to $\tau \leq H_{[a,b]^c}(X)$.*

Proof of Proposition 2.2.2. The original proof goes back to Monroe [Mon72a], who made an extensive use of the theory of barriers. It was then argued in a much simpler way by Chacon and Ghoussoub [CG79].
Suppose first that τ is minimal and $\rho \leq \tau$ with $X_\rho \sim X_\tau$. It follows that

$$\mathbb{E}[X_\tau - x; X_\tau \geq x] = \mathbb{E}[X_\rho - x; X_\rho \geq x] = \mathbb{E}[X_\tau - x; X_\rho \geq x], \quad \forall x \in \mathbb{R},$$

where the second equality follows from $\mathbb{E}[X_\tau | \mathcal{F}_\rho] = X_\rho$. However $\mathbb{E}[X_\tau - x; A] \leq \mathbb{E}[X_\tau - x; X_\tau \geq x]$ for any other A with $\mathbb{P}(A) = \mathbb{P}(X_\tau \geq x)$. It follows that

$$\{X_\tau > x\} \subset \{X_\rho \geq x\} \subset \{X_\tau \geq x\}, \quad \forall x \in \mathbb{R},$$

and hence $X_\tau = X_\rho$ a.s. By the UI of τ, for any stopping time σ, $\rho \leq \sigma \leq \tau$ we have

$$X_\sigma = \mathbb{E}[X_\tau | \mathcal{F}_\sigma] = \mathbb{E}[X_\rho | \mathcal{F}_\sigma] = X_\rho,$$

and therefore X is constant on $[\rho,\tau]$. We conclude that $\tau=\rho$ since X has no intervals of constancy.

□

If X has intervals of constancy which are known, e.g. X is constant for $t\in[1,2]$ and if $\mathbb{P}(\tau=1)>0$ then we can change τ on this event to take value 1.5 without affecting UI but destroying minimality. We note however that this is a rather artificial example. Naturally, if $\mathcal{L}(X_\tau)$ does not have a finite first moment or the means of X_0 and X_τ do not match then τ can not be UI so the notion of minimality is more general than that of UI. Nevertheless, in the cases which are most often studied the two notions coincide. The following proposition gives some tools to verify if a stopping time is minimal or UI. Equivalent conditions for minimality when $\mathcal{L}(X_\tau)$ has a different mean than X_0 were described in Cox and Hobson [CH06].

Proposition 2.2.3 (Azéma, Gundy and Yor [AGY80], Takaoka [Tak99])**.** *Let X be continuous local martingale with X_0 a constant. We assume X is not identically equal to X_0 and consider a stopping time τ such that $\mathcal{L}(X_\tau)$ has finite first moment. Then*

- *If $\mathbb{E}\,X_\tau=X_0$ and $\mathcal{L}(X_\tau)$ has bounded support with bounds $a<X_0<b$ then UI of τ is equivalent to $\tau\le H_{[a,b]^c}(X)$.*
- *UI of τ is equivalent to $\lim_{\lambda\to\infty}\lambda\,\mathbb{P}(\sup_{t\ge 0}|X_{t\wedge\tau}|>\lambda)=0$.*
- *UI of τ is equivalent to $\lim_{\lambda\to\infty}\lambda\,\mathbb{P}(\langle X\rangle_\tau^{1/2}>\lambda)=0$.*
- *In particular, if $\mathbb{E}\sup_{t\ge 0}|X_t^\tau|<\infty$ or if $\mathbb{E}\,\langle X\rangle_\tau^{1/2}<\infty$ then τ is minimal.*

We note that it is easy to show that $\mathbb{E}\,\langle X\rangle_\tau<\infty$ implies minimality of τ. Indeed, localising, applying the optional stopping theorem to $X^2-\langle X\rangle$, and taking limits, we see that then $\mathbb{E}\,\langle X\rangle_\tau=\mathbb{E}\,X_\tau^2$. It follows that if $\rho\le\tau$ embeds the same measure then $\langle X\rangle_\rho\le\langle X\rangle_\tau$ but they have the same expectation and hence are equal a.s. However for a continuous local martingale the intervals of constancy of X and of $\langle X\rangle$ coincide so if X has not intervals of constancy then $\rho=\tau$ a.s. and τ is minimal and hence also UI by Proposition 2.2.2 above. We can actually prove a stronger result (see Root [Roo69] and Sawyer [Saw74]):

Proposition 2.2.4. *Let τ be a stopping time such that $(X_{t\wedge\tau})_{t\ge 0}$ is a uniformly integrable martingale. Then there exist universal constants c_p, C_p such that*

$$c_p\,\mathbb{E}\left[\langle X\rangle_\tau^{p/2}\right]\le\mathbb{E}\left[|X_\tau|^p\right]\le C_p\,\mathbb{E}\left[\langle X\rangle_\tau^{p/2}\right]\quad \textit{for } p>1. \tag{2.3}$$

Proof. The Burkholder-Davis-Gundy inequalities (see Revuz and Yor [RY99] p. 160) guarantee existence of universal constants k_p and K_p such that $k_p\,\mathbb{E}\,\langle X\rangle_\tau^{p/2}\le$

$\mathbb{E}[(\sup_{u\le\tau}|X_u|)^p] < K_p\,\mathbb{E}\,\langle X\rangle_\tau^{p/2}$, for any $p>0$. As $(X_{t\wedge\tau}: t\ge 0)$ is a uniformly integrable martingale we have $\sup_t \mathbb{E}\,|X_{t\wedge\tau}|^p = \mathbb{E}\,|X_\tau|^p$, and Doob's L^p inequalities yield $\mathbb{E}\,|X_\tau|^p \le \mathbb{E}[(\sup_{u\le\tau}|X_u|)^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\,|X_\tau|^p$ for any $p>1$. The proof is thus completed taking $c_p = k_p\left(\frac{p}{p-1}\right)^p$ and $C_p = K_p$. □

We note that the above Proposition is not true for $p=1$. Indeed we can build a stopping time τ such that $(W_{t\wedge\tau}: t\ge 0)$ is a uniformly integrable martingale, $\mathbb{E}\,|W_\tau|<\infty$, and $\mathbb{E}\sqrt{\tau}=\infty$ (see Exercise II.3.15 in Revuz and Yor [RY99]).

Finally, we mention one more notion. A stopping time τ is called **standard** for X if there exists a sequence of bounded stopping times τ_n, with $\lim\tau_n=\tau$ a.s. and $\lim U\mathcal{L}(X_{\tau_n})(x) = U\mathcal{L}(X_\tau)(x) > -\infty$ for all $x\in\mathbb{R}$ (cf. Chacon [Cha77a], Chacon and Ghoussoub [CG79], Falkner [Fal80]). However this is yet equivalent to minimality in the interesting case covered by Proposition 2.2.2 and in general minimality appears to us as the most natural property to consider.

2.3 Changes of time

We start with deterministic considerations following [RY01, Sec 0.4]. Let A_t be a non–decreasing, non-negative, possibly infinite, right–continuous function on $[0,\infty)$. We write A_{t-} for its left-limit $\lim_{u\nearrow t} A_u$. Define

$$C_s := \inf\{t\ge 0 : A_t > s\},\quad s\ge 0,\quad \text{with } \inf\{\emptyset\}=\infty. \tag{2.4}$$

We have the following result

Lemma 2.3.1. *The function C is non-decreasing and right-continuous. We have*

$$C_{s-} = \inf\{t\ge 0 : A_t \ge s\},\quad A_t = \inf\{s\ge 0 : C_s > t\},\quad A_{t-} = \inf\{s\ge 0 : C_s\ge t\}.$$

A has a constant stretch at level s if and only if C jumps in s and then C_{s-} and C_s correspond to the endpoints of the interval on which A is equal to s. In particular, C is continuous if and only if A is strictly increasing.
We have $A(C_s)\ge s$ with strict inequality if and only if s belongs to the interior of an interval of constancy of C.

It is clear that A and C play exactly symmetrical roles, in particular constant stretches of C correspond to jumps of A. The above properties are easy to established and are best seen with a drawing.

Consider a continuous non-decreasing function $T(u)$ on $[a,b]$. The function $A_{T(u)}$ is again a right-continuous non-decreasing function which thus induces a measure which can be used for integration. We have

$$\int_{[a,b]} f(T(u))\mathrm{d}A_{T(u)} = \int_{[T(a),T(b)]} f(t)\mathrm{d}A_t, \quad \forall f \geq 0. \tag{2.5}$$

Applying this with $T(s) = C_s$ and letting $b \nearrow A_\infty$ yields a change of variables formula for Stieltjes integrals:

$$\int_{[0,\infty)} f(t)\mathrm{d}A_t = \int_0^\infty f(C_s)\mathbf{1}_{C_s<\infty}\mathrm{d}s, \quad \forall f \geq 0. \tag{2.6}$$

With this preparations, let us go back to the stochastic setup. From this point onwards we assume the usual hypothesis and that all processes considered have càdlàg paths.

Definition 2.3.2. *A family of stopping times $(\tau_t : t \geq 0)$ is called a* time-change *is the maps $t \to \tau_t$ are increasing and right-continuous a.s.*

A time-change defines $\hat{\mathcal{F}}_s := \mathcal{F}_{\tau_s}$ an increasing sequence of σ-fields, i.e. a filtration. If (X_t) is a processes adapted to the original filtration then $\hat{X}$, $\hat{X}_s := X_{\tau_s}$, is adapted to $(\hat{\mathcal{F}}_s)$. It is called the *time-changed process of* X. We propose to study now in some more detail examples of time-changes, the properties of Y and how to recover X from Y. The exposition follows [RY01, Sec V.1].

Consider an adapted, increasing, right–continuous process A with $A_0 = 0$.

Lemma 2.3.3. *The family (C_s) defined in (2.4) is an increasing right-continuous family of $(\mathcal{F}_t)$ stopping times. Moreover, for any $t \geq 0$, A_t is an $(\mathcal{F}_{C_s})$ stopping time.*

Proof. By (2.4), $C_s = H_{(s,\infty)}(A)$ which is a stopping time since the filtration is right continuous, A is right continuous and (s,∞) is an open set. By Lemma 2.3.1, an analogous reasoning holds with A and C exchanged. □

Given a process X, we can time-change X using C: $\hat{X}_s = X_{C_s}$. Note that C_s and hence $\hat{X}_s$ are defined for $s < A_\infty$. Observe also that if $C_s = \tau \wedge s$, for some stopping time τ, then $\hat{X}_s = X^\tau_s$ is the stopped process.

The fundamental property for our considerations is that the class of semimartingales is stable under time-changing. Note that this is not true for martingales or local martingales. Consider for a example a one dimensional Brownian motion (W_t) and its supremum process $\overline{W}_t = \sup_{u\leq t} W_u$. Then taking $A_t = \overline{W}_t$ we have $\hat{W}_s = W_{C_s} = A_{C_s} = s$. This motivates the following notion

Definition 2.3.4. *Let (C_s) be a time-change. We say that a process X is C-continuous if X is constant on any interval of time of the form $[C_{s-}, C_s]$.*

If X is C-continuous then the time-change and stochastic integration (and local martingale property) commute. We write $H \cdot X$ for the stochastic integral process $H \cdot X_t = \int_0^t H_u \mathrm{d}X_u$. For these notes we limit ourselves to the continuous case. We also assume $A_\infty = \infty$ so that C_s is finite for all $s \geq 0$. The results may be generalised to the case $A_\infty < \infty$ as long as the objects make sense. We will come back to this when considering an important example.

Proposition 2.3.5. *Let A, C be as above and X a continuous semimartingale which is C-continuous. Assume further that $A_\infty = \infty$. We have $\langle \hat{X} \rangle = \widehat{\langle X \rangle}$ and if X is an $(\mathcal{F}_t)$-local martingale then $(\hat{X}_s)$ is an $(\hat{F}_s)$-continuous local martingale.*
If H is an $(\mathcal{F}_t)$-progressively measurable process then $\hat{H}$ is $(\hat{F}_s)$-progressively measurable. Moreover if $\int_0^t H_u^2 \mathrm{d} \langle X \rangle_u < \infty$ a.s. for all t then $\int_0^s \hat{H}_v \mathrm{d}\langle \hat{X} \rangle_v < \infty$ a.s. for all s. In this case we have $\hat{H} \cdot \hat{X} = \widehat{H \cdot X}$.

In particular a suitably time-changed Brownian motion is a local martingale. Crucially, it turns out that the converse is true.

Theorem 2.3.6 (Dambis, Dubins, Schwarz)**.** *Let (X_t) be an $(\mathcal{F}_t)$–continuous local martingale vanishing at 0 with $\langle X \rangle_\infty = \infty$ a.s. If we set*

$$C_s := \inf\{t \geq 0 : \langle X \rangle_t > s\},$$

then $W_s := \hat{X}_s = X_{C_s}$ is an $(\mathcal{F}_{C_s})$ standard Brownian motion. Furthermore, $X_t = W_{\langle X \rangle_t}$.

Proof. First note that X is C-continuous because intervals of constancy of X are equal to the intervals of constancy of $\langle X \rangle$. Proposition 2.3.5 shows that W is a continuous local martingale with $\langle W \rangle_s = \langle X \rangle_{C_s} = s$, $s \geq 0$, and Lévy's charaterisation of Brownian motion implies that W is a Brownian motion. Finally, $W_{\langle X \rangle_t} = X_{C_{\langle X \rangle_t}}$ and although we might have $C_{\langle X \rangle_t} > t$ we have X_u being constant for $u \in [t, C_{\langle X \rangle_t}]$. □

A multi-dimensional analogue of the above result is due to Knight. It allows to represent a d-dimensional continuous local martingale with zero quadratic variation between its components, as a time changed d-dimensional Brownian motion. Further, an important refinement states that when $\langle X \rangle_\infty < \infty$ then, possibly on an extended probability space, we can define a Brownian motion (B_s) such that $W_s = B_{s \wedge \langle M \rangle_\infty}$. This follows since for a continuous local martingale X the sets $\{\langle X \rangle_\infty < \infty\}$ and $\{\lim_{t \to \infty} X_t \text{ exists}\}$ are equal a.s. In particular, if $\langle X \rangle_\infty < \infty$ then $W_{\langle X \rangle_\infty} = X_\infty$ is well defined and W is continuous at $\langle X \rangle_\infty$. Proposition 2.3.5 shows that W is a martingale when stopped at a bounded stopping time and we conclude easily that it is a continuous local martingale. Brownian motion B is obtained by "glueing" an independent Brownian motion at $W_{\langle M \rangle_\infty}$.

Finally, let us comment on measurability of X wrt to W and vice-versa. It follows from the construction that W is measurable w.r.t. $\mathcal{F}^X_\infty$, where $(\mathcal{F}^X_t)$ is the natural filtration of X taken right-continuous and complete. The converse is not necessarily true, i.e. X need not be measurable w.r.t. to $\mathcal{F}^W_\infty$. However it is true if X is a solution to a Stochastic Differential Equation (SDE).

In fact time-change ideas arise naturally when solving SDEs. These ideas were essential for the work of Wolfgang Doeblin, see [BY02], and offer an alternative perspective to the Itô-McKean approach. Suppose we are interested in a process X solving

$$X_t = \int_0^t \sigma(X_u)\mathrm{d}B_u,$$

where B is a Brownian motion and σ is non-zero function. Dambis-Dubins-Schwarz theorem shows that $W_s := X_{C_s}$ is a Brownian motion. Further, since $\sigma(X_u)^2 > 0$, $\langle X\rangle_t$ is continuous and strictly increasing and we have

$$s = \langle X\rangle_{C_s} = \int_0^{C_s} \sigma^2(X_u)\mathrm{d}u = \int_0^s \sigma^2(X_{C_v})\mathrm{d}C_v = \int_0^s \sigma^2(W_v)\mathrm{d}C_v, \quad s \geq 0.$$

It follows that $C_s = \int_0^s \sigma(W_v)^{-2}\mathrm{d}v$. If we take this as the definition of C_s and define A_t to be its inverse. We have $A_t = \langle X\rangle_t$ and $X_t = W_{A_t}$. This provides a constructive strong solution of our SDE if we can find W. However starting with a Brownian motion W, we can given a pathwise construction of a weak solution to the original SDE.

A time-changed Brownian motion is always a semimartingale. We close this section with a converse result of Monroe [Mon78].

Theorem 2.3.7. *Any real-valued semimartingale X may be represented, on a suitably enlarged probability space, as a time change of Brownian motion.*

In many ways time-changing has the same *good* properties as changing measure. Both preserve the semimartingale character and both preserve stochastic integrals. However time-change is able to perturb X to a much greater degree than (an equivalent) change of measure. The latter will add a drift to a continuous local martingale. The former can change both the drift and the volatility. If we consider time-change C such that X is C-continuous then $\hat{X}$ is still a local martingale but we are free to affect its quadratic variation process. In particular we can turn it into a Brownian motion. If we abandon the requirement of X being C-continuous then we can also add a drift to X. In return, change of time is in general harder to design and understand than a change of measure whose consequences are rather well understood through the Girsanov theorem.

Chapter 3

Elements of potential theory[1]

(mainly for real valued processes)

Potential theory has played a crucial role in a number of works about the Skorokhod embedding problem and we will also make a substantial use of it, especially in its one-dimensional context. We introduce here some potential theoretic objects which will be important for us. As this is not an introduction to Markovian theory we will omit certain details and assumptions. We refer to the classical work of Blumenthal and Getoor [BG68] for precise statements.

3.1 General theory

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual conditions consider a strong right-continuous Markov process $X = (X_t : t \geq 0)$ taking values in a locally compact space $(E, \mathcal{E})$ with a denumerable basis and a Borel σ-field. Associated with X is $(P_t^X : t \geq 0)$, a standard semigroup of submarkovian kernels. A natural interpretation is that $P_t^X \nu$ represents the law of X_t under the starting distribution $X_0 \sim \nu$. Define the potential kernel $\mathcal{U}^X$ through $\mathcal{U}^X = \int_0^\infty P_t^X dt$. This can be seen as a linear operator on the space of measures on E. The intuitive meaning is that $\mathcal{U}^X \nu$ represents the occupation measure for X along its trajectories, where $X_0 \sim \nu$.

If the potential operator is finite[2] it is not hard to see that for two bounded stopping times, $\rho \leq \tau$, we have $\mathcal{U}^X(P_\rho^X - P_\tau^X) \geq 0$. This explains how the potential can be used to keep track of the relative stage of the development of the process (see Chacon [Cha77a]). In fact, Rost established that it gives a necessary and sufficient condition for existence (of any and of minimal) embedding. We

[1] Large parts of this chapter are taken from Obłój [Obł04].

[2] That is, for any finite starting measure, it provides a finite measure.

state it here for a transient Markov process on a compact space since then things are (a little) easier. Transience of X is encoded in the requirement that the potential $\mathcal{U}^X \nu$ is σ-finite. The theorem remains true when E is not compact but the potentials are σ-finite.

Theorem 3.1.1 (Rost [Ros76a]). *Assume that E is compact and that $\mathcal{F}$ supports an atomfree random variable independent of X. Let ν be a probability measure with a σ–finite potential $\mathcal{U}^X \nu$. For a measure μ on E there exists a minimal stopping time τ such that $X_\tau \sim \mu$ if and only if $\mathcal{U}^X \mu \leq \mathcal{U}^X \nu$.*

3.2 One-dimensional potential theory

For $X = B$, a real-valued Brownian motion, as the process is recurrent, the potential $\mathcal{U}^B \nu$ is infinite for positive measures ν. However, if the measure ν is a signed measure with $\nu(\mathbb{R}) = 0$, and $\int |x|\nu(dx) < \infty$, then the potential $\mathcal{U}^B \nu$ is not only finite but also absolutely continuous with respect to Lebesgue measure. In order to compute it let us first recall that

$$\int_0^\infty \frac{1}{\sqrt{2\pi t}} \left(\mathrm{e}^{-\frac{\xi^2}{2t}} - 1 \right) \mathrm{d}t = -|\xi|, \quad \xi \in \mathbb{R}\,.$$

Write $\nu = \nu_+ - \nu_-$ for two positive measures ν_-, ν_+ with $\nu_+(\mathbb{R}) = \nu_-(\mathbb{R})$. Then, using Fubini-Tonelli theorem, we obtain

$$\begin{aligned}
\frac{\mathcal{U}^B \nu(\mathrm{d}x)}{\mathrm{d}x} &= \int_0^\infty \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-\frac{(y-x)^2}{2t}} \nu(\mathrm{d}y)\mathrm{d}t \\
&= \int_0^\infty \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \left(\mathrm{e}^{-\frac{(y-x)^2}{2t}} - 1 \right) \nu_+(\mathrm{d}y)\mathrm{d}t \\
&\quad - \int_0^\infty \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \left(\mathrm{e}^{-\frac{(y-x)^2}{2t}} - 1 \right) \nu_-(\mathrm{d}y)\mathrm{d}t \\
&= -\int |x-y|\nu(\mathrm{d}y).
\end{aligned} \tag{3.1}$$

The right hand side is well defined for any probability measure μ on $\mathbb{R}$ (instead of ν) with $\int |x|\mu(\mathrm{d}x) < \infty$ and, with a certain abuse of terminology, this quantity is called the one-dimensional potential of the measure μ:

Definition 3.2.1. *Denote by $\mathcal{M}^1$ the set of probability measures on $\mathbb{R}$ with finite first moment, $\mu \in \mathcal{M}$ iff $\int |x|\mu(\mathrm{d}x) < \infty$. Let $\mathcal{M}^1_m$ denote the subset of measures with expectation equal to m, $\mu \in \mathcal{M}^1_m$ iff $\int |x|\mu(\mathrm{d}x) < \infty$ and $\int x\mu(\mathrm{d}x) = m$. Naturally $\mathcal{M}^1 = \bigcup_{m\in\mathbb{R}} \mathcal{M}^1_m$. The one-dimensional potential operator U acting from $\mathcal{M}^1$ into the space of continuous, non-positive functions, $U : \mathcal{M}^1 \to C(\mathbb{R}, \mathbb{R}_-)$, is defined through $U\mu(x) := U\Big(\mu\Big)(x) = -\int_{\mathbb{R}} |x-y|\mu(\mathrm{d}y)$. We refer to $U\mu$ as to **the potential of** μ.*

We adopt the notation $U\mu$ to differentiate this case from the general case of a potential kernel $\mathcal{U}^X$ when $\mathcal{U}^X\,\mu$ is a measure and not a function. This simple operator enjoys some remarkable properties, which will be crucial for the Chacon-Walsh methodology (see Section 4.2), which in turn is our main tool in these notes. The following can be found in Chacon [Cha77a] and Chacon and Walsh [CW76a]:

Proposition 3.2.2. *Let $m \in \mathbb{R}$ and $\mu \in \mathcal{M}^1_m$. Then*

(i) *$U\mu$ is concave and Lipschitz-continuous with parameter 1, $U\mu'(x-) = 2\mu([x,\infty)) - 1$, $U\mu'(x+) = 2\mu((x,\infty)) - 1$ and $U\mu''(\mathrm{d}x) = -2\mu(\mathrm{d}x)$. In particular $U\mu$ has a kink at x if and only if $\mu(\{x\}) > 0$ and $U\mu$ is linear on $[a,b]$ if and only if $\mu((a,b)) = 0$;*

(ii) *$U\mu(x) \le U\delta_{\{m\}}(x) = -|x-m|$ and if $\nu \in \mathcal{M}^1$ and $U\nu \le U\mu$ then $\nu \in \mathcal{M}^1_m$;*

(iii) *for $\mu_1, \mu_2 \in \mathcal{M}^1_m$ $\lim_{|x|\to\infty} |U\mu_1(x) - U\mu_2(x)| = 0$, in particular $|x-m| + U\mu(x) \nearrow 0$ as either $x \to \infty$ or $x \to -\infty$;*

(iv) *for $\mu_n \in \mathcal{M}^1_m$, $U\mu_n(x) \xrightarrow[n\to\infty]{} U\mu(x)$ for all $x \in \mathbb{R}$ if and only if $\mu_n \Rightarrow \mu$ and $U\mu_n(x_0) \xrightarrow[n\to\infty]{} U\mu(x_0)$ for some $x_0 \in \mathbb{R}$.*
If $X_n \sim \mu_n$, $X \sim \mu$ are random variables $X_n \to X$ a.s. and $U\mu_n \to U\mu$ pointwise then $X_n \to X$ in L^1;

(v) *for $\nu \in \mathcal{M}^1_0$, if $\int_{\mathbb{R}} x^2\nu(dx) < \infty$, then $\int_{\mathbb{R}} x^2\nu(\mathrm{d}x) = \int_{\mathbb{R}} \Big||x| + U\nu(x)\Big| dx$;*

(vi) *for $\nu \in \mathcal{M}^1_m$, $U\nu|_{[b,\infty)} = U\mu|_{[b,\infty)}$ if and only if $\mu|_{(b,\infty)} \equiv \nu|_{(b,\infty)}$, likewise $U\nu|_{(-\infty,a])} = U\mu|_{(-\infty,a]}$ if and only if $\mu|_{(-\infty,a)} \equiv \nu|_{(-\infty,a)}$.*

Proof. For $x, z \in \mathbb{R}$ and $\lambda \in (0,1)$ we have

$$\begin{aligned}\lambda U\mu(x) + (1-\lambda)U\mu(z) &= -\int \left(\lambda|x-y| + (1-\lambda)|z-y|\right)\mu(\mathrm{d}y)\\ &\le -\int |\lambda x + (1-\lambda)z - y|\mu(\mathrm{d}y) = U\mu(\lambda x + (1-\lambda z)),\end{aligned}$$

which shows that $U\mu$ is concave and hence continuous. We rewrite the potential as:

$$\begin{aligned} U\mu(x) &= -\int_{\mathbb{R}} |x-y|d\mu(y) = -\int_{(-\infty,x)} (x-y)d\mu(y) - \int_{[x,\infty)} (y-x)d\mu(y)\\ &= x\Big(2\mu([x,\infty)) - 1\Big) + m - 2\int_{[x,\infty)} yd\mu(y), \end{aligned} \tag{3.2}$$

where, to obtain the third equality, we use the fact that $\mu \in \mathcal{M}^1_m$ and so $\int_{[x,\infty)} yd\mu(y) = m - \int_{(-\infty,x)} yd\mu(y)$. The above immediately shows that the left- and right- derivatives of $U\mu$ exist everywhere and are as in *(i)*. This gives us another proof of concavity of $U\mu$ and also shows that $U\mu''(\mathrm{d}x) = -2\mu(\mathrm{d}x)$. We also see that $U\mu'(x) \in [-1,1]$ and in particular $U\mu$ is 1–Lipschitz continuous. Other assertions in (i) are clear.

The first assertion in *(ii)* follows from Jensen's inequality as

$$U\mu(x) = -\int_{-\infty}^{\infty} |x-y| d\mu(y) \leq -\Big| \int_{-\infty}^{\infty} (x-y) d\mu(y) \Big| = -|x-m| = U\delta_m(x).$$

Further, for $\nu \in \mathcal{M}^1$ with $U\nu \leq U\mu$ we have $U\nu(x) \leq -|x-m|$. Using (3.2) above and letting $x \to -\infty$ we see that $\int_{\mathbb{R}} x d\nu(x) \geq m$ and, analogously, letting $x \to \infty$ we see that the reverse holds.

The formula displayed in (3.2) shows that the potential is linear on intervals $[a,b]$ such that $\mu((a,b)) = 0$. Furthermore, it shows that for $\nu \in \mathcal{M}^1_m$, $U\nu|_{[b,\infty)} = U\mu|_{[b,\infty)}$ if and only if $\mu|_{(b,\infty)} \equiv \nu|_{(b,\infty)}$. Note that the same is true with $[b,\infty)$ replaced by $(-\infty, a]$ and that in particular $U\nu \equiv U\mu$ if and only if $\nu \equiv \mu$. Points *(i),(ii),(vi)* are now established.

From (3.2) we have

$$U\mu_1(x) - U\mu_2(x) = 2x\Big(\mu_1([x,\infty)) - \mu_2([x,\infty))\Big) - 2\int_{[x,\infty)} y(\mu_1(\mathrm{d}y) - \mu_2(\mathrm{d}y))$$

and *(iii)* follows since μ_1, μ_2 have the same (well defined) mean.

Finally, *(v)* follows by integration by parts and it remains to argue *(iv)*. In fact, we will show a slightly stronger statement than in *(iv)*. Consider $\mu_n \Rightarrow \mu$, where μ is some non-negative measure on $\mathbb{R}$, and suppose $U\mu_n(x_0)$ converge. Take g continuous bounded with $-|x_0 - y| \leq g(y) \leq 0$. Then $U\mu_n(x_0) \leq \int g(y)\mu_n(\mathrm{d}y) \leq 0$ and Fatou Lemma gives that

$$\limsup_{n\to\infty} U\mu_n(x_0) \leq \int g(y)\mu(\mathrm{d}y).$$

Letting $g \searrow -|x_0 - y|$ we conclude that $U\mu(x) \geq \lim_{n\to\infty} U\mu_n(x_0)$ and in particular is finite. Finally note that by (3.2) for $x < x_0$ we have

$$U\mu_n(x) - U\mu_n(x_0) = (x - x_0)(2\overline{\mu}_n(x) - 1) - 2\int_{[x,x_0)} (y - x_0)\mu_n(\mathrm{d}y)$$

and we see that the right hand side converges for a.e. x to $U\mu(x) - U\mu(x_0)$. A similar argument holds for $x > x_0$ and we obtain that $U\mu_n$ converge pointwise and that $U\mu = \lim U\mu_n - K$ for positive constant $K = \lim U\mu_n(x_0) - U\mu(x_0)$. In particular, by above, if $\mu \in \mathcal{M}^1$ then $\mu \in \mathcal{M}^1_m$. Likewise if $U\mu_n \geq U\nu$ for some $\nu \in \mathcal{M}^1_m$ then also $U\mu \geq U\nu$ and hence $\mu \in \mathcal{M}^1_m$.

Conversely, suppose $U\mu_n(x) \to U\mu(x)$ for all $x \in \mathbb{R}$ and some $\mu_n, \mu \in \mathcal{M}^1_m$. Then, writing a C^2 function f with compact support as $f(y) = \frac{1}{2}\int |y -$

$x|f''(x)\mathrm{d}x$ we obtain

$$\int f(y)\mu_n(\mathrm{d}y) = \frac{1}{2}\int\int |x-y|\mu_n(\mathrm{d}y)f''(x)\mathrm{d}x = \frac{1}{2}\int U\mu_n(x)f''(x)\mathrm{d}x.$$

As $U\mu_n$ converge to $U\mu$ uniformly on compacts, the right hand side converges to $\frac{1}{2}\int U\mu(x)f''(x)\mathrm{d}x = \int f(y)\mu(\mathrm{d}y)$. Weak convergence of measures follows. For the final statement note that $\mathbb{E}|X_n| = -U\mu_n(0) \to -U\mu(0) = \mathbb{E}|X|$ and the result follows from Scheffe's lemma. □

Remark 3.2.3. *Observe that in (iv) in Proposition 3.2.2 it is not true*[3] *that* $U\mu_n(x) \xrightarrow[\mathrm{n}\to\infty]{} U\mu(x)$ *for all* $x \in \mathbb{R}$ *if and only if* $\mu_n \Rightarrow \mu$. *Consider for example* $\mu_n = \frac{1}{n}(\delta_{\{-n^2\}} + \delta_{\{n^2\}}) + \frac{n-2}{n}\delta_{\{0\}}$ *so that* $\mu_n \in \mathcal{M}_0^1$ *and* $\mu_n \Rightarrow \delta_{\{0\}}$. *We see however that* $U\mu_n(x) \to -\infty$ *for all* $x \in \mathbb{R}$.

In financial applications it is often natural to consider

$$C\mu(x) := \int_x^\infty (y-x)\mu(\mathrm{d}y) = \int_{\mathbb{R}} (y-x)^+\mu(\mathrm{d}y), \tag{3.3}$$

which we refer to as *call prices implied by* μ, see also (1.5) above. A direct computation shows that

$$U\mu(x) = -2C\mu(x) - x. \tag{3.4}$$

Similarly to the general theory, both the potential and the call function allow us to keep track of ordering of stopping times.

Lemma 3.2.4. *If* $\mu, \nu \in \mathcal{M}^1$ *then* $U\mu \le U\nu$ *is equivalent to* $\int f(x)\mu(\mathrm{d}x) \ge \int f(x)\nu(\mathrm{d}x)$ *for any positive convex function.*
If X *is a local martingale and* $\rho \le \tau < \infty$ *a.s. are two UI stopping times then* $U\mathcal{L}(X_\tau) \le U\mathcal{L}(X_\rho)$ *with a reversed inequality for the call functions.*
Proposition 2.2.2 holds if X_0 *is not a constant but* $\mathcal{L}(X_0) \in \mathcal{M}^1$ *and* $U\mathcal{L}(X_0) \ge U\mathcal{L}(X_\tau)$.

Proof. The first statement follows by (3.4) and by writing a positive convex function as a $f(x) = f(0) + \int_0^x (x-a)^+ f''(\mathrm{d}a)$. The second statement is an immediate consequence of Jensens inequality. The final assertion follows by inspection of the proof of Proposition 2.2.2. □

When τ is a first exit time after ρ then we can provide a precise description of $U\mathcal{L}(X_\tau)$. This is the crucial link which we will exploit in Section 4.2 to build stopping times embedding a given distribution.

[3]Contrary to what is stated in [Obł04]. We became aware of this while reading a recent paper of Hobson and Klimmek [HK13].

Proposition 3.2.5. *Let X be a continuous local martingale, ρ an a.s. finite stopping time and let $\nu = \mathcal{L}(X_\rho) \in \mathcal{M}^1$. Consider two numbers $a < b$, let $\tau := \inf\{t \geq \rho : X_t \notin (a,b)\}$ and assume $\tau < \infty$ a.s. A sufficient condition for this is that $\langle X \rangle_\infty = \infty$ a.s. Write $\mu = \mathcal{L}(X_\tau)$. Then $U\nu|_{(-\infty,a]\cup[b,\infty)} = U\mu|_{(-\infty,a]\cup[b,\infty)}$ and $U\mu$ is linear on $[a,b]$.*

Proof. The assertions follows from the fact that $X_\tau = X_\rho$ on $\{X_\rho \notin [a,b]\} = \{X_\tau \notin [a,b]\}$ and hence $\mu = \nu$ on $\mathbb{R} \setminus [a,b]$ and the fact that μ does not charge the interval (a,b). □

To illustrate some of the results established above let us compute the atom of μ in a in two ways. First of all, conditionally on $X_\rho \in [a,b]$, the process on $[\rho,\tau]$ is bounded and hence a martingale. It follows that

$$\mu(\{a\}) = \nu(\{a\}) + \int_{(a,b)} \frac{b-x}{b-a}\nu(\mathrm{d}x).$$

On the other hand from *(i)* in Proposition 3.2.2 and relation between $U\nu$ and $U\mu$ in Proposition 3.2.5 we obtain

$$\begin{aligned}
\mu(\{a\}) &= \frac{1}{2}(U\mu'(a-) - U\mu'(a+)) = \frac{1}{2}U\nu(a-) - \frac{1}{2}\frac{U\nu(b)-U\nu(a)}{b-a} \\
&= \nu([a,\infty)) - \frac{1}{2} - \frac{1}{2}\frac{2b\nu([b,\infty)) - 2a\nu([a,\infty)) + a - b + 2\int_{[a,b)} y\nu(\mathrm{d}y)}{b-a} \\
&= \nu([a,\infty)) + \frac{1}{2}\frac{2b\nu([a,b)) + 2(a-b)\nu([a,\infty)) - 2\int_{[a,b)} y\nu(\mathrm{d}y)}{b-a} \\
&= \nu(\{a\}) + \int_{(a,b)} \frac{b-x}{b-a}\nu(\mathrm{d}x),
\end{aligned}$$

as required.

Chapter 4

The Skorokhod embedding problem

4.1 Introduction

The so called *Skorokhod embedding problem* or *Skorokhod stopping problem* was first formulated and solved by Skorokhod in 1961 [Sko61] (English translation in 1965 [Sko65]). For a given centered probability measure μ with finite second moment and a Brownian motion B, one looks for an integrable stopping time T such that the distribution of B_T is μ. This original formulation has been changed, generalised or narrowed a great number of times. The problem has stimulated research in probability theory for over 50 years now. Here we will focus only on some aspects and we refer to Obłój [Obł04] for a survey. An excellent account of the domain with emphasis on the applications in mathematical finance is given in Hobson [Hob11].

Let us start with some history of the problem. Skorokhod's solution required a randomization external to B. Three years later another solution was proposed by Dubins [Dub68], which did not require any external randomization. Around the same time, a third solution was proposed by Root. It was part of his Ph.D. thesis and was then published in an article [Roo69].

Soon after, another doctoral dissertation was written on the subject by Monroe who developed a new approach using additive functionals. His results were published in 1972 [Mon72b]. Although he did not have any explicit formulae for the stopping times, his ideas proved fruitful as can be seen from the elegant solutions by Vallois [Val83a] and Bertoin and Le Jan [BLJ92a], which also use additive functionals.

The next landmark was set in 1971 with the work of Rost [Ros71b]. He generalised the problem by looking at any Markov process and not just Brownian motion. He gave an elegant necessary and sufficient condition for the existence of a solution to the embedding problem. Rost made extensive use of potential theory. This approach was also used a few years later by Chacon and Walsh [CW76a], who proposed a new solution to the original problem which included Dubins' solution as a special case.

By that time, Skorokhod embedding, also called the Skorokhod representation, had been successfully used to prove various invariance principles for random walks (Sawyer [Saw74]). It was the basic tool used to realize a discrete process as a stopped Brownian motion. This ended however with the Komlós, Major and Tusnády [KMT75] construction, which proved far better for this purpose. Still, the Skorokhod embedding problem continued to inspire researchers and found numerous new applications. In particular, starting with the crucial insight of Hobson [Hob98a], it has been applied to obtain robust pricing and hedging in mathematical finance, see our earlier discussion in Section 1.4.4.

The next development of the theory came in 1979 with a solution proposed by Azéma and Yor [AY79b]. Unlike Rost, they made use of martingale theory, rather than Markov and potential theory, and their solution was formulated for any recurrent, real-valued diffusion. We will see below that their solution can be obtained as a limit case of Chacon and Walsh's solution. Azéma and Yor's solution has interesting properties which are discussed as well. In particular the solution maximizes stochastically the law of the supremum up to the stopping time. This direction was continued by Perkins [Per86], who proposed his own solution in 1985, which in turn minimizes the law of the supremum and maximizes the law of the infimum.

Finally, yet another solution was proposed in 1983 by Bass [Bas83]. He used the stochastic calculus apparatus and defined the stopping time through a time-change procedure. His solution is also reported in Stroock ([Str03], p. 213–217).

Further developments can be classified broadly into two categories: works trying to extend older results or develop new solutions, and works investigating properties of particular solutions. The former category is well represented by papers following Monroe's approach: solution with local times by Vallois (1983) [Val83b] and the paper by Bertoin and Le Jan (1992) [BLJ92b], where they develop explicit formulae for a wide class of Markov processes. Azéma and Yor's construction was taken as a starting point by Grandits and Falkner [GF00] and then by Pedersen and Peskir [PP01] who worked with non-singular diffusions. Roynette, Vallois and Yor [RVY02] used Rost criteria to investigate Brownian motion and its local time [RVY02]. There were also some older works on n-dimensional Brownian motion (Falkner [Fal80]) and right processes (Falkner and Fitzsimmons [FF91]).

The number of works in the second category is greater and we will not try to describe it here. We only mention that the emphasis was placed on the one hand on the solution of Azéma and Yor and its accurate description and, on the other hand, following Perkins' work, on the control of the maximum and the minimum of the stopped martingale (Kertz and Rösler [KR90], Hobson [Hob98b], Cox and Hobson [CH04b]).

We finish this introduction by stating formally the problem we consider. It is not the original formulation (which featured Brownian motion) nor the most general formulation as is clear from the discussion above, in particular from Theorem 3.1.1. Instead we give a formulation which will be most relevant in the sequel.

Problem 4.1.1 (The Skorokhod embedding problem (SEP))**.** *Given a continuous local martingale X and a probability measure μ on $\mathbb{R}$ find a stopping time τ such that $X_\tau \sim \mu$ and τ is minimal.*

We will be mostly interested in the case when X_0 is a constant, $\int_{\mathbb{R}} |x|\mu(\mathrm{d}x) < \infty$ and $\int_{\mathbb{R}} x\mu(\mathrm{d}x) = \mathbb{E}[X_0]$. We will also consider the situation when X_0 is not constant but $U\mathcal{L}(X_0) \geq U\mu$. In both cases, by Proposition 2.2.2 minimality of τ is equivalent to X^τ being uniformly integrable.

The requirement that τ is "small" (made precise by the notion of minimality) is essential. Without it, there is a trivial solution to the above problem which, we believe, was first observed by Doob. Consider $X = B$ a standard Brownian motion and let $\tau = \inf\{t \geq 1 : B_t = \underline{\mu}^{-1}(\underline{\mathcal{N}}(B_1))\}$, where $\underline{\mu}^{-1}$ is the right-continuous inverse of the cumulative distribution function $\underline{\mu}$ of μ. Trivially τ embeds μ in Brownian motion, $B_\tau \sim \mu$. However, unless $\mu = \mathcal{N}(0,1)$, we have $\mathbb{E}\,\tau = \infty$.

In general, there are many solutions to the SEP for a given X and μ. We do not intend to discuss all of them in these notes. Our aim is to showcase main ideas and techniques. We organise the discussion in the following three sections.

4.2 Solutions to SEP via first exit times of an interval

We start by discussing how potential theory can be used to construct solutions to the embedding problem. This methods is due to Chacon and Walsh [CW76b] and Chacon [Cha77b].

Let X be a continuous local martingale with $\langle X\rangle_\infty = \infty$ a.s. For simplicity we assume that $X_0 = 0$ a.s. It will be clear that the problem and methods are invariant under a shift by a constant. Recall that $H_\Delta(X)$ is the first hitting time of a set Δ by X. Since X is fixed we write simply H_Δ. We write $H_{a,b}$ for $H_{(a,b)^c}$ the first exit from the interval (a, b), as in Figure 4.1.

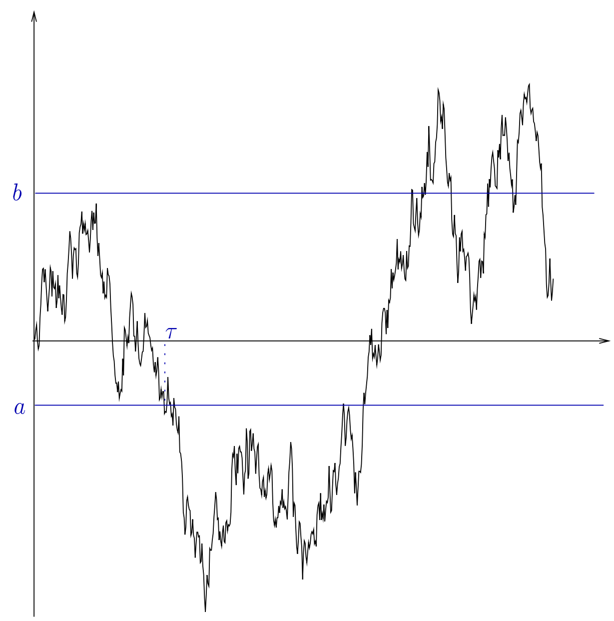


Figure 4.1: The first exit time from an interval (a, b).

Consider $\mu = \frac{1}{2}(\delta_{\{-1\}} + \delta_{\{1\}})$. It follows from Proposition 2.2.2 that the only minimal stopping time embedding μ is $H_{-1,1}$. Likewise, any centred distribution supported on two points: $\mu_{r,s} = \frac{s}{s-r}\delta_{\{r\}} + \frac{-r}{s-r}\delta_{\{s\}}$, $r < 0 < s$, is (uniquely in a minimal way) embedded by the first hitting time $H_{r,s}$. Note that we are using here the fact that a bounded local martingale is a true martingale and hence $\mathbb{E}\, X_{H_{a,b}} = X_0$ and that X has continuous paths so that $X_{H_{a,b}} \in \{a, b\}$.

The following idea appears naturally: express any measure μ as random mix of two-atomic measures $\mu_{r,s}$ and then use the first hitting time $H_{R,S}$, where R, S are independent are represent the mixing. Explicitly, following Breiman [Bre68] and Hall [Hal68], given $\mu \in \mathcal{M}_0^1$, define a distribution on $\mathbb{R}^2$ through

$$\rho_\mu^{Hall}(\mathrm{d}r, \mathrm{d}s) = \frac{(s-r)}{\int_0^\infty x\mu(\mathrm{d}x)} \mathbf{1}_{r \le 0 \le s} \mu(\mathrm{d}r)\mu(\mathrm{d}s). \tag{4.1}$$

Then $H_{R,S}$, where $(R, S) \sim \rho_\mu^{Hall}$ independent of B, embeds μ in B, i.e.

$B_{H_{R,S}} \sim \mu$. Indeed, for $f \geq 0$ we have

$$\begin{aligned}\mathbb{E}[f(X_{H_{R,S}})] &= \mathbb{E}\left[f(S)\frac{-R}{S-R} + f(R)\frac{S}{S-R}\right]\\ &= \int_{-\infty}^{0}\int_{0}^{\infty} -rf(s)\frac{\mu(\mathrm{d}r)\mu(\mathrm{d}s)}{\int_0^\infty x\mu(\mathrm{d}x)} + \int_{-\infty}^{0}\int_{0}^{\infty} sf(r)\frac{\mu(\mathrm{d}r)\mu(\mathrm{d}s)}{\int_0^\infty x\mu(\mathrm{d}x)}\\ &= \int_{-\infty}^{\infty} f(x)\mu(\mathrm{d}x),\end{aligned}$$

where we used $\mu \in \mathcal{M}_0^1$ so that $\int_0^\infty x\mu(\mathrm{d}x) = -\int_{-\infty}^0 x\mu(\mathrm{d}x)$.

We note that the same idea was in fact used in Skorokhod's original work with the difference that in Skorokhod's construction S is a deterministic function of R.

4.2.1 Measures with finite support

Suppose however that the filtration does not allow for a pair of variables independent of X. Can we still construct an embedding? We will start by discussing the case of μ with finite support. In fact, consider first $\mu = \frac{1}{4}(\delta_{\{-1\}} + \delta_{\{1\}}) + \frac{1}{2}\delta_{\{0\}}$. We write $\cdot \circ \theta_\rho$ for a shift by a stopping time ρ, i.e.

$$H_\Delta \circ \theta_\rho = \inf\{t \geq \rho : X_t \in \Delta\}.$$

This is a useful notation in the current setup, we note however that it is different from the standard shift operator used in Markovian context.

Let $\tau = H_{0,1} \circ \theta_{H_{-1,1/3}}$ and note that we have

$$\mathbb{P}(X_{H_{-1,\frac{1}{3}}} = -1) = \frac{1/3}{1/3+1} = \frac{1}{4}, \quad \mathbb{P}(X_\tau = 1 | X_{H_{-1,\frac{1}{3}}} = \frac{1}{3}) = \frac{1}{4}$$

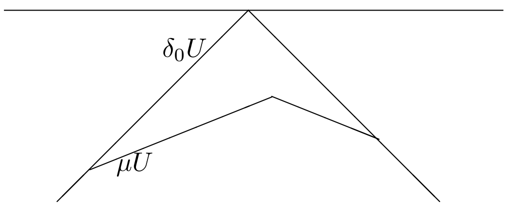
and it follows that $X_\tau \sim \mu$. Using Proposition 3.2.5, let us draw the potentials $U\delta_{\{0\}} \geq U\mathcal{L}(X_{H_{-1,1/3}}) \geq U\mathcal{L}(X_\tau)$:

FIGURE HERE

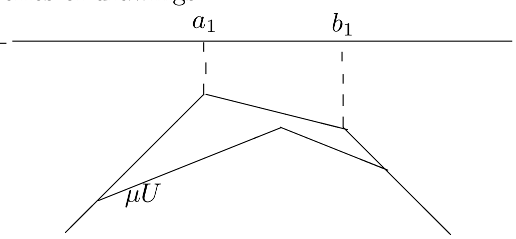
It is clear now why we took $1/3$ in the first place: it is the intersect of $U\delta_{\{0\}}$ with the tangent to $U\mu$ at points $(-1, 0)$. Naturally by symmetry $H_{-1,0}\circ\theta_{H_{-1/3,1}}$ also embeds μ. The general approach should now be clear: we can build a sequence of stopping times, in each step considering a first exit from an interval, which correspond to taking linear sections of the potential. If we design them well, e.g. taking tangents to the target potential, we build an embedding. This way it is not hard to show that any measure μ with support greater than two points can be embedded in an (uncountably) infinite number of minimal ways.

Proposition 4.2.1. *If support of $\mu \in \mathcal{M}_0^1$ is not equal to one or two points then there exists uncountably many different minimal stopping times τ which embed μ: $X_\tau \sim \mu$.*

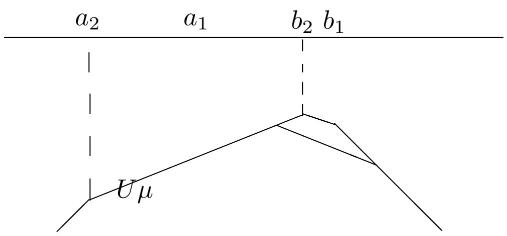
Proof. We consider the case of μ with three atoms in $x_1 < x_2 < x_3$. Note that $U\mu$ is equal to $-|x|$ on $\mathbb{R} \setminus (x_1, x_3)$, is piece-wise linear with kinks in x_1, x_2, x_3. The reasoning is easiest explained with a series of drawings:



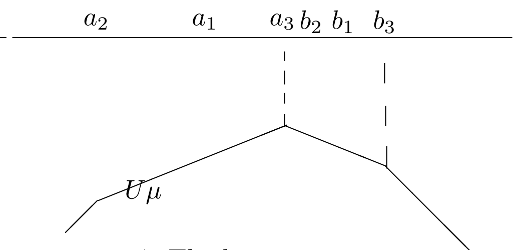
1. The initial situation



2. The first step



3. The second step



4. The last step

Note that $x_1 = a_2$, $x_2 = a_3$ and $x_3 = b_3$. It follows that with such choice $X_\tau \sim \mu$ where

$$\tau_i = H_{a_i,b_i} \circ \theta_{\tau_{i-1}},\ i = 1, 2, 3, \quad \tau_0 = 0 \text{ and } \tau := \tau_3.$$

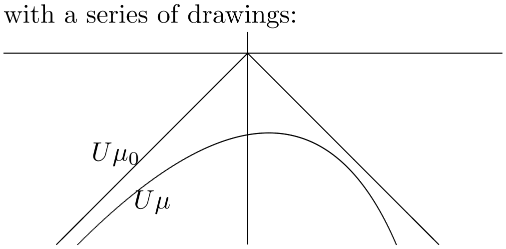
Consider the set of paths ω which are stopped in $a_2 = x_1$: $\{\omega : X(\omega)_{\tau(\omega)} = x_1\}$. It can be described as the set of paths that hit a_1 before b_1 and then a_2 before b_2. It is clear from our construction that we can choose these points in an infinite number of ways and each of them yields a different embedding (paths stopped at x_1 differ). Indeed as for any $a_1 \le \bar{a}_1 < 0 < b_2 < \bar{b}_2 < b_1$, $\mathbb{P}(H_{\{\bar{a}_1\}} < H_{\{\bar{b}_2\}} < H_{\{a_1\}} < H_{\{b_1\}}) > 0$, the result follows from the fact that if $\mathbb{P}(X_{\rho_1} \neq X_{\rho_2}) > 0$ then $\mathbb{P}(\rho_1 = \rho_2) < 1$ for two stopping times ρ_1, ρ_2. □

Each of the stopping times in the poof above is minimal. However only two of them are a composition of two exit times: for $a_1 = a_2 = x_1$ or for $b_1 = b_3 = x_3$, the rest requires considering three exit times. The former correspond to reducing the potential by taking tangents to the target potential and is the procedure suggested in the original work of Chacon and Walsh. In the case of μ with n atoms, the potential $U\mu$ has $(n-1)$ linear segments and we have to order them which can be done in $(n-1)!$ ways. Several of them are known as particular solutions. For example, taking tangents from left to right corresponds to Azéma-Yor [AY79b] solution to the SEP.

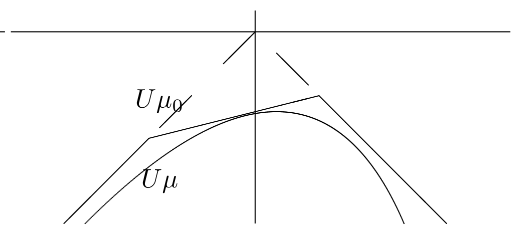
4.2.2 Arbitrary measures

Making use of potential theory on the real line, Chacon and Walsh [CW76a] gave an elegant and simple description of a general method to obtain a solution to SEP. Their work was based on an earlier paper of Chacon and Baxter [BC74], who worked with a more general setup and obtained results, for example, for n-dimensional Brownian motion. This approach proves very fruitful in one dimension as we saw already above in the case of measures with finite support. We now describe the solution for an arbitrary $\mu \in \mathcal{M}_0^1$ and a continuous local martingale X, $\langle X \rangle_\infty = \infty$ a.s.

We start by assuming $X_0 = 0$ and write $\mu_0 = \delta_0$. Choose a point x such that $U\mu_0(x) > U\mu(x)$ and draw a tangent to $U\mu$ through $(x, U\mu(x))$. This line cuts the potential $U\delta_0$ in two points $a_1 < 0 < b_1$. Naturally $x_1 \in [a_1, b_1]$. We consider the new potential $U\mu_1$ given by $U\mu_0$ on $(-\infty, a_1] \cup [b_1, \infty)$ and linear on $[a_1, b_1]$. We iterate the procedure taking tangents at x_n. The particular choice of tangents which we use to produce potentials $U\mu_n$ is not important. It suffices to see that we can indeed choose tangents[1] in such a way that $U\mu_n \to U\mu$ (and therefore $\mu_n \Rightarrow \mu$). This is true, as μU is a concave function and it can be represented as the infimum of a countable number of affine functions[2]. The stopping time is obtained therefore through a limit procedure. We have $\tau_1 = H_{a_1,b_1}$, $\tau_2 = H_{a_2,b_2} \circ \theta_{\tau_1}$, ..., $\tau_n = H_{a_n,b_n} \circ \theta_{\tau_{n-1}}$. Note that $\tau_n \leq H_{\underline{a}_n, \bar{b}_n} < \infty$ a.s., where $\underline{a}_n := \min\{a_i : i \leq n\}$, $\bar{b}_n := \max\{b_i : i \leq n\}$. The sequence (τ_n) is increasing and hence $\tau := \lim \tau_n$ is well defined. The solution is easily explained with a series of drawings:



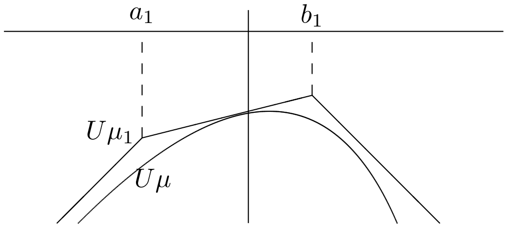
1. The initial situation



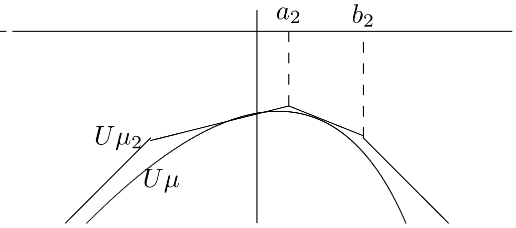
2. The first step

[1]It is convenient to consider tangents but in fact the construction works if, at n^{th} step, we consider an arbitrary line l_n which stays above $U\mu$, has gradient in $(-1, 1)$, and $U\mu_n \to U_\mu$.

[2]To be more specific observe that if we take tangents to $U\mu$ in x_1 and x_2 then for any $x \in (x_1, x_2)$ the distance of larger of the tangent lines at x to $U\mu(x)$ is less than the change in gradient of $U\mu$, which is less than 2, times the width $(x_2 - x_1)$. Recall also that $|U\mu(x) + |x|| \to 0$ as $|x| \to \infty$. Say in the n^{th} step we reduce $U\mu_{n-1}$ by taking its minimum with the tangent to $U\mu$ at x_n. Points $x_1, \dots, x_n$ provide a partition of $[\min_{i \leq n} x_i, \max_{i \leq n} x_i]$. It follows that if the mesh of this partition goes to zero and $\min_{i \leq n} x_i \to -\infty$ and $\max_{i \leq n} x_i \to \infty$ then $U\mu_n(x) \to U_\mu(x)$ for any $x \in \mathbb{R}$ and hence, by Proposition 3.2.2, $\mu_n \Rightarrow \mu$.



3. After the first step



4. After the second step

We observe that now that the assumption $X_0 = 0$ played no role. The same method works if $X_0 \sim \nu \in \mathcal{M}_0^1$ as long as $U\nu \geq U\mu$.

Theorem 4.2.2 (Chacon and Walsh [CW76a])**.** *Suppose X is a continuous local martingale, $\langle X\rangle_\infty = \infty$ a.s. and $X_0 \sim \nu \in \mathcal{M}_m^1$. Consider $\mu \in M_m^1$ with $U\mu \leq U\nu$. Let $U\mu_n$ be a sequence of potentials $\mu_0 = \nu$ and $U\mu_n = \min\{U\mu_{n-1}, \ell_n\}$, where ℓ_n is a tangent to $U\mu$ chosen so that $U\mu_n \searrow U\mu$. Let τ_n correspond to the composition of first exit times which embeds μ_n: $X_{\tau_n} \sim \mu_n$. Then $\tau_n \nearrow \tau < \infty$ a.s., $X_\tau \sim \mu$ and τ is minimal.*

Proof. We first show that $\tau < \infty$ a.s. Let $\underline{x}_n = \min\{x_i : i \leq n\}$ and $\overline{x}_n = \max\{x_i : i \leq n\}$ be the smallest and the largest tangential points chosen in the first n steps. Note that by construction, $\tau \leq H_{\underline{a}_n,\bar{b}_n}$ on the set $\{X_{\tau_n} \in [\underline{x}_n, \overline{x}_n]\}$. Since $U\mu_n$ is tangential to $U\mu$ in $\underline{x}_n$ and $\overline{x}_n$ it follows from (i) in Proposition 3.2.2 that

$$\mu([\underline{x}_n, \overline{x}_n]) \geq \mu_n([\underline{x}_n, \overline{x}_n]) \geq \mu((\underline{x}_n, \overline{x}_n)).$$

Putting the two facts together we obtain

$$\mathbb{P}(\tau \leq H_{\underline{a}_n,\bar{b}_n}) \geq \mu_n([\underline{x}_n, \overline{x}_n]) \geq \mu((\underline{x}_n, \overline{x}_n)) \nearrow 1,$$

by the convergence of potentials. We thus have

$$\mathbb{P}(\tau < \infty) = \mathbb{P}\left(\cup_n\{\tau \leq H_{\underline{a}_n,\bar{b}_n}\}\right) = 1.$$

By continuity of paths of X we also have $X_{\tau_n} \to X_\tau$ a.s. and $X_\tau \sim \mu$ since $U\mathcal{L}(X_\tau) = U\mu$. Each τ_n is UI and hence $(X_{\tau_n} : n \geq 0)$ is a martingale. Using (iv) in Proposition 3.2.2 we conclude that $X_{\tau_n} \to X_\tau$ in L^1 and hence $\mathbb{E}[X_\tau | \mathcal{F}_{\tau_n}] = X_{\tau_n}$. It follows that

$$\mathbb{E}[X_\tau | \mathcal{F}_{\tau\wedge t}] = \lim_{n\to\infty} \mathbb{E}[X_\tau | \mathcal{F}_{\tau_n\wedge t}] = \lim_{n\to\infty} \mathbb{E}[X_{\tau_n} | \mathcal{F}_{\tau_n\wedge t}] = \lim_{n\to\infty} X_{\tau_n\wedge t} = X_{\tau\wedge t}$$

and hence τ is UI and also minimal by Proposition 2.2.2. We note that if μ has second moment it follows directly from the convergence of potentials that $\mathbb{E}\,\langle X\rangle_\tau = \int x^2\mu(\mathrm{d}x)$ (via *(v)* in Proposition 3.2.2) and hence τ is UI. □

Dubins [Dub68] solution is a special case of this procedure. What Dubins proposes is actually a simple method of choosing the tangents. To obtain the potential $U\mu_1$ draw tangent at $(0, U\mu(0))$, which will intersect the potential $U\delta_0$ in two points $a < 0 < b$. Taking minimum of $U\delta_{\{0\}}$ and the tangents gives $U\mu_1$. Then draw tangents in $(a, U\mu(a))$ and $(b, U\mu(b))$. The lines will intersect the potential $U\mu_1$ in (at most) four points yielding the potential $U\mu_2$. Draw the tangents in those four points obtaining (at most) 8 intersections with $U\mu_2$ which give new coordinates for drawing tangents. Iterate.

4.3 Solutions to SEP via first entry times

We started the previous section with the example of $H_{-1,1}$ or $H_{a,b}$ in general, as the first simplest example of stopping time where we can easily describe the embedded distribution. We then generalised this to obtain a method which uses iteration of first exit times of intervals to embed an arbitrary distribution. In this section we take a different approach: we want to stick to the first exit/entry time but we consider more involved domains than just an interval.

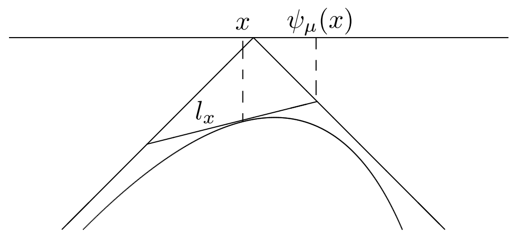
In fact a very fruitful idea is to consider a two-dimensional process (X_t, A_t), where A is an *auxiliary* process which provides a *natural clock* for X and to look at first entry times for (X, A) to a domain. It may be natural to expect such a solution to "optimaise", in some way, the distribution of the stopped auxiliary process. We consider here three examples of the auxiliary process: $A_t = t$ the usual time which leads to Root's and Rost's solutions to the SEP, $A_t = \overline{X}_t := \sup_{u \le t} X_u$ the running supremum process which leads to Azéma and Yor solution, and $A_t = L_t$ the local time in zero of X which corresponds to Vallois' solution. We will see that indeed in all of these cases the distribution of A_τ is extremal in an appropriate sense.

4.3.1 Azéma-Yor solution via potential theory

Let us go back to the case of a measure μ with n atoms, $\mu = \sum_{i=1}^n p_i \delta_{\{x_i\}}$, $x_1 \le \ldots \le x_n$. Consider Chacon-Walsh potential picture and consider the stopping time resulting from taking tangents to $U\mu$ from left to right. This corresponds to composition of consecutive first exit times from some intervals (a_i, b_i) where a_i, b_i are increasing sequences. In fact $a_i = x_i$ are subsequent atoms of μ and $b_i = \psi_\mu(x_i)$ are given in function of x_i. In the i^{th} step, if the process reaches x_i then it is stopped there. If it reaches b_i then it continues until it either hits x_{i+1} (stopped) or b_{i+1} (continues). We further have $b_n = x_n$ so that the final stopping time τ satisfies $\tau \le H_{x_1, x_n}$. Observe that a convenient way of keeping track of which step we are currently running is simply to look at the maximum of the process:

$$\tau = \inf\{t \ge 0 : \psi_\mu(X_t) \le \overline{X}_t\}, \tag{4.2}$$

where $\psi_\mu(x)$ is defined as the point in which the tangent to $U\mu$ at x, which we denote l_x, intersects $U\delta_{\{0\}}$. Note that at x_i we mean the left-tangent, i.e. the tangent which continues the linear segment of $U\mu$ between x_{i-1} and x_i. We go on to compute $\psi_\mu(x)$ in points of differentiability of $U\mu$ where the tangent is unique. It follows that we obtain ψ_μ for all x by making it left-continuous. The following drawing summarises these quantities:



The slope of l_x is given by $U\mu'(x) = 2\overline{\mu}(x) - 1$ by (3.2), where we recall $\overline{\mu}(x) = \mu([x,\infty))$. This gives the equation for $\psi_\mu(x)$, the intersection of l_x and $-|t|$:

$$\begin{aligned}
-\psi_\mu(x) &= U\mu(x) + (\psi_\mu(x) - x)U\mu'(x), \\
-\psi_\mu(x) &= x(2\overline{\mu}(x) - 1) - 2\int_{[x,\infty)} y\mu(\mathrm{d}y) + (\psi_\mu(x) - x)(2\overline{\mu}(x) - 1), \\
\psi_\mu(x) &= \frac{1}{\overline{\mu}(x)} \int_{[x,\infty)} y\mu(\mathrm{d}y), \qquad (4.3)
\end{aligned}$$

which is called the *barycentre function* defined by Azéma and Yor, and also called the Hardy-Littlewood (maximal) function. It is easy to see from the construction of $\psi_\mu(x)$ that it is a non-decreasing function which we took left-continuous above. The formula holds for x with $\overline{\mu}(x) \in (0,1)$. Note that if $\xi \sim \mu$ then

$$\psi_\mu(x) = \mathbb{E}[\xi|\xi \geq x],$$

which explains the name barycentre. We assumed above $X_0 = 0$ but it is easy to verify that if $X_0 \neq 0$ and $\mu \in \mathcal{M}^1_{X_0}$ then the intersection of tangent at x to $U\mu$ with $-|t - X_0|$ is still given by (4.3). To complete the description of ψ_μ consider x with $\overline{\mu}(x) = 1$. Then we simply have $U\mu(x) = x - X_0$ and so the tangent is just the line $\{(t, t - X_0) : t \in \mathbb{R}\}$, which intersects the line $\{(t, X_0 - t) : t \in \mathbb{R}\}$ in $(X_0, 0)$, so we put $\psi_\mu(x) = X_0$. In fact for such x, (4.3) is well defined and gives the same answer. For x with $\overline{\mu}(x) = 0$, we have $\mu U(x) = -x + X_0$ and so the very point $(x, \mu U(x))$ lies on the line $\{(t, -t + X_0) : t \in \mathbb{R}\}$ and we put $\psi_\mu(x) = x$. This completes the definition of the function ψ_μ, which coincides with the one given in [AY79b].

Theorem 4.3.1 (Azéma-Yor [AY79b]). *Let X be a continuous local martingale, $X_0 = 0$ and $\langle X\rangle_\infty = \infty$ a.s. Given $\mu \in \mathcal{M}_0^1$ define $\psi_\mu(x)$ via (4.3) for $x < \inf\{x : \overline{\mu}(x) = 0\}$ and $\psi_\mu(x) = x$ otherwise. Let*

$$\tau^\mu_{AY} := \inf\{t \geq 0 : \psi_\mu(X_t) \leq \overline{X}_t\} = \inf\{t \geq 0 : X_t \leq b_\mu(\overline{X}_t)\}, \tag{4.4}$$

where b_μ is the right-continuous inverse of ψ_μ. Then $X_{\tau^\mu_{AY}} \sim \mu$ and τ^μ_{AY} is minimal or, equivalently, $X^{\tau^\mu_{AY}}$ is a UI martingale.

It is possible to give a rigurous proof of the result relying only on potential theoretic arguments. However this requires a rather tedious construction of $U\mu_n \to U\mu$ such that barycentre functions converge. Instead, in the subsequent section, we give a proof which uses martingale arguments.

Sometimes it is convenient to reason in terms of b_μ the right-continuous inverse of ψ_μ. If we draw a tangent line to $U\mu$ which goes through $(y, -y)$ then $b_\mu(y)$ is the point at which the tangent first touches $U\mu$. Recall that the tangent has the smallest possible gradient among lines joining $(y, -y)$ and $U\mu$ which gives

$$b_\mu(y) = \max \operatorname{argmin} \frac{U\mu(x) + y}{y - x},$$

where max corresponds to the fact that we take b_μ right continuous so that if we have an interval of tangent points than we take the largest of them. Using (3.4) we can rewrite the above as

$$b_\mu(y) = \max \operatorname{argmin} \frac{U\mu(x) + x}{y - x} = \max \operatorname{argmin} \frac{C_\mu(x)}{y - x},$$

which shows that a tangent to C_μ in $b_\mu(y)$ goes through $(y, 0)$. Suppose that $\mu(\{b_\mu(y)\}) = 0$ so that the tangent has slope given uniquely by $C'_\mu(b_\mu(y)) = -\overline{\mu}(b_\mu(y))$. We conclude that

$$\overline{\mu}(b_\mu(y)) = \inf_{x<y} \frac{C_\mu(x)}{y - x} \tag{4.5}$$

which we will use later when showing that Azéma-Yor solution maximises the law of $\overline{X}_\tau$. Note also that a tangent to C_μ in $(x, C_\mu(x))$ crosses the x-axis in $\psi_\mu(x)$. Naturally (4.3) may be equally well derived from this description. This is sometime used in financial context.

Finally, similarly to Theorem 4.2.2, it is clear how to extend the Azéma-Yor embedding to the case of $X_0 \sim \nu$. It suffices to replace ψ_μ with $\psi^\nu_\mu(x)$ — the point of intersection of the tangent to $U\mu$ at x with $U\nu$.

4.3.2 Azéma-Yor solution via martingale arguments

The original proof of the embedding property in Azéma-Yor did not use the potential theoretic arguments. Instead it relied on martingale arguments. It is important to stress that above, in the potential theoretic picture, it was natural to approximate μ with atomic measures. Such measures have easy to describe embeddings and potential. In contrast, from martingale point of view, smooth measures are more natural. We will hence approximate arbitrary μ with μ_n such that ψ_{μ_n} are continuous and strictly increasing. This would correspond to taking $U\mu_n \searrow U\mu$ with $U\mu_n$ strictly concave and continuously differentiable.

The main protagonists of this section are the so-called Azéma-Yor martingales, see Carraro, El Karoui and Obłój [CEKO12] for a full account. For a C^2 function F, using Itô's formula and the that $(\overline{X}_t - X_t) \equiv 0 \ \mathrm{d}\overline{X}_t$ a.e. a.s., we obtain

$$M^F(X)_t := F(\overline{X}_t) - F'(\overline{X}_t)(\overline{X}_t - X_t) = F(X_0) + \int_0^t F'(\overline{X}_s)\mathrm{d}X_s.$$

In particular $M^F(X)$ is a local martingale. Monotone class argument shows the above holds for any locally bounded F'. Let τ be a stopping time such that X^τ is UI. Taking F with compact support, $F(0) = 0 = X_0$, we see that $M^F(X)^\tau$ is also UI and optional stopping theorem gives

$$\mathbb{E}\left[F(\overline{X}_\tau) - F'(\overline{X}_\tau)(\overline{X}_\tau - X_\tau)\right] = 0.$$

Now consider $\tau = \inf\{t \geq 0 : X_t \leq b(\overline{X}_t)\}$ so that $X_\tau = b(\overline{X}_\tau)$, where b is a right-continuous functions, $b(y) < y$ for $y < \overline{y} \leq \infty$ and $b(y) = y)$ for $y \geq \overline{y}$. The above becomes

$$\mathbb{E}\left[F(\overline{X}_\tau) - F'(\overline{X}_\tau)(\overline{X}_\tau - b(\overline{X}_\tau))\right] = 0.$$

Let $\nu := \mathcal{L}(\overline{X}_\tau)$ which is a probability measure on $[X_0, \overline{y}]$. Integrating by parts we obtain

$$\int_{\mathbb{R}} ((b(y) - y)\nu(\mathrm{d}y) - \overline{\nu}(y)\mathrm{d}y) = 0,$$

for all continuous F' with bounded support. We deduce that

$$(b(y) - y)\nu(\mathrm{d}y) - \overline{\nu}(y)\mathrm{d}y = 0, \quad y \geq 0, \quad \overline{\nu}(0) = 1.$$

which yields

$$\overline{\nu}(y) = \exp\left(-\int_0^y \frac{\mathrm{d}s}{s - b(s)}\right), \quad 0 \leq y \leq \overline{y}, \quad \overline{\nu}(\overline{y}+) = 0. \tag{4.6}$$

In particular, ν admits a density with respect to the Lebesque measure on $[0, \overline{y})$ and may have an atom at $\overline{y}$. The above expression is essentially an excursion theoretic result. We think of X as either attaining a new maximum or running an excursion away from its current maximum. We let X diffuse and each time it attains a new maximum m we setup a lower barrier, $b(m)$, such that if the next excursion away from the maximum reaches the barrier we stop. This way of thinking brings us back to our starting point: τ is the first entry time for $(X_t, \overline{X}_t)$ into a domain $D = \{(x, m) : x \leq b(m),\ m \geq 0\}$, see Figure 4.2.

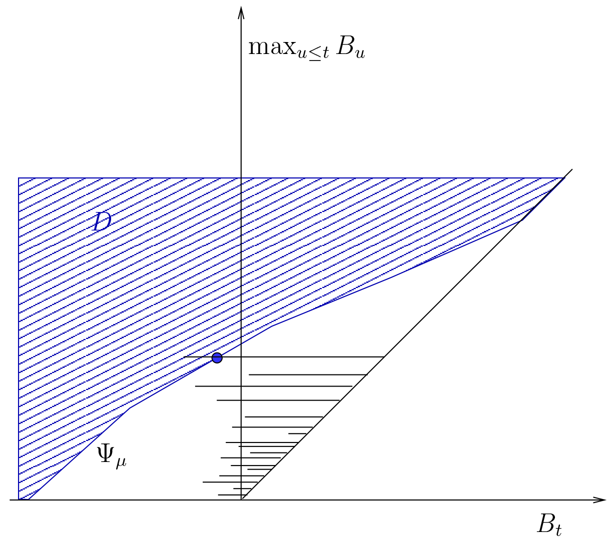


Figure 4.2: The Azéma-Yor stopping time, here $\Psi_\mu = \psi_\mu^{-1} = b_\mu$.

Proof of Theorem 4.3.1. Write $b = b_\mu$, $\psi = \psi_\mu$ and $\tau = \tau_{AY}^\mu$. Note that $\psi = b^{-1}$ the left-continuous inverse of b and, using Lemma 2.3.1, we have $\{y : b(y) \geq x\} = \{y : y \geq b^{-1}(x)\}$. The equality between the two stopping times in (4.4) follows given continuity of paths of X. Note that, by (4.6), we could also take b to be left-continuous since $\overline{X}_\tau$ does not charge points on $[0, \overline{y})$, where $\overline{y} = \overline{x} := \inf\{x : \overline{\mu}(x) = 0\}$. Suppose first that μ has compact support. Then $\tau \leq H_{\underline{x},\overline{x}}$, where $\underline{x} = \sup\{x : \overline{\mu}(x) = 1\}$, and hence τ is UI.

Using the relation between ψ and b and (4.6) we obtain

$$\begin{aligned}\mathbb{P}(X_\tau \geq x) &= \mathbb{P}(b(\overline{X}_\tau) \geq x) = \mathbb{P}(\overline{X}_\tau \geq \psi(x)) \\ &= \overline{\nu}(\psi(x))) = \exp\left(-\int_0^{\psi(x)} \frac{\mathrm{d}s}{s - b(s)}\right), \quad x \leq \overline{y}.\end{aligned} \tag{4.7}$$

We need to show that this is equal to $\overline{\mu}(x)$. Note that we have equality for $x = \underline{x}$ since $\psi(\underline{x}) = 0$. We compute directly for $\underline{x} \leq z < x \leq \overline{x}$, using the fact that $b(\psi(u)) = u\ \mathrm{d}\psi(u)$ a.e.:

$$\begin{aligned}\int_{\psi(z)}^{\psi(x)} \frac{\mathrm{d}s}{s-b(s)} &= \int_z^x \frac{\mathrm{d}\psi(u)}{\psi(u)-u} = \int_z^x \frac{\mathrm{d}(\psi(u)-u)}{\psi(u)-u} + \int_z^x \frac{\mathrm{d}u}{\psi(u)-u} \\ &= \log\frac{\psi(x)-x}{\psi(z)-z} - \int_z^x \frac{-\overline{\mu}(u)\mathrm{d}u}{\int_{[u,\infty)}(s-u)\mu(\mathrm{d}s)} \\ &= \log\frac{\overline{\mu}(z)}{\overline{\mu}(x)}\frac{\int_{[x,\infty)}(s-x)\mu(\mathrm{d}s)}{\int_{[z,\infty)}(s-z)\mu(\mathrm{d}s)} - \log\frac{\int_{[x,\infty)}(s-x)\mu(\mathrm{d}s)}{\int_{[z,\infty)}(s-z)\mu(\mathrm{d}s)} \\ &= \log\frac{\overline{\mu}(z)}{\overline{\mu}(x)}\end{aligned}$$

and the result follows.

Finally suppose μ has unbounded support. Note that for any $x > 0$ we have

$$\tau_{AY}^{\mu} \leq \inf\{t > H_{\{x\}} : X_t \leq b_\mu(x)\}$$

which is finite a.s. since $\langle X\rangle_\infty = \infty$ (the latter implies $\limsup_{t\to\infty} X_t = \infty$ and $\liminf_{t\to\infty} X_t = -\infty$). Let U_n be the potential which is equal to $U\mu$ on $(-n,n)$ and is piece-wise linear otherwise. To have a unique definition let us take U_n with no kink in $-n$ and n and μ_n the corresponding measure $U\mu_n = U_n$. Since μ_n has bounded support we have $\tau_{AY}^{\mu_n}$ is UI and embeds μ_n. We also have ψ_μ is equal to ψ_{μ_n} on $(-n,n)$ and it follows that the sets $\{\tau_{AY}^{\mu_n} \neq \tau_{AY}^{\mu}\}$ are decreasing and $\mathbb{P}(\tau_{AY}^{\mu_n} < \tau_{AY}^{\mu}) \leq \mu_n((-n,n)^c) \to 0$ and $\mathbb{P}(\tau_{AY}^{\mu_n} > \tau_{AY}^{\mu}) \leq \mathbb{P}(X_{H_{-n,\psi(-n)}} = -n) \to 0$ so that $\tau_{AY}^{\mu_n} \to \tau_{AY}^{\mu}$ a.s. Weak convergence $\mu_n \Rightarrow \mu$ implies $X_{\tau_{AY}^{\mu}} \sim \mu$.

Finally, let us verify that $\tau = \tau_{AY}^{\mu}$ is UI. Following Hobson [Hob11], we compute

$$\begin{aligned}\mathbb{P}(\sup_{t\geq 0}|X_{t\wedge\tau}| \geq \lambda) =& \mathbb{P}(H_{\{-\lambda\}} < H_{\{\psi_\mu(-\lambda)\}}) \\ &+ \mathbb{P}(H_{\{-\lambda\}} > H_{\{\psi_\mu(-\lambda)\}})\,\mathbb{P}\left(\sup_{t\geq 0} X_{t\wedge\tau} \geq \lambda \middle| H_{\{-\lambda\}} < H_{\{\psi_\mu(-\lambda)\}}\right) \\ =& \frac{\psi_\mu(-\lambda)}{\lambda+\psi_\mu(-\lambda)} + \frac{\lambda}{\lambda+\psi_\mu(-\lambda)}\frac{\mathbb{P}(X_\tau \geq b_\mu(\lambda))}{\mathbb{P}(X_\tau \geq -\lambda)}.\end{aligned}$$

Using $\psi_\mu(y) \to 0$ and $\overline{\mu}(y) \to 1$ as $y \to -\infty$ we obtain

$$\begin{aligned}\lim_{\lambda\to\infty} \mathbb{P}(\sup_{t\geq 0}|X_{t\wedge\tau}| \geq \lambda) &= \lim_{\lambda\to\infty} \lambda\overline{\mu}(b_\mu(\lambda)) \\ &= \lim_{x\to\infty} \psi_\mu(x)\overline{\mu}(x) = \lim_{x\to\infty}\int_{[x,\infty)} u\mu(\mathrm{d}u) = 0.\end{aligned}$$

Proposition 2.2.3 allows to conclude. □

Azéma-Yor stopping time compares the evolution of X and $\overline{X}$ and, as mentioned earlier, we would expect it to have optimal properties with respect to stopped maximum process. Indeed, it turns out τ^μ_{AY} maximises stochastically $\overline{X}_\tau$ among solutions to the SEP of μ in X. Azéma-Yor martingales provide a simple proof of this optimal property of τ_{AY}.

Proposition 4.3.2 (Azéma-Yor [AY79a]). *Let τ be a UI stopping time such that $X_\tau \sim \mu \in \mathcal{M}^1_0$. Suppose μ has a positive density on its support. Then*

$$\mathbb{P}(\overline{X}_\tau \geq y) \leq \mathbb{P}(\overline{X}_{\tau^\mu_{AY}} \geq y) = \overline{\mu}(b_\mu(y)), \quad y \geq 0.$$

Proof. We follow the arguments in Obłój and Yor [OY06]. Taking $F(x) = (x-y)^+$ we see that $(X_t - y)\mathbf{1}_{\overline{X}_t \geq y}$ is a local martingale. Stopped at τ it is a UI martingale and yields Doob's maximal equality

$$y\,\mathbb{P}(\overline{X}_\tau \geq y) = \mathbb{E}\left[X_\tau \mathbf{1}_{\overline{X}_\tau \geq y}\right].$$

Let $p := \mathbb{P}(\overline{X}_\tau \geq y)$ and continue the above

$$\begin{aligned} yp &\leq \mathbb{E}\left[X_\tau \mathbf{1}_{X_\tau \geq \overline{\mu}^{-1}(p)}\right] = p\psi_\mu(\overline{\mu}^{-1}(p)), \quad \text{and hence} \\ \mathbb{P}(\overline{X}_\tau \geq y) &= p \leq \overline{\mu}\left(b_\mu(y)\right) = \mathbb{P}(\overline{X}_{\tau^\mu_{AY}} \geq y), \end{aligned} \tag{4.8}$$

where we used that $\overline{\mu}$ is decreasing and that the last equality follows by construction since $\{\overline{X}_{\tau^\mu_{AY}} \geq y\} = \{X_{\tau^\mu_{AY}} \geq b_\mu(y)\}$. Note that here we use the fact that ψ_μ is strictly increasing and continuous on the support of μ. □

The above extends to arbitrary measures $\mu \in \mathcal{M}^1_0$. The distribution of $\overline{X}_{\tau^\mu_{AY}}$ is denoted μ^{HL}, the Hardy-Littlewood transform of μ. If μ has positive density than $\overline{\mu}^{HL}(y) = \overline{\mu}(b_\mu(y))$ as above. In general we have

$$\overline{\mu}^{HL}(y) = \frac{C_\mu(b_\mu(y))}{y - b_\mu(y)} = \inf_{K<y} \frac{C_\mu(K)}{y-K}. \tag{4.9}$$

To see this recall the relation between the call prices and the potential $U\mu$ given in (3.4). As discussed above, $b_\mu(y)$ is the point such that the tangent to $U\mu$ at $b_\mu(y)$ intersects the line $(x,-x)$ in $(y,-y)$. It follows that $b_\mu(y)$ may be described as the point in which the tangent to C_μ intersects the $x-$axis in $(y,0)$. If $\mu(\{b_\mu(y)\}) = 0$ then C_μ is differentiable in $b_\mu(y)$ and the slope of the tangent $-\frac{C_\mu(b_\mu(y))}{y-b_\mu(y)}$ is equal to the derivative given by $-\overline{\mu}(b_\mu(y))$ which is equal to $-\overline{\mu}^{HL}(y)$ as shown above. The last equality in (4.9) follows from the elementary properties of the tangent. Consider now an interval of constancy: $b_\mu(y) = x$ for $y \in [y_-, y_+) = [\psi_\mu(x), \psi_\mu(x+))$ which corresponds to $\mu(\{x\}) > 0$. Note that the arguments given above extend to $y = y_-$. For $y \in (y-, y_+)$ we compute using (4.6)

$$\begin{aligned} \mathbb{P}(\overline{X}_{\tau^\mu_{AY}} \geq y) &= \exp\left(-\int_0^y \frac{\mathrm{d}s}{s - b_\mu(s)}\right) = \overline{\mu}^{HL}(y_-)\exp\left(-\int_{y-}^y \frac{\mathrm{d}s}{s-b_\mu(s)}\right) \\ &= \overline{\mu}^{HL}(y_-)\exp\left(-\int_{y-}^y \frac{\mathrm{d}s}{s-x}\right) = \overline{\mu}^{HL}(y_-)\frac{y_- - x}{y-x} = \frac{C_\mu(x)}{y-x} \end{aligned} \tag{4.10}$$

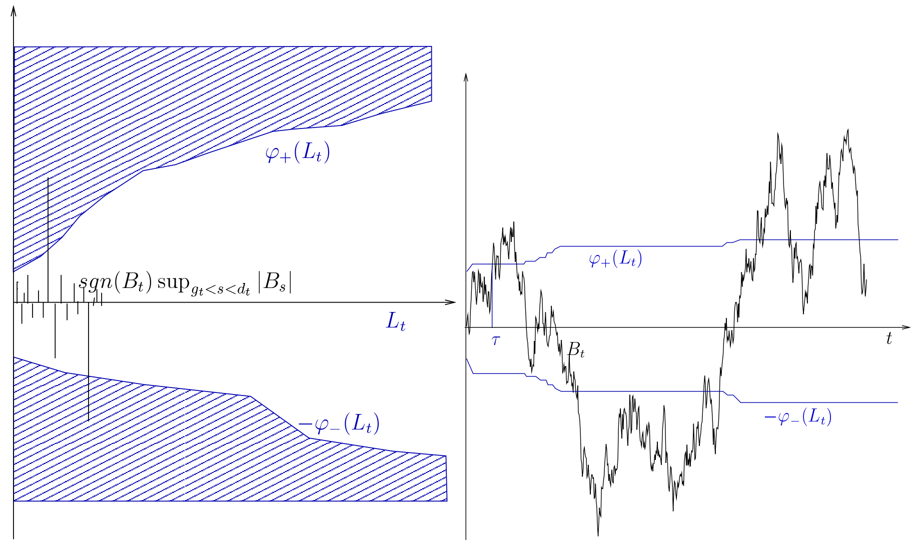


Figure 4.3: The Vallois stopping time for a Brownian motion $X = B$. The left pane presents the evolution as seen form an excursion theoretical perspective. Here $g_t = \sup\{s \leq t : B_s = 0\}$ and $d_t = \inf\{s \geq t : B_s = 0\}$.

and again the last equality in (4.9) follows from the properties of a tangent. For a different derivation of (4.9), many properties of μ^{HL} and a different proof of Proposition 4.3.2 for a general measure $\mu \in \mathcal{M}_0^1$ see Carraro, El Karoui and Obłój [CEKO12]. We do not continue this discussion here since we will see establish a generalisation of Proposition 4.3.2 in Chapter 5 with a proof exploiting pathwise inequalities.

4.3.3 Vallois' solution(s)

Vallois [Val83a] considered the first entry times for (X_t, L_t), where L_t is the local time in zero of X. We recall that L_t is increasing, supported by $\{t : X_t = 0\}$ and that $|X_t| - L_t$ is a local martingale. The essential difference with the Azéma-Yor construction is that X has excursions both below and above zero, while X has only excursions away (below) from its running maximum. We need to consider stopping times of the form

$$\inf\{t : X_t \notin (-\varphi_-(L_t), \varphi_+(L_t))\}.$$

Figure 4.3 presents an embedding of this type. In fact, Vallois showed that one can obtain closed form formulae taking $\varphi\pm$ both increasing or both decreasing. These solutions respectively maximise and minimise the stopped local time L_τ in convex order, among all UI embeddings which solve the SEP for μ in X.

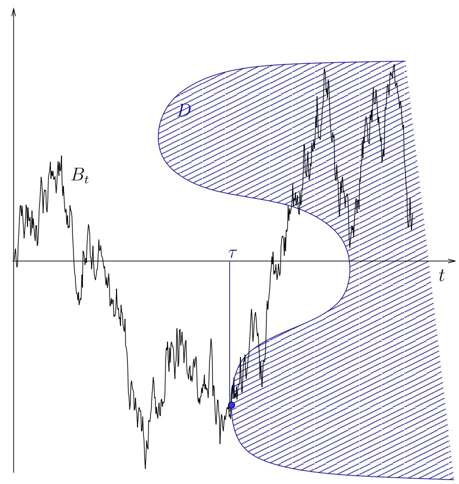


Figure 4.4: A first hitting time of a barrier by a Brownian motion B.

Similar ideas, in a much more general setup which includes both Vallois [Val83a] and Obłój and Yor [OY04] as special cases, were developed in Obłój [Obł07]. We refer to Obłój [Obł04] or to the original papers for details.

4.3.4 Root's solution

We close this short overview of solutions to the embedding problem with a pair of very natural, indeed possibly the most natural, solutions. The idea is simple: draw a region in time-space and stop when the process enters the region. For some special target measures this shape, assuming minimality of τ, will be essentially unique (in the sense that the the induced stopping time will be a.s. unique). This is true for $\mu \in \mathcal{M}_0^1$ supported on two points. However for more complex target measures there may be infinitely many different regions which yield different first entry times but embed the same distribution. Solutions of Root, and of Rost discussed below in Section 4.3.5, focus on two classes of regions such that, in essence, for any $\mu \in \mathcal{M}_0^1$ there is a unique region which embeds μ.

The key notion is that of a barrier, see Figure 4.4.

Definition 4.3.3. *A closed subset $\mathcal{R}$ of $[0, +\infty] \times [-\infty, +\infty]$ is a* barrier *if*

- $(+\infty, x) \in \mathcal{R}$ *for all* $x \in [-\infty, +\infty]$,
- $(t, \pm\infty) \in \mathcal{R}$ *for all* $t \in [0, +\infty]$,
- **if** $(t, x) \in \mathcal{R}$ **then** $(s, x) \in \mathcal{R}$ **whenever** $s > t$.

We made the last point stand out as it is the crucial defining property: if a point in time-space is in $\mathcal{R}$ then all the points to the right of it are also in $\mathcal{R}$. We state now the original result due to Root, which is formulated for a Brownian motion and μ which admits a second moment. We will discuss below generalisations where X is a diffusion and $\mu \in \mathcal{M}_0^1$.

Theorem 4.3.4 (Root [Roo69]). *Let X be a standard real valued Brownian motion. For any probability measure $\mu \in \mathcal{M}_0^2$ there exists a barrier $\mathcal{R}_\mu$ such that the first hitting time of the barrier*

$$\tau_{\mathcal{R}_\mu} = H_{\mathcal{R}_\mu}((t, X_t)) = \inf\{t \geq 0 : (t, X_t) \in \mathcal{R}_\mu\}, \tag{4.11}$$

solves the Skorokhod problem for μ, i.e. $X_{\tau_{\mathcal{R}_\mu}} \sim \mu$ and $\tau_{\mathcal{R}_\mu}$ is minimal.

Proof. We start with an atomic $\mu = \sum_{i=0}^{n+1} p_i \delta_{\{x_i\}} \in \mathcal{M}_0^2$. Given a vector $b = (b_1, \ldots, b_n) \in \mathbb{R}_+^n$ we amend it with $b_0 = b_{n+1} = 0$ and consider

$$\mathcal{R}_b := \bigcup_{i=0}^{n+1} [b_i, \infty) \times \{x_i\}.$$

Note that $\tau_{\mathcal{R}_b} \leq H_{x_0, x_{n+1}}$ and hence is a UI stopping time. We need to show that for some b we have $X_{\tau_{\mathcal{R}_b}} \sim \mu$. In fact it is then unique. Indeed if $\tau_{\mathcal{R}_b}$ and $\tau_{\mathcal{R}_b'}$ embed the same law then let $\Gamma := \{x_i : b_i < b_i'\}$. It follows that $\mathbb{P}(X_{\tau_{\mathcal{R}_b}} \in \Gamma) > \mathbb{P}(X_{\tau_{\mathcal{R}_b'}} \in \Gamma)$ if $\Gamma \neq \emptyset$ and hence Γ is empty. By symmetry we conclude $b = b'$.

Root [Roo69] used topological arguments: he shows that that a mapping $b \to \mathcal{L}(X_{\tau_{\mathcal{R}_b}})$, from set of points b to set of atomic measures on $\{x_0, \ldots, x_{n+1}\}$ endowed with suitable metrics, is continuous, maps boundary onto boundary and is one to one on the interior. It then follows that the mapping is globally onto and hence an embedding exists.

We present a different proof following Hobson [Hob11] and Cox, Hobson and Obłój [CHO10]. Let $\Delta_\mu = \{b : \mathbb{P}(X_{\tau_{\mathcal{R}_b}} = x_i) \leq p_i,\ 1 \leq i \leq n\}$ be the set of points which embed "less mass" in $x_1, \ldots, x_n$ than prescribed by μ. The excess mass is embedded in the endpoints x_0 and x_{n+1}. We claim that if $b, b' \in \Delta$ then $\underline{b} = b \wedge b'$, i.e. $b_i = b_i \wedge b_i'$, $1 \leq i \leq n$ also is an element of Δ_μ. Indeed, suppose $b \neq b'$ and fix an index i between 1 and n where they differ and, say, $b_i \wedge b_i' = b_i$. Note that $\tau_{\mathcal{R}_{\underline{b}}} \leq \tau_{\mathcal{R}_b}$ – we only extended the barrier to the left making stopping earlier more likely. However, if $X_{\tau_{\mathcal{R}_{\underline{b}}}} = x_i$ then, since $\underline{b}_i = b_i$, we also stop according to b: $X_{\tau_{\mathcal{R}_b}} = x_i$. This shows that $\mathbb{P}(X_{\tau_{\mathcal{R}_{\underline{b}}}} = x_i) \leq \mathbb{P}(X_{\tau_{\mathcal{R}_b}} = x_i) \leq p_i$ and since i was arbitrary it follows that $\underline{b} \in \Delta_\mu$.

In consequence Δ_μ has a minimal element b^*. Further, this element embeds μ. Otherwise, if we had $\mathbb{P}(X_{\tau_{\mathcal{R}_{b^*}}} = x_i) < p_i$ for some i then let $b' := (b_1^*, \ldots, b_{i-1}^*, b_i^* - \epsilon, b_{i+1}^*, \ldots, b_n^*)$. It follows that $\mathbb{P}(X_{\tau_{\mathcal{R}_{b'}}} = x_j) \leq \mathbb{P}(X_{\tau_{\mathcal{R}_{b^*}}} =$

$x_j) \leq p_j$ for $j \neq i$ and for ϵ small enough, by continuity, we still have $b' \in \Delta_\mu$ contradicting the minimality of b^* in Δ_μ.

The rest of the proof is carried out through a limit procedure. Let H denote the closed half plane, $H = [0, +\infty] \times [-\infty, \infty]$. Map H homeomorphically to a bounded rectangle F by $(t, x) \to (\frac{t}{1+t}, \frac{x}{1+|x|})$. Let F be endowed with the ordinary Euclidean metric ρ and denote by r the induced metric on H. Define the Hausdorff distance on $\mathcal{C}$, the set of closed subsets of H, via

$$r(C, D) = \max\{\sup_{x\in C} r(x, D), \sup_{y\in D} r(y, C)\}, \quad C, D \in \mathcal{C}. \tag{4.12}$$

Equipped with r, $\mathcal{C}$ is a separable, compact metric space and the subspace of all barriers is closed in $\mathcal{C}$ and hence compact. Furthermore, this metric allows us to deal with convergence in probability of first hitting times of barriers. More precisely, the application which associates with a barrier $\mathcal{R}$ its first hitting time, i.e. $\mathcal{R} \to \tau_{\mathcal{R}}$, is uniformly continuous:

Lemma 4.3.5 (Root [Roo69], Loynes [Loy70]). *Let $\mathcal{R}$ be a barrier with corresponding stopping time $\tau_{\mathcal{R}}$. If $\mathbb{P}(\tau_{\mathcal{R}} < \infty) = 1$, then for any $\epsilon > 0$ there exists $\delta > 0$ such that if $\mathcal{R}_1$ is barrier with $r(\mathcal{R}, \mathcal{R}_1) < \delta$ then $\mathbb{P}(|\tau_{\mathcal{R}} - \tau_{\mathcal{R}_1}| > \epsilon) < \epsilon$. If $\mathbb{P}(\tau_{\mathcal{R}} = \infty) = 1$ then for any $K > 0$ there exists $\delta > 0$ such that $r(\mathcal{R}, \mathcal{R}_1) < \delta$ implies $\mathbb{P}(\tau_{\mathcal{R}_1} < K) < \epsilon$.*

If a sequence of barriers converges, $\mathcal{R}_n \xrightarrow[n\to\infty]{r} \mathcal{R}$ with $\mathbb{E}\langle X\rangle_{\tau_{\mathcal{R}_n}} < K < \infty$, then $\mathbb{E}\langle X\rangle_{\tau_{\mathcal{R}}} \leq K$ and $\mathbb{P}(|\tau_{\mathcal{R}_n} - \tau_{\mathcal{R}}| > \epsilon) \xrightarrow[\mathrm{n}\to\infty]{} 0$.

With this lemma the theorem is proven for μ with finite second moment by taking a sequence μ_n of probability measures with finite supports converging to μ, $\mu_n \Rightarrow \mu$. Let $\mathcal{R}_n$ be the barrier corresponding to μ_n. Then, since the set of barriers is compact, we can extract a converging subsequence $\mathcal{R}_{n_k} \to \mathcal{R}$ and by the above lemma we conclude that $\tau_{\mathcal{R}}$ embeds μ. □

Remark 4.3.6. *The above theorem and its proof were given for a Brownian motion. However all arguments instantly extend to the case of X which solves $\mathrm{d}X_t = \sigma(X_t)\mathrm{d}W_t$, $t \geq 0$, where W is a standard Brownian motion and where σ is uniformly bounded away from zero, $\sigma(x) \geq \epsilon > 0$, $x \in \mathbb{R}$.*

Since we look at embeddings which use $A_t = t$ as the auxiliary process we may expect them to be extremal in terms of the distribution of the stopping time. Suppose $X = W$ is a Brownian motion and τ is a solution to the SEP embedding $\mu \in \mathcal{M}_0^1$ in W. Then $\mathbb{E}\,\tau = \int x^2\mu(\mathrm{d}x)$ is the same among all stopping minimal stopping times so it makes little sense to mininise or maximise $\mathbb{E}\,\tau$. We are naturally led to look at dispersion of the law of τ and try to minimise or maximise the variance of τ. It turns out Root's solution minimises the variance of τ. More generally we have a stronger result, conjectured by Kiefer [Kie72] and proved by Rost.

Theorem 4.3.7 (Rost [Ros76b]). *In the setting of Theorem 4.3.4, let $\mu \in \mathcal{M}_0^2$ and $\tau = \tau_{\mathcal{R}_\mu}$ the Root's embedding of μ. For any minimal stopping time ρ with $X_\rho \sim \mu$ we have*

$$\mathbb{E}\left[\int_{t\wedge\tau}^{\tau} h(X_s)\mathrm{d}s\right] \leq \mathbb{E}\left[\int_{t\wedge\rho}^{\rho} h(X_s)\mathrm{d}s\right], \quad \forall t \geq 0,\ h \geq 0. \tag{4.13}$$

In particular $\mathbb{E}\,\tau^2 \leq \mathbb{E}\,\rho^2$.

In fact it is enough to show the above with $h = 1$, i.e. to show that τ minimises $\mathbb{E}(\tau - \tau\wedge t) = \mathbb{E}(\tau - t)^+$. Such times are said to be *of minimal residual expectation* and Rost [Ros76a] shows that τ is of minimal residual expectation (essentially) *in and only if* it is a hitting time of a barrier. From the proof it then follows that the more general statement above is also true. In fact in Theorem 3.1.1 we can take τ to be of minimal residual expectation.

Strictly speaking, Rost proved the above result for measures with bounded support, since he considered transient processes. On the other hand, he extended the original results of Root to time-homogenuous Markov processes, see [Ros70]. Some of his results have been recently reinterpreted, and shown using entirely new techniques, via PDE methods. We come back to these developments below.

In financial terms the optimality of Root's embedding translates into minimising prices of options on variance. Recall the setup of Section 1.4.4. However instead of writing S_t as a time-change of a Brownian motion we will rather write it as a time change of a geometric Brownian motion. More precisely, we have that S is a strictly positive local martingale with $S_T \sim \mu$. Let C_s be the right-continuous inverse of $\rho_t := \langle \ln S\rangle_t$. One can verify that $X_t := S_{C_s}$ is a geometric Brownian motion (i.e. its stochastic logarithm is a Brownian motion) and we have $(X_{\rho_T}, \rho_T) = (S_T, \langle \ln S\rangle_T)$. Conversely, if we start with a geometric Brownian motion X, i.e. $\mathrm{d}X_t = X_t \mathrm{d}W_t$ for some Brownian motion W, and a UI stopping time τ such that $X_\tau \sim \mu$ then $S_t := X_{\tau\wedge t/(T-t)}$ is a continuous martingale with $S_T \sim \mu$ and $\langle \ln S\rangle_T = \tau$.

In particular, if we have an option on the variance paying $O(\langle \ln S\rangle_T)$ at maturity, a convex bounded below payoff, e.g. $(\langle \ln S\rangle_T - K)^+$ a call option on the variance, we obtain

$$\mathbb{E}[O(\langle \ln S\rangle_T)] = \mathbb{E}[O(\tau_T)] \geq \inf_{\mathrm{UI}\ \tau: X_\tau \sim S_T} \mathbb{E}[O(\rho)] = \mathbb{E}[O(\tau_{\mathcal{R}_\mu})],$$

where X is a geometric Brownian motion, $\tau_{\mathcal{R}_\mu}$ is the Root embedding of μ in X, and since by Theorem 4.3.7 Root's stopping time minimises $\mathbb{E}[O(\rho)]$ for a convex O bounded below. We note that so far we established existence of Root's stopping times for a local martingale with $\langle X\rangle_\infty = \infty$ a.s. In the case of geometric Brownian motion additional arguments are needed, see Section 4.4 below.

We have learned so far that Root's stopping time exists and has very desirable properties. The question is how to compute it? This has been first suggested by Dupire[3] and recently established in Cox and Wang [CW13]. We explain the construction by going back, again, to the potential picture. Consider a barrier $\mathcal{R}$ and $\tau = \tau_{\mathcal{R}}$. As $X_{t\wedge\tau}$ diffuses the potential $U\mathcal{L}(X_{t\wedge\tau})$ decreases. We want to control the diffusion of the potential so that it never goes below $U\mu$ and in the end is equal to $U\mu$ to achieve the embedding. Let $\underline{t}(x) := \inf\{t : (t,x) \in \mathcal{R}\}$ and consider x such that $\underline{t}(x) \in (0,\infty)$. Observe that from time $\underline{t}$ onwards the process can not cross the level x. In particular, if we are to achieve an UI embedding we have that

$$\mathbb{P}(X_{\tau\wedge\underline{t}(x)} < x) = \mu((-\infty, x)) \quad \text{and} \quad \mathbb{E}\, X_{\tau\wedge\underline{t}(x)}\mathbf{1}_{X_{\tau\wedge\underline{t}(x)}<x} = \int_{(-\infty,x)} \mu(\mathrm{d}y).$$

Further, again as no mass will cross x after $\underline{t}(x)$, UI of X^τ implies that the above remains true for all $t \geq \underline{t}(x)$ in place of $\underline{t}(x)$. By (3.2) it follows that $U\mathcal{L}(X_{\underline{t}\wedge\tau})(x) = U\mu(x)$ for all $t \geq \underline{t}(x)$. Finally, by the definition of $\underline{t}(x)$, the stopped process is free to diffuse in the neighbourhood of x before time $\underline{t}(x)$ and hence $U\mathcal{L}(X_{\tau\wedge t})(x)$ is decreasing and hence has to be strictly above $U\mu(x)$. We conclude that

$$\underline{t}(x) = \inf\{t : U\mathcal{L}(X_{\tau\wedge t})(x) = U\mu(x)\}.$$

Put differently we run the process, observe the potential and stop the process there and then when the current potential $U\mathcal{L}(X_{\tau\wedge t})$ touches $U\mu$. If X is a diffusion then the evolution of the potential $U\mathcal{L}(X_t)$ is governed by a PDE. This results in a free boundary problem for $\mathcal{R}$.

Theorem 4.3.8 (Cox and Wang). *Let X_t be a solution to an SDE: $\mathrm{d}X_t = \sigma(X_t)\mathrm{d}W_t$ with $X_0 \sim \nu \in \mathcal{M}^1_m$ and a smooth, bounded and bounded away from zero, Lipschitz function σ. Consider $\mu \in \mathcal{M}^1_m$ with $U\mu \leq U\nu$ and let $\mathcal{R}_\mu$ be Root's barrier such that $\tau_{\mathcal{R}_\mu}$ embeds μ. Let u, $u(0,x) = U\nu(x)$, be a solution to the following variational inequality*

$$\min\left\{\frac{\partial u}{\partial t}(t,x) - \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}(t,x), u(t,x) - U\mu(x)\right\} = 0,$$

taken in appropriate spaces. We then have that $u(t,x) = -\mathbb{E}\,|X_{\tau\wedge t} - x|$ and $\mathcal{R}_\mu = \{(t,x) : u(t,x) = U\mu(x)\}$.

The above result has been recently extended to more general, possible time–inhomogenuous, diffusions by Oberhauser and dos Reis [OdR14] using techniques of viscosity solutions to variational inequalities.

[3] Unpublished; see http://legacy.samsi.info/200506/fmse/transition/b.dupire.ppt

This understanding of Root's barrier allows us to deduce the optimal properties, at least informally. Observe that

$$\int_0^\tau h(X_s)\mathrm{d}s = \int_0^\tau \frac{h(X_s)}{\sigma^2(X_s)}\mathrm{d}\langle X\rangle_s = G(X_\tau) + \int_0^\tau G'(X_s)\mathrm{d}X_s,$$

for a suitable choice of G with $G'' = 2h/\sigma^2$. It follows that, at least for a large class of h, minimising $\int_{t\wedge\tau}^\tau h(X_s)\mathrm{d}s$, as in (4.13), is the same as maximising $\int_0^{t\wedge\tau} h(X_s)\mathrm{d}s$. Consider the latter for $h(x) = \mathbf{1}_{(x-\epsilon,x+\epsilon)}$. Intuitively, we want to allow X to diffuse in a neighbourhood of x for as long as possible. This is exactly what Root's embedding does – we restrict the diffusion of X in the neighbourhood of x only when we have to since the potential of $X_{\tau\wedge t}$ touches $U\mu$. We refer to Hobson [Hob11] for a different intuitive argument and to Cox and Wang [CW13] for a formal agument using pathwise inequalities as in Chapter 5.

Example. Consider $\mu = p\delta_{\{0\}} + \frac{1-p}{2}\left(\delta_{\{-1\}} + \delta_{\{1\}}\right)$, $p \in (0,1)$, and $X = W$ a Brownian motion. We know that

$$\mathcal{R}_\mu = \mathcal{R}_b = [b,\infty) \times \{0\} \cup [0,\infty) \times ((-\infty,1] \cup [1,\infty)).$$

The atom in zero of $X_{\tau_{\mathcal{R}_b}}$ is a continuous function of b, it is zero for $b=\infty$ and one for $b=0$. Hence there is a unique value $\underline{t}(0) = b$ for which μ is embedded. It follows from Theorem 4.3.8 that

$$b = \inf\{t : \mathbb{E}\,|W_{t\wedge H_{-1,1}}| = 1-p\},$$

but even this simple computation is not available in an explicit form (to the best of our knowledge). However, Theorem 4.3.8 provides us with a powerful tool to compute barriers numerically. Figure 4.3.4 gives an example uniform target law.

4.3.5 Rost

Rost in his works [Ros71a, Ros71b, Ros76a] made a fundamental contribution to the field of embedding problems. His main result establishes a necessary and sufficient condition for existence of a minimal embedding of μ in a Markov process X (taking values in some metric space). The result was based on the so-called *filling scheme* which became well known independently, see Meyer [Mey72].

The filling scheme implies the following reversed construction to Root. We say that $\mathcal{R}$ is a *reversed barrier* if it is closed and the closure of $\mathbb{R}_+ \times \mathbb{R} \setminus \mathcal{R}$ is a barrier. Essentially, a reversed barrier has the property that if $(t,x) \in \mathcal{R}$ then $(s,x) \in \mathcal{R}$ for all $0 \le s \le t$. Embedding of μ using a first hitting time of a reversed barrier is minimal and it *maximises* the variance of the stopping time

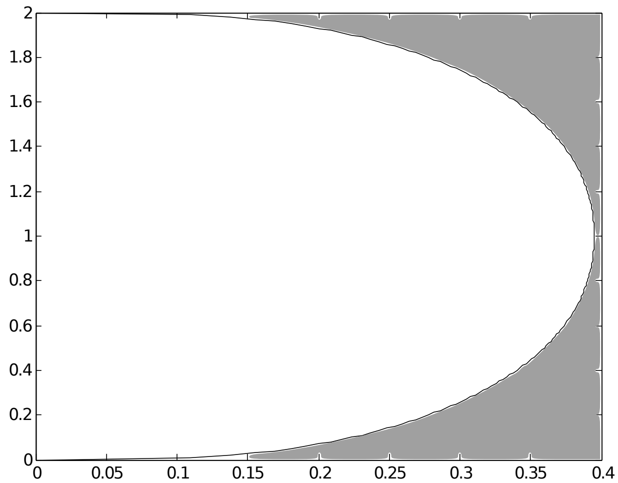


Figure 4.5: Root's Barrier embedding uniform distribution on $[0,2]$ in a Brownian motion starting in $W_0 = 1$ (generated using code by Di Huang).

among solutions to SEP for μ in X. More generally it maximises $\mathbb{E}\int_{t\wedge\tau}^{\tau} h(X_s)\mathrm{d}s$ which Root's stopping time minimises.

Existence and uniqueness of Rost embeddings has been established using purely probabilistic arguments (for Brownian motion) in a recent paper by Cox and Peskir [CP13]. Similarly to Root's barrier, Rost's reversed barrier arises as the free boundary in a suitable variational inequality (obstacle) problem.

4.4 Diffusions and the case $\langle X\rangle_\infty < \infty$

So far we have assumed that X was a continuous local martingale with $\langle X\rangle_\infty = \infty$ a.s. Recall that

$$\{\langle X\rangle_\infty < \infty\} = \{\lim_{t\to\infty} X_t \text{ exists }\}$$
$$\{\langle X\rangle_\infty = \infty\} = \{\limsup_{t\to\infty} X_t = \infty \text{ and } \liminf_{t\to\infty} X_t = -\infty\}$$

see e.g. Proposition V.1.8 in Revuz and Yor [RY01]. If we relax our setting and allow $\mathbb{P}(\langle X\rangle_\infty < \infty) > 0$ then the fundamental property that $H_{a,b} < \infty$ a.s. for any $a, b \in \mathbb{R}$, which we were using over and over again in this chapter, does not hold anymore. In consequence it may not be possible to embed any measure $\mu \in \mathcal{M}_0^1$ in X. For a trivial example consider a UI stopping time τ and the stopped process X^τ for which $\langle X^\tau\rangle_\infty = \langle X\rangle_\tau < \infty$ a.s. Clearly we can not embed in X^τ any measure ν with $U\nu \leq U\mu$.

In all the generality it is hard to say much. However in an important special case we can give a full characterisation of the embedding problem. Namely, suppose that $\langle X\rangle_\infty < \infty$ a.s. and that $X_t \to X_\infty \in \{\underline{x}, \overline{x}\}$ with X being constant for $t \geq H_{\underline{x},\overline{x}}$, for some $-\infty \leq \underline{x} < X_0 < \overline{x} \leq \infty$. Put differently, X stays between $\underline{x}$ and $\overline{x}$, is stopped upon hitting either of them and converges to one of the two as $t \to \infty$. Naturally if $\underline{x}$ or $\overline{x}$ is infinite that X has to converge to the other a.s. A motivating example is given by geometric Brownian motion for which $\underline{x} = 0 < 1 = X_0 < \infty = \overline{x}$.

In such a setting we have $H_{a,b} < \infty$ a.s. for any $\underline{x} < a < b < \overline{x}$. Theorem 4.2.2 and its proof extend immediately to the case of μ with $\mu((\underline{x}, \overline{x})) = 1$. However it is also possible to consider $\mu([\underline{x}, \overline{x}]) = 1$. Upon inspection we see that the results still extend to this case and we have $\mathbb{P}(\tau = \infty) = \mu(\{\underline{x}, \overline{x}\})$, where naturally we still assume that $\mu((-\infty, \infty)) = 1$. We conclude that any probability measure μ on $\mathbb{R}$ with $\mu \in \mathcal{M}^1_{X_0}$ with $\mu([\underline{x}, \overline{x}]) = 1$ can be embedded in X using a UI stopping time.

The above case allows one to deal with the situation when X is not a local martingale but a continuous diffusion process (with good boundary behaviour). Such a process admits a scale function s such that $Y_t = s(X_t)$, $t \geq 0$, is a continuous local martingale and falls into our discussion above. In particular we can describe precisely which measures can be embedded in Y and measures which can be embedded in X are simply the image by s^{-1}. We refer to Cox and Hobson [CH04a] for details.

4.5 Embedding processes

Suppose now that we want to construct a martingale X with given marginals μ_t for all times $t \geq 0$. A necessary a sufficient condition for existence of X, established by Kellerer [Kel72] is that $\mu_t\, \mathcal{M}^1_m$ and $U\mu_t \leq U\mu_s$ for any $t \geq s \geq 0$ and for some $m \in \mathbb{R}$. A natural constructive approach would be to solve the SEP for μ_t and try to use the solutions to define a time change. More precisely, let W be a standard Brownian motion and consider, for example, $\tau_t := \tau^{\mu_t}_{AY}$ the Azéma-Yor solution (4.4) to the SEP for μ_t in W. For simplicity assume $m = 0$ and $\mu_0 = \delta_{\{0\}}$. The family of stopping times (τ_t) is a time-change if and only if $\tau_t \geq \tau_s$ a.s. for any $t \geq s$. This in turn is equivalent to $\psi_{\mu_t} \geq \psi_{\mu_s}$ for any $t \geq s$. If this holds than $X_t := W_{\tau_t}$ is a martingale with $X_t \sim \mu_t$. Note however that typically W is not (τ_t)–continuous (recall Definition 2.3.4) and hence X is not a continuous martingale. This construction was first discussed by Madan and Yor [MY02] who also showed that X is a strong Markov process. As an example one can check that when μ_t are log-normal marginals of a Geometric Brownian motion then the above construction is feasible. The resulting martingale X has paths increasing along deterministic curves with downward jumps (random in size and position), see [Xu11]. However in general it may be hard to verify the monotonicity of the barycentre functions ψ_{μ_t}.

Naturally, any other embedding, e.g. Vallois' or Root's, maybe be used instead of Azéma-Yor construction above. Each embedding will lead to a possibly different class of marginals which may be embedded. We give one more example. Consider $\mu_t \sim t^c\xi$ for some random variable $\xi \sim \mu_1 \in \mathcal{M}_0^1$ and $0 < c < 1$. Recall Hall's solution to the SEP and let (R, S) independent of W with $(R, S) \sim \rho_{\mu_1}^{Hall}$ as given in (4.1). Then it is easy to see that $(t^c R, t^c S) \sim \rho_{\mu_t}^{Hall}$ for $t \geq 0$. In consequence

$$\tau_t := H_{t^c R, t^c S} = \inf\{s \geq 0 : W_s \notin (t^c R, t^c S)\} \text{ satisfies } X_t := W_{\tau_t} \sim \mu_t.$$

We stress again that so generated X, as in any such procedure described above, is typically a discontinuous martingale. To construct continuous martingales different tools are usually applied. These go back to Krylov [Kry85a, Kry85b] and [Gyö86]. In the context of financial mathematics such constructions were pioneered by Dupire [Dup94] and are referred to as local volatility models.

Chapter 5

Pathwise inequalities

5.1 From SEP to robust hedging

Robust pricing and hedging are dual to each other. We saw this in Section 1.4.3. So far we have only discussed methods that lead to robust bounds on prices of derivatives. However it is often the hedging strategies which are of more interest: not only they allow to recover the price bounds but also to enforce them. In practice the price bounds may be too wide to be of any use however the hedging methods may often outperform classical in-model hedging when there is even a small degree of model ambiguity and some market frictions, see Obłój and Ulmer [OU12].

So how one goes about constructing a robust hedge? How to guess the cheapest superreplicating strategy of an exotic option which uses all traded assets? The first answer is to go back to the embeddings. If we identify the optimal embedding then we know that in the model it induces, e.g. $S_t = W_{\tau\wedge t/(T-t)}$, our cheapest superhedge should be a perfect hedge. Hence to find the best superhedge we study the extremal model, identify the hedging strategy in this model and amend it adding positions with zero cost in the extremal model so that the hedge becomes a superhedge in all models.

The answer above, even though it sounds straightforward, still requires a certain level of craftmanship and good intuition. Recently, stochastic control methods have been successfully applied to find robust hedging strategies in several examples, see Galichon, Henry-Labordère and Touzi [GHLT14]. Crucially, they proved successful in attacking problems with intermediate marginal constraints where guessing the right pathwise inequality directly seems very hard, if not impossible, see Henry-Labordère et al. [HLOST14].

Here we only have time to discuss one simple example of robust hedging. Recall the setup of Section 1.4. We take $\mathfrak{P}$ to be all càdlàg functions with initial value of S_0 and assume $S^0 \equiv 1$ (no interest rates). We are given prices of all call options with maturity T: $\mathcal{X} = \{(S_T - K)^+ : K \geq 0\}$. We consider pricing and hedging of a one-touch option which pays 1 if a certain barrier is breached before the maturity T: $O_T := \mathbf{1}_{\overline{S}_T \geq y}$. Note that by Theorem 4.3.2, and the discussion following it, we have supremum of possible model prices of O_T in (1.4) is equal to $\underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) = \overline{\mu}^{HL}(y)$ given by (4.9), where μ is given by (1.5).

We are now interested in the cost of the cheapest superhedge, i.e. in $\overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$ in (1.3). We claim that for any $0 \leq K < b$ the following holds for any càdlàg path $(S_t)_{t\leq T}$

$$\mathbf{1}_{\overline{S}_T \geq y} \leq \frac{(S_T - K)^+}{y - K} + \frac{(y - S_T)}{y - K}\mathbf{1}_{\overline{S}_T \geq y}. \tag{5.1}$$

Indeed, this is easily seen by considering the two cases. If the indicator is zero then the LHS is zero and the RHS is non-negative. If the indicator is one then the RHS is greater than one (if $S_T < K$) or equal to one (if $S_T \geq K$). The terms of the RHS have a clear financial meaning of a *trading strategy*. We start by buying $1/(y - K)$ call options with strike K. When the asset reaches the barrier we sell $1/(y - K)$ forwards. This is done at zero cost and the payoff is

$$\frac{1}{y - K}(S_{H_{[y,\infty)}(S)} - S_T)\mathbf{1}_{H_{[y,\infty)}(S)\leq T}$$

where $S_{H_{[y,\infty)}(S)} \geq y$ with equality, for example, if S has continuous paths. In particular, the last term on the RHS of (5.1) is less than or equal to the payoff from the forward transaction. We conclude that

$$\overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) \leq \inf_{0\leq K<y} \frac{\mathcal{P}(S_T - K)^+}{y - K} = \inf_{0\leq K<y} \frac{C_\mu(K)}{y - K}$$

Comparing with (4.9) we deduce that

$$\overline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O) \leq \inf_{0\leq K<y} \frac{C_\mu(K)}{y - K} = \overline{\mu}^{HL}(y) \leq \underline{UB}_{(\mathfrak{P},\mathcal{X},\mathcal{P})}(O)$$

and hence we have equalities throughout. The bound is attained in the model $S_t = W_{\tau^\mu_{AY}\wedge t/(T-t)}$, where W is a standard Brownian motion starting from $W_0 = S_0$ and τ^μ_{AY} is the Azéma-Yor stopping time (4.4), which in particular provides a different proof of (a generalisation of) Proposition 4.3.2. Finally, in this model the RHS of (5.1) is a hedging strategy for the one-touch option O_T.

5.2 Martingale inequalities

We saw above that (pathwise) superhedging inequalities lead to robust hedging strategies. However they also appear to be a very natural tool for constructing martingale inequalities. We explain here how to obtain Doob's classical L^p–inequality using (5.1). Let $(X_t)_{t\le T}$ be a non-negative càdlàg submartingale defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual assumptions. For simplicity assume X_0 is a constant. Recall that the pathwise inequality (5.1) holds with X in place of S. Let $K = \alpha y$ and take expectations to see that

$$\mathbb{P}(\overline{X}_T \ge y) \le \frac{\mathbb{E}[(X_T - \alpha y)^+]}{(1-\alpha)y}, \quad \alpha \in (0,1).$$

Integrating by parts and using Fubini we have, for $p > 1$,

$$\begin{aligned}
\mathbb{E}[\bar{X}_T^p] - X_0^p = \int_{X_0}^\infty p y^{p-1}\, \mathbb{P}(\overline{X}_T \ge y)\mathrm{d}y \le \mathbb{E}\int_{X_0}^\infty p y^{p-1} \frac{(X_T-\alpha y)^+}{y-\alpha y}\mathrm{d}y& \\
= \quad \mathbb{E}\int_{X_0}^{\frac{X_T}{\alpha}\vee X_0} p y^{p-1}\frac{X_T - \alpha y}{y-\alpha y}\mathrm{d}y \quad \le \quad \mathbb{E}\int_{X_0}^{\frac{X_T}{\alpha}} p y^{p-1}\frac{X_T - \alpha y}{y-\alpha y}\mathrm{d}y& \\
= \quad \frac{p}{p-1}\frac{1}{1-\alpha}\,\mathbb{E}\left[\left\{\left(\frac{X_T}{\alpha}\right)^{p-1} - X_0^{p-1}\right\} X_T\right] - \frac{\alpha}{1-\alpha}\,\mathbb{E}\left[\left(\frac{X_T}{\alpha}\right)^p - X_0^p\right]& \\
\le \quad \frac{1}{p-1}\frac{1}{(1-\alpha)\alpha^{p-1}}\,\mathbb{E}[X_T^p] - \frac{p-\alpha(p-1)}{(p-1)(1-\alpha)}X_0^p,& \qquad (5.2)
\end{aligned}$$

where we used the submartingale property of X in the last inequality. Note that the function $\alpha \mapsto \frac{1}{(1-\alpha)\alpha^{p-1}}$ attains its minimum at $\alpha^\star = \frac{p-1}{p}$. Plugging $\alpha = \alpha^\star$ into the above yields

$$\mathbb{E}[\bar{X}_T^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[X_T^p] - \frac{p}{p-1}X_0^p,$$

which is the classical Doob's L^p–inequality.

The above is a striking simple example. More generally, pathwise inequalities obtained within the robust hedging framework can be used to obtain a verity of known martingale inequalities as well as new ones. Combined with robust pricing, they can also be shown to be tight. We refer to Obłój, Spoida and Touzi [OST14] for details.

Bibliography

[ABPS13] B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer. A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance*, 2013. DOI: 10.1111/mafi.12060.

[AGY80] J. Azéma, R. F. Gundy, and M. Yor. Sur l'intégrabilité uniforme des martingales continues. In *Seminar on Probability, XIV (Paris, 1978/1979) (French)*, volume 784 of *Lecture Notes in Math.*, pages 53–61. Springer, Berlin, 1980.

[ALP95] M. Avellaneda, A. Levy, and A. Parás. Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Math Finance*, 2(2):73–88, 1995.

[AY79a] J. Azéma and M. Yor. Le problème de Skorokhod: compléments à "Une solution simple au problème de Skorokhod". In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 625–633. Springer, Berlin, 1979.

[AY79b] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In C. Dellacherie, P.-A. Meyer, and M. Weil, editors, *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[Bas83] Richard F. Bass. Skorokhod imbedding via stochastic integrals. In *Séminaire de Probabilités, XVII*, volume 986 of *Lecture Notes in Math.*, pages 221–224. Springer, Berlin, 1983.

[BC74] J. R. Baxter and R. V. Chacon. Potentials of stopped distributions. *Illinois J. Math.*, 18:649–656, 1974.

[BG68] R. M. Blumenthal and R. K. Getoor. *Markov processes and potential theory.* Pure and Applied Mathematics, Vol. 29. Academic Press, New York, 1968.

[BHLP11] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. arXiv:1106:5929v1, 2011.

[BHR01] H. Brown, D. Hobson, and L. C. G. Rogers. Robust hedging of barrier options. *Math. Finance*, 11(3):285–314, 2001.

[BL78] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

[BLJ92a] J. Bertoin and Y. Le Jan. Representation of measures by balayage from a regular recurrent point. *Ann. Probab.*, 20(1):538–548, 1992.

[BLJ92b] J. Bertoin and Y. Le Jan. Representation of measures by balayage from a regular recurrent point. *Ann. Probab.*, 20(1):538–548, 1992.

[BN] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Prob.* To appear, available at arXiv:1305.6008.

[Bre68] Leo Breiman. *Probability.* Addison-Wesley Publishing Company, Reading, Mass., 1968.

[BS73] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[BW94] Avi Bick and Walter Willinger. Dynamic spanning without probabilities. *Stochastic Process. Appl.*, 50(2):349–374, 1994.

[BY02] Bernard Bru and Marc Yor. Comments on the life and mathematical legacy of Wolfgang Doeblin. *Finance Stoch.*, 6(1):3–47, 2002.

[Cas08] G. Cassese. Asset pricing with no exogenous probability measure. *Math. Finance*, 18(1):23–54, 2008.

[CEKO12] Laurent Carraro, Nicole El Karoui, and Jan Obłój. On Azéma-Yor processes, their optimal properties and the Bachelier-Drawdown equation. *The Annals of Probability*, 40(1):372–400, 2012.

[CG79] R. V. Chacon and N. A. Ghoussoub. Embeddings in Brownian motion. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 15(3):287–292, 1979.

[CH04a] A. M. G. Cox and D. G. Hobson. An optimal embedding for diffusions. *Stochastic Process. Appl.*, 111(1):17–39, 2004.

[CH04b] A.M.G. Cox and D.G. Hobson. An optimal Skorokhod embedding for diffusions. *Stochastic Process. Appl.*, 111(1):17–39, 2004.

[CH06] A. M. G. Cox and D. G. Hobson. Skorokhod embeddings, minimality and non-centred target distributions. *Probab. Theory Related Fields*, 135(3):395–414, 2006.

[Cha77a] R. V. Chacon. Potential processes. *Trans. Amer. Math. Soc.*, 226:39–58, 1977.

[Cha77b] R. V. Chacon. Potential processes. *Trans. Amer. Math. Soc.*, 226:39–58, 1977.

[CHO08] A. M. G. Cox, D. G. Hobson, and J. Obłój. Pathwise inequalities for local time: applications to Skorokhod embeddings and optimal stopping. *Ann. Appl. Probab.*, 18(5):1870–1896, 2008.

[CHO10] Alexander M G Cox, David Hobson, and Jan Obłój. Time-homogeneous diffusions with a given marginal at a random time. *ESAIM: Probability and Statistics*, 15:S11–S24, 2010.

[CHO14] A. M. G. Cox, Z. Hou, and J. Obłój. Robust pricing and hedging under trading restrictions and the emergence of local martingale models. arXiv:1406.0551, 2014.

[CN09] René Carmona and Sergey Nadtochiy. Local volatility dynamic models. *Finance and Stochastics*, 13:1–48, 2009.

[CO11a] A. M. G. Cox and J. Obłój. Robust hedging of double touch barrier options. *SIAM Journal on Financial Mathematics*, 2:141–182, 2011.

[CO11b] Alexander M G Cox and Jan Obłój. Robust pricing and hedging of double no-touch options. *Finance and Stochastics*, 15(3):573–605, 2011.

[CP13] A. M. G. Cox and G. Peskir. Embedding laws in diffusions by functions of time. arXiv:1201.5321v3, 2013.

[CW76a] R. V. Chacon and J. B. Walsh. One-dimensional potential embedding. In *Séminaire de Probabilités, X*, pages 19–23. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[CW76b] R. V. Chacon and J. B. Walsh. One-dimensional potential embedding. In *Séminaire de Probabilités, X (Prèmiere partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 19–23. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[CW13] A. M. G. Cox and J. Wang. Root's barrier: Construction, optimality and applications to variance options. *Annals of Applied Probability*, 23(3):859–894, March 2013.

[DE06] Mark Davis and Alison Etheridge. *Louis Bachelier's Theory of Speculation*. Princeton University Press, Princeton and Oxford, 2006.

[DH07] M. H. A. Davis and D. G. Hobson. The range of traded option prices. *Math. Finance*, 17(1):1–14, 2007.

[DM75] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel.* Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[DM80] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel. Chapitres V à VIII: Théorie des martingales.* Hermann, Paris, revised edition, 1980.

[DM06] Laurent Denis and Claude Martini. A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *Ann. Appl. Probab.*, 16(2):827–852, 2006.

[DOR14] M. H. A. Davis, J. Obłój, and V. Raval. Arbitrage bounds for prices of weighted variance swaps. *Mathematical Finance*, 24:821–854, 2014.

[DS94] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994.

[DS06] Freddy Delbaen and Walter Schachermayer. *The mathematics of arbitrage.* Springer Finance. Springer-Verlag, Berlin, 2006.

[DS14] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[Dub68] L. E. Dubins. On a theorem of Skorohod. *Ann. Math. Statist.*, 39:2094–2097, 1968.

[Dup94] B. Dupire. Pricing with a smile. *Risk*, 7:18–20, 1994.

[EKJPS98] N. El Karoui, M. Jeanblanc-Picqué, and S.E. Shreve. Robustness of the black and scholes formula. *Mathematical Finance*, 8(2):93–126, 1998.

[EKQ95] Nicole El Karoui and M.-C. Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. on Control and Optimization*, 33(1):29–66, 1995.

[Fal80] Neil Falkner. On Skorohod embedding in n-dimensional Brownian motion by means of natural stopping times. In *Séminaire de Probabilitées, XIV*, volume 784 of *Lecture Notes in Math.*, pages 357–391. Springer, Berlin, 1980.

[FF91] Neil Falkner and P. J. Fitzsimmons. Stopping distributions for right processes. *Probab. Theory Related Fields*, 89(3):301–318, 1991.

[FSW09] H. Föllmer, A. Schied, and S. Weber. Robust preferences and robust portfolio choice. volume 15 of *Handbook of Numerical Analysis*, pages 29–87. Elsevier, 2009.

[GF00] Peter Grandits and Neil Falkner. Embedding in Brownian motion with drift and the Azéma-Yor construction. *Stochastic Process. Appl.*, 85(2):249–254, 2000.

[GHLT14] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Prob*, 24(1):312–336, 2014.

[Gua06] P. Guasoni. No arbitrage with transaction costs, with fractional brownian motion and beyond. *Math. Finance*, 16(3):569–582, 2006.

[Gyö86] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Relat. Fields*, 71(4):501–516, 1986.

[Hal68] W.J. Hall. On the Skorokhod embedding theorem. Technical Report 33, Stanford Univ., Dept. of Stat., 1968.

[HK79] J. Michael Harrison and David M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3):381–408, 1979.

[HK13] D. Hobson and M. Klimmek. Maximising functionals of the maximum in the skorokhod embedding problem and an application to variance swaps. *Ann. Appl. Prob*, 23(5):2020–2052, 2013.

[HLOST14] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. The maximum maximum of a martingale with given n marginals. arXiv:1203.6877v3, 2014.

[Hob98a] D. G. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.

[Hob98b] David G. Hobson. The maximum maximum of a martingale. In *Séminaire de Probabilités, XXXII*, volume 1686 of *Lecture Notes in Math.*, pages 250–263. Springer, Berlin, 1998.

[Hob11] D. Hobson. The Skorokhod Embedding Problem and Model-Independent Bounds for Option Prices. In R.A. Carmona, I. Çinlar, E. amd Ekeland, E. Jouini, J.A. Scheinkman, and N. Touzi, editors, *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, 2011.

[HS10] Lars Peter Hansen and Thomas J. Sargent. Fragile beliefs and the price of uncertainty. *Quantitative Economics*, 1(1):129–162, 2010.

[JP10] Jean Jacod and Philip Protter. Risk neutral compatibility with option prices. *Finance and Stochastics*, 14(2):285–315, 2010.

[Kel72] Hans G. Kellerer. Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.*, 198:99–122, 1972.

[Kie72] J. Kiefer. Skorohod embedding of multivariate rv's, and the sample df. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 24:1–35, 1972.

[KMT75] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.

[Kni21] F. Knight. *Risk, Uncertainty and Profit.* Boston: Houghton Mifflin, 1921.

[KR90] Robert P. Kertz and Uwe Rösler. Martingales with given maxima and terminal distributions. *Israel J. Math.*, 69(2):173–192, 1990.

[Kra96] D. Kramkov. Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probability Theory and Related Fields*, 105:459–479, 1996.

[Kry85a] N V Krylov. On the relation between differential operators of second order and the solutions of differential equations. In N V Krylov, R S Lipster, and A A Novikov, editors, *Steklov Seminar 1984*, pages 214–229. Inc. Publications Division, New York, October 1985.

[Kry85b] N. V. Krylov. Once more about the connection between elliptic operators and Itô's stochastic equations. In *Statistics and control of stochastic processes (Moscow, 1984)*, Transl. Ser. Math. Engrg., pages 214–229. Optimization Software, New York, 1985.

[KS98] Ioannis Karatzas and Steven E. Shreve. *Methods of mathematical finance*, volume 39 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1998.

[Loy70] R. M. Loynes. Stopping times on Brownian motion: Some properties of Root's construction. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16:211–218, 1970.

[Lyo95] T. J. Lyons. Uncertain volatility and the risk-free synthesis of derivatives. *Applied Math Finance*, 2(2):117–133, 1995.

[Mer69] Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969.

[Mer71] Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, 3(4):373–413, 1971.

[Mey72] P. A. Meyer. Le schéma de remplissage en temps continu, d'après H. Rost. In *Séminaire de Probabilités, VI*, pages 130–150. Lecture Notes in Math., Vol. 258. Springer, Berlin, 1972.

[MMR06] Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica*, 74(6):1447–1498, 2006.

[Mon72a] Itrel Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.

[Mon72b] Itrel Monroe. Using additive functionals to embed preassigned distributions in symmetric stable processes. *Trans. Amer. Math. Soc.*, 163:131–146, 1972.

[Mon78] I. Monroe. Processes that can be embedded in Brownian motion. *Ann. Probability*, 6(1):42–56, 1978.

[MY02] D. B. Madan and M. Yor. Making Markov martingales meet marginals: with explicit constructions. *Bernoulli*, 8(4):509–536, 2002.

[NN13] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electron. J. Probab.*, 8(48):1–18, 2013.

[Nut12] M. Nutz. Pathwise construction of stochastic integrals. *Electron. Comm. Probab.*, 17(24):1–7, 2012.

[Obł04] Jan Obłój. The Skorokhod embedding problem and its offspring. *Probability Surveys*, 1:321–392, 2004.

[Obł07] Jan Obłój. An explicit solution to the Skorokhod embedding problem for functionals of excursions of Markov processes. *Stochastic Process. Appl.*, 117(4):409–431, 2007.

[OdR14] H. Oberhauser and G. dos Reis. Root's barrier, viscosity solutions of obstacle problems and reflected FBSDEs. arXiv:1301.3798v4, 2014.

[OST14] J. Obłój, P. Spoida, and N. Touzi. Martingale inequalities for the maximum via pathwise arguments. arXiv:1409.6255, 2014.

[OU12] J. Obłój and F. Ulmer. Performance of robust hedges for digital double barrier options. *International J.Theoretical and Applied Finance*, 15(1):1–34, 2012.

[OY04] J. Obłój and M. Yor. An explicit Skorokhod embedding for the age of Brownian excursions and Azéma martingale. *Stochastic Process. Appl.*, 110(1):83–110, 2004.

[OY06] Jan Obłój and Marc Yor. On local martingale and its supremum: harmonic functions and beyond. In *From stochastic calculus to mathematical finance*, pages 517–533. Springer, Berlin, 2006.

[Pen07] Shige Peng. G-Expectation, G-Brownian motion and related stochastic calculus of Itô type. In Fred Benth, Giulia Nunno, Tom Lindstrøm, Bernt Øksendal, and Tusheng Zhang, editors, *Stochastic Analysis and Applications*, volume 2 of *Abel Symposia*, pages 541–567. Springer Berlin Heidelberg, 2007.

[Per86] Edwin Perkins. The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In *Seminar on stochastic processes, 1985 (Gainesville, Fla., 1985)*, volume 12 of *Progr. Probab. Statist.*, pages 172–223. Birkhäuser Boston, Boston, MA, 1986.

[PP01] J. L. Pedersen and G. Peškir. The Azéma-Yor embedding in non-singular diffusions. *Stochastic Process. Appl.*, 96(2):305–312, 2001.

[PRT13] D. Possamaï, G. Royer, and N. Touzi. Superreplication under volatility uncertainty for measurable claims. arXiv:1302.1850, 2013.

[Roo69] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.

[Ros70] Hermann Rost. Die Stoppverteilungen eines Markoff-Prozesses mit lokalendlichem Potential. *Manuscripta Mathematica*, 3:321–329, 1970.

[Ros71a] H. Rost. The stopping distributions of a Markov Process. *Invent. Math.*, 14:1–16, 1971.

[Ros71b] Hermann Rost. The stopping distributions of a Markov Process. *Invent. Math.*, 14:1–16, 1971.

[Ros76a] H. Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 194–208. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[Ros76b] Hermann Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités, X*, pages 194–208. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[RVY02] B. Roynette, P. Vallois, and M. Yor. A solution to Skorokhod's embedding for linear Brownian motion and its local time. *Studia Sci. Math. Hungar.*, 39(1-2):97–127, 2002.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[RY01] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, revised Third edition, 2001.

[Sam65] P.A. Samuelson. Rational theory of warrant pricing. *Industrial Management Rev.*, 6:13–31, 1965.

[Saw74] Stanley Sawyer. The Skorokhod representation. *Rocky Mountain J. Math.*, 4:579–596, 1974. Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972).

[Sch92] Martin Schweizer. Mean-variance hedging for general claims. *The Annals of Applied Probability*, 2(1):171–179, 1992.

[Sch99] Philipp J. Schönbucher. A market model for stochastic implied volatility. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 357(1758):2071–2092, 1999.

[Sch07] Alexander Schied. Optimal investments for risk- and ambiguity-averse preferences: a duality approach. *Finance and Stochastics*, 11:107–129, 2007. 10.1007/s00780-006-0024-2.

[Sko61] A. V. Skorohod. *Issledovaniya po teorii sluchainykh protsessov (Stokhasticheskie differentsialnye uravneniya i predelnye teoremy dlya protsessov Markova).* Izdat. Kiev. Univ., Kiev, 1961.

[Sko65] A. V. Skorokhod. *Studies in the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[Str03] Daniel W. Stroock. *Markov processes from K. Itô's perspective*, volume 155 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2003.

[STZ11] M. Soner, N. Touzi, and J. Zhang. Quasi-sure stochastic analysis through aggregation. *Electron. J. Probab.*, 2011. To appear.

[SW08] Martin Schweizer and Johannes Wissel. Term structure of implied volatilities: absence of arbitrage and existence results. *Math. Finance*, 18:77–114, 2008.

[Tak99] Koichiro Takaoka. Some remarks on the uniform integrability of continuous martingales. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 327–333. Springer, Berlin, 1999.

[Val83a] P. Vallois. Le problème de Skorokhod sur $\mathcal{R}$: une approche avec le temps local. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 227–239. Springer, Berlin, 1983.

[Val83b] Pierre Vallois. Le problème de Skorokhod sur $\mathbf{R}$: une approche avec le temps local. In *Séminaire de Probabilités, XVII*, volume 986 of *Lecture Notes in Math.*, pages 227–239. Springer, Berlin, 1983.

[Vov09] Vladimir Vovk. Continuous-time trading and the emergence of randomness. *Stochastics*, 81(5):455–466, 2009.

[Xu11] Xingjian Xu. Fake Geometric Brownian Motion And Its Option Pricing. Master's thesis, Mathematical Institute, University of Oxford, 2011.