Electronic Journal of Probability

Electron. J. Probab. **23** (2018), no. 24, 1–43.
ISSN: 1083-6489 https://doi.org/10.1214/18-EJP150

Coupling polynomial Stratonovich integrals: the two-dimensional Brownian case*

Sayan Banerjee[†] Wilfrid Kendall[‡]

Abstract

We show how to build an immersion coupling of a two-dimensional Brownian motion (W_1, W_2) along with $\binom{n}{2} + n = \frac{1}{2}n(n+1)$. integrals of the form $\int W_1^i W_2^j \circ \mathrm{d}W_2$, where $j = 1, \ldots, n$ and $i = 0, \ldots, n-j$ for some fixed n. The resulting construction is applied to the study of couplings of certain hypoelliptic diffusions (driven by two-dimensional Brownian motion using polynomial vector fields). This work follows up previous studies concerning coupling of Brownian stochastic areas and time integrals (Ben Arous, Cranston and Kendall (1995), Kendall and Price (2004), Kendall (2007), Kendall (2009), Kendall (2013), Banerjee and Kendall (2015), Banerjee, Gordina and Mariano (2016)) and is part of an ongoing research programme aimed at gaining a better understanding of when it is possible to couple not only diffusions but also multiple selected integral functionals of the diffusions.

Keywords: Brownian motion; coupling; elliptic diffusion; faithful coupling; Heisenberg group; hypoelliptic diffusion; immersion coupling; Kolmogorov diffusion; Lévy stochastic area; Markovian coupling; monomial; nilpotent diffusion; parabolic Hörmander condition; reflection coupling; stochastic differential equation; Stratonovich integral; total variation distance.
AMS MSC 2010: 60J60; 60J65.
Submitted to EJP on May 3, 2017, final version accepted on February 11, 2018.

1 Introduction

A *coupling* of two probability measures μ_1 and μ_2, defined on respective measure spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, is a joint law μ defined on the product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ whose marginals are μ_1 and μ_2. A coupling of Markov processes X and $\widetilde{X}$ is an *immersion coupling* when the joint process

$$\{(X(t+s), \widetilde{X}(t+s)) : s \geq 0\} \text{ conditioned on } \mathcal{F}_t$$

is again a coupling of the laws of X and $\widetilde{X}$, but now starting from $(X(t), \widetilde{X}(t))$. This is also called a co-adapted coupling [19], or faithful coupling [29], and is very closely

*This work was supported by EPSRC Research Grant EP/K013939. This is a theoretical research paper and, as such, no new data were created during this study.

[†]University of North Carolina, Chapel Hill, USA. E-mail: sayan@email.unc.edu
[‡]University of Warwick, UK. E-mail: w.s.kendall@warwick.ac.uk

related to the near-equivalent notion of a Markovian coupling [7], which additionally constrains the joint process $(X(t), \widetilde{X}(t))$ to be Markovian with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Immersion couplings are typically very much easier to describe than general couplings, since they may be specified in causal ways using, for example, stochastic calculus.

In the following we consider couplings of smooth elliptic diffusions with d-dimensional state space $\mathbb{R}^d$. Specifically, for $d \geq 2$ and $1 \leq k \leq d$, consider the following Stratonovich stochastic differential equation on $\mathbb{R}^d$:

$$X(t) \quad = \quad x + \int_0^t V_0(X(s))\,\mathrm{d}s + \sum_{i=1}^k \int_0^t V_i(X(s)) \circ \mathrm{d}W_i(s)\,. \tag{1.1}$$

Here $x \in \mathbb{R}^d$ is the initial state, $V_1, \ldots, V_k$ are smooth vector fields, and $(W_1, \ldots, W_k)$ is a standard Brownian motion on $\mathbb{R}^k$. We will consider couplings of two copies X and $\widetilde{X}$ of this diffusion, starting from arbitrary distinct initial states $x, \widetilde{x} \in \mathbb{R}^d$. Interest focusses on the *coupling time* T, defined as

$$T \quad = \quad \inf\{t \geq 0 : X(s) = \widetilde{X}(s) \text{ for all } s \geq t\}\,.$$

The coupling is said to be *successful* if almost surely $T < \infty$ (where "almost surely" refers to the coupling measure μ). A major motivation to study couplings arises from the so-called Aldous' coupling inequality (see 1):

$$\mu(\tau > t) \quad \geq \quad \|\mu_t - \widetilde{\mu}_t\|_{TV}\,, \qquad \text{for all } t \geq 0\,, \tag{1.2}$$

where μ_t and $\widetilde{\mu}_t$ denote the laws of $X(t)$ and $\widetilde{X}(t)$ respectively and $||\cdot||_{TV}$ denotes the total variation distance between probability measures given by

$$\|\mu_1 - \mu_2\|_{TV} \quad = \quad \sup\{|\mu_1(A) - \mu_2(A)| : \text{ measurable } A\}\,.$$

Using inequality (1.2), construction of a coupling of X and $\widetilde{X}$ automatically bounds the total variation distance between the laws of the diffusions at time t. A *maximal coupling* is one for which the inequality (1.2) is actually an equality for all t. These have been shown to exist under very general conditions [14, 28, 13, 30, 11]. However in most cases the task of explicitly constructing such a maximal coupling is extremely hard, if not impossible. This provides strong motivation for considering immersion couplings which, although not maximal in most cases [4, 23], are easier to describe and can provide helpful bounds via (1.2).

Immersion couplings have been extensively studied for *elliptic* diffusions (which is to say diffusions given by (1.1) when $k = d$ and $\{V_1(x), \ldots, V_d(x)\}$ form a basis for $\mathbb{R}^d$ at each $x \in \mathbb{R}^d$). The simplest example of such a coupling is the *reflection coupling* of Euclidean Brownian motions starting from two different points: the second Brownian path is obtained from the first by reflecting the first path on the hyperplane bisecting the line joining the starting points until the first path (equivalently, the second, reflected, path) hits this hyperplane. This coupling turns out to be maximal as well as Markovian! [24] extended this reflection construction to produce successful Markovian couplings for *elliptic diffusions* on $\mathbb{R}^d$ with bounded and Lipschitz drift and diffusion coefficients when the diffusion matrix does not vary too much in space (see also 8). However in general the construction is not symmetric between the two coupled processes, and this method is not easily applicable when the diffusion matrix varies appreciably over space. A more geometric approach, depending symmetrically on the two coupled diffusions, is provided by the *Kendall-Cranston coupling* [18, 9]. Consider the positive-definite diffusion matrix $\sigma(x)$ formed with columns $V_1(x), \ldots, V_d(x)$. As x varies over $\mathbb{R}^d$, so this furnishes $\mathbb{R}^d$

with a Riemannian metric g given by $g(x) = (\sigma(x)\sigma(x)^\top)^{-1}$. With this intrinsic *diffusion metric*, $\mathbb{R}^d$ becomes a Riemannian manifold and the diffusion can be recognized as a Brownian motion with drift on this manifold. One then uses an appropriate generalization of reflection involving parallel transport along geodesics to obtain reflection couplings. These couplings are successful when (for example) the Ricci curvature of the manifold is non-negative and the drift vector field satisfies appropriate regularity conditions. We note here that the Kendall-Cranston coupling also works for elliptic diffusions whose state space is any smooth manifold, and can be applied to even more general situations [32].

The above techniques fail for diffusions which are not elliptic, as there is no natural Riemannian metric intrinsic to such diffusions. However, an important class of non-elliptic diffusions has attracted attention in recent times: namely, the *hypoelliptic diffusions*. These are diffusions $(X(t) : t \geq 0)$ such that $X(t)$ has a *smooth density with respect to Lebesgue measure* for each $t > 0$. They arise naturally in a variety of contexts: for example, modelling the motion of a particle following Newton's equations under a potential, white noise random forcing and linear friction (the *kinetic Fokker-Plank diffusion*, 31), describing stochastic oscillators (the *Kolmogorov diffusion*, 26), quantum mechanics and rough paths theory (*Brownian motion on the Heisenberg group*, 27, 12) and modelling of macromolecular systems [15]. All these examples place a premium on gaining a good understanding of the behaviour of hypoelliptic diffusions. In particular, the construction of successful couplings for these diffusions immediately implies, *via* Aldous' inequality (1.2), that the total variation distance between the laws of two such diffusions started from distinct points converges to zero as time goes to infinity. Furthermore, estimates on the coupling time distribution deliver bounds on the convergence rate. This, in turn, yields estimates of rate of convergence to stationarity, when a stationary measure exists. Moreover, these couplings can also be used to furnish gradient estimates for harmonic functions corresponding to the generators of the diffusions *via* purely probabilistic means [10, 9, 2].

At the time of writing, coupling of hypoelliptic diffusions have only been studied for rather specific examples. Hypoelliptic diffusions can be viewed as "high dimensional processes driven by low dimensional Brownian motions", which suggests that the goal of producing successful Markovian couplings of such diffusions may be best achieved by learning how to produce Markovian couplings of the driving Brownian motion together with a (typically finite) collection of path functionals. These couplings, sometimes described as *exotic couplings*, were first studied in [6]. This described successful Markovian couplings for the *Kolmogorov diffusion of order one* (given by a Brownian motion B along with its running time integral $\int_0^t B(s)\,\mathrm{d}s$) and *Brownian motion on the Heisenberg group* (a two-dimensional Brownian motion (B_1, B_2) together with its Lévy stochastic area $\int_0^t B_1(s)\,\mathrm{d}B_2(s) - \int_0^t B_2(s)\,\mathrm{d}B_1(s)$). [22] showed how to generate successful Markovian couplings for the Kolmogorov diffusion of any finite order n (a Brownian motion B along with its $n-1$ iterated time integrals $\int \cdots \int_{0\leq s_1 \leq \cdots \leq s_i \leq t} B(s_1)\,\mathrm{d}s_1\,\mathrm{d}s_2 \ldots \mathrm{d}s_i$ for $1 \leq i \leq n-1$). Later [19, 20] described a construction of a successful Markovian coupling of Brownian motion on the step-two free nilpotent Lie group of any underlying finite dimension n (corresponding to an n-dimensional Brownian motion $(B_1, \ldots, B_n)$ together with the $\binom{n}{2}$ stochastic areas $\int_0^t B_i(s)\,\mathrm{d}B_j(s) - \int_0^t B_j(s)\,\mathrm{d}B_i(s)$ for $1 \leq i < j \leq n$, or, using vector notation, $\underline{B}$ together with the alternating vector-product $\int \underline{B} \wedge \mathrm{d}\,\underline{B}$). [21] described how to couple scalar Brownian motion together with local time, and used this to couple a rather degenerate diffusion arising in stochastic control theory. Even in these rather simple examples, the coupling constructions turn out to be quite complicated. Simpler cases use careful combinations of reflection coupling and *synchronous coupling* (making Brownian increments agree): [19, 20] show that one also needs to use more varieties of coupling (for example what might be called "rotation couplings") when coupling all stochastic areas for Brownian motion in dimension of 3 or greater.

In this article, we provide constructions of immersion (in fact, Markovian) couplings for a considerable range of diffusions of the form given in (1.1) with $k = 2$, based on polynomial vector fields V_i. This is a significant step beyond [19, 20] in the development of the programme of understanding coupling for hypoelliptic diffusions, albeit limited here to the case of an underlying two-dimensional Brownian motion. Before going into the detailed description of the problem, we define the *parabolic Hörmander condition* which will be a crucial assumption in the coupling construction.

Consider the following sets of vector fields:

$$\mathcal{V}_0 \quad = \quad \{V_i : i \geq 1\}, \qquad\qquad \mathcal{V}_{j+1} \quad = \quad \{[U, V_i] : U \in \mathcal{V}_j,\ i \geq 0\} \text{ for } j \geq 0,$$

where $[U, V]$ denotes the Lie bracket of the vector fields U and V. Set $\mathbf{V}_j(x) = \operatorname{span}\{V(x),\ V \in \mathcal{V}_j\}$. We will make the following assumption:

(PHC) The vector fields $V_0, V_1, \ldots V_k$ satisfy the *parabolic Hörmander condition*, i.e., $\bigcup_{j\geq 0} \mathbf{V}_j(x) = \mathbb{R}^d$ for each $x \in \mathbb{R}^d$.

Subject to suitable regularity conditions, (PHC) is a *necessary* assumption if we want to construct successful couplings from arbitrary pairs of starting points. To see this, consider the distribution of sub-spaces $\{\mathcal{D}(x) = \bigcup_{k\geq 0} \mathbf{V}_j(x),\ x \in \mathbb{R}^d\}$ generated by $\bigcup_{j\geq 0} \mathcal{V}_j$ (that is, the smoothly varying subspace of the tangent space spanned by these vector fields). [25] showed that if $\mathcal{D}$ is of "locally finite type" (in particular, if the vector fields $V_0, V_1, \ldots, V_k$ are real analytic), then it has the *maximal integral manifold property*, i.e., for each point $x \in \mathbb{R}^d$, there exists an immersed submanifold S (called an integral manifold) containing x with the property that its tangent bundle coincides with the distribution $\mathcal{D}$. Moreover, S can be chosen so that any other integral manifold which intersects S must be an open submanifold of S (note that S need not be complete in $\mathbb{R}^d$). In this case, $\mathbb{R}^d$ splits into disjoint maximal integral manifolds. It follows from support theorems [16, for example] that if a diffusion starts from a point inside one maximal integral manifold then almost surely it must stay in this manifold for all time. Thus, under regularity conditions such as real analyticity, if (PHC) does not hold, then there must be at least two disjoint maximal integral manifolds. Consequently, two copies of the diffusion started from points in different maximal integral manifolds will almost surely never meet.

In order to make progress towards answering the general question of whether it is possible to construct successful immersion couplings of a diffusion satisfying (PHC) from arbitrary pairs of distinct starting points, this article considers a simplification. It will be convenient to view $\mathbb{R}^d = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{d-2}$, with corresponding coordinates $w = (w_1, w_2, w_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{d-2}$ *etc* (in a mild abuse of notation, w_3 denotes a $(d-2)$-dimensional vector). We assume that our diffusions satisfy (1.1) when the *drift vector field* $V_0 = 0$, the *driving Brownian motion is two-dimensional* (i.e. $k = 2$) and the *driving vector fields* V_1 *and* V_2 *are polynomial functions of the driving Brownian motion*. Specifically, for $d \geq 3$ and for each $w = (w_1, w_2, w_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{d-2}$, suppose that X can be written as

$$X(t) \quad = \quad (w_1 + W_1(t), w_2 + W_2(t), X_3(t)) \tag{1.3}$$

where (W_1, W_2) is a two-dimensional standard Brownian motion and X_3 can be written in vector format as satisfying the Stratonovich differential equation

$$X_3(t) \quad = \quad w_3 + \sum_{i=1}^{2} \int_0^t \sigma_i(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_i(s). \tag{1.4}$$

Here σ_1, σ_2 are ($(d-2)$-dimensional) vector-valued *polynomials*:

$$\sigma_i(x_1,x_2) \quad = \quad a_i^{0,0} + \sum_{1\le l+m\le n} a_i^{l,m} x_1^l x_2^m, \qquad \text{for } i=1,2\,,$$

with $(d-2)$-dimensional vector-valued coefficients $a_i^{l,m} = (a_{i,3}^{l,m},\ldots,a_{i,d}^{l,m})^\top \in \mathbb{R}^{d-2}$. For convenience, write

$$\sigma_i(x_1,x_2) \quad = \quad (\sigma_{i,3}(x_1,x_2),\ldots,\sigma_{i,d}(x_1,x_2))^\top\,.$$

Lemma 1 below describes exactly when the system (1.4) satisfies (PHC).

Several important examples of hypoelliptic diffusions fall in this category, including *Brownian motion on the Heisenberg group* [27, 6]. The problem of immersion coupling for diffusions in the form of (1.4) makes a useful next step in the bigger program of coupling hypoelliptic diffusions because of the following reasons. Firstly, the zero drift condition helps to simplify (PHC) and give a clearer exposition, although we believe that the methods developed here can be used even when the drift is non-zero but satisfies certain growth conditions. Secondly, the polynomial form of the driving vector fields ensures that X_3 can be written using linear combinations of monomial Stratonovich integrals of the form $(\int W_1^i W_2^j \circ \mathrm{d}W_2 \,:\, i+j\le n)$ and thus, the problem reduces to successfully coupling the driving Brownian motions along with these integrals. Moreover, these polynomial vector fields can be used to approximate a large class of real analytic and nilpotent vector fields and we hope that our technique will extend to more general diffusions driven by such vector fields. Thirdly, as proved in Lemma 1 below, (PHC) for this class of diffusions simplifies to a non-singularity condition for a matrix formed by the vectors $a_i^{l,m}$. Finally, as described in [19] in the simpler context of coupling stochastic areas, successful Markovian coupling strategies can be achieved using only reflection/synchronous coupling of Brownian motions when the driving Brownian motion is two-dimensional (for example, Brownian motion on the Heisenberg group), but for higher dimensional analogues it is necessary to employ rotation couplings (which is to say, control strategies using orthogonal matrices), and this complicates the coupling strategy considerably. As we will see, the restriction to a two-dimensional driving Brownian motion in the case of (1.4) similarly allows for a rather explicit coupling construction using only synchronous coupling of W_2 at all times together with judicious switching between synchronous and reflection phases for W_1. However, we anticipate that one of the challenges of dealing with higher-dimensional Brownian motions will be to deal with complexity entailed by no longer being able to keep one coordinate synchronously coupled and in agreement for all time. We plan to address the complexities of the higher dimensional case in a subsequent article.

In the remainder of this section, we will show that, in order to successfully couple two copies X and $\widetilde{X}$ of our diffusion (1.3) started from distinct points, it suffices successfully to couple simultaneously the driving Brownian motions along with integrals of the form $(\int W_1^i W_2^j \circ \mathrm{d}W_2 : i+j\le n)$. Define the ($(d-2)$-dimensional) vector-valued function

$$\phi(x_1,x_2) \quad = \quad \sigma_2(x_1,x_2) - \int_{w_1}^{x_1} \partial_2\sigma_1(u,x_2)\,\mathrm{d}u \tag{1.5}$$

where $(x_1,x_2)\in\mathbb{R}^2$ and ∂_i denotes the partial derivative with respect to the i^{th} coordinate $(i=1,2)$.

Set $\Psi_1(x_1,x_2) = \int_{w_1}^{x_1}\sigma_1(u,x_2)\,\mathrm{d}u$. Computing the Stratonovich differential of $\Psi_1(w_1 + W_1(t), w_2+W_2(t))$, and then integrating this differential, amounts to establishing an integration-by-parts relation between certain Stratonovich integrals with respect to W_1

and other Stratonovich integrals with respect to W_2, holding up to addition of a function of W_1 and W_2 whose coupling follows directly from coupling of (W_1, W_2):

$$\int_0^t \sigma_1(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_1(s) \quad = \quad \Psi_1(w_1 + W_1(t), w_2 + W_2(t)) - \int_0^t \partial_2 \Psi_1(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_2(s)\,.$$

Hence X_3 can be expressed as the sum of a function of W_1 and W_2 and a Stratonovich integral with respect to W_2 alone:

$$\begin{aligned} X_3(t) \quad &= \quad w_3 + \int_0^t \sigma_1(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_1(s) \\ &\qquad + \int_0^t \sigma_2(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_2(s) \\ &= \quad w_3 + \Psi_1(w_1 + W_1(t), w_2 + W_2(t)) \\ &\qquad + \int_0^t \left[\sigma_2(w_1 + W_1(s), w_2 + W_2(s)) - \partial_2 \Psi_1(w_1 + W_1(s), w_2 + W_2(s))\right] \circ \mathrm{d}W_2(s) \\ &= \quad w_3 + \Psi_1(w_1 + W_1(t), w_2 + W_2(t)) + \int_0^t \phi(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_2(s)\,. \end{aligned} \tag{1.6}$$

Let $\Sigma(x_1, x_2)$ denote the $(d-2) \times \frac{n(n-1)}{2}$ matrix formed by arranging in a row the $\frac{n(n-1)}{2}$ different $(d-2)$-dimensional vector-valued functions $\partial_1^{l+1}\partial_2^m \phi(x_1, x_2)$ (here $1 \leq l+1+m \leq n$). The condition (PHC) is equivalent to a rank condition on the matrix-valued function Σ:

Lemma 1. The diffusion X satisfies (PHC) if and only if $\Sigma(w_1, w_2)$ has full rank $d-2$, where w is the starting point of the diffusion in question.

Proof. Let ∂_1, ∂_2, ..., ∂_d represent the standard basis vectors of $\mathbb{R}^d$. The diffusion X can be expressed in the form

$$X(t) \quad = \quad w + \int_0^t V_1(X(s)) \circ \mathrm{d}W_1(s) + \int_0^t V_2(X(s)) \circ \mathrm{d}W_2(s)$$

where $w = (w_1, w_2, w_3)$ and the vector fields V_1, V_2 are given by

$$V_i(x) \quad = \quad \partial_i + \sum_{j=3}^d \sigma_{i,j}(x_1, x_2)\partial_j$$

for $i = 1, 2$ and $x = (x_1, x_2, x_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{d-2}$ (so x_3 is a $(d-2)$-dimensional vector). For any word $I = (i_1, \ldots, i_N) \in \{1, 2\}^N$, write $V_I = [V_{i_1}[V_{i_2}[\ldots[V_{i_{N-1}}, V_{i_N}]]\ldots]$. Denote by $n_1(I)$, $n_2(I)$ the number of occurrences of $1, 2$ respectively in the word I. We claim that if I is of length 2 or more and $i_{N-1} = 1$, $i_N = 2$ then

$$V_I(x) \quad = \quad \sum_{j=3}^d \partial_1^{n_1(I)} \partial_2^{n_2(I)-1} \phi_j(x_1, x_2)\partial_j\,. \tag{1.7}$$

We prove this by induction on the length of the word I. For length 2, (1.7) follows from the definition of ϕ in (1.5):

$$[V_1, V_2](x) \quad = \quad \sum_{j=3}^d \left(\partial_1 \sigma_{2,j}(x_1, x_2) - \partial_2 \sigma_{1,j}(x_1, x_2)\right)\partial_j \quad = \quad \sum_{j=3}^d \partial_1 \phi_j(x_1, x_2)\partial_j\,.$$

For any $N > 2$ assume that the induction hypothesis holds for words of length less than N. Consider $I \in \{1,2\}^{N-1}$ with $i_{N-2} = 1$, $i_{N-1} = 2$, and examine the case of $I^* = (i_0, I)$. By the induction hypothesis,

$$V_I(x) \quad = \quad \sum_{j=3}^{d} \partial_1^{n_1(I)} \partial_2^{n_2(I)-1} \phi_j(x_1,x_2) \partial_j \,.$$

Observe that σ_1, σ_2 depend only on x_1, x_2, so that the form of V_I implies that $V_I V_{i_0} = 0$ and $V_{i_0} V_I = \partial_{i_0} V_I$. Therefore

$$\begin{aligned} V_{I^*}(x) \quad &= \quad \partial_{i_0} V_I(x) \quad = \quad \sum_{j=3}^{d} \partial_{i_0} \partial_1^{n_1(I)} \partial_2^{n_2(I)-1} \phi_j(x_1,x_2) \partial_j \\ &= \quad \sum_{j=3}^{d} \partial_1^{n_1(I^*)} \partial_2^{n_2(I^*)-1} \phi_j(x_1,x_2) \partial_j \end{aligned}$$

proving (1.7) in that case also. Note that the coefficients of ∂_1 and ∂_2 are zero in $V_{i_1,\ldots,i_{N-2},1,2}$.

Now observe that $V_1(x)$ and $V_2(x)$ are linearly independent for each $x \in \mathbb{R}^d$. Furthermore, $V_{i_1,\ldots,i_{N-2},2,1} = -V_{i_1,\ldots,i_{N-2},1,2}$ while $V_{i_1,\ldots,i_{N-2},1,1} = V_{i_1,\ldots,i_{N-2},2,2} = 0$, so the coefficients of ∂_1 and ∂_2 for V_I are zero whenever the word I has length greater than or equal to 2. Thus, for (PHC) to hold, the subspace spanned by $\{V_I : \text{length}(I) \geq 2\}$ must have dimension $d-2$. By (1.7) and the definition of $\Sigma(x_1,x_2)$, this is equivalent to requiring that $\Sigma(x_1,x_2)$ has rank $d-2$. But ϕ is a vector polynomial, so $\Sigma(x_1,x_2)$ has rank $d-2$ if and only if $\Sigma(w_1,w_2)$ has rank $d-2$. To see this, note that if $\Sigma(x_1,x_2)$ has rank $d-2$ for some $x \in \mathbb{R}^d$, then there exists a $(d-2)\times(d-2)$ sub-matrix $\Sigma^*(x_1,x_2)$ of $\Sigma(x_1,x_2)$ which is non-singular. From the continuity of the determinant, we conclude that $\Sigma^*(y_1,y_2)$, and hence $\Sigma(y_1,y_2)$, has rank $d-2$ for all y in an open neighborhood of x. Conversely, if $\Sigma(x_1,x_2)$ has rank less than $d-2$ for some $x \in \mathbb{R}^d$, then there exist constants $c_1,\ldots,c_d$, not all zero, such that

$$\sum_{k=3}^{d} c_k \partial_1^{l+1} \partial_2^m \phi_k(x_1,x_2) \quad = \quad 0\,, \qquad \text{for } 1 \leq l+1+m \leq n\,.$$

As $\phi(x_1,x_2) = (\phi_3(x_1,x_2),\ldots,\phi_d(x_1,x_2))$ is a vector polynomial in x_1, x_2 of degree n,

$$\partial_1 \phi(y_1,y_2) \quad = \quad \sum_{0\leq l+m\leq n-1} \frac{\partial_1^{l+1}\partial_2^m \phi(x_1,x_2)}{l!\; m!} (y_1-x_1)^l (y_2-x_2)^m\,, \qquad \text{for } y \in \mathbb{R}^d\,.$$

Hence,

$$\sum_{k=3}^{d} c_k \partial_1 \phi_k(y_1,y_2) \quad = \quad \sum_{0\leq l+m\leq n-1} \sum_{k=3}^{d} c_k \frac{\partial_1^{l+1}\partial_2^m \phi_k(x_1,x_2)}{l!\; m!} (y_1-x_1)^l (y_2-x_2)^m \quad = \quad 0\,.$$

So further differentiation yields

$$\sum_{k=3}^{d} c_k \partial_1^{l+1} \partial_2^m \phi_k(y_1,y_2) \quad = \quad 0\,, \qquad \text{for } 1 \leq l+1+m \leq n, \;\; y \in \mathbb{R}^d\,.$$

Thus, $\Sigma(y_1,y_2)$ has rank less than $d-2$ for all $y \in \mathbb{R}^d$. From the connectedness of $\mathbb{R}^d$, $\Sigma(x_1,x_2)$ has rank $d-2$ for some $x \in \mathbb{R}^d$ if and only if $\Sigma(w_1,w_2)$ has rank $d-2$, completing the proof of the lemma. $\square$

Remark 2. It follows from the reasoning in the proof of Lemma 1 that $V_I = 0$ whenever the length of the word I strictly exceeds n (the maximal degree of the polynomial coefficients σ_1, σ_2), regardless of whether (PHC) is satisfied or not. From this observation it follows that any diffusion of the form implied by (1.4) is *nilpotent* [5]. Nilpotent diffusions serve as the starting point for many analyses of hypoelliptic diffusions owing to their simplicity.

Consider the task of immersion coupling two copies X and $\widetilde{X}$ of the diffusion starting from w and $\widetilde{w}$ respectively, using driving Brownian motions (W_1, W_2) and $(\widetilde{W}_1, \widetilde{W}_2)$. Reflection coupling can be used to bring together the driving Brownian motions first. Thus there is no loss of generality in assuming that the *Brownian* starting points agree: $(w_1, w_2) = (\widetilde{w}_1, \widetilde{w}_2)$. Referring to the representation (1.6), it suffices to couple the two diffusions

$$\begin{aligned} X^*(t) &= \left(w_1 + W_1(t), w_2 + W_2(t), w_3 + \int_0^t \phi(w_1 + W_1(s), w_2 + W_2(s)) \circ \mathrm{d}W_2(s)\right), \\ \widetilde{X}^*(t) &= \left(w_1 + \widetilde{W}_1(t), w_2 + \widetilde{W}_2(t), \widetilde{w}_3 + \int_0^t \phi(w_1 + \widetilde{W}_1(s), w_2 + \widetilde{W}_2(s)) \circ d\widetilde{W}_2(s)\right), \end{aligned}$$

with starting points differing only in the third, vectorial, coordinates w_3, $\widetilde{w}_3$. (This is because coupling of the summands $\Psi_1(w_1 + W_1(t), w_2 + W_2(t))$ and $\Psi_1(w_1 + \widetilde{W}_1(t), w_2 + \widetilde{W}_2(t))$ in (1.6) is immediately implied by coupling of (W_1, W_2) and $(\widetilde{W}_1, \widetilde{W}_2)$.) Denote by $\mathcal{I}(t)$ the vector formed by $\left(\frac{\int_0^t W_1(s)^{l+1} W_2(s)^m \circ \mathrm{d}W_2(s)}{(l+1)!\ m!} : 1 \le l+1+m \le n\right)$ and similarly $\widetilde{\mathcal{I}}(t)$. Decomposing the polynomials given by the integrands $\phi(w_1+W_1(s), w_2+W_2(s))$ and $\phi(w_1+\widetilde{W}_1(s), w_2+\widetilde{W}_2(s))$ according to whether or not monomials involve W_1 (respectively $\widetilde{W}_1$), the last $d-2$ coordinates of X^*, $\widetilde{X}^*$ can be written in vector form as X_3^*, $\widetilde{X}_3^*$ where

$$\begin{aligned} X_3^*(t) &= w_3 + P(w_2, w_2 + W_2(t)) + \Sigma(w_1, w_2)\mathcal{I}(t), \\ \widetilde{X}_3^*(t) &= \widetilde{w}_3 + P(w_2, w_2 + \widetilde{W}_2(t)) + \Sigma(w_1, w_2)\widetilde{\mathcal{I}}(t), \end{aligned}$$

where P is a polynomial that arises from Stratonovich integration (with respect to W_2) of monomials in $w_2 + W_2$ alone. By Lemma 1, (PHC) implies that $\Sigma(w_1, w_2)$ has rank $d-2$, hence $w_3, \widetilde{w}_3$ both lie in the space spanned by the columns of $\Sigma(w_1, w_2)$. Thus, there are $z_3^*, \widetilde{z}_3^* \in \mathbb{R}^{n(n-1)/2}$ such that $w_3 = \Sigma(w_1, w_2)z_3^*$, $\widetilde{w}_3 = \Sigma(w_1, w_2)\widetilde{z}_3^*$. Hence

$$\begin{aligned} X_3^*(t) &= P(w_2, w_2 + W_2(t)) + \Sigma(w_1, w_2)(z_3^* + \mathcal{I}(t)), \\ \widetilde{X}_3^*(t) &= P(w_2, w_2 + \widetilde{W}_2(t)) + \Sigma(w_1, w_2)(\widetilde{z}_3^* + \widetilde{\mathcal{I}}(t)). \end{aligned}$$

It follows that if we can successfully couple

$$\begin{aligned} &(w_1 + W_1(t), w_2 + W_2(t), z_3^* + \mathcal{I}(t)), \\ &(\widetilde{w}_1 + \widetilde{W}_1(t), \widetilde{w}_2 + \widetilde{W}_2(t), \widetilde{z}_3^* + \widetilde{\mathcal{I}}(t)) \end{aligned} \tag{1.8}$$

from arbitrary pairs of starting points $(w_1, w_2, z_3^*), (\widetilde{w}_1, \widetilde{w}_2, \widetilde{z}_3^*) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{n(n-1)/2}$, then we can successfully couple X and $\widetilde{X}$ from arbitrary pairs of starting points. In the next section, we construct a coupling of the above processes. We show in Theorem 10 that this coupling is indeed successful, and that moreover the coupling time has a power law tail. The existence of such a successful coupling immediately implies the following theorem.

Theorem 3. Consider the diffusion $X(t) = (X_1(t), X_2(t), X_3(t)) \in \mathbb{R}^d$ (for $d \ge 3$, considering X_3 as a $(d-2)$-dimensional process) defined by the following stochastic differential equation:

$$\begin{aligned}
\mathrm{d}X_1(t) &= \mathrm{d}W_1(t)\,,\\
\mathrm{d}X_2(t) &= \mathrm{d}W_2(t)\,,\\
\mathrm{d}X_3(t) &= \sigma_1(X_1(t),X_2(t))\circ \mathrm{d}W_1(t)+\sigma_2(X_1(t),X_2(t))\circ \mathrm{d}W_2(t)\,,
\end{aligned}$$

where (W_1, W_2) is a two-dimensional standard Brownian motion and σ_1, σ_2 are polynomial vector fields such that (PHC) holds. Then there exists a successful Markovian coupling of two copies of the above diffusion starting from any pair of distinct points.

Theorem 3 summarizes the qualitative content of (and is a direct consequence of) Theorem 10 stated and proved in Section 4 below, but omits the tail estimate on the coupling time distribution. It is stated here as a separate theorem in order to highlight how the Brownian integral couplings constructed in the subsequent sections connect to the general theme of coupling hypoelliptic diffusions.

2 Technical preliminaries

To facilitate inductive arguments in the following proofs, we fix a total ordering $\preceq$ of the discrete simplex $\Delta_n = \{(a,b)\in\mathbb{Z}^2 : 0\le a,b,a+b\le n\}$ (for some fixed $n\ge 1$). We achieve this by specifying a function $f:\Delta_n\to\mathbb{Z}$ and defining the order by $(a,b)\preceq(c,d)$ if $f(a,b)\le f(c,d)$.

We choose $f(a,b) = 2na+(2n+1)b$: totality of $\preceq$ follows since f takes values in the totally ordered set $\mathbb{Z}$; antisymmetry holds by a parity argument showing that $f(a,b)=f(c,d)$ if and only if $a=c$ and $b=d$; transitivity is immediate. Note that the $\preceq$-maximal element of Δ is $(0,n)$. We remark that $\preceq$ can be replaced by any other total ordering extending the partial ordering induced by considering $a+b$.

From here onwards, to save cumbersome notation, (W_1, W_2) will denote a two-dimensional Brownian motion starting from a general point $(W_1(0), W_2(0))\in\mathbb{R}^2$. Let $(a,b)\in\Delta_n$ be the index representing the Brownian Stratonovich integral $I_{(a,b)}(t) = I_{(a,b)}(0)+\int_0^t W_1^a W_2^b\circ \mathrm{d}W_2$ (so $I_{(0,0)}(t)=W_2$). We shall refer to such integrals as *monomial Stratonovich integrals*. Consider the $\preceq$-ordered collection of Brownian integrals (deeming W_1 to have precedence over all $I_{(i,j)}$ for $(i,j)\in\Delta$)

$$\mathbf{X}_{(a,b)} \quad=\quad \left(W_1, I_{(c,d)}; (c,d)\preceq(a,b),\ c\ge 1\right)\,.$$

In the following, it is only necessary to consider $c\ge1$; in the case $c=0$, $I_{(c,d)}$ reduces to a monomial in W_2 and so W_2 and its coupled counterpart will take equal values for all time in our coupling construction. Notice also the following: if $a=0$ and (a^-,b^-) is the predecessor of $(a,b)=(0,b)$ in the $\preceq$ ordering, then $\mathbf{X}_{(a^-,b^-)}=\mathbf{X}_{(a,b)}$. This is because $\{(c,d)\preceq(0,b), c\ge1\} = \{(c,d)\prec(0,b),c\ge1\}=\{(c,d)\preceq(a^-,b^-),c\ge1\}$.

Scaling arguments play a major rôle in the study of these couplings. The following lemma records a simple but crucial fact about scaling for Stratonovich integrals of Brownian motions. Consider the scaling transform $\mathcal{S}_r$, defined for any scalar r by

$$\mathcal{S}_r\left(\mathbf{X}_{(a,b)}\right) \quad=\quad \left(rW_1, r^{i+j+1}I_{(i,j)} : (i,j)\preceq(a,b), i\ge1\right)\,.$$

Further, define $W_i^{(r)}(t) = rW_i(0)+(W_i(t)-W_i(0))$ for $i=1,2$ and $I^{(r)}_{(a,b)}(t) = r^{a+b+1}I_{(a,b)}(0)+\int_0^t (W_1^{(r)})^a (W_2^{(r)})^b\circ \mathrm{d}W_2$ for $0\le a,b,a+b\le n$. Write

$$\mathbf{X}^{(r)}_{(a,b)} \quad=\quad \left(W_1^{(r)}, I^{(r)}_{(c,d)}; (c,d)\preceq(a,b),\ c\ge1\right)\,.$$

Lemma 4. The following distributional equality holds:

$$\left(\mathcal{S}_r\left(\mathbf{X}_{(a,b)}\right)(t) : t \geq 0\right) \quad \overset{\mathcal{D}}{=} \quad \left(\mathbf{X}^{(r)}_{(a,b)}(r^2 t) : t \geq 0\right) .$$

Proof. This is a direct consequence of linearity of Stratonovich integration taken together with the Brownian scaling property

$$(rW_i(t) : t \geq 0) \quad \overset{\mathcal{D}}{=} \quad \left(W_i^{(r)}(r^2 t) : t \geq 0\right) \qquad \text{for } i = 1, 2 . \quad \square$$

Note that $\preceq$ is a total ordering extension of the partial order on Δ_1 given by the *scaling degree*, $\deg((a,b)) = \deg(I_{(a,b)}) = a + b + 1$. Monomial Stratonovich integrals of lower degree evolve in time faster than those of higher degree; this is a key reason why our inductive arguments will work.

The following two technical lemmas complete the list of technical preliminaries.

Lemma 5. Let B_t be a standard Brownian motion adapted to a filtration $(\mathcal{F}_t : t \geq 0)$. Let Y_t be a random process and let τ be a stopping time, both adapted to the same filtration.

(i) Suppose there exists $\epsilon > 0$ and some constants C, $\alpha > 0, \beta > 0$, not depending on ε, such that for all $M \geq 1$ and all $t \geq \varepsilon$,

$$\begin{aligned} \mathbb{P}\left[\sup_{t \leq \tau} |Y_t| \geq M\right] &\leq CM^{-\alpha} , \\ \mathbb{P}\left[\tau \geq t\right] &\leq C \left(\frac{\varepsilon}{t}\right)^{\beta} . \end{aligned} \tag{2.1}$$

Then there exists a further constant C' not depending on ε, and positive indices γ', γ'' depending only on α, β such that

$$\begin{aligned} \mathbb{P}\left[\sup_{t \leq \tau}\left|\int_0^t Y_s \, \mathrm{d}B_s\right| \geq x\right] &\leq C' \left(\frac{\varepsilon}{x^4}\right)^{\gamma'} \qquad \text{for } x \geq \varepsilon^{1/4} , \\ \mathbb{P}\left[\sup_{t \leq \tau}\left|\int_0^t Y_s \, \mathrm{d}s\right| \geq x\right] &\leq C' \left(\frac{\varepsilon}{x^2}\right)^{\gamma''} \qquad \text{for } x \geq \varepsilon^{1/2} . \end{aligned} \tag{2.2}$$

Here we may take $\gamma' = (\alpha \wedge \beta \wedge 1)/8$ and $\gamma'' = (\alpha \wedge \beta \wedge 2)/8$.

(ii) Suppose there exists $\epsilon > 0$ and some constants C, $\alpha > 0, \beta > 0$, not depending on ε, such that for all $M \geq \varepsilon$ and all $t \geq 1$,

$$\begin{aligned} \mathbb{P}\left[\sup_{t \leq \tau} |Y_t| \geq M\right] &\leq C \left(\frac{\varepsilon}{M}\right)^{\alpha} , \\ \mathbb{P}\left[\tau \geq t\right] &\leq C\, t^{-\beta} . \end{aligned} \tag{2.3}$$

Then there exists a further constant C' not depending on ε, and positive indices γ', γ'' depending only on α, β such that

$$\begin{aligned} \mathbb{P}\left[\sup_{t \leq \tau}\left|\int_0^t Y_s \, \mathrm{d}B_s\right| \geq x\right] &\leq C' \left(\frac{\varepsilon}{x}\right)^{\gamma'} \text{ for } x \geq \varepsilon , \\ \mathbb{P}\left[\sup_{t \leq \tau}\left|\int_0^t Y_s \, \mathrm{d}s\right| \geq x\right] &\leq C' \left(\frac{\varepsilon}{x^2}\right)^{\gamma''} \text{ for } x \geq \varepsilon^{1/2} . \end{aligned} \tag{2.4}$$

Here we may take $\gamma' = (\alpha \wedge \beta \wedge 1)/2$ and $\gamma'' = (\alpha \wedge \beta \wedge 2)/8$.

Proof. Consider the stopping time

$$\sigma_M \quad = \quad \inf\{t > 0 : |Y_t| \geq M\}\,.$$

The Burholder-Davis-Gundy (BDG) inequality (see for example 17, p. 163), respectively the monotonicity of the Lebesgue integral, implies that, for any $M, T > 0$, there exists a constant $C'' > 0$ not depending on M, T such that

$$\begin{aligned} \mathbb{E}\left[\sup_{t\leq T\wedge\sigma_M}\left|\int_0^t Y_s\,\mathrm{d}B_s\right|^2\right] \quad &\leq \quad C''\,\mathbb{E}\left[\int_0^{T\wedge\sigma_M}(Y_s)^2\,\mathrm{d}s\right] \quad \leq \quad C''M^2T\,, \\ \mathbb{E}\left[\sup_{t\leq T\wedge\sigma_M}\left|\int_0^t Y_s\,\mathrm{d}s\right|\right] \quad &\leq \quad \mathbb{E}\left[\int_0^{T\wedge\sigma_M}|Y_s|\,\mathrm{d}s\right] \quad \leq \quad MT\,. \end{aligned}$$

Under the hypothesis of (i) it follows that, for arbitrary $M \geq 1$ and $T \geq \varepsilon$,

$$\begin{aligned} \mathbb{P}\left[\sup_{t\leq\tau}\left|\int_0^t Y_s\,\mathrm{d}B_s\right| > x\right] \quad &\leq \quad \mathbb{P}\left[\tau > \sigma_M\right] + \mathbb{P}\left[\tau > T\right] + \mathbb{P}\left[\sup_{t\leq T\wedge\sigma_M}\left|\int_0^t Y_s\,\mathrm{d}B_s\right| \geq x\right] \\ &\leq \quad CM^{-\alpha} + C\varepsilon^\beta T^{-\beta} + C''M^2Tx^{-2}\,, \end{aligned} \tag{2.5}$$

where the last inequality follows from the hypothesis of (i) together with a Markov inequality argument. Similarly,

$$\begin{aligned} \mathbb{P}\left[\sup_{t\leq\tau}\left|\int_0^t Y_s\,\mathrm{d}s\right| \geq x\right] \quad &\leq \quad \mathbb{P}\left[\tau > \sigma_M\right] + \mathbb{P}\left[\tau > T\right] + \mathbb{P}\left[\sup_{t\leq T\wedge\sigma_M}\left|\int_0^t Y_s\,\mathrm{d}s\right| \geq x\right] \\ &\leq \quad CM^{-\alpha} + C\varepsilon^\beta T^{-\beta} + MTx^{-1}\,. \end{aligned} \tag{2.6}$$

The first assertion of (i) now follows by optimization. To be explicit, set $T = \sqrt{\varepsilon x}$ and $M = \varepsilon^{-1/8}\sqrt{x}$ in (2.5) and use $\varepsilon \in (0,1)$ (second inequality) followed by $x \geq \varepsilon^{1/4}$ (third inequality) to obtain

$$\begin{aligned} &\mathbb{P}\left[\sup_{t\leq\tau}\left|\int_0^t Y_s\,\mathrm{d}B_s\right| > x\right] \quad \leq \quad C\left(\frac{\varepsilon^{1/4}}{x}\right)^{\alpha/2} + C\left(\frac{\varepsilon}{x}\right)^{\beta/2} + C''\left(\frac{\varepsilon^{1/2}}{x}\right)^{1/2} \\ &\leq \quad C\left(\frac{\varepsilon^{1/4}}{x}\right)^{\alpha/2} + C\left(\frac{\varepsilon^{1/4}}{x}\right)^{\beta/2} + C''\left(\frac{\varepsilon^{1/4}}{x}\right)^{1/2} \quad \leq \quad 3\max\{C, C''\}\left(\frac{\varepsilon^{1/4}}{x}\right)^{(\alpha\wedge\beta\wedge 1)/2}. \end{aligned}$$

The second assertion of (i) follows similarly: set $T = \sqrt{\varepsilon}x^{1/4}$ and $M = \varepsilon^{-1/4}x^{1/4}$ in (2.6), and use $\varepsilon \in (0,1)$ (second inequality) followed by $x \geq \varepsilon^{1/2}$ (third inequality) to obtain

$$\begin{aligned} &\mathbb{P}\left[\sup_{t\leq\tau}\left|\int_0^t Y_s\,\mathrm{d}s\right| \geq x\right] \quad \leq \quad C\left(\frac{\varepsilon}{x}\right)^{\alpha/4} + C\left(\frac{\varepsilon^2}{x}\right)^{\beta/4} + \left(\frac{\varepsilon^{1/2}}{x}\right)^{1/2} \\ &\leq \quad C\left(\frac{\varepsilon^{1/2}}{x}\right)^{\alpha/4} + C\left(\frac{\varepsilon^{1/2}}{x}\right)^{\beta/4} + \left(\frac{\varepsilon^{1/2}}{x}\right)^{1/2} \quad \leq \quad 3\max\{C, 1\}\left(\frac{\varepsilon^{1/2}}{x}\right)^{(\alpha\wedge\beta\wedge 2)/4}. \end{aligned}$$

The proof of (ii) follows along similar lines (using $M = \sqrt{\varepsilon x}$ and $T = \varepsilon^{-1/2}\sqrt{x}$ for the first assertion and $M = \sqrt{\varepsilon}x^{1/4}$ and $T = \varepsilon^{-1/4}x^{1/4}$ for the second assertion). ☐

Lemma 6. Let X_i, τ_i be non-negative random variables adapted to a given filtration $(\mathcal{F}_i : i \geq 1)$ and satisfying

$$\mathbb{P}\left[X_{i+1} > x \mid \mathcal{F}_i\right] \quad \leq \quad C_\alpha x^{-\alpha}\,, \tag{2.7}$$

$$\mathbb{P}\left[\tau_{i+1} > t \mid \mathcal{F}_i\right] \quad \leq \quad C_\beta t^{-\beta}\,, \tag{2.8}$$

for some $\alpha, \beta > 0$ and $x, t \geq 1$, where C_α, C_β are positive constants that do not depend on i. Then for any $\gamma < \alpha \wedge \beta$ there is $\varepsilon_0 > 0$ (depending on α, β and γ) such that

$$\mathbb{P}\left[\tau_1 + \sum_{k=1}^{\infty} \varepsilon_0^k (\Pi_{j=1}^k X_j)\tau_{k+1} > t\right] \quad \leq \quad C'_\beta t^{-\gamma}$$

for some constant C'_β depending only on β, and $t \geq 1$.

Proof. Set $\Pi_k = \Pi_{j=1}^k X_j$ for $k \geq 1$, with $\Pi_0 = 1$. Take any $\gamma < \alpha \wedge \beta$. Using $\mathbb{E}\left[X_{i+1}^\gamma \mid \mathcal{F}_i\right] \leq 1 + \int_1^\infty \mathbb{P}\left[X_{i+1}^\gamma > x \mid \mathcal{F}_i\right] dx$, it follows from (2.7) that $\mathbb{E}\left[X_{i+1}^\gamma \mid \mathcal{F}_i\right] \leq 1 + \frac{\gamma C_\alpha}{\alpha - \gamma} < \infty$. Hence, for $\varepsilon > 0$,

$$\mathbb{E}\left[(\varepsilon^k \Pi_k)^\gamma\right] \quad \leq \quad \left[\varepsilon\left(1 + \frac{\gamma C_\alpha}{\alpha - \gamma}\right)^{1/\gamma}\right]^{k\gamma}. \tag{2.9}$$

For any $\varepsilon > 0$, we can write for any $t \geq 1$,

$$\begin{aligned}\mathbb{P}\left[\sum_{k=0}^{\infty} \varepsilon^k \Pi_k \tau_{k+1} > t^{\beta/\gamma}\right] \quad &\leq \\ \mathbb{P}\left[\sum_{k=0}^{\infty} \varepsilon^k \Pi_k \tau_{k+1} > \sum_{k=0}^{\infty} 2^{-k} (\varepsilon^k \Pi_k)^{1-\frac{\gamma}{\beta}} t\right] &+ \mathbb{P}\left[\sum_{k=0}^{\infty} 2^{-k} (\varepsilon^k \Pi_k)^{1-\frac{\gamma}{\beta}} > t^{\frac{\beta}{\gamma}-1}\right]\end{aligned} \tag{2.10}$$

(adopting the convention that $\mathbb{P}[Z > \infty] = 0$ for any random variable Z). Take $\varepsilon_0 > 0$ satisfying $\varepsilon_0(1 + \frac{\gamma C_\alpha}{\alpha - \gamma})^{1/\gamma} = 4^{-\beta/\gamma}$. Then

$$\begin{aligned}\mathbb{P}\left[\sum_{k=0}^{\infty} \varepsilon_0^k \Pi_k \tau_{k+1} > \sum_{k=0}^{\infty} 2^{-k} (\varepsilon_0^k \Pi_k)^{1-\frac{\gamma}{\beta}} t\right] \quad &\leq \quad \sum_{k=0}^{\infty} \mathbb{P}\left[\tau_{k+1} > 2^{-k} (\varepsilon_0^k \Pi_k)^{-\gamma/\beta} t\right] \\ \leq \quad \max\{1, C_\beta\} t^{-\beta} \sum_{k=0}^{\infty} 2^{k\beta}\, \mathbb{E}\left[(\varepsilon_0^k \Pi_k)^\gamma\right] \quad &\leq \quad \max\{1, C_\beta\} t^{-\beta} \sum_{k=0}^{\infty} 2^{-k\beta} \quad = \quad C'_\beta t^{-\beta},\end{aligned}$$

where the second inequality is obtained using (2.8), and the third by using (2.9) together with the specific choice of ε_0. We have used $\max\{1, C_\beta\}$ in place of C_β to account for the situation when $2^{-k}(\varepsilon_0^k \Pi_k)^{-\gamma/\beta} t < 1$. Furthermore

$$\begin{aligned}\mathbb{P}\left[\sum_{k=1}^{\infty} 2^{-k} (\varepsilon_0^k \Pi_k)^{1-\frac{\gamma}{\beta}} > t^{\frac{\beta}{\gamma}-1}\right] \quad &\leq \quad \sum_{k=1}^{\infty} \mathbb{P}\left[(\varepsilon_0^k \Pi_k)^\gamma > t^\beta\right] \\ &\leq \quad t^{-\beta} \sum_{k=1}^{\infty} \mathbb{E}\left[(\varepsilon_0^k \Pi_k)^\gamma\right] \quad \leq \quad t^{-\beta} \sum_{k=1}^{\infty} 4^{-k\beta} = C'_\beta t^{-\beta},\end{aligned}$$

where we have used the Markov inequality to obtain the second inequality above and (2.9) together with the choice of ε_0 for the third inequality.

The above estimates can be used with (2.10) to show

$$\mathbb{P}\left[\sum_{k=1}^{\infty} \varepsilon_0^k \Pi_k \tau_{k+1} > t^{\beta/\gamma}\right] \quad \leq \quad 2C'_\beta t^{-\beta},$$

which proves the lemma. □

3 Coupling BM($\mathbb{R}^2$) and a single monomial Stratonovich integral

In this section, we construct couplings of $(W_1, W_2, I_{(a,b)})$ and $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(a,b)})$ for $(a, b) \in \Delta_n, a \geq 1, b \geq 0$. The cases $a + b = 1$ (which implies $a = 1$, $b = 0$) and $a + b > 1$ differ in complexity, so we first consider the simpler case $a + b = 1$ (Lemma 7). This case is significantly easier to describe, and corresponds to the case of Brownian motion on the Heisenberg group already treated in [6] and [19] as noted in Remark 8 below; however the present technique carries through to the case $a + b > 1$ (Lemma 9). Thus, the construction given in the simplest non-trivial case (Lemma 7) is a good model for the general approach. Lemma 9 deals with coupling just one monomial Stratonovich integral of more general form, but this is an essential component of the inductive argument that will be required to establish coupling for a finite set of monomial Stratonovich integrals in Section 4.

We will use some further notation, namely $\Delta W_1 = W_1 - \widetilde{W}_1$ and $\Delta I_{(a,b)} = I_{(a,b)} - \widetilde{I}_{(a,b)}$.

3.1 Case of simplest non-trivial monomial Stratonovich integral

The next lemma establishes a coupling result based on a driving 2-dimensional Brownian motion W_1, W_2 plus the single monomial stochastic integral $I_{(1,0)}$.

Lemma 7. For any $\gamma < \frac{1}{3}$, there exists a successful Markovian coupling $\mathbb{P}_\gamma$ of $(W_1, W_2, I_{(1,0)})$ and $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(1,0)})$ started from distinct points $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \widetilde{\mathfrak{i}})$ respectively, with coupling time T_γ, satisfying

$$\sup_{w_1, w_2, |\mathfrak{i} - \widetilde{\mathfrak{i}}| \leq 1} \mathbb{P}_\gamma \left[T_\gamma > t\right] \leq C_\gamma t^{-\gamma}, \quad t \geq 1. \tag{3.1}$$

Proof. We first outline the general proof strategy. At all times W_2 and $\widetilde{W}_2$ will be synchronously coupled; hence $W_2(t) = \widetilde{W}_2(t)$ for all $t \geq 0$. Brownian scaling as given in Lemma 4 can be used to re-scale to a unit difference between the two stochastic integrals, thus reducing all cases to the case of starting points $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \mathfrak{i} - 1)$ for $w, i \in \mathbb{R}$. The coupling decomposes naturally into disjoint cycles. Each cycle consists of a patterned alternation between phases of reflection and synchronous coupling for W_1 and $\widetilde{W}_1$, so that the distance between the coupled processes $(W_1, W_2, I_{(1,0)})$ and $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(1,0)})$ at the end of the cycle is roughly a fixed proportion of the distance between them at the start of the cycle. At the end of each cycle, the next cycle is constructed by applying the same coupling strategy as the previous cycle after appropriately re-scaling the coupled processes *via* Lemma 4, so that there is unit re-scaled distance between them at the start of the next cycle. Lemma 6 is then used to show that the end-points of these cycles have an accumulation point which corresponds to a finite coupling time. As the coupling strategy within each cycle is the same (modulo re-scaling), it is sufficient to describe in detail only the construction of the first cycle. Note that iterated cycles and re-scaling to achieve successful coupling have been used to couple Kolmogorov diffusions by Ben Arous et al. [6], Kendall and Price [22], Banerjee and Kendall [3].

A: Description of the first cycle

As noted before, the scaling argument represented by Lemma 4 shows there is no loss of generality in assuming that $|\Delta I_{(1,0)}(0)| = 1$. Choose and fix a constant $R > 1$. The estimates derived for the first cycle will be uniform with respect to $R > 1$ and an optimal choice of R will be made at the end of the proof. In the proof, $C, C_1, C_2, \ldots$ will denote generic positive constants whose values will not depend on $R, w_1, w_2, \mathfrak{i}$ and whose value might change from line to line. The first cycle consists of three phases whose end-points are defined by the following stopping times:

1:	T_1	$=$	$\inf\{t \geq 0 : \lvert\Delta W_1(t)\rvert = R^{-1}\}$,	reflection till $W_1 - \widetilde{W}_1$ hits $\pm R^{-1}$;
2:	T_2	$=$	$\inf\{t \geq T_1 : \Delta I_{(1,0)}(t) = 0\}$,	synchronous till $I_{(1,0)} - \widetilde{I}_{(1,0)}$ hits 0;
3:	T_3	$=$	$\inf\{t \geq T_2 : \Delta W_1(t) = 0\}$,	reflection till $W_1 - \widetilde{W}_1$ hits 0.

Phase 1: Using Brownian scaling, independent Brownian increments, and eigenvalues of the Laplacian with Dirichlet boundary conditions on $[-1,1]$, together with $W_1(0) = \widetilde{W}_1(0)$, it follows that

$$\mathbb{P}\left[T_1 > t/R^2\right] \quad \leq \quad C e^{-\pi^2 t/8} \,. \tag{3.2}$$

Now consider the increment of $\Delta I_{(a,b)}$ over the time interval $[0, T_1]$. Since the second Brownian coordinates satisfy $W_2 = \widetilde{W}_2$ throughout the entire coupling, and the first Brownian coordinates W_1, $\widetilde{W}_1$ are reflection coupled hence independent of $W_2 = \widetilde{W}_2$, we may re-write the Stratonovich integral for the increment as an Itô integral:

$$\Delta I_{(1,0)}(T_1) - \Delta I_{(1,0)}(0) \quad = \quad \int_0^{T_1} \Delta W_1(s)\, \mathrm{d}W_2(s) \,.$$

On the other hand, $\sup_{t\in[0,T_1]} |\Delta W_1(t)| = \frac{1}{R}$ by definition of T_1. Using the L^2-isometry of the Itô integral,

$$\mathbb{E}\left[\left|\int_0^{T_1} \Delta W_1(s)\, \mathrm{d}W_2(s)\right|^2\right] \quad = \quad \mathbb{E}\left[\int_0^{T_1} (\Delta W_1(s))^2\, \mathrm{d}s\right] \quad \leq \quad \frac{1}{R^2}\,\mathbb{E}\left[T_1\right] \quad \leq \quad \frac{C}{R^4} \,,$$

where the last inequality follows from (3.2) using $\mathbb{E}\left[T_1\right] = \int_0^\infty \mathbb{P}\left[T_1 > t\right]\, \mathrm{d}t$.

By a Markov inequality argument, it now follows for any $x > 0$ that

$$\mathbb{P}\left[|\Delta I_{(1,0)}(T_1) - \Delta I_{(1,0)}(0)| > x\right] \quad \leq \quad \frac{C}{R^4 x^2} \,.$$

But we have assumed that $|\Delta I_{(1,0)}(0)| = 1$, so for any $x \geq 2$

$$\begin{aligned}\mathbb{P}\left[|\Delta I_{(1,0)}(T_1)| > x\right] \quad &\leq \quad \mathbb{P}\left[|\Delta I_{(1,0)}(T_1) - \Delta I_{(1,0)}(0)| > x - 1\right] \quad \leq \quad \frac{C}{R^4 (x-1)^2} \\ &\leq \quad \frac{4C}{R^4 x^2} \,.\end{aligned} \tag{3.3}$$

Phase 2: Because we are constructing a Markovian coupling, we may condition on the past of the driving Brownian motions till time T_1. Under synchronous coupling the Stratonovich expression for the increment of $\Delta I_{(a,b)}$ over the time interval $[T_1, T_2]$ can again be re-written as an Itô integral, only now $\Delta W_1(s) = \Delta W_1(T_1)$ while $s \in [T_1, T_2]$. Thus in this time interval $\Delta I_{(a,b)}(s) - \Delta I_{(a,b)}(T_1) = \Delta W_1(T_1) \times (W_2(s) - W_2(T_1))$. Since $\Delta W_1(T_1) = R^{-1}$ by construction of T_1, it follows that $T_2 - T_1$ has the same distribution as the hitting time of a one-dimensional Brownian motion on the level $-R\operatorname{sgn}(\Delta W_1(T_1))\Delta I_{(1,0)}(T_1)$. Thus, for $x \geq 2$ and $t > 0$, we can assert that

$$\begin{aligned}\mathbb{P}\left[T_2 - T_1 > t\right] \quad &\leq \quad \mathbb{P}\left[|\Delta I_{(1,0)}(T_1)| > x\right] + \mathbb{P}\left[T_2 - T_1 > t \text{ and } |\Delta I_{(1,0)}(T_1)| \leq x\right] \\ &\leq \quad \frac{4C}{R^4 x^2} + \frac{CRx}{\sqrt{t}} \,,\end{aligned}$$

where the last inequality is a consequence of (3.3) and a hitting time estimate for Brownian motions derived from the reflection principle. Taking $x = t^{1/6}$ in the above expression and recalling that $R > 1$,

$$\mathbb{P}\left[T_2 - T_1 > t\right] \quad \leq \quad \frac{4C}{R^4 t^{1/3}} + \frac{CR}{t^{1/3}} \quad \leq \quad \frac{5CR}{t^{1/3}} \qquad \text{for } t \geq 2^6 \,.$$

The above expression gives a useful bound on the probability $\mathbb{P}[T_2 - T_1 > t]$ only when $t \geq (5CR)^3$. We therefore adjust the above bound (using C as a new generic positive constant):

$$\mathbb{P}[T_2 - T_1 > t] \quad \leq \quad \frac{CR}{t^{1/3}} \qquad \text{for } t \geq CR^3 \,. \tag{3.4}$$

Note that $\Delta I_{(1,0)}(T_2) = 0$ follows from the definition of T_2.

Phase 3: Using reflection coupling, and conditioning on the past at time T_2, we may view $T_3 - T_2$ as the hitting time of level $\frac{1}{2R}$ by a standard Brownian motion. Employing the reflection principle for Brownian motion

$$\mathbb{P}\left[T_3 - T_2 > t/R^2\right] \quad \leq \quad \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{t}} \qquad \text{for } t \geq 1 \,. \tag{3.5}$$

Moreover, for $x > R^{-2}, H > R^{-2}$, and once again re-writing the Stratonovich integral of $\Delta I_{(1,0)}$ as an Itô integral,

$$\begin{aligned}
\mathbb{P}\left[|\Delta I_{(1,0)}(T_3)| > x\right] \quad &\leq \quad \mathbb{P}[T_3 - T_2 > H] + \mathbb{P}\left[|\Delta I_{(1,0)}(T_3)| > x, T_3 - T_2 \leq H\right] \\
&\leq \quad \mathbb{P}[T_3 - T_2 > H] + \mathbb{P}\left[\sup_{T_2 \leq t \leq T_2 + H}\left|\int_{T_2}^{t} (\Delta W_1)(s)\,\mathrm{d}W_2(s)\right| > x\right] \\
&\leq \quad \sqrt{\frac{2}{\pi}}\frac{1}{R\sqrt{H}} + \frac{CH}{x^2}\,\mathbb{E}\left[\sup_{T_2 \leq t \leq T_2 + H} |\Delta W_1(t)|^2\right] \\
&\qquad\qquad \text{(by (3.5), Tchebychev and BDG inequalities)} \\
&\leq \quad \sqrt{\frac{2}{\pi}}\frac{1}{R\sqrt{H}} + \frac{CH}{x^2}\left(\frac{1}{R^2} + H\right) \\
&\qquad \text{(using } |\Delta W_1(T_2)| = R^{-1} \text{ and Doob's } L^2\text{-maximal inequality)} \\
&\leq \quad \sqrt{\frac{2}{\pi}}\frac{1}{R\sqrt{H}} + \frac{CH^2}{x^2} \quad \left(\text{as } H > \frac{1}{R^2}\right).
\end{aligned}$$

Taking $H = \frac{x^{4/5}}{R^{2/5}}$, we obtain a bound

$$\mathbb{P}\left[|\Delta I_{(1,0)}(T_3)| > x\right] \quad \leq \quad \frac{C}{(R^2 x)^{2/5}}\,, \qquad \text{for } x > \frac{1}{R^2}\,. \tag{3.6}$$

Moreover, the combined effect of the estimates in (3.2), (3.4) and (3.5) can be summarized as

$$\mathbb{P}[T_3 > t] \quad \leq \quad \frac{CR}{t^{1/3}}, \qquad \text{for } t \geq CR^3 \,. \tag{3.7}$$

The estimates (3.6) and (3.7) give bounds on the difference of the integrals $I_{(1,0)}$ and $\widetilde{I}_{(1,0)}$ at the end of the first cycle and the time taken to complete the first cycle respectively.

B: Describing subsequent cycles and successful coupling

For $t \geq T_3$, define further stopping times $T_k, k > 3$, such that for any $k \geq 1$,

$$|\Delta I_{(1,0)}(T_{3k})|^{-1} T_{3k+j}, \; j = 1, 2, 3,$$

is the time of completion of the j^{th} phase of the first cycle constructed above for the re-scaled processes

$$(|\Delta I_{(1,0)}(T_{3k})|^{-1/2} W_1(T_{3k}+t), |\Delta I_{(1,0)}(T_{3k})|^{-1/2} W_2(T_{3k}+t), |\Delta I_{(1,0)}(T_{3k})|^{-1} I_{(0,1)}(T_{3k}+t))_{t \geq 0}$$

and

$$(|\Delta I_{(1,0)}(T_{3k})|^{-1/2} \widetilde{W}_1(T_{3k}+t), |\Delta I_{(1,0)}(T_{3k})|^{-1/2} \widetilde{W}_2(T_{3k}+t), |\Delta I_{(1,0)}(T_{3k})|^{-1} \widetilde{I}_{(0,1)}(T_{3k}+t))_{t \geq 0}$$

in place of $(W_1(t), W_2(t), I_{(0,1)}(t))_{t\geq 0}$ and $(\widetilde{W}_1(t), \widetilde{W}_2(t), \widetilde{I}_{(0,1)}(t))_{t\geq 0}$ respectively.

The concatenation of these cycles does in fact lead to a successful coupling. The proof of this follows from two facts:

(i) $\lim_{k\to\infty} \Delta I_{(1,0)}(T_{3k}) = 0$, meaning that the coupled processes, observed at the end-points of the cycles, come arbitrarily close as the number of cycles becomes large, and

(ii) $\lim_{k\to\infty} T_{3k} < \infty$ almost surely, meaning that the end points of these cycles have a finite accumulation point T_∞, so that the concatenation completes in finite time.

Continuity of Brownian motion and stochastic integrals then implies successful coupling at time T_∞.

We now demonstrate that these two facts follow from Lemma 6. Define

$$\tau_k^* = \frac{T_{3k} - T_{3k-3}}{|\Delta I_{(1,0)}(T_{3k-3})|}, \quad X_k^* = \frac{R^2|\Delta I_{(1,0)}(T_{3k})|}{|\Delta I_{(1,0)}(T_{3k-3})|}, \quad \text{for } k \geq 1\,, \tag{3.8}$$

where we take $T_0 = 0$. Applying Lemma 4 to (3.6), the pair $(X_k^*, 1)$ satisfies the hypotheses of (X_k, τ_k) of Lemma 6 with $\alpha = 2/5$ and any $\beta > 0$. Thus, Lemma 6 implies that there is $R_0 > 1$ such that for all $R \geq R_0$,

$$\sum_{k=1}^{\infty} R^{-2k}\left(\Pi_{j=1}^{k} X_j^*\right) \quad < \quad \infty, \text{ almost surely.}$$

Choosing $R \geq R_0$ in the description of the first cycle, $|\Delta I_{(1,0)}(T_{3k})| = R^{-2k}\left(\Pi_{j=1}^{k} X_j^*\right)$ and consequently, $\lim_{k\to\infty} \Delta I_{(1,0)}(T_{3k}) = 0$ almost surely.

To prove that the coupling is successful in finite time almost surely and that the coupling time has a power law tail given by (3.1), apply Lemma 4 to (3.6) and (3.7): $(X_k^*, \tau_k^*/R^3)$ satisfies the hypotheses of (X_k, τ_k) of Lemma 6 with $\alpha = 2/5$ and $\beta = 1/3$. Thus, by Lemma 6, for any $0 < \gamma < 1/3$, there is $R_\gamma > 1$ such that for any $R \geq R_\gamma$,

$$\mathbb{P}\left[\tau_1^* + \sum_{k=1}^{\infty} R^{-2k}\left(\Pi_{j=1}^{k} X_j^*\right)\tau_{k+1}^* > R^3 t\right] \quad \leq \quad C t^{-\gamma}\,. \tag{3.9}$$

Now observe the following product collapses because of the definitions expressed by (3.8):

$$R^{-2k}\left(\Pi_{j=1}^{k} X_j^*\right)\tau_{k+1}^* \quad = \quad (T_{3(k+1)} - T_{3k})/\Delta I_{(0,1)}(T_0)\,.$$

Thus, for any $0 < \gamma < 1/3$, the above coupling construction with $R = \max\{R_0, R_\gamma\}$ gives the required successful coupling satisfying (3.1). □

Remark 8. Recall the Brownian motion in the Heisenberg group started at $(w_1, w_2, \mathfrak{i})$, defined as the $\mathbb{R}^3$ valued process given by

$$((W_1(t), W_2(t), \mathfrak{i} + \int_0^t W_1(s)\,\mathrm{d}W_2(s) - \int_0^t W_2(s)\,\mathrm{d}W_1(s)) : t \geq 0),$$

where (W_1, W_2) is a two-dimensional Brownian motion started at (w_1, w_2). Lemma 7 is of independent interest as it gives a successful Markovian coupling of Brownian motions on the Heisenberg group started at $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \widetilde{\mathfrak{i}})$ with explicit bounds on the tail probabilities of the coupling time. To see this, note that by the Itô formula,

$$W_1(t)W_2(t) - W_1(0)W_2(0) \quad = \quad \int_0^t W_1(s)\,\mathrm{d}W_2(s) + \int_0^t W_2(s)\,\mathrm{d}W_1(s)\,, \qquad \text{for } t \geq 0.$$

From this, we obtain

$$\begin{aligned}
\mathfrak{i} + \int_0^t W_1(s)\,\mathrm{d}W_2(s) - \int_0^t W_2(s)\,\mathrm{d}W_1(s) \\
&= 2\left(\frac{\mathfrak{i}}{2} + \int_0^t W_1(s)\,\mathrm{d}W_2(s)\right) \\
&\quad - \left(\int_0^t W_1(s)\,\mathrm{d}W_2(s) + \int_0^t W_2(s)\,\mathrm{d}W_1(s)\right) \\
&= 2\left(\frac{\mathfrak{i}}{2} + \int_0^t W_1(s)\,\mathrm{d}W_2(s)\right) - \left(W_1(t)W_2(t) - W_1(0)W_2(0)\right) .
\end{aligned}$$

Thus, the successful coupling construction given in Lemma 7 for $(W_1, W_2, I_{(1,0)})$ and $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(1,0)})$, started from $(w_1, w_2, \mathfrak{i}/2)$ and $(w_1, w_2, \widetilde{\mathfrak{i}}/2)$ respectively, is also a successful coupling of the corresponding Brownian motions on the Heisenberg group started from $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \widetilde{\mathfrak{i}})$. Couplings of Brownian motions on the Heisenberg group have appeared in several papers in recent times: [6] and [19] have constructed successful Markovian couplings of Brownian motions for the Heisenberg group. Kendall [20, Theorem 3.1] established some coupling time distribution asymptotics for the coupling constructed in [19], under some limiting operation on the starting points. But our result gives explicit bounds on the tail probabilities of the coupling time for each t and each pair of starting points $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \widetilde{\mathfrak{i}})$ (in fact, this coupling can be extended to general pairs of distinct starting points $(w_1, w_2, \mathfrak{i})$ and $(\widetilde{w}_1, \widetilde{w}_2, \widetilde{\mathfrak{i}})$ and associated bounds can be derived). Moreover, the tail probabilities of the coupling time of Kendall [20, Theorem 3.1] decay at best like $t^{-1/6}$; the rate in Lemma 7 is significantly better. We note here however (a) that the treatment by [6] and [19] uses an invariant difference that permits a generalization which couples all possible stochastic areas for a d-dimensional Brownian motion [19]; (b) that Banerjee et al. [2, Lemma 3.1] obtained a non-Markovian coupling for Brownian motions on the Heisenberg group started at $(w_1, w_2, \mathfrak{i})$ and $(w_1, w_2, \widetilde{\mathfrak{i}})$ that attains the total variation bound (the best possible bound on the tails of coupling time distribution), and decays like t^{-1}, which is significantly better than the bound in Lemma 7. Moreover Markovian couplings cannot reach a bound that decays faster than $t^{-1/2}$ [2, Remark 3.2]. It would be interesting to investigate whether the bound $t^{-1/2}$ can be attained, or whether $t^{-1/3}$ is the best bound for Markovian couplings.

3.2 Case of general monomial Stratonovich integral

The next lemma generalizes the previous coupling construction, establishing a coupling result based on a driving 2-dimensional Brownian motion plus a single monomial stochastic integral: $(W_1, W_2, I_{(a,b)})$ for a single fixed $(a,b) \in \Delta_n$ with $a \geq 1$, $b \geq 0$ and $a + b > 1$. Recall from Section 2 that $f(k,l) = 2nk + (2n+1)l$.

Lemma 9. For any $(a,b) \in \Delta_n$ with $a \geq 1$, $b \geq 0$, $a+b > 1$, there exists $R_0 > 1$ such that for each $R \geq R_0$, we can obtain a successful Markovian coupling construction $\mathbb{P}_R$ of $(W_1, W_2, I_{(a,b)})$ and $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(a,b)})$ starting from $(w, Rw, \mathfrak{i})$ and $(w, Rw, \widetilde{\mathfrak{i}})$ respectively, with coupling time $T_{R,(a,b)}$, such that:

(i) There are positive constants γ, C not depending on R such that, for large t,

$$\sup\left\{\mathbb{P}_R\left[T_{R,(a,b)} > R^{4n+2}t\right] \;:\; w, \mathfrak{i}, \widetilde{\mathfrak{i}} \in \mathbb{R}, |\mathfrak{i} - \widetilde{\mathfrak{i}}| \leq 1\right\} \quad \leq \quad Ct^{-\gamma} . \tag{3.10}$$

In the interval $[0, T_{R,(a,b)}]$ we identify the *active region* $S_{R,(a,b)}$,

$$S_{R,(a,b)} \quad = \quad \text{closure of } \{t \leq T_{R,(a,b)} : W_1(t) \neq \widetilde{W}_1(t)\}.$$

$S_{R,(a,b)}$ depends on R, a, b; both $[0, T_{R,(a,b)}) \cap S_{R,(a,b)}$ and $[0, T_{R,(a,b)}) \setminus$ (interior of $S_{R,(a,b)}$) are unions of countable sequences of disjoint random closed intervals, where for each sequence of intervals the left-end-points of the intervals form an increasing sequence. Writing the total length of $S_{R,(a,b)}$ by $|S_{R,(a,b)}|$, the following holds for large t,

$$\sup\left\{\mathbb{P}_R\left[|S_{R,(a,b)}| > t\right] \;:\; w, \mathfrak{i}, \widetilde{\mathfrak{i}} \in \mathbb{R}, |\mathfrak{i} - \widetilde{\mathfrak{i}}| \leq 1\right\} \quad \leq \quad Ct^{-\gamma}\,. \tag{3.11}$$

(ii) There are positive constants α, C not depending on R such that for large t,

$$\sup\left\{\mathbb{P}_R\left[\sup_{t \leq T_{R,(a,b)}} |W_1(t) - \widetilde{W_1}(t)| > x/R^{f(a-1,b)}\right] \;:\; w, \mathfrak{i}, \widetilde{\mathfrak{i}} \in \mathbb{R}, |\mathfrak{i} - \widetilde{\mathfrak{i}}| \leq 1\right\} \\ \leq \quad Cx^{-\alpha}\,. \tag{3.12}$$

For convenience, we will prove the inequalities (3.10), (3.11) and (3.12) for $t \geq 1$.

Proof. As before, W_2 and $\widetilde{W_2}$ will be synchronously coupled at all times so we may take $W_2 = \widetilde{W_2}$. Brownian scaling (Lemma 4) can be applied to ensure the monomial stochastic integrals differ by 1: so it suffices to consider starting points $(w, Rw, \mathfrak{i})$ and $(w, Rw, \mathfrak{i} - 1)$ for $w, \mathfrak{i} \in \mathbb{R}$. Let $\gamma, \delta, C, C_1, C_2 \ldots$ be generic positive constants not depending on $R, w, \mathfrak{i}$, (but often depending on a and b) whose values might change from line to line.

The proof uses some martingale estimates, so we will use the decomposition of the Stratonovich integral $I_{(a,b)}$ in Itô integral form:

$$I_{(a,b)}(t) \quad = \quad I_{(a,b)}(0) + \int_0^t W_1(s)^a W_2(s)^b \,\mathrm{d}W_2(s) + \frac{b}{2}\int_0^t W_1(s)^a W_2(s)^{b-1}\,\mathrm{d}s\,. \tag{3.13}$$

In contrast with the case of Lemma 7, here the Stratonovich integral has a drift component if $b \geq 1$.

As in the previous lemma, the coupling decomposes into disjoint cycles, and the successive cycles are connected via scaling. We describe the first cycle and then discusses the total effect of this and subsequent cycles on finiteness and moment estimates for the coupling time.

A: Description of the first cycle

The first cycle consists of five phases. The coupling strategy alternates between synchronous coupling and reflection coupling of W_1 and $\widetilde{W_1}$ between the phases. We will first describe each phase in terms of an arbitrary value of the tuning parameter $R \geq R_0 > 1$. The estimates derived for the first cycle will hold uniformly with respect to $R > 1$ and the appropriate lower bound R_0 for R will arise in the course of the proof and be specified at the end of the coupling construction. The end-points of the five phases are defined by the following stopping times. Initially $W_1(0) = \widetilde{W_1}(0)$ ($W_2 = \widetilde{W_2}$ throughout.)

1:	$\theta_1 = \inf\{t \geq 0 : W_2(t) = RW_1(t) \text{ and } \lvert W_1(t)\rvert \geq R^{2n}\}$,	synchronous till W_1 hits $R^{-1}W_2$ **and** $\lvert W_1\rvert \geq R^{2n}$, and note $W_1(\theta_1) = \widetilde{W_1}(\theta_1)$;
2:	$\tau_1 = \inf\left\{t \geq \theta_1 : \lvert\Delta W_1(t)\rvert = \frac{1}{\lvert W_1(\theta_1)\rvert^{a+b-1}R^b}\right\}$,	reflection till ΔW_1 hits $\pm\frac{1}{\lvert W_1(\theta_1)\rvert^{a+b-1}R^b}$, and note $W_1(\tau_1) - W_1(\theta_1) = -(\widetilde{W_1}(\tau_1) - \widetilde{W_1}(\theta_1))$;

3: $\eta_1 = \inf\{t \geq \tau_1 : W_2(t) = W_2(\tau_1) - a^{-1}\operatorname{sgn}(\Delta W_1(\tau_1))(\operatorname{sgn}(W_1(\theta_1)))^{a+b-1}\}$, synchronous till $W_2 - W_2(\tau_1)$ hits $-\frac{\operatorname{sgn}(\Delta W_1(\tau_1))(\operatorname{sgn}(W_1(\theta_1)))^{a+b-1}}{a}$, and note $W_1(\eta_1) - W_1(\tau_1) = \widetilde{W}_1(\eta_1) - \widetilde{W}_1(\tau_1)$;

4: $\lambda_1 = \inf\{t \geq \eta_1 : \Delta W_1(t) = 0\}$, reflection till ΔW_1 hits 0, and note $W_1(\lambda_1) = \widetilde{W}_1(\lambda_1)$;

5: $\beta_1 = \inf\{t \geq \lambda_1 : W_2(t) = RW_1(t)\}$, synchronous till $RW_1 - W_2$ hits 0, and note $W_1(\beta_1) = \widetilde{W}_1(\beta_1)$.

Note that the first and last phases both use synchronous coupling. However we do not amalgamate these across cycles, since at the β_k times we have $RW_1 = W_2$ as well as $W = \widetilde{W}$.

Phase 1: In this phase, synchronous coupling of W_1 and $\widetilde{W_1}$ is applied on $[0, \theta_1]$, where θ_1 is the stopping time defined above. The Brownian motions agree at time 0 and are synchronously coupled on $[0, \theta_1]$, so agree over the whole interval $[0, \theta_1]$. Therefore $\Delta I_{(a,b)}(\theta_1) = \Delta I_{(a,b)}(0) = 1$.

We first estimate the tail probability of θ_1 as follows. If $t \geq 1$ then

$$\mathbb{P}\left[\theta_1 > R^{4n+2}t\right] \quad \leq \quad \frac{C \log t}{\sqrt{t}}\,. \tag{3.14}$$

To see this, note that θ_1 is obtained by starting a planar Brownian motion located at distance $\sqrt{1+R^2}w$ along the line $W_2 = RW_1$ from the origin, and running it till it hits the diagonal $W_2 = RW_1$ at a distance at least R^{2n} from the origin. By Brownian scaling and rotational invariance of planar Brownian motion, θ_1 is stochastically dominated by $R^{4n+2}\theta_1'$, where

$$\theta_1' \quad = \quad \inf\{t > 0 : W_1^*(t) = 0, |W_2^*(t)| \geq \sqrt{2}\}\,.$$

Here (W_1^*, W_2^*) is a planar Brownian motion (W_1^*, W_2^*) with $W_1^*(0) = 0, W_2^*(0) = \frac{\sqrt{1+R^2}w}{R^{2n+1}}$; if $\sqrt{2}$ were replaced by $\sqrt{1+R^{-2}}$ then the stochastic domination would become an equality (recall, $R > 1$). If $L(t)$ denotes the local time of W_1^* at 0 at time t and $\zeta(t)$ denotes the inverse local time, then $W_2^*(\zeta(t)) = \mathcal{C}(t)$, where $\mathcal{C}$ is a Cauchy process starting at $\sqrt{1+R^2}w/R^{2n+1}$. If $L(\theta_1') > s$, then the continuity of L implies $\theta_1' > \zeta(s)$. The range of ζ is a subset of the set of times where the monotone function L increases (namely, the times where $W_1^* = 0$), so $\theta_1' > \zeta(s)$ yields $\sqrt{2} > W_2^*(\zeta(s))$ from the definition of θ_1'. Hence, for $t \geq 1$, and $u = C_2^{-1}\log t$ for a certain positive constant C_2,

$$\begin{aligned}\mathbb{P}\left[\theta_1' > t\right] \quad &\leq \quad \mathbb{P}\left[L(t) \leq u\right] + \mathbb{P}\left[\theta_1' > t, L(t) > u\right] \\ &\leq \quad \mathbb{P}\left[L(t) \leq u\right] + \mathbb{P}\left[L(\theta_1') > u\right] \quad \leq \quad \mathbb{P}\left[L(t) \leq u\right] + \mathbb{P}\left[\sup_{s \leq u} |\mathcal{C}(s)| \leq \sqrt{2}\right]\,. \end{aligned} \tag{3.15}$$

By the Lévy transform, the local time process $(L(s) : s \geq 0)$ has the distribution of the running supremum of Brownian motion, so

$$\mathbb{P}\left[L(t) \leq u\right] \quad \leq \quad \frac{2}{\pi}\frac{u}{\sqrt{t}}\,.$$

To bound the second probability in (3.15), recall that the Cauchy process $\mathcal{C}$ is a pure jump Lévy process. Consequently, the increments $(\mathcal{C}(j) - \mathcal{C}(j-1) : 1 \leq j \leq \lfloor u \rfloor)$ are i.i.d. with a common Cauchy distribution. If $\sup_{s\le u}|\mathcal{C}(s)|\le\sqrt{2}$ holds then $|\mathcal{C}(j)-\mathcal{C}(j-1)|\le 2\sqrt{2}$ for $1\le j\le\lfloor u\rfloor$, and therefore (using positive constants C_1, C_2 not depending on w)

$$\mathbb{P}\left[\sup_{s\le u}|\mathcal{C}(s)|\le\sqrt{2}\right] \quad\le\quad C_1 e^{-C_2 u}\,.$$

Applying these bounds to (3.15),

$$\mathbb{P}\left[\theta_1'>t\right] \quad\le\quad \frac{\sqrt{2}u}{\sqrt{\pi t}}+C_1 e^{-C_2 u}\,.$$

The required bound (3.14) follows by taking $u=C_2^{-1}\log t$ and choosing a suitable C bearing in mind that $t\ge 1$.

Phase 2: This phase employs reflection coupling between W_1 and $\widetilde{W}_1$, and runs from time θ_1 till the stopping time

$$\tau_1 \quad=\quad \inf\left\{t\ge\theta_1 : |\Delta W_1(t)|=\frac{1}{|W_1(\theta_1)|^{a+b-1}R^b}\right\}\,.$$

Phase 1 leaves $|W_1(\theta_1)|\ge R^{2n}$. Using $f(a-1,b)=2n(a-1)+(2n+1)b$,

$$\sup_{t\in[\theta_1,\tau_1]}|\Delta W_1(t)| \quad=\quad \frac{1}{|W_1(\theta_1)|^{a+b-1}R^b}\le\frac{1}{R^{2n(a+b-1)}R^b}=\frac{1}{R^{f(a-1,b)}}\,.$$

Thus, for $t\ge 1$, applying successively reflection coupling and Brownian scaling,

$$\begin{aligned}
\mathbb{P}\left[\tau_1-\theta_1>t/R^{2f(a-1,b)}\right] \quad&\le\quad \mathbb{P}\left[\sup_{0\le s-\theta_1\le t/R^{2f(a-1,b)}}|\Delta W_1(s)| \;\le\; \frac{1}{R^{f(a-1,b)}}\right]\\
&=\quad \mathbb{P}\left[\sup_{0\le s-\theta_1\le t/R^{2f(a-1,b)}}|W_1(s)-W_1(\theta_1)| \;\le\; \frac{1}{2R^{f(a-1,b)}}\right]\\
&=\quad \mathbb{P}\left[\sup_{s\in[0,t]}|W_1(s)| \;\le\; \frac{1}{2}\right] \quad\le\quad C_1 e^{-C_2 t}\,.
\end{aligned}\tag{3.16}$$

Consider the telescoping sum (for $s\ge 0$),

$$\begin{aligned}
&(W_1(s)^a-\widetilde{W}_1(s)^a)W_2(s)^b\\
&\qquad=\quad \Delta W_1(s)(W_1(s)^{a-1}+W_1(s)^{a-2}\widetilde{W}_1(s)+\cdots+\widetilde{W}_1(s)^{a-1})W_2(s)^b.
\end{aligned}\tag{3.17}$$

Since $W_2(\theta_1)=RW_1(\theta_1)$, we know

$$\sup_{s\in[\theta_1,\tau_1]}|\Delta W_1(s)|=\frac{1}{|W_1(\theta_1)|^{a+b-1}R^b}=\frac{1}{|W_1(\theta_1)|^{a-1}|W_2(\theta_1)|^b}.$$

So, for $1\le k\le a$,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\theta_1\le s\le\tau_1}|\Delta W_1(s)||W_1(s)|^{a-k}|\widetilde{W}_1(s)|^{k-1}|W_2(s)|^b>x\right] \quad\le\\
&\qquad\qquad \mathbb{P}\left[\sup_{\theta_1\le s\le\tau_1}\left|\frac{W_1(s)}{W_1(\theta_1)}\right|^{a-k}\left|\frac{\widetilde{W}_2(s)}{\widetilde{W}_1(\theta_1)}\right|^{k-1}\left|\frac{W_2(s)}{W_2(\theta_1)}\right|^b>x\right]\,.
\end{aligned}\tag{3.18}$$

As W_1 and $\widetilde{W}_1$ are reflection coupled in $[\theta_1,\tau_1]$, therefore

$$\sup_{\theta_1\le s\le\tau_1}|W_1(s)-W_1(\theta_1)| \;=\; \sup_{\theta_1\le s\le\tau_1}|\widetilde{W}_1(s)-\widetilde{W}_1(\theta_1)| \;=\; \frac{1}{2|W_1(\theta_1)|^{a+b-1}R^b} \;\le\; \frac{1}{2R^{f(a-1,b)}}\,.\tag{3.19}$$

Writing $W_1(s) = (W_1(s) - W_1(\theta_1)) + W_1(\theta_1)$ and using $R > 1$ and $|W_1(\theta_1)| \geq R^{2n}$ as well as (3.19),

$$\sup_{\theta_1 \leq s \leq \tau_1} \left|\frac{W_1(s)}{W_1(\theta_1)}\right| \quad \leq \quad 2\,, \qquad \sup_{\theta_1 \leq s \leq \tau_1} \left|\frac{\widetilde{W}_1(s)}{\widetilde{W}_1(\theta_1)}\right| \quad \leq \quad 2\,. \tag{3.20}$$

If $b = 0$ then the right-hand side of (3.18) simplifies, and for $x > 2^{a-1}$ it is immediate that

$$\mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |\Delta W_1(s)||W_1(s)|^{a-k}|\widetilde{W}_1(s)|^{k-1}|W_2(s)|^b > x\right] \quad \leq \\ \mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} \left|\frac{W_1(s)}{W_1(\theta_1)}\right|^{a-k} \left|\frac{\widetilde{W}_1(s)}{\widetilde{W}_1(\theta_1)}\right|^{k-1} > x\right] \quad = \quad 0\,. \tag{3.21}$$

If $b \geq 1$, we can use (3.18) with (3.20) to obtain

$$\begin{aligned} &\mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |\Delta W_1(s)||W_1(s)|^{a-k}|\widetilde{W}_1(s)|^{k-1}|W_2(s)|^b > x\right] \\ &\leq \quad \mathbb{P}\left[2^{a-1} \sup_{\theta_1 \leq s \leq \tau_1} \left|\frac{W_2(s)}{W_2(\theta_1)}\right|^b > x\right] \quad = \quad \mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |W_2(s)| > (x^{1/b}/2^{(a-1)/b})|W_2(\theta_1)|\right] \\ &\qquad \leq \quad \mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |W_2(s) - W_2(\theta_1)| > \left((x^{1/b}/2^{(a-1)/b}) - 1\right)|W_2(\theta_1)|\right]\,. \end{aligned}$$

Now introduce the requirement that $x \geq 2^{a+b-1}$, so that $(x^{1/b}/2^{(a-1)/b}) - 1 \geq (x/2^{a+b-1})^{1/b}$ for $x \geq 2^{a+b-1}$. Applying this together with $|W_2(\theta_1)| \geq R^{2n+1}$, and then applying a Markov inequality argument, followed by an application of the BDG inequality [17, p. 163] after conditioning on $\sigma\{(W_1(s), W_2(s)) : s \leq \theta_1\}$,

$$\begin{aligned} &\mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |\Delta W_1(s)||W_1(s)|^{a-k}|\widetilde{W}_1(s)|^{k-1}|W_2(s)|^b > x\right] \\ &\quad \leq \quad \mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |W_2(s) - W_2(\theta_1)| > (x/2^{a+b-1})^{1/b} R^{2n+1}\right] \\ &\quad \leq \quad \frac{2^{2(a+b-1)/b}\, \mathbb{E}\left[\sup_{\theta_1 \leq s \leq \tau_1} |W_2(s) - W_2(\theta_1)|\right]^2}{x^{2/b} R^{4n+2}} \\ &\quad \leq \quad \frac{2^{2(a+b-1)/b}\, \mathbb{E}\left[\tau_1 - \theta_1\right]}{x^{2/b} R^{4n+2}}\, C \qquad \text{when } x \geq 2^{a+b-1}\,. \end{aligned} \tag{3.22}$$

From (3.16), since $f(a-1, b) \geq 2n+1$ for $a, b \geq 1$,

$$\mathbb{E}\left[\tau_1 - \theta_1\right] \quad \leq \quad \frac{C}{R^{2f(a-1,b)}} \quad \leq \quad \frac{C}{R^{4n+2}}\,. \tag{3.23}$$

Using this estimate in (3.22), and using a new constant C, we obtain the following when $b \geq 1$, when $x \geq 2^{a+b-1}$,

$$\mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} |\Delta W_1(s)||W_1(s)|^{a-k}|\widetilde{W}_1(t)|^{k-1}|W_2(s)|^b > x\right] \quad \leq \quad \frac{2^{2(a+b-1)/b}}{R^{8n+4} x^{2/b}}\, C\,. \tag{3.24}$$

Note that (3.21) yields an upper bound of 0 when $b = 0$ (and $x > 2^{a-1}$). Using (3.24) in (3.17), for whatever b, and writing $x = 2^{a+b-1}M$ for future convenience of exposition, if $M > 1$ then

$$\mathbb{P}\left[\sup_{\theta_1 \leq s \leq \tau_1} \left|\frac{(W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^b}{a 2^{a+b-1}}\right| > M\right] \quad \leq \quad \begin{cases} \frac{a}{R^{8n+4} M^{2/b}}\, C & \text{if } b \geq 1\,, \\ 0 & \text{if } b = 0\,. \end{cases} \tag{3.25}$$

We now rewrite (3.25) and (3.16) to match the first assertion in part (i) of Lemma 5 (after conditioning on $\sigma\{(W_1(s), W_2(s)) : s \leq \theta_1\}$). For $s > \theta_1$, we set $t = s - \theta_1$, $B(t) = W_2(t + \theta_1) - W_2(\theta_1)$, $Y_t = \frac{(W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^b}{a2^{a+b-1}}$, $\tau = \tau_1 - \theta_1$, $\varepsilon = R^{-1}$. To match the indices in part (i) of Lemma 5, set $\alpha = 2/b, \beta = 2$ if $b \geq 1$, and choose any $\beta > 0$ if $b = 0$. Then (3.25) is equivalent to the following, holding when $M > 1$:

$$\mathbb{P}\left[\sup_{0\leq t\leq\tau} |Y_t| > M\right] \quad \leq \quad \begin{cases} \frac{a\varepsilon^{8n+4}}{M^\alpha}\, C & \text{if } b \geq 1\,, \\ 0 & \text{if } b = 0\,. \end{cases}$$

Note that $\varepsilon < 1$ (since $R > 1$), so the above implies the weaker inequality, if $M > 1$ then

$$\mathbb{P}\left[\sup_{0\leq t\leq\tau} |Y_t| > M\right] \quad \leq \quad aC\, M^{-\alpha}\,.$$

On the other hand (3.16) becomes

$$\mathbb{P}\left[\tau > t\varepsilon^{4n(a+b-1)+2b}\right] \quad \leq \quad C_1 e^{-C_2 t}\,.$$

Noting $e^{-C_2 t} \leq 1/(C_2 t)^2$ for $t > 0$, and then re-scaling time and using $n \geq 1, a + b > 1$, we obtain

$$\mathbb{P}\left[\tau > t\right] \quad \leq \quad C_1 e^{-C_2 t/\varepsilon^{4n(a+b-1)+2b}} \quad \leq \quad (C_1/C_2^2)\varepsilon^{8n(a+b-1)+4b}t^{-2} \quad \leq \quad (C_1/C_2^2)\varepsilon^2/t^2\,.$$

We can now apply the first assertion in part (i) of Lemma 5 to deduce the following. For $z > \varepsilon^{1/4}$,

$$\mathbb{P}\left[\left|\int_0^\tau Y_s \,\mathrm{d}B_s\right| \geq z\right] \quad \leq \quad \frac{C'\ \varepsilon^{1/(4(b\vee 2))}}{z^{1/(b\vee 2)}}\,.$$

Writing Y_s in full, this amounts to the following: when $x > a2^{a+b-1}R^{-1/4}$, and taking $\gamma' = 1/(b \vee 2)$,

$$\mathbb{P}\left[\left|\int_{\theta_1}^{\tau_1} (W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^b \,\mathrm{d}W_2(s)\right| > x\right] \quad \leq \quad \frac{C''}{(R^{1/4}x)^{\gamma'}}\,. \tag{3.26}$$

A similar procedure leads to a bound concerning $\int_{\theta_1}^{\tau_1} (W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^{b-1}\,\mathrm{d}s$. Here we need only argue for the case $b \geq 1$, as the time integral does not appear for $I_{(a,0)}$. Referring to (3.25), but using $b - 1$ instead of b, if $M > 1$ then we obtain

$$\mathbb{P}\left[\sup_{\theta_1\leq s\leq\tau_1} \left|\frac{(W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^{b-1}}{a2^{a+b-2}}\right| > M\right] \quad \leq \quad \begin{cases} \frac{a}{R^{8n+4}M^{2/(b-1)}}\, C & \text{if } b \geq 2\,, \\ 0 & \text{if } b = 1\,. \end{cases} \tag{3.27}$$

Choosing B, ε and τ as before, and again conditioning on $\sigma\{(W_1(s), W_2(s)) : s \leq \theta_1\}$, but now setting $Y_t = \frac{(W_1(t)^a - \widetilde{W}_1(t)^a)W_2(t)^{b-1}}{a2^{a+b-2}}$. To match the indices in part (i) of Lemma 5, set $\alpha = 2/(b-1)$ (for $b > 1$), and take any $\beta > \alpha$. When $M > 1$,

$$\mathbb{P}\left[\sup_{0\leq t\leq\tau} |Y_t| > M\right] \quad \leq \quad \begin{cases} \frac{a\varepsilon^{8n+4}}{M^\alpha}\, C & \text{if } b \geq 2\,, \\ 0 & \text{if } b = 1\,. \end{cases}$$

Applying the second assertion in part (i) of Lemma 5, and using $\gamma'' = \frac{1}{2}(1/(1 \vee (b-1)))$,

$$\mathbb{P}\left[\left|\int_{\theta_1}^{\tau_1} (W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^{b-1}\,\mathrm{d}s\right| > x\right] \quad \leq \quad \frac{C}{(R^{1/2}x)^{\gamma''}} \quad \text{for } x > a2^{a+b-2}R^{-1/2}. \tag{3.28}$$

Applying the inequalities (3.26) and (3.28) to the Itô representation of $I_{(a,b)}$ given in (3.13), we conclude that for any $a \geq 1$, $b \geq 0$, $a+b>1$, if $x > a2^{a+b-1}R^{-1/4}$ then

$$\mathbb{P}\left[|\Delta I_{(a,b)}(\tau_1) - \Delta I_{(a,b)}(\theta_1)| > x\right] \quad \leq \quad \frac{C}{(R^{1/4}x)^{\gamma' \wedge \gamma''}}\,. \tag{3.29}$$

Phase 3: Now, we address the time interval $[\tau_1, \eta_1]$. In this phase, starting at time τ_1, synchronous coupling is employed to the driving Brownian motions till $W_2(t+\tau_1) - W_2(\tau_1)$ hits the level $-a^{-1}\operatorname{sgn}(\Delta W_1(\tau_1))(\operatorname{sgn}(W_1(\theta_1)))^{a+b-1}$. Applying the reflection principle to $(W_1(t+\tau_1) - W_1(\tau_1) : t \geq 0)$, we can deduce the following estimate related to hitting times of Brownian motion:

$$\mathbb{P}\left[\eta_1 - \tau_1 > t\right] \quad \leq \quad Ct^{-1/2}\,. \tag{3.30}$$

Consider the fluctuations of $\Delta I_{(a,b)}$ on this interval. Using (3.17), it suffices to address the integrals

$$\int_{\tau_1}^{\eta_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^b \,\mathrm{d}W_2(s)$$

and (for $b \geq 1$ and $1 \leq k \leq a$)

$$\int_{\tau_1}^{\eta_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^{b-1} \,\mathrm{d}s\,.$$

As this is a synchronous coupling phase, $|\Delta W_1(t)| = |\Delta W_1(\tau_1)| = \frac{1}{|W_1(\theta_1)|^{a+b-1}R^b}$ for all $t \in [\tau_1, \eta_1]$. Observe that

$$\begin{aligned}\operatorname{sgn}\Delta W_1(\tau_1)(\operatorname{sgn} W_1(\theta_1))^{a+b-1}\Delta W_1(t) \quad &= \quad (\operatorname{sgn} W_1(\theta_1))^{a+b-1}|\Delta W_1(t)| \\ &= \quad (\operatorname{sgn} W_1(\theta_1))^{a+b-1}\frac{1}{|W_1(\theta_1)|^{a+b-1}R^b} \quad = \quad \frac{1}{W_1(\theta_1)^{a+b-1}R^b}\,.\end{aligned}$$

Combining this with the facts that $W_1(\theta_1) = \widetilde{W}_1(\theta_1)$ and $W_2(\theta_1) = RW_1(\theta_1)$, if $t \in [\tau_1, \eta_1]$ then

$$\begin{aligned}&\operatorname{sgn}\Delta W_1(\tau_1)(\operatorname{sgn} W_1(\theta_1))^{a+b-1} \times \Delta W_1(t) W_1(t)^{a-k}\widetilde{W}_1(t)^{k-1}W_2(t)^b \\ &\qquad = \quad \left(\frac{W_1(t)}{W_1(\theta_1)}\right)^{a-k}\left(\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right)^{k-1}\left(\frac{W_2(t)}{W_2(\theta_1)}\right)^{b}.\end{aligned} \tag{3.31}$$

Set $A_1(t) = \frac{W_1(t)}{W_1(\theta_1)}$, $\widetilde{A}_1(t) = \frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}$ and $A_2(t) = \frac{W_2(t)}{W_2(\theta_1)}$. Observe that for $1 \leq k \leq a$

$$\begin{aligned}&|A_1(t)^{a-k}\widetilde{A}_1(t)^{k-1}A_2(t)^b - 1| \quad \leq \\ &\qquad |A_1(t)^{a-k} - 1||\widetilde{A}_1(t)^{k-1}||A_2(t)^b| + |\widetilde{A}_1(t)^{k-1} - 1||A_2(t)^b| + |A_2(t)^b - 1|\,.\end{aligned} \tag{3.32}$$

We will show that the first term above is small with high probability. If $a = 1$, or more generally if $k = a$, then the first term is identically zero. If $a \geq 2$ and $k \leq a-1$ then

$$|A_1(t)^{a-k} - 1||\widetilde{A}_1(t)^{k-1}||A_2(t)^b| \quad \leq \quad \sum_{j=1}^{a-k}|A_1(t) - 1||A_1(t)|^{a-k-j}|\widetilde{A}_1(t)|^{k-1}|A_2(t)|^b\,. \tag{3.33}$$

Fix $x \geq 1/R^{2n}$. Recall that $|W_1(\theta_1)| \geq R^{2n}$, and note firstly that for $x \geq 1/R^{2n}$, by the reflection coupling implications summarized in (3.19),

$$\mathbb{P}\left[|W_1(\tau_1) - W_1(\theta_1)| > xR^{2n}/2\right] \quad \leq \quad \mathbb{P}\left[|W_1(\tau_1) - W_1(\theta_1)| > 1/2\right] = 0\,.$$

and secondly by a Tchebychev inequality argument and Doob's L^2-maximal inequality

$$\mathbb{P}\left[\sup_{\tau_1 \le t \le \tau_1 + T} |W_1(t) - W_1(\tau_1)| \quad > \quad xR^{2n}/2\right] \le \frac{CT}{x^2 R^{4n}}\,.$$

Then

$$\begin{aligned}
\mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |A_1(t) - 1| > x\right] \quad &\le \quad \mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |W_1(t) - W_1(\theta_1)| > xR^{2n}\right] \\
&\le \quad \mathbb{P}\left[\eta_1 - \tau_1 > T\right] + \mathbb{P}\left[\sup_{\tau_1 \le t \le \tau_1 + T} |W_1(t) - W_1(\theta_1)| > xR^{2n}\right] \\
&\le \quad \mathbb{P}\left[\eta_1 - \tau_1 > T\right] + \mathbb{P}\left[\sup_{\tau_1 \le t \le \tau_1 + T} |W_1(t) - W_1(\tau_1)| > xR^{2n}/2\right] \\
+ \, \mathbb{P}\left[|W_1(\tau_1) - W_1(\theta_1)| > xR^{2n}/2\right] \quad &\le \quad \frac{C}{\sqrt{T}} + \frac{CT}{x^2 R^{4n}} \quad \le \quad \frac{C}{(xR^{2n})^{2/3}}\,, \qquad (3.34)
\end{aligned}$$

where the last inequality follows by taking $T = (xR^{2n})^{4/3}$.

Similarly, for $x \ge 2$,

$$\begin{aligned}
\mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |A_1(t)| > x\right] \quad &= \quad \mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |W_1(t)| > x|W_1(\theta_1)|\right] \\
&\le \quad \mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |W_1(t) - W_1(\theta_1)| > x|W_1(\theta_1)|/2\right] \\
&\le \quad \mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |W_1(t) - W_1(\theta_1)| > xR^{2n}/2\right] \le \frac{C}{(xR^{2n})^{2/3}}\,, \qquad (3.35)
\end{aligned}$$

where the last inequality follows from the computations performed to obtain (3.34). A similar estimate for $\mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |\widetilde{A}_1(t)| > x\right]$ holds by replacing W_1 with $\widetilde{W}_1$ in the above calculations. To derive an analogous estimate for $\mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |A_2(t)| > x\right]$, first observe that

$$\mathbb{E}\left[|W_2(\tau_1) - W_2(\theta_1)|\right]^2 = \mathbb{E}\left[\tau_1 - \theta_1\right] \le C/R^{4n+2},$$

where the first equality is because conditional on $\sigma\{(W_1(s), W_2(s)) : s \le \theta_1\}$, $W_2 - W_2(\theta_1)$ is independent of $\tau_1 - \theta_1$ and the last inequality follows from (3.23). Using this observation along with the Tchebychev inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\tau_1 \le t \le \eta_1} |W_2(t) - W_2(\theta_1)| > \frac{xR^{2n+1}}{2}\right] \\
&\qquad \le \quad \mathbb{P}\left[\eta_1 - \tau_1 > T\right] + \mathbb{P}\left[\sup_{\tau_1 \le t \le \tau_1 + T} |W_2(t) - W_2(\theta_1)| > \frac{xR^{2n+1}}{2}\right] \\
&\qquad \le \quad \mathbb{P}\left[\eta_1 - \tau_1 > T\right] + \mathbb{P}\left[\sup_{\tau_1 \le t \le \tau_1 + T} |W_2(t) - W_2(\tau_1)| > \frac{xR^{2n+1}}{4}\right] \\
&\qquad\qquad + \mathbb{P}\left[|W_2(\tau_1) - W_2(\theta_1)| > \frac{xR^{2n+1}}{4}\right] \\
&\qquad\qquad \le \quad \frac{C}{\sqrt{T}} + \frac{CT}{x^2 R^{4n+2}} + \frac{C}{x^2 R^{8n+4}} \quad \le \quad \frac{C}{(xR^{2n})^{2/3}}\,, \qquad (3.36)
\end{aligned}$$

where the last inequality follows by taking $T = (xR^{2n})^{4/3}$. Using (3.36) and recalling

$|W_2(\theta_1)| \geq R^{2n+1}$,

$$\begin{aligned}\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_2(t)| > x\right] &\leq \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |W_2(t) - W_2(\theta_1)| > x|W_2(\theta_1)|/2\right] \\ &\leq \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |W_2(t) - W_2(\theta_1)| > xR^{2n+1}/2\right] \leq \frac{C}{(xR^{2n})^{2/3}}. \quad (3.37)\end{aligned}$$

From the above estimates, we can argue the following in case $x \geq 2^{2(a+b-1)}/R^{2n}$:

$$\begin{aligned}&\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t) - 1||A_1(t)|^{a-k-j}|\widetilde{A}_1(t)|^{k-1}|A_2(t)|^b > x\right] \\ &\quad\leq \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t) - 1| > \frac{\sqrt{x}}{\sqrt{R^{2n}}}\right] + \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t)| > (\sqrt{R^{2n}x})^{\frac{1}{a+b-1-j}}\right] \\ &\quad + \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |\widetilde{A}_1(t)| > (\sqrt{R^{2n}x})^{\frac{1}{a+b-1-j}}\right] + \mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_2(t)| > (\sqrt{R^{2n}x})^{\frac{1}{a+b-1-j}}\right] \\ &\quad\leq C\frac{1}{(R^{2n}x)^{\gamma}}\end{aligned}$$

for some $\gamma > 0$ (in fact $\gamma = 1/3$) that does not depend on R (the last three probabilities appearing after the first inequality above can be taken to be zero if $a + b - 1 - j = 0$).

By applying the above argument to each term on the right hand side of (3.33), we obtain

$$\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t)^{a-k} - 1||\widetilde{A}_1(t)^{k-1}||A_2(t)^b| > x\right] \leq \frac{C}{(R^{2n}x)^{\gamma}} \quad \text{for } x \geq (a-k)2^{2(a+b-1)}/R^{2n}.$$

The terms $|\widetilde{A}_1(t)^{k-1}-1||A_2(t)^b|$ and $|A_2(t)^b-1|$ appearing in (3.32) are subject to estimates of the same form, based on $\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |\widetilde{A}_1(t) - 1| > x\right]$ and $\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_2(t) - 1| > x\right]$ respectively in place of $\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t) - 1| > x\right]$, but otherwise using the same arguments. Hence (3.32) and the above estimates yield the following for $x \geq (a + b - 1)2^{2(a+b-1)}/R^{2n}$:

$$\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |A_1(t)^{a-k}\widetilde{A}_1(t)^{k-1}A_2(t)^b - 1| > x\right] \leq \frac{C}{(R^{2n}x)^{\gamma}}. \quad (3.38)$$

Thus (3.31) yields (when $x \geq (a + b - 1)2^{2(a+b-1)}/R^{2n}$)

$$\begin{aligned}&\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} \left|\Delta W_1(t)W_1(t)^{a-k}\widetilde{W}_1(t)^{k-1}W_2(t)^b - \operatorname{sgn}\Delta W_1(\tau_1)(\operatorname{sgn} W_1(\theta_1))^{a+b-1}\right| > x\right] \\ &\qquad\leq \frac{C}{(R^{2n}x)^{\gamma}}.\end{aligned}$$

The above holds for all $1 \leq k \leq a$; consequently (3.17) implies that, for $x \geq a(a + b - 1)2^{2(a+b-1)}/R^{2n}$,

$$\begin{aligned}&\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} \left|(W_1(t)^a - \widetilde{W}_1(t)^a)W_2(t)^b - a\operatorname{sgn}\Delta W_1(\tau_1)(\operatorname{sgn} W_1(\theta_1))^{a+b-1}\right| > x\right] \\ &\qquad\leq \frac{C}{(R^{2n}x)^{\gamma}}. \quad (3.39)\end{aligned}$$

Now $\mathbb{P}\left[\eta_1 - \tau_1 > t\right] \leq Ct^{-1/2}$; so the first assertion in part (ii) of Lemma 5 implies there is $\gamma' > 0$, not depending on R, such that for $x \geq a(a+b-1)2^{2(a+b-1)}/R^{2n}$

$$\begin{aligned}\mathbb{P}\Big[\Big|\int_{\tau_1}^{\eta_1} (W_1(s)^a - \widetilde{W}_1(s)^a) W_2(s)^b \,\mathrm{d}W_2(s) \\ - a \operatorname{sgn} \Delta W_1(\tau_1)(\operatorname{sgn} W_1(\theta_1))^{a+b-1}(W_2(\eta_1) - W_2(\tau_1))\Big| > x\Big] \\ \leq \quad \frac{C}{(R^{2n}x)^{\gamma'}}\,.\end{aligned}$$

But it follows from the definition of η_1 that

$$W_2(\eta_1) - W_2(\tau_1) = -a^{-1} \operatorname{sgn}(\Delta W_1(\tau_1))(\operatorname{sgn}(W_1(\theta_1)))^{a+b-1}\,.$$

Together with the above inequality this yields, for $x \geq a(a+b-1)2^{2(a+b-1)}/R^{2n}$,

$$\mathbb{P}\left[\left|\int_{\tau_1}^{\eta_1} (W_1(s)^a - \widetilde{W}_1(s)^a) W_2(s)^b \,\mathrm{d}W_2(s) + 1\right| > x\right] \quad \leq \quad \frac{C}{(R^{2n}x)^{\gamma}}\,. \tag{3.40}$$

To estimate the integral $\int_{\tau_1}^{\eta_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^{b-1}\,\mathrm{d}s$ for $b \geq 1$, we can once more use the synchronous coupling of W_1, $\widetilde{W}_1$ on $[\tau_1, \eta_1]$ to show that for any $t \in [\tau_1, \eta_1]$,

$$|\Delta W_1(t) W_1(t)^{a-k} \widetilde{W}_1(t)^{k-1} W_2(t)^{b-1}| \quad \leq \quad \frac{1}{R^{2n+1}} \left|\frac{W_1(t)}{W_1(\theta_1)}\right|^{a-k} \left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right|^{k-1} \left|\frac{W_2(t)}{W_2(\theta_1)}\right|^{b-1}.$$

For $b \geq 1$ we may use (3.38) to show, for $x \geq 2^{2(a+b-1)}$,

$$\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} \left|\frac{W_1(t)}{W_1(\theta_1)}\right|^{a-k} \left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right|^{k-1} \left|\frac{W_2(t)}{W_2(\theta_1)}\right|^{b-1} > x\right] \quad \leq \quad \frac{C}{(R^{2n}x)^{\gamma}}\,.$$

Thus, for $x \geq 2^{2(a+b-1)}/R^{2n+1}$,

$$\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |\Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^{b-1}| > x\right] \quad \leq \quad \frac{C}{(R^{4n+1}x)^{\gamma}}\,.$$

Using the above and the fact that $\mathbb{P}\left[\eta_1 - \tau_1 > t\right] \leq Ct^{-1/2}$ in the second assertion in part (ii) of Lemma 5, we obtain $\gamma, \delta > 0$ not depending on R such that

$$\mathbb{P}\left[\left|\int_{\tau_1}^{\eta_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^{b-1}\,\mathrm{d}s\right| > x\right] \leq \frac{C}{(R^{\delta}x)^{\gamma}}$$

for $x \geq (a+b-1)2^{2(a+b-1)}R^{-\delta}$. But then we can use (3.17) to deduce

$$\mathbb{P}\left[\left|\int_{\tau_1}^{\eta_1} (W_1(s)^a - \widetilde{W}_1(s)^a) W_2(s)^{b-1}\,\mathrm{d}s\right| > x\right] \quad \leq \quad \frac{C}{(R^{\delta}x)^{\gamma}} \tag{3.41}$$

for $x \geq a(a+b-1)2^{2(a+b-1)}R^{-\delta}$. Recalling the expression of $I_{(a,b)}$ in terms of the Itô integral and the time integral given in (3.13), we obtain from (3.40) and (3.41),

$$\mathbb{P}\left[\left|\Delta I_{(a,b)}(\eta_1) - \Delta I_{(a,b)}(\tau_1) + 1\right| > x\right] \quad \leq \quad \frac{C}{(R^{\delta}x)^{\gamma}} \tag{3.42}$$

for $x \geq a(a+b-1)2^{2(a+b-1)}R^{-\delta}$.

Phase 4: The next phase occurs in the time interval $[\eta_1, \lambda_1]$. In this phase, after time η_1 the Brownian motions W_1 and $\widetilde{W}_1$ are subjected to reflection coupling till they meet. Applying the reflection principle, and using the fact that $|\Delta W(\eta_1)| = 1/(|W(\theta_1)|^{a+b-1}R^b)$ together with other consequences of the definitions of the stopping times θ_1 and τ_1, we see that when $t > 0$

$$\mathbb{P}\left[\lambda_1 - \eta_1 > t/R^{2f(a-1,b)}\right] \quad \leq \quad Ct^{-1/2}\,. \tag{3.43}$$

Once again (3.17) can be applied, so it suffices to consider the integrals

$$\int_{\eta_1}^{\lambda_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^b \,\mathrm{d}W_2(s)$$

and

$$\int_{\eta_1}^{\lambda_1} \Delta W_1(s) W_1(s)^{a-k} \widetilde{W}_1(s)^{k-1} W_2(s)^{b-1} \,\mathrm{d}s$$

for $1 \leq k \leq a$.

For $\eta_1 \leq t \leq \lambda_1$ we can write

$$\begin{aligned} |\Delta W_1(t)||W_1(t)|^{a-k}|\widetilde{W}_1(t)|^{k-1}|W_2(t)|^b \quad &= \\ \left(|\Delta W_1(t)||W_1(\theta_1)|^{a+b-1}R^b\right)&\left|\frac{W_1(t)}{W_1(\theta_1)}\right|^{a-k}\left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right|^{k-1}\left|\frac{W_2(t)}{W_2(\theta_1)}\right|^b . \end{aligned} \tag{3.44}$$

Recalling that $|\Delta W_1(\eta_1)| = |\Delta W_1(\tau_1)| = \frac{1}{|W_1(\theta_1)|^{a+b-1}R^b}$, and bearing in mind that W_1 and $\widetilde{W}_1$ are reflection coupled on $[\eta_1, \lambda_1]$, when $x \geq 1$ it follows that

$$\begin{aligned} &\mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} |\Delta W_1(t)||W_1(\theta_1)|^{a+b-1}R^b \geq x\right] \\ &\quad = \quad \mathbb{P}\left[\text{Brownian motion starting from } \frac{1}{2} \text{ hits } \frac{x}{2} \text{ before zero}\right] \quad = \quad \frac{1}{x}\,, \end{aligned} \tag{3.45}$$

where the last equality follows from the optional stopping theorem. Fixing $x \geq 2$, we can employ (3.43) and a Tchebychev inequality argument to show

$$\begin{aligned} \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1}\left|\frac{W_1(t)}{W_1(\theta_1)}\right| > x\right] \quad &\leq \quad \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} |W_1(t) - W_1(\theta_1)| > xR^{2n}/2\right] \\ \leq \quad & \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} |W_1(t) - W_1(\eta_1)| > xR^{2n}/4\right] + \mathbb{P}\left[|W_1(\eta_1) - W_1(\theta_1)| > xR^{2n}/4\right] \\ \leq \quad & \mathbb{P}\left[\lambda_1 - \eta_1 > T\right] + \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \eta_1 + T} |W_1(t) - W_1(\eta_1)| > xR^{2n}/4\right] \\ & + \mathbb{P}\left[|W_1(\eta_1) - W_1(\theta_1)| > xR^{2n}/4\right] \\ \leq \quad & \frac{C}{R^{f(a-1,b)}\sqrt{T}} + \frac{CT}{x^2 R^{4n}} + \frac{C}{(xR^{2n})^{2/3}}. \end{aligned}$$

The bound $\mathbb{P}\left[|W_1(\eta_1) - W_1(\theta_1)| > xR^{2n}/4\right] \leq \frac{C}{(xR^{2n})^{2/3}}$ follows from the calculations leading to (3.34), where, in fact, we obtained the following bound when $x \geq \frac{1}{R^{2n}}$:

$$\mathbb{P}\left[\sup_{\tau_1 \leq t \leq \eta_1} |W_1(t) - W_1(\theta_1)| > xR^{2n}\right] \quad \leq \quad \frac{C}{(xR^{2n})^{2/3}}\,.$$

Taking $T = (xR^{2n})^{4/3}$, we obtain the following when $x \geq 2$:

$$\mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} \left|\frac{W_1(t)}{W_1(\theta_1)}\right| > x\right] \quad \leq \quad \frac{C}{(xR^{2n})^{2/3}}\,. \tag{3.46}$$

Similar estimates for $\mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} \left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right| > x\right]$ and $\mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} \left|\frac{W_2(t)}{W_2(\theta_1)}\right| > x\right]$ follow by replacing W_1 by $\widetilde{W}_1$ and W_2 respectively in the above calculations (in the latter case, we use (3.36)). Using these estimates along with (3.46) and (3.44), we obtain for $x \geq 2^{a+b}$,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} |\Delta W_1(t)||W_1(t)|^{a-k}|\widetilde{W}_1(t)|^{k-1}|W_2(t)|^b > x\right] \\
&\quad = \quad \mathbb{P}\left[\left(|\Delta W_1(t)||W_1(\theta_1)|^{a+b-1}R^b\right)\left|\frac{W_1(t)}{W_1(\theta_1)}\right|^{a-k}\left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right|^{k-1}\left|\frac{W_2(t)}{W_2(\theta_1)}\right|^{b} > x\right] \\
&\quad \leq \quad \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1} |\Delta W_1(t)||W_1(\theta_1)|^{a+b-1}R^b \geq x^{\frac{1}{a+b}}\right] + \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1}\left|\frac{W_1(t)}{W_1(\theta_1)}\right| > x^{\frac{1}{a+b}}\right] \\
&\qquad\qquad + \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1}\left|\frac{\widetilde{W}_1(t)}{\widetilde{W}_1(\theta_1)}\right| > x^{\frac{1}{a+b}}\right] + \mathbb{P}\left[\sup_{\eta_1 \leq t \leq \lambda_1}\left|\frac{W_2(t)}{W_2(\theta_1)}\right| > x^{\frac{1}{a+b}}\right] \\
&\qquad\qquad\qquad \leq \quad \frac{1}{x^{\frac{1}{a+b}}} + \frac{C}{(x^{\frac{1}{a+b}}R^{2n})^{2/3}} \quad \leq \quad Cx^{-\frac{2}{3(a+b)}}\,,
\end{aligned} \tag{3.47}$$

where the last step follows as $R > 1$. From (3.43), (3.47) and the first assertion in part (i) of Lemma 5, it follows that there are $\delta, \gamma > 0$ not depending on R such that

$$\mathbb{P}\left[\left|\int_{\eta_1}^{\lambda_1} (W_1(s)^a - \widetilde{W}_1(s)^a)W_2(s)^b \,\mathrm{d}W_2(s)\right| > x\right] \quad \leq \quad \frac{C}{(R^\delta x)^\gamma}$$

for $x \geq 2^{a+b}/R^\delta$. Arguing as above, using (3.43), and (3.47) but with $b-1$ replacing b, and appealing to the second assertion in part (i) of Lemma 5, if $b \geq 1$ then

$$\mathbb{P}\left[\left|\int_{\eta_1}^{\lambda_1} \Delta W_1(s)W_1(s)^{a-k}\widetilde{W}_1(s)^{k-1}W_2(s)^{b-1}\,\mathrm{d}s\right| > x\right] \quad \leq \quad \frac{C}{(R^\delta x)^\gamma}$$

for $x \geq 2^{a+b-1}/R^\delta$. From the above two bounds, if $x \geq 2^{a+b}/R^\delta$ then

$$\mathbb{P}\left[|\Delta I_{(a,b)}(\lambda_1) - \Delta I_{(a,b)}(\eta_1)| > x\right] \quad \leq \quad \frac{C}{(R^\delta x)^\gamma}\,. \tag{3.48}$$

Phase 5: The final phase concerns the interval $[\lambda_1, \beta_1]$, in which the Brownian motions (W_1, W_2) and $(\widetilde{W}_1, \widetilde{W}_2)$ are coupled synchronously till the time β_1 when $(W_1, W_2) = \widetilde{W}_1, \widetilde{W}_2)$ hits the line $u_2 = Ru_1$. Since $(W_1, W_2) = (\widetilde{W}_1, \widetilde{W}_2)$ during this time interval, $\Delta I_{(a,b)}(\beta_1) = \Delta I_{(a,b)}(\lambda_1)$. We claim there is a positive constant C not depending on R such that

$$\mathbb{P}\left[\beta_1 - \lambda_1 > t\right] \quad \leq \quad Ct^{-1/6}\,. \tag{3.49}$$

To see this, observe that $\beta_1 - \lambda_1$ depends on how far away the Brownian motion (W_1, W_2) is from the line $u_2 = Ru_1$ at time λ_1. As $W_2(\theta_1) = RW_1(\theta_1)$, this distance, in turn, depends on the size of the total duration $\lambda_1 - \theta_1$ of the previous three phases. Indeed, for any $\alpha, x > 0$ (to be chosen later),

$$\begin{aligned}
\mathbb{P}\left[\beta_1 - \lambda_1 > t\right] &\leq \mathbb{P}\left[\beta_1 - \lambda_1 > t,\ \lambda_1 - \theta_1 \leq t^\alpha\right] + \mathbb{P}\left[\lambda_1 - \theta_1 > t^\alpha\right] \\
&\leq \mathbb{P}\left[|(W_1, W_2)(\lambda_1) - (W_1, W_2)(\theta_1)| > x,\ \lambda_1 - \theta_1 \leq t^\alpha\right] \\
&\quad + \mathbb{P}\left[|(W_1, W_2)(\lambda_1) - (W_1, W_2)(\theta_1)| \leq x,\ \beta_1 - \lambda_1 > t\right] \\
&\quad + \mathbb{P}\left[\lambda_1 - \theta_1 > t^\alpha\right].
\end{aligned}$$

To estimate the first probability above, note that an application of the strong Markov property at time θ_1 allows us to deduce

$$\begin{aligned}
&\mathbb{P}\left[|(W_1, W_2)(\lambda_1) - (W_1, W_2)(\theta_1)| > x,\ \lambda_1 - \theta_1 \leq t^\alpha\right] \\
&\qquad \leq \mathbb{P}\left[\sup_{s \in [\theta_1, \theta_1 + t^\alpha]} |(W_1, W_2)(s) - (W_1, W_2)(\theta_1)| > x\right] \leq C\frac{t^{\alpha/2}}{x},
\end{aligned}$$

where the last inequality follows from Doob's submartingale inequality applied to the radial part of two-dimensional Brownian motion. The second probability is controlled by conditioning on the past event $[|(W_1, W_2)(\lambda_1) - (W_1, W_2)(\theta_1)| \leq x]$ and using the strong Markov property to argue that the hitting time on the line $u_2 = Ru_1$ by the Brownian motion $((W_1, W_2)(t) - (W_1, W_2)(\lambda_1) : t \geq \lambda_1)$ is stochastically dominated by the hitting time on zero by a one dimensional Brownian motion starting from x. Therefore,

$$\mathbb{P}\left[|(W_1, W_2)(\lambda_1) - (W_1, W_2)(\theta_1)| \leq x,\ \beta_1 - \lambda_1 > t\right] \leq C\frac{x}{\sqrt{t}}.$$

From (3.16), (3.30) and (3.43), we deduce that

$$\mathbb{P}\left[\lambda_1 - \theta_1 > t^\alpha\right] \leq Ct^{-\alpha/2}. \tag{3.50}$$

Putting these bounds together, it follows that

$$\mathbb{P}\left[\beta_1 - \lambda_1 > t\right] \leq C\frac{t^{\alpha/2}}{x} + C\frac{x}{\sqrt{t}} + Ct^{-\alpha/2}.$$

The target inequality (3.49) is obtained by taking $\alpha = 1/3$ and $x = t^{1/3}$ in the above bound.

From (3.29), (3.42) and (3.48), we see that there exist positive constants C_1, C_2, δ, γ not depending on R, w, i such that

$$\mathbb{P}\left[\left|\Delta I_{(a,b)}(\beta_1)\right| > x\right] \leq \frac{C_1}{(R^\delta x)^\gamma} \qquad \text{for } x \geq C_2/R^\delta. \tag{3.51}$$

B: Describing subsequent cycles and successful coupling

The above account gives a description of the five phases that constitute the first cycle. Subsequent cycles are defined similarly as follows:

For $t \geq \beta_1$, we apply scaling using Lemma 4 with $r = \left|\Delta I_{(a,b)}(\beta_1)\right|^{-1/(a+b+1)}$ and define further stopping times $\theta_2, \ldots, \beta_2$ corresponding to $\theta_1, \ldots, \beta_1$ for the scaled process, and continue in this fashion to obtain successive cycles. As in the proof of Lemma 7, in order to show that constructing these cycles leads to a successful coupling we need to show that $\lim_{k\to\infty} \Delta I_{(a,b)}(\beta_k) = 0$, and $\lim_{k\to\infty} \beta_k < \infty$ almost surely. This would imply that the end points of these cycles have an accumulation point and thus that the coupling is successful in finite time. We now demonstrate that these facts follow from the estimates obtained above, *via* Lemma 6.

For $k \geq 1$, if $|\Delta I_{(a,b)}(\beta_{k-1})| = 0$, then the coupling is successful. If the coupling is not successful, define $X_k = \dfrac{R^\delta |\Delta I_{(a,b)}(\beta_k)|}{|\Delta I_{(a,b)}(\beta_{k-1})|}$, where δ is as used in (3.51) and we adopt the convention that $\beta_0 = 0$. Taking $\tau_k = 1$ for $k \geq 1$, we see that X_k, τ_k satisfy the hypotheses of Lemma 6, and thus we obtain $R_0' > 1$ such that for all $R \geq R_0'$,

$$\sum_{k=1}^{\infty} R^{-k\delta} \left(\Pi_{j=1}^k X_j\right) \quad < \quad \infty$$

almost surely. In particular this implies that almost surely

$$\lim_{k\to\infty} R^{-k\delta} \left(\Pi_{j=1}^k X_j\right) \quad = \quad \lim_{k\to\infty} |\Delta I_{(a,b)}(\beta_k)| \quad = \quad 0\,.$$

Choose and fix any $R \geq R_0'$. From (3.14), (3.16), (3.30), (3.43) and (3.49), we have $\alpha > 0$ such that

$$\mathbb{P}\left[\beta_1 > R^{4n+2} t\right] \quad \leq \quad C t^{-\alpha} \tag{3.52}$$

for $t \geq 1$. Write $\mathcal{F}_k = \sigma\{(W_1(s), W_2(s)) \,:\, s \leq \beta_k\}$. Define $\tau_k^* = \dfrac{|\beta_k - \beta_{k-1}|}{\left|\Delta I_{(a,b)}(\beta_{k-1})\right|^{2/(a+b+1)}}$ for $k \geq 1$. By (3.52), τ_k^* satisfies

$$\mathbb{P}\left[\tau_{k+1}^* > R^{4n+2} t \mid \mathcal{F}_k\right] \quad \leq \quad C t^{-\alpha}\,.$$

for $t \geq 1$. Define $X_k^* = \dfrac{R^{2\delta/(a+b+1)} |\Delta I_{(a,b)}(\beta_k)|^{2/(a+b+1)}}{|\Delta I_{(a,b)}(\beta_{k-1})|^{2/(a+b+1)}}$ for $k \geq 1$, where δ is the same as that used in (3.51). By (3.51), observe that

$$\mathbb{P}\left[X_{k+1}^* > x \mid \mathcal{F}_k\right] \quad \leq \quad C x^{-\gamma(a+b+1)/2} \qquad \text{for } x \geq 1.$$

The following holds:

$$\beta_{k+1} \quad = \quad \tau_1^* + \sum_{l=1}^{k} R^{-2l\delta/(a+b+1)} \left(\Pi_{j=1}^l X_j^*\right) \tau_{l+1}^*\,.$$

Thus, for any $\gamma' < \alpha \wedge \frac{\gamma(a+b+1)}{2}$, using Lemma 6 with $(X_i^*, \tau_i^*/R^{4n+2})$ in place of (X_i, τ_i), we obtain $R_0'' \geq R_0'$ such that for every $R \geq R_0''$,

$$\mathbb{P}\left[\tau_1^* + \sum_{l=1}^{\infty} R^{-2l\delta/(a+b+1)} \left(\Pi_{j=1}^l X_j^*\right) \tau_{l+1}^* > R^{4n+2} t\right] \quad \leq \quad C t^{-\gamma'}\,.$$

This shows that the coupling construction represented by $\mathbb{P}_R$ yields an almost surely successful coupling with coupling time given by $T_{R,(a,b)} = \lim_{k\to\infty} \beta_k$. R_0 claimed in the theorem can be taken to be R_0''.

From the coupling construction, we see that the *active region* $S_{R,(a,b)}$ referred to in the theorem can be written as

$$S_{R,(a,b)} \quad = \quad \bigcup_{k=1}^{\infty} [\theta_k, \lambda_k]\,.$$

The estimate on the tail probabilities of $|S_{R,(a,b)}|$, claimed in the statement of the lemma, follows from Lemma 6 using an argument similar to that given above, after re-scaling by considering $\frac{|\lambda_k - \theta_k|}{\left|\Delta I_{(a,b)}(\beta_{k-1})\right|^{2/(a+b+1)}}$ for τ_k^* (in fact, it follows from (3.50) that the tail estimate holds for any $\gamma < 1/2$).

Assertion (ii) claimed in the lemma follows first from observing that

$$\sup_{t\leq T_{R,(a,b)}}|\Delta W_1(t)| \quad\leq\quad \sum_{l=1}^{\infty}\sup_{t\in[\theta_l,\lambda_l]}|\Delta W_1(t)|$$

and then from applying Lemma 6 with $(\sqrt{X_k^*}, M_k^*)$ in place of (X_k, τ_k), where

$$M_k^* \quad=\quad R^{f(a-1,b)}\frac{\sup_{t\in[\theta_k,\lambda_k]}|\Delta W_1(t)|}{|\Delta I_{(a,b)}(\beta_{k-1})|^{1/(a+b+1)}}\,.$$

The tail estimates for M_k^* needed to apply Lemma 6 are derived by recalling $|W_1(\theta_1)|\geq R^{2n}$ and applying scaling to deduce for $x\geq 1$

$$\begin{aligned}&\mathbb{P}\left[M_{k+1}^*>x\mid\mathcal{F}_k\right]\leq\mathbb{P}\left[R^{f(a-1,b)}\sup_{t\in[\theta_1,\lambda_1]}|\Delta W_1(t)|>x\right]\\&=\mathbb{P}\left[R^{f(a-1,b)}\sup_{t\in[\eta_1,\lambda_1]}|\Delta W_1(t)|>x\right]\leq\mathbb{P}\left[\sup_{t\in[\eta_1,\lambda_1]}|\Delta W_1(t)||W_1(\theta_1)|^{a+b-1}R^b>x\right]=x^{-1},\end{aligned}$$

where the last step follows from (3.45). □

4 Simultaneously coupling multiple monomial Stratonovich integrals

This section describes the construction of a successful coupling based on a driving 2-dimensional Brownian motion (W_1, W_2) and the complete finite set of monomial stochastic integrals up to a given scaling degree n, given by $(I_{(a,b)} : a\geq 1,\ b\geq 0,\ a+b\leq n)$. The construction uses an inductive strategy; coupling first at the level of monomial stochastic integrals $I_{(k,l)}$ for all $(k,l)\prec(a,b)$ and then coupling $I_{(a,b)}$ while ensuring that the lower order integrals do not deviate too far from coupled agreement.

Recall

$$\mathbf{X}_{(a,b)} \quad=\quad \left(W_1, I_{(c,d)}; (c,d)\preceq(a,b),\ c\geq 1\right)\,.$$

We will abbreviate the complete set of monomial stochastic integrals (up to $I_{(0,n)}$) as

$$\mathbf{X}(t) \quad=\quad \mathbf{X}_{(0,n)}(t),\ \ t\geq 0\,.$$

$\widetilde{\mathbf{X}}_{(a,b)}$ and $\widetilde{\mathbf{X}}$ are defined in a similar manner.

The main theorem of this article states the existence of this successful coupling and estimates the rate at which it happens. In the following, we will need a simple norm on quantities such as $\mathbf{X}$; we use Euclidean norm viewing $\mathbf{X}$ as a vector in the Euclidean space of appropriate dimension.

Theorem 10. For any pair of starting points $\mathbf{X}(0)$ and $\widetilde{\mathbf{X}}(0)$ there exists a successful Markovian coupling construction $\mathbb{P}$ of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$, with coupling time T satisfying the following rate estimate:
There are positive constants C, γ such that if $t\geq 1$ then

$$\sup\{\mathbb{P}\left[T>t\right]\ :\ |\mathbf{X}(0)|\leq 1\,,\ |\widetilde{\mathbf{X}}(0)|\leq 1\} \quad\leq\quad Ct^{-\gamma}\,.$$

Proof. As before, C,γ will denote generic positive constants whose values will change from line to line. The constant $R>1$ is a tuning parameter for the coupling construction: its value will be specified later.

By a combination of reflection coupling and then synchronous coupling, we may assume that the starting points satisfy $(W_1, W_2)(0) = (\widetilde{W}_1, \widetilde{W}_2)(0)$ and $W_2(0) = RW_1(0)$. We will write this as $(\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}$ where

$$\mathcal{R} \;=\; \{(w, \widetilde{w}) : (w_1, w_2) = (\widetilde{w}_1, \widetilde{w}_2), w_2 = Rw_1, \text{ remaining coordinates of } w, \widetilde{w} \text{ unconstrained}\}\;.$$

At the end of the proof we will check that the rate of coupling is not affected by the time taken to arrange for this.

The main body of the proof is based on induction on the number of $\preceq$-ordered monomial Stratonovich integrals to be coupled.
Induction hypothesis: Define $\Delta\mathbf{X}_{(a,b)} = \mathbf{X}_{(a,b)} - \widetilde{\mathbf{X}}_{(a,b)}$. For any $(a, b) \in \Delta_n$, there exists a successful Markovian coupling between the arrays of monomial Stratonovich integrals $\mathbf{X}_{(a,b)}$ and $\widetilde{\mathbf{X}}_{(a,b)}$, and between W and $\widetilde{W}$, with coupling time $T_{(a,b)}$ such that for all $t \geq 1$

$$\sup\{\mathbb{P}\left[T_{(a,b)} > t\right] \;:\; |\Delta\mathbf{X}_{(a,b)}(0)| \leq 1,\; (\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}\} \quad\leq\quad Ct^{-\gamma}\,.$$

for positive constants C, γ.

By Lemma 4 there is no loss of generality in assuming that (a) the starting points $\mathbf{X}_{(a,b)}(0)$ and $\widetilde{\mathbf{X}}_{(a,b)}(0)$ satisfy $(\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}$ and (b) $|\Delta\mathbf{X}_{(a,b)}(0)| = 1$.

Lemma 7 establishes the inductive hypothesis in the initial case of $(a, b) = (1, 0)$, since then $(W_1, W_2)(0) = (\widetilde{W}_1, \widetilde{W}_2)(0)$ and $|\Delta I_{(1,0)}(0)| \leq |\Delta\mathbf{X}_{(a,b)}(0)| = 1$.

Consider $(a, b) \in \Delta_n$ such that $(1, 0) \prec (a, b)$ (equivalently, $a + b > 1$). Let (a^-, b^-) be the $\preceq$ predecessor of (a, b). The inductive step of the proof is as follows: suppose the induction hypothesis is true for (a^-, b^-); then it is required to show that the hypothesis is also true for (a, b). The key to this is to conduct a careful analysis of the cycles described informally above. By scaling arguments, it is sufficient to do this for the first cycle, and then to show how scaling arguments can be used to establish suitable convergence over the whole sequence of cycles.

If $a = 0$, then from the definition of $\mathbf{X}_{(a,b)}$, $\mathbf{X}_{(a,b)} = \mathbf{X}_{(a^-,b^-)}$ (as remarked in Section 2) and there is nothing to prove. Therefore, we assume $a \geq 1$.

A: Description of the first cycle
We can write $\mathbf{X}_{(a,b)} = (\mathbf{X}_{(a^-,b^-)}, I_{(a,b)})$. The three phases of the first cycle have end-points given by the following stopping times.

1: $\sigma_1 = \inf\{t \geq 0 : \Delta\mathbf{X}_{(a^-,b^-)}(t) = 0\}$,	Coupling of $\mathbf{X}_{(a^-,b^-)}$ and $\widetilde{\mathbf{X}}_{(a^-,b^-)}$ derived from inductive hypothesis, and note $\mathbf{X}_{(a^-,b^-)}(\sigma_1) = \widetilde{\mathbf{X}}_{(a^-,b^-)}(\sigma_1)$;
2: $\sigma_2 = \inf\{t \geq \sigma_1 :\; RW_1(t) = W_2(t)\}$	synchronous coupling till $RW_1 = W_2$;
3: $\sigma_3 > \sigma_2$	Coupling strategy of Lemma 9 after re-scaling $\mathbf{X}_{(a,b)}$, $\widetilde{\mathbf{X}}_{(a,b)}$ using $\lvert\Delta I_{(a,b)}(\sigma_1)\rvert^{-1/(a+b+1)}$ and note $(W_1, W_2)(\sigma_3) = (\widetilde{W}_1, \widetilde{W}_2)(\sigma_3)$, $I_{(a,b)}(\sigma_3) = \widetilde{I}_{(a,b)}(\sigma_3)$.

Between σ_1 and σ_2, the two Brownian motions are coalesced and synchronously coupled. Therefore $\Delta\mathbf{X}_{(a,b)}(\sigma_2) = (0, \Delta I_{(a,b)}(\sigma_1))$.

Phase 1: At the end of this phase $\Delta\mathbf{X}_{(a,b)}(\sigma_1) = (0, \Delta I_{(a,b)}(\sigma_1))$. By the induction hypothesis

$$\mathbb{P}\left[\sigma_1 > t\right] \quad\leq\quad Ct^{-\gamma} \qquad \text{for } t \geq 1\,. \tag{4.1}$$

We need a tail bound on $\mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x\right]$ for $x \geq 2$. Using (4.1), $x \geq 2$ and $t \geq 1$,

$$\begin{aligned}\mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x\right] \quad &\leq \quad \mathbb{P}\left[\sigma_1 > t\right] + \mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x,\ \sigma_1 \leq t\right] \\ &\leq \quad Ct^{-\gamma} + \mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x,\ \sigma_1 \leq t\right] . \end{aligned} \tag{4.2}$$

Since $x \geq 2$ and $|\Delta I_{(a,b)}(0)| \leq 1$, the second probability satisfies

$$\begin{aligned}&\mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x,\ \sigma_1 \leq t\right] \quad \leq \\ &\quad \mathbb{P}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\circ \mathrm{d}W_2(s) - \int_0^u \widetilde{W}_1^a(s)W_2^b(s)\circ \mathrm{d}W_2(s) + \Delta I_{(a,b)}(0)\right| > x\right] \\ &\qquad\qquad \leq \quad 2\,\mathbb{P}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\circ \mathrm{d}W_2(s)\right| > x/4\right] . \end{aligned} \tag{4.3}$$

Using the Itô representation of $I_{(a,b)}$ (Equation (3.13)),

$$\begin{aligned}&\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\circ \mathrm{d}W_2(s)\right| \quad \leq \\ &\qquad \sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\,\mathrm{d}W_2(s)\right| + \frac{b}{2}\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^{b-1}(s)\,\mathrm{d}s\right| .\end{aligned}$$

Now Cauchy-Schwarz arguments yield

$$\begin{aligned}&\mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\circ \mathrm{d}W_2(s)\right|^2\right] \\ &\quad\leq \quad 2\,\mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\,\mathrm{d}W_2(s)\right|^2\right] + \frac{b^2}{2}\,\mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^{b-1}(s)\,\mathrm{d}s\right|^2\right] \\ &\quad\leq \quad 2\,\mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\,\mathrm{d}W_2(s)\right|^2\right] + \frac{b^2}{2}\,\mathbb{E}\left[\left(\int_0^t |W_1^a(s)||W_2^{b-1}(s)|\,\mathrm{d}s\right)^2\right] .\end{aligned}$$

By the BDG inequality,

$$\begin{aligned}&\mathbb{E}\left[\sup_{u\leq t}\left|\int_0^u W_1^a(s)W_2^b(s)\,\mathrm{d}W_2(s)\right|^2\right] \\ &\qquad \leq \quad C\,\mathbb{E}\left[\int_0^t W_1^{2a}(s)W_2^{2b}(s)\,\mathrm{d}s\right] \quad \leq \quad C\int_0^t s^a s^b\,\mathrm{d}s \quad = \quad \frac{Ct^{a+b+1}}{a+b+1} ,\end{aligned}$$

while a further application of the Cauchy Schwarz inequality yields

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t |W_1^a(s)||W_2^{b-1}(s)|\,\mathrm{d}s\right)^2\right] \quad &\leq \quad t\,\mathbb{E}\left[\int_0^t W_1^{2a}(s)W_2^{2b-2}(s)\,\mathrm{d}s\right] \\ &\leq \quad Ct\int_0^t s^a s^{b-1}\,\mathrm{d}s \quad = \quad \frac{Ct^{a+b+1}}{a+b} .\end{aligned}$$

Using the Tchebychev inequality and the above two bounds together with Equation (4.3), if $x \geq 2$ and $t \geq 1$ then

$$\mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x,\ \sigma_1 \leq t\right] \quad \leq \quad \frac{Ct^{a+b+1}}{x^2} . \tag{4.4}$$

Combining inequalities (4.2) and (4.4), and choosing $t = x^{2/(a+b+1+\gamma)}$, if $x \geq 2$ then

$$\mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)| > x\right] \quad \leq \quad Cx^{-2\gamma/(a+b+1+\gamma)} . \tag{4.5}$$

Phase 2: During this phase, the driving Brownian motions are synchronously coupled till (W_1, W_2) hits the line $y = Rx$. This is done to get to the starting configuration of the coupled processes in Lemma 9. Between σ_1 and σ_2, the two Brownian motions are coalesced and synchronously coupled and hence $\Delta\mathbf{X}_{(a,b)}(\sigma_2) = (0, \Delta I_{(a,b)}(\sigma_1))$.

To get a bound on the tail of the distribution of $\sigma_2 - \sigma_1$, we rewrite it as follows, using (4.1):

$$\begin{aligned} \mathbb{P}\left[\sigma_2 - \sigma_1 > t\right] &\leq \mathbb{P}\left[\sigma_1 > t^\alpha\right] + \mathbb{P}\left[\sigma_2 - \sigma_1 > t, \sigma_1 \leq t^\alpha\right] \\ &\leq Ct^{-\alpha\gamma} + \mathbb{P}\left[\sigma_2 - \sigma_1 > t, \sigma_1 \leq t^\alpha\right] , \end{aligned}$$

for $t \geq 1$, and arbitrary $\alpha \in (0,1)$. The second term above can be estimated in terms of the distance of (W_1, W_2) from the line $y = Rx$ at time σ_1, in fact following the lines of the proof of (3.49):

$$\begin{aligned} \mathbb{P}\left[\sigma_2 - \sigma_1 > t,\ \sigma_1 \leq t^\alpha\right] \leq\ & \mathbb{P}\left[|(W_1, W_2)(\sigma_1) - (W_1, W_2)(0)| > x,\ \sigma_1 \leq t^\alpha\right] \\ & + \mathbb{P}\left[|(W_1, W_2)(\sigma_1) - (W_1, W_2)(0)| \leq x,\ \sigma_2 - \sigma_1 > t\right] , \end{aligned} \tag{4.6}$$

where $x, \alpha > 0$ will be chosen appropriately to optimize the bounds. To estimate the first probability in (4.6), note that

$$\begin{aligned} &\mathbb{P}\left[|(W_1, W_2)(\sigma_1) - (W_1, W_2)(0)| > x,\ \sigma_1 \leq t^\alpha\right] \\ &\qquad \leq \mathbb{P}\left[\sup_{s\in[0,t^\alpha]} |(W_1, W_2)(s) - (W_1, W_2)(0)| > x\right] \leq Ct^{\alpha/2}x^{-1} . \end{aligned}$$

To control the second probability in (4.6), condition on the event $[|(W_1, W_2)(\sigma_1) - (W_1, W_2)(0)| \leq x]$ and use the strong Markov property to argue that the hitting time on the line $y = Rx$ by the Brownian motion $((W_1, W_2)(t) - (W_1, W_2)(\sigma_1) : t \geq \sigma_1)$ is stochastically dominated by the hitting time on zero by a one dimensional Brownian motion starting from x. Therefore,

$$\mathbb{P}\left[|(W_1, W_2)(\sigma_1) - (W_1, W_2)(0)| \leq x, \sigma_2 - \sigma_1 > t\right] \leq C\frac{x}{\sqrt{t}} .$$

Using the above estimates in (4.6), we obtain

$$\mathbb{P}\left[\sigma_2 - \sigma_1 > t,\ \sigma_1 \leq t^\alpha\right] \leq C\frac{t^{\alpha/2}}{x} + C\frac{x}{\sqrt{t}} .$$

Using this and (4.1), and choosing suitable values of x and α, we obtain $\gamma > 0$ such that

$$\mathbb{P}\left[\sigma_2 - \sigma_1 > t\right] \leq Ct^{-\gamma} \quad \text{for } t \geq 1 . \tag{4.7}$$

Phase 3: In this phase, Lemma 9 is used to couple $(W_1, W_2, I_{(a,b)})$ with $(\widetilde{W}_1, \widetilde{W}_2, \widetilde{I}_{(a,b)})$ while controlling the difference between the lower order integrals of the coupled processes.

Note that at time σ_3 the array $\Delta\mathbf{X}_{(a,b)}(\sigma_3)$ of monomial Stratonovich integrals is obtained by appending $\Delta I_{(a,b)}(\sigma_3) = 0$ to the array $\Delta\mathbf{X}_{(a^-,b^-)}(\sigma_3)$. Now $\Delta\mathbf{X}_{(a^-,b^-)}(\sigma_3)$ will be the discrepancy between the coupled sets of integrals at the end of the third phase and thus, it is necessary to control its size. We do this by controlling the size (in an appropriate sense) of each individual integral appearing in $\Delta\mathbf{X}_{(a^-,b^-)}(\sigma_3)$ and showing the coupling strategy of Lemma 9 does not make this size large. Fix any $(k,l) \preceq (a^-, b^-)$. Note that, as $\Delta I_{(k,l)}(\sigma_2) = \Delta I_{(k,l)}(\sigma_1) = 0$, scaling yields the following distributional equality (where the second equality simply involves rewriting the Stratonovich integral as the sum of an Itô integral and a time integral):

$$
\begin{aligned}
\Delta I_{(k,l)}(\sigma_3) &\overset{\mathcal{D}}{=} U\int_0^{T_{R,(a,b)}} (B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\circ \mathrm{d}B_2(s) \\
&= U\left(\int_0^{T_{R,(a,b)}} (B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\,\mathrm{d}B_2(s)+\frac{l}{2}\int_0^{T_{R,(a,b)}} (B_1^k(s)-\widetilde{B}_1^k(s))B_2^{l-1}(s)\,\mathrm{d}s\right)
\end{aligned}
\tag{4.8}
$$

where U has the same distribution as $|\Delta I_{(a,b)}(\sigma_1)|^{(k+l+1)/(a+b+1)}$, and (B_1,B_2) and $(\widetilde{B}_1,\widetilde{B}_2)$ are two-dimensional Brownian motions starting respectively from

$$
\left(\frac{W_1(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|^{1/(a+b+1)}},\frac{W_2(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|^{1/(a+b+1)}}\right)
$$

and

$$
\left(\frac{\widetilde{W}_1(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|^{1/(a+b+1)}},\frac{\widetilde{W}_2(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|^{1/(a+b+1)}}\right),
$$

and $T_{R,(a,b)}$ is the coupling time for the coupling construction of

$$
\left(B_1,B_2,\frac{I_{(a,b)}(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|}+\int B_1^aB_2^b\circ \mathrm{d}B_2\right)
$$

and

$$
\left(\widetilde{B}_1,\widetilde{B}_2,\frac{\widetilde{I}_{(a,b)}(\sigma_2)}{|\Delta I_{(a,b)}(\sigma_1)|}+\int \widetilde{B}_1^a\widetilde{B}_2^b\circ \mathrm{d}\widetilde{B}_2\right)
$$

given in Lemma 9. Furthermore, U is independent of $((B_1(t)-B_1(0),B_2(t)-B_2(0)):t\geq 0)$ and $((\widetilde{B}_1(t)-\widetilde{B}_1(0),\widetilde{B}_2(t)-\widetilde{B}_2(0)):t\geq 0)$.

Define stopping times $\theta_j,\tau_j,\eta_j,\lambda_j,\beta_j,j\geq 1$ in the time interval $[0,T_{R,(a,b)}]$ as in the proof of Lemma 9. As the Brownian motions move together on the intervals $[\beta_{j-1},\theta_j]$ and $[\lambda_j,\beta_j]$, the monomial Stratonovich integral $\Delta I_{(k,l)}$ does not change on these intervals.

On $[\theta_1,\lambda_1]$, the Itô integral in (4.8) can be written as

$$
\begin{aligned}
\int_{\theta_1}^{\lambda_1}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\,\mathrm{d}B_2(s) &= \int_{\theta_1}^{\lambda_1}\Delta B_1(s)\sum_{j=1}^{k-1}B_1^{k-j}(s)\widetilde{B}_1^{j-1}(s)B_2^l(s)\,\mathrm{d}B_2(s) \\
= \int_{\theta_1}^{\lambda_1}(\Delta B_1(s)B_1^{k+l-1}(\theta_1)R^l)&\sum_{j=1}^{k-1}\left(\frac{B_1(s)}{B_1(\theta_1)}\right)^{k-j}\left(\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right)^{j-1}\left(\frac{B_2(s)}{B_2(\theta_1)}\right)^{l}\mathrm{d}B_2(s) \\
= \left(\int_{\theta_1}^{\lambda_1}(\Delta B_1(s)B_1^{a+b-1}(\theta_1)R^b)\right.&\left.\sum_{j=1}^{k-1}\left(\frac{B_1(s)}{B_1(\theta_1)}\right)^{k-j}\left(\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right)^{j-1}\left(\frac{B_2(s)}{B_2(\theta_1)}\right)^{l}\mathrm{d}B_2(s)\right) \\
&\times\frac{1}{(B_1(\theta_1))^{(a+b)-(k+l)}R^{b-l}}\,.
\end{aligned}
\tag{4.9}
$$

Using $\sup_{t\in[\theta_1,\tau_1]}|\Delta B_1(t)|=\frac{1}{|B_1(\theta_1)|^{a+b-1}R^b}$ and $\sup_{t\in[\tau_1,\eta_1]}|\Delta B_1(t)|=|\Delta B_1(\tau_1)|$ together with (3.45),

$$
\mathbb{P}\left[\sup_{\theta_1\leq t\leq\lambda_1}|\Delta B_1(s)||B_1^{a+b-1}(\theta_1)|R^b\geq x\right]\quad\leq\quad\frac{1}{x}\qquad\text{for } x\geq 1\,. \tag{4.10}
$$

The same line of argument employed to obtain (3.24) can be used to get the following bound. For $1 \le j \le k-1$, $x > 2^{k+l-1}$:

$$\mathbb{P}\left[\sup_{\theta_1 \le s \le \tau_1} \left|\frac{W_1(s)}{W_1(\theta_1)}\right|^{k-j} \left|\frac{\widetilde{W}_2(s)}{\widetilde{W}_1(\theta_1)}\right|^{j-1} \left|\frac{W_2(s)}{W_2(\theta_1)}\right|^{l} > x\right] \quad \le \quad \frac{2^{2(k+l-1)/l}}{R^{8n+4}x^{2/l}}\, C\,.$$

(where the above bound is taken to be zero if $l = 0$). As a result, using (3.35) and (3.46) (bounding $\sup_{\tau_1 \le t \le \eta_1} \left|\frac{B_1(t)}{B_1(\theta_1)}\right|$ and $\sup_{\eta_1 \le t \le \lambda_1} \left|\frac{B_1(t)}{B_1(\theta_1)}\right|$; the same bounds can be shown to hold when B_1 is replaced by $\widetilde{B}_1$ and B_2), we can pick $\gamma > 0$ such that, for any $1 \le j \le k-1$ and all $x > 2^{k+l-1}$,

$$\mathbb{P}\left[\sup_{\theta_1 \le t \le \lambda_1} \left|\frac{B_1(s)}{B_1(\theta_1)}\right|^{k-j} \left|\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right|^{j-1} \left|\frac{B_2(s)}{B_2(\theta_1)}\right|^{l} > x\right] \quad \le \quad \frac{C}{(R^{2n}x)^{\gamma}}\,. \tag{4.11}$$

(4.10) and (4.11) together imply that there exists $\gamma > 0$ such that for any $1 \le j \le k-1$ and all $x > 2^{k+l}$,

$$\mathbb{P}\left[\sup_{\theta_1 \le s \le \lambda_1} |\Delta B_1(s)||B^{a+b-1}(\theta_1)|R^b \left|\frac{B_1(s)}{B_1(\theta_1)}\right|^{k-j} \left|\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right|^{j-1} \left|\frac{B_2(s)}{B_2(\theta_1)}\right|^{l} > x\right] \quad \le \quad \frac{C}{x^{\gamma}}\,. \tag{4.12}$$

Also, recall from (3.50) (using $\alpha = 1$) that

$$\mathbb{P}\left[\lambda_1 - \theta_1 > t\right] \quad \le \quad Ct^{-1/2} \qquad \text{for } t \ge 1\,. \tag{4.13}$$

Using (4.12) and (4.13) in Lemma 5, we can chose positive γ, x_0 such that (for $x \ge x_0$)

$$\begin{aligned}
&\mathbb{P}\left[\left|\int_{\theta_1}^{\lambda_1} (\Delta B_1(s) B^{a+b-1}(\theta_1) R^b) \sum_{j=1}^{k-1} \left(\frac{B_1(s)}{B_1(\theta_1)}\right)^{k-j} \left(\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right)^{j-1} \left(\frac{B_2(s)}{B_2(\theta_1)}\right)^{l} \mathrm{d}B_2(s)\right| > x\right] \\
&\le \sum_{j=1}^{k-1} \mathbb{P}\left[\left|\int_{\theta_1}^{\lambda_1} (\Delta B_1(s) B^{a+b-1}(\theta_1) R^b) \left(\frac{B_1(s)}{B_1(\theta_1)}\right)^{k-j} \left(\frac{\widetilde{B}_1(s)}{\widetilde{B}_1(\theta_1)}\right)^{j-1} \left(\frac{B_2(s)}{B_2(\theta_1)}\right)^{l} \mathrm{d}B_2(s)\right| > \frac{x}{k}\right] \\
&\le \quad Cx^{-\gamma}\,.
\end{aligned} \tag{4.14}$$

Furthermore, by virtue of the definition of the stopping time θ_1 in the coupling construction, and because $(k,l) \prec (a,b)$,

$$\frac{1}{|(B_1(\theta_1))|^{(a+b)-(k+l)} R^{b-l}} \quad \le \quad \frac{1}{R^{f(a,b)-f(k,l)}} \quad \le \quad \frac{1}{R}\,. \tag{4.15}$$

Using (4.14) and (4.15) in (4.9), for $x \ge x_0/R$ we obtain

$$\mathbb{P}\left[\left|\int_{\theta_1}^{\lambda_1} (B_1^k(s) - \widetilde{B}_1^k(s)) B_2^l(s)\, \mathrm{d}B_2(s)\right| > x\right] \quad \le \quad \frac{C}{(Rx)^{\gamma}}\,. \tag{4.16}$$

Recall from Lemma 9 that $S_{R,(a,b)} = \bigcup_{j=1}^{\infty} [\theta_j, \lambda_j]$ and

$$\int_0^{T_{R,(a,b)}} (B_1^k(s) - \widetilde{B}_1^k(s)) B_2^l(s)\, \mathrm{d}B_2(s) = \int_{S_{R,(a,b)}} (B_1^k(s) - \widetilde{B}_1^k(s)) B_2^l(s)\, \mathrm{d}B_2(s).$$

Now apply Lemma 6 to the sum on the right hand side of

$$R\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\,\mathrm{d}B_2(s)=\tau_1^{**}+\sum_{j=1}^{\infty}R^{-(k+l+1)j\delta/(a+b+1)}\left(\Pi_{m=1}^{j}X_m^{**}\right)\tau_{j+1}^{**}$$

with (X_j^{**},τ_j^{**}) replacing (X_j,τ_j) in the statement of the lemma, where

$$\tau_j^{**} \quad = \quad \frac{R\left|\int_{\theta_j}^{\lambda_j}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\,\mathrm{d}B_2(s)\right|}{|\Delta I_{(a,b)}(\beta_{j-1})|^{(k+l+1)/(a+b+1)}}$$

and

$$X_j^{**} \quad = \quad \frac{R^{(k+l+1)\delta/(a+b+1)}|\Delta I_{(a,b)}(\beta_j)|^{(k+l+1)/(a+b+1)}}{|\Delta I_{(a,b)}(\beta_{j-1})|^{(k+l+1)/(a+b+1)}}\,.$$

Taking δ to be the same as that used in (3.51), and setting $\beta_0=0$, we can find $\gamma'>0$ for which

$$\mathbb{P}\left[\left|\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\,\mathrm{d}B_2(s)\right|>x\right] \quad \leq \quad \frac{C}{(Rx)^{\gamma'}} \tag{4.17}$$

for $x\geq x_0/R$, when R is sufficiently large, where x_0 is the same as that used in (4.16). Using the same estimates as above and using the assertions of Lemma 5 involving time integrals, for $l\geq 1$ it follows in a similar way that there is $\gamma''>0, x_1>0$ such that

$$\mathbb{P}\left[\left|\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^{l-1}(s)\,\mathrm{d}s\right|>x\right] \quad \leq \quad \frac{C}{(R^{2n+2}x)^{\gamma''}} \tag{4.18}$$

for $x\geq x_1/R^{2n+2}$, for R sufficiently large.

Using the distributional equality (4.8), when $x\geq\max\{x_0^2,x_1^2,4\}/R$ and U is as defined for that equation, we can find positive constants γ,γ_1,γ_2 such that

$$\begin{aligned}
\mathbb{P}\left[|\Delta I_{(k,l)}(\sigma_3)|>x\right] \quad &= \quad \mathbb{P}\left[U\left|\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\circ\,\mathrm{d}B_2(s)\right|>x\right]\\
&\leq \quad \mathbb{P}\left[U\geq\sqrt{xR}\right]+\mathbb{P}\left[\left|\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\circ\,\mathrm{d}B_2(s)\right|>\sqrt{x/R}\right]\\
&= \quad \mathbb{P}\left[|\Delta I_{(a,b)}(\sigma_1)|>(\sqrt{xR})^{(a+b+1)/(k+l+1)}\right]\\
&\quad +\mathbb{P}\left[\left|\int_0^{T_{R,(a,b)}}(B_1^k(s)-\widetilde{B}_1^k(s))B_2^l(s)\circ\,\mathrm{d}B_2(s)\right|>\sqrt{x/R}\right]\\
&\leq \quad \frac{C}{(xR)^{\gamma_1}}+\frac{C}{(xR)^{\gamma_2}} \quad \leq \quad \frac{C}{(xR)^{\gamma}}\,.
\end{aligned}$$

The last step above follows from (4.5), (4.17) and (4.18). As this bound can be chosen to hold for all $(k,l)\prec(a,b)$ and $\Delta I_{(a,b)}(\sigma_3)=0$, we can choose positive constants C,γ such that

$$\mathbb{P}\left[|\Delta\mathbf{X}_{(a,b)}(\sigma_3)|>x\right] \quad \leq \quad \frac{C}{(xR)^{\gamma}} \qquad \text{when } x\geq C/R\,. \tag{4.19}$$

A bound on the tail of the law of σ_3 can be obtained using Lemma 9: if $t\geq 1$ then

$$\mathbb{P}\left[(\sigma_3-\sigma_2)/|\Delta I_{(a,b)}(\sigma_1)|^{2/(a+b+1)}>R^{4n+2}t\right] \quad \leq \quad Ct^{-\gamma}\,.$$

Together with (4.5) this implies that when $t \geq 1$

$$\mathbb{P}\left[\sigma_3 - \sigma_2 > R^{4n+2}t\right] \quad \leq \quad Ct^{-\gamma}\,. \tag{4.20}$$

Thus, from (4.1), (4.7) and (4.20), when $t \geq 1$

$$\mathbb{P}\left[\sigma_3 > R^{4n+2}t\right] \quad \leq \quad Ct^{-\gamma}\,. \tag{4.21}$$

B: Describing subsequent cycles and successful coupling

After completion of the first cycle, at time σ_3, we re-scale $\mathbf{X}_{(a,b)}$ and $\widetilde{\mathbf{X}}_{(a,b)}$ according to Lemma 4 by a (random) scaling $\mathcal{S}_{R_1}$ such that

$$|\mathcal{S}_{R_1}\left(\Delta\mathbf{X}_{(a,b)}\right)(\sigma_3)| \quad = \quad 1\,.$$

Define $\sigma_4, \sigma_5, \sigma_6$ (for the original process) corresponding to $\sigma_1, \sigma_2, \sigma_3$ for the coupled process after scaling exactly as before, and so on. At each stopping time σ_{3k}, $k \geq 1$, we denote by $\mathcal{S}_{R_k}$ the (random) scaling that renormalizes at 1 the norm of the difference of the re-scaled processes.

For any $r \geq 1$, $t \geq 0$,

$$r|\Delta\mathbf{X}_{(a,b)}(t)| \quad \leq \quad |\mathcal{S}_r\left(\Delta\mathbf{X}_{(a,b)}\right)(t)| \quad \leq \quad r^{a+b+1}|\Delta\mathbf{X}_{(a,b)}(t)|\,,$$

with inequalities reversed if $r \leq 1$.

For each $k \geq 1$,

$$\left|\mathcal{S}_{(\Pi_{j=1}^k R_j)}\left(\Delta\mathbf{X}_{(a,b)}\right)(\sigma_{3k})\right| \quad = \quad 1\,.$$

Therefore, if it can be shown that $\lim_{k\to\infty}\Pi_{j=1}^k R_j = \infty$ then it follows that

$$\lim_{k\to\infty}|\Delta X_{(a,b)}(\sigma_{3k})| = 0.$$

We achieve this by estimating the tail of the distribution of R_1^{-1}. Note that if $R_1^{-1} \leq 1$, then from the above relations

$$\frac{R^{1/(a+b+1)}}{R_1} \quad \leq \quad |R\Delta\mathbf{X}_{(a,b)}(\sigma_3)|^{1/(a+b+1)}$$

while if $R_1^{-1} \geq 1$ then

$$\frac{R}{R_1} \quad \leq \quad |R\Delta\mathbf{X}_{(a,b)}(\sigma_3)|\,.$$

Thus, for $x \geq 1$, if $R^{1/(a+b+1)} > x$ then

$$\begin{aligned}
\mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \geq x\right] \quad &= \quad \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \in \left[x, R^{1/(a+b+1)}\right]\right] + \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} > R^{1/(a+b+1)}\right] \\
&\leq \quad \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \in \left[x, R^{1/(a+b+1)}\right]\right] + \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} > x,\ R_1^{-1} > 1\right] \\
&\leq \quad \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \in \left[x, R^{1/(a+b+1)}\right]\right] + \mathbb{P}\left[\frac{R}{R_1} > x,\ R_1^{-1} > 1\right] \\
&\leq \quad \mathbb{P}\left[|R\Delta\mathbf{X}_{(a,b)}(\sigma_3)| \geq x^{a+b+1}\right] + \mathbb{P}\left[|R\Delta\mathbf{X}_{(a,b)}(\sigma_3)| \geq x\right] \\
&\leq \quad Cx^{-\gamma}\,,
\end{aligned}$$

where the last inequality is a consequence of (4.19).

On the other hand, if $R^{1/(a+b+1)} \leq x$,

$$\begin{aligned}\mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \geq x\right] &= \mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \geq x,\ R_1^{-1} \geq 1\right] \\ &\leq \mathbb{P}\left[\frac{R}{R_1} \geq x,\ R_1^{-1} \geq 1\right] \quad \leq \quad \mathbb{P}\left[|R\Delta\mathbf{X}_{(a,b)}(\sigma_3)| \geq x\right] \\ &\leq Cx^{-\gamma}\,.\end{aligned}$$

Combining the above two bounds, if $x \geq 1$ then

$$\mathbb{P}\left[\frac{R^{1/(a+b+1)}}{R_1} \geq x\right] \quad \leq \quad Cx^{-\gamma}\,. \tag{4.22}$$

Applying Lemma 6 with $X_k = \dfrac{R^{1/(a+b+1)}}{R_k}$ and $\tau_k = 1$, we obtain

$$\sum_{k=1}^{\infty} R^{-k/(a+b+1)}(\Pi_{j=1}^{k} X_j) \quad < \quad \infty$$

almost surely for sufficiently large R. In particular, $\lim_{k\to\infty} \Pi_{j=1}^{k} R_j = \infty$ and consequently,

$$\lim_{k\to\infty} |\Delta\mathbf{X}_{(a,b)}(\sigma_{3k})| = 0.$$

Finally, to show that the coupling is successful and to verify the induction hypothesis for (a, b), it is necessary to show that $\lim_{k\to\infty} \sigma_{3k}$ is almost surely finite and that its law has a power law tail.

This follows by applying Lemma 6 to the sum on the right hand side of the expression

$$\sigma_{3k+3} \quad = \quad \hat{\tau}_1 + \sum_{l=1}^{k} R^{-2l/(a+b+1)}(\Pi_{j=1}^{l} X_j^2)\hat{\tau}_{l+1}, \qquad k \geq 1,$$

with $(X_k^2, \hat{\tau}_k)$ in place of (X_j, τ_j) in the lemma, where $X_k = \dfrac{R^{1/(a+b+1)}}{R_k}$, defined for $k \geq 1$, and $\hat{\tau}_k = \left(\Pi_{j=1}^{k-1} R_j^2\right)(\sigma_{3k} - \sigma_{3k-3})$, defined for $k \geq 2$, and $\sigma_0 = 0$. As the law of $\hat{\tau}_k$ has the same tail as that of σ_3, it follows from (4.21) and (4.22) that if $t \geq 1$ then

$$\mathbb{P}\left[\lim_{k\to\infty} \sigma_{3k} > R^{4n+2}t\right] \quad \leq \quad Ct^{-\gamma}\,,$$

for sufficiently large R. This establishes the induction hypothesis, and so completes the construction of a successful coupling when the starting points of the coupled Brownian motions satisfy $(W_1, W_2)(0) = (\widetilde{W}_1, \widetilde{W}_2)(0)$ and $W_2(0) = RW_1(0)$.

The argument is completed by showing how to construct the coupling from arbitrary starting points $\mathbf{X}(0)$ and $\widetilde{\mathbf{X}}(0)$ satisfying $|\mathbf{X}(0)| \leq 1$, $|\widetilde{\mathbf{X}}(0)| \leq 1$. To do this, define the stopping times

$$\begin{aligned}\sigma_{-1} &= \inf\{t \geq 0 :\ \text{using reflection coupling},\ (W_1, W_2)(t) = (\widetilde{W}_1, \widetilde{W}_2)(t)\}\,, \\ \sigma_0 &= \inf\{t \geq \sigma_{-1} :\ \text{using synchronous coupling},\ W_2(t) = RW_1(t)\} \\ T &= \inf\{t \geq \sigma_0 :\ \text{coupling strategy constructed above},\ \mathbf{X}(t) = \widetilde{\mathbf{X}}(t)\}\,.\end{aligned}$$

Using Brownian hitting time estimates derived from the reflection principle, when $t \geq 1$

$$\mathbb{P}\left[\sigma_{-1} > t\right] \quad \leq \quad \frac{|(W_1, W_2)(0) - (\widetilde{W}_1, \widetilde{W}_2)(0)|}{\sqrt{t}} \quad \leq \quad \frac{2}{\sqrt{t}}\,.$$

Now, using the fact that $|(W_1,W_2)(0)-(\widetilde{W}_1,\widetilde{W}_2)(0)|\leq 2$, and controlling the distance of the Brownian motion (W_1,W_2) from the line $y=Rx$ at time σ_{-1} as in the proof of (3.49), we obtain a constant C that does not depend on the starting points such that

$$\mathbb{P}\left[\sigma_0-\sigma_{-1}>t\right] \quad\leq\quad Ct^{-1/6}\,.$$

Consequently, if $t\geq 1$ then

$$\sup\{\mathbb{P}\left[\sigma_0>t\right]\;:\;|\mathbf{X}(0)|\leq 1,\;|\widetilde{\mathbf{X}}(0)|\leq 1\}\quad\leq\quad Ct^{-1/6}\,. \tag{4.23}$$

Furthermore, for $x\geq 4$ and arbitrary $t>0$ to be chosen later,

$$\begin{aligned}
\mathbb{P}\left[|\Delta\mathbf{X}(\sigma_0)|>x\right] \quad&=\quad \mathbb{P}\left[|\Delta\mathbf{X}(\sigma_{-1})|>x\right]\\
\leq\quad & \mathbb{P}\left[|\mathbf{X}(\sigma_{-1})-\mathbf{X}(0)|+|\widetilde{\mathbf{X}}(\sigma_{-1})-\widetilde{\mathbf{X}}(0)|+|\mathbf{X}(0)-\widetilde{\mathbf{X}}(0)|>x\right]\\
\leq\quad & \mathbb{P}\left[|\mathbf{X}(\sigma_{-1})-\mathbf{X}(0)|+|\widetilde{\mathbf{X}}(\sigma_{-1})-\widetilde{\mathbf{X}}(0)|>x/2\right] \quad (\text{ as } |\mathbf{X}(0)-\widetilde{\mathbf{X}}(0)|\leq 2 \text{ and } x\geq 4)\\
\leq\quad & 2\,\mathbb{P}\left[|\mathbf{X}(\sigma_{-1})-\mathbf{X}(0)|>x/4\right]\\
\leq\quad & 2\,\mathbb{P}\left[\sigma_{-1}>t\right]+2\,\mathbb{P}\left[\sup_{s\leq t}|\mathbf{X}(s)-\mathbf{X}(0)|>x/4\right]\\
\leq\quad & \frac{4}{\sqrt{t}}+C\frac{\mathbb{E}\left[\sum_{(k,l)\in\Delta_n}\int_0^t W_1(s)^{2k}W_2(s)^{2l}\,\mathrm{d}s\right]}{x^2}\\
& \quad +Ct\frac{\mathbb{E}\left[\sum_{(k,l)\in\Delta_n} l^2\int_0^t W_1(s)^{2k}W_2(s)^{2l-2}\,\mathrm{d}s\right]}{x^2}\\
\leq\quad & \frac{2}{\sqrt{t}}+C\frac{\sum_{(k,l)\in\Delta_n}\int_0^t s^{k+l}\,\mathrm{d}s}{x^2}+Ct\frac{\sum_{(k,l)\in\Delta_n} l^2\int_0^t s^{k+l-1}\,\mathrm{d}s}{x^2}\quad\leq\quad\frac{2}{\sqrt{t}}+C\frac{t^{n+1}}{x^2}\\
\leq\quad & Cx^{-\gamma}\,.
\end{aligned}$$

Here the last step holds for some $\gamma>0$, and for t chosen appropriately in terms of x. Moreover the final constant C above does not depend on the starting points so long as they lie in the unit ball. Hence, for $x\geq 4$,

$$\sup\{\mathbb{P}\left[|\Delta\mathbf{X}(\sigma_0)|>x\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\quad\leq\quad Cx^{-\gamma}\,. \tag{4.24}$$

Thus, for $t\geq 2$, $x\geq 4$,

$$\begin{aligned}
&\sup\{\mathbb{P}\left[T>t\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\quad\leq\\
&\qquad\sup\{\mathbb{P}\left[T>t,\;|\Delta X(\sigma_0)|\leq x,\;\sigma_0\leq t/2\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\\
&\qquad\qquad+\sup\{\mathbb{P}\left[|\Delta X(\sigma_0)|>x\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\\
&\qquad\qquad\qquad+\sup\{\mathbb{P}\left[\sigma_0>t/2\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\,.
\end{aligned}$$

The second and third terms above are already estimated in (4.24) and (4.23) respectively. To bound the first term above, we apply the strong Markov property at σ_0 to obtain

$$\begin{aligned}
&\sup\{\mathbb{P}\left[T>t,|\Delta X(\sigma_0)|\leq x,\sigma_0\leq t/2\right]\;:\;|\mathbf{X}(0)|\leq 1,|\widetilde{\mathbf{X}}(0)|\leq 1\}\quad\leq\\
&\qquad\qquad\sup\{\mathbb{P}\left[T>t/2\right]\;:\;|\Delta\mathbf{X}(0)|\leq x,(\mathbf{X}(0),\widetilde{\mathbf{X}}(0))\in\mathcal{R}\}
\end{aligned}$$

Apply Lemma 4 to the right hand side above, taking $r=x^{-1}$. Now use the fact that if $|\Delta\mathbf{X}(0)|\leq x$, then as $x>1$,

$$|\mathcal{S}_{x^{-1}}\left(\Delta\mathbf{X}\right)(0)|\quad\leq\quad x^{-1}|\Delta\mathbf{X}(0)|\quad\leq\quad 1\,.$$

Hence, we obtain

$$\sup\{\mathbb{P}\left[T > t/2\right] \;:\; |\Delta\mathbf{X}(0)| \leq x, (\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}\} \quad \leq \\ \sup\{\mathbb{P}\left[T > t/(2x^2)\right] \;:\; |\Delta\mathbf{X}(0)| \leq 1, (\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}\}\,.$$

If $t \geq 2x^2$ then the tail estimate of the law of the above coupling time gives

$$\sup\{\mathbb{P}\left[T > t/(2x^2)\right] \;:\; |\Delta\mathbf{X}(0)| \leq 1, (\mathbf{X}(0), \widetilde{\mathbf{X}}(0)) \in \mathcal{R}\} \quad \leq \quad C\frac{x^{2\gamma}}{t^\gamma}\,.$$

This estimate, along with (4.23) and (4.24), yields

$$\sup\{\mathbb{P}\left[T > t\right] \;:\; |\mathbf{X}(0)| \leq 1, |\widetilde{\mathbf{X}}(0)| \leq 1\} \quad \leq \quad C\frac{x^{2\gamma}}{t^\gamma} + Cx^{-\gamma} + Ct^{-1/6}\,.$$

Choosing x appropriately in terms of t, we see that for sufficiently large t (and consequently for all $t \geq 1$ by readjusting the constants),

$$\sup\{\mathbb{P}\left[T > t\right] \;:\; |\mathbf{X}(0)| \leq 1, |\widetilde{\mathbf{X}}(0)| \leq 1\} \quad \leq \quad Ct^{-\gamma}\,,$$

and this proves the theorem. □

5 Conclusion

In this article, we have constructed a successful Markovian coupling for the two-dimensional Brownian motion along with a finite collection of its monomial Stratonovich integrals. In the context provided by Theorem 3, this is a further step in the direction of extending Markovian coupling techniques beyond the realm of specific examples towards a more general context. Our method shares some features with an iterative coupling scheme employed in [22] for coupling iterated Kolmogorov diffusions, though the inductive strategy described in the current paper seems to be more robust as one can build iterations within iterations into the coupling, exploiting the inductive approach described here. A natural next step for the general program of coupling hypoelliptic diffusions would be to couple diffusions driven by *nilpotent vector fields* which do not just depend on the driving Brownian motion but the entire diffusion. The Baker-Campbell-Hausdorff formula can be employed to show that such a coupling can be achieved if one can construct successful Markovian couplings for *Brownian motion on the free Carnot group* of finite order [5]. The geometry of the Carnot group seems to lend itself particularly to our inductive approach: the Lie algebra $\mathcal{U}$ of the Carnot group has a graded structure given by $\mathcal{U} = \mathcal{U}_1 \oplus \mathcal{U}_2 \cdots \oplus \mathcal{U}_N$ and there are dilation operators δ_t that act by multiplication by t^i on the elements of $\mathcal{U}_i$ while preserving the graded structure. A possible strategy for constructing the coupling in this case would be to use the graded structure in the induction hypothesis and to use the dilation operator to implement the scaling strategy used repeatedly in the above arguments. We will investigate this in future work.

The current article also provides quantitative bounds on the distribution of the coupling time. These can be used to obtain estimates on the total variation distance between the laws of the diffusions. Employed in conjunction with the scaling property (Lemma 4), this would lead to gradient estimates for heat kernels and harmonic functions corresponding to the generator of the diffusion [9, 10, 2]. We note here that it was shown in recent work [3, 2] that optimal bounds on total variation distance and good gradient estimates, especially in the case of hypoelliptic diffusions, require non-immersion couplings. However, so far, it has been possible to provide explicit constructions of these couplings only in rather special examples: (generalized) Kolmogorov diffusions [3] and Brownian motion on the Heisenberg group [2]. An important challenge is to find robust non-immersion coupling constructions applicable to a wider framework and then to compare their performance with analogous immersion or Markovian couplings.

References

[1] Aldous, D. J. (1983). Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de probabilités de Strasbourg 17*, 243–297. MR-770418

[2] Banerjee, S., M. Gordina, and P. Mariano (2016). Coupling in the Heisenberg group and its applications to gradient estimates. *The Annals of Probability* (to appear). arXiv:1610.06430, 33pp.

[3] Banerjee, S. and W. S. Kendall (2015). Coupling the Kolmogorov Diffusion: maximality and efficiency considerations. *Advances in Applied Probability 48*(A), 15–35. MR-3539295

[4] Banerjee, S. and W. S. Kendall (2017). Rigidity for Markovian maximal couplings of elliptic diffusions. *Probability Theory and Related Fields 168*(1), 55–112. MR-3651049

[5] Baudoin, F. (2004). *An introduction to the geometry of stochastic flows.* Imperial College Press. MR-2154760

[6] Ben Arous, G., M. Cranston, and W. S. Kendall (1995). Coupling constructions for hypoelliptic diffusions: Two examples. In M. Cranston and M. Pinsky (Eds.), *Stochastic Analysis: Proceedings of Symposia in Pure Mathematics 57*, Volume 57, Providence, RI Providence, pp. 193–212. American Mathematical Society. MR-1335472

[7] Burdzy, K. and W. S. Kendall (2000). Efficient Markovian couplings: examples and counterexamples. *The Annals of Applied Probability 10*(2), 362–409. MR-1768241

[8] Chen, M.-F. and S.-F. Li (1989). Coupling methods for multidimensional diffusion processes. *The Annals of Probability*, 151–177. MR-972776

[9] Cranston, M. (1991). Gradient estimates on manifolds using coupling. *Journal of Functional Analysis 99*(1), 110–124. MR-1120916

[10] Cranston, M. (1992). A probabilistic approach to gradient estimates. *Canad. Math. Bull 35*(1), 46–55. MR-1157463

[11] Ernst, P., W. S. Kendall, G. O. Roberts, and J. S. Rosenthal (2017). MEXIT: Maximal uncoupling times for Markov processes. arXiv:1702.03917, 34pp.

[12] Friz, P. K. and M. Hairer (2014). *A Short Course on Rough Paths: With an Introduction to Regularity Structures.* Universitext. Springer International Publishing.

[13] Goldstein, S. (1979). Maximal coupling. *Probability Theory and Related Fields 46*(2), 193–204. MR-516740

[14] Griffeath, D. (1975). A maximal coupling for Markov chains. *Probability Theory and Related Fields 31*(2), 95–106. MR-0370771

[15] Grubmüller, H. and P. Tavan (1994). Molecular dynamics of conformational substates for a simplified protein model. *The Journal of chemical physics 101*(6), 5047–5057.

[16] Ikeda, N. and S. Watanabe (1981). *Stochastic differential equations and diffusion processes.* Amsterdam: North Holland / Kodansha. MR-1011252

[17] Karatzas, I. and S. Shreve (2012). *Brownian motion and stochastic calculus*, Volume 113. Springer Science & Business Media. MR-1121940

[18] Kendall, W. S. (1986). Nonnegative Ricci curvature and the Brownian coupling property. *Stochastics: An International Journal of Probability and Stochastic Processes 19*(1–2), 111–129. MR-864339

[19] Kendall, W. S. (2007). Coupling all the Lévy stochastic areas of multidimensional Brownian motion. *The Annals of Probability*, 935–953. MR-2319712

[20] Kendall, W. S. (2010). Coupling time distribution asymptotics for some couplings of the Lévy stochastic area. In N. H. Bingham and C. M. Goldie (Eds.), *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, London Mathematical Society Lecture Note Series, Chapter 19, pp. 446–463. Cambridge: Cambridge University Press. MR-2744251

[21] Kendall, W. S. (2015). Coupling, local times, immersions. *Bernoulli 21*(2), 1014–1046. MR-3338655

[22] Kendall, W. S. and C. J. Price (2004). Coupling iterated Kolmogorov diffusions. *Electronic Journal of Probability 9*(Paper 13), 382–410. MR-2080604

[23] Kuwada, K. (2009). Characterization of maximal Markovian couplings for diffusion processes. *Electron. J. Probab 14*(25), 633–662. MR-2486817

[24] Lindvall, T. and L. C. G. Rogers (1986). Coupling of multidimensional diffusions by reflection. *The Annals of Probability*, 860–872. MR-841588

[25] Lobry, C. (1970). Contrôlabilité des systèmes non linéaires. *SIAM Journal on Control and Optimization 8*(4), 573–605. MR-0430926

[26] Markus, L. and A. Weerasinghe (1988). Stochastic oscillators. *Journal of Differential Equations 71*(2), 288–314. MR-927004

[27] Neuenschwander, D. (1996). *Probabilities on the Heisenberg group*, Volume 1630 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag. MR-1439509

[28] Pitman, J. W. (1976). On coupling of Markov chains. *Probability Theory and Related Fields 35*(4), 315–322. MR-0415775

[29] Rosenthal, J. S. (1997). Faithful Couplings of Markov Chains: Now Equals Forever. *Advances in Applied Mathematics 18*(3), 372–381. MR-1436487

[30] Sverchkov, M. Y. and S. N. Smirnov (1990). Maximal coupling for processes in $D[0, \infty]$. In *Soviet Math. Dokl*, Volume 41, pp. 352–354. MR-1072656

[31] Villani, C. (2006). Hypocoercive diffusion operators. In *International Congress of Mathematicians*, Volume 3, pp. 473–498. MR-2275692

[32] Von Renesse, M.-K. (2004). Intrinsic Coupling on Riemannian Manifolds and Polyhedra. *Electronic Journal of Probability 9*, 411–435. MR-2080605

Acknowledgments. We are grateful to the anonymous referees for their careful reading of the manuscript and suggestions that greatly improved the presentation of the article.