COUPLING THE KOLMOGOROV DIFFUSION: MAXIMALITY AND EFFICIENCY CONSIDERATIONS

By SAYAN BANERJEE AND WILFRID S. KENDALL

Abstract

This is a case study concerning the rate at which probabilistic coupling occurs for nilpotent diffusions. We focus on the simplest case of Kolmogorov diffusion (Brownian motion together with its time integral or, more generally, together with a finite number of iterated time integrals). We show that in this case there can be no Markovian maximal coupling. Indeed, there can be no *efficient* Markovian coupling strategy (efficient for all pairs of distinct starting values), where the notion of efficiency extends the terminology of Burdzy and Kendall (2000). Finally, at least in the classical case of a single time integral, it is not possible to choose a Markovian coupling that is optimal in the sense of simultaneously minimizing the probability of failing to couple by time t for all positive t. In recompense for all these negative results, we exhibit a simple efficient non-Markovian coupling strategy.

1. Introduction

This paper is written in homage and thanks to our friend and colleague Nick Bingham, who has always made it his mission to encourage and to spur on younger colleagues. It is a case study of probabilistic coupling for a particular simple nonelliptic diffusion, namely the Kolmogorov diffusion $(B, \int B \, dt)$ (Brownian motion together with its time integral), studied for example in McKean's celebrated stochastic oscillator paper [20]. The work forms part of a long-running programme of study of coupling for nilpotent diffusions, by which we mean diffusions with infinitesimal generators of the form $\chi_0 + \sum_{i=1}^{k} \chi_i^2$ for smooth vector fields $\chi_0, \chi_1, \chi_2, \ldots, \chi_k$ such that the Lie algebra generated by the vector fields is nilpotent (iterated Lie brackets of the vector fields vanish at a sufficiently high order of iteration), so that consequently the diffusion has smooth positive probability transition densities. (Note that there is a fully worked-out structure theory for diffusions for which the vector fields form a nilpotent Lie algebra: see, for example, [18, Section 4.9].) As a case study, this work can be usefully compared with the study [16] of probabilistic coupling for scalar Brownian motion together with local time at the origin. In this introductory section, we begin by making a careful definition of an iterated generalization of $(B, \int B \, dt)$, and then introduce some key coupling concepts.

15

1.1. The Kolmogorov diffusion

The classical two-component Kolmogorov diffusion is obtained by pairing a real Brownian motion B with its time integral $\int B \, \mathrm{d}t$. The vector fields in question are $\chi_0 = x_1 \partial/\partial x_2$ and $\chi_1 = \partial/\partial x_1$, and nilpotence follows from the fact that $[\chi_1, \chi_0] = \chi_2 = \partial/\partial x_2$, and noting that $[\chi_1, \chi_2] = [\chi_0, \chi_2] = 0$. The diffusion $(B, \int B \, \mathrm{d}t)$ was studied, for example, by McKean [20] as a simple example of a stochastic oscillator; see also [1]. Distributional properties have been investigated, for example, in [11], [13], and [26], while L^p estimates are studied in [27] when Brownian motion is replaced by a continuous local martingale. It has arisen in statistical studies: see [2] for an application to polynomial regression. There are potential applications (for example, to model relativistic diffusion of photons [3], also to model the motion of a tracer in fluid flow [12]); however, its main interest is as a simple model for a nonelliptic diffusion. Coupling properties have been studied in [5] and numerically in [12], also (when supplemented by further iterated time integrals) in [17]; this problem provides the simplest nontrivial example of a diffusion with nilpotent group symmetries which admits a Markovian or immersed coupling. Most interest to date has focussed on the classical two-component Kolmogorov diffusion. Here we follow [17] in considering coupling for (in the most part) the case of index k (k iterated time integrals), since the general structure adds clarity to the arguments.

We begin with an explicit definition. Given B, a standard real-valued Brownian motion (hence begun at 0), we define the *standard* generalized Kolmogorov diffusion $\widetilde{\boldsymbol{I}} = (\widetilde{I}_0, \widetilde{I}_1, \ldots, \widetilde{I}_k)^\top$ of index k by the finite recursion

$$\widetilde{I}_0(t) = B(t), \qquad \widetilde{I}_r(t) = \int_0^t \widetilde{I}_{r-1}(s) \, \mathrm{d}s \quad \text{for } r = 1, 2, \ldots, k. \tag{1}$$

Thus, $\widetilde{I}_r$ is simply the r-fold iterated time integral of Brownian motion. We shall refer to the index-1 case as the *classical* Kolmogorov diffusion, written in column vector form as $(\widetilde{I}_0, \widetilde{I}_1)^\top = (B, \int B \, \mathrm{d}t)^\top$.

Nilpotence for $\widetilde{\boldsymbol{I}}$ follows by considering the infinitesimal generator $\chi_0 + \frac{1}{2}\chi_1^2$, where $\chi_1 = \partial/\partial x_1$ (differentiation in the 'Brownian' direction, temporarily deviating from the indexing convention of (1) for the sake of convenience of exposition) but $\chi_0 = x_1\chi_2 + x_2\chi_3 + \cdots + x_{k-1}\chi_k$, where $\chi_r = \partial/\partial x_r$. An inductive argument based on $[\chi_r, \chi_0] = \chi_{r+1}$ (for $r = 1, \ldots, k-1$) shows that kth-iterated Lie brackets vanish.

Bearing in mind the form of the law of $\widetilde{\boldsymbol{I}}(t+s)$ conditional on $\mathcal{F}_t = \sigma\{B(u) : u \leq t\}$, we define the index-$k$ Kolmogorov diffusion $\boldsymbol{I}$, begun at $\boldsymbol{x} = (x_0, x_1, \ldots, x_k)^\top$, using

$$I_r(t) = \widetilde{I}_r(t) + x_r + tx_{r-1} + \frac{t^2}{2}x_{r-2} + \cdots + \frac{t^r}{r!}x_0. \tag{2}$$

This definition arises from considering the integral curve of χ_0 determined by the initial values $\boldsymbol{x} = (x_0, x_1, \ldots, x_k)^\top$ of $\boldsymbol{I}$ at time 0, and using linearity.

The Kolmogorov diffusion is simple enough to allow for explicit calculations, and yet exhibits interesting properties from the point of view of probabilistic coupling. Much of the interest of the index-k Kolmogorov diffusion $\boldsymbol{I}$ lies in the observation that the random vector $\boldsymbol{I}(t)$ has a positive continuous density over all $\mathbb{R}^{k+1}$ for any positive time $t > 0$. This is related of course to Hörmander's hypoellipticity theorem (since $\boldsymbol{I}$ is indeed a hypoelliptic diffusion), but can be seen much more directly by calculating the covariance structure of $\boldsymbol{I}(t)$ and noting that the resulting variance–covariance matrix is nonsingular. This computation is carried out in Section 2, but first we introduce some relevant concepts from coupling.

1.2. Maximal couplings

The technique of probabilistic coupling dates back to Doeblin [8]. However, Griffeath [10] was the first to prove a remarkable optimization result: it is possible to construct a coupling in a *maximal* way, in the following sense. Given two coupled copies X and $\widetilde{X}$ of a random process, let T be the (random) *coupling time*, namely $T = \inf\{t > 0 : X_s = \widetilde{X}_s \text{ for all } s \geq t\}$. Then a given coupling is *maximal* if it simultaneously minimizes $\mathbb{P}[T > t]$ for all $t > 0$; in particular, $\mathbb{P}[T > t]$ is then given by the total variation distance between the distributions of X_t and $\widetilde{X}_t$. Thus, a maximal coupling occurs at altogether the fastest possible rate. Griffeath proved the existence of maximal couplings for discrete-time Markov chains. Pitman [21] gave an elegant explicit construction for homogeneous discrete-time Markov chains on a countable state space (and in fact his construction generalizes easily), while Goldstein [9] extended the result to general discrete-time random processes. For further generalizations, see, e.g. [24].

The notion of coupling contains many subtleties. For example, in general (and in contrast to the Markovian case described below), the simpler random time $\inf\{t > 0 : X_t = \widetilde{X}_t\}$ may fail to produce a coupling time; this relates to the notion of *faithful coupling* [22] (note however that constructions of Pitman type deliver maximal couplings for which this simpler random time does produce a coupling). We note in passing that one can relax the definition of coupling to allow for arbitrary time shifts: this is the notion of *shift coupling* [25]. In general, it is hard to produce explicit constructions of maximal couplings. On the other hand, it is much easier to build *Markovian couplings*: couplings which jointly produce a Markov process, whose transition-probability kernel has as marginals the transition-probability kernels of the coupled diffusions. There is an important detail here: the marginals must be transition-probability kernels *with respect to the natural filtration of the coupled process*. A slightly more general notion of coupling, that of *immersed coupling* (also called *co-adapted coupling*) can be described succinctly and with more clarity: the martingales for the filtrations of X and $\widetilde{X}$ remain martingales for the joint filtration of $(X, \widetilde{X})$ [16], so that the filtrations of X and $\widetilde{X}$ are immersed in the joint filtration of X and $\widetilde{X}$. However, in the following we will restrict ourselves to consideration of Markovian couplings.

Note that the above constructions beg the question of whether the coupling time T can be chosen to be almost surely finite, in which case the coupling is said to be *successful*. In the case of the Kolmogorov diffusion it can be shown that successful Markovian couplings exist [5], [17]; the extent to which this is the case for a general nilpotent diffusion is an interesting open question (but see [5], [14], [15], and [16]).

Markovian couplings are relatively easy to construct and verify (consider for example the classic reflection coupling of Brownian motion); however, in general, they will not be maximal (the Brownian reflection coupling provides a rare exception). In the case of driftless Brownian motions on rather general spaces, Kuwada [19] showed that maximal Markovian couplings occur only in highly symmetrical cases; it is shown in [4] that if a smooth elliptic diffusion on a Riemannian manifold admits a Markovian maximal coupling then the manifold must be a space form, the diffusion must be Brownian motion plus drift, and the drift must arise from a continuous one-parameter group of symmetries of the space form (possibly augmented by dilations in the Euclidean case). Thus (at least in the elliptic case), Markovian maximal couplings are very rare.

1.3. Efficient couplings

We have asserted that we should not expect maximal couplings to be readily constructible. But, for practical purposes, it will often suffice to obtain a coupling such that the probability of

failing to couple by time t is comparable (asymptotically in t) to the probability of failing to couple maximally by t.

Efficient couplings were introduced in [6] for the case of Markov chains in which there is rapid convergence to a stationary distribution: a coupling is *efficient* if the (presumed exponential) rate of coupling from generic initial states equals the (presumed exponential) rate of convergence to stationarity. Here we generalize to cases where a stationary distribution need not exist: instead of considering exponential rates of convergence, we consider the rate of coupling compared to the maximum possible rate.

Definition 1. Let $\mu_{x,y}$ be a successful coupling of two Markov processes X and Y with state space $\mathscr{S}$. Suppose that X and Y start from distinct points $x, y \in \mathscr{S}$ respectively, with coupling time τ, and let $d_t(x, y)$ denote the corresponding total variation distance between their distributions. We call $\mu_{x,y}$ an *efficient coupling* if there exists a positive constant $C(x, y)$ such that

$$\frac{\mu_{x,y}\{\tau > t\}}{d_t(x, y)} \leq C(x, y)$$

for all $t > 0$ (note that $d_t(x, y) \leq \mu_{x,y}\{\tau > t\}$ follows from Aldous's inequality).

We call the family of couplings $\{\mu_{x,y} : x, y \in \mathscr{S}\}$ an *efficient coupling strategy* if $\mu_{x,y}$ is an efficient coupling for every pair of distinct starting points $x, y \in \mathscr{S}$.

Note that the constant $C(x, y)$ and the maximal coupling rate may, and often do, depend on initial conditions. Note also that it is entirely possible for a diffusion to exhibit efficient couplings from some distinct pairs of starting points, while still failing to possess an efficient coupling strategy from all possible distinct pairs of starting points. (Examples can be constructed using diffusions which are ordered pairs of independent component diffusions, of which the first coordinate is efficient, and the second inefficient.)

Even when maximal or efficient Markovian couplings do not exist, it is still possible that there may be a Markovian coupling which is *optimal*, in the sense of simultaneously minimizing $\mathbb{P}[T > t]$ for all $t > 0$ over the class of all Markovian couplings. Such a coupling exists in the case of simultaneous coupling of Brownian motion and its local time at zero [16].

1.4. Questions

The purpose of this paper is to address the following questions.

(Q1) Is there a maximal Markovian coupling for the generalized Kolmogorov diffusion $\boldsymbol{I}$? (The work of [4] is primarily concerned with smooth elliptic diffusions and its main results do not directly apply.)

(Q2) If not, is there an efficient Markovian coupling for the generalized Kolmogorov diffusion $\boldsymbol{I}$?

(Q3) Maximal couplings are typically hard to compute. Is there an efficient non-Markovian coupling for the generalized Kolmogorov diffusion $\boldsymbol{I}$ which is relatively easy to describe?

We shall also address the question of whether there might be an optimal Markovian coupling, but only in the case of the classic Kolmogorov diffusion $(B, \int B \, \mathrm{d}t)$. We restrict ourselves here to the classic case; indeed, in the case of generalized Kolmogorov diffusions $\boldsymbol{I}$ of index exceeding 2 we know only of implicit and indirect constructions of Markovian couplings [17], while construction of a successful Markovian coupling at index 2 is direct, but requires a somewhat involved analysis.

In Section 1 we describe basic coupling concepts and give an exact definition of the generalized Kolmogorov diffusion. In Section 2 we carry out some basic calculations for the generalized Kolmogorov diffusion, from which is derived the straightforward Theorem 1, which establishes the rate of maximal coupling. In Section 3 we establish a sequence of negative results: for the generalized Kolmogorov diffusion, it is not possible to construct a Markovian maximal coupling (Theorem 2), nor are there any efficient Markovian coupling strategies (Theorem 3). Moreover, for the classical Kolmogorov diffusion, Theorem 5 (by way of the analytical Theorem 4 giving coupling rates for the coupling described in [5]) shows that there can be no optimal Markovian strategy (we restrict here to the classical case of $(B, \int B \, dt)$ to avoid considerable potential complexity). In Section 4 we give a more positive result in the form of Theorem 6; working in the class of non-Markovian couplings which look ahead only over bounded intervals of time, this theorem exhibits a simple but efficient non-Markovian coupling strategy. Finally, we discuss possible future research directions in Section 5.

2. Explicit calculations for the Kolmogorov diffusion

The Kolmogorov diffusion is a linear Gaussian diffusion, and therefore permits explicit calculation. From (2), the Kolmogorov diffusion $\boldsymbol{I}(t)$ of index k, using $\boldsymbol{x}$ as the initial configuration, can usefully be expressed in terms of the standard Kolmogorov diffusion in vector form:

$$\boldsymbol{I}(t) = \widetilde{\boldsymbol{I}}(t) + \boldsymbol{H}(t)\boldsymbol{x}.$$

Here the $(k+1) \times (k+1)$ lower-triangular matrix $\boldsymbol{H}(t)$ can be written as $\boldsymbol{H}(t) = \boldsymbol{D}(t)\boldsymbol{H}\boldsymbol{D}(t^{-1})$, where $\boldsymbol{D}(t)$ is the $(k+1) \times (k+1)$ diagonal matrix with entries $1, t, \ldots, t^k$ running down the diagonal, and

$$H_{a,b} = \begin{cases} 1/(a-b)! & \text{if } a \geq b, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that $\boldsymbol{H}(t+s) = \boldsymbol{H}(t)\boldsymbol{H}(s)$. Note also that $\det \boldsymbol{H} = 1$, so that $\boldsymbol{H}$ and $\boldsymbol{H}(t)$ (for $t > 0$) are nonsingular matrices.

Consider the Kolmogorov diffusion $\widetilde{\boldsymbol{I}}$ of index k and begun at $\boldsymbol{0}$. Mathematical induction establishes the following linear relationship of Volterra-integral type:

$$\widetilde{I}_k(t) = \int_0^t \frac{(t-s)^{k-1}}{(k-1)!} B(s) \, ds \quad \text{for } k > 0.$$

Fixing T, and defining $F(T-t) = (T-t)^k/k!$ for $k > 0$, an application of Itô's formula to $F(T-t)B(t)$ for standard Brownian motion B shows that

$$\widetilde{I}_k(T) = \int_0^T \frac{(T-t)^k}{k!} \, dB(t).$$

In fact, this holds for all $k \geq 0$. Applying L^2 isometry for Itô integrals, we find that, for all $a, b \geq 0$,

$$\mathbb{E}[\widetilde{I}_a(T)\widetilde{I}_b(T)] = \int_0^T \frac{(T-t)^a}{a!} \frac{(T-t)^b}{b!} \, dt = \binom{a+b}{a} \frac{T^{a+b+1}}{(a+b+1)!}.$$

Thus, the variance–covariance matrix $\boldsymbol{V}(T)$ for $\widetilde{\boldsymbol{I}}(T)$, equivalently $\boldsymbol{I}(T)$, is given by $\boldsymbol{V}(T) = T\boldsymbol{D}(T)\boldsymbol{V}\boldsymbol{D}(T)$, where

$$V_{a,b} = \binom{a+b}{a} \frac{1}{(a+b+1)!} \quad \text{for } a, b \geq 0. \tag{4}$$

Note that V is nonsingular: given a vector a of coefficients, the L^2-isometry implies that

$$a^\top V a = \mathrm{var}\left[\int_0^1 \sum_{k=0}^n a_k \frac{(1-t)^k}{k!}\, \mathrm{d}B(t)\right] = \int_0^1 \left(\sum_{k=0}^n a_k \frac{(1-t)^k}{k!}\right)^2 \mathrm{d}t = \int_0^1 \left(\sum_{k=0}^n a_k \frac{t^k}{k!}\right)^2 \mathrm{d}t,$$

and this integral is zero only if the polynomial $\sum_{k=0}^n a_k t^k/k!$ vanishes identically, forcing $a = 0$. Thus, V is symmetric positive definite.

Consider the Cholesky decomposition of the symmetric matrix V (unique up to '±' sign, since V is positive definite). This provides a lower-triangular nonsingular matrix L such that

$$V = LL^\top. \tag{5}$$

Applying (4), (5), and the lower-triangular nature of L, note for future use that the top row of L can be taken to be $(1, 0, \ldots, 0)$. There follows a representation of the distribution of $I(T)$ in terms of a vector W filled with $k+1$ independent standard Brownian motions: for fixed T, we obtain

$$I(T) \stackrel{\mathrm{D}}{=} D(T)(HD(T^{-1})x + LW(T)). \tag{6}$$

(Note however that this equality in distribution cannot be translated into a sample-path equality as T varies.)

Suppose that $I^{(1)}$ and $I^{(2)}$ are begun at $x^{(1)}$ and $x^{(2)}$, respectively. Then (6) can be used to compute the total variation distance between the Gaussian distributions $\mathcal{L}(I^{(1)}(T))$ and $\mathcal{L}(I^{(2)}(T))$. These Gaussian distributions have the same variance–covariance matrix, and therefore the total variation distance is given by the following expression, with $z = x^{(1)} - x^{(2)}$:

$$\mathrm{dist}_{\mathrm{TV}}(\mathcal{L}(I^{(1)}(T)), \mathcal{L}(I^{(2)}(T))) = \mathbb{P}\left[|N(0,1)| \le \frac{1}{2\sqrt{T}} \|L^{-1}HD(T^{-1})z\|\right]$$
$$\le \frac{\|L^{-1}HD(T^{-1})z\|}{\sqrt{2\pi T}}. \tag{7}$$

For large T, the bound and the total variation distance are asymptotically equivalent.

The relationship between coupling, maximal coupling, and the total variation distance [10], [21], [9] (in particular Aldous's inequality) immediately establishes sharp bounds on the coupling rate for Kolmogorov diffusions, with the coupling rate depending on the extent of agreement between low-index initial conditions.

Theorem 1. (Maximal coupling rate for the generalized Kolmogorov diffusion.) *If $I^{(1)}$ and $I^{(2)}$ are coupled copies of the generalized Kolmogorov diffusion begun at $x^{(1)}$ and $x^{(2)}$, respectively, with $z_0 = \cdots = z_{r-1} = 0$ and $z_r \ne 0$ for $z = x^{(1)} - x^{(2)}$, and τ denotes the coupling time, then*

$$\mathbb{P}[\tau > T] \ge \mathbb{P}\left[|N(0,1)| \le \frac{1}{2\sqrt{T}} \|L^{-1}HD(T^{-1})z\|\right] \sim \mathrm{O}\left(\frac{1}{T^{r+1/2}}\right),$$

and this sharp lower bound is achieved by a maximal coupling between $I^{(1)}$ and $I^{(2)}$.

Proof. It is classical that the maximal coupling achieves the upper bound provided by the total variation distance [10], [21], [9]. Hence, this result follows directly from (7) and the lower-triangular nature of H and L, and, hence, of $L^{-1}H$. From (7), $\mathbb{P}[\tau > T]$ is controlled by $\mathbb{P}[|N(0,1)| \le \|L^{-1}HD(T^{-1})z\|/(2\sqrt{T})]$. But if the first r indices of z vanish then $\|L^{-1}HD(T^{-1})z\| = \mathrm{O}(T^{-r})$, and the result follows from the naïve bound

$$\sqrt{\frac{2}{\pi}} \ell e^{-\ell^2/2} \le \frac{1}{\sqrt{2\pi}} \int_{-\ell}^{\ell} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \le \sqrt{\frac{2}{\pi}} \ell. \qquad \square$$

This has implications for efficient (albeit possibly non-Markovian) coupling strategies for the generalized Kolmogorov diffusion: coupling will occur much faster if the initial states give the same initial Brownian locations and agree up to the first $r - 1 \geq 0$ iterated integrals.

3. Markovian couplings of Kolmogorov diffusions

In this section we consider the rate of coupling for *Markovian* couplings of Kolmogorov diffusions: can Markovian couplings be maximal in this case? Or efficient? Or, failing either of these, can there be Markovian maximal couplings of Kolmogorov diffusions which are optimal in the sense of coupling faster than any other Markovian couplings? We shall see that the answers to the first two of these questions are negative for Kolmogorov diffusions of whatever positive index. For the third question, we shall consider only classic (index-one) Kolmogorov diffusions, and show that at least in this case the answer is again negative.

3.1. Markovian Kolmogorov couplings are never maximal

For a very general class of Markov processes, if a Markovian maximal coupling exists for two copies of such a process started from distinct points, then this coupling must satisfy some specific properties. Maximal coupling has to occur at the space–time interface defined by equality of the transition probabilities of the two coupled diffusions. If, in addition, the coupling is Markovian then Varadhan asymptotics show that at a given time the interface must be of a specific form (a hyperplane, if the diffusive part of the diffusion is Brownian). Both properties follow from 'soft' arguments and do not depend explicitly on the specific process being considered. This is the content of Section 1.1 of [4] and the results there can be used to show the following.

Theorem 2. *There is no Markovian maximal coupling for the Kolmogorov diffusion of any positive index, started from any pair of distinct points.*

Proof. Consider the implications of the existence of a Markovian maximal coupling between two index-k generalized Kolmogorov diffusions $\boldsymbol{I}^+$ and $\boldsymbol{I}^-$, begun at different starting points, $\boldsymbol{x}_0^+$ and $\boldsymbol{x}_0^-$, respectively. Suppose that the coupling time is τ.

First consider the case where the last coordinates of $\boldsymbol{x}_0^+$ and $\boldsymbol{x}_0^-$ differ. By linearity, we may suppose that $\boldsymbol{x}_0^+ + \boldsymbol{x}_0^- = \boldsymbol{0}$ (the zero vector). By (6), for any fixed $t > 0$,

$$\boldsymbol{I}^{\pm}(t) \stackrel{\text{D}}{=} \boldsymbol{D}(t)(\boldsymbol{L}\boldsymbol{W}^{\pm}(t) + \boldsymbol{H}\boldsymbol{D}(t^{-1})\boldsymbol{x}_0^{\pm}).$$

Here $\boldsymbol{W}^{\pm}$ are $(k + 1)$-dimensional Brownian motions which are coupled (not necessarily in a Markovian manner). The Gaussian distributions of $\boldsymbol{I}^{\pm}(t)$ have the same nonsingular variance matrix $t\boldsymbol{V}(t) = t\boldsymbol{D}(t)\boldsymbol{V}\boldsymbol{D}(t)$: the corresponding probability densities agree on a hyperplane $\mathcal{H}(t; \boldsymbol{x}_0^+, \boldsymbol{x}_0^-)$ which runs through $\boldsymbol{0}$ and is orthogonal to the vector given by the vector expression

$$(t\boldsymbol{D}(t)\boldsymbol{V}\boldsymbol{D}(t))^{-1}\boldsymbol{D}(t)\boldsymbol{H}\boldsymbol{D}(t^{-1})\boldsymbol{z} = t^{-1}\boldsymbol{D}(t^{-1})\boldsymbol{V}^{-1}\boldsymbol{H}\boldsymbol{D}(t^{-1})\boldsymbol{z},$$

where $\boldsymbol{z} = \boldsymbol{x}_0^+ - \boldsymbol{x}_0^-$. Note that invertibility of $\boldsymbol{V}$ and $\boldsymbol{H}$ imply that this vector is nonzero.

It is convenient to scale this vector expression by t^{1+2k}, and so to deduce that the hyperplane $\mathcal{H}(t; \boldsymbol{x}_0^+, \boldsymbol{x}_0^-)$ is normal to the vector

$$(t^k \boldsymbol{D}(t^{-1}))\boldsymbol{V}^{-1}\boldsymbol{H}(t^k \boldsymbol{D}(t^{-1}))\boldsymbol{z}, \tag{8}$$

where $t^k \boldsymbol{D}(t^{-1})$ is a diagonal $(k + 1) \times (k + 1)$ matrix whose diagonal is composed of $t^k, t^{k-1}, \ldots, t, 1$.

The hyperplane $\mathcal{H}(t; x_0^+, x_0^-)$ separates $\mathbb{R}^{k+1}$ into disjoint half-spaces, one containing $I^+(0)$ and the other containing $I^-(0)$. We call the respective half-spaces $\mathcal{H}^+(t; x_0^+, x_0^-)$ and $\mathcal{H}^-(t; x_0^+, x_0^-)$. Now the following observations must necessarily hold for a Markovian maximal coupling μ of $I^\pm$ with coupling time τ.

1. Denote the probability densities corresponding to $I^\pm(t)$ by $p_t^\pm(x_0^\pm, \cdot)$. Let $\alpha_t(\cdot) = p_t^+(x_0^+, \cdot) - p_t^-(x_0^-, \cdot)$. Then [4, Lemma 2] states that, for any Borel-measurable set $A \in \mathbb{R}^{k+1}$,

$$\mu\{I^\pm(t) \in A, \tau > t\} = \int_A \alpha_t^\pm(z) \, dz,$$

   where

$$\alpha^+(z) = \max\{p_t^+(x_0^+, z) - p_t^-(x_0^-, z), 0\}$$

$$\text{and} \quad \alpha^-(z) = \max\{p_t^-(x_0^-, z) - p_t^+(x_0^+, z), 0\}.$$

   Thus, if coupling has not been successful by time t, then at time t the diffusions $I^\pm(t)$ must lie on different sides of the hyperplane $\mathcal{H}(t; x_0^+, x_0^-)$. Moreover, we may use the Gaussian nature of $I^\pm(t)$ to deduce that, under the conditioning $\tau > t$, the support of $I^+(t)$ (respectively $I^-(t)$) must be the whole of $\mathcal{H}^+(t; x_0^+, x_0^-)$ (respectively $\mathcal{H}^-(t; x_0^+, x_0^-)$).

2. Let μ_t denote the joint law of the coupled $I^\pm$ evaluated at time t. The coupling is Markovian and maximal, so [4, Lemma 3] shows that, for almost every pair (x^+, x^-) of distinct points in the support of μ_t, it will be the case that the forward processes $I^\pm(t + \cdot)$ must generate a new Markovian maximal coupling starting from x^+ and x^-, respectively. We will denote the set of such (x^+, x^-) by $\mathcal{M}(\mu_t)$.

The first observation shows that, for any positive t and s, when conditioned on $\tau > t + s$, the support of the conditional distribution of $I^+(t + s)$ must be the whole of $\mathcal{H}^+(t + s; x_0^+, x_0^-)$. Adding the second observation, we may deduce that, for all $(x^+, x^-) \in \mathcal{M}(\mu_t)$, when conditioning on $I^+(t) = x^+$ and $I^-(t) = x^-$, the support of the conditional distribution of $I^+(t + s)$ (also given $\tau > t + s$) must be the whole of $\mathcal{H}^+(s; x^+, x^-)$, where $\mathcal{H}(s; x^+, x^-)$ is the hyperplane on which the densities at time s agree for two generalized Kolmogorov diffusions begun at x^+ and x^-, respectively. Thus, for $(x^+, x^-) \in \mathcal{M}(\mu_t)$, we must have $\mathcal{H}^+(t + s; x_0^+, x_0^-) = \mathcal{H}^+(s; x^+, x^-)$.

Arguing as above, the common hyperplane must be normal to the vector

$$(s^k D(s^{-1})) V^{-1} H(s^k D(s^{-1}))(x^+ - x^-)$$
$$= (s^k D(s^{-1})) V^{-1}(s^k D(s^{-1})) \times (s^{-k} D(s)) H(s^k D(s^{-1}))(x^+ - x^-).$$

Now consider the limit of this vector as $s \downarrow 0$. We see that $(s^k D(s^{-1})) V^{-1}(s^k D(s^{-1}))$ converges to a matrix all of whose entries are zero save for the (k, k) position. Moreover, this entry is that same as the (k, k) entry of V^{-1}: since V is symmetric positive definite, it follows that V^{-1} is also symmetric positive definite, and therefore its (k, k) entry must be positive.

On the other hand, consider the lower-triangular matrix $(s^{-k} D(s)) H(s^k D(s^{-1}))$. All off-diagonal entries must tend to zero with s; the on-diagonal entries are not affected and, by (3), all are equal to 1.

These facts, along with the assumption that the last coordinate of z is nonzero, imply that as $s \downarrow 0$ so $(s^k D(s^{-1})) V^{-1}(s^k D(s^{-1}))(x^+ - x^-)$ converges to a vector parallel to $(0, 0, \ldots, 0, 1)^\top$, and this vector is nonzero if $s > 0$. Consequently, $\mathcal{H}(t; x_0^+, x_0^-)$ is normal to $e_k = (0, 0, \ldots, 0, 1)^\top$ for all positive t.

Together with (8), this tells us that, for each $t > 0$, there is a nonzero scalar λ_t such that

$$(t^k \boldsymbol{D}(t^{-1})) \boldsymbol{V}^{-1} (t^k \boldsymbol{D}(t^{-1})) \times (t^{-k} \boldsymbol{D}(t)) \boldsymbol{H}(t^k \boldsymbol{D}(t^{-1})) z = \lambda_t e_k,$$

or, equivalently,

$$(t^{-k} \boldsymbol{D}(t)) \boldsymbol{H}(t^k \boldsymbol{D}(t^{-1})) z = \lambda_t (t^{-k} \boldsymbol{D}(t)) \boldsymbol{V}(t^{-k} \boldsymbol{D}(t)) e_k. \tag{9}$$

For any $0 \le i \le k$, (9) yields

$$\sum_{j=0}^{i} t^{i-j} H_{i,j} z_j = t^{-k+i} V_{i,k} \lambda_t. \tag{10}$$

Putting $i = k$ into (10), we find that

$$\lambda_t = V_{k,k}^{-1} \sum_{j=0}^{k} t^{k-j} H_{k,j} z_j.$$

Substituting this value of λ_t back into (10), we obtain the following equation:

$$\sum_{j=0}^{i} t^{i-j} H_{i,j} z_j = \frac{V_{i,k}}{V_{k,k}} \sum_{j=0}^{k} t^{i-j} H_{k,j} z_j.$$

Setting $i = 0$ in the above equation and comparing coefficients of inverse powers of t, the explicit formulæ (3) for $\boldsymbol{H}$ and (4) for $\boldsymbol{V}$ lead to $z = 0$, and we thus obtain a contradiction of the initial hypothesis that $z_k \ne 0$.

For the case when the last coordinates of $\boldsymbol{x}_0^+$ and $\boldsymbol{x}_0^-$ agree, consider the largest $j < k$ such that $I_j^+(0) \ne I_j^-(0)$. Without loss of generality, suppose that $I_j^+(0) > I_j^-(0)$. Path continuity of the diffusion shows that there must exist $T > 0$ such that the measure value $\mu\{I_j^+(t) - I_j^-(t) > 0 \text{ for all } t \le T\}$ is positive. But

$$I_k^+(t) - I_k^-(t) = \int_0^t \int_0^{s_{k-1}} \cdots \int_0^{s_{j+1}} (I_j^+(s_j) - I_j^-(s_j)) \mathrm{d}s_j \cdots \mathrm{d}s_{k-1},$$

and, therefore, $\mu\{I_k^+(t) - I_k^-(t) > 0 \text{ for all } 0 < t \le T\} > 0$.

In particular, $\mu_T\{I_k^+(T) \ne I_k^-(T)\} > 0$. Under the hypothesis of existence of a Markovian maximal coupling starting from $\boldsymbol{x}_0^+$ and $\boldsymbol{x}_0^-$, and using [4, Lemma 3], we can find $(\boldsymbol{x}_T^+, \boldsymbol{x}_T^-) \in \mathcal{M}(\mu_T)$ such that the coupled forward processes $(I_k^+(T + \cdot), I_k^-(T + \cdot))$ started from $(\boldsymbol{x}_T^+, \boldsymbol{x}_T^-)$ create a Markovian maximal coupling. The previous argument applied to this coupling then leads to a contradiction. □

3.2. Markovian Kolmogorov couplings are never efficient

As before, let $d_t(\boldsymbol{x}, \boldsymbol{y})$ denote the total variation distance between the laws of generalized Kolmogorov diffusions (of the same positive index k) started from $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. From Theorem 1, if the first coordinates of $\boldsymbol{x}$ and $\boldsymbol{y}$ disagree then $d_t(\boldsymbol{x}, \boldsymbol{y}) \sim t^{-1/2}$, while if they agree then $d_t(\boldsymbol{x}, \boldsymbol{y}) \sim t^{-3/2}$ or even higher powers of t^{-1} (if further coordinates agree). That is, for efficient (possibly non-Markovian) coupling strategies (see Definition 1), the coupling happens much faster when the scalar Brownian motions, driving the coupled Kolmogorov diffusions of index k, both start from the same point. This turns out to be the key observation in proving Theorem 3. The theorem, on the absence of efficient Markovian coupling strategies, will follow as a direct corollary of the following lemma.

Lemma 1. *Let μ be any successful Markovian coupling of Kolmogorov diffusions starting from distinct points $\boldsymbol{x}$ and $\boldsymbol{y}$ whose first coordinates agree. Let τ denote the coupling time. Then there are positive constants C_μ and t_μ (possibly depending on the coupling μ) such that, for all $t \geq t_\mu$, $\mu\{\tau > t\} \geq C_\mu/\sqrt{t}$.*

Proof. Let $\boldsymbol{I}^{(1)}$ and $\boldsymbol{I}^{(2)}$ denote specific Markovian-coupled copies of the generalized Kolmogorov diffusions starting from $\boldsymbol{x}$ and $\boldsymbol{y}$ (with first coordinates agreeing but second coordinates disagreeing). This corresponds to a Markovian coupling of the driving Brownian motions B and $\tilde{B}$. Let $\mathcal{F}_t = \sigma\{\boldsymbol{I}^{(1)}(s), \boldsymbol{I}^{(2)}(s) \colon s \leq t\}$ be the corresponding σ-algebra of events determined by time t. Consider the possibility that $\mu\{B(t) = \tilde{B}(t)\} = 1$ for all t. A Fubini argument then shows that

$$\mu\{B(t) = \tilde{B}(t) \text{ for almost every } t > 0\} = 1.$$

By path continuity of Brownian motion, the above would imply that B and $\widetilde{B}$ are synchronously coupled μ-almost surely and, hence, $\mu\{\boldsymbol{I}^{(1)}(t) - \boldsymbol{I}^{(2)}(t) = \boldsymbol{x} - \boldsymbol{y}$ for all $t > 0\} = 1$. So a successful Markovian coupling cannot be obtained in this manner. Thus, for a successful Markovian coupling μ, there must exist $t_0 > 0$ such that $\mu\{B(t_0) \neq \tilde{B}(t_0)\} > 0$. Now, for every $t > t_0$,

$$\mu\{\tau > t\} = \mathbb{E}_\mu[\mathbb{E}_\mu[\mathbb{I}(\tau > t) \mid \mathcal{F}_{t_0}]],$$

where $\mathbb{E}_\mu$ represents expectation with respect to the probability measure μ. Introduce the ordinary Brownian coupling time $\tau^* = \inf\{s > t_0 \colon B(s) = \tilde{B}(s)\}$. Evidently this happens no later than the Kolmogorov coupling time, so $\mathbb{I}(\tau > t) \geq \mathbb{I}(\tau^* > t)$. As the coupling is Markovian, the shifted process $((\theta_{t_0} B(s), \theta_{t_0} \tilde{B}(s)) \colon s > 0)$ (when conditioned on $\mathcal{F}_{t_0}$) gives a coupling of Brownian motions starting from $(B(t_0), \tilde{B}(t_0))$. If $d_t^{\mathrm{BR}}(x, y)$ represents the total variation distance between the distributions of Brownian motions starting from x and y at time t, then we know that for all $t > 0$ and all x, y satisfying $|x - y| \leq 2\sqrt{t}$,

$$d_t^{\mathrm{BR}}(x, y) = \mathbb{P}\left[|N(0, 1)| \leq \frac{|x - y|}{2\sqrt{t}}\right] \geq \frac{1}{\sqrt{2\pi\mathrm{e}}} \frac{|x - y|}{\sqrt{t}}.$$

There exists $t_\mu > t_0$ such that, for all $t \geq t_\mu$,

$$\mathbb{E}_\mu[|B(t_0) - \tilde{B}(t_0)|; |B(t_0) - \tilde{B}(t_0)| \leq 2\sqrt{t - t_0}] \geq \sqrt{2\pi\mathrm{e}}\,C_\mu,$$

where $C_\mu = \mathbb{E}_\mu|B(t_0) - \tilde{B}(t_0)|/(2\sqrt{2\pi\mathrm{e}})$. (The left-hand side converges to $2\sqrt{2\pi\mathrm{e}}\,C_\mu$ as $t \to \infty$.) Thus, for $t \geq t_\mu$,

$$\begin{aligned}
\mathbb{E}_\mu[\mathbb{E}_\mu[\mathbb{I}(\tau > t) \mid \mathcal{F}_{t_0}]] &\geq \mathbb{E}_\mu[\mathbb{E}_\mu[\mathbb{I}(\tau^* > t) \mid \mathcal{F}_{t_0}]] \\
&\geq \mathbb{E}_\mu\big[d_{t-t_0}^{\mathrm{BR}}(B(t_0), \tilde{B}(t_0))\big] \\
&\geq \frac{1}{\sqrt{2\pi\mathrm{e}}}\mathbb{E}_\mu\left[\frac{|B(t_0) - \tilde{B}(t_0)|}{\sqrt{t - t_0}}; \frac{|B(t_0) - \tilde{B}(t_0)|}{\sqrt{t - t_0}} \leq 2\right] \\
&\geq \frac{C_\mu}{\sqrt{t - t_0}} \\
&\geq \frac{C_\mu}{\sqrt{t}},
\end{aligned}$$

which proves the lemma. $\qquad\square$

Theorem 3. *There exists no efficient Markovian coupling strategy (in the sense of Definition 1) for the Kolmogorov diffusion of index k (for $k > 0$).*

Proof. We argue by contradiction. If there is such an efficient Markovian coupling strategy, then there must exist an efficient Markovian coupling μ of the generalized Kolmogorov diffusions starting from x and y (with first coordinates agreeing but second coordinates disagreeing) with coupling time τ. Then, by Lemma 1, there exist positive constants C_μ and t_μ such that $\mu\{\tau > t\} \geq C_\mu t^{-1/2}$ for all $t \geq t_\mu$.

But, as noted in the discussion preceding Lemma 1, since the first coordinates of x and y agree, we have $d_t(x, y) \sim t^{-3/2}$, or even faster. Thus, efficiency of the coupling strategy (in the sense of Definition 1) must fail, proving the theorem. $\square$

Remark 1. Theorem 3 shows that any Markovian coupling strategy is nonefficient, in the sense that there exist some pairs of distinct starting points from which it is impossible to construct efficient couplings. But it does not imply Theorem 2, which shows nonmaximality of Markovian couplings from *any pair of distinct starting points*.

3.3. Markovian classic Kolmogorov couplings cannot be optimal

To begin with, recall the simplest version of the Markovian coupling for the classic Kolmogorov diffusion [5], [17]. Write U for the process corresponding to the difference between the Brownian motions, and V for the difference in the time integrals of the Brownian motions. We assume that $U_0 = 1$ and $V_0 = 0$; generalization to the case of arbitrary distinct starting points should be clear. We write the coupling probability measure as $\hat{\mu}_{(u,v)}$ when starting values are $U_0 = u$ and $V_0 = v$.

We describe the coupling strategy of [5] and [17] for $\hat{\mu}_{(1,0)}$. When U and V have the same sign, we apply reflection coupling (so that U evolves as a Brownian motion run at rate 4). Thus, the visits of (U, V) to the axis $V = 0$ have to occur at isolated instants of time: between each pair of visits the particle (U, V) describes a 'half-cycle' about the origin. Over the kth cycle, we actually apply reflection coupling until U hits $U = \pm 2^{-k}$ or $V = 0$, taking the sign '$\pm$' as the opposite of the sign of U at the start of the half-cycle. We then apply synchronous coupling till $V = 0$, so that U is held constant. As a result, $|U|$ will be no larger than 2^{-k} at the end of the kth half-cycle.

It is convenient to introduce some notation. Suppose that the kth cycle begins at time S_{k-1} (so $S_0 = 0$). Then $S_k = \inf\{t > S_{k-1}: V(t) = 0\}$. Let $T_k = \min\{S_k, \inf\{t > S_{k-1}: U(t) = (-1)^k 2^{-k}\}\}$ be the first time that U hits $U = (-1)^k 2^{-k}$, or S_k if that happens first. Thus, $[T_k, S_k]$ is the part of the half-cycle in which synchronous coupling is applied: one might think of this as the ballistic phase, while $[S_{k-1}, T_k)$ is the Brownian phase. Note that the ballistic phase will be trivial if the Brownian phase hits the $(V = 0)$-axis (that is, $T_k = S_k$).

The time of successful coupling is the time at which (U, V) hits the origin, and thus it is given by

$$\tau = \lim_{k \to \infty} S_k = \sum_{k=0}^{\infty} (S_{k+1} - S_k).$$

Borel–Cantelli arguments show that this limit is finite [5]. Our first task is to determine the precise rate of decay of the probability of failing to couple by time t.

Theorem 4. *Under the coupling $\hat{\mu}_{(1,0)}$ described above, the coupling time τ satisfies*

$$\frac{C_1}{t^{1/3}} \leq \hat{\mu}_{(1,0)}\{\tau > t\} \leq \frac{C_2}{t^{1/3}} \quad \text{for } t \geq 1, \tag{11}$$

for some positive constants C_1 and C_2.

Proof. Note that scaling arguments show that $S_{k+1} - S_k$ has the same distribution as $2^{-k}S_1$, so that the principal computation concerns the rate of decay of $\hat{\mu}_{(1,0)}\{S_1 > t\}$.

Now $S_1 = T_1 + 2V(T_1) = 2\int_0^{T_1}(\frac{1}{2} + U(s))\,\mathrm{d}s$, since either $V(T_1) = 0$ (when $T_1 = S_1$) or the velocity of V over $[T_1, S_1]$ is fixed at $-\frac{1}{2}$. Writing $\frac{1}{2} + U = 2\widetilde{W}$ for a Brownian motion $\widetilde{W}$ started from $\frac{3}{4}$ (over the time interval $[S_0, T_1]$), and noting that $S_1 = T_1 + 2V(T_1) = 4\int_0^{T_1}\widetilde{W}\,\mathrm{d}t$ (since $T_1 = \inf\{t > 0: \widetilde{W}_t = 0\}$), we see that $S_1/4$ can be viewed as having the density of the random area under a one-dimensional Brownian motion started from a positive level and measured till the first time it hits zero.

The density for this random area can be obtained from the calculations of [13, Equation (12)]. Consider a standard (rate-1) Brownian motion W started from $a > 0$. If $\sigma_a = \inf\{t > 0: W_t = 0\}$, then the martingale $\mathbb{E}[\exp(-\lambda\int_0^{\sigma_a} W(s)\,\mathrm{d}s) \mid \mathcal{F}_t]$ can be used to show that the moment generating function of $\int_0^{\sigma_a} W(s)\,\mathrm{d}s$ (viewed as a function of the starting position a) must solve an Airy partial differential equation. The moment generating function can be inverted to yield the distribution of $\int_0^{\sigma_a} W(s)\,\mathrm{d}s$:

$$\mathbb{P}\left[\int_0^{\sigma_a} W(s)\,\mathrm{d}s \in \mathrm{d}u\right] = \frac{2^{1/3}}{3^{2/3}\Gamma(1/3)}\frac{a}{u^{4/3}}\exp\left(-\frac{2a^3}{9u}\right)\mathrm{d}u. \tag{12}$$

From (12), it follows that, for $t \geq 1$,

$$\frac{6^{1/3}e^{-2a^3/9}}{\Gamma(1/3)}\frac{1}{t^{1/3}} \leq \mathbb{P}\left[\int_0^{\sigma_a} W(s)\,\mathrm{d}s > t\right] \leq \frac{6^{1/3}}{\Gamma(1/3)}\frac{1}{t^{1/3}}. \tag{13}$$

As $S_1/4$ has the density (12) with $a = \frac{3}{4}$, (13) gives us the following for $t \geq 1$:

$$\frac{24^{1/3}e^{-3/32}}{\Gamma(1/3)}\frac{1}{t^{1/3}} \leq \hat{\mu}_{(1,0)}\{S_1 > t\} \leq \frac{24^{1/3}}{\Gamma(1/3)}\frac{1}{t^{1/3}}. \tag{14}$$

We now apply this rate computation to $\tau = \sum_{k=0}^{\infty}(S_{k+1} - S_k)$. By Boole's inequality, scaling, and (14), if $t \geq 1$ then

$$\hat{\mu}_{(1,0)}\left\{\sum_{k=0}^{\infty}(S_{k+1} - S_k) > t\right\} \leq \sum_{k=0}^{\infty}\hat{\mu}_{(1,0)}\left\{S_{k+1} - S_k > \frac{\sqrt{2}-1}{\sqrt{2}}2^{-k/2}t\right\}$$

$$= \sum_{k=0}^{\infty}\hat{\mu}_{(1,0)}\left\{S_1 > \frac{\sqrt{2}-1}{\sqrt{2}}2^{k/2}t\right\}$$

$$\leq \frac{24^{1/3}}{\Gamma(1/3)}\left(\frac{\sqrt{2}}{\sqrt{2}-1}\right)^{1/3}\frac{1}{1 - 2^{-1/6}}\frac{1}{t^{1/3}},$$

and this establishes the right-hand inequality in (11). On the other hand, (14) shows that if $t \geq 1$ then

$$\hat{\mu}_{(1,0)}\left\{\sum_{k=0}^{\infty}(S_{k+1} - S_k) > t\right\} \geq \hat{\mu}_{(1,0)}\{S_1 > t\} \geq \frac{24^{1/3}e^{-3/32}}{\Gamma(1/3)}\frac{1}{t^{1/3}},$$

and this establishes the left-hand inequality in (11), thus completing the proof. □

The order of decay of the failure probability of coupling by time t is $t^{-1/3}$; therefore, this Markovian coupling is far from efficient even when the Brownian motions start from different points. In principle similar analyses should be possible for Markovian couplings of generalized

Kolmogorov diffusions, though in such cases progress is difficult because we know only of implicitly defined Markovian coupling constructions for index exceeding 2, while analysis for the case of index 2 is complicated [17].

We now consider the interesting question, can there be an *optimal coupling rate*, optimizing over all Markovian couplings but obtained by a single Markovian coupling of a Kolmogorov diffusion? For a study of this question in the different context of coupling for Brownian motion together with a local time, see [16]. In contrast to the case of [16], we will prove that such a coupling cannot exist for the classical Kolmogorov diffusion.

Theorem 5. *There exists no optimal Markovian coupling strategy for two copies of the classical Kolmogorov diffusion.*

Proof. Fix attention on classical Kolmogorov diffusions which initially differ only in their second coordinate. Denote by U the difference between the Brownian motions and by V the difference in the time integrals. By scaling and stopping arguments, we may concentrate on the situation in which $U_0 = 0$ and $V_0 = 1$. Our strategy will be to obtain, for each $t > 0$, a specific Markovian coupling $\mu^{(t)}$ with coupling time $\tau^{(t)}$ such that $\mu^{(t)}\{\tau^{(t)} > t\} \leq C/t$ for all $t > 0$, where the constant C does not depend on t. If there were to exist an optimal Markovian coupling ν with coupling time τ^*, then the above would imply that it should satisfy $\nu\{\tau^* > t\} \leq C/t$ for all $t > 0$. This would contradict Lemma 1, which shows that, for any (successful) Markovian coupling ν, there must exist positive constants C_ν and t_ν such that, for all $t \geq t_\nu$, $\nu\{\tau^* > t\} \geq C_\nu t^{-1/2}$.

First we must describe the coupling $\mu^{(t)}$. The coupling $\mu^{(t)}$ differs from the coupling described at the start of this subsection essentially only by early completion of the Brownian phase of its initial half-cycle.

1. Couple the Brownian motions by reflection till time

$$T_1' = \inf\left\{t > 0 \colon U_t = -\frac{4}{t}\right\} \wedge \inf\{t > 0 \colon V_t = 0\}.$$

    Then couple synchronously till time $S_1 = \inf\{t \geq T_1' \colon V_t = 0\}$.

2. Starting from $(U_{S_1}, 0)$, employ the coupling $\hat{\mu}_{(U_{S_1},0)}$ described at the start of this subsection.

In the following, $C_1, C_2, \ldots$ will be positive constants that do not depend on t.

First observe that T_1' is smaller than the hitting time of the (rate-4) Brownian motion U on the level $4/t$ when started at 0. Therefore, arguments using the reflection principle show that $\mu^{(t)}\{T_1' > 1\} \leq C_1/t$. On the other hand,

$$V(T_1') = \int_0^{T_1'} \left(U(s) + \frac{4}{t}\right)ds - \frac{4}{t}T_1' \leq \int_0^{T_1'} \left(U(s) + \frac{4}{t}\right)ds,$$

and we can apply (12) to deduce that $\mu^{(t)}\{V_{T_1'} > 2\} \leq C_2/t$. Now $S_1 - T_1' \leq \frac{1}{2}t$ on the event $\{V_{T_1'} \leq 2\}$. By scaling arguments and Theorem 4,

$$\mu^{(t)}\left\{\tau^{(t)} - S_1 > \frac{t}{2}\right\} \leq \hat{\mu}_{(4/t,0)}\left\{\tau > \frac{t}{2}\right\} = \hat{\mu}_{(1,0)}\left\{\frac{4^2}{t^2}\tau > \frac{t}{2}\right\} \leq \frac{C_3}{t},$$

where τ denotes the coupling time under the coupling strategy described at the start of this subsection.

Combining the above facts, we obtain $\mu^{(t)}\{\tau^{(t)} > t + 1\} \leq C_4/t$. The theorem follows. $\qquad\square$

4. Efficiency for the finite-look-ahead coupling

Despite these negative results, nevertheless it is possible to exhibit a simple and explicit efficient coupling strategy if we allow couplings which are allowed finite but varying amounts of precognition. In this section, we will describe a simple non-Markovian coupling strategy which achieves efficiency. Our approach will be to divide time into successive intervals $[S_n, S_{n+1}]$ of growing size and then to couple the driving Brownian motions (and hence the Kolmogorov diffusions) according to a non-Markovian recipe on each such interval. We call this coupling a *finite-look-ahead coupling* as the coupling construction on each interval $[S_n, S_{n+1}]$, although non-Markovian, requires information on the driving Brownian paths only till time S_{n+1}. Furthermore, the coupling of the future paths of the Kolmogorov diffusions conditional on the paths run up till time S_n depends on the past only through the values taken at time S_n. Thus, the coupling restricted to times S_n (for $n = 0, 1, \ldots$) can be considered to have a Markovian property.

Recall that in (6) we wrote down a representation for the Kolmogorov diffusion in terms of a Brownian motion vector $\boldsymbol{W}$ at time T. Note in particular that the lower-triangular form of $\boldsymbol{L}$ means that (6) represents $I_0(T)$ by $I_0(T) \overset{\mathrm{D}}{=} x_0 + W_0(T)$. However, this holds only for the stipulated fixed time T; we cannot maintain the representation (6) of $\boldsymbol{I}(T)$ in terms of the full vector $\boldsymbol{W}$ of independent standard Brownian motions while simultaneously writing $I_0(t) = x_0 + W_0(t)$ for $0 \le t \le T$. Representation (6) can be best understood as a fragment of an infinite-dimensional representation as follows. Consider the initial Brownian path $\{B(t) : 0 \le t \le T\}$ as a Gaussian vector in the infinite-dimensional space $C([0, T])$ of continuous paths. Realize this as the evaluation at time T of an infinite-dimensional Brownian motion $\mathscr{B}$ starting at the zero path and taking values in the Banach space $C([0, T])$, and evolving in 'algorithmic time' (as opposed to the 'process time' t used for the stochastic process $\{B(t) : 0 \le t \le T\}$). A candidate for this is given by the Karhunen–Loève expansion

$$\mathscr{B}(\zeta, t) = x_0 + \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, w_k(\zeta) f_k\left(\frac{t}{T}\right) \qquad (15)$$

for $\zeta, t \in [0, T]$, where $\lambda_k = 1/((k - \frac{1}{2})^2 \pi^2)$ and $f_k(t) = \sqrt{2} \sin((k - \frac{1}{2})\pi t)$ respectively denote the eigenvalues and eigenfunctions of the covariance kernel of Brownian motion viewed as a bounded operator acting on $L^2[0, 1]$, and the w_k are independent standard Brownian motions. Here ζ represents the algorithmic time and t represents the process time. (See [7] for more on stochastic analysis on infinite-dimensional spaces.)

Set $I_0(t) = x_0 + \mathscr{B}(T, t)$. For $1 \le r \le k$ and $t \in [0, T]$, the iterated integrals $I_r(t)$ can be viewed as $\mathscr{I}_r(T, t)$ where $\{\mathscr{I}_r(\zeta, t) : \zeta, t \in [0, T]\}$ have the representation

$$I_r(\zeta, t) = x_r + t x_{r-1} + \frac{t^2}{2} x_{r-2} + \cdots + \frac{t^r}{r!} x_0 + T^r \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, w_k(\zeta) f_{r,k}\left(\frac{t}{T}\right) \qquad (16)$$

with $f_{r,k}(t) = \int_0^t \cdots \int_0^{s_1} f_k(s_0)\, \mathrm{d}s_0\, \mathrm{d}s_1 \cdots \mathrm{d}s_{r-1}$.

It follows from (6) and (16) that, for a fixed process time T, the random process $\{\boldsymbol{W}(\zeta) : \zeta \in [0, T]\}$ obtained by

$$\boldsymbol{W}(\zeta) = \boldsymbol{L}^{-1} \boldsymbol{D}(T^{-1})(\boldsymbol{I}(\zeta, T) - \boldsymbol{D}(T) \boldsymbol{H} \boldsymbol{D}(T^{-1}) \boldsymbol{x}) \qquad (17)$$

is a standard $(k + 1)$-dimensional Brownian motion evolving in algorithmic time up to time T. Furthermore, from the representation in (16), it follows that the Brownian motion $\boldsymbol{W}$, obtained in this way, does not depend on T. Thus we can write the components W_j of $\boldsymbol{W}$ as linear

combinations of the Brownian motions w_k, i.e.

$$W_j(\zeta) = \sum_{k=1}^{\infty} w_k(\zeta) e_{j,k}, \tag{18}$$

where the $e_{j,k}$ do not depend on T.

It will be convenient to define the infinite matrix E whose (j, k)th entry is given by $e_{j,k}$, and to note that its rows must form an independent orthonormal set in the sequence space l^2.

We now describe the *finite-look-ahead coupling* by constructing the coupled paths in each of the successive time intervals (look-ahead-blocks) $[T_1 + \cdots + T_n, T_1 + \cdots + T_{n+1}]$ using the Brownian motion $\mathcal{B}_n$ on the infinite-dimensional space $C[0, T_n]$ described in (15), for $T_n = \alpha^n$ for some fixed $\alpha > 1$. In the following, set $S_0 = 0$ and $S_n = T_1 + \cdots + T_n$ to be the time of initiation of the nth look-ahead-block.

Before commencing detailed analysis, we provide a brief heuristic description of the coupling. Recall that the reflection coupling of Brownian motions started from distinct points gives a maximal coupling by using reflection on one Brownian path to produce the other (till the coupling time), using the hyperplane that bisects the line joining their respective starting points. Although the Kolmogorov diffusion $\{I(t) \colon t \in [0, T]\}$ is not a Brownian motion in process time, the process $\{W(\zeta) \colon \zeta \in [0, T]\}$ obtained from $I(\zeta, T)$ in (17) for a fixed process time T evolves as a Brownian motion in algorithmic time. With this observation in mind, we couple two Kolmogorov diffusions $I^{(1)}$ and $I^{(2)}$ started from $x^{(1)}$ and $x^{(2)}$, respectively, on the block $[0, T]$ as follows.

1. Couple two infinite-dimensional Brownian motions $\{\mathcal{B}_n^{(1)}(\zeta, t) \colon \zeta, T \in [0, T]\}$ and $\{\mathcal{B}_n^{(2)}(\zeta, t) \colon \zeta, T \in [0, T]\}$ in such a way that the processes $W^{(1)}$ and $W^{(2)}$ obtained from $\mathcal{B}_n^{(1)}$ and $\mathcal{B}_n^{(2)}$, respectively, via (16) and (17) are reflection coupled (in algorithmic time) by reflecting in the hyperplane bisecting the initial discrepancy vector $W^{(1)}(0) - W^{(2)}(0)$.

2. Repeat this construction on each block $[S_n, S_{n+1}]$, updating the starting points of the respective Kolmogorov diffusions to $I^{(1)}(S_n)$ and $I^{(2)}(S_n)$.

This coupling of the infinite-dimensional Brownian motions projects down to a coupling of the corresponding driving Brownian motions (and hence the Kolmogorov diffusions) by setting

$$I_0^{(i)}(t) = I_0^{(i)}(S_n) + \mathcal{B}_n^{(i)}(T_{n+1}, t - S_n) \tag{19}$$

for $t \in [S_n, S_{n+1}]$ and $i = 1, 2$.

It is reasonable to expect that if such a coupling is achieved then it will be efficient, as we are using reflection coupling of Brownian motions (in algorithmic time) in each block. The coupling is analysed in what follows, and efficiency of the coupling is shown in Theorem 6.

We will give an inductive description of the coupling for Kolmogorov diffusions $I^{(1)}$ and $I^{(2)}$ started from $x^{(1)}$ and $x^{(2)}$, respectively. Suppose that we have constructed the coupling on $[0, S_n]$. If $I^{(1)}(S_n) = I^{(2)}(S_n)$, we can synchronously couple the Brownian motions $I_0^{(1)}$ and $I_0^{(2)}$ after S_n. So, henceforth, we assume that $I^{(1)}(S_n) \neq I^{(2)}(S_n)$. We will couple two infinite-dimensional Brownian motions $\mathcal{B}_n^{(1)}$ and $\mathcal{B}_n^{(2)}$ on the block $[0, T_{n+1}]$, which are represented by the Karhunen–Loève expansion

$$\mathcal{B}_n^{(i)}(\zeta, t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, w_{n,k}^{(i)}(\zeta) f_k\left(\frac{t}{T_{n+1}}\right) \tag{20}$$

for $\zeta, t \in [0, T_{n+1}]$ and $i = 1, 2$. The corresponding coupling for the Brownian motions $I_0^{(1)}$ and $I_0^{(2)}$ (and thus for the entire diffusions $\boldsymbol{I}^{(1)}$ and $\boldsymbol{I}^{(2)}$) on the block $[S_n, S_{n+1}]$ can then be obtained by (19).

The $(k + 1)$-dimensional Brownian motions running in algorithmic time obtained from the iterated integrals by (17) are denoted by $\boldsymbol{W}_n^{(1)}$ and $\boldsymbol{W}_n^{(2)}$, respectively.

Write $\boldsymbol{Z}_n = \boldsymbol{I}^{(1)}(S_n) - \boldsymbol{I}^{(2)}(S_n)$. Define the unit vectors

$$\boldsymbol{v}_n = \frac{\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n}{\|\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n\|}$$

(taking $\boldsymbol{v}_0 = \boldsymbol{L}^{-1}\boldsymbol{z}/\|\boldsymbol{L}^{-1}\boldsymbol{z}\|$) and

$$\boldsymbol{\eta}_n = \frac{(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_n}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_n\|}.$$

Let $\{B_{n,j}\colon j \geq 1\}$ be independent standard scalar Brownian motions. Recall the matrix $\boldsymbol{E}$ defined just after (18). The infinite vector $\boldsymbol{v}_{n,1} = \boldsymbol{E}^T \boldsymbol{\eta}_n$ has l^2-norm 1. We can extend it to an orthonormal basis of l^2, say $\{\boldsymbol{v}_{n,1}, \boldsymbol{v}_{n,2}, \dots\}$. Let $\boldsymbol{P}$ denote the (infinite) orthogonal matrix with columns formed by these vectors. Note that $\boldsymbol{P}$ is a unitary matrix, and therefore the rows of $\boldsymbol{P}$ are also orthonormal. Now, we can define a coupling of the component Brownian motions $w_{n,k}^{(i)}$ in terms of the matrix $\boldsymbol{P}$ and the Brownian motions $B_{n,j}$. Define the stopping time

$$\sigma_n = \inf\{\zeta > 0\colon B_{n,1}(\zeta) = -\tfrac{1}{2}\|\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(T_{n+1}^{-1})\boldsymbol{Z}_n\|\}.$$

Then the coupling is a reflection coupling as follows:

$$w_{n,k}^{(i)}(\zeta) = \begin{cases} (-1)^{i+1}P_{k,1}B_{n,1}(\zeta) + \sum_{j=2}^{\infty} P_{k,j}B_{n,j}(\zeta) & \text{when } \zeta \leq \sigma_n, \\ w_{n,k}^{(i)}(\sigma_n) + \sum_{j=1}^{\infty} P_{k,j}(B_{n,j}(\zeta) - B_{n,j}(\sigma_n)) & \text{when } \zeta > \sigma_n, \end{cases} \quad (21)$$

for $i = 1, 2$. This gives the coupling between $\mathscr{B}_n^{(1)}$ and $\mathscr{B}_n^{(2)}$, and thus the corresponding Kolmogorov diffusions $\boldsymbol{I}^{(1)}$ and $\boldsymbol{I}^{(2)}$, via (20).

Note that the reflection coupling recipe (21) along with (18) give us

$$\boldsymbol{W}_n^{(1)}(\zeta) - \boldsymbol{W}_n^{(2)}(\zeta) = 2B_{n,1}(\zeta \wedge \sigma_n)\boldsymbol{\eta}_n. \quad (22)$$

Thus, the $(k+1)$-dimensional Brownian motions $\boldsymbol{W}_n^{(1)}$ and $\boldsymbol{W}_n^{(2)}$ (running in algorithmic time) are coupled in such a way that their difference is a rate-4 scalar Brownian motion running along the vector $\boldsymbol{\eta}_n$. This will be the crucial fact we use to prove efficiency of this coupling strategy in the following theorem.

Theorem 6. (Efficient finite-look-ahead coupling strategies.) *The finite-look-ahead coupling of the Kolmogorov diffusion of index k, constructed as above over successive intervals of lengths $T_n = \alpha^n$ for some fixed $\alpha > 1$, provides an efficient coupling strategy for the Kolmogorov diffusion.*

Remark 2. An efficient coupling *strategy* for the Kolmogorov diffusion has to couple at a much faster rate when the initial discrepancy vector $\boldsymbol{z}$ has an initial segment $z_0, z_1, \dots, z_{r-1}$ which is zero: see Theorem 1.

Proof of Theorem 6. Consider the above coupling between two index-k generalized Kolmogorov diffusions $\boldsymbol{I}^{(1)}$ and $\boldsymbol{I}^{(2)}$, begun at $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, respectively, with $\boldsymbol{z} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}$. Setting $\boldsymbol{Z}_0 = \boldsymbol{z} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}$ and $S_0 = T_0 = 0$, and employing the representation (17), we may write, for $n \geq 1$,

$$
\begin{aligned}
\boldsymbol{Z}_n(\zeta) &= \boldsymbol{I}^{(1)}(S_{n-1} + \zeta) - \boldsymbol{I}^{(2)}(S_{n-1} + \zeta) \\
&= \boldsymbol{D}(T_n)(\boldsymbol{H}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_{n-1}(T_{n-1}) + \boldsymbol{L}\Delta_n\boldsymbol{W}(\zeta)),
\end{aligned}
\tag{23}
$$

where $\Delta_n\boldsymbol{W}(\zeta) = \boldsymbol{W}_{n-1}^{(1)}(\zeta) - \boldsymbol{W}_{n-1}^{(2)}(\zeta)$ and $\zeta \in [0, T_n]$. We will write $\boldsymbol{Z}_n$ for $\boldsymbol{Z}_n(T_n)$ and $\Delta_n\boldsymbol{W}$ for $\Delta_n\boldsymbol{W}(T_n)$.

Set $F_0(0) = \|\boldsymbol{L}^{-1}\boldsymbol{z}\|$ and $\boldsymbol{v}_0 = \boldsymbol{L}^{-1}\boldsymbol{z}/\|\boldsymbol{L}^{-1}\boldsymbol{z}\|$. For $n \geq 1$, write

$$
F_n(\zeta) = \|\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n(\zeta)\|,
$$

and recall the unit vectors

$$
\boldsymbol{v}_n = \frac{\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n}{\|\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n\|} \quad \text{and} \quad \eta_n = \frac{(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_n}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_n\|}
$$

(defined when $\boldsymbol{I}^{(1)}(S_n) \neq \boldsymbol{I}^{(2)}(S_n)$). As before, we will write F_n for $F_n(T_n)$.

Then (23) gives us the following:

$$
\begin{aligned}
F_n(\zeta) &= \|\boldsymbol{L}^{-1}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_n(\zeta)\| \\
&= \|\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(T_n^{-1})\boldsymbol{Z}_{n-1} + \Delta_n\boldsymbol{W}(\zeta)\| \\
&= \|F_{n-1}(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1} + \Delta_n\boldsymbol{W}(\zeta)\|.
\end{aligned}
\tag{24}
$$

Suppose that $\boldsymbol{I}^{(1)}(S_{n-1}) \neq \boldsymbol{I}^{(2)}(S_{n-1})$. Then from (22) we have

$$
\Delta_n\boldsymbol{W}(\zeta) = 2B_{n-1,1}(\zeta \wedge \sigma_{n-1})\eta_{n-1}.
$$

Substituting this into (24), and noting the definition of the stopping time σ_{n-1}, we get

$$
F_n(\zeta) = \left(F_{n-1} + 2\frac{B_{n-1,1}(\zeta \wedge \sigma_{n-1})}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}\|}\right)\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}\|.
\tag{25}
$$

Suppose that the coupling has not been successful in $[0, S_n]$. Then putting $\zeta = T_n$ in (24) and using (22) again, we proceed similarly as above to obtain the following recursive relation:

$$
F_n\boldsymbol{v}_n = \left(F_{n-1} + 2\frac{B_{n-1,1}(T_n \wedge \sigma_{n-1})}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}\|}\right)(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}.
\tag{26}
$$

Thus, for this sequential coupling scheme, it follows from (26) that the vector $\boldsymbol{v}_n$ is deterministic, and indeed

$$
\boldsymbol{v}_n = \frac{(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}\|} = \cdots = \frac{(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n\boldsymbol{L})\boldsymbol{v}_0}{\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n\boldsymbol{L})\boldsymbol{v}_0\|}.
\tag{27}
$$

So we can re-express (25) in terms of $\boldsymbol{v}_0$:

$$
F_n(\zeta) = \left(F_{n-1} + 2\frac{B_{n-1,1}(\zeta \wedge \sigma_{n-1})}{\|(\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L})\boldsymbol{v}_{n-1}\|}\right)\frac{\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n\boldsymbol{L})\boldsymbol{v}_0\|}{\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^{n-1}\boldsymbol{L})\boldsymbol{v}_0\|}.
\tag{28}
$$

Consequently, if we define

$$
G_n(\zeta) = \frac{F_n(\zeta)}{\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n\boldsymbol{L})\boldsymbol{v}_0\|} \quad \text{for } \zeta \in [0, T_n],
$$

then (28) gives us

$$G_n(\zeta) = G_{n-1} + 2 \frac{B_{n-1,1}(\zeta \wedge \sigma_{n-1})}{\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n \boldsymbol{L})\boldsymbol{v}_0\|} \tag{29}$$

for $\zeta \in [0, T_n]$, where, as before, we write G_{n-1} for $G_{n-1}(T_{n-1})$.

Consider now the (deterministic) evolution of the length $\|(\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n \boldsymbol{L})\boldsymbol{v}_0\|$. The matrices $\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L}$ and $\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})$ share the same eigenvalues. Since $\boldsymbol{H}$ is lower-triangular, and has values 1 along the diagonal, and since $\alpha^{-1} \in (0, 1)$, it follows that the eigenvalues of $\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L}$ are distinct and positive, and are $1, \alpha^{-1}, \ldots, \alpha^{-k}$. Consequently, it follows that $\boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{D}(\alpha^{-1})\boldsymbol{L}$ has $k + 1$ distinct eigenvectors; let $\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ be the eigenvectors corresponding to the eigenvalues $1, \alpha^{-1}, \ldots, \alpha^{-k}$; we suppose these to be of unit Euclidean norm. Note in particular that $\boldsymbol{e}_0 = (1, 0, \ldots, 0)^\top$, while $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k$ all have zero initial coordinate (this follows from the lower-triangular nature of the matrices $\boldsymbol{L}, \boldsymbol{H}, \boldsymbol{D}(\alpha^{-1})$; moreover, the explicit construction of $\boldsymbol{L}$ implies that its top row is given by $(1, 0, \ldots, 0)$). Write $\boldsymbol{v}_0 = \gamma_0 \boldsymbol{e}_0 + \gamma_1 \boldsymbol{e}_1 + \cdots + \gamma_k \boldsymbol{e}_k$.

We suppose first that $\gamma_0 \neq 0$. From the above, it follows that

$$\|\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n \boldsymbol{L}\boldsymbol{v}_0\| = \|\gamma_0 \boldsymbol{e}_0 + \gamma_1 \alpha^{-n} \boldsymbol{e}_1 + \cdots + \gamma_k \alpha^{-kn} \boldsymbol{e}_k\|,$$

which by the triangle inequality lies in the range

$$|\gamma_0| - \max\{|\gamma_i| : i = 1, \ldots, k\} \alpha^{-n}$$
$$\leq \|\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n \boldsymbol{L}\boldsymbol{v}_0\| \leq |\gamma_0| + \max\{|\gamma_i| : i = 1, \ldots, k\} \alpha^{-n}. \tag{30}$$

Choose n_0 such that the lower bound of this range exceeds 0 for $n \geq n_0$.

Observe from the 'algorithmic time' interpolation (29) that if we set $G(0) = F_0(0) = \|\boldsymbol{L}^{-1}\boldsymbol{z}\|$ and $G(\zeta) = G_n(\zeta - S_{n-1})$ for $\zeta \in [S_{n-1}, S_n]$, then G is driven by a time-changed Brownian motion built up from the increments

$$2 \frac{B_{n-1,1}((\zeta - S_{n-1}) \wedge \sigma_{n-1})}{\|\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^n \boldsymbol{L}\boldsymbol{v}_0\|} \quad \text{for } S_{n-1} < \zeta \leq S_n.$$

Moreover, $G(\zeta) = 0$ for some $\zeta \in [S_{n-1}, S_n]$ (this happens when $\zeta - S_{n-1} = \sigma_{n-1}$) if and only if $F_n = 0$, and thus the Kolmogorov diffusions $\boldsymbol{I}^{(1)}$ and $\boldsymbol{I}^{(2)}$ (running in 'process time') couple by time S_n. The ratio S_n/S_{n-1} always equals α, so a simple asymptotic analysis shows that, for coupling-efficiency considerations, it is sufficient to analyse the hitting time of zero for the process G.

We see from (30) that by time $S_n = \alpha + \cdots + \alpha^n$ the intrinsic time of the time-changed Brownian motion will lie between the two values

$$\sum_{m=1}^{n_0-1} \frac{4\alpha^m}{\|\boldsymbol{L}^{-1}(\boldsymbol{H}\boldsymbol{D}(\alpha^{-1}))^m \boldsymbol{L}\boldsymbol{v}_0\|^2} + \sum_{m=n_0}^{n} \frac{4\alpha^m}{(|\gamma_0| \mp \max\{|\gamma_i| : i = 1, \ldots, k\} \alpha^{-m})^2},$$

and so the probability of G not yet having hit zero will be of order

$$\frac{C_1}{\sqrt{S_n}} \leq \left(C_3 + \sum_{m=n_0}^{n} \frac{4\alpha^m}{(|\gamma_0| \mp \max\{|\gamma_i| : i = 1, \ldots, k\} \alpha^{-m})^2} \right)^{-1/2} \leq \frac{C_2}{\sqrt{S_n}}$$

for positive constants C_1, C_2, and C_3.

Asymptotic analysis for large n now shows that efficiency follows in the case $\gamma_0 \neq 0$.

Suppose that $\gamma_0 = 0$, so the first coordinate of z must vanish. In case the first $r > 0$ coordinates of z vanish, the deterministic evolution of v_n given by (27) together with the lower-triangular nature of $L^{-1} H D(\alpha^{-1}) L$ ensures that the first r coordinates of v_n also vanish, for all n. Hence,

$$\|(L^{-1} H D(\alpha^{-1}) L) v_n\| \leq \alpha^{-r}. \tag{31}$$

We now argue much as above, but working with the rather simpler comparison process

$$\widetilde{G}(t) = 2\alpha^{nr} B_{n-1,1}((\zeta - S_{n-1}) \wedge \sigma_{n-1}) + \alpha^r \|(L^{-1} H D(\alpha^{-1}) L) v_{n-1}\| \widetilde{G}(S_{n-1})$$

for $S_{n-1} < \zeta \leq S_n$, where $\widetilde{G}(0) = \|L^{-1} z\|$.

From (31), $\alpha^r \|(L^{-1} H D(\alpha^{-1}) L) v_{n-1}\| \leq 1$. Thus, the jumps of the process $\widetilde{G}$ at the times $\{S_i\}_{i \geq 0}$ decrease the value of $\widetilde{G}$. Hence, the hitting time of zero for $\frac{1}{2} \widetilde{G}(S_n)$ is dominated by that of a new Brownian motion run until time

$$\alpha^{2r} T_1 + \cdots + \alpha^{2nr} T_n = \alpha^{2r+1} + \cdots + \alpha^{n(2r+1)} \sim C_1^{-2} \alpha^{(n+1)(2r+1)}.$$

This avoids 0 up to this time with probability of order $C_1 / \alpha^{(n+1)(2r+1)/2}$. Since $S_n \sim C_2 \alpha^{n+1}$, we deduce that the asymptotics for the probability of avoiding zero before S_n are of order at most $1 / S_n^{r+1/2}$.

As a consequence, it follows that this finite-look-ahead coupling provides an efficient coupling strategy, taking efficient advantage of faster coupling when an initial set of coordinates of z vanish. This completes the proof. □

Suppose that we require the range of precognition to be bounded throughout the coupling procedure: is it still possible to produce efficient couplings, so long as the Brownian components of the two coupled Kolmogorov diffusions start at different points? We do not yet have a general answer to this question, but can show that the obvious approach cannot work. Consider the above finite-look-ahead coupling of the Kolmogorov diffusion of index k, defined over successive intervals of length $T_n = 1$ (so in particular the look-ahead is *bounded*). Then this coupling may not even succeed!

We justify this assertion by proceeding as in the proof of Theorem 6, but taking $T_n = 1$, so $S_n = n$. Then (25) and (27) simplify: we obtain

$$F_n(\zeta) = \|(L^{-1} H L) v_{n-1}\| F_{n-1} + 2 B_{n-1,1}(\zeta \wedge \sigma_{n-1}),$$

$$v_n = \frac{(L^{-1} H L) v_{n-1}}{\|(L^{-1} H L) v_{n-1}\|} = \frac{L^{-1} H(n) z}{\|L^{-1} H(n) z\|}.$$

Accordingly, we obtain

$$F_n(\zeta) = \frac{\|L^{-1} H(n) z\|}{\|L^{-1} H(n-1) z\|} F_{n-1} + 2 B_{n-1,1}(\zeta \wedge \sigma_{n-1}).$$

But, asymptotically (considering the action of $H(n)$ on z for large n),

$$\frac{\|L^{-1} H(n) z\|}{\|L^{-1} H(n-1) z\|} \sim \left(\frac{n}{n-1}\right)^k$$

(note that the index k satisfies $k \geq 1$). Therefore, the time-changed Brownian interpolation of the F_n / n^k (following the proof of Theorem 6) may be compared with a scalar Brownian motion run up to time $\sum_n 1/n^{2k}$. This sum converges, and therefore there is a positive probability of the Brownian motion not reaching zero.

5. Conclusion

In this paper we have studied rates at which the (generalized) Kolmogorov diffusion can be coupled using Markovian or near-Markovian couplings. While this diffusion does possess successful Markovian couplings [5], [17], such couplings cannot be maximal (Theorem 2), nor can there be an efficient Markovian coupling strategy (Theorem 3). Moreover, at least in the classical case, there can be no optimal Markovian coupling (Theorem 5). By way of compensation, it *is* possible to exhibit a simple efficient finite-look-ahead coupling strategy even for the generalized Kolmogorov diffusion (Theorem 6). Thus, a controlled amount of anticipation suffices to obtain efficiency of coupling.

The results of [4] show that smooth elliptic diffusions admit Markovian maximal couplings only in cases of very special geometry. The Kolmogorov diffusion is the simplest nontrivial nilpotent diffusion, and so its failure to admit Markovian maximal couplings suggests the conjecture that no smooth nontrivial nilpotent diffusions can admit Markovian maximal couplings. Certainly this seems plausible for the case of planar Brownian motion plus Itô stochastic area [5], [14], [15]: perhaps the conjecture can be resolved in the manner of [4], using ideas from the geometry of nilpotent groups.

The above results can be compared with those of [16], concerning the coupling of scalar Brownian motion together with its local time at zero. That case lies outside the range of nilpotent diffusions; however, it does seem to provide relevant insights. For the local-time coupling, a (partly numerical) argument shows that there is no Markovian maximal coupling, but a control-theoretic argument shows that a simple reflection/synchronous coupling is optimal Markovian. It would be most interesting to determine whether the existence of efficient Markovian couplings for a given smooth diffusion, or of optimal Markovian couplings, can enforce geometric rigidity in a manner similar to the existence of maximal Markovian couplings. (The results in [6, Sections 3, 4] can be viewed as providing a very preliminary exploration of this kind of problem in the special case of reflecting Brownian motion in a compact convex domain.) This would further elucidate the enigmatic rôle of geometry in probabilistic coupling theory.

A further detailed question for future research is whether it is possible to attain efficiency for a *bounded-horizon* finite-look-ahead coupling for a Kolmogorov diffusion. The coupling described in Theorem 6 has a horizon which extends at a geometric rate over successive blocks: as noted after the proof of Theorem 6, the obvious bounded-horizon coupling does not even have the property of being successful. There are instances in which a small amount of look-ahead can have a dramatic influence on coupling rate—see the work of Smith [23] on coupling for a Gibbs' sampler on the simplex—and it would be very interesting if one could map out the circumstances in which this might apply in the relatively well-behaved case of smooth diffusions. As exemplified in [23], probabilistic coupling has a large part to play in the analysis of random algorithms; it would be of considerable advantage to gain some case history for the potential of modestly non-Markovian coupling to deliver faster couplings in the amenable instance of smooth diffusions.

Acknowledgements

We are grateful to an anonymous referee whose suggestions greatly improved the paper.

References

[1] ALILI, L. AND WU, C. T. (2014). Müntz linear transforms of Brownian motion. *Electron. J. Prob.* **19,** 15pp.
[2] AUE, A., HORVÁTH, L. AND HUŠKOVÁ, M. (2009). Extreme value theory for stochastic integrals of Legendre polynomials. *J. Multivariate Anal.* **100,** 1029–1043.

[3] BAILLEUL, I. (2008). Poisson boundary of a relativistic diffusion. *Prob. Theory Relat. Fields* **141,** 283–329.

[4] BANERJEE, S. AND KENDALL, W. S. (2014). Rigidity for Markovian maximal couplings of elliptic diffusions. Preprint. Available at http://arxiv.org/abs/1412.2647v1.

[5] BEN AROUS, G., CRANSTON, M. AND KENDALL, W. S. (1995). Coupling constructions for hypoelliptic diffusions: two examples. In *Stochastic Analysis* (Ithaca, NY, 1993; Proc. Sympos. Pure Math. **57**), eds M. C. Cranston and M. A. Pinsky, American Mathematical Society, Providence, RI, pp. 193–212.

[6] BURDZY, K. AND KENDALL, W. S. (2000). Efficient Markovian couplings: examples and counterexamples. *Ann. Appl. Prob.* **10,** 362–409.

[7] DA PRATO, G. AND ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions* (Encyclopedia Math. Appl. **152**), 2nd edn. Cambridge University Press.

[8] DOEBLIN, W. (1938). Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'états. *Revue Math. de l'Union Interbalkanique* **2,** 77–105.

[9] GOLDSTEIN, S. (1979). Maximal coupling. *Prob. Theory Relat. Fields* **46,** 193–204.

[10] GRIFFEATH, D. (1975). A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitsth.* **31,** 95–106.

[11] GROENEBOOM, P., JONGBLOED, G. AND WELLNER, J. A. (1999). Integrated Brownian motion, conditioned to be positive. *Ann. Prob.* **27,** 1283–1303.

[12] JANSONS, K. M. AND METCALFE, P. D. (2007). Optimally coupling the Kolmogorov diffusion, and related optimal control problems. *LMS J. Comput. Math.* **10,** 1–20.

[13] KEARNEY, M. J. AND MAJUMDAR, S. N. (2005). On the area under a continuous time Brownian motion till its first-passage time. *J. Phys. A* **38,** 4097–4104.

[14] KENDALL, W. S. (2007). Coupling all the Lévy stochastic areas of multidimensional Brownian motion. *Ann. Prob.* **35,** 935–953.

[15] KENDALL, W. S. (2010). Coupling time distribution asymptotics for some couplings of the Lévy stochastic area. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman* (London Math. Soc. Lecture Note Ser. **378**), eds N. H. Bingham and C. M. Goldie, Cambridge University Press, pp. 446–463.

[16] KENDALL, W. S. (2015). Coupling, local times, immersions. *Bernoulli* **21,** 1014–1046.

[17] KENDALL, W. S. AND PRICE, C. J. (2004). Coupling iterated Kolmogorov diffusions. *Electron. J. Prob.* **9,** 382–410.

[18] KUNITA, H. (1997). *Stochastic Flows and Stochastic Differential Equations* (Camb. Stud. Adv. Math. **24**). Cambridge University Press.

[19] KUWADA, K. (2009). Characterization of maximal Markovian couplings for diffusion processes. *Electron. J. Prob.* **14,** 633–662.

[20] MCKEAN, H. P., JR. (1963). A winding problem for a resonator driven by a white noise. *J. Math. Kyoto Univ.* **2,** 227–235.

[21] PITMAN, J. W. (1976). On coupling of Markov chains. *Z. Wahrscheinlichkeitsth.* **35,** 315–322.

[22] ROSENTHAL, J. S. (1997). Faithful couplings of Markov chains: now equals forever. *Adv. Appl. Math.* **18,** 372–381.

[23] SMITH, A. M. (2013). A Gibbs sampler on the n-simplex. *Ann. Appl. Prob.* **24,** 114–130.

[24] SVERCHKOV, M. YU. AND SMIRNOV, S. N. (1990). Maximal coupling for processes in D[0,∞]. *Dokl. Akad. Nauk SSSR* **311,** 1059–1061 (in Russian). English translation: *Soviet Math. Dokl.* **41** (1990), 352–354.

[25] THORISSON, H. (1994). Shift-coupling in continuous time. *Prob. Theory Relat. Fields* **99,** 477–483.

[26] WELLNER, J. A. AND SMYTHE, R. T. (2002). Computing the covariance of two Brownian area integrals. *Statist. Neerlandica* **56,** 101–109.

[27] YAN, L. (2004). Two inequalities for iterated stochastic integrals. *Arch. Math. (Basel)* **82,** 377–384.

SAYAN BANERJEE, *University of Warwick*

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.
Email address: sayan.banerjee@warwick.ac.uk

WILFRID S. KENDALL, *University of Warwick*

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK.
Email address: w.s.kendall@warwick.ac.uk