MIMICKING MARTINGALES

BY DAVID HOBSON

University of Warwick

Given the univariate marginals of a real-valued, continuous-time martingale, (resp., a family of measures parameterised by $t \in [0, T]$ which is increasing in convex order, or a double continuum of call prices), we construct a family of pure-jump martingales which mimic that martingale (resp., are consistent with the family of measures, or call prices). As an example, we construct a fake Brownian motion. Then, under a further "dispersion" assumption, we construct the martingale which (within the family of martingales which are consistent with a given set of measures) has the smallest expected total variation. We also give a pathwise inequality, which in the mathematical finance context yields a model-independent sub-hedge for an exotic security with payoff equal to the total variation of the price process.

1. The problem. In this article, we are concerned with the following problem: given a martingale $M = (M_t)_{0 \leq t \leq T}$ on a stochastic basis $(\Omega^M, \mathcal{F}^M, \mathbb{F}^M, \mathbb{P}^M)$, construct a *fake version* of M (or equivalently a process which *mimics M*), that is, construct [potentially on a new filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$] a stochastic process $X = (X_t)_{0 \leq t \leq T}$ such that X is a $\mathbb{F}$-martingale and the univariate marginals of X are the same as those of M, but such that the joint marginals are different.

The problem can be reformulated in two further ways. First, instead of beginning with a martingale M, we can begin with a family of laws $(\mu_t)_{0 \leq t \leq T}$ which are increasing in convex order. Then the aim is to construct a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and a $\mathbb{F}$-martingale $X = (X_t)_{0 \leq t \leq T}$ on that space such that $\mathbb{P}(X_t \leq x) = \mu_t((-\infty, x])$ for all t and all x. Then we say that X is consistent with the measures $(\mu_t)_{0 \leq t \leq T}$. Second, but closely related, instead of beginning with a process or a set of measures we can work in the setting of mathematical finance and start with a double continuum of European call prices $\{C(t, k); 0 \leq t \leq T, 0 \leq k < \infty\}$ which satisfy no-arbitrage conditions. [We assume that we are working with discounted prices, and then the no-arbitrage conditions are that: for each t, $C(t, k)$ is a decreasing convex function with $C'(t, 0+) \geq 1$ and $\lim_{k \uparrow \infty} C(t, k) = 0$; $C(t, 0)$ is a positive constant, independent of t; and $C(t, k)$ is nondecreasing in t, for all k.] Then the aim is to find a model which is consistent with the given call prices, that is, a filtered probability space and a martingale X

on that space such that $\mathbb{E}[(X_t - k)^+] = C(t, k)$ for all t and all k. Then we say X is consistent with the family of call prices $\{C(t, k)\}$.

For a given martingale $(M_t)_{0 \leq t \leq T}$ [or family of laws $(\mu_t)_{0 \leq t \leq T}$ or call prices $(C(t, k))_{0 \leq t \leq T, 0 \leq k < \infty}$], there will in general be many martingales X which mimic M (are consistent with the family of measures or call prices). (One rare exception is if μ_t is the uniform measure on the two point set $\{-t, t\}$ for each t.) Then an extension of the mimicking martingale problem is to find an extremal martingale with the correct marginals. Thus, given a path-dependent functional, for example, the total variation of the path, the aim is to find the martingale which minimises (the expected value of) the path-dependent functional within the class of martingales which have the given marginals. A strong motivation for considering problems of this type is from mathematical finance where there is a literature on robust (model-independent) pricing and hedging. Typically, in the modelling of financial security prices we do not know the underlying dynamics, but we can observe put and call prices, and from these we can infer the market-implied law of the security price at maturity times t. If a double continuum of European vanilla option prices is available, then we may assume that the family of laws $(\mu_t)_{0 \leq t \leq T}$ is given and the problem is to construct processes which are consistent with this family. Further, since the mathematical theory of option pricing postulates that security prices are given as discounted expectations under an equivalent martingale measure, a natural additional requirement is that the price process is a martingale. Then the lowest expected value of the path-dependent functional (taken over models consistent with call prices) is identified with the lowest arbitrage-free price for an exotic derivative.

The literature on these problems can be traced back to Gyöngy [14], Dupire [9], Madan and Yor [29], the literature on fake Brownian motions (Hamza and Klebaner [15], Albin [2], Oleszkiewicz [31]), PCOCs (Hirsch et al. [19]) and most recently and most relevantly for this paper, Henry-Labordère et al. [17] and Källblad et al. [27]. In addition, there are strong connections with the literatures on the Skorokhod embedding problem (SEP); see Obłój [30] and Hobson [20] for surveys, martingale optimal transport (MOT, Beiglböck et al. [4], Beiglböck and Juillet [5]), and robust hedging of options (Hobson [20, 23]).

Given an Itô process, Gyöngy [14] showed how to construct a diffusion X with the same univariate marginals. Dupire [9] took a finance viewpoint and argued that there is a unique martingale diffusion process which is consistent with any sufficiently regular set of discounted call prices. Madan and Yor [29] looked for Markov martingales which are consistent with a given set of univariate marginals and give several constructions including one based on the Azéma–Yor solution of the SEP. In the case of Gaussian laws, this construction yields a discontinuous fake Brownian motion. Unaware of this result, Hamza and Klebaner [15] introduced the fake Brownian motion problem, and gave a further example of a discontinuous martingale with $N(0, t)$ marginals. Their paper inspired Albin [2] and Oleszkiewicz [31] to produce very elegant constructions of continuous non-Markovian martingales

with $N(0, t)$ marginals. More generally, the problem of constructing processes consistent with a given set of marginals (which are increasing in convex order) was called the PCOC problem by Hirsch et al. [19]. They give several general methods and examples. However, there is much scope for further solutions, especially those with optimality properties.

From the mathematical finance perspective, the study of extremal consistent martingales can be traced back to Hobson [23]; see also the survey article [20]. In this literature, a path-dependent functional is interpreted as the payoff of an exotic option and then the lowest price which is consistent with a given set of call option prices and no-arbitrage is the minimum over all consistent martingale models of the expected value of the path-dependent functional. One of the significant achievements of this literature is to understand the relationship between this primal pricing problem and a dual hedging problem. In the dual problem, the aim is to find strategies whereby the payoffs satisfy

$$\text{Exotic option} \geq \text{Vanilla options} + \text{Self-financing strategy}.$$

Then we can represent the lower bound on the price of the exotic option in terms of the most expensive sub-hedging vanilla option portfolio (the self-financing part being costless under any martingale model). The fact that European call options (which have prices which depend on univariate marginals) are often liquidly traded, whereas derivatives with payoffs depending on two-or-more maturities (e.g., forward starting straddles or barrier options) are relatively less liquidly traded, provides a strong motivation for considering mimicking problems where the aim is to match univariate marginals, but not joint marginals.

One approach to solving the mimicking problem is to exploit links with the SEP. Given a stochastic process Y and a target law μ, the SEP is to construct a stopping time τ such that $Y_\tau \sim \mu$. The classic setting is when Y (which we now write as W) is a Brownian motion null at 0 and μ is a centred probability measure. Then, using the correspondence between right-continuous martingales and minimal time-changes of Brownian motion, given τ such that $W_\tau \sim \mu$ we can define $X = (X_t)_{0 \leq t \leq T} = (W_{A_t})_{0 \leq t \leq T}$ such that X has law μ at a fixed time: simply take $A_t = \frac{t}{T-t} \wedge \tau$. Then we have a martingale with marginal μ_T at time T. (Another approach to the same result, based on stochastic control, is described by Galichon et al. [13].) Moreover, given solutions of the SEP for nontrivial initial laws (the forward-starting SEP) we can use concatenation to construct processes with prescribed univariate marginals at a finite set of times $0 < t_1 < t_2 < \cdots < t_n \leq T$.

Following work of Beiglböck et al. [4], the problem of constructing martingales with given laws at two times (whether via a connection with the SEP or otherwise) is sometimes called the martingale optimal transport problem (MOT). Given a pair (μ, ν) of measures which are increasing in convex order, the MOT is to construct a coupling or joint law with marginals (μ, ν) which respects the martingale property.

MOT theory has proved extremely useful both in constructing extremal consistent models, and in proving martingale inequalities (see, e.g., Acciaio et al. [1]). Again, from MOT based on pairs of measures, there is a natural extension to finite families of measures.

The problem of constructing a fake version of a continuous-time martingale can be thought of as the infinite-marginal generalisation of the multi-marginal SEP or MOT. Indeed, the existence of solutions of these problems underpins a natural approach to the mimicking problem: discretize time using a sequence of partitions to reduce the continuous-time problem based on $(\mu_t)_{0 \le t \le T}$ to a sequence of discrete-time problems with marginals $(\tilde{\mu}_k^{(n)})_{0 \le k \le K(n)}$ (where $\tilde{\mu}_k^{(n)} = \mu_{kT/K(n)}$); construct a sequence of discrete-time martingales $\tilde{X}^{(n)}$ with marginals $(\tilde{\mu}_k^{(n)})_{0 \le k \le K(n)}$; reinterpret these processes as skeletons of piecewise constant continuous-time martingales $X^{(n)}$; and finally show that the processes $X^{(n)}$ converge to a consistent martingale. It is this last step which is potentially difficult, especially if we want to show that some optimality property is preserved in the limit. This is the approach taken in Henry-Labordère et al. [17], Juillet [26] and Källblad et al. [27] (based on solutions of the SEP/MOT due to Beiglböck and Juillet [5] in the first two cases and Henry-Labordère et al. [16] in the third).

The approach in this article is different. Rather than using an approximation technique we aim to write down the mimicking martingale directly. Conceptually, this is more direct than the approximation procedure used in [17] and [27]. There is one key idea which allows us to do this: our solutions are based on the principle that whenever possible we leave the location of mass unchanged. The processes we construct have finitely many jumps in every interval $[\varepsilon, T]$ and are constant between jumps.

Our main contributions are as follows. First, under fairly weak assumptions we give a family of martingales whose univariate marginals match a given set of measures. Second, under an extra simplifying assumption—the extension of the "dispersion assumption" in [21] from pairs of marginals to a continuum of marginals—we identify the consistent martingale with smallest expected total variation. (Note that if our goal is to mimic a martingale diffusion M then necessarily M has infinite variation, almost surely.) Third, we give a pathwise inequality which can be interpreted as a model independent sub-hedge in the exotic option pricing setting.

In particular, we show that there is a martingale with Gausian univariate marginals (mean zero, variance t at time t) which unlike Brownian motion has finite expected total variation. More generally, suppose $(\mu_t)_{0 \le t \le T}$ is a family of probability measures which is increasing in convex order, and suppose that μ_t has a continuous density $\rho(t, x)$ and is such that ρ is differentiable in t (and such that ρ and $\dot{\rho}$ have certain integrability properties). Now add the hypothesis that $\dot{\rho}(t, y) \le K(t)\rho(t, y)$ for some decreasing function $K(t)$: in the Brownian case we can take $K(t) = 1/2t$. In this setting, we show how to construct a consistent martingale from a Poisson point process and a family of martingale couplings.

Different martingale couplings will lead to different mimicking processes, so this construction potentially yields many consistent martingales. Further, since the dispersion assumption is satisfied in the Brownian case, we can construct the fake Brownian motion of minimal expected total variation.

2. Construction of a family of mimicking process.

2.1. *Set-up, assumptions and examples.* Fix ε, T with $0 \leq \varepsilon < T < \infty$ and let $\mathbb{T}$ be the open interval $\mathbb{T} = (\varepsilon, T)$ and $\overline{\mathbb{T}} = [\varepsilon, T]$. We suppose we are given $(\mu_t)_{t \in \overline{\mathbb{T}}}$; our aim is to construct a martingale $X = (X_t)_{t \in \overline{\mathbb{T}}}$ with marginals $(\mu_t)_{t \in \overline{\mathbb{T}}}$.

We begin by stating our assumptions on the family of measures $(\mu_t)_{t \in \overline{\mathbb{T}}}$. The first assumption is necessary for the family $(\mu_t)_{t \in \overline{\mathbb{T}}}$ to be increasing in convex order, and hence by Kellerer's theorem [28] for there to exist a martingale with this family of marginal distributions. A key quantity is $U(t, x) := \int |y - x| \mu_t(dy)$ which is (minus) the potential of μ_t. Then U is nondecreasing in t for each fixed x if and only if the family $(\mu_t)_{t \in \overline{\mathbb{T}}}$ is increasing in convex order.

ASSUMPTION 2.1. (a) For all $t \in \overline{\mathbb{T}}$, $\int \mu_t(dy) = 1$; $\int |y| \mu_t(dy) < \infty$; $\int y \mu_t(dy) = \overline{\mu}$.
(b) $U(t, x) := \int |y - x| \mu_t(dy)$ is nondecreasing in $t \in \overline{\mathbb{T}}$ for each x.

Now we impose some further regularity conditions on the potential.

ASSUMPTION 2.2. (a) The potential $U(t, x)$ is continuous in t for $t \in \overline{\mathbb{T}}$.
(b) $Q_t(x) = Q(t, x) := \frac{1}{2} \dot{U}(t, x)$ exists for each $t \in \mathbb{T}$ and is continuous in x and t. Also, $Q_t(x)$ satisfies $\lim_{x \to \pm\infty} Q_t(x) = 0$ and $\lim_{x \to \pm\infty} Q_t(x) - x Q'_t(x) = 0$. Since U is nondecreasing in t, we have $Q_t \geq 0$.
(c) For each $t \in \mathbb{T}$, $Q_t(\cdot)$ is the difference of two positive, decreasing convex functions. In particular, $Q_t(x) = \Lambda_t(x) - \Gamma_t(x)$ where $\lim_{x \uparrow \infty} \Lambda_t(x) = 0$. Further, Λ_t is such that $m(t) := \lim_{x \downarrow -\infty} \Lambda_t(x)/|x|$ and $n(t) := \lim_{x \downarrow -\infty} \Lambda_t(x) - x \Lambda'_t(x)$ are finite. Finally, $\int_{t \in \mathbb{T}} m(t) \, dt < \infty$.

Since Λ_t and Γ_t are convex we can write $\lambda_t = \Lambda''_t$ and $\gamma_t = \Gamma''_t$, whence γ_t and λ_t are measures [which we argue below have the same finite total mass $m(t)$]. For uniqueness, we insist that λ_t and γ_t are orthogonal. This is equivalent to choosing the decomposition of $Q_t = \Lambda_t - \Gamma_t$ such that Λ_t and Γ_t are as small as possible. Then $q_t = Q''_t$ may be interpreted as a signed measure and $q_t(dx) = \lambda_t(dx) - \gamma_t(dx)$.
(d) For each $t \in \mathbb{T}$, γ_t is absolutely continuous with respect to μ_t and the Radon–Nykodym derivative $R_t(x) = \gamma_t(dx)/\mu_t(dx)$ is measurable in x and t and is bounded in the sense that $R_t(x) \leq K(t)$ for some function $K(t)$ which is bounded on $\mathbb{T}$ by $\overline{K}$.
(e) For any measurable set B and for all $t \in \overline{\mathbb{T}}$, $\mu_t(B) = \mu_\varepsilon(B) + \int_\varepsilon^t q_s(B) \, ds$.

Fix $t \in \mathbb{T}$. By assumption, $\lim_{x \to \pm\infty} Q_t(x) = 0$ and since Q_t is the difference of convex functions and since $\lim_{x \uparrow \infty} \Lambda_t(x) = 0$ we have $\lim_{x \uparrow \infty} \Gamma_t(x) = 0$ also.

The fact that $\lim_{x\uparrow\infty} \Lambda_t(x) - x\Lambda'_t(x) = 0$ follows from the convexity of Λ_t as does the existence of $\lim_{x\downarrow-\infty} \Lambda_t(x) - x\Lambda'_t(x)$ in $[0, \infty]$. Hence, the content of the assumption on $n(t)$ is that $n(t)$ is finite.

Convexity of Λ_t implies that $\lim_{x\downarrow-\infty} \Lambda_t(x)/|x|$ exists in $[0, \infty]$, so again, the content of the assumption on $m(t)$ is that this limit is finite. In addition, $\lim_{x\downarrow-\infty} Q_t(x) = 0$ by hypothesis and we conclude that $\lim_{x\downarrow-\infty} \Gamma_t(x)/|x| = m(t)$ also. In particular, γ_t and λ_t have the same total mass, and q_t has zero total mass.

Finally, since $\lim_{x\downarrow-\infty} Q_t(x) - xQ'_t(x) = 0$ and $\lim_{x\downarrow-\infty} \Lambda_t(x) - x\Lambda'_t(x) = n(t)$, we conclude that $\lim_{x\downarrow-\infty} \Gamma_t(x) - x\Gamma'_t(x) = n(t)$ also. Hence, λ_t and γ_t have the same mean. The fact that $Q_t \geq 0$ implies that the pair (γ_t, λ_t) is increasing in convex order.

REMARK 2.1. Our ultimate goal is to construct a process on $[0, T]$. However, it is clear from the examples below that often there is no universal bound $\overline{K}$ for $K(t)$ on intervals of the form $\mathbb{T} = (0, T)$, and the processes we build will have infinitely many jumps on such intervals. To circumvent any problems related to the presence of infinitely many jumps near $t = 0$ we first construct processes with time parameter set $\overline{\mathbb{T}} = [\varepsilon, T]$, and then extend to larger intervals such as $[0, T]$ by concatenation. See Section 3.

REMARK 2.2. The requirement that Q exists for all $t \in \mathbb{T}$ can be relaxed to an assumption that Q exists and has the properties of parts (b), (c) and (d) of Assumption 2.2 on a subset $\mathbb{T}_0 \subseteq \overline{\mathbb{T}}$ of full Lebesgue measure.

For each $t \in \overline{\mathbb{T}}$, let I_t^μ be the smallest interval such that μ_t has support in I_t^μ. Let the endpoints of I_t^μ be given as $\{\ell_\mu(t), r_\mu(t)\}$. Let $I = \bigcup_{t\in\overline{\mathbb{T}}} I_t^\mu$. Similarly, let I_t^λ with endpoints $\{\ell_\lambda(t), r_\lambda(t)\}$ denote the smallest interval such that λ_t has support in I_t^λ, with similar conventions for other measures. Then, since the pair (γ_t, λ_t) is increasing in convex order we must have $\ell_\mu(t) \leq \ell_\lambda(t) \leq \ell_\gamma(t) \leq r_\gamma(t) \leq r_\lambda(t) \leq r_\mu(t)$.

DEFINITION 2.1. By the regular case, we mean that $U(t, x) \in C^{1,2}$ [so that the density $\rho(t, x)$ of μ_t and $\dot\rho(t, x)$ exist and are continuous], that for each $t \in \mathbb{T}$, $\int |\dot\rho(t, x)|\, dx < \infty$ and $\int |x||\dot\rho(t, x)|\, dx < \infty$ and also $\int_{\mathbb{T}} dt \int |\dot\rho(t, x)|\, dx < \infty$.

REMARK 2.3. Suppose we are in the regular case. Then $\Gamma_t(x) = \int_x^\infty (y - x)\dot\rho(t, x)^-\, dx$ and $\Lambda_t(x) = \int_x^\infty (y - x)\dot\rho(t, x)^+\, dx$. Since $\int |\dot\rho(t, x)|\, dx < \infty$ and $\int |x||\dot\rho(t, x)|\, dx < \infty$ it follows that $m(t)$ and $n(t)$ are well-defined and finite. Then using $m(t) = \lim_{x\downarrow-\infty} \Gamma_t(x)/|x|$ we have

$$m(t) = \lim_{x\downarrow-\infty} \left\{ \int_x^\infty \frac{y\dot\rho(t, y)^-}{|x|}\, dy + \int_x^\infty \dot\rho(t, y)^-\, dy \right\} = \int_{-\infty}^\infty \dot\rho(t, y)^-\, dy$$

and $n(t) = \int_{-\infty}^\infty y\dot\rho(t, y)^-\, dy$.

It follows from the condition that $\int_x |\dot{\rho}(t,x)|\,dx$ is integrable on $\mathbb{T}$ and an application of Fubini's theorem that $0 = \int_{-\infty}^{\infty} \dot{\rho}(s,x)\,dx$, and hence that $m(t) = \int \dot{\rho}(t,x)^-\,dx = \int \dot{\rho}(t,x)^+\,dx$. Similarly, we find that $n(t) = \int x\dot{\rho}(t,x)^-\,dx = \int x\dot{\rho}(t,x)^+\,dx$. If H is bounded and measurable, we have

$$\int_{t_0}^{t} du \int H(x) q_u(dx) = \int H(x) \int_{t_0}^{t} \dot{\rho}(u,x)\,dx$$

$$= \int H(x) \mu_t(dx) - \int H(x) \mu_{t_0}(dx)$$

and taking $H(x) = I_{\{x \in B\}}$ we find $\frac{d}{dt}\mu_t(B) = q_t(B)$. Finally, in the regular case, $R_t(x) = \dot{\rho}(t,x)^-/\rho(t,x)$, and Assumption 2.2(d) becomes $\dot{\rho}(t,x)^- \le K(t)\rho(t,x)$ uniformly in $x \in I$, for some decreasing function K. In particular, in the regular case verifying that the conditions of Assumption 2.2 are satisfied reduces to verifying that $\dot{\rho}(t,x)^- \le \rho(t,x)K(t)$.

We give illustrative five examples. The first two examples are natural examples where the situation is regular in the sense of Definition 2.1. The third example is also completely natural, but is such that although μ_t and γ_t have nice densities the measure λ_t is atomic. The fourth example has a reverse structure in which γ_t is purely atomic (and μ_t has an atom), but λ_t has a density. The final example is beyond the scope of our methods: the issue is not that μ_t has atoms, but rather that the atoms of μ_t move over time.

Let δ_y denote the point mass at y so that $\delta_y(B) = I_{\{y \in B\}}$. Define $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy$.

EXAMPLE 2.1 (Brownian motion). Let μ_t be the law of a centred Gaussian variable with variance t. Then the family $(\mu_t)_{t \ge 0}$ is the family of marginals of Brownian motion. We have

$$U(t,x) = 2\sqrt{t}\left\{ \frac{e^{-x^2/2t}}{\sqrt{2\pi}} + \frac{x}{\sqrt{t}}\Phi\left(\frac{x}{\sqrt{t}}\right) - \frac{x}{2\sqrt{t}} \right\};$$

$$\rho(t,x) = \frac{1}{2}U''(t,x) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t};$$

$$Q(t,x) = \frac{1}{2}\dot{U}(t,x) = \frac{1}{2\sqrt{t}}\frac{e^{-x^2/2t}}{\sqrt{2\pi}};$$

$$\dot{\rho}(t,x) = \frac{\rho(t,x)}{2t^2}(x^2 - t);$$

$$R_t(x) = \frac{(t - x^2)^+}{2t^2} \le \frac{1}{2t} =: K(t).$$

For this example, $m(t) = \frac{1}{2t}\int_{-1}^{1}(1 - y^2)e^{-y^2/2}\frac{dy}{\sqrt{2\pi}} = \frac{1}{t}\frac{e^{-1/2}}{\sqrt{2\pi}}$ and $n(t) = 0$.

EXAMPLE 2.2 (Exponential Brownian motion). Let μ_t be the time-t law of exponential Brownian motion.

$$U(t, x) = 2\Phi\left(\frac{-\ln x}{\sqrt{t}} + \frac{\sqrt{t}}{2}\right) - 2x\Phi\left(\frac{-\ln x}{\sqrt{t}} - \frac{\sqrt{t}}{2}\right) + x;$$

$$\rho(t, x) = \frac{1}{2}U''(t, x) = \frac{1}{x^{3/2}\sqrt{2\pi t}}e^{-(\ln x)^2/2t}e^{-t/8};$$

$$Q(t, x) = \frac{1}{2}\dot{U}(t, x) = \frac{\sqrt{x}}{2\sqrt{2\pi t}}e^{-(\ln x)^2/2t}e^{-t/8} = \frac{x^2}{2}\rho(t, x);$$

$$\dot{\rho}(t, x) = \frac{\rho(t, x)}{2t^2}[(\ln x)^2 - t - t^2/4];$$

$$R_t(x) = \left[\frac{1}{8} + \frac{1}{2t} - \frac{(\ln x)^2}{2t^2}\right]^+ \leq 1/(2t) + 1/8 =: K(t).$$

Note that this example is asymmetric about the mean 1 and $n(t) \neq 1$.

In general, there are few examples of fake exponential Brownian motions—a rare exception is [24]. The lognormal law does not have the scaling properties of the normal distribution, and as a consequence it seems harder to construct elegant mimicking processes.

EXAMPLE 2.3 (Continuous uniform). Let $\mu_t \sim U[-t, t]$. Although γ_t has a density, λ_t consists of a pair of point masses.

$$U(t, x) = \frac{(t^2 + x^2)}{2t}I_{\{-t < x < t\}} + |x|I_{\{|x| \geq t\}};$$

$$\rho(t, x) = \frac{1}{2}U''(t, x) = \frac{1}{2t}I_{\{-t < x < t\}};$$

$$Q(t, x) = \frac{1}{2}\dot{U}(t, x) = \frac{1}{4t^2}(t^2 - x^2)I_{\{-t < x < t\}};$$

$$\gamma_t(dx) = \frac{dx}{2t^2}I_{\{-t < x < t\}};$$

$$\lambda_t = \frac{1}{2t}\delta_{-t} + \frac{1}{2t}\delta_t;$$

$$R_t(x) = \frac{1}{t}I_{\{-t < x < t\}} \leq \frac{1}{t} := K(t).$$

In this case, $\gamma_t = \frac{1}{t}U[-t, t]$ and $\lambda_t = \frac{1}{t}U\{-t, t\}$ so that the signed measure q_t does not have a density. Nonetheless, for $B = (\underline{b}, \overline{b})$ with $0 < \underline{b} < \overline{b}$ (for example), we find $\mu_t(B) = \frac{\overline{b} \wedge t - \underline{b} \wedge t}{2t}$ and $\frac{d}{dt}\mu_t(B) = \frac{1}{2t}I_{\{\underline{b} < t < \overline{b}\}} - \frac{\overline{b} \wedge t - \underline{b} \wedge t}{2t^2} = q_t(B)$ for all t except $t = \underline{b}$ and $t = \overline{b}$. Hence, $\mu_t((\underline{b}, \overline{b})) = \mu_\varepsilon((\underline{b}, \overline{b})) + \int_\varepsilon^t q_s((\underline{b}, \overline{b}))\, ds$.

EXAMPLE 2.4. Suppose $\mu_t \sim e^{-t}\delta_0 + (1 - e^{-t})U[-1, 1]$.

$$U(t, x) = \left[\frac{(1 + x^2)}{2}(1 - e^{-t}) + e^{-t}|x|\right]I_{\{-1<x<1\}} + |x|I_{\{|x|\geq 1\}};$$

$$Q(t, x) = \frac{1}{2}\dot{U}(t, x) = \frac{e^{-t}}{4}(1 + x^2 - 2|x|)I_{\{-1<x<1\}};$$

$$\gamma_t = e^{-t}\delta_0;$$

$$\lambda_t(dx) = \frac{e^{-t}}{2}I_{\{-1<x<1\}}\,dx;$$

$$R_t(x) = I_{\{x=0\}} \leq 1 := K(t).$$

Note that in the first three of these examples if $\mathbb{T} = (\varepsilon, T)$ with $\varepsilon > 0$ then we may take $\overline{K} = K(\varepsilon)$, but if $\varepsilon = 0$ then there is no global bound on $R_t(x)$.

EXAMPLE 2.5 (Two point Discrete Uniform). Suppose $\mu_t \sim \frac{1}{2}\delta_t + \frac{1}{2}\delta_{-t}$. Then $U(t, x) = \max\{t, |x|\}$ and $Q(t, x) = I_{\{-t<x<t\}}$. This indicator function cannot be written as the difference of two convex functions, and this example lies out the scope of our analysis. Note, it is not the presence of atoms in this example which means that our approach does not work, but rather the fact that the locations of the atoms move over time.

2.2. *The construction.* Our goal is to construct a pure-jump Markov martingale with the required marginals. The idea behind the process X we construct is that it stays where it is whenever and wherever possible. In Section 4, this will be one of the features which leads to X having minimal variation. It also leads to a simple construction: most especially we can write down the process directly, and no limiting procedures are used.

The construction is based upon the twin ingredients of a Poisson point process and a family of martingale transports. The Poisson point process determines whether there is a jump, and then conditional on there being a jump the martingale transports determine the new location of the process. Our problem is to specify the ingredients in such a way that the resulting jump process has the desired marginals.

First consider the point process. Let $(N_t)_{t>0}$ be a Poisson point process on $(0, \infty) \times (0, \infty) \times (0, 1)$ with density $ds \times dh \times du$. Let $\mathcal{N}$ be the set of locations of events of the Poisson process. We think of the three coordinates as time, height and label. If we want to restrict the time domain of N to an interval A, for example, $A = (\varepsilon, T]$, then we write N^A for the Poisson point process $(N_t)_{t\in A}$ with density $ds \times dh \times du$ on $A \times (0, \infty) \times (0, 1)$.

Conditional on there being an event of the Poisson process at (s, h, u) for which $R_s(X_s) < h$, then the process jumps at time s. The location of the jump destination depends on the label u and the martingale transport. If the process is

at x at time s, if the first two coordinates of N determine that there should be a jump, and if π_s^x denotes the regular conditional distribution of the post-jump location, then we set the post-jump location to be $F_{\pi_s^x}^{-1}(u)$ where u is the label and $F_{\pi_s^x}(y) = \pi_s^x((-\infty, y])$. By convention (this will not impact on whether X has the required marginals) at the time of the jump X takes the value it jumps to rather than the value it jumps from. Hence, the process X we construct is cádlág.

Now consider the family of martingale transports. Suppose ν_0 and ν_1 are probability measures on I. Let $\mathcal{M}_1(\nu_0, \nu_1)$ be the set of martingale couplings or martingale transports from ν_0 to ν_1. In particular, $\pi \in \mathcal{M}_1(\nu_0, \nu_1)$ is a probability measure on $I \times I$ which satisfies the marginal constraints $\int_y \pi(dx, dy) = \nu_0(dx)$ and $\int_x \pi(dx, dy) = \nu_1(dy)$ and the martingale constraint $\int_y (y - x)\pi(dx, dy) = 0$. Then π has an interpretation as a rule for the "transport" of mass from ν_0 to ν_1 which respects the martingale property. We have that $\mathcal{M}_1(\nu_0, \nu_1)$ is nonempty if and only if ν_0 and ν_1 are increasing in convex order which we now assume. Any $\pi \in \mathcal{M}_1(\nu_0, \nu_1)$ can be written in disintegration form $\pi(dx, dy) = \pi^x(dy)\nu_0(dx)$. Then $\int_y \pi^x(dy) = 1$, $\int_y y\pi^x(dy) = x$ and $\int_x \pi^x(dy)\nu_0(dx) = \nu_1(dy)$. If ν_0 and ν_1 are orthogonal, then $\pi^x(\{x\}) = 0$.

The notion of a martingale transport generalises easily to any pair $(\tilde{\nu}_0, \tilde{\nu}_1)$ of measures with the same finite total mass which are in convex order. Let m be the mass of $\tilde{\nu}_0$. If $\hat{\nu}_i = \tilde{\nu}_i / m$ and if $\hat{\pi} \in \mathcal{M}_1(\hat{\nu}_0, \hat{\nu}_1)$ has disintegration representation $\hat{\pi}(dx, dy) = \hat{\nu}_0(dx)\pi^x(dy)$, then $\tilde{\pi}$ given by $\tilde{\pi}(dx, dy) = m\hat{\pi}(dx, dy) = m\hat{\nu}_0(dx)\pi^x(dy) = \tilde{\nu}_0(dx)\pi^x(dy)$ is a martingale transport of $\tilde{\nu}_0$ to $\tilde{\nu}_1$. Let $\mathcal{M} = \mathcal{M}(\tilde{\nu}_0, \tilde{\nu}_1)$ denote the set of martingale transports in this wider sense.

Several martingale couplings are known and can be described in more-or-less explicit terms. Many of the couplings are associated with solutions of a Skorokhod embedding problem for Brownian motion for a nontrivial initial law. For a pair of probability measures in increasing convex order, Chacon and Walsh [8] give a general family of solutions, of which the construction in Hobson [23] is a special case related to the Azéma–Yor [3] embedding (the Azéma–Yor embedding covers the case of an initial law which is a point mass). The Hobson [23] solution inherits from the Azéma–Yor embedding the property of having a very explicit construction. Bertoin and Le Jan [7] give a martingale coupling based on local times in the case where ν_0 and ν_1 are orthogonal. Hobson and Neuberger [22] gave the coupling which maximises the L^1 norm of the increment. Recently, Beiglböck and Juillet [5] introduced the "left-curtain" martingale coupling. For our purposes, the construction of Hobson and Pedersen [25] has certain desirable properties in that the first step of the construction is to fix any mass in common between ν_0 and ν_1, consistent with the general idea of leaving mass unmoved wherever possible. The Hobson–Pedersen construction is one of the most explicit martingale coupling for general measures, and we describe it in detail in Section 2.3 below. In Section 4, we will exploit the coupling of Hobson and Klimmek [21] which like the constructions of [22] and [5] has certain optimality properties.

Suppose we have our Poisson point process and a family of martingale transports $(\pi_t)_{t\in\mathbb{T}}$ with $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$. We begin by explaining, first informally and in the regular case, why the process $X = (X_t)_{t\in\mathbb{T}}$ we construct has density $\rho(t, x)$.

Suppose that at time $v \in \mathbb{T}$ the process is at x and x is in the support of γ_v and more generally that x is in the support of γ_s for $v \le s \le T$. Then $(X_t)_{v\le t}$ stays constant until $\tau = \inf\{s > v; \exists (s, h, u) \in \mathcal{N} \cap (h \le R_s(x))\}$. Note that since $R_s(x) < K(s) \le \overline{K}$ we have that $\tau > v$ almost surely. Then, from the properties of the Poisson process, for fixed v, t and x,

$$\mathbb{P}^{v,x}(\{\mathcal{N} \cap (v \le s \le t) \cap (h \le R_s(x))\} = \varnothing)$$

$$= \exp\left(-\int_v^t R_s(x)\,ds\right) = \exp\left(\int_v^t \frac{\dot\rho(s, x)}{\rho(s, x)}\,ds\right) = \frac{\rho(t, x)}{\rho(v, x)},$$

and it follows that

$$\mathbb{P}(X_t \in dx) = \mathbb{P}(X_v \in dx)\frac{\rho(t, x)}{\rho(v, x)} = \rho(t, x)\,dx.$$

Now suppose that x is not in the support of γ_v. Consider the rate of change in density at x at time v, assuming that X_v has density $\rho(v, x)$. We have

$$(1) \qquad \mathbb{P}(X_{v+dv} \in dx) = \mathbb{P}(X_v \in dx) + \int_{z\in I_v^\gamma} \mathbb{P}(X_v \in dz)\,dv\,R_v(z)\pi_v^z(dx).$$

If $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$ then the right-hand side of (1) can be rewritten as

$$\rho(v, x)\,dx + \int_{z\in I_v^\gamma} \rho(v, z)R_v(z)\,dz\pi_v^z(dx)\,dv$$

$$= \rho(v, x)\,dx + \int_{z\in I_v^\gamma} \gamma_v(dz)\pi_v^z(dx)\,dv$$

$$= \rho(v, x)\,dx + \lambda_v(dx)\,dv$$

$$= \rho(v, x)\,dx + \dot\rho(v, x)\,dx\,dv,$$

and hence $\mathbb{P}(X_{v+dv} \in dx) = \rho(v + dv, x)\,dx$ as required.

Our goal now is to extend this heuristic argument to prove that X_t has law μ_t. For $t \in \mathbb{T}$, define the operator $\mathcal{A}_t^{\pi_t}$ by

$$\mathcal{A}_t^{\pi_t}h(x) = R_t(x)\left\{\int_y \pi_t^x(dy)h(y) - h(x)\right\}$$

and recall that this is nonzero only if $x \in \operatorname{supp}(\gamma_t)$. Suppose $(\pi_t)_{t\in\mathbb{T}}$ is chosen with $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$ and let $X = (X_t)_{t\in\mathbb{T}}$ be the Markov process with generator $\mathcal{A}_t^{\pi_t}$ and initial law $X_\varepsilon \sim \mu_\varepsilon$. Then X is a pure-jump Markov process.

ASSUMPTION 2.3. Suppose the family $(\pi_t)_{t\in\mathbb{T}}$ is such that $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$ has representation $\pi_t(dx, dy) = \gamma_t(dx)\pi_t^x(dy)$ and for all measurable sets B, $\pi_t^x(B)$ is measurable in x and t.

Our main theorem says that under the assumptions of this section, the process $X = (X_t)_{t \in \mathbb{T}}$ has marginals $(\mu_t)_{t \in \mathbb{T}}$. The proof of this result relies on results of Feller [12] and Feinberg et al. [11] on the existence and uniqueness of solutions of the Kolmogorov forward equations for pure-jump Markov processes. Feller [12] proves existence under an assumption that jump rates are continuous in t, and this condition is weakened to measurability in [11]. The idea in these papers is to use the fact that there are only finitely many jumps (almost surely, in $\mathbb{T}$) to construct the process and its transition functions as a limit of processes which have at most n jumps. These pure-jump processes with a bounded number of jumps are simple to characterise, and it is possible to write down the probability transition functions in terms of integrals involving the transition rates.

THEOREM 2.1. *Suppose Assumptions 2.1, 2.2 and 2.3 hold. Suppose $\mathbb{T} = (\varepsilon, T)$ and suppose $X_\varepsilon \sim \mu_\varepsilon$. Then for $t \in \overline{\mathbb{T}}$, $X_t \sim \mu_t$.*

PROOF.   For $t \in \overline{\mathbb{T}}$, let ν_t denote the law of X_t. Then $\nu_\varepsilon = \mu_\varepsilon$. Further, since $R_t(x)$ is bounded, I is q-bounded in the language of Feinberg et al. [11], and by Feller [12], equation (37), or Feinberg et al. [11], Corollary 4.2, for measurable sets B,

$$(2) \quad \nu_t(B) = \mu_\varepsilon(B) + \int_\varepsilon^t ds \int_{x \in I} \nu_s(dx) R_s(x) \pi_s^x(B) - \int_\varepsilon^t \int_B R_s(x) \nu_s(dx).$$

Conversely, by construction and design the family $(\mu_t)_{t \in \mathbb{T}}$ is such that

$$\mu_t(B) = \mu_\varepsilon(B) + \int_\varepsilon^t ds\, q_s(B)$$

$$(3) \quad = \mu_\varepsilon(B) + \int_\varepsilon^t ds \int_B \lambda_s(dy) - \int_\varepsilon^t ds \int_B \gamma_s(dx)$$

$$= \mu_\varepsilon(B) + \int_\varepsilon^t ds \int_{x \in I} \int_{y \in B} \gamma_s(dx) \pi_s^x(dy) - \int_\varepsilon^t ds \int_B \gamma_s(dx)$$

$$= \mu_\varepsilon(B) + \int_\varepsilon^t ds \int_{x \in I} \mu_s(dx) R_s(x) \pi_s^x(B) - \int_\varepsilon^t ds \int_B R_s(x) \mu_s(dx).$$

Let $(\eta_t)_{t \in \mathbb{T}}$ be the family of signed measures given by $\eta_t = \nu_t - \mu_t$, and note that $\eta_t(I) = 0$. Let Δ_t denote the total variation distance between ν_t and μ_t so that $\Delta_t = \sup_B |\eta_t(B)|$, and let $\overline{\Delta}_t = \sup_{\varepsilon \leq s \leq t} \Delta_s$. Then, comparing (2) and (3)

$$\eta_t(B) = \int_\varepsilon^t ds \int_{x \in I} \eta_s(dx) R_s(x) \left[\int_{x \in I} \pi_t^x(B) - I_{\{x \in B\}} \right]$$

and it follows that

$$|\eta_t(B)| \leq \int_\varepsilon^t ds \int_{x \in I} |\eta_s(dx)| K(s) \leq 2 \int_\varepsilon^t K(s) \Delta_s\, ds \leq 2\overline{K} \int_\varepsilon^t \overline{\Delta}_s\, ds.$$

Hence, $\Delta_t \leq 2\overline{K} \int_\varepsilon^t \overline{\Delta}_s \, ds$ and since the right-hand side of this last expression is increasing, $\overline{\Delta}_t \leq 2\overline{K} \int_\varepsilon^t \overline{\Delta}_s \, ds$. It follows from Gronwall's lemma (Ethier and Kurtz [10], Appendix, Theorem 5.1) that $\Delta_t = \overline{\Delta}_t = 0$ and $X_t \sim \mu_t$ for $t \in \mathbb{T}$. Finally, since the rate of jumps is bounded, $\mathbb{P}(X_T \in B) = \lim_{s \uparrow T} \mathbb{P}(X_s \in B) = \lim_{s \uparrow T} \mu_s(B) = \mu_T(B)$ by Assumption 2.2(e). □

2.3. *An explicit martingale transport.* The goal of this section is to show that even when γ_t and λ_t are arbitrary (integrable) measures we can write down a martingale coupling.

Fix t and write γ as shorthand for γ_t, and similarly λ as shorthand for λ_t, where the orthogonal measures γ and λ have the same finite total mass and the pair (γ, λ) is increasing in convex order. Our aim is to write down a martingale transport $\pi(dx, dy)$ which has marginals γ and λ and respects the martingale condition. In fact, we write down $\pi^x(dy)$; then $\pi(dx, dy) = \gamma(dx)\pi^x(dy)$. The construction we give is based on the solution to the Skorokhod embedding problem with nontrivial initial law given in Hobson and Pedersen [25]. The only difference with respect to that paper is that there γ and λ are probability measures, but the extension to measures with finite mass does not introduce major complications.

Let $Q(x) = \frac{1}{2}\{\mathbb{E}^{X \sim \lambda}[|X - x|] - \mathbb{E}^{X \sim \gamma}[|X - x|]\}$. Since γ and λ are in convex order, we have $Q \geq 0$, and Q is continuous in x. Define

$$\xi(z) = \sup_{w:w<z} \frac{Q(z) - Q(w)}{z - w}$$

and let $g(z) \leq z$ be the largest of those values where the supremum is attained, or if the supremum is not attained then $\xi(z) = Q'(z-)$ and we set $g(z) = z$. Now define $\Xi(z) = \lambda((-\infty, z)) - \xi(z)$. Then (Hobson and Pedersen [25], Theorem 2.1) Ξ is an increasing, left-continuous function.

Decompose λ into a continuous part and a sum of atoms: $\lambda(A) = \int_{z \in A} \lambda^c(dz) + \sum_{z \in A} \lambda(\{z\})$. For $z > x$ define $\Upsilon^x(z)$ by

$$\Upsilon^x(z) = \exp\left(-\int_{(x,z)} \frac{\lambda^c(dw)}{\gamma((-\infty, w)) - \Xi(w)}\right)$$

$$\times \prod_{u \in [x,z)} \left(1 - \frac{\lambda(\{u\})}{\gamma((-\infty, u)) - \Xi(u)}\right)^+.$$

Υ^x can be thought of as a survivor function of a probability measure.

For x in the support of γ, we define π^x separately on $z > x$ and $z < x$. For $z > x$, we set

$$(4) \qquad \pi^x([z, \infty)) = \int_z^\infty dw \exp\left(-\int_x^w \frac{dv}{v - g(v)}\right)|d\Upsilon^x(w)|$$

and for $z < x$,

$$(5) \quad \pi^x((-\infty, z]) = \int_x^\infty I_{\{w:g(w)\leq z\}} \frac{dw}{w - g(w)} \exp\left(-\int_x^w \frac{du}{u - g(u)}\right)\Upsilon^x(w).$$

The following result is an immediate consequence of Proposition 5.4 in [25].

PROPOSITION 2.1 (Hobson and Pedersen [25]). *With π^x as above and $\pi(dx, dy) = \gamma(dx)\pi^x(dy)$ we have $\pi \in \mathcal{M}(\gamma, \lambda)$.*

EXAMPLE 2.6. Suppose $\gamma \sim \frac{1}{t}U[-t, t]$ and $\lambda \sim \frac{1}{2t}\delta_{-t} + \frac{1}{2t}\delta_t$ (as is the case for Example 2.3). Then $Q(x) = \frac{1}{4t^2}(t^2 - x^2)I_{\{-t < x < t\}}$.

We find $\xi(z) = \frac{1}{4t^2}(t - z)$ for $-z < t \le z$ (and $\xi(z) = 0$ otherwise), and $g(z) = -t$ for $-z < t \le z$ (and $g(z) = z$ otherwise). Then $\Xi(z) = \frac{1}{4t^2}(t + z)^+$ for $z \le t$ and for $x \in (-t, t)$, $\Upsilon^x(z) = I_{\{z < t\}}$. Note that for $z \le t$, $\int_x^z \frac{dw}{w - g(w)} = \int_x^z \frac{dw}{t + w} = \ln \frac{t+z}{t+x}$, and hence $\exp(-\int_x^z \frac{dw}{w - g(w)}) = \frac{t+x}{t+z}$. Then for $-t < x < t$ and $z > x$, we have $\pi^x([z, \infty)) = \frac{t+x}{2t}I_{\{z \le t\}}$.

For $-t < x < t$ and $z < x$, we find $\pi^x((-\infty, z]) = I_{\{-t \le z\}}\int_x^t \frac{dw}{t+w}\frac{t+x}{t+w} = I_{\{-t \le z\}}[-\frac{t+x}{t+w}]_{w=x}^t = I_{\{-t \le z\}}\frac{t-x}{2t}$.

The situation can be summarised as $\pi^x = \frac{t-x}{2t}\delta_{-t} + \frac{t+x}{2t}\delta_t$.

3. Extension to $[0, T]$. In many of our examples, the bound $K(t)$ diverges at $t = 0$ and the methods of the previous section are only sufficient to construct a process on $[\varepsilon, T]$ for $\varepsilon > 0$. (A rare exception is Example 2.4.) In the case where $K(t)$ diverges at zero, we give two approaches to constructing a process on $[0, T]$. The first approach is based on the Kolmogorov consistency condition and is non-constructive. For related ideas in a similar context, see Oleszkiewicz [31]. The second approach requires extra assumptions, and makes use of time-reversal.

3.1. *Construction via the Daniell–Kolmogorov Consistency Theorem.* Let ε be a parameter in $(0, T)$. Let $X^{(\varepsilon)} = (X_t^{(\varepsilon)})_{\varepsilon \le t \le T}$ be the Markov martingale with initial law μ_ε and marginals $(\mu_t)_{\varepsilon \le t \le T}$ constructed as in the previous section from the family of martingale transports $(\pi_t^{(\varepsilon)})_{\varepsilon < t \le T}$ and the realisation $\mathcal{N}^{(\varepsilon)} = (\mathcal{N}_t^{(\varepsilon)})_{\varepsilon < t \le T}$ of the Poisson point process $(N_t^{(\varepsilon)})_{\varepsilon < t \le T}$. Write $X^{(\varepsilon)} = (X_t^{(\varepsilon)})_{\varepsilon \le t \le T} = X^{(\varepsilon)}(X_\varepsilon^{(\varepsilon)}, (\mathcal{N}_t^{(\varepsilon)})_{\varepsilon < t \le T})$ to indicate that given the family of martingale transports the path $(X_t^{(\varepsilon)})_{\varepsilon \le t \le T}$ is a deterministic function of the initial value (at time ε) and the realisation of the Poisson point process.

Suppose now that we are given a family of martingale transports $(\pi_t)_{0 < t \le T}$ and a Poisson point process N on $(0, T] \times (0, \infty) \times (0, 1)$. For any δ in $(0, T)$, we can consider the restrictions of $(\pi_t)_{0 < t \le T}$ and $(N_t)_{0 < t \le T}$ to $(\delta, T]$. Write $\pi^{(\delta)} = (\pi_t^{(\delta)})_{\delta < t \le T} := (\pi_t)_{\delta < t \le T}$ and $N^{(\delta)} = (N_t^{(\delta)})_{\delta < t \le T} := (N_t)_{\delta < t \le T}$ for these restrictions. Then define $X^{(\delta)} = (X_t^{(\delta)})_{\delta \le t \le T} = X^{(\delta)}(X_\delta^{(\delta)}, (\mathcal{N}_t^{(\delta)})_{\delta < t \le T})$ where $X_\delta^{(\delta)}$ is a random variable with law μ_δ. It follows that if $\delta < \varepsilon$ and $X_\varepsilon^{(\varepsilon)}$ is taken to equal $X_\varepsilon^{(\delta)}$ then there is an identification or coupling between processes started at different times and $(X_t^{(\varepsilon)})_{\varepsilon \le t \le T}$ and $(X_t^{(\delta)})_{\varepsilon \le t \le T}$ are identical, omega by omega.

For $0 < t_1 < t_2 < \cdots < t_n \leq T$, define

(6) $$\mu_{t_1, t_2, \ldots, t_n} = \mathcal{L}\big(X_{t_1}^{(\varepsilon)}, X_{t_2}^{(\varepsilon)}, \ldots, X_{t_n}^{(\varepsilon)}\big).$$

By the above comments, $\mu_{t_1, t_2, \ldots, t_n}$ does not depend on ε (provided $\varepsilon \leq t_1$). Also it is clear that the measures $(\mu_{t_1, t_2, \ldots, t_n})_{n \in \mathbb{N}, 0 < t_1 < t_2 < \cdots < t_n \leq T}$ are consistent, since any pair are the finite dimensional distributions of a stochastic process $(X_t^{(\varepsilon)})_{\varepsilon \leq t \leq T}$ for some ε sufficiently small. Hence, by the Kolmogorov consistency theorem, there exists a stochastic process $\hat{X} = (\hat{X}_t)_{0 < t \leq T}$ such that the finite dimensional distributions of $\hat{X}$ are given by the measures in (6).

Let $\mathbb{D}_{k,T} = \{j2^{-k}T; 1 \leq j \leq 2^k\}$ and let $\mathbb{D}_T = \bigcup_{k \geq 1} \mathbb{D}_{k,T}$ so that $\mathbb{D}_T$ is the set of dyadic rationals on $(0, 1]$ rescaled to $(0, T]$. Consider $(\hat{X}_t)_{t \in \mathbb{D}_T}$. Fix $0 < t_0 < t_1 \leq T$ with $t_i \in \mathbb{D}_T$. Then with $\varepsilon < t_0$, and using the right-continuity of $X^{(\varepsilon)}$,

$$\begin{aligned}
\mathbb{P}\big(\hat{X}_t = \hat{X}_{t_0}, \forall t \in \mathbb{D}_T \cap [t_0, t_1]\big) &= \lim_k \mathbb{P}\big(\hat{X}_t = \hat{X}_{t_0}, \forall t \in \mathbb{D}_{k,T} \cap [t_0, t_1]\big) \\
&= \lim_k \mathbb{P}\big(X_t^{(\varepsilon)} = X_{t_0}^{(\varepsilon)}, \forall t \in \mathbb{D}_{k,T} \cap [t_0, t_1]\big) \\
&= \mathbb{P}\big(X_t^{(\varepsilon)} = X_{t_0}^{(\varepsilon)}, \forall t \in [t_0, t_1]\big) \\
&= \mathbb{E}\bigg[\exp\bigg(-\int_{t_0}^{t_1} R_t\big(X_{t_0}^{(\varepsilon)}\big)\, dt\bigg)\bigg] \\
&\geq \exp\big(-\overline{K}(t_1 - t_0)\big),
\end{aligned}$$

for some constant $\overline{K}$. In particular, on a set Ω_1 of probability 1, $\hat{X} : \mathbb{D}_T \mapsto \mathbb{R}$ is piecewise constant and has finitely many jumps on any closed sub-interval of $(0, T]$.

On Ω_1 let $X : (0, T] \mapsto \mathbb{R}$ be defined by $X_t = \lim_{s \downarrow t, s \in \mathbb{D}_T} \hat{X}_s$, and on the null set Ω_1^c let $X_t = \overline{\mu}$. Then every path of X is piecewise constant and right-continuous on $(0, T]$ and has finitely many jumps on any closed sub-interval of $(0, T]$.

Suppose either that $K(t)$ is bounded on $(0, T)$ or that $\int \mu_t(dx)|x - \overline{\mu}| \to 0$. In the former case, let $X_0 = \lim_{s \downarrow 0, s \in \mathbb{D}_T} \hat{X}_s$ (which exists since $\hat{X}$ has only finitely many jumps in $(0, T]$ if K is bounded) and in the latter case, let $X_0 = \overline{\mu}$.

It remains to check that X has the desired marginals, but for $t > 0$ this is immediate given Assumption 2.2(e):

$$\mathbb{P}(X_t \in B) = \lim_{s \downarrow t, s \in \mathbb{D}_T} \mathbb{P}(\hat{X}_s \in B) = \lim_{s \downarrow t, s \in \mathbb{D}_T} \mu_s(B) = \mu_t(B).$$

When $t = 0$, if $K(t)$ is bounded on $(0, T)$ then the same argument applies, otherwise $X_0 \sim \mu_0 \sim \delta_{\overline{\mu}}$ by definition.

The fact that X is a martingale (relative to its own natural filtration) follows as in Oleszkiewicz [31].

3.2. *Extension to* $[0, T]$ *via time reversal.*   Under some extra conditions, it is possible to sketch an alternative and more direct construction of a process on $[0, T]$. In particular in this section, in addition to the hypotheses of Assumptions 2.1, 2.2 and 2.3 we assume that the family of marginals $(\mu_t)_{t \in [0,T]}$ is such that $\lim_{t \downarrow 0} U(t, x) = |x - \overline{\mu}|$ (and then $\mu_0 \sim \delta_{\overline{\mu}}$) and further that the Radon–Nykodym derivative $\lambda_t(dy)/\mu_t(dy)$ is bounded in y by a function $\tilde{K}(t)$ which is bounded on every interval of the form $[\varepsilon, T]$ with $\varepsilon > 0$.

Throughout this section, we assume T is fixed.

Temporarily fix ε. Using the ideas of the previous section, we can construct $(X_t^{(\varepsilon)})_{\varepsilon \leq t \leq T}$ such that $X_t^{(\varepsilon)} \sim \mu_t$. Now let $\tilde{X}^{(\varepsilon)} = (\tilde{X}_s^{(\varepsilon)})_{0 \leq s \leq T - \varepsilon}$ be given by $\tilde{X}_s^{(\varepsilon)} = X_{T-s}^{(\varepsilon)}$. Then $\tilde{X}^{(\varepsilon)}$ is the time-reversal of $X^{(\varepsilon)}$ and $\tilde{X}_s^{(\varepsilon)} = X_{T-s}^{(\varepsilon)} \sim \mu_{T-s} := \tilde{\mu}_s$. Indeed, the process $\tilde{X}$ is a pure-jump Markov process with state-dependent jump intensities $\tilde{R}_s(y)$ and conditional jump measures $\tilde{\pi}_s^y$ where

$$\tilde{R}_s(y) := \frac{\lambda_{T-s}(dy)}{\mu_{T-s}(dy)}, \qquad \tilde{\pi}_s^y(dx) := \frac{\pi_{T-s}(dx, dy)}{\lambda_{T-s}(dy)} = \frac{\gamma_{T-s}(dx)\pi_{T-s}^x(dy)}{\lambda_{T-s}(dy)}.$$

In particular, $\tilde{X}^{(\varepsilon)}$ does not depend on ε in the sense that if $\tilde{X}_0^{(\varepsilon)} = \tilde{X}_0^{(\delta)}$ then $\tilde{X}^{(\varepsilon)}$ and $\tilde{X}^{(\delta)}$ are pathwise identical on the set $[0, (T - \delta) \wedge (T - \varepsilon)]$ on which they are both defined.

Hence, we may drop the superscript. Let $\tilde{X}$ be the pure-jump Markov process with initial law $\tilde{X}_0 \sim \tilde{\mu}_0 = \mu_T$ defined using $\tilde{R}_s$ and $\tilde{\pi}_s^y$ as in the Section 2.2, except that here we take $\tilde{X}$ to be left-continuous. (Note that $\tilde{\pi}_s^y$ does not respect the martingale property, so that the process $\tilde{X}$ will not be a martingale.) Then $\tilde{X}$ is defined on $[0, T)$, and $\tilde{X}_s \sim \tilde{\mu}_s = \mu_{T-s}$. Now set $\overline{X}_0 \sim \delta_{\overline{\mu}}$ and $\overline{X}_t = \tilde{X}_{T-t}$. Then $\overline{X} = (\overline{X}_t)_{0 \leq t \leq T}$ is well-defined on $[0, T]$ and $\overline{X}_t \sim \mu_t$. Further, $\overline{X}$ agrees with $X^{(\varepsilon)}$ on $[\varepsilon, T]$ and so for any $0 < s < t$ we can take $\varepsilon < s$ so that $\mathbb{E}[\overline{X}_t | \overline{X}_s] = \mathbb{E}[X_t^{(\varepsilon)} | X_s^{(\varepsilon)}] = X_s^{(\varepsilon)} = \overline{X}_s$ and $\overline{X}$ inherits the martingale property from $X^{(\varepsilon)}$.

EXAMPLE 3.1.   Suppose $\mu_t \sim e^{-t}\delta_0 + (1 - e^{-t})U[-1, 1]$. Then $\gamma_t \sim e^{-t}\delta_t$ and $\lambda_t \sim e^{-t}U[-1, 1]$. We find that for $y \in (-1, 1) \setminus \{0\}$, $\tilde{R}_s(y) = \frac{e^{-(T-s)}}{1-e^{-(T-s)}}$ and $\tilde{\pi}_s^y = \delta_0$.

Then the time reversed process is well-defined on $[0, T]$. The process starts at 0 with probability e^{-T} and if it starts at 0 then it remains constant at zero throughout. Otherwise it starts at a nonzero point chosen according to the uniform distribution on $[-1, 0) \cup (0, 1]$, and stays constant until the time of the first event of a time-inhomogeneous Poisson process with rate $\frac{e^{-(T-s)}}{1-e^{-(T-s)}}$, at which point it jumps to zero. Zero acts as an absorbing state.

In forward time, the process starts at 0 and jumps to a uniformly distributed location on $[-1, 1] \setminus \{0\}$ at rate 1.

EXAMPLE 3.2. Suppose that for $t \in [0, T]$, $\mu_t \sim U[-t, t]$. In this case, λ_{T-t} is not absolutely equivalent with respect to μ_{T-t}, and all calculations for this example should be viewed as formal calculations. (Nonetheless, this example is sufficiently concrete that it is easy to verify that the proposed processes have all the right properties.) We find $\tilde{R}_s(y) = 0$ except at $y = \pm(T - s)$, and at those points it is infinite. The interpretation is that $\tilde{X}$ only jumps when $\tilde{X}_s = T - s$. Conditional on their being a jump, we find $\tilde{\pi}_s^{T-s}(dx) = \frac{(T-s+x)}{2(T-s)^2} dx$ and $\tilde{\pi}_s^{-(T-s)}(dx) = \frac{(T-s-x)}{2(T-s)^2} dx$.

The process $\tilde{X}$ can be described as follows. The initial value x_0 is uniform on $[-T, T]$. The process remains constant until $\tau_1 = T - |x_0|$. At time τ_1 the process jumps to a new point x_1 in the interval $(-(T - \tau_1), T - \tau_1)$ chosen according to the density $f(x_1) = \frac{(T-|x_0|+x_1 \, \mathrm{sgn}(x_0))}{2(T-|x_0|)^2}$. The process can then be constructed inductively. Suppose that after k jumps and at time s the process is at x_k where necessarily $x_k \in [-(T - s), T - s]$. The process remains constant until $\tau_{k+1} = T - |x_k|$ at which point the process jumps to a new point in $(-|x_k|, |x_k|) = (-(T - \tau_{k+1}), T - \tau_{k+1})$ chosen with density $\frac{(T-|x_k|+x_{k+1} \, \mathrm{sgn}(x_k))}{2(T-|x_k|)^2}$.

4. Minimising the expected total variation.

In this section, we want to explain how the ideas of Section 2 can be used to construct an extremal process consistent with a given set of marginals under an additional assumption on the relationship between the measures γ_t and λ_t. In particular, the process we construct will have smallest expected total variation amongst all processes consistent with a given set of marginals. As an application, we construct the fake Brownian motion with smallest expected total variation.

Our analysis is motivated by the martingale coupling constructed in Hobson and Klimmek [21]. A key assumption in the construction of that paper is that mass is moved using a martingale transport from an interval in the centre to the tails. We make a similar assumption, which is augmented with some further regularity conditions. A similar condition to Assumption 4.1(a) is stated in Henry-Labordère and Touzi [18] as the difference in the cumulative distribution functions (of λ_t and γ_t) having a unique local maximum.

ASSUMPTION 4.1. (a) (Dispersion assumption) There exists a family of open intervals $(E_t)_{t \in \mathbb{T}}$ such that $\mathrm{supp}(\gamma_t) \subseteq E_t$ and $\mathrm{supp}(\lambda_t) \subseteq E_t^c$. Moreover, the intervals $(E_t)_{t \in \mathbb{T}}$ are increasing: for $s < t$ we have $E_s \subseteq E_t$;

(b) (Regularity) $U(t, x) \in C^{1,2}$ and $Q_t(x)$, $Q_t'(x)$ and $\Gamma_t'(x)$ are continuous in t for all x.

(c) (Strict positivity of densities on their domains) For each $t \in \mathbb{T}$ $\dot{\rho}(t, x) < 0$ for $x \in E_t$ [and then $E_t = (\ell_\gamma(t), r_\gamma(t))$] and $\dot{\rho}(t, x) > 0$ for $x \in (\ell_\lambda(t), \ell_\gamma(t)) \cup (r_\gamma(t), r_\lambda(t))$;

(d) (Finite activity) $\int(|x| + 1)|\dot{\rho}(t, x)| \, dx < \infty$ for each $t \in \mathbb{T}$ and $\int_{t \in \mathbb{T}} dt \times \int(|x| + 1)|\dot{\rho}(t, x)| \, dx < \infty$.

Consider the Brownian motion (Example 2.1) and exponential Brownian motion (Example 2.2) examples. For the former, $E(t) = (-\sqrt{t}, \sqrt{t})$ and $I_t = \mathbb{R}$. For the latter, $E(t) = (\exp(-\sqrt{t + t^2/4}), \exp(\sqrt{t + t^2/4}))$ and $I_t = (0, \infty)$. In both cases, all the elements of Assumption 4.1 are satisfied. However, in the uniform example (Example 2.3) and in Example 2.4, we find $U(t, x) \notin C^{1,2}$, and these examples do not satisfy the regularity conditions (b) and (c) of Assumption 4.1. Nonetheless, the dispersion assumption [Assumption 4.1(a)] is satisfied in Example 2.3: mass is moving from a central region to the tails and there exists an open interval $E_t = (\ell_\gamma(t), r_\gamma(t)) = (-t, t)$ such that $\dot\rho(t, x) < 0$ on E_t, and $\mathrm{supp}(\lambda_t) \subseteq E_t^c$. In Example 2.4, the spirit of Assumption 4.1(a) is satisfied in the sense that mass is moving from a central region to the tails, but in this case $E_t = \{0\}$ is not open. For these examples, it is possible to prove optimality directly; see the discussion in Example 4.2 below.

4.1. *Construction of the martingale transport for fixed t.* In this section, we will consider t to be fixed and generally write t as a subscript. Later we will consider t as a parameter.

Our goal is to show how to write down a martingale transport $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$ under Assumption 4.1. In particular, the idea is that π_t has a decomposition $\pi_t(dx, dy) = \gamma_t(dx)\pi_t^x(dy)$ where π_t^x has a binomial distribution:

$$
(7) \qquad \pi_t^x = \frac{b_t(x) - x}{b_t(x) - a_t(x)}\delta_{a_t(x)} + \frac{x - a_t(x)}{b_t(x) - a_t(x)}\delta_{b_t(x)}.
$$

Thus, for this candidate martingale transport, mass at x moves down to $a_t(x)$ or up to $b_t(x)$, where $a_t(x) < x < b_t(x)$, and the probabilities of up and down moves are such that the martingale property is satisfied. Our goal is to explain how to choose $a_t(x)$ and $b_t(x)$ such that $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$. A key element of the construction will be that $a_t(x)$ and $b_t(x)$ are decreasing functions of x.

Following Hobson and Klimmek [21] for $x \in E_t$ and $y > x$, define $\mathcal{Q}_{t,x}(y) = Q_t(x) + Q_t'(x)(y - x) - Q_t(y)$. For $x \in E_t$, define

$$
(8) \qquad \phi_t(x) = \max_{\alpha < x < \beta}\left\{\frac{\mathcal{Q}_{t,x}(\beta) - Q_t(\alpha)}{\beta - \alpha}\right\}
$$

and let the maximum be attained at the pair $(\alpha = a_t(x), \beta = b_t(x))$. Under Assumption 4.1 (and especially the dispersion and strict positivity elements), Q_t is strictly convex and increasing on $(\ell_\lambda(t), \ell_\gamma(t))$, strictly concave on $(\ell_\gamma(t), r_\gamma(t))$ and strictly convex and decreasing on $(r_\gamma(t), r_\lambda(t))$. This guarantees the existence and uniqueness of optimisers $a_t(x)$ and $b_t(x)$ in the right-hand side of (8) with $a_t(x) < \ell_\gamma(t)$ and $b_t(x) > r_\gamma(t)$.

PROPOSITION 4.1. $a_t(\cdot)$ *is continuously differentiable in x on E_t and satisfies*

$$
a_t'(x)\dot\rho(t, a_t(x)) = \frac{b_t(x) - x}{b_t(x) - a_t(x)}\dot\rho(t, x).
$$

Similarly $b_t(\cdot)$ is continuously differentiable on E_t and satisfies

$$b_t'(x)\dot{\rho}\big(t, b_t(x)\big) = -\frac{x - a_t(x)}{b_t(x) - a_t(x)}\dot{\rho}(t, x).$$

PROOF. We prove the result for a_t, the proof for b_t being similar.

The construction of a_t, b_t given in Hobson and Klimmek [21] is a martingale transport of probability measures, but generalises easily to pairs of measures with the same total mass which are increasing in convex order. By construction, $a_t(x) = (Q_t')^{-1} \circ \phi_t \circ \gamma_t((-\infty, x])$. By the continuity and strict positivity of $\dot{\rho}(t, x)$ on $(\ell_\lambda(t), \ell_\gamma(t))$, $(Q_t')^{-1}$ is continuously differentiable, as is $\gamma_t((-\infty, x])$. Further, by Lemma 5.1 and the argument after Lemma 5.2 of Hobson and Klimmek [21], ϕ_t is continuously differentiable. Hence, $a_t(\cdot)$ is continuously differentiable. The expression for the derivative is taken from Section 3.1 of [21]. Note that [21] covers the case of general probability measures [satisfying Assumption 4.1(a)] including measures with atoms. □

COROLLARY 4.1. *Define π_t by $\pi_t(dx, dy) = \gamma_t(dx)\pi_t^x(dy)$ where π_t^x is as given in (7). Then $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$ and $\pi_t^x(B)$ is measurable in x for all measurable sets $B \subseteq I$.*

PROOF. Suppose H is a bounded, continuous test function. Then, using Proposition 4.1 in the second line, and a change of variables in the third,

$$\int_x \int_y \gamma_t(dx)\pi_t^x(dy)H(y)$$

$$= \int_{E_t} \dot{\rho}(t, x)^- dx\left(\frac{(b_t(x) - x)}{b_t(x) - a_t(x)}H\big(a_t(x)\big) + \frac{(x - a_t(x))}{b_t(x) - a_t(x)}H\big(b_t(x)\big)\right)$$

$$= -\int_{E_t} H\big(a_t(x)\big)a_t'(x)\dot{\rho}\big(t, a_t(x)\big) dx - \int_{E_t} H\big(b_t(x)\big)b_t'(x)\dot{\rho}\big(t, b_t(x)\big) dx$$

$$= \int_{(\ell_\lambda(t), \ell_\gamma(t))} H(y)\dot{\rho}(t, y) dy + \int_{(r_\gamma(t), r_\lambda(t))} H(y)\dot{\rho}(t, y) dy$$

$$= \int_{I_t} \lambda_t(dy)H(y)$$

and it follows that $\pi_t \in \mathcal{M}(\gamma_t, \lambda_t)$. Measurability in x of $\pi_t^x(B)$ is clear from the monotonicity of a_t and b_t. □

Now we introduce some auxiliary variables which will play a role in the evaluation of the minimal expected total variation. Fix $x_0 \in \bigcap_{t \in \mathbb{T}} E_t$. For $x \in \overline{E}_t = [\ell_\gamma(t), r_\gamma(t)]$, define

$$\theta_t(x) = \int_{x_0}^x \frac{2}{b_t(z) - a_t(z)} dz, \qquad \psi_t(x) = \int_{x_0}^x \frac{2(x - x_0) - b_t(z) - a_t(z)}{b_t(z) - a_t(z)} dz.$$

Extend these definitions to $x \notin \overline{E}_t$ by setting

$$\theta_t(x) = \begin{cases} \theta_t(a_t^{-1}(x)), & x < \ell_\gamma(t), \\ \theta_t(b_t^{-1}(x)), & x > r_\gamma(t), \end{cases}$$

$$\psi_t(x) = \begin{cases} \psi_t(a_t^{-1}(x)) + (a_t^{-1}(x) - x)(1 - \theta_t(a_t^{-1}(x))), & x < \ell_\gamma(t), \\ \psi_t(b_t^{-1}(x)) + (x - b_t^{-1}(x))(1 + \theta_t(b_t^{-1}(x))), & x > r_\gamma(t). \end{cases}$$

Set $L_t(x, y) = L_{\psi_t, \theta_t}(x, y) = |y - x| + \psi_t(x) + \theta_t(x)(y - x) - \psi_t(y)$.

PROPOSITION 4.2 (Hobson and Klimmek [21]). $L_t(x, y) \geq 0$, *with equality for* $y \in \{a_t(x), x, b_t(x)\}$. *Further,* $\psi_t(x) \leq |x - x_0|$ *and* ψ_t *is convex on* E_t *and concave on* $(\ell_\lambda(t), \ell_\gamma(t))$ *and* $(r_\gamma(t), r_\lambda(t))$.

PROOF.   The only part of this result which is not contained in Theorem 4.5 of Hobson and Klimmek [21] is the fact that $\psi_t(x) \leq |x - x_0|$. But

$$0 \leq L_t(x_0, x) = |x - x_0| + \psi_t(x_0) + \theta_t(x_0)(x - x_0) - \psi_t(x)$$
$$= |x - x_0| - \psi_t(x). \qquad \square$$

COROLLARY 4.2.   *For any random variables* $X, Y \in L^1$ *with* $\mathbb{E}[(Y - X)|X] = 0$, *we have* $\mathbb{E}[\psi_t(Y) - \psi_t(X)] \leq \mathbb{E}[|Y - X|]$.

4.2. *Continuity in* t.   Now we consider the dependence of the various quantities on t. Where appropriate we switch to writing t as an argument rather than as a subscript; for example, we now write $a(t, x)$ in place of $a_t(x)$.

LEMMA 4.1. $a(t, x)$ *and* $b(t, x)$ *are continuous in* t. *Further,* $\psi(t, x)$ *is continuous in* t.

PROOF.   We have the representation $a(t, x) = (Q_t')^{-1} \circ \phi_t \circ \gamma_t((-\infty, x])$. By assumption, Q_t' and Γ_t' are continuous in t. Since $\phi(t, x)$ is continuous in x, continuity of a will follow if $\phi(t, x)$ is continuous in t.

Recall (8) and the fact that $\phi(t, x)$ is the slope from $(a(t, x), Q(t, a(t, x)))$ to $(b(t, x), \mathcal{Q}_{t,x}(b(t, x)))$. Then, using continuity of Q_t and Q_t' in t [Assumption 4.1(b)]

$$\phi(t', x) = \sup_{\alpha < x < \beta} \frac{\mathcal{Q}_{t',x}(\beta) - Q(t', \alpha)}{\beta - \alpha} \geq \frac{\mathcal{Q}_{t',x}(b(t, x)) - Q(t', a(t, x))}{b(t, x) - a(t, x)}$$

$$\xrightarrow{t' \to t} \frac{\mathcal{Q}_{t,x}(b(t, x)) - Q(t, a(t, x))}{b(t, x) - a(t, x)} = \phi(t, x).$$

Conversely, given $\eta > 0$, for t' sufficiently close to t we have $Q(t, y) > Q(t', y) - \eta$ and $\mathcal{Q}_{t,x}(y) < \mathcal{Q}_{t,x}(y) + \eta(1 + (y - x))$, and then if $r_\gamma(t') - \ell_\gamma(t') > r_\gamma(t-) - \ell_\gamma(t-) - \eta$,

$$\phi(t', x) = \sup_{\alpha < x < \beta} \frac{\mathcal{Q}_{t',x}(\beta) - Q(t', \alpha)}{\beta - \alpha} < \sup_{\alpha < x < \beta} \frac{\mathcal{Q}_{t,x}(\beta) - Q(t, \alpha)}{\beta - \alpha} + \eta + \frac{2\eta}{\beta - \alpha}$$

$$< \phi(t, x) + \eta + \frac{2\eta}{r_\gamma(t-) - \ell_\gamma(t-) - \eta}.$$

Hence, for fixed x, $\phi(t, x)$ is continuous in t.

Continuity of $b(t, x)$ in t is similar, from which continuity of $\psi(t, x)$ follows. □

COROLLARY 4.3. *Suppose Assumptions 2.1, 2.2 and 4.1 hold. Then $X_t \sim \mu_t$ for $t \in \overline{\mathbb{T}}$.*

PROOF. It is clear that under Assumption 4.1 $R_t(x)$ and π_t^x are such that the measurability requirements of Assumption 2.2(d) and Assumption 2.3 are satisfied and hence by Theorem 2.1 that $X_t \sim \mu_t$. □

4.3. *Minimising expected total variation*: *The primal approach.* The idea now is to show that, under Assumption 4.1, the process we have constructed has the smallest possible expected total-variation in the class of martingales with the specified marginals.

Suppose $\varepsilon \geq 0$ and let $\overline{\mathbb{T}} = [\varepsilon, T]$. Let $P^{\overline{\mathbb{T}}} = \{t_0, t_1, \ldots, t_N\}$ be a partition of $\overline{\mathbb{T}}$ of dimension N, that is, a vector of length $N + 1$ with $t_0 = \varepsilon$, $t_k > t_{k-1}$ and $t_N = T$. We say $\mathcal{P}^{\overline{\mathbb{T}}} = (P_n^{\overline{\mathbb{T}}})_{n \geq 1}$ is a dense sequence of nested partitions of $[\varepsilon, T]$, if $P_n^{\overline{\mathbb{T}}} = \{t_0^n, t_1^n, \ldots, t_{N(n)}^n\}$ is a partition for each n, $P_n^{\overline{\mathbb{T}}} \subseteq \{t_0^{n+1}, t_1^{n+1}, \ldots, t_{N(n+1)}^{n+1}\} = P_{n+1}^{\overline{\mathbb{T}}}$ and $\lim_{n \uparrow \infty} \max_{1 \leq k \leq N(n)} |t_k^n - t_{k-1}^n| = 0$.

Let $P_{U,n}^{\overline{\mathbb{T}}} = \{t_0^{U,n}, t_1^{U,n}, \ldots, t_{2^n}^{U,n}\}$ be the uniform partition of $[\varepsilon, T]$ of dimension 2^n in which $t_k^{U,n} = \varepsilon + (T - \varepsilon)k2^{-n}$, and let $\mathcal{P}_U^{\overline{\mathbb{T}}} = (P_{U,n}^{\overline{\mathbb{T}}})_{n \geq 1}$. Then $\mathcal{P}_U^{\overline{\mathbb{T}}}$ is a dense sequence of nested partitions.

DEFINITION 4.1. Let $f : [\varepsilon, T] \mapsto \mathbb{R}$ be any function. The total variation of f along a partition $P^{\overline{\mathbb{T}}}$ is

$$\mathcal{V}(P^{\overline{\mathbb{T}}}, f) = \sum_{k=1}^{N} |f(t_k) - f(t_{k-1})|.$$

The total variation of f along a dense sequence of nested partitions $\mathcal{P}^{\overline{\mathbb{T}}}$ is

$$\mathcal{V}(\mathcal{P}^{\overline{\mathbb{T}}}, f) = \lim_n \sum_{k=1}^{N(n)} |f(t_k^n) - f(t_{k-1}^n)|.$$

The fact that the sequence is nested means that the limit exists.

The total variation of f is $\mathcal{V}^{\overline{\mathbb{T}}}(f) = \sup \mathcal{V}(\mathcal{P}^{\overline{\mathbb{T}}}, f)$ where the supremum is taken over dense sequences of nested partitions. Note that $\mathcal{V}^{\overline{\mathbb{T}}}(f) \geq \mathcal{V}(\mathcal{P}_U^{\overline{\mathbb{T}}}, f)$.

DEFINITION 4.2. Let $Z = (Z_t)_{t \in \overline{\mathbb{T}}}$ be a stochastic process. Define the expected total variation $V^{\overline{\mathbb{T}}}(Z)$ of Z by $V^{\overline{\mathbb{T}}}(Z) = \mathbb{E}[\mathcal{V}^{\overline{\mathbb{T}}}(Z)] = \mathbb{E}[\sup \mathcal{V}(P^{\overline{\mathbb{T}}}, Z)]$, where the supremum is taken over partitions which may depend upon the realisation of Z.

For the sequence of uniform partitions $\mathcal{P}_U^{\overline{\mathbb{T}}}$, define the expected total variation along the uniform partition $V(\mathcal{P}_U^{\overline{\mathbb{T}}}, Z)$ by $V(\mathcal{P}_U^{\overline{\mathbb{T}}}, Z) = \mathbb{E}[\lim_n \mathcal{V}(P_{U,n}^{\overline{\mathbb{T}}}, Z)]$. By the final remark in the previous definition, $V^{\overline{\mathbb{T}}}(Z) \geq V(\mathcal{P}_U^{\overline{\mathbb{T}}}, Z)$.

Suppose $\varepsilon > 0$ and $\mathbb{T} = (\varepsilon, T)$.

THEOREM 4.1. *Suppose Assumptions* 2.1, 2.2 *and* 4.1 *hold. Let Y be any martingale with marginals $(\mu_t)_{t \in \overline{\mathbb{T}}}$. Then*

$$(9) \quad V^{\overline{\mathbb{T}}}(Y) \geq V(\mathcal{P}_U^{\overline{\mathbb{T}}}; Y) \geq \int_{\mathbb{T}} dt \int dx \psi_t(x) \dot{\rho}(t, x) = \int_{\mathbb{T}} dt \int \psi_t(x) q_t(dx).$$

PROOF. We only need to prove the second inequality. Let $[t]_n = \max\{t_k \in P_{U,n}^{\overline{\mathbb{T}}}; t_k \leq t\}$. Then $[t]_n \uparrow t$ and by Corollary 4.2, for $s < t$ we have, $\mathbb{E}[|Y_t - Y_s|] \geq \mathbb{E}[\psi(s, Y_t) - \psi(s, Y_s)]$,

$$
\begin{aligned}
V(\mathcal{P}_{U,n}^{\overline{\mathbb{T}}}, Y) &= \mathbb{E}\left[\sum_{k=1}^{2^n} |Y_{t_k^{U,n}} - Y_{t_{k-1}^{U,n}}|\right] \\
&\geq \sum_{k=1}^{2^n} \mathbb{E}[(\psi(t_{k-1}^{U,n}, Y_{t_k^{U,n}}) - \psi(t_{k-1}^{U,n}, Y_{t_{k-1}^{U,n}}))] \\
&= \sum_{k=1}^{2^n} \int_I dx \psi(t_{k-1}^{U,n}, x)[\rho(t_k^{U,n}, x) - \rho(t_{k-1}^{U,n}, x)] \\
&= \int dx \int_{\mathbb{T}} \psi([t]_n, x) \, dt \dot{\rho}(t, x) \, dt \\
&\longrightarrow \int dx \int_\varepsilon^T dt \psi(t, x) \dot{\rho}(t, x) \, dt,
\end{aligned}
$$

where we use the fact that $\psi([t]_n, x) \to \psi(t, x)$, $\psi(t, x) \leq |x - x_0|$, $\int_{\mathbb{T}} \int (|x| + 1)|\dot{\rho}(t, x)| \, dx < \infty$ and dominated convergence. $\square$

It remains to show that this bound is best possible, or equivalently for the process constructed in this section there is equality in (9). This follows from the fact that $L_t(x, y) \equiv 0$ for $y \in \{a(t, x), x, b(t, x)\}$.

THEOREM 4.2. *Let X be the process constructed in Section 2 using the family of martingale transports given in (7). Then*

$$V^{\overline{\mathbb{T}}}(X) = \int_{\mathbb{T}} dt \int dx \psi(t,x)\dot{\rho}(t,x). \tag{10}$$

Hence, in the class of martingales with marginals $(\mu_t)_{t \in \overline{\mathbb{T}}}$, X minimises the expected total variation.

PROOF. First, note that from easy properties about zero mean random variables taking only two values

$$\mathbb{E}[|X_{t+dt} - X_t||X_t = x] = \text{Rate of jumps} \times dt \times \mathbb{E}[|\Delta X_t||\text{Jump at } t, X_t = x]$$

$$= R_t(x) \frac{2(b(t,x) - x)(x - a(t,x))}{b(t,x) - a(t,x)} dt.$$

Recall that by construction X is right-continuous, and almost surely has only finitely many jumps in $\overline{\mathbb{T}} = [\varepsilon, T]$. We may assume that the jump times are elements of the partition and then

$$V^{\overline{\mathbb{T}}}(X) = \mathbb{E}\left[\sum_{\varepsilon < t \le T} |\Delta X_t|\right]$$

$$= \int_{\varepsilon}^{T} dt \int_{E_t} \rho(t,x) R_t(x) \frac{2(b(t,x) - x)(x - a(t,x))}{b(t,x) - a(t,x)} dx$$

$$= \int_{\varepsilon}^{T} dt \int_{E_t} 2\dot{\rho}(t,x)^{-} \frac{(b(t,x) - x)(x - a(t,x))}{b(t,x) - a(t,x)} dx.$$

The result will follow if we can show that

$$\int dx \psi_t(x)\dot{\rho}(t,x) = 2\int_{E_t} \dot{\rho}(t,x)^{-} \frac{(b(t,x) - x)(x - a(t,x))}{b(t,x) - a(t,x)} dx. \tag{11}$$

We have

$$\int dx \psi_t(x)\dot{\rho}(t,x) dx = -\int_{E_t} \psi(t,x)\dot{\rho}(t,x)^{-} dx + \int_{E_t^c} \psi(t,y)\dot{\rho}(t,y) dx. \tag{12}$$

Then, from the fact that $L_t(x,y) = 0$ for $y \in \{a_t(x), x, b_t(x)\}$, for $y = a(t,x)$ we have $\psi_t(y) = \psi(t,y) = \psi(t,x) + (x - y)(1 - \theta(t,x))$ and for $y = b(t,x)$ we have $\psi_t(y) = \psi(t,y) = \psi(t,x) + (x - y)(-1 - \theta(t,x))$. Then, by Proposition 4.1,

$$\int_{y \le \ell_\gamma(t)} dy \psi(t,y)\dot{\rho}(t,y)$$

$$= -\int_{E_t} dx a'(t,x)\psi(t,a(t,x))\dot{\rho}(t,a(t,x))$$

$$= \int_{E_t} dx [\psi(t,x) + (x - a(t,x)) - (x - a(t,x))\theta(t,x)]$$

$$\times \frac{b(t,x) - x}{b(t,x) - a(t,x)}\dot{\rho}(t,x)^{-}.$$

Similarly,

$$\int_{y \geq r_\gamma(t)} dy\, \psi(t, y)\dot\rho(t, y)$$

$$= \int_{E_t} dx\big[\psi(t, x) - (x - b(t, x)) - (x - b(t, x))\theta(t, x)\big]$$

$$\times \frac{x - a(t, x)}{b(t, x) - a(t, x)} \dot\rho(t, x)^-.$$

Substituting these last two expressions into (12), we find that the terms involving $\psi(t, x)$ and $\theta(t, x)$ cancel and (11) follows. $\quad\square$

REMARK 4.1.   Let $H_t = \int \psi(t, x)\rho(t, x)\,dx$. Suppose that $\int_{t \in \mathbb{T}} dt \int_x dx \times |\dot\psi(t, x)|\rho(t, x) < \infty$ and note that Assumption 4.1(d) implies $\int_{t \in \mathbb{T}} dt \times \int_x dx |\psi(t, x)||\dot\rho(t, x)| < \infty$. Then

$$H_T = H_\varepsilon + \int_\varepsilon^T \dot H_t\, dt = H_\varepsilon + \int_\varepsilon^T dt \int \dot\psi(t, x)\rho(t, x)\, dx$$

$$+ \int_\varepsilon^T dt \int \psi(t, x)\dot\rho(t, x)\, dx,$$

and hence if $Y = (Y_t)_{t \in \overline{\mathbb{T}}}$ is a martingale with marginals $(\mu_t)_{t \in \overline{\mathbb{T}}}$ then

$$V^{\overline{\mathbb{T}}}(Y) \geq \int \psi_T(x)\rho(T, x)\, dx - \int \psi_\varepsilon(x)\rho(\varepsilon, x)\, dx$$

(13)
$$- \int_\varepsilon^T dt \int \dot\psi(t, x)\rho(t, x)\, dx.$$

Now consider the case where $\varepsilon = 0$ and $\mathbb{T} = (0, T)$. Results in this case follow on taking the limit $\varepsilon \downarrow 0$.

COROLLARY 4.4.   *Suppose* $\int_0^T (|x| \vee 1)\xi_t(dx) < \infty$. *Let* $Y = (Y_t)_{t \in [0,T]}$ *be any martingale with marginals* $(\mu_t)_{t \in [0,T]}$, *and let* X *be the martingale constructed using the martingale transports in* (7). *Then* $V^{[0,T]}(Y) \geq V^{[0,T]}(X) = \int_0^T dt \int dx\, \psi(t, x)\dot\rho(t, x)$ *and in the class of martingales with marginals* $(\mu_t)_{t \in [0,T]}$, X *has minimal expected total variation.*

Suppose further that $\int \psi_\varepsilon(x)\rho(\varepsilon, x)\, dx \overset{\varepsilon \downarrow 0}{\to} 0$ *and* $\int_0^T dt \int_x dx |\dot\psi(t, x)|\rho(t, x) < \infty$. *Then*

$$V^{[0,T]}(Y) \geq V^{[0,T]}(X) = \int_0^T dt \int \psi(t, x)q_t(dx)$$

(14)
$$= \int \psi_T(x)\mu_T(dx) - \int_0^T dt \int \dot\psi(t, x)\mu_t(dx).$$

4.4. *The dual problem and a pathwise representation.* The inequalities (9) and (13) are lower bounds on expected total variation. It is well known from the literature on model-free option pricing (see, e.g., Hobson [20, 23] and Beiglböck and Nutz [6]) that such bounds are often related to pathwise martingale inequalities. Again the key to this result is the inequality $L_t(x, y) \geq 0$.

Note that in the following theorem we do not make any assumption that Z is a martingale, or that Z has the correct marginals.

THEOREM 4.3. *Suppose $\mathbb{T} = (\varepsilon, T)$ with $\varepsilon > 0$. Suppose $\dot{\psi}(t, x)$ and $\theta(t, x)$ are continuous and bounded on $\mathbb{T} \times I$ and suppose $L_t(x, y) = L_{\psi_t, \theta_t}(x, y) \geq 0$ for all $(x, y) \in I \times I$ and all $t \in \mathbb{T}$. Then for all paths $Z(\omega) = (Z_u(\omega))_{u \in \overline{\mathbb{T}}}$ of finite variation*

$$\mathcal{V}^{\overline{\mathbb{T}}}(Z(\omega)) \geq \psi(T, Z_T(\omega)) - \psi(\varepsilon, Z_\varepsilon(\omega)) - \int_\varepsilon^T \dot{\psi}(u, Z_u(\omega))\, du$$

(15)
$$- \int_\varepsilon^T \theta(u, Z_u(\omega))\, dZ_u(\omega).$$

Note that if Z is not of finite variation then the left-hand side of the above inequality is infinite, and in this sense (15) is trivially satisfied. However, the final integral on the right-hand side may not be well-defined in this case.

PROOF OF THEOREM 4.3. Fix ω and suppose $Z = (Z_t(\omega))_{t \in \overline{\mathbb{T}}}$ is of finite variation. We have $L_{\psi_t, \theta_t}(Z_t, Z_{t+h}) \geq 0$, and hence

(16) $\psi(t, Z_{t+h}) - \psi(t, Z_t) - \theta(t, Z_t)(Z_{t+h} - Z_t) \leq |Z_{t+h} - Z_t|.$

Consider the uniform partition $P_{U,n}^{\mathbb{T}} = \{t_k^{U,n}; 0 \leq k \leq 2^n\}$ where $t_k^{U,n} = \varepsilon + k2^{-n}(T - \varepsilon)$ and abbreviate $t_k^{U,n}$ to t_k^n. Then, applying (16) repeatedly,

$$\sum_{k=1}^{2^n} [\psi(t_{k-1}^n, Z_{t_k^n}) - \psi(t_k^n, Z_{t_k^n})] + \sum_{k=1}^{2^n} [\psi(t_k^n, Z_{t_k^n}) - \psi(t_{k-1}^n, Z_{t_{k-1}^n})]$$

$$- \sum_{k=1}^{2^n} \theta(t_{k-1}^n, Z_{t_{k-1}^n})(Z_{t_k^n} - Z_{t_{k-1}^n}) \leq \sum_{k=1}^{2^n} |Z_{t_k^n} - Z_{t_{k-1}^n}|.$$

The second sum on the top line telescopes. By monotonicity, as we take finer and finer uniform partitions, the second term in the second line converges: $\sum_{k=1}^{2^n} |Z_{t_k^n} - Z_{t_{k-1}^n}| \to \mathcal{V}(\mathcal{P}_U^{\overline{\mathbb{T}}}, Z) \leq \mathcal{V}^{\overline{\mathbb{T}}}(Z)$. For the remaining terms, recalling $[t]_n = \max\{t_n^k \in P_n^{\mathbb{T}} : t_n^k \leq t\}$, and setting $[t]^n = \min\{t_k^n \in P_n^{\overline{\mathbb{T}}} : t_k^n \geq t\}$

$$\sum_{k=1}^{2^n} \theta(t_{k-1}^n, Z_{t_{k-1}^n})(Z_{t_k^n} - Z_{t_{k-1}^n}) = \int_\varepsilon^T \theta([u]_n, Z_{[u]_n})\, dZ_u \to \int_\varepsilon^T \theta(u, Z_u)\, dZ_u$$

by bounded convergence, since θ is continuous and bounded and Z is of finite variation, and

$$\sum_{k=1}^{2^n}(\psi(t_k^n, Z_{t_k^n}) - \psi(t_{k-1}^n, Z_{t_k^n})) = \int_\varepsilon^T \dot\psi(u, Z_{[u]^n})\,du \to \int_\varepsilon^T \dot\psi(u, Z_u)\,du$$

again by bounded convergence, using the continuity and boundedness of $\dot\psi$. $\square$

COROLLARY 4.5.   *Suppose $\overline{\mu} \in \bigcap_{0 \le t \le T} E_t$ and $\mu_0 = \delta_{\overline{\mu}}$. If $\dot\psi(t, x)$ and $\theta(t, x)$ are continuous and bounded on $(0, T) \times I$ and if $Z(\omega)$ is of finite variation with $Z_0(\omega) = \overline{\mu}$ then*

(17)
$$\begin{aligned}\mathcal{V}^{[0,T]}&(Z(\omega)) \\ &\ge \psi(T, Z_T(\omega)) - \int_0^T \dot\psi(u, Z_u(\omega))\,du - \int_0^T \theta(u, Z_u(\omega))\,dZ_u(\omega).\end{aligned}$$

4.5. *Self-similar marginals.*   Suppose that in addition to being increasing in convex order, the family $(\mu_t)_{t \in [0,T]}$ is self-similar in the sense that there exists a centred continuous random variable Z and some positive exponent α such that $\mu_t = \mathcal{L}(t^\alpha Z)$. Examples include the Gaussian (Example 2.1) and continuous uniform (Example 2.3) specifications. Note that we must have $\mu_0 \sim \delta_0$.

Let μ_Z be the law of Z, with potential U_Z, and let I_Z with endpoints $\{\ell_Z, r_Z\}$ be the smallest interval such that the support of μ_Z is contained in I_Z. Let ρ_Z be the associated density and suppose that the product $z\rho_Z(z)$ is continuously differentiable on I_Z and that $\rho_Z(z) + z\rho_Z'(z)$ is strictly positive on an open interval $E_Z = (\ell_E, r_E)$ with $\ell_E < 0 < r_E$ and negative on $(\ell_Z, \ell_E) \cup (r_E, r_Z)$. Suppose further that $\int_{I_Z}(1 + |z|^2)(\rho_Z(z) + |\rho_Z'(z)|)\,dz < \infty$. Finally, define $\zeta_Z(z) = -\alpha(\rho_Z(z) + z\rho_Z'(z))$ and suppose $\sup_{\ell_E < z < r_E} \frac{\zeta_Z(z)^-}{\rho_Z(z)} < K_Z$ for some finite constant K_Z.

If $\rho(t, y)$ and $U(t, y)$ are the density and potential of μ_t then we have $\rho(t, y) = t^{-\alpha}\rho_Z(yt^{-\alpha})$ and $U(t, y) = t^\alpha U_Z(yt^{-\alpha})$. Further, if we define $P_Z(y) = \frac{1}{2}\alpha\{U_Z(y) - yU_Z'(y)\}$, then

$$Q(t, y) := \tfrac{1}{2}\dot U(t, y) = t^{\alpha-1}P_Z(yt^{-\alpha}) = \tfrac{1}{2}\alpha t^{\alpha-1}\{U_Z(yt^{-\alpha}) - yt^{-\alpha}U_Z'(yt^{-\alpha})\}$$

is positive since U_Z is the potential of a centred random variable, and then, since $P_Z'' = \xi_Z$,

$$\dot\rho(t, y) = Q''(t, y) = t^{-(\alpha+1)}\zeta_Z(yt^{-\alpha}).$$

It follows that we can define $\gamma_t(dy) = \dot\rho(t, y)^-\,dy$ and $\lambda_t(dy) = \dot\rho(t, y)^+\,dy$, and that $m(t) = \int \dot\rho(t, y)^-\,dy = \frac{1}{t}\int_{I_Z}\zeta(z)^-\,dz$ and $n(t) = t^{\alpha-1}\int_{I_Z} z\zeta_Z(z)^-\,dz$ are both finite and well-defined.

By definition $R_t(y) = \frac{\dot\rho(t, y)^-}{\rho(t, y)} = \frac{1}{t}\frac{\zeta_Z(yt^{-\alpha})^-}{\rho_Z(yt^{-\alpha})} < \frac{1}{t}K_Z$.

For $y \in E_Z$ let $A(y) = A_Z(y)$ and $B(y) = B_Z(y)$ with $A(y) < y < B(y)$ attain the supremum in

$$\sup_{a < y < b} \left\{ \frac{P_Z(y) + (b - y) P_Z'(y) - P_Z(b) - P_Z(a)}{b - a} \right\}.$$

Then, as in Section 4.1 $A : E_Z \mapsto (\ell_Z, \ell_E)$ and $B : E_Z \mapsto (r_E, r_Z)$ are decreasing.

For $y \in \overline{E}_Z$, define

$$\Theta(y) = \int_0^y \frac{2}{B(z) - A(z)} \, dz; \qquad \Psi(y) = \int_0^y \frac{2y - B(z) - A(z)}{B(z) - A(z)} \, dz.$$

Extend these definitions to $y \notin \overline{E}_Z$ by setting

(18)
$$\Theta(y) = \begin{cases} \Theta(A^{-1}(y)), & y < \ell_E, \\ \Theta(B^{-1}(y)), & y > r_E, \end{cases}$$

$$\Psi(y) = \begin{cases} \Psi(A^{-1}(y)) + (A^{-1}(y) - y)(1 - \Theta(A^{-1}(y))), & y < \ell_E, \\ \Psi(B^{-1}(y)) + (y - B^{-1}(y))(1 + \Theta(B^{-1}(y))), & y > r_E. \end{cases}$$

Finally, define $a_t(y) = t^\alpha A(yt^{-\alpha})$, $b_t(y) = t^\alpha B(yt^{-\alpha})$, $\theta_t(y) = \Theta(yt^{-\alpha})$ and $\psi_t(y) = t^\alpha \Psi(yt^{-\alpha})$. Then, for $y \in E_t = (t^\alpha \ell_E, t^\alpha r_E)$,

$$\theta_t(y) = \int_0^y \frac{2}{b_t(z) - a_t(z)} \, dz, \qquad \psi_t(y) = \int_0^y \frac{2y - b_t(z) - a_t(z)}{b_t(z) - a_t(z)} \, dz,$$

and $\psi_t(y) \leq \psi_t(x) + (y - x)\theta_t(x) + |y - x|$.

Then, for any process Y with marginals $(\mu_t)_{t \in [\varepsilon, T]}$,

$$V^{[\varepsilon, T]}(Y) \geq \int_\varepsilon^T dt \int dy \psi_t(y) \dot{\rho}(t, y)$$

$$= \int_\varepsilon^T dt \int dy t^\alpha \Psi(yt^{-\alpha}) t^{-(\alpha+1)} \zeta_Z(yt^{-\alpha})$$

$$= \int_\varepsilon^T dt t^{\alpha-1} \int dz \Psi(z) \zeta_Z(z)$$

$$= \frac{(T^\alpha - \varepsilon^\alpha)}{\alpha} \int dz \Psi(z) \zeta_Z(z)$$

with equality for the process X constructed in Section 2.2 using the martingale transports defined via (7). Letting $\varepsilon \downarrow 0$, we obtain $V^{[0,T]}(Y) \geq \frac{T^\alpha}{\alpha} \int dz \Psi(z) \zeta_Z(z)$.

REMARK 4.2. Note that $\dot{\psi}(t, x) = \alpha t^{\alpha-1} [\Psi(xt^{-\alpha}) - xt^{-\alpha} \Psi'(xt^{-\alpha})] \leq \alpha t^{\alpha-1} J$ where $J = \sup_y |\Psi(y) - y\Psi'(y)|$. Then from the shape of Ψ (concave, then convex, then concave again) we find $J = \max\{\Psi(\ell_E) + (\Theta(\ell_E) + 1)|\ell_E|; \Psi(r_E) + (1 - \Theta(r_E))r_E; \Psi'(\ell_E)\ell_E - \Psi(\ell_E); \Psi'(r_E)r_E - \Psi(r_E)\} < \infty$. Similarly $\sup_x |\theta_t(x)| = \sup_x |\Theta(x)| \leq 2$.

Hence, on each interval of the form $[\varepsilon, T]$, with $\varepsilon > 0$ (15) holds [and if $\varepsilon = 0$ then (17) holds] and we have a pathwise inequality to complement the inequality in expectation.

EXAMPLE 4.1 (Brownian motion). For Brownian motion, there do not exist explicit forms for Θ or Ψ. Nonetheless, we find $V^{[0,T]}(Y) \geq C\sqrt{T}$ where C is a finite constant which can evaluated numerically:

$$C = 2 \int \Psi(y)\zeta_Z(y) \, dy = 2 \int \Psi(y)\big(-y\rho_Z(y)\big)' dy.$$

The inequality is tight; in particular, there exists a fake Brownian motion with finite total variation on $[0, T]$. Moreover, C is bounded above by

$$C \leq 4 \int_1^\infty \Psi(y)\big(-y\rho_Z(y)\big)' dy = 4 \int_1^\infty (z^3 - z)e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} = \sqrt{\frac{32}{\pi}} e^{-1/2}.$$

Here, we use symmetry about zero, the bounds $0 \leq \Psi(y) \leq |y|$ and the fact that $(-y\rho_Z(y))' = (y^2 - 1)e^{-y^2/2}/\sqrt{2\pi}$ is negative on $|y| \leq 1$.

EXAMPLE 4.2. Consider the continuous uniform example (Example 2.3) in which $\mu_t \sim U[-t, t]$. In this case, we can write down values for Θ and Ψ directly. Note that in (19) below we have exploited the fact that when $x \notin [-1, 1]$ the inequality $L_t(x, y) = 0$ is only tight at $y = x$ to give a simpler form for Θ than that given in (18).

Let $\Psi(x) = x^2 \wedge 1$ and $\Theta(x) = xI_{\{-1 < x < 1\}}$. Then, for all $x, y \in \mathbb{R}$,

$$(19) \qquad \Psi(y) \leq \Psi(x) + \Theta(x)(y - x) + |y - x|,$$

or equivalently, $|y - x| \geq \Psi(y) - \Psi(x) - \Theta(x)(y - x)$. The inequality is easily proved by considering the different cases: for example, if $-1 < x \leq y < 1$,

$$\Psi(x) + \Theta(x)(y - x) + |y - x|$$
$$= x^2 + x(y - x) + (y - x) = y^2 + (y - x)(1 - y) \geq y^2 = \Psi(y)$$

and if $-1 < x < 1 \leq y$,

$$\Psi(x) + \Theta(x)(y - x) + |y - x|$$
$$= x^2 + x(y - x) + (y - x) = 1 + (x + 1)(y - 1) \geq 1 = \Psi(y).$$

Now define $\psi(t, x) = t\Psi(x/t)$ and $\theta(t, x) = \Theta(x/t)$. It follows that

$$\psi(t, y) = t\Psi(y/t) \leq t\left\{ \Psi(x/t) + \Theta(x/t)\left(\frac{y}{t} - \frac{x}{t}\right) + \left|\frac{y}{t} - \frac{x}{t}\right| \right\}$$
$$= \psi(t, x) + \theta(t, x)(y - x) + |y - x|.$$

Let $Y = (Y_t)_{0 \leq t \leq T}$ be any martingale with uniform marginals $Y_t \sim U[-t, t]$. Then, disregarding for the moment that the regularity elements of Assumption 4.1 are not satisfied, from (9) we have

$$V^{[0,T]}(Y) \geq \lim_{\varepsilon \downarrow 0} \int_\varepsilon^T dt \int \frac{x^2}{t} q_t(dx) = \int_0^T dt \left[t \frac{1}{2t} + t \frac{1}{2t} - \int_{-t}^t \frac{x^2}{t} \frac{dx}{2t^2} \right] = \frac{2T}{3},$$

or alternatively, using the second representation in (14),

$$V^{[0,T]}(Y) \geq \frac{1}{2T^2} \int_{-T}^T x^2 \, dx + \int_0^T dt \int_{-T}^T \frac{x^2}{T^2} \frac{dx}{2T} = \frac{2T}{3}.$$

To give a direct proof of this bound, we argue as follows: omitting the superscript U, n on $t_k^{U,n}$,

$$V^{[0,T]}(Y) \geq V\left(P_{U,n}^{[0,T]}, Y\right) \geq \sum_{k=1}^{2^n} \mathbb{E}\left[\frac{Y_{t_k}^2}{t_{k-1}} \wedge t_{k-1} - \frac{Y_{t_{k-1}}^2}{t_{k-1}} \wedge t_{k-1} \right]$$

$$= \sum_{k=1}^{2^n} \left[\int_{-t_k}^{t_k} \frac{dy}{2t_k} \frac{y^2}{t_{k-1}} \wedge t_{k-1} - \int_{-t_{k-1}}^{t_{k-1}} \frac{dy}{2t_{k-1}} \frac{y^2}{t_{k-1}} \right]$$

$$= \sum_{k=1}^{2^n} \left[\int_{t_{k-1}}^{t_k} \frac{dy}{t_k} t_{k-1} + \int_0^{t_{k-1}} dy \left\{ \frac{1}{t_k} - \frac{1}{t_{k-1}} \right\} \frac{y^2}{t_{k-1}} \right]$$

$$= \sum_{k=1}^{2^n} \left[\frac{t_{k-1}(t_k - t_{k-1})}{t_k} - \frac{(t_k - t_{k-1})}{t_k t_{k-1}} \frac{t_{k-1}^2}{3} \right]$$

$$= \sum_{k=1}^{2^n} \frac{2}{3} \frac{t_{k-1}}{t_k} (t_k - t_{k-1}) \xrightarrow{n \uparrow \infty} \frac{2T}{3}.$$

REFERENCES

[1] ACCIAIO, B., BEIGLBÖCK, M., PENKNER, F., SCHACHERMAYER, W. and TEMME, J. (2013). A trajectorial interpretation of Doob's martingale inequalities. *Ann. Appl. Probab.* **23** 1494–1505. MR3098440

[2] ALBIN, J. M. P. (2008). A continuous non-Brownian motion martingale with Brownian motion marginal distributions. *Statist. Probab. Lett.* **78** 682–686. MR2409532

[3] AZÉMA, J. and YOR, M. (1979). Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78). Lecture Notes in Math.* **721** 90–115. Springer, Berlin. MR0544782

[4] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985

[5] BEIGLBÖCK, M. and JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332

[6] BEIGLBÖCK, M. and NUTZ, M. (2014). Martingale inequalities and deterministic counterparts. *Electron. J. Probab.* **19** no. 95, 15. MR3272328

[7] BERTOIN, J. and LE JAN, Y. (1992). Representation of measures by balayage from a regular recurrent point. *Ann. Probab.* **20** 538–548. MR1143434

[8] CHACON, R. V. and WALSH, J. B. (1976). One-dimensional potential embedding. In *Séminaire de Probabilités, X* (*Prèmiere Partie, Univ. Strasbourg, Strasbourg, Année Universitaire* 1974/1975). *Lecture Notes in Math.* **511**. 19–23. Springer, Berlin. MR0445598

[9] DUPIRE, B. (1997). Pricing and hedging with smiles. In *Mathematics of Derivative Securities* (*Cambridge*, 1995). *Publ. Newton Inst.* **15** 103–111. Cambridge Univ. Press, Cambridge. MR1491370

[10] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

[11] FEINBERG, E. A., MANDAVA, M. and SHIRYAEV, A. N. (2014). On solutions of Kolmogorov's equations for nonhomogeneous jump Markov processes. *J. Math. Anal. Appl.* **411** 261–270. MR3118483

[12] FELLER, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48** 488–515. MR0002697

[13] GALICHON, A., HENRY-LABORDÈRE, P. and TOUZI, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.* **24** 312–336. MR3161649

[14] GYÖNGY, I. (1986). Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Related Fields* **71** 501–516. MR0833267

[15] HAMZA, K. and KLEBANER, F. C. (2007). A family of non-Gaussian martingales with Gaussian marginals. *J. Appl. Math. Stoch. Anal.* Art. ID 92723, 19. MR2335977

[16] HENRY-LABORDÈRE, P., OBŁÓJ, J., SPOIDA, P. and TOUZI, N. (2016). Maximum maximum of martingales given marginals. *Ann. Appl. Probab.* **26** 1–44.

[17] HENRY-LABORDÈRE, P., TAN, X. and TOUZI, N. (2014). An explicit martingale version of the one-dimensional Brenier's theorem with full marginals constraint. Preprint.

[18] HENRY-LABORDÈRE, P. and TOUZI, N. (2014). An explicit martingale version of Brenier's theorem. Preprint.

[19] HIRSCH, F., PROFETA, C., ROYNETTE, B. and YOR, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions. Bocconi & Springer Series* **3**. Springer, Milan. MR2808243

[20] HOBSON, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance* 2010. *Lecture Notes in Math.* **2003** 267–318. Springer, Berlin. MR2762363

[21] HOBSON, D. and KLIMMEK, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.* **19** 189–214. MR3292129

[22] HOBSON, D. and NEUBERGER, A. (2012). Robust bounds for forward start options. *Math. Finance* **22** 31–56. MR2881879

[23] HOBSON, D. G. (1998). Robust hedging of the lookback option. *Finance Stoch.* **2** 329–347.

[24] HOBSON, D. G. (2013). Fake exponential Brownian motion. *Statist. Probab. Lett.* **83** 2386–2390. MR3093829

[25] HOBSON, D. G. and PEDERSEN, J. L. (2002). The minimum maximum of a continuous martingale with given initial and terminal laws. *Ann. Probab.* **30** 978–999. MR1906424

[26] JUILLET, N. (2014). Stability of the shadow projection and the left curtain coupling. Preprint. Available at arXiv:1407.8009v2.

[27] KÄLLBLAD, S., TAN, X. and TOUZI, N. (2015). Optimal Skorokhod embedding given full martinagles and Azéma-Yor peacocks. Preprint. Available at arXiv:1503.00500v1.

[28] KELLERER, H. G. (1972). Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.* **198** 99–122. MR0356250

[29] MADAN, D. B. and YOR, M. (2002). Making Markov martingales meet marginals: With explicit constructions. *Bernoulli* **8** 509–536. MR1914701

[30] Obłój, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–390. MR2068476

[31] Oleszkiewicz, K. (2008). On fake Brownian motions. *Statist. Probab. Lett.* **78** 1251–1254. MR2444313

Department of Statistics
University of Warwick
Coventry, CV4 7AL
United Kingdom
E-mail: D.Hobson@warwick.ac.uk