MEASURE-VALUED MARTINGALES AND OPTIMALITY OF BASS-TYPE SOLUTIONS TO THE SKOROKHOD EMBEDDING PROBLEM

MATHIAS BEIGLBÖCK, ALEXANDER M. G. COX, MARTIN HUESMANN,
AND SIGRID KÄLLBLAD

ABSTRACT. In this paper we consider (probability-)measure valued processes, which we call MVMs, which have a natural martingale structure. Following previous work of Eldan and Cox-Källblad, these processes are known to have a close connection to the solutions to the Skorokhod Embedding Problem. In this paper, we consider properties of these processes, and in particular, we are able to show that the MVMs connected to the Bass and Root embeddings have natural measure-valued analogues which also possess natural optimality properties. We also introduce a new MVM which is a generalisation of both the Bass and Root MVMs.

1. INTRODUCTION

We consider here an alternative approach to the Skorokhod embedding problem (SEP) which is based on viewing real-valued processes as measure-valued stochastic processes. The Skorokhod Embedding Problem (SEP) is a longstanding and classical problem in probability; we refer the reader to the survey articles [23, 15], and the recent paper [6] for a more recent approach inspired by methods from Optimal Transport.

The observation which connects the SEP and measure-valued processes is the following: that, modulo technicalities, there is a one-to-one correspondence between:

- the set of (continuous, UI) martingales M_t with $M_0 = 0$ and $M_\infty \sim \mu$;
- the set of (continuous) measure-valued martingales (MVMs) ξ_t with $\xi_0 = \mu$ and $\xi_\infty = \delta_y$ for some $y \in \mathbb{R}^d$.

The first formulation is (up to a time change, and for 1-dimensional processes), equivalent to the SEP, and we say that a measure-valued process ξ_t is a measure-valued martingale (MVM) if, for any $A \in \mathcal{B}(\mathbb{R}^d)$, $\xi_t(A)$ is a martingale. To see the correspondence, given M_t, we may define ξ_t by $\xi_t(A) := \mathbb{P}(M_\infty \in A | \mathcal{F}_t)$ and, conversely, $M_t := \int x \, \xi_t(\mathrm{d}x)$; see Section 2 for details.

While there is a one-to-one correspondence between processes and MVMs on $[0, \infty)$, for any $t < \infty$, the process ξ_s, $s \leq t$, naturally carries more information than the corresponding process M_s, $s \leq t$, since also the marginal distribution to eventually be embedded is known. In [9], this was used to formulate optimal embedding problems as dynamic programming problems, exploiting the fact that the MVM approach allows one to include the terminal embedding constraint in the current state of the controlled process; see also [4, 5, 18]. MVMs and their connection to the SEP have also appeared in [12]. There is also a stream of literature dating back to at least [26] who look at *martingale measures*, where the assumption that the measures are probability measures is typically dropped. Subsequently, a

substantial literature on these processes has developed, although much of it under the assumption that the martingale measures are either *orthogonal*, or have *nuclear covariance*, e.g. [11]. Such processes have been useful in a variety of applications; see e.g. [17].

Here, we consider an application of the measure-valued viewpoint that truly exploits the metric structure of the underlying space of measures. Specifically, motivated by the theory of optimal transport, we equip the space of measures with the 1-Wasserstein metric. The metric structure then allows for a study of functions defined on this space. First, we study the evolution over time of the Wasserstein distance between to measure-valued martingales: we show that for any two MVMs η_t and ξ_t, the process $t \mapsto \mathcal{W}(\xi_t, \eta_t)$ is a sub-martingale for $t \geq 0$. Second, we introduce a suitable notion of 'speed'. That is, a means by which one can measure how fast a measure-valued process evolves over time – while the speed of an MVM ξ_t is always bounded from below by the quadratic variation of its mean process M_t, these quantities will in general differ.

By use of these observations, we then obtain two optimality properties of the well-known Bass [3] solution to the SEP. We show that when η_t is the MVM associated with a Brownian motion, then the above-mentioned process measuring the distance of η_t to ξ_t, is in fact a martingale – as opposed to a sub-martingale – when evaluated for the Bass solution. On the other hand, the Bass MVM also miminises the speed among all MVMs. We also obtain an optimality property of the Root [24] solution. In fact, modulo a suitable scaling of space, the Bass and Root solutions solve the same optimisation problem. We emphasise that these optimality properties genuinely exploit the properties of the embeddings viewed as measure-valued martingales rather than stopped processes. We also discuss Markov properties of the involved MVMs.

1.1. **Notation.** We recall some useful definitions and notation relating to probability measures. Specifically, we recall that the pushforward of a probability measure λ on X by a function $f : X \to Y$, is the measure μ given by $\mu(A) = \lambda(f^{-1}(A))$, and is denoted $f_\#\lambda$. Further, on the space of probability measures, we introduce the functional $\mathbb{M}(\mu) = \int x \, \mu(\mathrm{d}x)$ for the mean. For a semimartingale X, we write $\langle X \rangle$ for the quadratic variation process.

2. Basics on Measure-Valued Martingales

For any $p \geq 1$, we introduce the set of p-integrable probability measures:

$$(2.1) \qquad \mathcal{P}_p := \left\{ \mu \in \mathcal{M}_+(\mathbb{R}^d) : \mu(\mathbb{R}^d) = 1, \int |x|^p \, \mu(\mathrm{d}x) < \infty \right\},$$

where $\mathcal{M}_+(X)$ is the set of non-negative measures on X; we will mostly be interested in the set $\mathcal{P}_1$.

We consider a fixed underlying filtered probability space, $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, satisfying the usual conditions.

Definition 2.1. We say that an adapted process $(\xi_t)_{t \geq 0}$ with $\xi_t \in \mathcal{P}_1$, $t \geq 0$, is a *measure-valued martingale* if, for any $f \in C_b(\mathbb{R}^d)$, $\xi_\cdot(f) := \int f(x)\xi_\cdot(\mathrm{d}x)$ is a martingale.

Note, trivially, that for $f \in C_b(\mathbb{R}^d)$, the martingale $\xi_\cdot(f)$ is uniformly integrable with well defined limit $\xi_\infty(f)$ (in particular, ξ_∞ is a measure; see [16, Proposition 2.1]). We also note that an adapted process $\xi_\cdot \in \mathcal{P}_1$ is a measure-valued martingale, if and only if, $\xi_\cdot(A)$ is a martingale for any $A \in \mathcal{B}(\mathbb{R})$, and then, in fact, $\xi_\cdot(f)$ is a martingale for any (non-negative) measurable function; see [9, Remark 2].

We also wish to discuss continuity of measure-valued martingales. In order to do this, we make the following definition:

Definition 2.2. We say that a measure-valued martingale is *continuous* if, for any 1-Lipschitz function f, $\xi.(f) = \int f(x)\,\xi.(\mathrm{d}x)$ is a continuous process.

It immediately follows that $\mathbb{M}(\xi.)$ is a continuous process whenever $\xi.$ is continuous; conversely, whenever $\xi.$ is continuous in the sense of weak convergence of measures, and also its mean is continuous, it is continuous in the above sense. Specifically, Definition 2.2 is equivalent to requiring continuity of $t \mapsto \xi_t$ in the topology of $\mathcal{W}_1$, the L^1 Wasserstein metric (cf. (4.2) below), by the duality of the Wasserstein distance [1, Theorem 6.1.1].

Since our underlying probability space is assumed to satisfy the usual conditions, of course, for every $f \in C_b(\mathbb{R}_+)$, the martingale $\xi.(f)$ has a càdlàg version. More pertinently, according to [16, Theorem 2.5] and [9, Remark 4], we can choose a version of any given measure-valued martingale ξ, such that $\xi.(f)$ is càdlàg for every 1-Lipschitz function f. In what follows, we will assume that we always take this càdlàg (in the sense of Definition 2.2) version of ξ. Moreover, by a slight modification of [9, Remark 4], given that $\xi_0 \in \mathcal{P}_p$, for some $p \geq 1$, there exists a version of ξ which is càdlàg with respect to continuity in the topology induced by $\mathcal{W}_p$; whenever $\xi_0 \in \mathcal{P}_p$ we will assume that we take this version.

We need one further concept. Consider the set of singular measures on $\mathbb{R}^d$: $\mathcal{P}^s := \{\mu \in \mathcal{P}_1 : \mu = \delta_y, y \in \mathbb{R}^d\}$. Motivated by the fact that the support of a measure-valued martingale can only ever decrease – if $\xi_{t_0}(A) = 0$, then $\xi_t(A) = 0$ for all $t \geq t_0$ – we define as follows:

Definition 2.3. We say that a measure-valued martingale $\xi.$ is *terminating* if $\xi_t \to \xi_\infty \in \mathcal{P}^s$ a.s. as $t \to \infty$, where the convergence is in the sense of weak convergence of measures. It is *finitely terminating* if $\tau_s := \inf\{t \geq 0 : \xi_t \in \mathcal{P}^s\}$ is finite a.s.

Lemma 2.4. *Suppose $\xi.$ is a terminating measure-valued martingale with $\xi_0 = \mu$. Then $X. := \mathbb{M}(\xi.)$ is a UI martingale with $X_\infty \sim \mu$.*

Proof. First, observe that by the martingale property, for $f \in C_b(\mathbb{R}^d)$,

$$\mathbb{E}\left[f(X_\infty)\right] = \mathbb{E}\left[\int f(x)\,\xi_\infty(\mathrm{d}x)\right] = \mathbb{E}\left[\int f(x)\,\xi_0(\mathrm{d}x)\right] = \int f(x)\,\mu(\mathrm{d}x).$$

and so $X_\infty \sim \mu$. In particular, X_∞ is integrable.

Now observe that the martingale statement can be shown for each co-ordinate individually. Write $f_K(x) := (x \wedge K) \vee (-K)$. Then $\xi_t \in \mathcal{P}_1$ implies that

$$\int f_K(x)\,\xi_t(\mathrm{d}x) \to \int x\,\xi_t(\mathrm{d}x) = X_t$$

as $K \to \infty$. Moreover, $\int f_K(x)\,\xi_t(\mathrm{d}x) = \mathbb{E}[f_K(X_\infty)|\mathcal{F}_t] \to \mathbb{E}[X_\infty|\mathcal{F}_t]$ by the (conditional) Dominated Convergence Theorem. Hence $X_t = \mathbb{E}[X_\infty|\mathcal{F}_t]$ for all $t \geq 0$, and it follows that X_t is a UI martingale. $\square$

Corollary 2.5. *If $\xi.$ is a terminating measure-valued martingale with $\xi_0 = \mu$, then for every 1-Lipschitz function f, $X^f. := \xi.(f)$ is a uniformly integrable martingale with $X_0^f = \int f\,\mathrm{d}\mu$ and $X_\infty^f \sim f_\#\mu$.*

3. EXAMPLES OF MVMS

In this section we will introduce certain natural MVMs, which will also become an important focus for our later optimality results. We will pay particular attention to MVMs associated with various solutions to the Skorokhod embedding problem. When doing so, we restrict to the case of one dimension, $d = 1$, and suppose that the given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ supports a Brownian motion $(B_t)_{t \geq 0}$.

Recall that for a given centred distribution $\mu \in \mathcal{P}_1(\mathbb{R})$, a solution to the SEP is a stopping time τ such that $B_\tau \sim \mu$ and $(B_{t \wedge \tau})_{t \geq 0}$ is uniformly integrable. Following

[12, 9], the SEP is equivalent to the problem of finding a terminating MVM with $\xi_0 = \mu$ and such that $\mathbb{M}(\xi_t)$ is a Brownian motion up to time $\tau := \inf\{t \geq 0 : \xi_s \in \mathcal{P}^s\}$. Specifically, if we let

$$(3.1) \qquad \xi_t(A) := \mathbb{P}\left(B_\tau \in A | \mathcal{F}_t\right), \quad A \in \mathcal{B}(\mathbb{R}),$$

then the process ξ_t is an MVM with $\xi_0 = \mu$. Note that by definition $\xi_0 = \mu$ and ξ_t is terminating.

We now consider some examples (see also [12]).

• **The Root MVM:** Root [24] (see also [19, 25]) showed that there exists a barrier $\mathcal{R}$ such that $\tau_\mathcal{R} := \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$ is a solution to the SEP; we denote the associated MVM defined via (3.1) by $\xi^\mathcal{R}$. More precisely, a barrier is a subset of $\mathbb{R}_+ \times \mathbb{R}$ such that $(t, x) \in \mathcal{R}$ implies $(s, x) \in \mathcal{R}$ for all $s \geq t$. The results of Root (and subsequent authors) then show that for any centred, integrable probability measure μ, there exists a barrier $\mathcal{R}$ such that $B_{\tau_\mathcal{R}} \sim \mu$, and $B_{t \wedge \tau_\mathcal{R}}$ is uniformly integrable. Moreover, by [22], such a barrier is unique. Given the barrier $\mathcal{R}$, we can further define stopping times $\tau_\mathcal{R}^{t,x} := \inf\{s \geq 0 : (t + s, x + W_s) \in \mathcal{R}\}$ for some Brownian motion W, $W_0 = 0$. Then $\mu_\mathcal{R}^{t,x}(A) := \mathbb{P}(x + W_{\tau_\mathcal{R}^{t,x}} \in A)$ defines a class of probability measures, and we have $\xi_t^\mathcal{R} = \mu_\mathcal{R}^{t,B_t}$, a.s. Note that $\xi^\mathcal{R}$ is a terminating MVM (since $\tau_\mathcal{R}$ is almost surely finite) and that $B_{t \wedge \tau_\mathcal{R}} = \int x \, \xi_t^\mathcal{R}(\mathrm{d}x)$.

• **The Bass MVM:** A second natural construction of MVMs, which has a strong 'transport' influence, and is closely related to the construction of a solution to the SEP due to [3] (see also [12, 13]), is described as follows. Let $(B_t)_{t \in [0,1]}$ be a Brownian motion started at 0, and observe then that $\eta_t(A) := \mathbb{P}(B_1 \in A | \mathcal{F}_t)$ is an MVM given by $\eta_t = \mathcal{N}(B_t, 1 - t)$. Let μ be a given (integrable, centred) probability measure. Write F_μ, F_{η_t} for the c.d.f.s of μ and η_t respectively, and write $h = F_\mu^{-1} \circ F_{\eta_0}$. We will call h the *scale function* of the MVM. Then $h(B_1) \sim \mu$. Moreover, it is easily seen that the measure-valued process ξ^h defined by $\xi_t^h := h_\# \eta_t$, or equivalently

$$\int f(x) \, \xi_t^h(\mathrm{d}x) := \int (f \circ h)(x) \, \eta_t(\mathrm{d}x) = \int f(x) \, (h_\# \eta_t)(\mathrm{d}x) = \mathbb{E}\left[f(h(B_1)) | \mathcal{F}_t\right],$$

for $f \in C_b(\mathbb{R}^d)$, defines an MVM with $\xi_0^h = \mu$.

Strictly speaking, to recover Bass' solution to the SEP, one needs to show how to convert this MVM into a stopping time for a Brownian motion. Since ξ^h is an MVM, then $M^h := \mathbb{M}(\xi_\cdot^h)$ is a (continuous) martingale. In particular, there exists a time-change $\tau_t := \inf\{s \geq 0 : \langle M^h \rangle_s \geq t\}$ such that M_{τ_t} is a Brownian motion (with respect to the filtration $\mathcal{G}_t := \mathcal{F}_{\tau_t}$) up to the ($\mathcal{G}$-)stopping time $\sigma := \tau_1^{-1} = \sup\{s \geq 0 : \tau_s < 1\}$, and $B_\sigma \sim \mu$. Bass' solution is then the stopping time σ (which can also be constructed directly through stochastic calculus arguments). Of course, the MVM $\tilde{\xi}_t^h := \xi_{\tau_t}^h$ is another example of an MVM. To distinguish, we will call ξ^h the *canonical-time* Bass embedding, and $\tilde{\xi}^h$ the *natural-time* Bass embedding. We will typically be interested in the canonical-time MVM, and we will often call this MVM simply the Bass MVM.

Note that in both cases, the 'canonical' choice of $h = F_\mu^{-1} \circ F_{\eta_0}$ is not necessary to deduce that the resulting process is an MVM. In fact, an arbitrary h such that $h_\# \eta_0 = \mu$ can be chosen to determine an MVM.

• **The Azéma-Yor MVM:** The Azéma-Yor solution to the SEP, [2], is given by $\tau_{AY} := \inf\{t \geq 0 : B_t \leq \psi(S_t)\}$, where ψ is the inverse Barycentre function of μ and $S_t := \sup_{s \leq t} B_s$. To be specific, we introduce the following notation from [10]: given a (probability) measure μ, and $p \in [0, 1]$, we write μ_p for the (sub-probability) measure given by $\mu_p((-\infty, x]) = (\mu((-\infty, x]) - p)_+$. Then it

is clear that $p \in [0,1) \mapsto \frac{1}{p} \int x \, \mu_p(\mathrm{d}x)$ is a continuous, increasing function, and in fact strictly increasing as long as $p \leq \mu(\{\sup \operatorname{supp} \mu\})$. In particular, writing $\sup \operatorname{supp} \mu = r$, and $\inf \operatorname{supp} \mu = \ell$ we see that for all $x \in \left[\int x \, \mu(\mathrm{d}x), r \right)$ there exists $\pi(x) \in [0,1)$ such that $x(1 - \pi(x)) = \int x \, \mu_p(\mathrm{d}x)$. The inverse barycentre function for μ, ψ, is then given by $\psi(x) = F_\mu^{-1}(\pi(x))$ for $x < r$, and $\psi(r) = r$. Moreover, the corresponding MVM can then be written as:

$$\xi_t^{AY} = \begin{cases} \frac{S_t - B_t}{S_t - \psi(S_t)} \delta_{\psi(S_t)} + \frac{B_t - \psi(S_t)}{S_t - \psi(S_t)} \frac{\mu_{\pi(S_t)}}{1 - \pi(S_t)}, & t < \tau_{AY} \\ \delta_{\psi(S_{\tau_{AY}})}, & t \geq \tau_{AY}. \end{cases}$$

• **The Bass-Root MVM:** The final example we give is a combination of the Bass and Root cases. Specifically, let $\mathcal{R}$ be the Root barrier associated with a non-atomic measure λ. Given a measure μ, choose the function $\kappa = F_\mu^{-1} \circ F_\lambda$, and set $\xi_\cdot^{\kappa,\mathcal{R}} = \kappa_\# \xi^{\mathcal{R}}$. It follows that $\xi_0^{\kappa,\mathcal{R}} = \mu$, and $\xi^{\kappa,\mathcal{R}}$ is a terminating MVM. As above, we will call this time-scale the *canonical-time* Bass-Root embedding (associated to the barrier $\mathcal{R}$), and note that there is a corresponding *natural-time* Bass-Root embedding. As above, we will typically work with the canonical-time Bass-Root MVM.

Of course, both the Bass and Root MVMs are special cases of the Bass-Root MVM (corresponding, respectively, to the cases where $\mathcal{R} = \{(t,x) : t \geq 1\}$ and $h(x) = x$, or, equivalently, $\lambda = \eta_0$ and $\lambda = \mu$).

4. MVMs in Wasserstein space and the first optimality property

We will now study the properties of MVMs in Wasserstein spaces. To this end we define the set of *transport plans* from $\mathbb{R}^d$ to $\mathbb{R}^d$ which couple measures $\lambda, \mu \in \mathcal{P}_1$:

$$(4.1) \qquad \Pi(\lambda, \mu) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \nu(A \times \mathbb{R}^d) = \lambda(A), \ \nu(\mathbb{R}^d \times A) = \mu(A) \right\}.$$

For any $p \geq 1$, the p-Wasserstein metric (see (7.1.1) in [1]) on $\mathcal{P}_p$ is then given by:

$$(4.2) \qquad \mathcal{W}_p^p(\lambda, \mu) := \inf \left\{ \int |x_1 - x_2|^p \, \mathrm{d}\nu(x_1, x_2) : \nu \in \Pi(\lambda, \mu) \right\}.$$

We note that the infimum is attained, in particular, the set $\Pi_o^p(\lambda, \mu)$ of optimisers is non-empty, closed and compact in the weak topology.

In this section we study the evolution over time of the Wasserstein distance between two MVMs; specifically, for any two MVMs $\xi, \eta \in \mathcal{P}_p$, $p \geq 1$, we are interested in the stochastic process $t \mapsto \mathcal{W}_p(\xi_t, \eta_t)$. It is clearly adapted. Moreover, recall from Section 2 that we may – and do – consider versions of η and ξ which are càdlàg in the topology induced by $\mathcal{W}_p$; an application of the triangle inequality then yields that the process $\mathcal{W}_p(\eta_\cdot, \xi_\cdot)$ is right-continuous. We further adopt the following convention: for any two *random* measures λ and μ on $\mathbb{R}^d$, we say that a (random) measure on $\mathbb{R}^d \times \mathbb{R}^d$ is a transport plan from λ to μ, and write $\nu \in \Pi(\lambda, \mu)$, if the assertions in (4.1) hold a.s. For any $t \geq 0$, we then have

$$(4.3) \qquad \mathcal{W}_p^p(\xi_t, \eta_t) = \operatorname{ess\,inf} \left\{ \int |x_1 - x_2|^p \, \mathrm{d}\nu(x_1, x_2) : \nu \in \Pi(\xi_t, \eta_t) \right\} \quad \text{a.s.}$$

The first result we have is the following:

Proposition 4.1. *Let $p \geq 1$ and let $\xi_t^1, \xi_t^2 \in \mathcal{P}_p(\mathbb{R}^d)$, $t \geq 0$, be two measure-valued martingales. Then,*

    i) the process $\mathcal{W}_p^p(\xi_t^1, \xi_t^2)$, $t \geq 0$, is a sub-martingale;

    ii) if ξ^1 and ξ^2 satisfy, for some family of measures $m(x, \mathrm{d}y)$, $x \in \mathbb{R}^d$,

$$(4.4) \qquad \xi_t^2(\mathrm{d}y) = \int \xi_t^1(\mathrm{d}x) m(x, \mathrm{d}y), \quad t \geq 0,$$

    then $\mathcal{W}_p^p(\xi_t^1, \xi_t^2)$, $t \geq 0$, is a martingale;

iii) for any MVM $\xi_t^1 \in \mathcal{P}_p(\mathbb{R}^d)$, $t \geq 0$, and measure $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, there exists an MVM $\xi_t^2 \in \mathcal{P}_p(\mathbb{R}^d)$, $t \geq 0$, of the form (4.4) with $\xi_0^2 = \mu$.

Proof. i)   Let $0 < s < t$ and note that for any $\nu(\mathrm{d}x_1, \mathrm{d}x_2) \in \Pi(\xi_t^1, \xi_t^2)$, by use of Fubini's theorem, we may define an $\mathcal{F}_s$-measurable random measure ν_s by setting for bounded and measurable $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$(4.5) \qquad \int f(x_1, x_2) \, \nu_s(\mathrm{d}x_1, \mathrm{d}x_2) = \mathbb{E}\left[\int f(x_1, x_2) \, \nu(\mathrm{d}x_1, \mathrm{d}x_2) | \mathcal{F}_s\right].$$

Then, in particular,

$$\nu_s(A \times \mathbb{R}^d) = \mathbb{E}[\nu(A \times \mathbb{R}^d)|\mathcal{F}_s] = \mathbb{E}[\xi_t^1(A)|\mathcal{F}_s] = \xi_s^1(A), \quad \text{a.s.,} \quad A \in \mathcal{B}(\mathbb{R}^d),$$

and it follows that ν_s is an element of $\Pi(\xi_s^1, \xi_s^2)$. Next, since the set $\{\int |x_1 - x_2|^p \, \mathrm{d}\nu(x_1, x_2) : \nu \in \Pi(\xi_t^1, \xi_t^2)\}$ is directed downwards, there exists a sequence $\nu^n \in \Pi(\xi_t^1, \xi_t^2)$ such that $\int |x_1 - x_2|^p \, \mathrm{d}\nu^n(x_1, x_2) \searrow \mathcal{W}_p^p(\xi_t^1, \xi_t^2)$ a.s. By use of the monotone convergence theorem, we thus obtain

$$\mathbb{E}\left[\mathcal{W}_p^p(\xi_t^1, \xi_t^2)|\mathcal{F}_s\right] = \mathbb{E}\left[\lim_{n \to \infty} \int |x_1 - x_2|^p \, \nu^n(\mathrm{d}x_1, \mathrm{d}x_2)|\mathcal{F}_s\right]$$

$$= \lim_{n \to \infty} \int |x_1 - x_2|^p \, \nu_s^n(\mathrm{d}x_1, \mathrm{d}x_2) \ \geq \ \mathcal{W}_p^p(\xi_s^1, \xi_s^2),$$

where ν_s^n is defined via (4.5) with respect to ν^n.

ii) Let $t > 0$ and define $\nu(\mathrm{d}x, \mathrm{d}y) := \xi_t^1(\mathrm{d}x)m(x, \mathrm{d}y)$; by (4.4) we have that $\nu \in \Pi(\xi_t^1, \xi_t^2)$. By use of i) we then obtain

$$\mathcal{W}_p^p\left(\xi_0^1, \mu\right) = \int |x - y|^p \xi_0^1(\mathrm{d}x)m(x, \mathrm{d}y)$$

$$= \mathbb{E}\left[\int |x - y|^p \xi_t^1(\mathrm{d}x)m(x, \mathrm{d}y)\right] \ = \ \mathbb{E}\left[\mathcal{W}_p^p\left(\xi_t^1, \xi_t^2\right)\right].$$

iii) Recall that the set of minimizers for (4.2) is non-empty and let $\nu_0 \in \Pi_o^p(\xi_0^1, \mu)$. It follows by disintegration (see e.g. [1, Theorem 5.3.1]) that there exists a family of measures $m^0(x, \mathrm{d}y)$, $x \in \mathbb{R}$, such that $\nu_0(\mathrm{d}x, \mathrm{d}y) = \xi_0^1(\mathrm{d}x)m^0(x, \mathrm{d}y)$. Then, define

$$\xi_t^2 := \int \xi_t^1(\mathrm{d}x)m^0(x, \mathrm{d}y), \quad t \geq 0.$$

Since

$$\int \xi_s^1(\mathrm{d}x)m(x, \mathrm{d}y) \ = \ \mathbb{E}\left[\int \xi_t^1(\mathrm{d}x)m(x, \mathrm{d}y)|\mathcal{F}_s\right], \quad s \leq t,$$

the thus defined process $\xi^2 \in \mathcal{P}_p$ is indeed a measure-valued martingale; by definition, it satisfies (4.4) and the condition $\xi_0^2 = \mu$. $\qquad \square$

We note that the above result does not require the measure-valued martingales to be terminating; in particular, it holds also for $\xi_t^2 = \lambda \in \mathcal{P}_1$ constant.

4.1. The first optimisation problem: optimality of the Bass embedding.
We first recall the following notation from Section 3: B_t denotes a 1-dimensional Brownian motion, $\eta_t := \mathrm{Law}(B_1|\mathcal{F}_t)$ and for any $\mu \in \mathcal{P}_1$ and $h : \mathbb{R} \to \mathbb{R}$ such that $h_\# \eta_0 = \mu$, $\xi_t^h := \mathrm{Law}(h(B_1)|\mathcal{F}_t)$ defines a finitely terminating measure-valued martingale with $\xi_0^h = \mu$ – the canonical-time Bass MVM. In particular,

$$(4.6) \qquad\qquad\qquad\qquad \xi_t^h = h_\# \eta_t.$$

As a consequence of Proposition 4.1, we then obtain the following characterisation of the Bass embedding as the solution to a particular optimisation problem:

Theorem 4.2. *Given $p \geq 1$, let $\mu \in \mathcal{P}_p(\mathbb{R})$ be a given atomless measure. Define $h(x) := F_\mu^{-1} \circ F_{\eta_0}(x)$, $x \in \mathbb{R}$, and let $\xi_\cdot^h$ be the associated canonical-time Bass MVM. Then, $\xi_\cdot^h$ minimises, simultaneously for all weight-functions $w : [0, 1] \to \mathbb{R}$,*

$$(4.7) \qquad \mathbb{E}\left[\int_0^1 w(t) \mathcal{W}_p^p(\eta_t, \xi_t) \mathrm{d}t\right],$$

over all measure-valued martingales ξ_t with $\xi_0 = \mu$.

Proof. According to part ii) of Proposition 4.1, $\mathcal{W}_p^p(\eta_t, \xi_t^h)$ is a martingale; specifically, for $t \geq 0$, $(\mathrm{Id} \times h)_\# \eta_t \in \Pi(\eta_t, \xi_t^h)$ and

$$\mathcal{W}_p^p(\eta_0, \mu) = \int |x - h(x)|^p \eta_0(\mathrm{d}x) = \mathbb{E}\left[\int |x - h(x)|^p \eta_t(\mathrm{d}x)\right] = \mathbb{E}\left[\mathcal{W}_p^p(\eta_t, \xi_t^h)\right].$$

On the other hand, according to Proposition 4.1 i), $\mathcal{W}_p^p(\eta_t, \xi_t)$ is a sub-martingale for any other measure-valued martingale ξ. Hence, given that $\xi_0 = \mu$, $\mathbb{E}[\mathcal{W}_p^p(\eta_t, \xi_t)] \geq \mathbb{E}[\mathcal{W}_p^p(\eta_t, \xi_t^h)]$, for $t \geq 0$. Integration with respect to $w(t)$ and use of Fubini yields the result. $\square$

Remark 4.3. We note that for $w(t) = \delta_1$, the criterion in (4.7) reduces to minimising $\mathbb{E}[(B_1 - \mathbb{M}(\xi_1))^p]$ over measure-valued martingales with $\xi_0 = \mu$ and terminating before $t = 1$. In particular, for the case $p = 2$, since the law of B_1 and $\mathbb{M}(\xi_1)$ are fixed, this is equivalent to maximising $\mathbb{E}[B_1 \mathbb{M}(\xi_1)]$.

5. Markov properties of MVMs

In this section, we consider certain natural 'Markov-like' properties of MVMs. Clearly, one can simply ask that the MVM is itself a Markov process, in the usual sense. This is closely related to the definition of the Markov property due to Eldan [12], who gave a definition of the (time-homogenous) Markov property for the (closely related) notion of a 'Skorokhod embedding scheme'. An example of an embedding scheme/MVM that has this Markovianity is the Root embedding/MVM, where the process is completely determined by the current value of the MVM, through the corresponding Root barrier. As the MVM evolves, the barrier will move to reflect the current measure, but will evolve consistently. Note that this evolution can be defined from any given starting measure. (We note that [12] also required a shift invariance, so that the evolution of the process was invariant to constant shifts).

In fact, we will need a different form of the Markov property: specifically, we want to be able to control the evolution of the MVM in terms of the mean of the process, in such a way that if the mean of the process is at a given level at a given time, then we can conclude the value of the process at that time; however we then require a property relating the motion of the process to the motion of its mean which will imply that the mean process is (time-inhomogeneous) Markovian. Note that in this formulation, it will be important to have a fixed starting point for the process.

The required property is:

Definition 5.1 (Lipschitz-Markov)**.** We say that a MVM is Lipschitz-Markov if

$$(5.1) \qquad \mathcal{W}_1(\xi_t(\omega), \xi_t(\omega')) = |\mathbb{M}(\xi_t(\omega)) - \mathbb{M}(\xi_t(\omega'))|.$$

We note that (5.1) holds as an inequality for any MVM ξ. Indeed, it follows from the duality of the Wasserstein distance (Theorem 6.1.1, [1]) that for any two

measures $\lambda, \mu \in \mathcal{P}_1$,

$$\mathcal{W}_1(\lambda, \mu) = \sup_{f \text{ 1-Lipschitz}} \left\{ \int f(x) \, \mathrm{d}(\lambda - \mu) \right\}$$

$$(5.2) \qquad \geq \max \left\{ \int x \, \mathrm{d}(\lambda - \mu), \int (-x) \, \mathrm{d}(\lambda - \mu) \right\} = |\mathbb{M}(\lambda) - \mathbb{M}(\mu)|.$$

The Lipschitz-Markov property has appeared in e.g. [14, 21], in the following form: an adapted process, say M, is called Lipschitz-Markov if, for any bounded 1-Lipschitz function $f : \mathbb{R} \to \mathbb{R}$, there exists a 1-Lipschitz function $g : \mathbb{R} \to \mathbb{R}$, such that

$$(5.3) \qquad\qquad g(M_s) = \mathbb{E}[f(M_t) | \mathcal{F}_s], \qquad s \leq t.$$

In the form of Definition 5.1, the Lipschitz-Markov property first appeared in [7], where, in particular, its relation to the Root embedding was studied. It follows from the dual representation of the $\mathcal{W}_1$-distance, that a MVM ξ_t is Lipschitz-Markov if and only if $\mathbb{M}(\xi_t)$ is Lipschitz-Markov in the sense of [14, 21]; see [7, Lemma 5].

In particular, every Lipschitz-Markov MVM has a Markov mean process (see [20, Remark 1.70]). We also have the following result:

Lemma 5.2. *A MVM is Lipschitz-Markov if and only if it is of the form $\xi_t(\omega) = m_t(\mathbb{M}(\xi_t(\omega)))$ for some function m_t and there is an isotonic map T (i.e. $x \leq T(x)$), which may depend on t, x and y, such that $m_t(x) = T_\# m_t(y)$ whenever $x > y$.*

In particular, the Root, Bass and Bass-Root MVMs are Lipschitz-Markov.

Proof. To argue the sufficiency, suppose w.l.o.g. that $\mathbb{M}(\xi_t(\omega')) \geq \mathbb{M}(\xi_t(\omega))$. Then, the existence of an isotonic map such that $\xi_t(\omega') = T_\# \xi_t(\omega)$, implies that

$$\mathcal{W}_1(\xi_t(\omega'), \xi_t(\omega)) \leq \int |T(x) - x| \mathrm{d}\xi_t(\omega)$$

$$= \int T(x) \mathrm{d}\xi_t(\omega) - \int x \mathrm{d}\xi_t(\omega) = |\mathbb{M}(\xi_t(\omega')) - \mathbb{M}(\xi_t(\omega))|;$$

equalities must thus hold throughout according to (5.2). For the necessity, we refer to the discussion immediately preceding Lemma 4 of [7].

That the Root MVM is Lipschitz-Markov then follows as in the proof of this same result.

To prove that the Bass MVM is Lipschitz-Markov, recall that $\xi_t^h = h_\# \eta_t = h_\# \mathcal{N}(\mathbb{M}(\eta_t), 1-t)$; if h is monotone, then so is $\mathbb{M}(\xi_t^h)$ viewed as a function of $\mathbb{M}(\eta_t)$. Therefore $\mathbb{M}(\eta_t)$, and thus also $h_\# \eta_t$, may be recovered from $\mathbb{M}(\xi_t^h)$. Moreover, the existence of the isotone map T, then follows from the monotonicity of this mapping combined with the monotonicity of h itself. The Bass-Root case follows by combining the two arguments. $\qquad\square$

Remark 5.3. We note that we have several notions of Markovianity here. To try and clarify the situation, we discuss this property in relation to the MVMs introduced earlier. We consider first the Markov property of the MVM as a process in the space of probability measures (analogous to the case considered by [12]). In this case, it is easy to see that the Root embedding is a time-homogenous Markov process (this is essentially shown in [12]). If we consider the Bass MVM in the form $\xi_t^h = h_\# \eta_t$, where $\eta_t = \mathcal{N}(B_t, 1 - t)$, then it is easy to see that ξ_t^h is a time-inhomogeneous Markov process. On the other hand, in the time-changed version where $\tilde{\xi}_t^h := \xi_{\tau_t}^h$ is chosen such that $\langle \mathbb{M}(\tilde{\xi}_\cdot^h) \rangle_t = t$, then again, arguments of Eldan show that $\tilde{\xi}_t^h$ is a *time-homogenous* Markov process.

However, if we fix the starting law ξ_0, and hence h, and consider just the mean process $\mathbb{M}(\xi_t^h)$, then it is possible to check that this process is a time-inhomogeneous Markov process. On the other hand, the process $\mathbb{M}(\tilde{\xi}_t^h)$ is not a Markov process.

Similar arguments show that the mean process corresponding to the Bass and Bass-Root (on the underlying Brownian time-scale) are also both time-inhomogenous Markov processes. In the alternative Bass-Root case, where the time-scale is fixed so that the mean is a Brownian motion up to stopping, the resulting process is not Markov (in particular, the stopping time is not generally adapted to the filtration of the mean process).

6. The notion of 'Speed' of MVMs and the second optimality property

6.1. **The speed of an MVM.** A key question is how to define a suitable notion of 'speed' for MVMs. That is, a means by which one can measure how fast a MVM evolves over time. We want a notion of speed which replaces the usual notion of quadratic variation, but which is strictly increasing whenever the MVM evolves. As we will see below, this is not the case for the quadratic variation of the mean process $\mathbb{M}(\xi_t)$. However, we will be able to define a notion of speed such that, in certain circumstances, the speed of our MVMs and the quadratic variation of the mean process coincide, while in general the speed will dominate the quadratic variation. These features will be important for our second optimality property.

We first make some definitions relating to Lipschitz functions: we write $C^{0,1}(\mathbb{R}^d)$ for the set of Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ equipped with the norm:

$$||f|| = \sup_{x \in \mathbb{R}^d} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

In addition, we write $\mathcal{L}ip^1$ for the (closed) subset $\left\{ f \in C^{0,1} \, | \, \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \leq 1 \right\}$.

We now introduce the notion of a simple Lipschitz-valued process: we say that the process $(f_t)_{t \geq 0}$ is in the set $\mathcal{SL}^1$ if

$$f_t(x) = \sum_{n=0}^{\infty} f_i(x) \mathbf{1}_{\{t \in (\tau_i, \tau_{i+1}]\}}$$

where $f_i \in \mathcal{L}ip^1$ is $\mathcal{F}_{\tau_i}$-measurable, and $(\tau_i)_i$ is a sequence of increasing stopping times such that $\mathbb{P}(\lim_{n \to \infty} \tau_n = \infty) = 1$.

To a process $(f_t)_t \in \mathcal{SL}^1$ we associate a martingale (effectively, a stochastic integral against ξ_t)[1] defined by:

$$I_t := \sum_i \left(\int f_i(x) \, \xi_{\tau_{i+1} \wedge t}(\mathrm{d}x) - \int f_i(x) \, \xi_{\tau_i \wedge t}(\mathrm{d}x) \right).$$

Since $\xi_t \in \mathcal{P}_1$ for each t, and f_i are 1-Lipschitz, this sum is well defined and the resulting process is a martingale. Then we can associate an increasing process A_t^f to each $f \in \mathcal{SL}^1$ by defining A^f to be the compensator of the supermartingale I_t^2.

Definition 6.1. We define the *speed* of the MVM ξ_t to be the smallest increasing, càdlàg, adapted process $[\xi]_t$ such that $[\xi]_t \geq A_t^f$ almost surely, for every $f \in \mathcal{SL}^1$.

Note that such a process is trivially well defined by the fact that the set of increasing processes is directed downwards, so if $A_t, \tilde{A}_t$ are both increasing candidates for the speed, then so too is $A_t \wedge \tilde{A}_t$. In particular, it is easy to see that $\mathrm{d}[\xi]_t \geq \mathrm{d}A_t^f$, for any $f \in \mathcal{SL}^1$.

---

[1]There is a substantial literature on the construction of stochastic integrals for the closely related case of martingale measures (effectively, when ξ_t can be a signed measure, instead of a probability measure), however this literature is dominated by the cases where either ξ_t satisfies an orthogonality condition, or possesses *nuclear covariance*; in our setup, neither condition is natural. We refer to [26] for details. A natural question is whether the stochastic integral defined for the class $\mathcal{SL}^1$ of simple processes can be extended to a natural limit.

Proposition 6.2. *Suppose ξ is an MVM with speed $[\xi]$. Then*

(i) $[\xi]_t \geq \langle \xi \rangle_t := \langle \mathbb{M}(\xi) \rangle_t$, *the quadratic variation of the mean process;*
(ii) $[\xi]_t \equiv 0$ *if and only if ξ is constant;*
(iii) *if $(\tau_t)_{t \geq 0}$ is a time-change, and $\tilde{\xi}_t := \xi_{\tau_t}$, then $[\tilde{\xi}]_t = [\xi]_{\tau_t}$, that is, the speed is invariant under time-change.*

We note that the condition in (*ii*) is not true for the quadratic variation process, as the subsequent example demonstrates:

Example 6.3. Let ξ_t be an MVM in a Brownian filtration with $\xi_0 = U([-1,1])$. Between time 0 and the first hitting time of $\pm \varepsilon$, we use the Brownian motion to flip an artificial coin, based on which we have either $\xi = U([-1, -1/\sqrt{2}] \cup [0, 1/\sqrt{2}])$ or the converse. Then for this example, the mean process does not move at all, but the distance changes a lot. Variants on this can be used to get small movement in the quadratic variation, but large movements in speed in various natural ways. In particular, we see that it is certainly not always true that $[\xi]_t = \langle \xi \rangle_t$.

Proof of Proposition 6.2. The first part of the result follows immediately from considering a simple process such that $f_t(x) = x$ for all t.

To see (*ii*), note that if ξ is not constant almost surely, then there exist $h \in \mathcal{L}ip^1$ and $t_1 < t_2$ such that $\int h(x)\, \xi_{t_1}(\mathrm{d}x) \neq \int h(x)\, \xi_{t_2}(\mathrm{d}x)$ with positive probability. Taking $f_t(x) = h(x)\mathbf{1}_{\{t \in (t_1, t_2]\}}$, then $f \in \mathcal{SL}^1$ gives a non-trivial martingale, and hence a non-zero compensator.

Finally, the last part follows immediately from the definition of the speed of the process. $\square$

6.2. **The class of consistent MVMs.** As demonstrated by Example 6.3, the speed of an MVM will typically not coincide with the quadratic variation of its mean process. In this section we study a class of MVMs for which this is however the case. Specifically, the following result yields a sufficient condition for the speed and quadratic variation of a process to actually coincide:

Definition 6.4. We call a Lipschitz-Markov MVM ξ_t consistent if its associated function m_t satisfies $\mathcal{W}_1(m_s(x), m_t(x)) \leq \alpha(\varepsilon)$ for all x and $|t - s| < \varepsilon$, for some function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\alpha(\varepsilon)\varepsilon^{-1/2} \to 0$ as $\varepsilon \to 0$.

Lemma 6.5. *If ξ_t is a consistent Lipschitz-Markov MVM, then $[\xi]_t = \langle \xi \rangle_t$.*

Proof. Let $g \in \mathcal{L}ip^1$. By the Lipschitz-Markov property of $\xi_\cdot$, for $s < t$ we get

$$
\left| \int g(x)\, (\xi_t - \xi_s)\, (\mathrm{d}x) \right| \leq \mathcal{W}_1(\xi_t, \xi_s)
$$
$$
\leq \mathcal{W}_1(\xi_t, m_t(\mathbb{M}(\xi_s))) + \mathcal{W}_1(m_t(\mathbb{M}(\xi_s)), \xi_s)
$$
$$
\leq |\mathbb{M}(\xi_t) - \mathbb{M}(\xi_s)| + \alpha(t - s).
$$

Now take $f \in \mathcal{SL}^1$. Possibly taking a finer (stochastic) partition $\tau_1 \leq \tau_2 \leq \ldots$, which includes the original partition associated with f, and also the time t, and

using the fact that $\alpha(\varepsilon)^2 < \delta\varepsilon$, for some $\delta > 0^2$, we have:

$$\sum_i \left(\int f_i(x) \left(\xi_{\tau_{i+1} \wedge t} - \xi_{\tau_i \wedge t} \right) (\mathrm{d}x) \right)^2 \tag{6.1}$$

$$\leq \sum_{i:\tau_i < t} \left[\alpha(\tau_{i+1} - \tau_i)^2 + 2\alpha(\tau_{i+1} - \tau_i)|\mathbb{M}(\xi_{\tau_{i+1}}) - \mathbb{M}(\xi_{\tau_i})| \right.$$

$$\left. + |\mathbb{M}(\xi_{\tau_{i+1}}) - \mathbb{M}(\xi_{\tau_i})|^2 \right]$$

$$\leq \delta t + 2 \left(\sum_{i:\tau_i < t} \alpha(\tau_{i+1} - \tau_i)^2 \right)^{\frac{1}{2}} \left(\sum_{i:\tau_i < t} |\mathbb{M}(\xi_{\tau_{i+1}}) - \mathbb{M}(\xi_{\tau_i})|^2 \right)^{\frac{1}{2}}$$

$$+ \sum_i |\mathbb{M}(\xi_{\tau_{i+1} \wedge t}) - \mathbb{M}(\xi_{\tau_i \wedge t})|^2$$

$$\leq \delta t + 2 \sqrt{\delta t \sum_i |\mathbb{M}(\xi_{\tau_{i+1} \wedge t}) - \mathbb{M}(\xi_{\tau_i \wedge t})|^2} + \sum_i |\mathbb{M}(\xi_{\tau_{i+1} \wedge t}) - \mathbb{M}(\xi_{\tau_i \wedge t})|^2,$$

where we used Cauchy-Schwarz. Note that for $\delta > 0$ arbitrarily small, (6.1) still holds for ε sufficiently small. Further, we may find a sequence of partitions, with $|\Pi_n| \to 0$, such that $A_t^f = \lim_{n \uparrow \infty} \sum |I_{\tau_{i+1}^n \wedge t} - I_{\tau_i^n \wedge t}|^2$ and $\langle \xi \rangle_t = \lim_{n \uparrow \infty} \sum |\mathbb{M}(\xi_{\tau_{i+1}^n \wedge t}) - \mathbb{M}(\xi_{\tau_i^n \wedge t})|^2$, a.s.. Hence, we deduce that $A_t^f \leq \langle \xi \rangle_t$, $t \geq 0$. Since $f \in \mathcal{SL}^1$ was arbitrary, we thus have that $\langle \xi \rangle_t$ is an increasing process dominating A_t^f for all $f \in \mathcal{SL}^1$. It follows that $[\xi]_t \leq \langle \xi \rangle_t$, and applying Proposition 6.2 we conclude. $\square$

Corollary 6.6. *A consistent Lipschitz-Markov MVM ξ is continuous if its mean process, $\mathbb{M}(\xi)$, is continuous.*

Proof. Let $f \in \mathcal{SL}^1$ correspond to the constant process, $f_i = g \in \mathcal{L}ip^1$ for all i. It is sufficient to show that A^f has no jumps, but by Lemma 6.5, $\mathrm{d}A_t^f \leq \mathrm{d}\langle \xi \rangle_t$, and the conclusion follows. $\square$

Proposition 6.7. *The Bass-Root MVMs with Lipschitz scale function κ (and hence also the Bass and Root MVMs) are continuous Lipschitz-Markov MVMs with*

$$\langle \xi^{\kappa,\mathcal{R}} \rangle_t = [\xi^{\kappa,\mathcal{R}}]_t.$$

Proof. Consider a Bass-Root MVM, with K-Lipschitz scale function κ, that is $\xi_t^{\kappa,\mathcal{R}} = \kappa_{\#}\xi_t^{\mathcal{R}}$ for an appropriate Root MVM $\xi_t^{\mathcal{R}}$ with barrier $\mathcal{R}$; recall that $\tau_{\mathcal{R}} = \inf\{t \geq 0 : \xi_t^{\mathcal{R}} \in \mathcal{P}^s\}$. Let $t > 0$ and $x \in \mathbb{R}$ be such that $d((t,x),\mathcal{R}) > 0$, where d denotes the Euclidean distance, and let $0 < s < t$. In turn, let $B^{s,x}$ and $B^{t,x}$ be two independent BMs starting respectively in (s,x) and (t,x). Define $m_s(x) := \mathcal{L}(\kappa(B_{\tau^s}^{s,x}))$, with τ^s the first time $B^{s,x}$ hits the barrier $\mathcal{R}$, and let $m_t(x)$ be analogously defined.

We first derive an appropriate bound on $\mathcal{W}_1(m_s(x), m_t(x))$. To this end, let $\delta > 0$ and $a > 0$ such that the box $[s, t+\delta] \times [x-a, x+a]$ is disjoint from $\mathcal{R}$. Note that we have the following finite bound

$$\beta_0 := \sup \left\{ \mathbb{E}^{r,y}\left[|\kappa(B_{\tau^r}^{r,y})| \right] : (r,y) \in [s, t+\delta] \times [x-a, x+a] \right\} < \infty. \tag{6.2}$$

In order to compare the laws of $\kappa(B_{\tau^s}^{s,x})$ and $\kappa(B_{\tau^t}^{t,x})$, we couple the paths originating from (s,x) and (t,x), respectively, and exiting the box $[s, t+\delta] \times [x-a, x+a]$ at the same points as those will necessarily contribute to the same law. It then suffices to find an appropriate bound on the mass that we may not couple in this way. To this end, let $\beta > 0$, and note that according to Lemma A.2 (i), the

---

[2]For later use, we note that this identity holds even if δ is a fixed constant.

probability of $B^{s,x}$ exiting the box before t is bounded by $2\beta(t-s)$, provided that $t-s < \Delta$, for some sufficiently small $\Delta > 0$. Further, applying Lemma A.2 (iii), choosing if necessary δ and Δ even smaller, the total mass corresponding to paths exiting the box at the 'top' and 'bottom' but not being coupled is bounded by $\beta(t-s)$ for $t-s < \Delta$. Finally, applying Lemma A.1, we may bound the total mass corresponding to paths exiting the box along the line $(x, t+\delta)$, $x \in (-a, a)$, and not being coupled, by $\sqrt{2/\pi}\frac{t-s}{\delta}$; indeed, although the presence of the 'top' and 'bottom' of the box naturally modifies the distribution of $B^{s,x}$ and $B^{t,x}$ along this line, the corresponding contribution to the total variation distance is compensated by the fact that the 'tails' for $x \notin (-a, a)$ are cut off. Summing up and applying (6.2), we obtain that there exists $\delta > 0$ and $\Delta > 0$, such that for $t-s < \Delta$,

$$\mathcal{W}_1\left(\mathcal{L}\left(\kappa(B^{s,x}_{\tau^s})\right), \mathcal{L}\left(\kappa(B^{t,x}_{\tau^t})\right)\right) \leq \beta_0\left(3\beta + \sqrt{2/\pi}\delta^{-1}\right)(t-s).$$

In consequence, for any $d > 0$, there exist $\Delta^d > 0$ and $\beta^d > 0$ such that for any $0 < s < t$ and $x > 0$ with $d((t,x), \mathcal{R})) > d$ and $t-s < \Delta^d$,

$$\mathcal{W}_1\left(m_s(x), m_t(x)\right) \leq \beta^d(t-s).$$

Applying Lemma 6.5, we obtain $[\xi^{\kappa,\mathcal{R}}]_t = \langle\xi^{\kappa,\mathcal{R}}\rangle_t$, for $t \leq \tau^d := \inf\{t \geq 0 : d((\mathbb{M}(\xi^{\mathcal{R}}_t), t), \mathcal{R}) \leq d\}$. Since $d > 0$ was arbitrarily chosen, it follows that $[\xi^{\kappa,\mathcal{R}}]_t = \langle\xi^{\kappa,\mathcal{R}}\rangle_t$, for $t < \tau_{\mathcal{R}}$.

It only remains to argue that the speed $[\xi^{\kappa,\mathcal{R}}]$. does not possess a jump at $t = \tau_{\mathcal{R}}$. To this end, consider the Root barrier corresponding to a single atom. Let $s < t$ and consider now the same BM starting in (s, x) and (t, x) and couple the corresponding paths. Then, using that κ is K-Lipschitz we have

$$\mathcal{W}_1(m_s(x), m_t(x)) \leq K\mathbb{E}[|B^{s,x}_{\tau^s} - B^{t,x}_{\tau^t}|] \leq K\mathbb{E}[L^0_{t-s}] \leq K\sqrt{t-s};$$

we note that this bound is improved for any other Root barrier. In particular, the first inequality in (6.1) holds for $\alpha(t) = t$, and it follows that $\mathrm{d}[\xi^{\kappa,\mathcal{R}}]_t \leq \tilde{\beta}(\mathrm{d}t + \mathrm{d}\langle\xi^{\kappa,\mathcal{R}}\rangle_t)$, for some $\tilde{\beta} > 0$, and since $\langle\xi^{\kappa,\mathcal{R}}\rangle_t$ is continuous for the Bass-Root embedding we may conclude. □

Remark 6.8. Yet another natural candidate for measuring the 'speed' of a MVM, is given by the following increasing process:

$$(6.3) \qquad\qquad \widetilde{[\xi]}_t = \liminf_{|\Pi_n|\to 0} \sum_{i:t_{i+1}<t} \mathcal{W}_1^2(\xi_{t_{i+1}}, \xi_{t_i}),$$

where the limit is taken over a fixed sequence $\Pi_n = \{0 = t_0 \leq t_1 \leq \ldots, t_k \to \infty$ as $k \to \infty\}$ of partitions of $\mathbb{R}_+$ with $|\Pi_n| = \sup_i |t_{i+1} - t_i|$. Recall that by the duality of the Wasserstein distance, $|\mathcal{W}_1(\xi_{t_{i+1}}, \xi_{t_i})| = \sup\{|\xi_{t_{i+1}}(f) - \xi_{t_i}(f)| : f \in \mathcal{L}ip^1(\mathcal{F}_{t_{i+1}})\}$, which yields $\widetilde{[\xi]}_t \geq [\xi]_t$. For consistent MVMs, we immediately see that equality holds. More generally, regardless of whether the speed coincides with the quadratic variation or not, we expect equality to hold for a large class of MVMs being adapted to the Brownian filtration.

6.3. **The ρ-speed of an MVM.** We observe that, although important in some of our definitions (in particular, through the use of the Wasserstein topology), the metric of the underlying space ($\mathbb{R}^d$) so far has not been too significant. In this section, we consider one way of transforming the underlying geometry of the space, and examine the consequences for the processes. The starting point for this observation is that if $(\xi_t)_t$ is an MVM, and ρ is an increasing function such that $|\rho(x)| \leq K(1 + |x|)$ for some $K > 0$, then $\chi_t := \rho_\#\xi_t$ is also a MVM. This follows immediately from the definition of the pushforward, so that $\int f(x)\chi_t(\mathrm{d}x) = \int f \circ \rho(x)\xi_t(\mathrm{d}x)$.

Of course, to the process χ, we can associate a speed process, $[\chi]_t$, and we can think of this process as the ρ-speed of the process ξ. This corresponds to the following definition:

We say that $f \in \mathcal{L}ip^{1,\rho}$ if and only if $f = g \circ \rho$, where $g \in \mathcal{L}ip^1$. Then $(f_t) \in \mathcal{SL}^{1,\rho}$ if f is of the form

$$f_t(x) = \sum_{n=0}^{\infty} f_i(x) \mathbf{1}_{\{t \in (\tau_i, \tau_{i+1}]\}}$$

where $f_i \in \mathcal{L}ip^{1,\rho}$ is $\mathcal{F}_{\tau_i}$-measurable, and (τ_i) is a sequence of increasing stopping times such that $\mathbb{P}(\lim_{n \to \infty} \tau_n = \infty) = 1$. Then we have:

Definition 6.9. The ρ-speed of the MVM ξ is the smallest increasing, càdlàg, adapted process $[\xi]_t^\rho$ such that $[\xi]_t^\rho \geq A_t^f$ almost surely, for every $f \in \mathcal{SL}^{1,\rho}$.

It is a simple consequence of the definitions and the argument that $[\xi]_t^\rho \geq \langle \mathbb{M}^\rho(\xi) \rangle_t$, where $\mathbb{M}^\rho(\xi) := \int \rho(x) \, \xi(\mathrm{d}x)$.

6.4. **The second optimisation problem: optimality of the Bass and modified Root solution.** As a consequence of the above properties of the (ρ)-speed, we obtain yet another optimality property of the Bass embedding. More generally, our next result provides an optimality criterion which is minimised by the (canonical-time) Bass-Root MVM:

Theorem 6.10. Let $\mu, \lambda \in \mathcal{P}^1(\mathbb{R})$ be given atomless probability measures, and let $\xi^\mathcal{R}$ be the Root MVM corresponding to embedding λ. Define $\kappa = F_\mu^{-1} \circ F_\lambda$ and write $\rho = \kappa^{-1}$ for its right continuous inverse. Then the canonical-time Bass-Root MVM, $\xi^{\kappa,\mathcal{R}} = \kappa_\# \xi^\mathcal{R}$, is a terminating, continuous MVM with $\xi_0^{\kappa,\mathcal{R}} = \mu$, which minimises

$$(6.4) \qquad\qquad \mathbb{E}\left[F\left([\xi]_\infty^\rho\right)\right]$$

for any increasing, convex function F, over the class of continuous, terminating MVMs with $\xi_0 = \mu$.

In particular, we see that the (canonical-time) Bass MVM minimises (6.4) for the choice of $\rho(x) = h^{-1}(x)$ with h given as in Section 3, whereas the Root MVM minimises (6.4) with the ρ-speed replaced by the non-modified speed $[\xi.]$.

Proof of Theorem 6.10. The MVM $\xi^{\kappa,\mathcal{R}}$ is terminating with $\xi_0^{\kappa,\mathcal{R}} = \mu$ by definition, and continuous by Corollary 6.6 and Proposition 6.7. Moreover, $\rho_\# \xi^{\kappa,\mathcal{R}} = \xi^\mathcal{R}$, and so $[\xi^{\kappa,\mathcal{R}}]_\infty^\rho = [\xi^\mathcal{R}]_\infty$. But by Proposition 6.7, we have $[\xi^\mathcal{R}]_\infty = \langle \xi^\mathcal{R} \rangle_\infty$, and so $\mathbb{E}[F([\xi^{\kappa,\mathcal{R}}]_\infty^\rho)] = \mathbb{E}[F(\langle \xi^\mathcal{R} \rangle_\infty)]$. Note in particular, that $M^\mathcal{R} := \mathbb{M}(\xi^\mathcal{R})$ is a stopped Brownian motion, stopped according to the first hitting time of a Root barrier, and also has stopped law λ.

Now consider any other ξ satisfying the conditions of the theorem. As above, we have $\rho_\# \xi. = \chi.$ for some terminating MVM χ. Then $[\xi]_t^\rho = [\chi]_t \geq \langle \chi \rangle_t$, by Proposition 6.2 (*i*). In particular, $M. := \mathbb{M}(\chi.)$ is a uniformly integrable, continuous martingale, with $M_\infty \sim \lambda$, and so $\mathbb{E}\left[F\left([\xi]_\infty^\rho\right)\right] \geq \mathbb{E}[F(\langle M \rangle_\infty)] \geq \mathbb{E}[F(\langle M^\mathcal{R} \rangle_\infty)]$, where the first step follows from the increase of F, and the second from convexity of F and the optimality of the Root embedding, which completes the proof. $\quad\square$

APPENDIX A. AUXILIARY ESTIMATES

Lemma A.1. Let $0 < u < v$, then

$$||\mathcal{N}(0, u) - \mathcal{N}(0, v)||_{\mathsf{TV}} \leq \sqrt{\frac{2}{\pi}} \frac{v - u}{u}.$$

Proof. Denote the densities corresponding to $\mathcal{N}(0, u)$ and $\mathcal{N}(0, v)$ by p_u and p_v, and denote by $\delta > 0$ the value for which $p_u(\delta) = p_v(\delta)$. Then, the total variation distance between $\mathcal{N}(0, u)$ and $\mathcal{N}(0, v)$ is bounded by

$$||p_u - p_v||_{\mathsf{TV}} \leq 4 \int_\delta^\infty p_v(\mathrm{d}x) - p_u(\mathrm{d}x)$$

$$= 4 \left(p_v([\delta, \infty)) - p_v \left([\delta\sqrt{\frac{v}{u}}, \infty) \right) \right)$$

$$\text{(A.1)} \qquad = 4p_v \left([\delta, \delta\sqrt{\frac{v}{u}}) \right) \ \leq \ 4p_v(0)\delta \left(\sqrt{\frac{v}{u}} - 1 \right).$$

We note that δ solves $v^{-\frac{1}{2}}e^{-\frac{1}{2}\frac{x^2}{v}} = u^{-\frac{1}{2}}e^{-\frac{1}{2}\frac{x^2}{u}}$, which implies that

$$\delta = \sqrt{\frac{\ln(\frac{v}{u})}{\frac{1}{u} - \frac{1}{v}}} = \sqrt{\frac{\ln(1 + \frac{v-u}{u})}{\frac{1}{u} - \frac{1}{v}}} \leq \sqrt{\frac{\frac{v-u}{u}}{\frac{v-u}{uv}}} = \sqrt{v}.$$

Moreover,

$$\sqrt{\frac{v}{u}} = \sqrt{1 + \frac{v - u}{u}} \leq 1 + \frac{1}{2}\frac{v - u}{u}.$$

Putting the above bounds for δ and $\sqrt{v/u}$ back into (A.1) and using that $p_v(0) = (2\pi v)^{-\frac{1}{2}}$, we easily conclude. $\qquad\square$

Lemma A.2. *Let B_t, $t \geq 0$ be a BM and define $\tau_a := \inf\{t \geq 0 : B_t \geq a\}$ and $\tau_{a,a} := \inf\{t \geq 0 : B_t \notin (-a, a)\}$; the corresponding density functions exist and we denote them by p^a and $p^{a,a}$. Then the following holds:*

i) For any $\beta > 0$, there exists $\Delta > 0$ such that for $\varepsilon < \Delta$,

$$\mathbb{P}(\tau_a \leq \varepsilon) \leq \beta\varepsilon;$$

ii) For any $\beta > 0$, there exists $\delta > 0$ and $\Delta > 0$, such that for $\varepsilon < \Delta$,

$$\int_0^\delta |p^a(t + \varepsilon) - p^a(t)|\mathrm{d}t \leq \beta\varepsilon.$$

iii) For any $\beta > 0$, there exists $\delta > 0$ and $\Delta > 0$, such that for $\varepsilon < \Delta$,

$$\int_0^\delta |p^{a,a}(t + \varepsilon) - p^{a,a}(t)|\mathrm{d}t \leq \beta\varepsilon.$$

Proof. i) Recall that by the reflection principle

$$\mathbb{P}(\tau_a \leq t) = \mathbb{P}(|B_t| \geq a) = 2 \int_a^\infty \frac{1}{\sqrt{2\pi t}}e^{\frac{-x^2}{2t}} \, \mathrm{d}x;$$

differentiating twice in t we obtain

$$\text{(A.2)} \qquad p^a(t) = \frac{a}{\sqrt{2\pi t^3}}e^{\frac{-a^2}{2t}}. \quad \text{and} \quad \frac{\partial p^a}{\partial t}(t) = \frac{a(a^2 - 3t)}{\sqrt{4\pi t^7}}e^{\frac{-a^2}{2t}}.$$

Next, since $t \mapsto \mathbb{P}(\tau_a \leq t)$ is differentiable, we have $\mathbb{P}(\tau_a \leq \varepsilon) = p^a(s)\varepsilon$ for some $s < \varepsilon$. We note from (A.2) that p^a is continuous and that $\lim_{t\to 0} p^a(t) = 0$; hence, for any $\beta > 0$, by choosing Δ small enough, we may ensure that $p^a(s) \leq \beta$, for all $s < \Delta$ which yields the claim.

ii) We see from (A.2) that $p^a(t)$ is increasing for $t \leq \bar{t}$ and decreasing for $t \geq \bar{t}$, where $\bar{t} = a^2/3$. Let $\Delta < \bar{t}$ and $\delta := \bar{t} - \Delta$. Then, for any $\varepsilon < \Delta$,

$$\int_0^\delta |p^a(t + \varepsilon) - p^a(t)|\mathrm{d}t \leq \mathbb{P}(\tau_a \leq \delta + \varepsilon) - \mathbb{P}(\tau_a \leq \delta) = p^a(s)\varepsilon,$$

for some $s \in (\delta, \bar{t})$. By use of the same arguments as in i), for any $\beta > 0$, choosing if necessary $\bar{t}$ even smaller, we may ensure that $p^a(s) \leq \beta$ for all $s < \bar{t}$ which yields the claim.

iii) According to [8] (see pp. 355 and 641),

$$p^{a,a}(t) = \sum_{-\infty}^{\infty} (-1)^k p^{a(1+2k)}(t) = 2 \sum_{-\infty}^{\infty} p^{a(1+4k)}(t);$$

hence $p^{a,a}$ is continuously differentiable and $\lim_{t \to 0} p^{a,a}(t) = 0$. Since it is further a density function and thus non-negative, there must exist $\bar{t} > 0$ such that $p^{a,a}(t)$ is non-decreasing for $t \leq \bar{t}$. We may thus apply the same argument as in ii) to conclude. $\qquad\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows - In Metric Spaces and in the Space of Probability Measures*. Second Edition. Birkhäuser, 2008.

[2] J. Azéma and M. Yor. "Une solution simple au problème de Skorokhod". *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*. Vol. 721. 1979, pp. 90–115.

[3] R. F. Bass. "Skorokhod imbedding via stochastic integrals". *Séminaire de Probabilités XVII 1981/82*. 986. 1983, pp. 221–224.

[4] E. Bayraktar, A. Cox, and Y. Stoev. "Martingale optimal transport with stopping". 2017. URL: http://arxiv.org/abs/1701.06231.

[5] E. Bayraktar and C. W. Miller. "Distribution-Constrained Optimal Stopping". 11, 2016. URL: http://arxiv.org/abs/1604.03042.

[6] M. Beiglböck, A. M. G. Cox, and M. Huesmann. "Optimal transport and Skorokhod embedding". *Inventiones mathematicae* (18, 2016), pp. 1–74. DOI: 10.1007/s00222-016-0692-2.

[7] M. Beiglböck, M. Huesmann, and F. Stebegg. "Root to Kellerer". *Séminaire de Probabilités XLVIII*. 2016, pp. 1–12. DOI: 10.1007/978-3-319-44465-9_1.

[8] A. N. Borodin and P. Salminen. *Handbook of Brownian motion—facts and formulae*. Second Edition. Birkhäuser Verlag, Basel, 2002, pp. xvi+672. DOI: 10.1007/978-3-0348-8163-0.

[9] A. M. G. Cox and S. Källblad. "Model-independent bounds for Asian options: a dynamic programming approach". To Appear, SIAM Journal on Control and Optimisation. URL: https://arxiv.org/abs/1507.02651.

[10] A. M. G. Cox and J. Obłój. "On joint distributions of the maximum, minimum and terminal value of a continuous uniformly integrable martingale". *Stochastic Processes and their Applications* 125.8 (2015), pp. 3280–3300. DOI: 10.1016/j.spa.2015.03.005.

[11] N. El Karoui and S. Méléard. "Martingale measures and stochastic calculus". *Probability Theory and Related Fields* 84.1 (1990), pp. 83–101. DOI: 10.1007/BF01288560.

[12] R. Eldan. "Skorokhod embeddings via stochastic flows on the space of Gaussian measures". *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 52.3 (2016), pp. 1259–1280. DOI: 10.1214/15-AIHP682.

[13] Y. Hariya. "A connection of the Brascamp-Lieb inequality with Skorokhod embedding". *Electronic Communications in Probability* 19.0 (2014). DOI: 10.1214/ECP.v19-3025.

[14] F. Hirsch, B. Roynette, and M. Yor. "Kellerer's Theorem Revisited". *Asymptotic Laws and Methods in Stochastics*. DOI: 10.1007/978-1-4939-3076-0_18. 2015, pp. 347–363.

[15] D. Hobson. "The Skorokhod Embedding Problem and Model-Independent Bounds for Option Prices". A. Cousin et al. *Paris-Princeton Lectures on Mathematical Finance 2010*. Red. by J.-M. Morel, F. Takens, and B. Teissier. Vol. 2003. 2011, pp. 267–318.

[16] J. Horowitz. "Measure-valued random processes". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 70.2 (1, 1985), pp. 213–236. DOI: 10.1007/BF02451429.

[17] P. Imkeller, M. Pontier, and F. Weisz. "Free lunch and arbitrage possibilities in a financial market model with an insider". *Stochastic Process. Appl.* 92.1 (2001), pp. 103–130. DOI: 10.1016/S0304-4149(00)00071-5.

[18] S. Källblad. "A Dynamic Programming Principle for Distribution-Constrained Optimal Stopping". 24, 2017. URL: http://arxiv.org/abs/1703.08534.

[19] J. Kiefer. "Skorohod embedding of multivariate RV's, and the sample DF". *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 24.1 (1972), pp. 1–35. DOI: 10.1007/BF00532460.

[20] T. M. Liggett. *Continuous Time Markov Processes: An Introduction*. American Mathematical Soc., 2010. 290 pp.

[21] G. Lowther. "Limits of one-dimensional diffusions". *Ann. Probab.* 37.1 (2009), pp. 78–106. DOI: 10.1214/08-AOP397.

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

[23] J. Obłój. "The Skorokhod embedding problem and its offspring". *Probability Surveys* 1 (2004), pp. 321–392.

[24] D. H. Root. "The existence of certain stopping times on Brownian motion". *Ann. Math. Statist.* 40 (1969), pp. 715–718.

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

[26] J. B. Walsh. "An introduction to stochastic partial differential equations". *École d'Été de Probabilités de Saint Flour XIV - 1984*. 1180. 1986, pp. 265–439.

Technische Universität Wien, Vienna, Austria. e-mail: mathias.beiglboeck@tuwien.ac.at.

University of Bath., Bath, U. K.. e-mail: a.m.g.cox@bath.ac.uk.

University of Bonn, Bonn, Germany. e-mail: huesmann@iam.uni-bonn.de.

Technische Universität Wien, Vienna, Austria. e-mail: sigrid.kaellblad@tuwien.ac.at.