



---

**Universidad de Valladolid**

FACULTAD DE CIENCIAS

DEPARTAMENTO DE MATEMÁTICA APLICADA

PHD THESIS:

**WORD SERIES FOR THE NUMERICAL  
INTEGRATION OF STOCHASTIC  
DIFFERENTIAL EQUATIONS**

A Thesis submitted by ALFONSO ÁLAMO ZAPATERO for the  
degree of Doctor of Mathematics

---

Supervised by:  
JESÚS MARÍA SANZ SERNA





*A mis padres, Antonio y Pilar, Pilar y Antonio.*



# Agradecimientos

La realización de una tesis doctoral supone la culminación de muchas aspiraciones, no sólo intelectuales. Y desde luego que no es un camino fácil. Sin embargo, hay personas que prestan cosas tan valiosas como su tiempo, dedicación, atención, paciencia, conocimientos, comprensión, cariño... A todos ellos me gustaría dedicarles al menos unas pocas líneas.

- JESÚS MARÍA SANZ SERNA. Sin el apoyo, dedicación, paciencia y un sin fin de cosas más este trabajo no habría sido posible. Él es el mayor culpable. Es todo un privilegio trabajar con él. Agradezco mucho toda la ayuda que me ha brindado, como por ejemplo el círculo de científicos que he conocido gracias a él. De nuevo, muchas gracias por tus conocimientos, trato, comprensión, tiempo...
- ANTONIO ÁLAMO Y PILAR ZAPATERO. Mis padres. Son los otros causantes de que esté escribiendo estas líneas ahora mismo. Tener unos buenos padres en la vida es una suerte inmensa. Muchas gracias por todo vuestro cariño y por todos los medios que habéis puesto siempre a mi alcance.
- FÉLIX DELGADO. Coordinador de la escuela de doctorado. Por todas las facilidades que nos ha brindado a todos y por su dedicación, además de haber sido un excelente profesor. No me quisiera olvidarme de aquel curso de teoría de Galois, el cual me sirvió para entender la importancia de la teoría de grupos, hecho crucial en las páginas que siguen. Muchas gracias.
- MARI PAZ CALVO. Por todo lo que me ha guiado siempre, no sólo como profesora o en asuntos de la propia tesis (como por ejemplo para aprender más sobre integradores splitting), sino también en mi experiencia docente. Muchas gracias por tantas cosas.
- A LOS MIEMBROS DEL DEPARTAMENTO DE MATEMÁTICA APLICADA, especialmente a Víctor Gatón y a Sagrario Carranza. Muchas gracias.
- A MIS ANTIGUOS PROFESORES, especialmente a Antonio Campillo, del que aprendí gran parte del álgebra que sé y quien tanto cariño me ha dado

## VI

desde los dieciocho años, y a Tasio del Barrio, quien dirigió con mucha dedicación mi trabajo de fin de carrera y del que aprendí la teoría de procesos estocásticos. También quisiera guardar un pequeño espacio para personas como Jesús Domínguez, Javier Sanz, Manuel Núñez, Begoña Cano, Isaías Alonso, José y Felipe Cano, Fernando Gómez, Luis Tristán, Carlos Matrán, Ana Reguera, Miguel Ángel Revilla, José Manuel Aroca... de los que tanto aprendí.

- A VARIOS MATEMÁTICOS DE OTRAS UNIVERSIDADES, especialmente a
  - ANDER MURUA. Este trabajo debe mucho a las ideas de Ander. Los cimientos de esta tesis están en una serie de artículos que escribieron conjuntamente él y Jesús. Muchas gracias por sus aclaraciones cuando nos veíamos, por su amabilidad conmigo y por haberme brindado oportunidades magníficas.
  - NAWAF BOU-RABEE por las discusiones que hemos tenido, por su amabilidad cuando le pedía desde el otro lado del océano información y por su interés en los resultados. Muchas gracias.
  - FERNANDO CASAS Y SERGIO BLANES, por las oportunidades que me han brindado y por su amabilidad en los que fueron mis primeros congresos. Muchas gracias.

Y por último, lejos de las matemáticas, a dos personas muy especiales que conocí durante estos tres años que han sido capitales para mí:

- JAVIER MUÉLLEDES. Es imposible describir con palabras todo lo que me has aportado durante estos años. Muchas gracias por tanto. He aprendido muchísimo de ti y estoy muy orgulloso de ser tu amigo. Me alegra enormemente que por casualidad hayamos coincidido. Ojalá hubiésemos coincidido antes.
- MARÍA ÁLVAREZ. Mi profesora particular de asturiano y la mejor compañía posible durante tantas y tantas tardes. Además, muchas gracias por toda tu ayuda siendo mi médico particular. De otra forma, es seguro que muchas cosas no estarían tan bien como lo están ahora. Aún no sé cómo agradecerte toda esa dedicación. Muchísimas gracias por todo tu cariño y por ser como eres.

# Introducción y resumen

El presente trabajo tiene dos objetivos bien diferenciados. Por una parte, busca establecer un marco algebraico que permita describir la combinatoria que atañe a los integradores splitting de ecuaciones diferenciales y, por otra, busca probar la eficacia del citado enfoque a través de la obtención de resultados teóricos referentes a dichos integradores, como son, por ejemplo, las condiciones de orden o las ecuaciones modificadas. Las llamadas *word series*, *series de palabras*, son sin duda un instrumento de análisis extremadamente útil a la hora de analizar integradores splitting para ecuaciones diferenciales estocásticas en el sentido de Ito y en el de Stratonovich. El presente trabajo podría ser visto como una continuación de los trabajos de Murua y Sanz-Serna [41], [42], [44], [51]. La utilización de métodos combinatorios por parte de Butcher con el fin de obtener las condiciones de orden de los métodos Runge-Kutta (cf. [13]) supuso un punto de inflexión en el análisis numérico dando paso a un nuevo paradigma. Butcher, en su teoría, utiliza las llamadas diferenciales elementales (término acuñado por él mismo) con el fin de desarrollar en serie tanto la solución numérica como la solución teórica. En tales desarrollos, las diferenciales elementales aparecen flanqueadas por los pesos elementales: números reales que varían según el integrador pero no con el sistema a integrar. A cada árbol con raíz (grafo conexo no exento de otras ciertas peculiaridades), se le asocia un peso y una diferencial elemental. La propia estructura del árbol permite obtener de forma inmediata la expresión tanto del peso como de la diferencial a partir de ciertas fórmulas. Una de las claves de esta teoría estriba en que las diferenciales elementales, si bien cambian con el propio sistema a integrar, resultan ser comunes tanto para todos los métodos Runge-Kutta como para la propia solución teórica. Este hecho se traduce en que a través de una simple inspección de los pesos elementales asociados a un integrador sería posible obtener directamente, entre otras cosas, las condiciones de orden. Las B-series [28], series formales que combinan diferenciales elementales con coeficientes arbitrarios, son una forma de presentar de forma sistemática el enfoque de Butcher, extendiéndolo incluso a un contexto mucho más general. Un resultado clave en [28] es la célebre regla de composición de B-series. La aplicación de B-series en el análisis numérico es extraordinariamente fértil, no sólo en el campo de la

## VIII

integración geométrica, donde referencias distinguidas son [15], [50], [27], sino también en otros ámbitos, como por ejemplo en la obtención de ecuaciones modificadas, véase [14]. Para aplicaciones de B-series fuera del ámbito del análisis numérico puede consultarse [16], [17]. En [51], Sanz-Serna y Murua analizan en profundidad el desarrollo de tales descubrimientos. Cabe destacar que Burrage y Burrage [12] han analizado a la Butcher errores fuertes y débiles de integradores Runge-Kutta diseñados para ecuaciones diferenciales estocásticas, SDE's.

Las series de palabras encuentran sus orígenes en el llamado *mould calculus*, desarrollado por Écalle (cf. [22], 1981, 1985) con miras a la resolución de ciertos problemas de renormalización en geometría algebraica. La utilización de series de palabras como alternativa a las B-series es extremadamente reciente: [2], [38], [18], [19], [41], [42], [43]. Conviene destacar que el ámbito de aplicabilidad de las series de palabras es bastante más limitado que el de las B-series: los integradores splitting sí forman parte de dicho dominio pero, por contra, los métodos Runge-Kutta en general no. Siempre que sea posible la aplicación de las series de palabras, se recomienda su uso por encima de las B-series. Por un lado, las series de palabras son mucho más compactas que las B-series, y, por otro, su regla de composición (véase el teorema 2.2) es mucho más sencilla que la requerida por las B-series. En [2] el presente autor junto con Sanz-Serna mostraron cómo extender las series de palabras al contexto de las ecuaciones diferenciales estocásticas de Stratonovich, obteniendo las condiciones de orden para integradores splitting.

La importancia de los integradores splitting [7] ha ido *in crescendo* durante los últimos años principalmente debido a su capacidad para explotar la propia estructura del problema a integrar. Dentro del contexto determinista, dejando por un momento de lado a las series de palabras, (principalmente) se tienen dos caminos para el estudio de la consistencia de integradores splitting:

- La técnica más popular, descrita por ejemplo en [49], utiliza la llamada fórmula de *Baker-Campbell-Hausdorff*. Este procedimiento tiene ciertas inconveniencias fruto de la propia complejidad de que rodea a la fórmula BCH (véase [8] y [40]).
- Introducida en [40], encontramos una técnica que paraleliza el trabajo combinatorio de Butcher para el análisis de la consistencia de los métodos Runge-Kutta (puede encontrarse un resumen en [27, Sección III.3]). Tal y como sucede en el trabajo de Butcher, el enfoque está basado en el uso de árboles con raíz. La utilización de B-series resulta ser de gran importancia en esta teoría.

Las series de palabras constituyen un tercer enfoque diferente, siendo precisamente este el objeto de estudio de los próximos capítulos.



Por simplicidad, asumimos que todas las funciones involucradas son lo suficientemente suaves. Por supuesto, si este no fuera el caso, las fórmulas que se presentan tan sólo resultarían ser válidas hasta los términos que involucran las derivadas que existiesen. Las series que se presentan en este trabajo tienen carácter formal y en general no convergen. De ninguna manera esto disminuye su utilidad: al truncar dichas series se obtienen los polinomios de Taylor necesarios para analizar los integradores. Obsérvese que las series de palabras proporcionan un excelente punto de partida para derivar estimaciones rigurosas de los errores, véase e.g. [18], [19] y, específicamente para algoritmos splitting, [41].

## Resultados

Los principales resultados alcanzados son:

- Obtención de las condiciones de orden fuertes y débiles para integradores splitting en los casos de Ito y Stratonovich.
- Descripción algebraica del promedio de observables actuando en integradores splitting en los casos de Ito y Stratonovich.
- Renormalización de SDE's en el caso de Ito y de Stratonovich, incluyendo una renormalización en forma de ODE: la derandomización.
- Obtención de ecuaciones modificadas en el sentido fuerte para integradores splitting en los casos de Ito y Stratonovich.
- Obtención de una prueba con argumentos probabilísticos del isomorfismo de Hoffman entre el shuffle y el quasishuffle, ver [29]. Además, se presenta la construcción de la dualidad estocástica y de un alfabeto auxiliar, proporcionando ambos una descripción completa de la conexión entre las interpretaciones de Ito y Stratonovich de una SDE. Similarmente, se muestra la relación entre el shuffle y el quasishuffle.
- Análisis de un sistema de ecuaciones diferenciales estocásticas de particular importancia, las ecuaciones de Langevin, via series de palabras.
  - Obtención de las condiciones de orden para un amplio número de integradores.
  - Introducción y estudio de nuevos métodos con orden fuerte dos.
  - Comparación entre los desarrollos en series de palabras de los integradores BAOAB y ABOBA, obteniendo evidencias en favor de BAOAB, tal y como ha sido corroborado en [32], [33], [34].

## X

- Un teorema sobre Hamiltonianos modificados e integradores splitting, el cual de nuevo proporciona más razones a favor de BAOAB (y a favor de otros métodos originales).
- Introducción del procedimiento *tuning*: una forma sistemática de mejorar el orden fuerte de convergencia de ciertos integradores splitting. El procedimiento tuning no es exclusivo de las ecuaciones de Langevin, pudiéndose extender a otras ecuaciones diferenciales estocásticas tales que sus campos satisfagan ciertas propiedades.
- Comparación numérica entre varios integradores a la hora de reproducir propiedades fuertes y débiles de la solución de las ecuaciones de Langevin. Se constata que los hechos observados teóricamente se vislumbran con claridad en la práctica.

# Introduction and summary

The present work has two specific goals. On the one hand, to set up an algebraic framework to cater for the needs of split-step stochastic integrators and on the other hand, to prove the usefulness of such a framework by obtaining results such as order conditions and modified equations. The so-called word series turn out to be an extremely useful tool to analyze this kind of integrators in both the Stratonovich and Ito contexts. The present work may be seen as patterned after the work by Murua and Sanz-Serna, see for example [41], [42], [44], [51]. Butcher's pioneer work on the *combinatorics* of the order conditions for Runge-Kutta deterministic integrators (cf. [13]) stated a brand-new branch of numerical analysis. As is well known, in the combinatorial theory developed by Butcher, the numerical and true solution are expanded with the help of vector-valued mappings called elementary differentials. In the expansions, the elementary differentials are weighted by so-called elementary weights. These are real numbers that change with the integrator but are independent of the system being integrated. There is an elementary differential and an elementary weight for each rooted tree and both are easily written down because their structure is a transcription of that of the rooted tree. The elementary differentials change with the differential system being integrated but are common for all Runge-Kutta integrators and also for the true solution; this has important implications because when designing new integrators or comparing different integrators one may focus on the elementary weights. B-series [28], series of elementary differential with arbitrary coefficients, are a way of systematizing Butcher's approach and extending it to more general integrators. A key result in [28] is the rule for composing B-series. B-series have found many applications in numerical analysis, in particular in relation with modified equations [14] and geometric integration [15], [50], [27]. For applications of B-series outside numerical mathematics see [16], [17]. Burrage and Burrage [12] have analyzed à la Butcher weak and strong errors of Runge-Kutta integrators for SDEs. Sanz-Serna and Murua's paper [51] surveys the history of these developments.

Word series have their roots in the so-called mould calculus, developed by Écalle in a series of papers (cf. [22], 1981, 1985) with a view to solving renormalization problems in algebraic geometry. The use of *word series expansions*

as an alternative to B-series is quite recent: [2], [38], [18], [19], [41], [42], [43]. The scope of applicability of word series is narrower than that of B-series: splitting methods may be treated with word series but Runge-Kutta formulas may not. When applicable, word series are more convenient than B-series. They are more compact than B-series and have a composition rule (see Theorem 2.2) much simpler than the recipe used to compose B-series. In [2] the present author jointly with Sanz-Serna formulated several word-basis-based results that showed how to extend deterministic word series to the stochastic context. In particular, the order conditions of split-step integrators for Stratonovich calculus were also stated.

The importance of splitting algorithms [7] has been increasing in recent years, essentially as a consequence of their capability of exploiting the structure of the problem being integrated. In the deterministic case, apart from the word series techniques, there are two main ways of investigating the consistency properties of a splitting integrator:

- The best known technique, described in e.g. [49], applies the *Baker-Campbell-Hausdorff* formula. This method has several shortcomings, including the combinatorial complexity of the BCH formula itself (see [8] and [40] for a discussion).
- An approach that parallels Butcher's treatment of Runge-Kutta formulas has been introduced in [40] (a summary is available in [27, Section III.3]). As in Butcher's work, the approach is based on the use of rooted trees. The *B-series* expansions found in this way are also made of elementary differentials and scalar coefficients.

Word series represent a different approach and that is what we will study in detail throughout the following chapters.

For simplicity all mappings are assumed to be smooth. Of course, when that is not the case, the formulas presented only make sense up to the order where the derivatives that appear exist. The series presented are formal and in general do not converge. This does not diminish their usefulness: by truncating the series one obtains the Taylor polynomials that are needed to analyze the integrators. Note that word series actually provide a good starting point to derive sharp rigorous errors estimates, see e.g. [18], [19] and, specifically for splitting algorithms, [41].

## Results

The main results achieved are:

- Strong and weak order conditions of split-step integrators for both the Stratonovich (which was already published in [2]) and Ito case.

- Algebraic description of the expectation of observables acting on the split-step integrators, including the Stratonovich and Ito case.
- Renormalization of Stratonovich and Ito SDE's, including a random-ordinary-differential-equation normalization: the derandomization.
- Strong modified equations for split-step integrators. Stratonovich and Ito case.
- Probabilistic proofs of Hoffman's quasishuffle-shuffle morphism, see [29]. The construction of the stochastic duality and auxiliary alphabet, which provides a complete description of the Stratonovich and Ito connections with their shuffle-quasishuffle counterparts.
- Analysis of a system of stochastic differential equations of particular importance: the Langevin equations. The following results are of particular interest:
  - Obtention in a systematic way of the order conditions for known methods by means of word series.
  - Introduction and analysis of new methods with high strong order.
  - Comparison between word series expansions of known methods, which yields several reasons in favour of the so-called BAOAB, just like Matthews and Leimkuhler noticed experimentally, see [32], [33], [34].
  - A theorem on modified Hamiltonian dynamics, which provides more reasons in favour of the use of BAOAB and BUB instead of ABOBA and UBU.
  - Introduction of the tuning procedure: a systematic way to enhance the strong order of some split-step integrators by using the additive nature of the random noise. It is worth remarking that it may also be applied to other stochastic differential for which we can find simplifications among its word basis functions.
  - Numerical comparison between the strong and weak performance of some split-step integrators. Several facts noted above are empirically checked.



# Contents

<b>I</b>	<b>Word series for deterministic and Stratonovich calculus</b>	<b>3</b>
<b>1</b>	<b>Combinatorial algebra 1</b>	<b>5</b>
1.1	Motivation . . . . .	5
1.2	Algebraic framework . . . . .	10
1.2.1	Letters . . . . .	10
1.2.2	Algebra structure . . . . .	11
1.2.3	Shuffle product . . . . .	13
1.2.4	Algebra and group-like elements . . . . .	15
1.3	Linear differential equations in $\mathcal{G}_{sh}$ . . . . .	18
1.3.1	Autonomous linear differential equations . . . . .	20
1.3.2	Nonautonomous linear differential equations . . . . .	22
1.3.3	Linear problems with several integrands . . . . .	28
1.4	Letters and dynamical systems . . . . .	32
1.4.1	Multiplicative operators and derivations . . . . .	33
1.5	Word Series . . . . .	36
1.5.1	Composition of word series and action of operators . . . . .	37
1.6	Word series and initial value problems . . . . .	38
<b>2</b>	<b>Words for Stratonovich SDE</b>	<b>41</b>
2.1	Words for Stratonovich SDE's . . . . .	41
2.2	Solution operator . . . . .	44
2.2.1	First approach: Picard Iterations . . . . .	44
2.2.2	Second approach. Word series operator ansatz . . . . .	45
2.2.3	The formal series $J(t; t_0)$ . . . . .	46
2.3	Expectation of the solution . . . . .	47
2.4	Renormalization of SDE's . . . . .	50
2.4.1	Derandomization . . . . .	50
2.4.2	Drift-free SDE . . . . .	55

<b>3</b>	<b>Splitting integrators for Stratonovich SDE's</b>	<b>61</b>
3.1	Expansion of the numerical integrator . . . . .	61
3.1.1	Composition of word series: Shuffle Group . . . . .	61
3.1.2	Solution Operator approach . . . . .	65
3.2	Word series for split-step integrators . . . . .	66
3.2.1	Properties of $\tilde{J}(t_0 + h; t_0)$ and $\mathbb{E}\tilde{J}(t_0 + h; t_0)$ . . . . .	67
3.3	Strong and weak order conditions . . . . .	69
3.3.1	Independent order conditions . . . . .	71
3.4	Error bounds . . . . .	72
3.5	Modified equations . . . . .	74
3.5.1	Derandomization . . . . .	74
3.5.2	Drift-modified SDE's . . . . .	75
3.5.3	Drift-free modified SDE . . . . .	75
<b>II</b>	<b>Word series for Ito calculus</b>	<b>77</b>
<b>4</b>	<b>Combinatorial algebra 2</b>	<b>79</b>
4.1	The Ito extension of an alphabet . . . . .	79
4.1.1	Iterated integrals . . . . .	81
4.1.2	The $\theta$ mapping . . . . .	82
4.1.3	Completeness . . . . .	85
4.2	Products and coproducts on $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$ . . . . .	86
4.2.1	Shuffle product . . . . .	86
4.2.2	Quasishuffle product . . . . .	86
4.3	Algebra and group-like elements . . . . .	88
4.4	Linear differential equations in $\mathcal{G}_{qsh}$ . . . . .	91
4.5	Stochastic duality . . . . .	92
4.5.1	The auxiliary alphabet . . . . .	92
4.5.2	Iterated integrals on $\mathcal{A}^*$ and $\theta$ morphism . . . . .	93
4.5.3	Shuffle $\sqcup^*$ and quasishuffle $\bowtie^*$ . . . . .	95
4.5.4	Stochastic duality theorem . . . . .	96
<b>5</b>	<b>Words for Ito SDE</b>	<b>97</b>
5.1	Extended alphabets for Ito SDE's . . . . .	97
5.2	Word basis operators . . . . .	98
5.2.1	Multiplicative operators and derivations . . . . .	99
5.2.2	Word basis functions and composition rules . . . . .	101
5.3	Solution operator . . . . .	103
5.3.1	The formal series $I(t; t_0)$ . . . . .	106
5.4	Expectation of the solution . . . . .	107



5.5	Ito-Stratonovich equivalence . . . . .	110
5.6	Derandomization . . . . .	112
<b>6</b>	<b>Splitting integrators for Ito SDE's</b>	<b>117</b>
6.1	Numerical solution . . . . .	117
6.1.1	First approach: Quasishuffle Group . . . . .	118
6.1.2	Second approach: Solution Operators . . . . .	121
6.2	Word series for split-step integrators theorem, Ito case . . . . .	121
6.2.1	Properties of $\tilde{I}(t_0 + h; t_0)$ and $\mathbb{E}\tilde{I}(t_0 + h; t_0)$ . . . . .	122
6.3	Strong and weak order conditions . . . . .	123
6.4	Error bounds . . . . .	124
6.5	Modified equations for Ito split-step integrators . . . . .	125
6.5.1	Derandomization . . . . .	125
6.5.2	Drift-modified SDE's . . . . .	126
<b>III</b>	<b>Application of word series to Langevin Dynamics</b>	<b>127</b>
<b>7</b>	<b>Langevin dynamics</b>	<b>129</b>
7.1	Langevin dynamics . . . . .	129
7.2	Split-step algorithms . . . . .	129
7.2.1	ABOBA . . . . .	130
7.2.2	BAOAB . . . . .	131
7.2.3	BUB . . . . .	132
7.2.4	UBU . . . . .	133
7.3	Letters for the Langevin Dynamics . . . . .	134
7.3.1	Word basis functions . . . . .	135
7.3.2	Unidimensional calculations . . . . .	136
7.4	Word series expansions and comparisons . . . . .	138
7.4.1	Word series expansion . . . . .	138
7.4.2	ABOBA and BAOAB comparison . . . . .	141
7.5	Ito-Stratonovich comparison . . . . .	143
7.6	Modified equations . . . . .	145
7.6.1	Drift modified equations . . . . .	145
7.6.2	Perturbed Hamiltonian dynamics . . . . .	148
7.7	Appendix . . . . .	151
<b>8</b>	<b>Tuned methods</b>	<b>153</b>
8.1	Tuning procedure . . . . .	153
8.2	ABOBA and BAOAB tuned . . . . .	154
8.3	Deterministic integrators tuned . . . . .	156

*CONTENTS*

1

8.3.1	Tuning damped Hamiltonian . . . . .	156
8.3.2	Tuning perturbed Hamiltonian dynamics . . . . .	159
8.4	Further considerations . . . . .	160
<b>9</b>	<b>Numerical experiments</b>	<b>161</b>
9.1	Strong errors . . . . .	162
9.2	Weak errors . . . . .	162



# **Part I**

## **Word series for deterministic and Stratonovich calculus**



# Chapter 1

## Combinatorial algebra 1

### 1.1 Motivation

Fix a natural number  $d > 0$  and consider the Euclidean Space  $\mathbb{R}^d$ , which we shall call *space of states*. Differential systems in the space of states may also be interpreted as differential equations describing the evolution of suitably chosen linear operators. Roughly speaking, the main idea is to write formally the flow of the system as the exponential of the Lie derivative of the vector field.

The  $\mathbb{R}$ -algebra  $\mathcal{O} := \mathcal{C}^\infty(\mathbb{R}^d)$  comprises all the smooth  $\mathbb{R}$ -valued functions  $\chi$  defined in the space of states. Sometimes these functions are called *observables*. Recall that the product in this algebra is nothing but the pointwise product. The study of observables leads to another important algebra: the linear endomorphisms of  $\mathcal{O}$ , to be denoted by  $\mathcal{L}(\mathcal{O})$ . The product in this algebra is the composition of operators. Obviously, the unit element is the identity operator  $Id : \chi \rightarrow \chi$ . We shall be concerned with two closely related subsets of  $\mathcal{L}(\mathcal{O})$ :

- **Derivations.** Motivated by the role played by the derivative, we encounter the *derivations*. An element  $D \in \mathcal{L}(\mathcal{O})$  in the algebra is said to be a derivation if

$$D(\chi\chi') = \chi D(\chi') + \chi' D(\chi), \quad (1.1)$$

for any pair of observables  $\chi, \chi'$ . The set of derivations, to be denoted by  $Der(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ , is a linear space: given  $\alpha, \beta \in \mathbb{R}$ ,  $D_1, D_2 \in Der(\mathcal{O})$ ,  $\alpha D_1 + \beta D_2 \in Der(\mathcal{O})$ . It also can be endowed with a *Lie algebra structure* by taking the commutator  $[D_1, D_2] = D_1 D_2 - D_2 D_1 \in Der(\mathcal{O})$  as a product.

- **Algebra automorphisms.** The algebra structure of  $\mathcal{O}$  can be taken into

account by considering the bijections  $T \in \mathcal{L}(\mathcal{O})$  such that <sup>1</sup>

$$T(\chi\chi') = T(\chi)T(\chi') \quad (1.2)$$

for any pair of observables  $\chi, \chi'$ , i.e. the *multiplicative operators*. This set of operators will be denoted by  $Aut(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ . Considering the composition of operators as a product,  $Aut(\mathcal{O})$  may be endowed with a group structure:  $TR(\chi\chi') := TR(\chi)TR(\chi')$ .

Consider the initial value problem in  $\mathbb{R}^d$

$$\frac{d}{dt}x(t) = f(x(t), t), \quad x(0) = x_0, \quad (1.3)$$

with  $f$  smooth. We shall assume that for each time  $t$  and  $x_0 \in \mathbb{R}^d$ , the solution  $x(t)$  exists. The interpretation of the differential system (1.3) in terms of the evolution of linear operators is as follows. Define first two time-dependent families in  $Aut(\mathcal{O})$  and  $Der(\mathcal{O})$ :

- For each frozen  $t$  consider  $\mathbb{F}(t) \in Der(\mathcal{O})$  defined as

$$\mathbb{F}(t)(\chi)(x) = \nabla\chi(x) \cdot f(t, x), \quad (1.4)$$

where  $\nabla\chi(x)$  stands for the differential of  $\chi$  at the point  $x \in \mathbb{R}^d$ . The basic identity  $\nabla(\chi\chi') = \chi\nabla\chi' + \chi'\nabla\chi$  ensures that  $\mathbb{F}(t)$  is a derivation for each  $t$ .

- For each frozen  $t$ , define  $\mathbb{X}(t) \in Aut(\mathcal{O})$  as follows: for any given observable  $\chi \in \mathcal{O}$ ,

$$\mathbb{X}(t)(\chi)(x(0)) := \chi(x(t)), \quad (1.5)$$

for each  $x(0) \in \mathbb{R}^d$ . By construction,  $\mathbb{X}(t)$  is obviously a multiplicative operator.

Focus next on the equality

$$\frac{d}{dt}\chi(x(t)) = \nabla\chi(x(t)) \cdot f(x(t), t),$$

and rewrite it in terms of the former operators:

$$\frac{d}{dt}\mathbb{X}(t)\chi = \mathbb{X}(t)\mathbb{F}(t)(\chi), \quad \mathbb{X}(0)(\chi) = \chi.$$

---

<sup>1</sup>This condition implies that  $T^{-1}$  is also multiplicative.

As the above equation holds for arbitrary observables  $\chi$ , we can omit its dependence on them in order to reach the celebrated interpretation of the initial value problem for a differential equation in terms of differential operators:

$$\boxed{\frac{d}{dt}\mathbb{X}(t) = \mathbb{X}(t)\mathbb{F}(t), \quad \mathbb{X}(0) = Id.} \quad (1.6)$$

This idea is of critical importance in the development of the theory. Observe (1.6) might also be recast in terms of integrals

$$\boxed{\mathbb{X}(t) = Id + \int_0^t \mathbb{X}(s)\mathbb{F}(s)ds.} \quad (1.7)$$

Note that (1.6) is *linear in  $X$  even if (1.3) is not linear in  $x$* .

Let us look at two important particular cases.

- $f(t, x)$  is time independent, i.e. *the autonomous case*. Iterating (1.6), for each  $k \geq 1$ :

$$\frac{d^k}{dt^k}\mathbb{X}(t) = \mathbb{X}(t)\mathbb{F}^k,$$

so that, at least formally (recall that  $\mathbb{X}(0) = Id$ ):

$$\mathbb{X}(t) = \sum_{n=0}^{\infty} \left( \frac{d^n}{dt^n}\mathbb{X}(0) \right) \frac{t^n}{n!} = \sum_{n=0}^{\infty} \frac{(t\mathbb{F})^n}{n!} = \exp(t\mathbb{F}). \quad (1.8)$$

- $f(x, t) = \sum_{a \in \mathcal{A}} \lambda_a(t) f_a(x)$ , for some finite set of indices  $\mathcal{A}$ . If  $D_a$  denotes the Lie derivative corresponding to the vector field  $f_a$ , i.e.

$$D_a \chi = f_a \cdot \nabla \chi,$$

the derivation  $\mathbb{F}(t)$  is

$$\mathbb{F}(t) = \sum_{a \in \mathcal{A}} \lambda_a(t) D_a. \quad (1.9)$$

In this case we can formally solve (1.7) by means of Picard iterations. Plugging the value

$$\mathbb{X}(s) = Id + \int_0^s \mathbb{X}(r)\mathbb{F}(r)dr$$

into equation (1.7), we express the solution operator as follows:

$$\mathbb{X}(t) = Id + \int_0^t \mathbb{X}(s)\mathbb{F}(s)ds = Id + \int_0^t \left( Id + \int_0^s \mathbb{X}(r)\mathbb{F}(r)dr \right) \mathbb{F}(s)ds,$$



i.e.

$$\mathbb{X}(t) = Id + \int_0^t \mathbb{F}(s)ds + \int_0^t \left( \int_0^s \mathbb{X}(r)\mathbb{F}(r)dr \right) \mathbb{F}(s)ds. \quad (1.10)$$

By integrating equality (1.9)

$$\int_0^t \mathbb{F}(s)ds = \sum_{a \in \mathcal{A}} \left( \int_0^t \lambda_a(s)ds \right) D_a,$$

we rewrite (1.10) compactly:

$$\mathbb{X}(t) = Id + \sum_{a \in \mathcal{A}} \left( \int_0^t \lambda_a(s)ds \right) D_a + O(t^2).$$

The differential operator  $\mathbb{X}(t)$  is expressed as a perturbation of the identity operator whose leading term is a combination of the Lie Derivatives of the vector fields  $f_a$ . We may continue to perform these iterations in order to obtain arbitrary high order approximations. To obtain a second order approximation we need only insert

$$\mathbb{X}(r) = Id + \int_0^r \mathbb{X}(u)\mathbb{F}(u)du$$

in (1.10):

$$\begin{aligned} \mathbb{X}(t) &= Id + \sum_{a \in \mathcal{A}} \left( \int_0^t \lambda_a(s)ds \right) D_a + \int_0^t \left( \int_0^s \mathbb{X}(r)\mathbb{F}(r)dr \right) \mathbb{F}(s)ds. \\ &= Id + \sum_{a \in \mathcal{A}} \left( \int_0^t \lambda_a(s)ds \right) D_a + \\ &\quad + \int_0^t \left( \int_0^s \left( Id + \int_0^r \mathbb{X}(u)\mathbb{F}(u)du \right) \mathbb{F}(r)dr \right) \mathbb{F}(s)ds. \end{aligned}$$

Notice the last integral is of the form

$$\int_0^t \left( \int_0^s \mathbb{F}(r)dr \right) \mathbb{F}(s)ds + O(t^3).$$

In particular,

$$\begin{aligned}
\int_0^t \left( \int_0^s \mathbb{F}(r) dr \right) &= \int_0^t \left( \sum_{a \in \mathcal{A}} \left( \int_0^s \lambda_a(r) dr \right) D_a \right) \mathbb{F}(s) ds \\
&= \int_0^t \left( \sum_{a \in \mathcal{A}} \left( \int_0^s \lambda_a(r) dr \right) D_a \right) \sum_{b \in \mathcal{A}} \lambda_b(s) D_b ds \\
&= \int_0^t \sum_{a, b \in \mathcal{A}} \left( \int_0^s \lambda_a(r) dr \right) \lambda_b(s) D_a \circ D_b ds \\
&= \sum_{a, b \in \mathcal{A}} \left( \int_0^t \int_0^s \lambda_a(r) dr \lambda_b(s) ds \right) D_a \circ D_b \\
&= \sum_{w \in \mathcal{A} \times \mathcal{A}} J_w(t) D_w,
\end{aligned}$$

where, for  $w = (a, b) \in \mathcal{A} \times \mathcal{A}$ ,  $D_w = D_a \circ D_b$  (composition of operators) and the coefficients  $J_w(t)$  are nothing but

$$J_w(t) = \int_0^t \left( \int_0^s \lambda_a(r) dr \right) \lambda_b(s) ds.$$

Observe that  $D_w$  is no longer a derivation but a second order differential operator. To sum up, what we have computed is that

$$\mathbb{X}(t) = Id + \sum_{a \in \mathcal{A}} \left( \int_0^t \lambda_a(s) ds \right) D_a + \sum_{w \in \mathcal{A} \times \mathcal{A}} J_w(t) D_w + O(t^3).$$

There is no restriction to stop at any power of  $t^n$ . The same argument also works, although the more steps in the Taylor expansion are taken, the more intricate the combinatorics are. In particular, for  $w = (a_1, a_2, \dots, a_n) \in \mathcal{A}^n$ , defining the *iterated operators*

$$D_w = D_{a_1} \circ D_{a_2} \circ \dots \circ D_{a_n} \quad (1.11)$$

and the *iterated integrals*

$$J_w(t) = \int_0^t \int_0^{s_n} \dots \int_0^{s_2} \lambda_{a_1}(s_1) ds_1 \lambda_{a_2}(s_2) ds_2 \dots \lambda_{a_n}(s_n) ds_n, \quad (1.12)$$

we may write that

$$\mathbb{X}(t) = Id + \sum_{n=1}^{\infty} \sum_{w \in \mathcal{A}^n} J_w(t) D_w, \quad (1.13)$$

so that

$$\chi(x(t)) = \chi(x_0) + \sum_{n=1}^{\infty} \sum_{w \in \mathcal{A}^n} J_w(t) D_w \chi(x_0), \quad (1.14)$$

for any given observable  $\chi$ . In particular, considering the  $d$  canonical projections  $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ , we obtain

$$x(t) = x_0 + \sum_{n=1}^{\infty} \sum_{w \in \mathcal{A}^n} J_w(t) f_w(x_0), \quad (1.15)$$

with  $f_w := (D_w(\pi_1), D_w(\pi_2), \dots, D_w(\pi_d)) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ .

Conversely assume that  $\mathbb{F}(t) : \mathbb{R} \rightarrow \text{Der}(\mathcal{O})$  is such that there exists a one parameter family  $\mathbb{X}(t) : \mathbb{R} \rightarrow \mathcal{L}(\mathcal{O})$  meeting condition (1.7). Then, setting for each  $t$

$$f(t, \cdot) := (\mathbb{F}(t)\pi_1, \mathbb{F}(t)\pi_2, \dots, \mathbb{F}(t)\pi_d),$$

it is easily checked that the initial value problem has, for each  $x_0$ , a solution  $x(t)$  defined for all real  $t$ . Moreover,

$$x(t) = (\mathbb{X}(t)\pi_1, \mathbb{X}(t)\pi_2, \dots, \mathbb{X}(t)\pi_d).$$

Note that we are studying differential/integral equations as the evolution of a certain differential operator which is also multiplicative, i.e. we study the evolution of a curve  $t \mapsto \mathbb{X}(t)$  in a group of multiplicative operators. Even in the case in which these formal constructions cannot be associated to any particular differential operator, the formal solution will be of extreme importance if precise estimates of the “true” solution are to be derived. All these considerations motivate the introduction of the following algebraic concepts.

## 1.2 Algebraic framework

### 1.2.1 Letters

Let  $\mathcal{A}$  be a finite set, which we shall call the *alphabet*. Its elements will be called *letters*. A *word* on  $\mathcal{A}$  is any concatenation of letters, including the so-called empty word, to be denoted by 1. The infinite set of words will be denoted by  $\mathcal{W}$ . We shall identify each letter with the one-letter word comprising that letter so that  $\mathcal{A} \subset \mathcal{W}$ . In particular, any word is of the form

$$w = a_1 a_2 \dots a_n$$

with  $a_i$  letters on  $\mathcal{A}$ . The number  $n$  is referred as the *length* of the word. The length induces a (locally finite) *grading* in the set of words: a word is said to have degree  $n$  if its length is  $n$ . Locally finite means that despite the fact that there is an infinite number of words, only a finite amount of them have degree  $n$ , which is ensured by the finitude of the alphabet.

We shall denote the set of real *polynomials on  $\mathcal{A}$  with non-commutative indeterminates* by  $\mathbb{R}\langle\mathcal{A}\rangle$  and the real *formal series on  $\mathcal{A}$  with non-commutative indeterminates* by  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , see [48]. Obviously  $\mathbb{R}\langle\mathcal{A}\rangle \subset \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ . The above mentioned non-commutative-indeterminate condition implies that polynomials like

$$f = 3a_1a_2 + 3a_2a_1, \quad g = 6a_1a_2,$$

are different. Denoting  $a^n = \underbrace{a \dots a}_{n \text{ times}}$  and  $a^0 = 1$ , the element

$$\sum_{n=0}^{\infty} 3^n a^n$$

is an example of element belonging to  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  but not to  $\mathbb{R}\langle\mathcal{A}\rangle$ . A generic element of either these two sets will be simply written

$$S = \sum_w S_w w,$$

where we are omitting the fact that  $w$  ranges in the set  $\mathcal{W}$ . If  $S \in \mathbb{R}\langle\mathcal{A}\rangle$ , almost all  $S_w$  must be zero.

### 1.2.2 Algebra structure

Both  $\mathbb{R}\langle\mathcal{A}\rangle$  and  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  are vector spaces that inherit the (locally finite) algebra structure with the linear extension of the concatenation as a product. For instance, in the  $\mathbb{R}\langle\mathcal{A}\rangle$  case, given  $S = \sum_w S_w w, R = \sum_w R_w w \in \mathbb{R}\langle\mathcal{A}\rangle$ , the element  $SR \in \mathbb{R}\langle\mathcal{A}\rangle$  is defined by

$$SR = \sum_w (SR)_w w = S_1 R_1 1 + \sum_{w \neq 1} (SR)_w w \in \mathbb{R}\langle\mathcal{A}\rangle,$$

where for a non-empty word  $w = a_1 a_2 \dots a_n$ ,

$$(SR)_w = S_1 R_w + \sum_{j=1}^{n-1} S_{a_1 a_2 \dots a_j} R_{a_{j+1} \dots a_n} + S_w R_1.$$

The product in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is defined in a similar fashion. Occasionally, we may only be interested in the coefficients  $(S_w)_w \in \mathbb{R}^{\mathcal{W}}$  rather than in the series themselves.

The notation  $\mathbb{R}^{\mathcal{W}}$  stands for the direct product  $\prod_{w \in \mathcal{W}} \mathbb{R}$ . The product we are considering in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is readily identified with the so-called *convolution product* (cf. [41]). Consider  $\sigma = (\sigma_w)_w \in \mathbb{R}^{\mathcal{W}}$ , and  $\rho = (\rho_w)_w \in \mathbb{R}^{\mathcal{W}}$ . Then, its convolution product  $\sigma \star \rho \in \mathbb{R}^{\mathcal{W}}$  is such that,

$$\boxed{(\sigma \star \rho)_w = \sigma_1 \rho_w + \sum_{j=1}^{n-1} \sigma_{a_1 a_2 \dots a_j} \rho_{a_{j+1} \dots a_n} + \sigma_w \rho_1} \quad (1.16)$$

for non-empty words  $w = a_1 a_2 \dots a_n$ . In other case,  $(\sigma \star \rho)_1 = \sigma_1 \rho_1$ . Hence, given  $\alpha = (\alpha_w)_w \in \mathbb{R}^{\mathcal{W}}$  and  $\beta = (\beta_w)_w \in \mathbb{R}^{\mathcal{W}}$ ,

$$\boxed{\sum_w \alpha_w w \sum_w \beta_w w = \sum_w (\alpha \star \beta)_w w.} \quad (1.17)$$

The product of series is associative and so is the convolution product. Observe that  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is a local algebra whose maximal ideal is the set

$$\mathfrak{m} = \{S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : S_1 = 0\},$$

i.e.  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is invertible if and only if  $S \notin \mathfrak{m}$ . Obviously  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle/\mathfrak{m} \cong \mathbb{R}$ .

The degree of a non zero polynomial  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is defined as

$$\deg(S) = \sup\{\deg(w) : S_w \neq 0\}.$$

In this way, the algebra  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is graded: its graded component of degree  $n = 0, 1, 2, \dots$  consists of the linear combinations of words with  $n$  letters.

Following [48] we define a bilinear<sup>2</sup> *pairing*  $(, ) : \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \times \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \rightarrow \mathbb{R}$  such that for any pair of words  $u, v$ :

$$(u, v) = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{if } u \neq v \end{cases}. \quad (1.18)$$

Note that for  $S = \sum_w S_w w \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  and any word  $w$ ,  $(S, w) = S_w$ . Given  $f \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , we shall use the notation

$$S_f := (S, f).$$

The pairing allows us to identify  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  with the algebraic dual  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle^*$ : the action of  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  on  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is

$$f \mapsto (f, S)$$

for every  $f \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ . We shall say that the linear mapping  $\Phi : \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is the *transpose* of the linear map  $\phi : \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  if for any two words  $w, u$ :

$$(\phi u, w) = (u, \Phi w). \quad (1.19)$$

<sup>2</sup>It must be understood that bilinear means *infinite bilinear*

### 1.2.3 Shuffle product

Given two words  $w$  and  $w'$  with  $m$  and  $n$  letters respectively, their *shuffle product*  $w \sqcup w' \in \mathbb{R}\langle \mathcal{A} \rangle$  is the formal sum of the  $(m+n)!/(m!n!)$  words with  $m+n$  letters that may be obtained by interleaving the letters of  $w$  and  $w'$  while preserving the order in which the letters appear in  $w$  and  $w'$ . For instance,  $a \sqcup b = ab + ba$ ,  $a \sqcup a = aa + aa = 2aa$ ,  $ab \sqcup c = abc + acb + cab$ ,  $ab \sqcup cd = abcd + acbd + cabd + acdb + cadb + cdab$ . In particular,  $1 \sqcup 1 := 1$ . The locally finite property of  $\mathbb{R}\langle \mathcal{A} \rangle$  allows us to extend the shuffle product to  $\mathbb{R}\langle \langle \mathcal{A} \rangle \rangle$ .<sup>3</sup>

Take words  $u, v, w$  and a letter  $a$ . If

$$(wa, u \sqcup v) \neq 0$$

holds, the definition of the shuffle product guarantees that at least one of  $u, v$ , have  $a$  as its last letter. In particular, for two not necessarily distinct given letters  $x, y$ , the words  $w$  such that  $(w, ux \sqcup vy) \neq 0$  are rearranged in two groups: those whose last letter comes from the word  $ux$ , in which case they are of the form  $w'x$ , and thus  $w'$  satisfies

$$(w', (u \sqcup vy)x) = (w, ux \sqcup vy),$$

and those whose last letter comes from  $vy$ , which are of the form  $w''$  with

$$(w'', (ux \sqcup v)y) = (w, ux \sqcup vy).$$

This combinatorial analysis may be recast in a compact way:

**Lemma 1.1 (Shuffle relation)** *Given two words  $u, v$  and letters  $x, y$ , the shuffle product satisfies the following recurrence:*

$$\boxed{ux \sqcup vy = (ux \sqcup v)y + (u \sqcup vy)x.} \quad (1.20)$$

*In a similar way, the shuffle product also satisfies the recurrence*

$$\boxed{xu \sqcup yv = y(xu \sqcup v) + x(u \sqcup yv).} \quad (1.21)$$

□

The shuffle product possesses a *coproduct*, which will be denoted by  $\delta$ . This coproduct is sometimes called *de-shuffle*. A detailed analysis of this coproduct may be found [48], Chapter 1.  $\delta$  is the homomorphism of  $\mathbb{R}$ -algebras  $\mathbb{R}\langle \mathcal{A} \rangle \rightarrow \mathbb{R}\langle \mathcal{A} \rangle \otimes \mathbb{R}\langle \mathcal{A} \rangle$  mapping each letter  $x$  into

$$\delta x = x \otimes 1 + 1 \otimes x. \quad (1.22)$$

<sup>3</sup>i.e.  $\sqcup : \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle \times \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle \rightarrow \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle$  bilinear, which is nothing but consider a linear mapping  $\sqcup : \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle \otimes \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle \rightarrow \mathbb{R}\langle \langle \mathcal{A} \rangle \rangle$ .

As an algebra homomorphism,  $\delta 1 = 1 \otimes 1$ . For example, given two letters  $x, y$ :

$$\delta xy = \delta x \delta y = (x \otimes 1 + 1 \otimes x)(y \otimes 1 + 1 \otimes y) = xy \otimes 1 + x \otimes y + y \otimes x + 1 \otimes xy.$$

which is symmetric in  $x$  and  $y$ . Similarly

$$\delta(xy - yx) = \delta xy - \delta yx = (xy - yx) \otimes 1 + 1 \otimes (xy - yx).$$

That  $\delta$  is the coproduct of the shuffle product just amounts to saying that for any words  $u, v, w$ ,

$$\boxed{(u \sqcup v, w) = (u \otimes v, \delta w)}, \quad (1.23)$$

where the pairing has been naturally extended to  $\mathbb{R}\langle \mathcal{A} \rangle \otimes \mathbb{R}\langle \mathcal{A} \rangle$ :

$$(u \otimes w, v \otimes w') = \begin{cases} 1 & \text{if } u \otimes w = v \otimes w' \\ 0 & \text{if } u \otimes w \neq v \otimes w' \end{cases}. \quad (1.24)$$

Let us check (1.23) (see [48] for further details). Firstly, assuming that the sum ranges over all pairs of words  $u, v$ , we obtain

$$\delta w = \sum_{u,v} (\delta w, u \otimes v) u \otimes v.$$

We then must focus on proving the following equation

$$\delta w = \sum_{u,v} (w, u \sqcup v) u \otimes v. \quad (1.25)$$

We shall proceed by induction on the length of the word  $w$ . Choose  $w = w'x$  and assume (1.25) to hold for the length of  $w'$ . Then,

$$\delta w = \delta w' \delta x = \left( \sum_{u,v} (w', u \sqcup v) u \otimes v \right) (x \otimes 1 + 1 \otimes x).$$

Thus,

$$\delta w = \sum_{u,v} (w', u \sqcup v) ux \otimes v + \sum_{u,v} (w', u \sqcup v) u \otimes vx.$$

Observe that  $(w', u \sqcup v) = (w'x, ux \sqcup v)$  and that  $(w', u \sqcup v) = (w'x, u \sqcup vx)$ . The shuffle recurrence (1.20) allows us to conclude.  $\square$ .

Both the shuffle product and the shuffle coproduct may be extended to the completion of  $\mathbb{R}\langle \mathcal{A} \rangle \otimes \mathbb{R}\langle \mathcal{A} \rangle$  by infinite linearity, see [48] for further details.

### 1.2.4 Algebra and group-like elements

It is time for us to focus on two important, closely related subset of  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ .

- The **algebra-like elements** (sh stands for shuffle):

$$\mathfrak{g}_{sh} = \{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \delta R = R \otimes 1 + 1 \otimes R\}.$$

They are connected with derivations  $Der(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ . Observe that we are dealing with a linear space: given  $\alpha, \beta \in \mathbb{R}, S, R \in \mathfrak{g}_{sh}$ :

$$\delta(\alpha R + \beta S) = \alpha \delta R + \beta \delta S = (\alpha R + \beta S) \otimes 1 + 1 \otimes (\alpha R + \beta S).$$

Moreover,  $\mathfrak{g}_{sh}$  is endowed with a Lie algebra structure by considering the bracket

$$[S, R] = SR - RS.$$

Recalling that  $\delta$  is an algebra morphism, we readily check that  $[S, R] \in \mathfrak{g}_{sh}$  provided  $S, R \in \mathfrak{g}_{sh}$ :  $\delta[S, R] = [\delta S, \delta R] = [S, R] \otimes 1 + 1 \otimes [S, R]$ , i.e. Sometimes the elements of  $\mathfrak{g}_{sh}$  are called *primitive* or *shuffle-primitive*.

- The **group-like elements** (again, sh stands for shuffle):

$$\mathcal{G}_{sh} = \{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \delta R = R \otimes R\}.$$

This set plays a similar role to that of  $Aut(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ . It has a group structure:  $\delta(SR) = \delta(S)\delta(R)$  so that  $\delta(S)^{-1} = \delta(S^{-1})$ . The existence of  $S^{-1}$  is guaranteed by the condition  $S_1 = 1$ .

Observe that the bracket is bilinear, skew-symmetric ( $[S, R] = -[R, S]$ ) and satisfies the Jacobi identity:

$$[Q, [S, R]] + [S, [R, Q]] + [R, [Q, S]] = 0,$$

just as the commutator of derivations does. We now present two useful results:

**Theorem 1.1 (Characterization of group-like elements)** *Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ :  $S \in \mathcal{G}_{sh}$  if and only if  $S_{u \sqcup v} = S_u S_v$  for all words  $u, v$  and  $S_1 = 1$ .*

**Proof:** On the one hand,

$$\delta S = \sum_{u,v} (\delta S, u \otimes v) u \otimes v = \sum_{u,v} (S, u \sqcup v) u \otimes v = \sum_{u,v} S_{u \sqcup v} u \otimes v,$$

and on the other hand

$$S \otimes S = \sum_{u,v} S_u S_v u \otimes v,$$

so that the claim holds.  $\square$



**Theorem 1.2 (Characterization of algebra-like elements)** *Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ :  $S \in \mathfrak{g}_{sh}$  if and only if  $S_{u \sqcup v} = 0$  for all words  $u, v$  ( $S$  is orthogonal to all shuffles).*

**Proof:** Again, on the one hand

$$\delta S = S_1 1 \otimes 1 + \sum_{u \neq 1} S_u u \otimes 1 + \sum_{v \neq 1} S_v 1 \otimes v + \sum_{u, v \neq 1} S_{u \sqcup v} u \otimes v$$

and on the other,

$$S \otimes 1 + 1 \otimes S = \sum_{u \neq 1} S_u u \otimes 1 + \sum_{v \neq 1} S_v 1 \otimes v + 2S_1 1 \otimes 1.$$

□

Following [48], we define the *exponential mapping*. Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  with  $S_1 = 0$  (i.e.  $S$  lies in the maximal ideal of  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ ), its exponential is defined as

$$\exp(S) = \sum_{n=0}^{\infty} \frac{S^n}{n!} \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle.$$

If  $S \notin \mathfrak{m}$ , we define

$$\exp(S) := e^{S_1} \exp(S - S_1).$$

Observe that every  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  has an exponential. Similarly, given  $R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  with  $R_1 = 1$  we define its *logarithm* as

$$\log(R) = \sum_{n=0}^{\infty} \frac{(R - 1)^n}{n} \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle.$$

If  $R_1 \neq 0$ , we define

$$\log(R) := \log(R_1) + \log(R/R_1).$$

The previous sums are well defined because of the locally finite nature of  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ . Note that  $\exp(S)_1 = e^{S_1}$ ,  $\log(S)_1 = \log(S_1)$ . A formal calculation shows that

$$\exp(\log(S)) = S$$

if  $S_1 = 1$  and

$$\log(\exp(R)) = R$$

if  $R_0 = 0$ , see [48]. For example; consider the series

$$S = 1 + S_a a + S_b b + S_{aa} a^2 + S_{ab} ab + S_{ba} ba + S_{bb} b^2 + O(3),$$

where  $O(3)$  stands for terms with degree greater than or equal to 3. Then,

$$\begin{aligned}
\log(S) &= (S-1) - \frac{(S-1)^2}{2} + \frac{(S-1)^3}{3} + \dots \\
&= S_a a + S_b b + S_{aa} a^2 + S_{ab} ab + S_{ba} ba + S_{bb} b^2 \\
&\quad - \frac{1}{2} (S_a a + S_b b + S_{aa} a^2 + S_{ab} ab + S_{bb} b^2)^2 + \\
&\quad + \frac{1}{3} (S_a a + S_b b + S_{aa} a^2 + S_{ab} ab + S_{bb} b^2)^3 + \dots \\
&= S_a a + S_b b + S_{aa} a^2 + S_{ab} ab + S_{ba} ba + S_{bb} b^2 + \\
&\quad - \frac{1}{2} (S_a^2 a^2 + S_b^2 b^2 + S_a S_b ab + S_a S_b ba) + O(3).
\end{aligned}$$

Rearrange next its terms according to the grading:

$$\begin{aligned}
\log(S) &= S_a a + S_b b + (S_{aa} - \frac{S_a^2}{2}) a^2 + (S_{bb} - \frac{S_b^2}{2}) b^2 + (S_{ab} - \frac{S_a S_b}{2}) ab \\
&\quad + (S_{ba} - \frac{S_a S_b}{2}) ba + O(3).
\end{aligned}$$

Setting  $R := \log(S)$ , we observe that

$$R_{ab} + R_{ba} = R_{ab+ba} = R_{a \sqcup b} = S_{ab+ba} - S_a S_b = S_{a \sqcup b} - S_a S_b.$$

In particular, if  $S \in \mathcal{G}_{sh}$ ,

- $2R_{aa} = R_{a \sqcup a} = 0$ ,
- $2R_{bb} = R_{b \sqcup b} = 0$ ,
- $R_{ab+ba} = R_{a \sqcup b} = 0$ .

The following theorem (cf.[48]) ensures that  $R_{u \sqcup v} = 0$  for any non-empty words  $u, v$ , i.e.  $R \in \mathfrak{g}_{sh}$ . As the exponential is a bijection, the reciprocal statement also holds: if  $R \in \mathfrak{g}_{sh}$ ,  $\exp(R) \in \mathcal{G}_{sh}$ .

**Theorem 1.3** *The results*

$$\begin{aligned}
\exp(\mathfrak{g}_{sh}) &= \mathcal{G}_{sh}, \\
\log(\mathcal{G}_{sh}) &= \mathfrak{g}_{sh}.
\end{aligned}$$

*hold.*

**Proof:** Choose  $S \in \mathfrak{g}_{sh}$ , then

$$\delta \exp(S) = \delta \sum_{n=0}^{\infty} \frac{S^n}{n!} = \sum_{n=0}^{\infty} \frac{(\delta S)^n}{n!} = \exp(\delta S) = \exp(S \otimes 1 + 1 \otimes S).$$

But  $S \otimes 1$  and  $1 \otimes S$  commute so that

$$\begin{aligned} \exp(S \otimes 1 + 1 \otimes S) &= \exp(S \otimes 1) \exp(1 \otimes S) = (\exp(S) \otimes 1)(1 \otimes \exp(S)) \\ &= \exp(S) \otimes \exp(S). \end{aligned}$$

A similar argument is valid for the group-like elements.  $\square$

### 1.3 Linear differential equations in $\mathcal{G}_{sh}$

Consider  $S(t) \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$  such that every  $S_w(t)$  is differentiable. We then define  $S'(t) \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$  as

$$S'(t) := \sum_w S'_w(t)w.$$

In a similar way, if  $S(t)$  is such that  $S_w(t)$  is integrable in an interval  $[a, b]$  for every word  $w$ , we define

$$\int_a^b S(t)dt := \sum_w \left( \int_a^b S_w(t)dt \right) w,$$

which we quite often write in the following short way

$$dS(t) := \sum_w dS_w(t)w.$$

The following result (cf. [41]) leads us to think of  $\mathcal{G}_{sh}$  as a formal Lie group whose Lie algebra is  $\mathfrak{g}_{sh}$

**Theorem 1.4** *The elements of  $\mathcal{G}_{sh}$  coincide with the velocities at the identity  $1 \in \mathcal{G}_{sh}$  of curves in  $\mathcal{G}_{sh}$ .*

**Proof:**

- For any  $R \in \mathfrak{g}_{sh}$ , the curve

$$\gamma(t) = \exp(tR)$$

lies in the group for each  $t \in \mathbb{R}$  and is such that

$$\gamma'(0) = R.$$

- Choose a curve  $\gamma(t) \in \mathcal{G}_{sh}$  such that  $\gamma(0) = 1$ . Then, for any non-empty words  $u, v$ :

$$\gamma_u(t)\gamma_v(t) = \gamma_{u \sqcup v}(t),$$

which leads to the key relation

$$\gamma'_u(t)\gamma_v(t) + \gamma_u(t)\gamma'_v(t) = \gamma'_{u \sqcup v}(t).$$

Set  $t = 0$  above to conclude

$$\gamma'_u(0)\gamma_v(0) + \gamma_u(0)\gamma'_v(0) = 0 = \gamma'_{u \sqcup v}(0).$$

The computations above show that  $\gamma'(0)$  is ortogonal to shuffles, which in tandem with  $\gamma'_1(0) = 0$  (because  $\gamma_1(t) = 1$ ) yields  $\gamma'(0) \in \mathfrak{g}_{sh}$ .

□

The tangent space of the group  $\mathcal{G}_{sh}$  at an element  $S \in \mathcal{G}_{sh}$  is

$$S\mathfrak{g}_{sh} := \{SR \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle, R \in \mathfrak{g}_{sh}\}.$$

Observe that

$$\delta(SR) = (S \otimes S)(R \otimes 1 + 1 \otimes R) = SR \otimes S + S \otimes SR.$$

Reciprocally, for any  $Q \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  lying in the tangent space at  $S$  (in this case  $QS^{-1} \in \mathfrak{g}_{sh}$ ):

$$\begin{aligned} Q \otimes S + S \otimes Q &= QS^{-1}S \otimes S + S \otimes QS^{-1}S \\ &= (QS^{-1} \otimes 1 + 1 \otimes QS^{-1})(S \otimes S) = \delta(QS^{-1}S) = \delta Q. \end{aligned}$$

We have then obtained a characterization of the tangent space:

$$\mathcal{T}_S(\mathcal{G}_{sh}) = \{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \delta R = R \otimes S + S \otimes R\}.$$

It also may be of interest to consider the *following heuristic interpretation*. Define

$$\begin{aligned} \Phi : \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle &\rightarrow \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \otimes \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \\ S &\mapsto \delta S - S \otimes S, \end{aligned}$$

which measures in some sense the *distance of an element  $S$  to the group*. This mapping will allow us to gain insight into some features of the elements of the

group  $\mathcal{G}_{sh}$ . Observe that  $\mathcal{G}_{sh} = \{S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \Phi(S) = 0\}$ . Now, at the point  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , any increment of  $\Phi$  in the “ $H \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  direction” is:

$$\begin{aligned}\Phi(S + H) - \Phi(S) &= \delta(S + H) - \delta(S) - (S + H) \otimes (S + H) + S \otimes S \\ &= \delta H - H \otimes S - S \otimes H - H \otimes H.\end{aligned}$$

Discarding the “quadratic” term  $H \otimes H$ , for each  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  we ought to define the linear mapping

$$\begin{aligned}d\Phi_S : \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle &\rightarrow \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \otimes \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle \\ H &\mapsto \delta H - H \otimes S - S \otimes H\end{aligned}$$

in an attempt to consider the differential of  $\Phi$ . In such a way, the tangent space of  $\mathcal{G}_{sh}$  at the point  $S \in \mathcal{G}_{sh}$  should be the set

$$\{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : d\Phi_S(R) = 0\} = \{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \delta R = R \otimes S + S \otimes R\}.$$

In particular, the tangent space at 1 is

$$\mathfrak{g}_{sh} = \{R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle : \delta R = R \otimes 1 + 1 \otimes R\}.$$

### 1.3.1 Autonomous linear differential equations

Prior to handling the nonautonomous case, we shall be concerned with the simpler linear initial value problem

$$S'(t) = S(t)R, \quad S(0) = X, \quad (1.26)$$

where  $X, R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  and, in particular,  $R \in \mathfrak{m}$ , i.e.  $R_1 = 0^4$ , are given. Although its solution simply reads

$$S(t) = X \exp(tR),$$

we want to stress some important of facts:

**Theorem 1.5 (Autonomous linear initial value problems)** *The following statements hold:*

- *The initial value problem (1.26) has a unique solution.*

---

<sup>4</sup>This particular assumption is not groundless. Practical applications of these “abstract” differential equations often require conditions such that  $R \in \mathfrak{g}_{sh}$ . At this point, for this kind of differential equations in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , there is no such restriction: when considering an arbitrary  $R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , the conclusions of the next theorem are still true, and its proof, with suitable changes in the details, also works.

- The structure of (1.26) itself enables to compute recursively the coefficients of  $S(t)$ .
- Assume  $X \in \mathcal{G}_{sh}$ . Then for each time  $t$ ,  $S(t) \in \mathcal{G}_{sh}$  if and only if  $R \in \mathfrak{g}_{sh}$ .

**Proof:** Let us first solve the differential equation (1.26). Once solved, the first two claims will readily follow.

- For the empty word:

$$S'_1(t) = (S(t)R)_1 = S_1(t)R_1 = 0,$$

so that  $S_1(t) = X_1$ .

- For a letter  $a$ :

$$S'_a(t) = (S(t)R)_a = S_1(t)R_a + S_a(t)R_1 = R_a X_1,$$

i.e.  $S_a(t) = X_a + R_a X_1 t$ .

- Given a non-empty word  $w = a_1 a_2 \dots a_n$ ,

$$\begin{aligned} S'_w(t) = (S(t)R)_w &= S_{a_1 a_2 \dots a_{n-1}}(t)R_{a_n} + S_{a_1 a_2 \dots a_{n-2}}(t)R_{a_{n-1} a_n} + \dots \\ &+ S_1(t)R_{a_1 a_2 \dots a_n}, \end{aligned}$$

which is again a linear differential equation in  $S_w(t)$  with initial condition  $S_w(0) = X_w$ .

For the third statement it is enough to look at the formula

$$S(t) = X \exp(tR).$$

If  $S(t) \in \mathcal{G}_{sh}$ , then

$$S(t)X^{-1} = \exp(tR) \in \mathcal{G}_{sh},$$

i.e.  $R \in \mathfrak{g}_{sh}$ . Conversely, if  $R \in \mathfrak{g}_{sh}$ ,  $\exp(tR) \in \mathcal{G}_{sh}$  and the result holds.  $\square$

We are not interested in cases in which  $R \notin \mathfrak{m}$ , although the same results of the preceding theorem also holds.

### 1.3.2 Nonautonomous linear differential equations

Consider now the nonautonomous initial value problem (cf.[41], although the formulation there relies on elements  $\alpha \in \mathbb{R}^{\mathcal{W}}$ )

$$\boxed{S'(t) = S(t)R(t), \quad S(0) = 1,} \quad (1.27)$$

with  $R(t) \in \mathfrak{m} \subset \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  (i.e.  $R_1(t) = 0$ ) given and  $S(t) \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  unknown. Again, the structure of the problem itself delineates a profound connection between the group and the algebra: the condition  $S(t) \in \mathcal{G}_{sh}$  clearly looks related to  $R(t) \in \mathfrak{g}_{sh}$ . Indeed, assume  $t \mapsto S(t)$  is a curve in the group. Then, its velocity at time  $t$ , simply reading  $S'(t) = S(t)R(t)$ , must be an element of the group tangent space at the point  $S(t)$ , which is nothing but an algebra element translated by  $S(t)$ . This makes us think that  $R(t) \in \mathfrak{g}_{sh}$ . And conversely, recall the mapping (1.26) measures the “de-shuffle-deviation” of an element  $S$  from the group. We would like to examine under which conditions  $\Phi(S(t))$  is constant. Observe that

$$\Phi(S(0)) = 1 \otimes 1 - 1 \otimes 1 = 0,$$

so that we are seeking conditions under which  $\Phi(S(t)) = 0 \leftrightarrow S(t) \in \mathcal{G}_{sh}$ . Compute next heuristically  $(\Phi S(t))'$ :

$$(\Phi S(t))' = d\Phi_{S(t)}(S'(t)) = d\Phi_{S(t)}(S(t)R(t)),$$

which is equal to (ommiting the time dependence):

$$\delta(SR) - S \otimes SR - SR \otimes S = \delta S \delta R - (S \otimes S)(R \otimes 1 + 1 \otimes R).$$

As a result:  $S(t) \in \mathcal{G}_{sh}$  if and only if  $R(t) \in \mathfrak{g}_{sh}$ .

The next theorem (cf. [41]) puts these arguments on a solid mathematical footing. We decided to include the proof because it will surely shed some lighth on the way we tackle similar future results.

**Theorem 1.6 (Nonautonomous linear initial value problems)** *The following statements hold:*

- *The initial value problem (1.27) has a unique solution.*
- *The structure of (1.27) itself makes it possible to compute recursively the coefficients of  $S(t)$ .*
- *Assume  $R(t) \in \mathfrak{g}_{sh}$ . Then  $S(t) \in \mathcal{G}_{sh}$  for all  $t \in \mathbb{R}$ .*

**Proof:** The argument given in the autonomous case is enough to prove the first two statements. As distinct from the autonomous case, we do not have an explicit formula for the solution and therefore we have to work additionally to obtain a proof for the third item.

**Lemma 1.2** *Assume that  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is such that, for some positive integer  $n$ , and for each pair of words  $u, v$  such that the sum of their lengths, say  $l, m$ , satisfy  $l + m \leq n$ , it holds that:*

$$S_u S_v = S_{u \sqcup v}$$

Then, for  $R \in \mathfrak{g}_{sh}$  and words  $u, v$  with lengths  $l, m$  such that  $l + m \leq n + 1$ :

$$(SR)_{u \sqcup v} = S_u (SR)_v + (SR)_u S_v$$

The proof of the lemma is as follows: the element

$$\exp(S) = \sum_w c_w w$$

is ortogonal to the shuffles  $u \sqcup v$  for words  $u, v$  with lengths  $l + m \leq n$ . In particular, the polynomial

$$X = \sum_{w:|w|\leq n} c_w w \in \mathbb{R}\langle\mathcal{A}\rangle$$

is ortogonal to all the shuffles by construction. Recall that  $|w|$  stands for the length of  $w$ . In particular  $X \in \mathfrak{g}_{sh}$  so that

$$\hat{S} = \log(X) \in \mathcal{G}_{sh}$$

and coincides with  $S$  up to words of length  $n$ . Now, choose a curve  $\gamma(t) \in \mathcal{G}_{sh}$  such that  $\gamma'(0) = R$ . Then

$$(\hat{S}\gamma(t))_{u \sqcup v} = (\hat{S}\gamma(t))_u (\hat{S}\gamma(t))_v$$

for any pair of words  $u, v$ , thus

$$(\hat{S}R)_{u \sqcup v} = (\hat{S}R)_u \hat{S}_v + (\hat{S}R)_v \hat{S}_u.$$

Now take words  $u, v$  with lengths  $l + m \leq n + 1$ . We may assume that both of them are non-empty (in other case the result holds). As  $R \in \mathfrak{g}_{sh}$ ,  $R_1 = 0$  and therefore

$$(\hat{S}R)_{u \sqcup v} = (SR)_{u \sqcup v},$$

and  $\hat{S}_u = S_u, \hat{S}_v = S_v$ . In this case

$$(SR)_{u \sqcup v} = (SR)_u S_v + (SR)_v S_u$$



holds, as desired.

Now, an induction argument on  $n$  will show that  $S(t)$  satisfies the shuffle relations up to words of length  $n$ : for non-empty words  $u, v$  with lengths  $l + m \leq n$ ,

$$\partial_t \left( S_{u \sqcup v}(t) - S_u(t)S_v(t) \right) = (S(t)R(t))_{u \sqcup v} - S'_u(t)S_v(t) - S_u(t)S'_v(t).$$

The lemma implies that the last quantity vanishes, hence the function

$$\varphi(t) = S_{u \sqcup v}(t) - S_u(t)S_v(t)$$

is constant and  $\varphi(0) = 0$  because  $S(0) = 1$ , and the result holds.  $\square$

We have seen that the condition  $R(t) \in \mathfrak{g}_{sh}$  for all  $t$  implies that  $S(t) \in \mathcal{G}_{sh}$  for all  $t$ . The converse is also true: if  $S(t) \in \mathcal{G}_{sh}$  for all  $t$ , then  $R(t) \in \mathfrak{g}_{sh}$  for all  $t$ . But the time-dependent nature of  $R(t)$  is enough to make some simplifications.

**Theorem 1.7** *For the initial value problem (1.27) the following statements hold:*

- *If  $S(t) \in \mathcal{G}_{sh}$  for all  $t$ , then  $R(t) \in \mathfrak{g}_{sh}$  for all  $t$ .*
- *If  $R(t)$  is constant, then  $S(t) \in \mathcal{G}_{sh}$  for all  $t$  if and only if there exist  $h > 0$  such that  $S(h) \in \mathcal{G}_{sh}$ . In particular,  $R(t) = R \in \mathfrak{g}_{sh}$ .*
- *If  $R(t) \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle[t]$ , i.e. there exist  $R_0, R_1, \dots, R_n \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$  such that*

$$R(t) = R_0 + R_1 t + \dots + R_n t^n,$$

*then  $S(t) \in \mathcal{G}_{sh}$  for all  $t$  if and only if there exist  $n + 1$  positive times  $t_0, t_1, \dots, t_n$  such that  $S(t_i) \in \mathcal{G}_{sh}$ ,  $0 \leq i \leq n$ . In particular  $R(t) \in \mathfrak{g}_{sh}$  for all  $t$ .*

- *If  $R(t) = Z(t)RZ^{-1}(t)$  for some  $Z(t) \in \mathcal{G}_{sh}$ , then  $S(t) \in \mathcal{G}_{sh}$  for all  $t$  if and only if there exist  $h > 0$  such that  $S(h) \in \mathcal{G}_{sh}$ .*

**Proof:** Consider first the initial value problem (1.27):

$$S'(t) = S(t)R(t), \quad S(0) = 1,$$

with  $R(t) \in \mathfrak{m}$  and assume that  $S(t) \in \mathcal{G}_{sh}$  for all  $t$ . We shall prove that

$$R_{u \sqcup v}(t) = 0 \tag{1.28}$$

for non-empty words  $u, v$  inductively on  $n = |u| + |v|$ . If  $n = 2$ , it is easy to check that

$$R_{ab}(t) = -R_{ba}(t)$$

for each pair of letters  $a, b$ , thus the result holds. Assume (1.28) holds for  $n \geq 2$ . Take the  $n$ -degree truncation of  $R$ , i.e.

$$\hat{R}(t) = \sum_{w:|w|\leq n} R_w(t)w.$$

For words  $u, v$  such that  $|u| + |v| \geq n + 1$ ,

$$\hat{R}_{u\sqcup v}(t) = 0.$$

If  $|u| + |v| < n + 1$ ,

$$\hat{R}_{u\sqcup v}(t) = R_{u\sqcup v}(t) = 0,$$

so that  $\hat{R}(t) \in \mathfrak{g}_{sh}$ . Consider now the problem

$$\hat{S}'(t) = \hat{S}(t)\hat{R}(t), \quad \hat{S}(0) = 1.$$

Theorem 1.6 ensures that  $\hat{S}(t) \in \mathcal{G}_{sh}$  for all  $t$ . Observe that  $\hat{S}_w(t) = S_w(t)$  for words with  $|w| \leq n$  because  $\hat{R}_w(t) = R_w(t)$ . Choose  $w \in \mathcal{W}$  such that  $|w| = n + 1$ . Then, from the equations

$$\hat{S}'_w(t) = (\hat{S}(t)\hat{R}(t))_w,$$

$$S'_w(t) = (S(t)R(t))_w,$$

we deduce that

$$(S(t) - \hat{S}(t))'_w = R_w(t) - \hat{R}_w(t),$$

but  $\hat{R}_w(t) = 0$  and so

$$R_w(t) = (S(t) - \hat{S}(t))'_w,$$

or

$$S_w(t) - \hat{S}_w(t) = \int_0^t R_w(s)ds.$$

Observe that, as both  $S(t)$  and  $\hat{S}(t)$  satisfy the shuffle relations, the left-hand side of the previous equation only depends on words of length  $< n + 1$ , and so does the right-hand side. Let us take advantage of this: choose non-empty words  $u, v \in \mathcal{W}$  such that  $|u| + |v| = n + 1$ . Observe now that

$$R_{u\sqcup v}(t) = (S_{u\sqcup v}(t) - S_u(t)S_v(t))',$$

which also may be written in an integral way

$$S_{u\sqcup v}(t) - S_u(t)S_v(t) = \int_0^t R_{u\sqcup v}(s)ds,$$

because of the linearity of the shuffle and the fact that  $\hat{S}(t) \in \mathcal{G}_{sh}$  so that it satisfies the shuffle relations  $\hat{S}_{u \sqcup v}(t) = \hat{S}_u(t)\hat{S}_v(t)$  and coincides with  $S(t)$  up to words of length  $n$ . In particular, if  $S(t) \in \mathcal{G}_{sh}$  for all  $t$ , the function  $t \mapsto R_{u \sqcup v}(t)$  vanishes identically.

Assume now that  $R(t)$  is constant. By repeating the same argument, we see that the function

$$\varphi(t) := S_{u \sqcup v}(t) - S_u(t)S_v(t) = \int_0^t R_{u \sqcup v}(s)ds = tR_{u \sqcup v}$$

is linear in  $t$ . If there exists  $h > 0$  such that  $S(h) \in \mathcal{G}_{sh}$ , then  $\varphi(h) = 0$  so that  $R_{u \sqcup v} = 0$  as desired.

The same argument is valid in the case in which  $R(t)$  is a polynomial in  $t$ , thus the third statement holds.

Assume now that  $R(t) = Z(t)RZ^{-1}(t)$ , with  $Z(t) \in \mathcal{G}_{sh}$  for all  $t$ . We may repeat the argument used for the first statement with a new different  $\hat{R}(t)$ : truncate  $R$  instead of  $R(t)$  up to degree  $n$ , i.e.

$$\hat{R} = \sum_{w:|w| \leq n} R_w w \in \mathfrak{g}_{sh},$$

and then set  $\hat{R}(t) = Z(t)\hat{R}Z^{-1}(t)$ . Repeat the same argument and consider the problem

$$\hat{S}'(t) = \hat{S}(t)\hat{R}(t), \quad \hat{S}(0) = 1.$$

Obviously  $\hat{S}(t) \in \mathcal{G}_{sh}$  for all  $t$  and  $\hat{S}_w(t) = S_w(t)$  for words with  $|w| \leq n$ . Take  $w \in \mathcal{W}$  such that  $|w| = n + 1$  to conclude that

$$R_w(t) - \hat{R}_w(t) = (S(t) - \hat{S}(t))'_w.$$

By construction,  $R_w(t) - \hat{R}_w(t) = R_w - \hat{R}_w = R_w$ , thus

$$R_w = (S(t) - \hat{S}(t))'_w.$$

Now,

$$\varphi(t) = S_{u \sqcup v}(t) - S_u(t)S_v(t) = tR_w.$$

Again,  $\varphi$  is linear in  $t$ . Thus, one value of time  $h > 0$  such that  $\varphi(h) = 0$  is enough to ensure that  $\varphi = 0$ , as desired.  $\square$

The preceding results allows us to provide the following analysis of boundary value-value problem in  $\mathcal{G}_{sh}$ .

**Theorem 1.8 (Boundary value problems)** Consider the autonomous two-point boundary value problem

$$S'(t) = S(t)R \quad (1.29)$$

- i)  $S(t_0) = 1$ ,
- ii)  $S(t_0 + h) = X$ , for some  $h > 0$  and  $X \in \mathcal{G}_{sh}$ ,

with only  $X$  given. Then:

- There exists a unique choice of  $S(t)$  for all  $t$  and  $R$  such that (4.21) holds. In particular  $S(t) \in \mathcal{G}_{sh}$  for all  $t$  and  $R \in \mathfrak{g}_{sh}$ .
- The structure of (4.21) makes it possible to compute recursively the coefficients of  $S(t)$ .
- The unique solution is given explicitly by

$$S(t) = \exp((t - t_0)R), \quad R = \frac{1}{h} \log(X).$$

**Proof:** For simplicity, we shall assume that  $t_0 = 0$ . Firstly, let us solve the boundary value problem:

- For the empty word:

$$S'_1(t) = (S(t)R)_1 = S_1(t)R_1$$

which yields  $S_1(t) = e^{tR_1}$ . As  $S_1(h) = 1$ ,  $R_1 = 0$  necessarily. Hence,  $S_1(t) = 1$ .

- For a letter  $a$ :

$$S'_a(t) = (S(t)R)_a = S_1(t)R_a + S_a(t)R_1 = R_a,$$

i.e.  $S_a(t) = R_a t$ . Now,  $S_a(h) = R_a h = X_a$ , so  $S_a(t) = \frac{t}{h} X_a$  and  $R_a = X_a/h$ .

- Given a non-empty word  $w = a_1 a_2 \dots a_n$ ,

$$S'_w(t) = (S(t)R)_w = S_{a_1 a_2 \dots a_{n-1}}(t)R_{a_n} S_{a_1 a_2 \dots a_{n-2}}(t)R_{a_{n-1} a_n} + \dots + R_{a_1 a_2 \dots a_n}.$$

We check straightforwardly that  $S'_w(t) = tq(t) + R_w$  with  $q(t) \in \mathbb{R}[t]$  known. In particular

$$S_w(t) = S_w(0) + Q(t) + R_w t$$

for a suitable known polynomial  $Q(t) \in \mathbb{R}[t]$  with degree higher than 1. The condition  $S_w(h) = X_w$  allows us to obtain both  $R_w$  and  $S_w(t)$

Observe that

$$S(t) = \exp((t - t_0)R), \quad R = \frac{1}{h} \log(X)$$

satisfy the differential equation, whose solution is unique. Hence

$$R \in \mathfrak{g}_{sh}, \quad S(t) \in \mathcal{G}_{sh},$$

for all time  $t$ .

□

### 1.3.3 Linear problems with several integrands

Assume that we have functions  $x_1(t), x_2(t), \dots$  such that the integral

$$f \mapsto \int_0^t f(s) dx_i(s)$$

is defined under some assumptions on the integrand  $f$  and satisfies the product rule<sup>5</sup>

$$\int f(s) dx(s) = x(s)f(s) - \int x(s) df(s). \quad (1.30)$$

Assume also that  $x_i(0) = 0$ . For example,  $x(t) = [t]$ , i.e. the integer part of  $t$ . In such a way, for  $f$  continuously differentiable

$$\int_0^t f(s) dx(s) = \sum_{0 \leq n \leq t} f(n),$$

or  $x(t) = t$ , which provides the integration with respect to the Lebesgue measure or  $x(t) = \mathcal{B}(t)$  where  $\mathcal{B}$  stands for a Brownian motion and the integral is defined in the Stratonovich sense, (see [45], [35]). The Stratonovich integration, to be denoted by  $\circ d\mathcal{B}(t)$  satisfies the product rule. For example

$$\int \mathcal{B}(s) \circ d\mathcal{B}(s) = \mathcal{B}^2(s) - \int \mathcal{B}(s) \circ d\mathcal{B}(s),$$

---

<sup>5</sup>This is consideration is the one that motivates the use of the shuffle. More light will be shed on this idea. See for example Theorem 1.15 and its proof.

so that

$$\int \mathcal{B}(s) \circ d\mathcal{B}(s) = \frac{\mathcal{B}^2(s)}{2}.$$

We shall be concerned with initial value problems of the form

$$\boxed{dS(t) = \sum_{i=1}^n S(t)R^{(i)}(t)dx_i(t), \quad S(0) = 1.} \quad (1.31)$$

Recall that this notation is a shorthand for the integral equation

$$S(t) = 1 + \sum_{i=1}^n \int_0^t S(s)R^{(i)}(s)dx_i(s).$$

**Theorem 1.9 (Nonautonomous multiintegrand initial value problems)** *The following statements hold:*

- *The initial value problem (1.31) has a unique solution.*
- *The structure of (1.31) makes it possible to compute recursively the coefficients of  $S(t)$ .*
- *Assume  $R(t) \in \mathfrak{g}_{sh}$ . Then  $S(t) \in \mathcal{G}_{sh}$  for all  $t \in \mathbb{R}$ .*

**Proof:**

Observe that Theorem 1.6 is nothing but the case  $n = 1$ . The road map to prove all the claims is essentially the same, but let us check the third statement. Firstly, by using the arguments of Lemma 1.2, prove the following natural extension:

**Lemma 1.3** *Assume that  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is such that, for some positive integer  $n$ , and for each pair of words  $u, v$  such that their lengths  $l, m$  satisfy  $l + m \leq n$ , it holds that:*

$$S_u S_v = S_{u \sqcup v}.$$

*Then, for  $R^{(i)} \in \mathfrak{g}_{sh}$ ,  $1 \leq i \leq n$  and words  $u, v$  with lengths  $l, m$  such that  $l + m \leq n + 1$ :*

$$\sum_{i=1}^n (SR^{(i)})_{u \sqcup v} dx_i = \sum_{i=1}^n (S_u (SR^{(i)})_v dx_i + (SR^{(i)})_u S_v dx_i).$$

Consider next the element

$$\exp(tS) = \sum_w c_w w$$

and truncate it as follows:

$$X = \sum_{w:|w|\leq n} c_w w \in \mathfrak{g}_{sh}.$$

Again,

$$\hat{S} = \log(X) \in \mathcal{G}_{sh}$$

coincides with  $S$  up to words of length  $n$ . Now, for each  $1 \leq i \leq n$  choose a curve  $\gamma_i(t) \in \mathcal{G}_{sh}$  such that  $d\gamma_i(0) = R^{(i)} dx_i$ , for instance

$$\gamma_i(t) = \exp(R^{(i)} x_i(t)).$$

Then

$$\sum_{i=1}^n (\hat{S}\gamma_i(t))_{u \sqcup v} = \sum_{i=1}^n (\hat{S}\gamma_i(t))_u (\hat{S}\gamma_i(t))_v$$

for any pair of words  $u, v$ , so that

$$\sum_{i=1}^n (\hat{S}R^{(i)})_{u \sqcup v} dx_i = \sum_{i=1}^n (\hat{S}R^{(i)})_u \hat{S}_v dx_i + \sum_{i=1}^n (\hat{S}R^{(i)})_v \hat{S}_u dx_i.$$

Now take non-empty words  $u, v$  with lengths  $l + m \leq n + 1$ . As  $R^{(i)} \in \mathfrak{g}_{sh}$ ,  $R_1^{(i)} = 0$

$$\sum_{i=1}^n (\hat{S}R^{(i)})_{u \sqcup v} dx_i = \sum_{i=1}^n (SR^{(i)})_{u \sqcup v} dx_i.$$

and  $\hat{S}_u = S_u, \hat{S}_v = S_v$ . Thus

$$\sum_{i=1}^n (SR^{(i)})_{u \sqcup v} dx_i = \sum_{i=1}^n (SR^{(i)})_u S_v dx_i + \sum_{i=1}^n (SR^{(i)})_v dx_i S_u$$

holds, as desired.

Again, an induction argument on  $|u| + |v| = n$  will show that  $S_{u \sqcup v}(t) = S_u(t)S_v(t)$ . Assume that  $S(t)$  satisfies the shuffle relations up to total length  $n$ . Then, for non-empty words such that  $|u| + |v| = n + 1$  define

$$\varphi(t) := S_{u \sqcup v}(t) - S_u(t)S_v(t).$$

Since  $\varphi(0) = 0$ , it is enough to check that  $d\varphi = 0$ , but

$$d\varphi(t) = dS_{u \sqcup v}(t) - S_u(t)dS_v(t) - dS_u(t)S_v(t).$$

The structure itself of the initial value problem

$$dS(t) = \sum_{1 \leq i \leq n} S(t)R^{(i)}(t)dx_i(t),$$

allows us to conclude by using the lemma.  $\square$

Let us handle autonomous boundary value problems in the several integrands case. Observe that the following theorem generalizes Theorem 1.8.

**Theorem 1.10 (Boundary value problems with several integrands)** *Assume that  $R^{(1)}, R^{(2)}, \dots, R^{(M-1)} \in \mathfrak{g}_{sh}$  and  $X \in \mathcal{G}_{sh}$  are given. Then, for any election of  $x_i(t)$   $1 \leq i \leq M$  and  $h > 0$ , there exists a unique  $R^{(M)} \in \mathfrak{g}_{sh}$  and a curve in the group  $t \mapsto S(t)$  such that*

$$dS(t) = \sum_{i=1}^M S(t)R^{(i)}dx_i(t) \quad (1.32)$$

and

$$i) S(0) = 1,$$

$$ii) S(h) = X.$$

Moreover, the structure of (4.21) makes it possible to compute recursively the coefficients of  $S(t)$  and  $R$ .

**Proof:** Solving the boundary value problem (4.21) recursively in the length  $n = |w|$ , we obtain both  $S(t)$  and  $R$ . We only have to prove that  $S(t) \in \mathcal{G}_{sh}$  and  $R \in \mathfrak{g}_{sh}$ . As stated in the last theorems, it is enough to check that  $R \in \mathfrak{g}_{sh}$ . Let  $t \mapsto Z(t)$  the (unique) solution of the initial value problem

$$dZ(t) = \sum_{i=1}^{M-1} Z(t)R^{(i)}dx_i(t) \quad (1.33)$$

with  $Z(0) = 1$ . We already know that  $Z(t) \in \mathcal{G}_{sh}$  for all  $t$ . Consider  $t \mapsto Y(t)$  defined as

$$S(t) = Y(t)Z(t), \quad Y(t) = S(t)Z^{-1}(t).$$

Now, the product rule of the functions  $x_i(t)$  ensures that

$$dS(t) = Y(t)dZ(t) + dY(t)Z(t),$$

i.e.

$$dS = \sum_{i=1}^{M-1} Y(t)Z(t)R^{(i)}dx_i(t) + dY(t)Z(t) = \sum_{i=1}^{M-1} S(t)R^{(i)}dx_i(t) + dY(t)Z(t).$$



Thus,

$$dY(t) = S(t)R^{(M)}Z^{-1}(t)dx_i(M) = Y(t)Z(t)R^{(M)}Z^{-1}(t)dx_i(t).$$

If we set  $R(t) := Z(t)R^{(M)}Z^{-1}(t)$ ,  $Y(t)$  satisfies the following initial value problem:

$$dY(t) = Y(t)R(t)dx_i(t),$$

- $Y(0) = 1$ .
- $Y(h) = S(h)Z^{-1}(h) = XZ^{-1}(h) \in \mathcal{G}_{sh}$ .

The argument of the fourth paragraph of Theorem 1.7 may be used to check that  $Y(t) \in \mathcal{G}_{sh}$ , so that  $R(t) \in \mathfrak{g}_{sh}$ . In particular,  $R \in \mathfrak{g}_{sh}$ .  $\square$

## 1.4 Letters and dynamical systems

Fix  $d > 0$  integer and let  $\mathcal{A}$  be a finite alphabet. Assume that for each letter  $a \in \mathcal{A}$  there exist a smooth map  $f_a : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Consider the  $\mathbb{R}$ -algebra morphism

$$\begin{aligned} D : \mathbb{R}\langle \mathcal{A} \rangle &\rightarrow \mathcal{L}(\mathcal{O}) \\ p &\mapsto D_p \end{aligned} \tag{1.34}$$

defined on each letter  $l$  by

$$D_l \chi = \nabla \chi \cdot f_l = \chi'(f_l) = \sum_{i=1}^d f_l^i \frac{\partial \chi}{\partial x^i}.$$

Obviously  $\chi'$  denotes the differential of  $\chi$ , and so on, i.e.  $\chi''$  is the second differential, etc. Observe that  $D_a$  is the *Lie derivative* of the vector field  $f_a$ . In particular, for each letter  $a$ ,  $D_a$  is not only a linear operator but also a derivation. The algebra structures in  $\mathbb{R}\langle \mathcal{A} \rangle$  and  $\mathcal{L}(\mathcal{O})$  allow to calculate  $D_w$  for any word  $w = a_1 a_2 \dots a_n$  with length greater than 1:

$$D_w = D_{a_1} \circ D_{a_2} \circ \dots \circ D_{a_n}. \tag{1.35}$$

As a composition of linear operators,  $D_w$  is also a linear operator, but it is no longer a derivation. For example, consider  $w = ab$ ,

$$D_w \chi = D_{ab} \chi = D_a(D_b \chi) = D_a(\chi'(f_b)) = \chi''(f_a, f_b) + \chi'(f'_b(f_a)),$$

which in general is not a first order differential operator. It is worth noting that the second order term, i.e.  $\chi''(f_a, f_b)$ , is invariant under the swapping  $a \leftrightarrow b$ . In such a way, the operator

$$D_{ab-ba}\chi = D_{ab}\chi - D_{ba}\chi = \chi'(f'_a(f_b)) - \chi'(f'_b(f_a)) = \chi'(f'_a(f_b) - f'_b(f_a)),$$

is not only anti-symmetric, but also a derivation. Moreover, observe that it is the Lie derivative of the vector field

$$[f_a, f_b] = f'_a f_b - f'_b f_a.$$

The algebra morphism  $D$  may be extended from  $\mathbb{R}\langle\mathcal{A}\rangle$  to  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  by infinite linearity given rise to *formal differential operators*, i.e. formal operators of the form

$$D_S = \sum_w S_w D_w, \quad D_S(\chi) = \sum_w S_w D_w(\chi),$$

with  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  and  $\chi \in \mathcal{O}$ . The set of formal differential operators acting on observables will be denoted by  $\mathcal{FL}(\mathcal{O})$ . In the same way, formal derivations are denoted by  $\mathcal{FDer}(\mathcal{O})$  and formal automorphism by  $\mathcal{FAut}(\mathcal{O})$ .

### 1.4.1 Multiplicative operators and derivations

The subsequent lines are devoted to studying the decomposition of  $D_w(\chi\psi)$ , as a sum of products of operators of the form  $D_u$  acting on the observables  $\chi, \psi$  themselves.

**Theorem 1.11 (Leibniz's Rule, shuffle case)** *Given any polynomial  $f \in \mathbb{R}\langle\mathcal{A}\rangle$  and any two observables  $\chi, \psi$ :*

$$D_f(\chi\psi) = \sum_{u,v} (\delta f, u \otimes v) D_u \chi D_v \psi = \sum_{u,v} (f, u \sqcup v) D_u \chi D_v \psi.$$

**Proof:** We shall only be concerned with the case in which  $f$  is a word. Linearity will handle the rest. Let  $a$  be a letter. Obviously

$$\delta a = a \otimes 1 + 1 \otimes a,$$

and

$$D_a(\chi\psi) = \chi D_a \psi + \psi D_a \chi = \sum_{u,v} (\delta a, u \otimes v) D_u \chi D_v \psi,$$

which suggests we perform a recursion argument on the length of  $w$ . Assume that  $w = aw'$  and that the theorem holds for  $w'$ . Now,

$$D_{aw'}(\chi\psi) = D_a \left( \sum_{u,v} (w', u \sqcup v) D_u \chi D_v \psi \right) = \sum_{u,v} (w', u \sqcup v) D_a \left( D_u \chi D_v \psi \right),$$

but  $D_a$  is a derivation so that

$$D_a(D_u\chi D_v\psi) = D_{au}\chi D_v\psi + D_u\chi D_{av}\psi.$$

We find thus the equality

$$D_{aw'}(\chi\psi) = \sum_{u,v} (w', u \sqcup v) (D_{au}\chi D_v\psi + D_u\chi D_{av}\psi), \quad (1.36)$$

Observe that the sum

$$\sum_{u,v} (w', u \sqcup v) (au \otimes v + u \otimes av),$$

is a shorthand for (1.36) (we are omitting the reference to the  $D$ 's and  $\chi, \psi$ ). Plugging the straightforward equality  $(w', u \sqcup v) = (aw', au \sqcup v) = (aw', u \sqcup av)$  into the last sum, we obtain

$$\sum_{u,v} (aw', au \sqcup v) au \otimes v + \sum_{u,v} (aw', u \sqcup av) u \otimes av. \quad (1.37)$$

By using the same argument as that of the shuffle recurrence (1.20), we may write (1.37) in a different way

$$\sum_{u,v} (aw', u \sqcup v) u \otimes v = \sum_{u,v} (aw', au \sqcup v) au \otimes v + \sum_{u,v} (aw', u \sqcup av) u \otimes av,$$

which is nothing but what we were looking for:

$$D_{aw'}(\chi\psi) = \sum_{u,v} (aw', u \sqcup v) D_u(\chi) D_v(\psi),$$

and the result is proved.  $\square$

An infinite linearity argument yields the following corollary:

**Corollary 1.1** *Given any series  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  and any two observables  $\chi, \psi$ :*

$$D_S(\chi\psi) = \sum_{u,v} (\delta S, u \otimes v) D_u\chi D_v\psi = \sum_{u,v} (S, u \sqcup v) D_u\chi D_v\psi.$$

**Proof:** Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ ,

$$\begin{aligned} D_S(\chi\psi) &= \sum_w S_w D_w(\chi\psi) = \sum_w S_w \sum_{u,v} (\delta w, u \otimes v) D_u\chi D_v\psi \\ &= \sum_{u,v} \left( \sum_w S_w (\delta w, u \otimes v) \right) D_u\chi D_v\psi \\ &= \sum_{u,v} \left( \sum_w S_w \delta w, u \otimes v \right) D_u\chi D_v\psi \\ &= \sum_{u,v} (\delta \sum_w S_w w, u \otimes v) D_u\chi D_v\psi = \sum_{u,v} (\delta S, u \otimes v) D_u\chi D_v\psi, \end{aligned}$$

and the result is proved.  $\square$

The last corollary may be interpreted as follows. Take any  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ . Then

$$\delta S = \sum_{i,j} c_{i,j} S_i \otimes S_j$$

for some  $c_{i,j} \in \mathbb{R}$ , if and only if, for any pair of observables  $\chi, \psi$ :

$$D_S(\chi\psi) = \sum_{i,j} c_{i,j} D_{S_i}\chi D_{S_j}\psi.$$

In particular, if  $D_S$  is multiplicative, i.e.

$$D_S(\chi\psi) = D_S(\chi)D_S(\psi).$$

we deduce that

$$\delta S = S \otimes S,$$

i.e.  $S$  is a group-like element. If a derivation,

$$D_S(\chi\psi) = D_S(\chi)\psi + D_S(\psi)\chi,$$

we get the decomposition

$$\delta S = S \otimes 1 + 1 \otimes S,$$

i.e.  $S$  is an algebra-like element. These results are stated in the following theorem:

**Theorem 1.12** *Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ :*

- *The formal operator  $D_S$  is a formal automorphism of  $\mathcal{L}(\mathcal{O})$  if and only if  $S \in \mathcal{G}_{sh}$ .*
- *The formal operator  $D_S$  is a formal derivation of  $\mathcal{L}(\mathcal{O})$  if and only if  $S \in \mathfrak{g}_{sh}$ .*

$\square$

The following table emphasises some facts of the *dictionary correspondence* deliberately created by the algebraic framework we have set up.

$\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$	$\mathcal{FL}(\mathcal{O})$
Product of series	Composition of operators
$S \in \mathcal{G}_{sh}$	$D_S \in \mathcal{FAut}(\mathcal{O})$
$S \in \mathfrak{g}_{sh}$	$D_S \in \mathcal{FDer}(\mathcal{O})$

## 1.5 Word Series

Not only will we often be concerned with formal series of operators  $D_S$ , but also with formal series of functions. This kind of series, key in the numerical analysis of some integrators, will be called *word series*. Word series expansions were recently suggested by Murua and Sanz-Serna as an alternative to B-series to handle several problems in dynamical systems, see for example [41], [44], [42], [43], [51], [39]. Apart from formal series  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , the main ingredients of word series expansions are the so called *word basis functions*. Let  $Id$  denote the identity function on the space of states  $x \in \mathbb{R}^d \mapsto x \in \mathbb{R}^d$ . Then, for each word  $w$  we define the word basis function  $f_w : \mathbb{R}^d \rightarrow \mathbb{R}^d$  as

$$\boxed{f_w := D_w(Id)}, \quad (1.38)$$

where  $D_w(Id)$  is understood as acting component-wise. Sometimes it will be of particular importance to consider

$$\boxed{f_p := D_p(Id)}, \quad (1.39)$$

for polynomials  $p$  of positive degree. Formal series in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  are indexed by words. As in the operator case, this is the cue for us to define *word series*.

**Definition 1.1 (Word Series (cf. [41]))** *Given  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , we define its word series as the formal mapping  $\mathbb{R}^d \rightarrow \mathbb{R}^d$*

$$W_S(x_0) = \sum_w S_w f_w(x_0). \quad (1.40)$$

Observe that we may employ a simple recursion to construct all the word basis functions. Suppose  $w = xu \in \mathcal{W}$  for a letter  $x$  and a non-empty word  $u$ . Then, the concatenation recursion  $D_w = D_x \circ D_u$  provides the recurrence

$$f_w = D_x(f_u) = f'_u(f_x). \quad (1.41)$$

For example:

$$\begin{aligned} f_{ab} &= f'_b(f_a), \\ f_{abc} &= f'_{bc}(f_a) = f''_b(f_a, f_c) + f'_b(f'_c(f_a)). \end{aligned}$$

If  $f_a$  is constant,

$$f_{wa} = 0$$

for each word  $w$ . Observe that, if  $w = 1$ , i.e.  $w$  is the empty word, its word basis function is defined as

$$f_1(x) = x.$$

### 1.5.1 Composition of word series and action of operators

**Definition 1.2** (*h-coherence*) Assume that a curve  $h \mapsto S(h) \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is given. We say that the curve is *h-coherent* with respect to the grading if  $S_1$  does not depend on  $h$  and there exists  $N \in \mathbb{Z}$  such that, for every word  $w$ ,

$$h^N S_w = O(h^{|w|}), \quad h \rightarrow 0^+.$$

**Theorem 1.13 (Operator acting on word series)** Choose  $h \mapsto S(h) \in \mathcal{G}_{sh} \subset \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  *h-coherent*. Then, for any  $\chi \in \mathcal{O}$  and  $x_0 \in \mathbb{R}^d$ :

$$\boxed{D_{S(h)}(\chi)(x_0) = \chi(W_{S(h)}(x_0))}$$

for  $h$  small enough.

**Proof:** We are only concerned with the formal case so that we shall only prove the above relation in case  $\chi$  is a polynomial.

$$W_{S(h)} = D_{S(h)}(Id).$$

Thus,

$$\pi_j(W_{S(h)}) = D_{S(h)}(\pi_j),$$

where  $\pi_j$  is the  $j$ -projection. Now,

$$\pi_j \pi_k(W_{S(h)}) = \pi_j(W_{S(h)}) \pi_k(W_{S(h)}) = D_{S(h)}(\pi_j) D_{S(h)}(\pi_k) = D_{S(h)}(\pi_j \pi_k),$$

because  $D_{S(h)}$  is multiplicative. Then, linearity shows that for every polynomial  $p \in \mathcal{O}$ ,

$$p(W_{S(h)}) = D_{S(h)}(p),$$

as desired.  $\square$

Note that the same result also holds when observables  $\chi$  are replaced by smooth mappings  $\mathbb{R}^d \rightarrow \mathbb{R}^d$ .

**Theorem 1.14 (Composition of word series)** Choose  $h \mapsto S(h) \in \mathcal{G}_{sh}$  and  $h \mapsto R(h) \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  *h-coherent*. Then,

$$\boxed{W_{R(h)}(W_{S(h)}(x_0)) = W_{S(h)R(h)}(x_0)}.$$

**Proof:** The result is an immediate corollary of the preceding theorem. For convenience, we shall drop the  $h$  dependence from the righthand side of the following equation:

$$\begin{aligned} W_{R(h)}(W_{S(h)}(x_0)) &= \sum_w R_w f_w(W_S(x_0)) = \sum_w R_w D_S(f_w)(x_0) = \\ &= \sum_w R_w \sum_u S_u D_u(f_w)(x_0) = \sum_{u,w} R_w S_u f_{uw}(x_0) = \\ &= \sum_{w'} (SR)_{w'} f_{w'}. \end{aligned}$$

□

A similar rule exists for B-series, but the recipe involves a product far more intricate than the series product in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , see [27].

## 1.6 Word series and initial value problems

Let us go back to the initial value problem (1.3):

$$\frac{d}{dt}x(t) = f(x(t), t), \quad x(t_0) = x_0.$$

Assume that

$$f(t, x) = \sum_{a \in \mathcal{A}} \lambda_a(t) f_a(x),$$

where the alphabet  $\mathcal{A}$  is finite. The analysis we made previously can be now reformulated in terms of the algebraic framework that we have developed. Firstly, for every non-empty word  $w = a_1 a_2 \dots a_n$ , we consider the *iterated integrals*

$$J_w(t; t_0) = \int_{t_0}^t \int_{t_0}^{s_n} \dots \int_{t_0}^{s_2} \lambda_{a_1}(s_1) ds_1 \lambda_{a_2}(s_2) ds_2 \dots \lambda_{a_n}(s_n) ds_n \quad (1.42)$$

and

$$J_1(t; t_0) = 1.$$

The formal series

$$J(t; t_0) = \sum_w J_w(t; t_0) w \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$$

is sometimes called the *Chen series*, (cf. [20]). Similarly, we construct the word basis operators  $D_w$  based on the vector fields  $\{f_a\}_{a \in \mathcal{A}}$ . Above, the solution of

(1.3) was described via formula (1.13). Now, it may be written compactly as the word series<sup>6</sup>

$$\mathbb{X}(t) = \sum_w J_w(t) D_w = D_{J(t)}.$$

Furthermore, the solution (1.15)

$$x(t) = x_0 + \sum_{w \neq 1} J_w(t) f_w(x_0),$$

is nothing but the word series corresponding to  $J(t)$ :

$$x(t) = W_{J(t)}(x_0).$$

Note the following result holds:

**Theorem 1.15** For each  $t \geq 0$ ,

$$t \mapsto J(t) = \sum_w J_w(t) w \in \mathcal{G}_{sh}.$$

**Proof:**<sup>7</sup> Let us check that

$$J_w J_{w'} = J_{w \sqcup w'}.$$

It suffices an induction argument on  $N = |u| + |v|$  to prove the claim, but first let us drop the time dependency on  $J$ . Take words  $u, v$  and letters  $a, b$  such that  $|ua| + |vb| = N + 1$  and assume the statement to hold for  $N$ . Now, use the following calculus-based argument:

$$J_{ua} J_{vb} = \int d(J_{ua} J_{vb}) = \int J_{ua} dJ_{vb} + \int J_{vb} dJ_{ua} = \int J_u J_{vb} \lambda_a dt + \int J_{ua} J_v \lambda_b dt,$$

i.e.

$$J_{ua} J_{vb} = \int J_{u \sqcup vb} \lambda_a dt + \int J_{ua \sqcup v} \lambda_b dt = J_{(u \sqcup vb)a + (ua \sqcup v)b}.$$

This is the shuffle recursion (1.20):

$$ua \sqcup vb = (ua \sqcup v)b + (u \sqcup vb)a.$$

Thus,

$$J_{ua} J_{vb} = J_{ua \sqcup vb},$$

as desired.  $\square$

<sup>6</sup>For notational convenience we omit the  $t_0$  dependence

<sup>7</sup>The present theorem links the shuffle product with the product rule of differential calculus.



For instance, take  $p = 3A - 2a$ ,  $q = 2AA$ ,  $p \sqcup q = 18AAA - 4aAA - 4AaA - 4AAa$  and  $J_p = 3J_A - 2J_a$ ,  $J_q = 2J_{AA}$ . Then,

$$(3J_A - 2J_a)2J_{AA} = 18J_{AAA} - 4J_{aAA} - 4J_{AaA} - 4J_{AAa}.$$

Observe that  $t \mapsto J(t)$  is also  $t$ -coherent. Furthermore, in this case Theorem 2.1 just amounts to saying that:

$$\chi(W_{J(t)}(x_0)) = \chi(x(t)) = \mathbb{X}(t)(\chi)(x_0) = D_{J(t)}(\chi)(x_0).$$

# Chapter 2

## Words for Stochastic Differential Equations. The Stratonovich case

### 2.1 Words for Stratonovich SDE's

In this section we shall be concerned with the Stratonovich stochastic initial value problem in  $\mathbb{R}^d$  (see [45], [35]):

$$dx(t) = \sum_{a \in \mathcal{A}_{det}} f_a(x(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(x(t)) \circ d\mathcal{B}_A(t), \quad x(t_0) = x_0, \quad (2.1)$$

where

- The sets  $\mathcal{A}_{det}$  (det stands for *deterministic*) and  $\mathcal{A}_{sto}$  (sto stands for *stochastic*) are disjoint. We also assume that both of them are finite.
- For each  $A \in \mathcal{A}_{sto}$ ,  $\mathcal{B}_A(t)$  is a Brownian motion. All of them are defined on the same filtered probability space and are mutually independent.
- The symbol  $\circ$  denotes the Stratonovich interpretation.
- The vector fields  $f_i, l \in \mathcal{A}_{det} \cup \mathcal{A}_{sto}$  are sufficiently smooth

The finite set  $\mathcal{A} = \mathcal{A}_{det} \cup \mathcal{A}_{sto}$  will play the role of alphabet. This motivation is often stressed by saying that we are dealing with an *SDE-based alphabet*. The letters in  $\mathcal{A}_{sto}$  are called *stochastic*. On the other hand, the letters in  $\mathcal{A}_{det}$  are called *deterministic*. Again, the sets of polynomials  $\mathbb{R}\langle \mathcal{A} \rangle$  and series  $\mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$  will be the key to further developments. Instead of the length of the words, we shall use a different grading  $\|\cdot\|$  called the *stochastic weight* or simply the *weight*. This is based on the fact

$$\mathbb{E}\mathcal{B}_A^2(t) = t$$

for each stochastic letter  $A$  where  $\mathbb{E}$  stands for expectation. The stochastic weight of a letter  $l$  is defined as

- If  $l \in \mathcal{A}_{det}$ ,  $\|l\| = 1$ .
- If  $l \in \mathcal{A}_{sto}$ ,  $\|l\| = \frac{1}{2}$ .

The weight of words is defined in a similar way to what we did in the deterministic case:

**Definition 2.1 (Stochastic weight)** *Given a word  $w = a_1 a_2 \dots a_n$ , its stochastic weight is defined as*

$$\|w\| = \sum_{i=1}^n \|a_i\| \in \frac{\mathbb{Z}}{2}.$$

As an example, take  $\mathcal{A} = \{a, A\}$  with  $a$  deterministic and  $A$  stochastic. Then, the degrees of the components of an element  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  are

$$S = \underbrace{S_1}_0 + \underbrace{S_A A}_{0.5} + \underbrace{S_a a + S_{A^2} A^2}_1 + \underbrace{S_{aA} aA + S_{Aa} Aa + S_{A^3} A^3}_{1.5} + \dots$$

Obviously, *all the constructions that we made in the general case*, such as the group  $\mathcal{G}_{sh}$  or the Lie algebra  $\mathfrak{g}_{sh}$ , or simply the shuffle product  $u \sqcup v$  or the exponential mapping *may be particularized to this alphabet*. We shall be interested in the algebra morphism (1.34), i.e.

$$\begin{aligned} D : \mathbb{R}\langle\mathcal{A}\rangle &\rightarrow \mathcal{L}(\mathcal{O}) \\ p &\mapsto D_p \end{aligned}$$

which is defined for each letter  $l$  by

$$D_l \chi = \nabla \chi \cdot f_l = \chi'(f_l) = \sum_{i=1}^d f_l^i \frac{\partial \chi}{\partial x^i}.$$

and then extended to words by composition of operators, etc. Recall that *word basis functions* were defined as  $f_w := D_w(Id)$  with the possibility of generalize them to polynomials  $p$ , where  $D_w(Id)$  is understood as acting component-wise and  $Id : x \mapsto x$  is the identity map of the space of states  $\mathbb{R}^d$ . If  $w = lu$  for a letter  $l$  and a non-empty word  $u$ , the concatenation recursion  $D_w = D_l \circ D_u$  yields the useful recurrence

$$f_w = D_l(f_u) = f'_u(f_l). \quad (2.2)$$

Recall that we defined

$$f_1(x) = x.$$

In many cases we need to consider elements  $S \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  whose coefficients  $S_w$  are random variables, all of them defined in the same probability space. We shall call these elements *random series*.

From now on, we shall assume that the concept of  $h$  coherence refers to the weight grading rather than to the length grading.

**Definition 2.2 ( $h$ -coherence)** *Assume that a non-necessarily deterministic curve  $h \mapsto S(h) \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  is given. We say that the curve is  $h$ -coherent with respect to the weight grading (or simply w.r.t the weight) if  $S_1$  does not depend on  $h$  and there exists  $N \in \mathbb{Z}/2$ , such that for every word  $w$ ,*

$$\boxed{h^N \mathbb{E}|S_w| = O(h^{\|w\|}), \quad h \rightarrow 0^+.$$

If  $S$  is not random  $\mathbb{E}|S_w| = |S_w|$  so that the concept coincides with that defined previously, see Definition 1.2.

**Proposition 2.1** *Assume that  $X$  in Theorem 1.8 is  $h$ -coherent. Then  $R$  is also  $h$ -coherent with  $N = 1$ .*

**Proof:** Observe that

$$hR = \log(X), \quad S(t) = \exp((t - t_0)R),$$

so that the claim holds. Another approach is to compute recursively the coefficients of both  $R$  and  $S(t)$  and use an induction argument on the length of the words.  $\square$

To prove the following theorem we proceed similarly.

**Proposition 2.2** *Assume that  $h \mapsto X(h)$  in Theorem 1.10 is  $h$ -coherent. If  $x_M(t) = t$ , then  $R^{(M)}(h)$  is also  $h$ -coherent with  $N = 1$ . In case of  $x_M(t) = \mathcal{B}_A(t)$ ,  $R^{(M)}(h)$  is  $h$ -coherent with  $N = 1/2$ .*

The following statements are also true.

**Theorem 2.1 (Operators acting on word series)** *Choose  $S = S(h) \in \mathcal{G}_{sh} \subset \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$   $h$ -coherent w.r.t. the weight. Then, for any  $\chi \in \mathcal{O}$  and  $x_0 \in \mathbb{R}^d$ :*

$$\boxed{D_S(\chi)(x_0) = \chi(W_S(x_0))}.$$

**Theorem 2.2 (Composition of word series)** *Take  $S, R \in \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  with  $S \in \mathcal{G}_{sh}$  and both of them  $h$ -coherent w.r.t. the weight. Then,*

$$\boxed{W_{R(h)}(W_{S(h)}(x_0)) = W_{S(h)R(h)}(x_0)}.$$

## 2.2 Solution operator

We next present two approaches to compute the pullback corresponding to the solution operator  $\mathbb{X}(t)$  of the initial value problem (2.3):

$$dx(t) = \sum_{a \in \mathcal{A}_{det}} f_a(x(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(x(t)) \circ d\mathcal{B}_A(t), \quad x(t_0) = x_0. \quad (2.3)$$

### 2.2.1 First approach: Picard Iterations

Fortunately, the Stratonovich calculus follows the same rules as the deterministic infinitesimal calculus, so that the same steps we took in the deterministic case may be taken if the solution operator is to be analyzed. As all computations required have been done already, the path is quite straightforward. Indeed, let us begin defining the iterated integrals, which turn out to be the so-called *Stratonovich iterated integrals*:

$$\begin{aligned} J_1(t; t_0) &= 1, \\ J_a(t; t_0) &= \int_{t_0}^t ds = t - t_0, \quad a \in \mathcal{A}_{det}, \\ J_A(t; t_0) &= \int_{t_0}^t \circ d\mathcal{B}_A(s) = \mathcal{B}_A(t) - \mathcal{B}_A(t_0), \quad A \in \mathcal{A}_{sto}, \\ J_{wa}(t, t_0) &= \int_{t_0}^t J_w(s; t_0) ds, \quad a \in \mathcal{A}_{det}, \\ J_{wA}(t; t_0) &= \int_{t_0}^t J_w(s; t_0) \circ d\mathcal{B}_A(s), \quad A \in \mathcal{A}_{sto}. \end{aligned}$$

Repeating the Picard iterations we performed in the first chapter, we readily check that

$$\mathbb{X}(t) = \sum_w J_w(t; t_0) D_w = D_{J(t; t_0)}. \quad (2.4)$$

Thus

$$x(t) = \mathbb{X}(t)(Id) = D_{J(t; t_0)}(Id) = \sum_w J_w(t; t_0) f_w(x_0) = W_{J(t; t_0)}(x_0). \quad (2.5)$$

Notice that the following important result holds:

**Theorem 2.3** *The Stratonovich iterated stochastic integrals belong to the shuffle group, i.e. for each event  $\omega$ ,*

$$t \mapsto \sum_w J_w(t; t_0) w \in \mathcal{G}_{sh}.$$

**Proof:**<sup>1</sup> Firstly, let us drop the time dependence from the  $J$ 's. We are going to check that

$$J_w J_{w'} = J_{w \sqcup w'}. \quad (2.6)$$

Again, a simple induction argument on  $N = |u| + |v|$  is enough to conclude. Take words  $u, v$  and letters  $x, y$  such that  $|ux| + |vy| = N + 1$  and assume the statement to hold for  $N$ . For simplicity, if the letter  $l$  is deterministic, we set  $d\mathcal{B}_l(t) := dt$ . Then,

$$\begin{aligned} J_{ux} J_{vy} &= \int d(J_{ux} J_{vy}) = \int J_{ux} dJ_{vy} + \int J_{vy} dJ_{ux} \\ &= \int J_u J_{vy} \lambda_x(t) \circ d\mathcal{B}_x(t) + \int J_{ux} J_v \lambda_y(t) \circ d\mathcal{B}_y(t), \end{aligned}$$

which just amounts to saying that

$$J_{ua} J_{vb} = \int J_{u \sqcup vb} \lambda_a \circ d\mathcal{B}_a(t) + \int J_{ua \sqcup v} \lambda_b \circ d\mathcal{B}_b(t) = J_{(u \sqcup vb)a + (ua \sqcup v)b}.$$

Invoking once again the shuffle recursion (1.20), we check that equality holds in (2.6) as desired.  $\square$

Obviously  $h \mapsto J(t_0 + h; t_0)$  is also  $h$ -coherent w.r.t. the stochastic weight for every  $h > 0$ ,  $\omega \in \Omega$  and  $t_0 \geq 0$ .

### 2.2.2 Second approach. Word series operator ansatz

The idea is to use the ansatz

$$\mathbb{X}(t) = \sum_w S_w(t; t_0) D_w = D_{S(t; t_0)}.$$

Given an observable  $\chi$ ,

$$d\chi(x(t)) = \chi'(dx(t)) = \sum_{a \in \mathcal{A}_{det}} \chi'(f_a(x(t))) dt + \sum_{A \in \mathcal{A}_{sto}} \chi'(f_A(x(t))) \circ d\mathcal{B}_A(t).$$

But *this is easily translated into words language:*

$$d\mathbb{X}(t) = \sum_{a \in \mathcal{A}_{det}} \mathbb{X}(t) D_a dt + \sum_{A \in \mathcal{A}_{sto}} \mathbb{X}(t) D_A \circ d\mathcal{B}_A(t). \quad (2.7)$$

---

<sup>1</sup>We shall give the same proof as the one we gave in Chapter 1. This is because both Stratonovich and deterministic calculus satisfy the same rules, tailored to the shuffle product.

Set  $\mathbb{X}(t) = D_{S(t;t_0)}$ . Then,

$$dS(t; t_0) = \sum_{a \in \mathcal{A}_{det}} S(t; t_0) a dt + \sum_{A \in \mathcal{A}_{sto}} S(t; t_0) A \circ d\mathcal{B}_A(t), \quad (2.8)$$

with  $S(0) = 1$ . Observe that  $\mathcal{A} \subset \mathfrak{g}_{sh}$  because the equality

$$\delta x = x \otimes 1 + 1 \otimes x,$$

holds for each letter  $x$ . According to Theorem 1.6, there exists a unique mapping  $t \mapsto S(t; t_0) \in \mathcal{G}_{sh}$  solving (2.8). As it is easily checked that  $J(t; t_0)$  is a solution, we have then found a different way to prove that

$$J(t; t_0) \in \mathcal{G}_{sh}.$$

For convenience, we phrase this important result in a theorem.

**Theorem 2.4 (Stratonovich solution)** *The solution operator of (5.4) admits a word basis operator expansion; in particular, it may be identified with the curve  $t \mapsto J(t; t_0) \in \mathcal{G}_{sh}$  via*

$$\mathbb{X}(t) = D_{J(t;t_0)}.$$

### 2.2.3 The formal series $J(t; t_0)$

In view of the previous sections, the formal series  $J(t; t_0) \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$  is enough to describe the solution of the equation (2.3) so that it would be advisable to study it carefully. Although some of the following results have already been stated, it is useful to present all of them in a compact way.

**Proposition 2.3** *The iterated Stratonovich integrals  $J_w(t; t_0)$  possess the following properties.*

- $J(t; t_0) \in \mathcal{G}_{sh}$  for each event  $\omega$ .
- The joint distribution of any finite subfamily of the family of random variables  $\{h^{-\|w\|} J_w(t_0 + h; t_0)\}_{w \in \mathcal{W}}$  is independent of  $t_0 \geq 0$  and  $h > 0$ .
- $\mathbb{E} | J_w(t_0 + h; t_0) |^p < \infty$ , for each  $w \in \mathcal{W}$ ,  $t_0 \geq 0$ ,  $h > 0$  and  $p \in [0, \infty)$ .
- For each  $w \in \mathcal{W}$  and any finite  $p \geq 1$ , the ( $t_0$ -independent)  $L^p$  norm of the random variable  $J_w(t_0 + h; t_0)$  is  $\mathcal{O}(h^{\|w\|})$ , as  $h \downarrow 0$ .
- $\mathbb{E} J_w(t_0 + h; t_0) = 0$  whenever  $\|w\|$  is not an integer.

**Proof:** The first assertion has already been proved in two different ways. See [5], [6] for the last four items. In the following section we shall prove some of them with different techniques.  $\square$

## 2.3 Expectation of the solution

We are not only interested in *strong* properties of the solution, but also in its weak features. We shall also consider the following differential operator:

$$\mathbb{E}\mathbb{X}(t)\chi(x_0) := \mathbb{E}^{x_0}\chi(x(t)) = \mathbb{E}(\chi(x(t))|x(t_0) = x_0). \quad (2.9)$$

The above equation may be rewritten in terms of word basis operators:

$$\mathbb{E}\mathbb{X}(t) = \mathbb{E} \sum_w J_w(t; t_0) D_w = \sum_w \mathbb{E} J_w(t; t_0) D_w = D_{\mathbb{E}J(t; t_0)}.$$

Consider the formal series

$$\mathbb{E}J(t; t_0) = 1 + (t - t_0) \left( \sum_{a \in \mathcal{A}_{det}} a + \frac{1}{2} \sum_{A \in \mathcal{A}_{sto}} A^2 \right) + O(2),$$

and then set

$$\mathfrak{f} := \sum_{a \in \mathcal{A}_{det}} a + \frac{1}{2} \sum_{A \in \mathcal{A}_{sto}} A^2. \quad (2.10)$$

The above series are related via the following theorem:

**Proposition 2.4** *The formal series  $\mathbb{E}J(t; t_0)$  possesses the following properties;*

- If  $\mathcal{A}_{sto}$  is not empty,  $\mathbb{E}J(t; t_0) \notin \mathcal{G}_{sh}$ , i.e.  $\mathbb{E}\mathbb{X}(t)$  is not a multiplicative operator.
- Kolmogorov Backward equation:

$$\mathbb{E}J(t; t_0) = \exp((t - t_0)\mathfrak{f}), \quad (2.11)$$

i.e.

$$\mathbb{E}\mathbb{X}(t) = \exp((t - t_0)D_{\mathfrak{f}}), \quad (2.12)$$

where

$$D_{\mathfrak{f}} = \sum_{a \in \mathcal{A}_{det}} D_a + \frac{1}{2} \sum_{A \in \mathcal{A}_{sto}} D_{A^2}.$$

is the so-called infinitesimal generator of the SDE (2.3), see [45], [35].

- $\mathbb{E}J(t; t_0)$  comprises no element of fractional weight: for  $w$  such that  $\|w\| \notin \mathbb{Z}$ ,  $\mathbb{E}J_w(t; t_0) = 0$ .
- $h \mapsto \mathbb{E}J(t_0 + h; t_0)$  is  $h$ -coherent.



**Proof:**

- Observe that  $\mathbb{E}J_{2AA}(t; t_0) = t - t_0$  but  $\mathbb{E}J_A(t; t_0) = 0$ , so  $\mathbb{E}J_{2AA}(t; t_0) = \mathbb{E}J_{A \sqcup A}(t; t_0) \neq (\mathbb{E}J(t; t_0))_A^2$ . In particular,

$$\mathbb{E}J(t; t_0) \notin \mathcal{G}_{sh}$$

if the stochastic alphabet is not empty. In other case,  $\mathbb{E}J(t; t_0) = J(t; t_0) \in \mathcal{G}_{sh}$ .

- As  $\mathbb{E}J_1(t; t_0) = 1$ ,  $\mathbb{E}J(t; t_0)$  has a logarithm in  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$ , i.e. there exist  $R \in \mathfrak{m} \subset \mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  such that

$$\mathbb{E}J(t; t_0) = \exp(R(t; t_0)).$$

Use the ansatz  $R(t; t_0) = (t - t_0)X$ . If  $t = t_0 + h$ ,

$$\partial_h \mathbb{E}J(t_0 + h; t_0) = \partial_h \exp(hX) = X \exp(hX).$$

On the one hand, if we set above  $h = 0$ ,

$$\partial_{h|_{h=0}} \mathbb{E}J(t_0 + h; t_0) = X.$$

On the other hand,

$$\partial_{h|_{h=0}} \mathbb{E}J(t_0 + h; t_0) = \sum_w \partial_{h|_{h=0}} \mathbb{E}J_w(t_0 + h; t_0)w.$$

$h$ -coherence ensures that  $\partial_{h|_{h=0}} \mathbb{E}J_w(t_0 + h; t_0) = 0$  provided  $\|w\| \geq 2$ . In other case,

$$\partial_{h|_{h=0}} \mathbb{E}J_a(t_0 + h; t_0) = 1,$$

and

$$\partial_{h|_{h=0}} \mathbb{E}J_{A^2}(t_0 + h; t_0) = \frac{1}{2}.$$

To sum up,

$$X = \sum_{a \in \mathcal{A}_{det}} a + \frac{1}{2} \sum_{A \in \mathcal{A}_{sto}} A^2 = \mathfrak{f},$$

and

$$\mathbb{E}J(t_0 + h; t_0) = \exp(\mathfrak{f}), \quad \mathbb{E}X(t) = \exp(D_{\mathfrak{f}}).$$

This is the so-called backward Kolmogorov's equation or the Feymann-Kac formula, see again [45], [35].

- Kolmogorov's backward equation leads to

$$\boxed{\mathbb{E}J_w(t_0 + h; t_0) = (\mathbb{E}J(t_0 + h; t_0), w) = (\exp(h\mathfrak{f}), w)}.$$

Let  $\pi(w)$  denote the number of disjoint pairs of the form  $A^2$ ,  $A \in \mathcal{A}_{sto}$  in the word  $w$ . For example:  $\pi(A^2) = 1$ ,  $\pi(A^3) = 2$ ,  $\pi(A^4) = 3$ . Then

$$\mathbb{E}J_w(t_0 + h; t_0) = \left( \sum_{n=0}^{\infty} \frac{h^n (\sum_{a \in \mathcal{A}_{det}} a + \sum_{A \in \mathcal{A}_{sto}} \frac{A^2}{2})^n}{n!}, w \right).$$

Let  $\mathcal{W}' \subset \mathcal{W}$  be the set of words generated by concatenation of deterministic letters and the pairs  $A^2$  with  $A \in \mathcal{A}_{sto}$ . Observe that for words  $w \in \mathcal{W}'$ ,  $\|w\| \in \mathbb{Z}$ . Obviously, if  $w \notin \mathcal{W}'$ ,

$$\left( \sum_{n=0}^{\infty} \frac{h^n (\sum_{a \in \mathcal{A}_{det}} a + \sum_{A \in \mathcal{A}_{sto}} \frac{A^2}{2})^n}{n!}, w \right) = 0,$$

and thus  $\mathbb{E}J_w(t_0 + h; t_0) = 0$ . In other case,

$$\mathbb{E}J_w(t_0 + h; t_0) = \frac{1}{2^{\pi(w)}} \frac{h^{\|w\|}}{\|w\|!}.$$

- If  $\|w\| \notin \mathbb{Z}$ ,

$$\mathbb{E}J_w(t_0 + h; t_0) = 0 = O(h^{\|w\|}).$$

In other case,

$$\mathbb{E}J_w(t_0 + h; t_0) = \frac{1}{2^{\pi(w)}} \frac{h^{\|w\|}}{\|w\|!} = O(h^{\|w\|}).$$

□

Observe that  $\mathfrak{f} \notin \mathfrak{g}_{sh}$  so that  $\mathbb{E}J(t; t_0) \notin \mathcal{G}_{sh}$ , which just amounts to saying that  $\mathbb{E}\mathbb{X}(t)$  is not a multiplicative operator:

$$\begin{aligned} \mathbb{E}\mathbb{X}(t)(\chi\chi')(x_0) &= \mathbb{E}^{x_0}(\chi(x(t))\chi'(x(t))) \\ &= \mathbb{E}^{x_0}(\chi(x(t)))\mathbb{E}^{x_0}(\chi'(x(t))) + Cov^{x_0}(\chi(x(t))\chi'(x(t))) \\ &= \mathbb{E}\mathbb{X}(t)(\chi)(x_0)\mathbb{E}\mathbb{X}(t)(\chi')(x_0) + Cov^{x_0}(\chi(x(t))\chi'(x(t))). \end{aligned}$$

## 2.4 Renormalization of SDE's

The following paragraphs are devoted to showing some applications of Theorems 1.10 and 1.8 to the initial value problem (2.3) in  $\mathbb{R}^d$ :

$$dx(t) = \sum_{a \in \mathcal{A}_{det}} f_a(x(t))dt + \sum_{a \in \mathcal{A}_{sto}} f_A(x(t)) \circ d\mathcal{B}_A(t), \quad x(t_0) = x_0.$$

The freedom present in the hypotheses of Theorem 1.10 offers many possibilities to construct modified equations for the SDE.

### 2.4.1 Derandomization

Choose  $h > 0$  and consider Theorem 1.8. Set  $X := J(t_0 + h; t_0) \in \mathcal{G}_{sh}$ . Then, for every event  $\omega$  there exist  $R = R_h \in \mathfrak{g}_{sh}$  and a curve  $t \mapsto S(t) = S_h(t) \in \mathcal{G}_{sh}$  with end-points  $S(t_0) = 1, S(t_0 + h) = X$ , both connected via the equation  $S'(t) = S(t)R$ . Theorem 2.1 provides an explicit formula for both  $R$  and  $S(t)$ . For  $R$ :

$$\begin{aligned} R &= \frac{1}{h} \log(X) = \frac{1}{h} \log(J(t_0 + h; t_0)) \\ &= \frac{1}{h} (J(t_0 + h; t_0) - 1) - \frac{1}{2h} (J(t_0 + h; t_0) - 1)^2 + \frac{1}{3h} (J(t_0 + h; t_0) - 1)^3 \\ &\quad + \dots \\ &= \sum_{A \in \mathcal{A}_{sto}} \frac{J_A(t_0 + h; t_0)}{h} A + \sum_{a \in \mathcal{A}_{det}} a + O(0.5), \end{aligned}$$

where  $O(0.5)$  stands for terms  $R_w$  such that  $\mathbb{E}|R_w| = O(\sqrt{h})$ . For  $S(t)$ :

$$S(t) = S_h(t) = \exp((t - t_0)R).$$

Observe that when considering  $h$  as a parameter,  $h \mapsto R = R_h$  and  $h \mapsto S_h(t)$  are  $h$ -coherent,  $R$  with parameter  $N = 1$  and  $S$  with  $N = 0$ . The word series  $y(t) := W_{S(t)}(x_0)$  and the formal vector field

$$\hat{f}(x) := W_R(x) = \sum_w R_w f_w(x)$$

are connected via the formal random ordinary differential equation

$$\boxed{y'(t) = \hat{f}(y(t))}, \quad (2.13)$$

which satisfies  $y(t_0 + h) = x(t_0 + h)$ , provided that the initial conditions coincide. Note that

$$h\hat{f}(x) = \sum_{A \in \mathcal{A}_{sto}} J_A(t_0 + h; t_0) f_A(x) + h \sum_{a \in \mathcal{A}_{det}} f_a(x) + \mathcal{V},$$

where  $\mathcal{V}$  is a random variable with  $\mathbb{E}|\mathcal{V}| = O(h^{1.5})$ . In a particular problem, many word basis functions may be expected to vanish thus reducing the number of coefficients of  $R_w$  that must be computed. We have then *obtained a formal differential equation that interpolates the SDE (2.3) at time  $x(t_0 + h)$* . Note that

$$y'(t) = \partial_t \sum_w S(t) f_w(x_0) = \sum_w S'(t) f_w(x_0) = \sum_w S(t) R f_w(x_0) = W_{S(t)R}(x_0),$$

but Theorem 2.2 states that

$$W_{S(t)R}(x_0) = W_R(W_{S(t)}(x_0)) = \hat{f}(y(t)),$$

leading us to the following theorem.

**Theorem 2.5 (Derandomization)** *Take  $h > 0$  and consider the SDE (2.3). Then there exists a random formal vector field  $\hat{f}(x) = W_{R_h}(x)$  with  $R_h \in \mathfrak{g}_{sh}$  such that the random ordinary differential equation*

$$y'(t) = \hat{f}(y(t)),$$

*is a modified equation for (2.3), i.e. whenever the initial condition  $y(t_0) = x_0$  is set, the equation at time  $t_0 + h$  reaches the value*

$$\boxed{y(t_0 + h) = x(t_0 + h).}$$

*Furthermore, the algebra point  $R_h$  is readily computed via*

$$R_h = \frac{1}{h} \log(J(t_0 + h; t_0)).$$

A few remarks:

- This construction shows that, if the value  $x(t_0 + h)$  is to be computed approximately, there is no need for solving the SDE (2.3) itself: the random ordinary differential equation interpolates the SDE at time  $t_0$  and the only randomness involved in it is that of the random variables  $J_w(t_0 + h; t_0)$ .
- The RODE, i.e. random ordinary differential equation, is not adapted to the natural filtration  $(\mathcal{F}_t)_{t \geq t_0}$  of the Brownian motions. In particular, the vector field  $\hat{f}$  is  $\mathcal{F}_{t_0+h}$ -measurable.

**Examples**

- **Brownian Motion.** Consider  $d = 1$  and

$$\boxed{dx(t) = d\mathcal{B}(t), \quad x(0) = 0.}$$

Obviously  $x(t) = \mathcal{B}(t)$ . Note that the alphabet only comprises the letter  $A$ , set then  $f_A(x) = 1$ . Observe that  $f_{A^2} = f'_A(f_A) = 0$  so that  $f_{A^n} = f_A = 0$ . In particular, the word series for this alphabet are of the form

$$W_S(x_0) = \sum_w S_w f_w(x_0) = S_1 x_0 + S_A.$$

The equation  $J_{A^n}(t_0 + h; t_0) = \mathcal{B}^n(h)/n!$  is critical to check that

$$J(t_0 + h; t_0) = \sum_{n=0}^{\infty} J_{A^n}(t_0 + h; t_0) A^n = \sum_{n=0}^{\infty} \frac{\mathcal{B}^n(h)}{n!} A^n = \exp(\mathcal{B}(h)A).$$

Observe that  $A \in \mathfrak{g}_{sh}$ , thus  $J(t_0 + h; t_0) \in \mathcal{G}_{sh}$ . Following Theorem 2.1, we obtain  $R = R_h$ :

$$R_h := \frac{1}{h} \log(J(t_0 + h; t_0)) = \frac{1}{h} \log(\exp(\mathcal{B}(h)A)) = \frac{\mathcal{B}(h)}{h} A,$$

so that  $W_{R_h}(x_0) = \mathcal{B}(h)/h$ . The random differential equation (2.13) now reads

$$y'(t) = W_{R_h}(y(t)) = \frac{\mathcal{B}(h)}{h}.$$

The above equation in tandem with the condition  $y(0) = x(0) = 0$  proves that  $y(t)$  is linear in  $t$ . To obtain the slope of  $y(t)$  we use the interpolation condition  $y(t_0 + h) = \mathcal{B}(h)$ . In this manner,

$$\boxed{y(t) = \frac{\mathcal{B}(h)}{h} t, \quad x(t) = \mathcal{B}(t).} \quad (2.14)$$

Observe that the vector field  $f_{R_h}$  is not adapted but is  $\mathcal{F}_h$  measurable.

- **Exponential Brownian Motion.** Consider now

$$\boxed{dx(t) = x(t) \circ d\mathcal{B}(t), \quad x(0) = 1.}$$

The solution is  $x(t) = e^{\mathcal{B}(t)}$ . Set  $f_A(x) = x$  so that, once again, the alphabet only comprises one letter. For this case,

$$f_{A^2} = f'_A(f_A) = x.$$

Thus  $f_{A^n} = x$  for  $n \geq 0$ . The word series for this alphabet are of the form

$$W_S(x_0) = \sum_w S_w f_w(x_0) = x_0 \sum_w S_w.$$

Again,

$$J(t_0 + h; t_0) = \sum_{n=0}^{\infty} \frac{\mathcal{B}^n(h)}{n!} A^n = \exp(\mathcal{B}(h)A) \in \mathcal{G}_{sh}.$$

With a view to obtaining the formal vector field we must find the formal series  $R_h$ , which again is far from being difficult:

$$R_h = \frac{1}{h} \log(J(t_0 + h; t_0)) = \frac{1}{h} \log(\exp(\mathcal{B}(h)A)) = \frac{\mathcal{B}(h)}{h} A.$$

Thus,

$$\hat{f}(x) = W_{R_h}(x) = \frac{\mathcal{B}(h)}{h} x.$$

The random differential equation that interpolates the SDE is

$$y'(t) = W_R(y(t)) = \frac{\mathcal{B}(h)}{h} y(t),$$

which provides the comparison

$$\boxed{y(t) = \exp\left(\frac{\mathcal{B}(h)}{h}t\right), \quad x(t) = \exp(\mathcal{B}(t)).} \quad (2.15)$$

Again,  $y(h) = \exp(\mathcal{B}(h)) = x(h)$ .

- Consider next

$$\boxed{dx(t) = x(t)dt + x(t) \circ d\mathcal{B}(t), \quad x(0) = 1,}$$

whose solution is  $x(t) = e^{\mathcal{B}(t)+t}$ . Now the alphabet comprises two different letters, say  $a$  and  $A$ . Set  $f_a(x) = f_A(x) := x$ . Notice that  $f_w(x) = x$  for all words  $w$  so that any word series is of the form

$$W_S(x) = x \sum_w S_w.$$

Now the formula for  $R_h$  does not admit a compact expression such as those above, although it can always be computed recursively. From now on, for

notational convenience, we drop the  $h$  dependence of  $R_h$ . It is worth noticing that  $R$  is  $h$ -coherent with  $N = 1$ , i.e.  $R_w = O(h^{|w|-1})$ . The element  $R$  is of the form

$$R = R_A A + R_a a + R_{A^2} A^2 + R_{A^3} A^3 + R_{aA} aA + R_{Aa} Aa + R_{a^2} a^2 + R_{A^4} A^4 + R_{aAA} aAA + R_{AAa} AAa + R_{AAa} AAa + O(1.5).$$

Recall that  $R \in \mathfrak{g}_{sh}$ . Therefore it is orthogonal to shuffles and hence  $R_{A^2} = R_{A^3} = \dots = 0$ ,  $R_{aA} = -R_{AA}$  and so on. Hence,

$$W_R(x) = x \sum_w R_w = x((R_A + R_a + O(1.5))).$$

Moreover, choose a word  $w$  with  $n$  letters. Let  $a(w)$  denote the number of  $a$ 's on  $w$ . If  $a(w) = 0$ , i.e.  $w = A^n$ , then  $nw = A \sqcup A^{n-1}$ , so that  $R_w = 0$ . In other case, for  $1 \leq j \leq n$ ,

$$\sum_{w: a(w)=j, |w|=n} w = a^j \sqcup A^{n-j}. \quad (2.16)$$

Thus,

$$\sum_{w: a(w)=j, |w|=n} R_w = 0.$$

We are now in a position to obtain the following equation

$$\sum_{w: |w|=n} R_w = \sum_{j=0}^n \sum_{w: a(w)=j, |w|=n} R_w = 0, \quad (2.17)$$

which relieves us from calculating all coefficients excepting those of  $R_A$  and  $R_a$ . In this way,

$$\hat{f}(x) = x \sum_w R_w = x(R_A + R_a),$$

and

$$y'(t) = y(t) \left(1 + \frac{\mathcal{B}(h)}{h}\right).$$

The comparison of solutions is

$$\boxed{y(t) = \exp\left(t + \frac{\mathcal{B}(h)}{h}t\right), \quad x(t) = \exp(t + \mathcal{B}(t)).}$$

### 2.4.2 Drift-free SDE

The Derandomization is nothing but removing the Brownian vector fields in (2.3). Theorem 1.10 allows to eliminate in several ways the vector fields associated with the drift term.

#### First approach: same Brownian motions

Assume that  $\mathcal{A}_{sto}$  comprises  $N$  letters,  $A_1, A_2 \dots A_N$ . Take  $h > 0$  and set  $M = N$  and  $R^{(i)} = A_i \in \mathfrak{g}_{sh}$  for  $2 \leq i \leq N$ . According to Theorem 1.10, there exist  $R = R_h \in \mathfrak{g}_{sh}$  and a curve  $t \mapsto S(t) \in \mathcal{G}_{sh}$ , such that for each event  $\omega$ ,

$$dS(t) = S(t)R \circ d\mathcal{B}_{A_1}(t) + \sum_{i=2}^N S(t)A_i \circ d\mathcal{B}_{A_i}(t), \quad (2.18)$$

and  $S(t_0) = 1, S(t_0+h) = J(t_0+h; t_0)$ . Recall that  $R$  depends on  $h$  and it may be explicitly computed by using (2.18). After setting  $\hat{f}(x) = W_R(x)$ , the drift-free formal SDE

$$dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A_1}(t) + \sum_{i=2}^N f_{A_i}(y(t)) \circ d\mathcal{B}_{A_i}(t) \quad (2.19)$$

is a modified equation for the SDE (2.3):

$$dx(t) = \sum_{a \in \mathcal{A}_{det}} f_a(y(t))dt + \sum_{i=1}^N f_{A_i}(y(t)) \circ d\mathcal{B}_{A_i}(t),$$

which just amounts to saying that the value of their solutions at time  $t = t_0 + h$  coincide provided that  $y(t_0) = x(t_0) = x_0$ .

**Theorem 2.6 (Drift-free modified SDE's, first version)** *Take  $h > 0$  and consider the SDE (2.3). Then, for any stochastic letter  $A' \in \mathcal{A}_{sto}$ , there exists a random formal vector field  $\hat{f}(x) = W_{R_h}(x)$  with  $R_h \in \mathfrak{g}_{sh}$  random such that the stochastic differential equation*

$$dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}'_{A'}(t) + \sum_{A \in \mathcal{A}_{sto}, A \neq A'} f_A(y(t)) \circ d\mathcal{B}_A(t)$$

*is a modified equation for (2.3). Furthermore,  $R_h$  is explicitly computable via Theorem 1.10*



A few comments:

- The drift-free SDE is defined almost surely due to the random nature of  $J_A(t_0 + h; t_0) = \mathcal{B}_A(t_0 + h) - \mathcal{B}_A(t_0) \sim \mathcal{N}(0, h)$ . Fortunately,  $P(\mathcal{B}_A(t_0 + h) = \mathcal{B}_A(t_0)) = 0$  so that  $\hat{f}$  is defined almost surely.
- Again, the drift-free SDE fails to be adapted to the filtration of the natural Brownian motion;  $\hat{f}$  is  $\mathcal{F}_{t_0+h}$ -measurable.

### Examples

- Consider

$$\boxed{dx(t) = dt + x(t)d\mathcal{B}_A(t) + x(t)d\mathcal{B}_B(t), \quad x(0) = 0,}$$

whose solution is  $x(t) = e^{t+\mathcal{B}_A(t)+\mathcal{B}_B(t)}$  and take  $A' := A$ . Set  $f_a(x) = 1$ ,  $f_A(x) = f_B(x) = x$ . Recall that  $a(w)$  is the number of  $a$ 's in the word  $w$ . It is easy to check that

$$f_w(x) = \begin{cases} x & \text{if } a(w) = 0 \\ 0 & \text{if } a(w) \neq 0 \end{cases} \quad (2.20)$$

for words  $|w| \geq 2$ . In a similar fashion to (5.6), we obtain

$$\sum_{w:a(w)=0,|w|=n} R_w = 0$$

for all  $n \geq 1$ . Thus,

$$\hat{f}(x) = W_R(x) = R_A x + R_B x + R_a.$$

It takes no time to compute the first terms of  $R$ :  $R_a = h/\mathcal{B}_A(h)$ ,  $R_A = 1$ ,  $R_B = 0$ , so that

$$\hat{f}(x) = W_R(x) = x + \frac{h}{\mathcal{B}_A(h)}.$$

Observe that the above vector field is defined almost surely. The drift-free SDE is

$$dy(t) = \left(y(t) + \frac{h}{\mathcal{B}_A(h)}\right) \circ d\mathcal{B}_A(t) + y(t) \circ d\mathcal{B}_B(t).$$

In conclusion:

$$\boxed{y(t) = \exp\left(\left(1 + \frac{h}{\mathcal{B}_A(h)}\right)\mathcal{B}_A(t) + \mathcal{B}_B(t)\right), \quad x(t) = \exp\left(t + \mathcal{B}_A(t) + \mathcal{B}_B(t)\right).}$$

**Second approach: only one Brownian motion involved**

Fix  $h > 0$ , take a stochastic letter  $A'$  and consider Theorem 1.10. Take  $M = 1$ ,  $x_1(t) = \mathcal{B}_{A'}(t)$  and  $X := J(t_0 + h; t_0)$ . Then, there exist a random  $R = R^{(1)} \in \mathfrak{g}_{sh}$  and a curve  $t \mapsto S(t) \in \mathcal{G}_{sh}$ , both depending on  $h$ , such that equality holds in

$$dS(t) = S(t)R \circ d\mathcal{B}_{A'}(t). \quad (2.21)$$

Again, the end-points of the curve are  $S(t_0) = 1$ ,  $S(t_0 + h) = X = J(t_0 + h; t_0)$ . In this case, we are obviously looking for a renormalization of the SDE (2.3) of the form

$$y'(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A'}(t),$$

where  $\hat{f}(x) = W_R(x)$ . As stated in Theorem 2.1, equation (2.21) enables us to compute recursively both  $R$  and  $S(t)$ , even though the simple structure of the boundary value problem provides the straightforward formulae:

- $R = \frac{1}{J_{A'}(t_0 + h; t_0)} \log(J(t_0 + h; t_0)) = \frac{1}{\mathcal{B}_{A'}(t_0 + h) - \mathcal{B}_{A'}(t_0)} \log(J(t_0 + h; t_0))$
- $S(h) = \exp((\mathcal{B}_{A'}(t_0 + h) - \mathcal{B}_{A'}(t_0))R)$ .

Observe too that  $R \in \mathfrak{g}_{sh}$  is defined almost surely and that  $R$  is  $h$ -coherent with  $N = 1/2$ , i.e.  $\mathbb{E}|R_w| = O(h^{\|w\| - 1/2})$ . The word series  $y(t) := W_{S(t)}(x_0)$  and the *formal vector field* (defined almost surely)

$$\hat{f}(x) := W_R(x) = \sum_w R_w f_w(x),$$

whose first terms are

$$\hat{f} = \frac{1}{J_{A'}(t_0 + h; t_0)} \sum_{A \in \mathcal{A}_{sto}} J_A(t_0 + h; t_0) f_A + \frac{h}{J_{A'}(t_0 + h; t_0)} \sum_{a \in \mathcal{A}_{det}} f_a + \mathcal{V}_1,$$

where  $\mathcal{V}$  is a random variable with  $\mathbb{E}|\mathcal{V}_1| = O(h)$ , satisfy the *formal stochastic differential equation*

$$\boxed{dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A'}(t)}, \quad (2.22)$$

which is defined almost surely. The point is that

$$\boxed{y(t_0 + h) = x(t_0 + h)},$$

provided  $x(t_0) = y(t_0)$ . Again, the former statement is proved by the following calculations:

$$\begin{aligned} dy(t) &= d \sum_w S(t) f_w(x_0) = \sum_w dS(t) f_w(x_0) = \\ &= \sum_w S(t) R \circ d\mathcal{B}_{A'}(t) f_w(x_0) = W_{S(t)R}(x_0) \circ d\mathcal{B}_{A'}(t) = \hat{f}(y(t)). \end{aligned}$$

**Theorem 2.7 (Drift-free modified SDE's, second version)** *Take  $h > 0$  and consider the SDE (2.3). Then, for any stochastic letter  $A' \in \mathcal{A}_{sto}$ , there exists a random formal vector field  $\hat{f}(x) = W_{R_h}(x)$  with  $R_h \in \mathfrak{g}_{sh}$  such that the stochastic differential equation*

$$dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A'}(t)$$

*is a modified equation for (2.3). The series  $R_h \in \mathfrak{g}_{sh}$  is explicitly computable via Theorem 1.10.*

### Examples

- **Independent Brownian Motions.** Consider

$$dx(t) = d\mathcal{B}_A(t) + d\mathcal{B}_B(t), \quad x(0) = 0.$$

Obviously  $x(t) = \mathcal{B}_A(t) + \mathcal{B}_B(t)$ . Take, for example,  $A' := A$ . Set  $f_A(x) = f_B(x) = 1$ . This condition means that  $f_w = 0$  for words with  $|w| \geq 1$ , so that any word series is of the form

$$W_S(x) = xS_1 + S_A + S_B.$$

We already know that  $R_1 = 0, R_A = 1, R_B = \mathcal{B}_B(h)/\mathcal{B}_A(h)$ , i.e.  $\hat{f}(x) = \mathcal{B}_B(h)/\mathcal{B}_A(h)$ . Observe that both  $R$  and  $\hat{f}$  are defined almost surely. The drift-free SDE is

$$dy(t) = \left(1 + \frac{\mathcal{B}_B(h)}{\mathcal{B}_A(h)}\right) \circ d\mathcal{B}_A(t).$$

In this case,

$$\boxed{x(t) = \mathcal{B}_A(t) + \mathcal{B}_B(t), \quad y(t) = \left(1 + \frac{\mathcal{B}_B(h)}{\mathcal{B}_A(h)}\right) \mathcal{B}_A(t).}$$

- **Exponential function.** Consider now

$$\boxed{x'(t) = x(t), \quad x(t_0) = 1.}$$

The solution is  $x(t) = e^{t-t_0}$ . Set  $f_a(x) = x$  and consider, as in Theorem 1.10, a Brownian motion  $\mathcal{B}(t)$  defined on some probability space. This Brownian motion will play the role of integrator. In this case, the alphabet only comprises one letter, which is deterministic. Because of the structure of  $f_a$ , it takes no time to infer that  $f_{a^n}(x) = x$  for  $n \geq 0$ . Again,

$$J(t_0 + h; t_0) = \sum_{n=0}^{\infty} \frac{h^n}{n!} a^n = \exp(ha) \in \mathcal{G}_{sh}.$$

Compute next the formal series

$$\begin{aligned} R &:= \frac{1}{\mathcal{B}(t_0 + h) - \mathcal{B}(t_0)} \log(J(t_0 + h; t_0)) = \frac{1}{\mathcal{B}(t_0 + h) - \mathcal{B}(t_0)} \log(\exp(ha)) \\ &= \frac{ha}{\mathcal{B}(t_0 + h) - \mathcal{B}(t_0)}. \end{aligned}$$

Thus, the formal vector field simply reads  $W_R(x) = xh/(\mathcal{B}(t_0+h) - \mathcal{B}(t_0))$  and the *drift-free SDE* that interpolates the ODE is

$$dy(t) = W_R(y(t)) \circ d\mathcal{B}(t) = y(t) \left( \frac{h}{\mathcal{B}(t_0 + h) - \mathcal{B}(t_0)} \right) \circ d\mathcal{B}(t).$$

Thus,

$$\boxed{y(t) = \exp \left( h \frac{\mathcal{B}(t) - \mathcal{B}(t_0)}{\mathcal{B}(t_0 + h) - \mathcal{B}(t_0)} \right), \quad x(t) = \exp(t - t_0)} \quad (2.23)$$

Again,  $y(h) = e^h = x(h)$ .

- Consider

$$\boxed{dx(t) = x(t)dt + x(t) \circ d\mathcal{B}(t), \quad x(0) = 1,}$$

whose solution is  $x(t) = e^{\mathcal{B}(t)+t}$ . We already know that  $f_a(x) = f_A(x) = f_w(x) = x$  for every word  $w$ , hence

$$W_S(x) = x \sum_w S_w$$

for every  $S \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle$ . In the same way, we have already proved that, for any  $Q \in \mathfrak{g}_{sh}$ ,

$$W_Q(x) = x(Q_A + Q_a).$$

As  $R = A + h/\mathcal{B}(h)a$ , the drift-free renormalization of the SDE is

$$dy(t) = y(t) \left( 1 + \frac{h}{\mathcal{B}(h)} \right) \circ d\mathcal{B}(t).$$

In this case:

$$\boxed{y(t) = \exp \left( \mathcal{B}(t) \left( 1 + \frac{h}{\mathcal{B}(h)} \right) \right), \quad x(t) = \exp(\mathcal{B}(t) + t).}$$



# Chapter 3

## Splitting integrators for Stratonovich SDE's

### 3.1 Expansion of the numerical integrator

The following paragraphs are devoted to expressing the “pullback” of split-step integrators in terms of word series. It is obvious that the splitting procedure may also be applied to stochastic differential equations. Split-step integrators turn to be an extremely important tool when trying to solve a wide range of differential equations, from partial differential equations to stochastic differential equations. The main ideas of split-step integrators are present in the case of ordinary differential equations, and key references for this case are [27] and [7].

#### 3.1.1 Composition of word series: Shuffle Group

Theorem 2.2 leads to a general technique to represent the local error of splitting integrators. For notational convenience, we will only present a particular simple example, even though the idea is completely general.

Consider the particular Stratonovich initial value problem

$$dx = f_a(x)dt + f_b(x)dt + f_A(x) \circ d\mathcal{B}(t), \quad x(t_0) = x_0. \quad (3.1)$$

Its associated alphabet consists of three letters  $\mathcal{A} = \{a, b, A\}$ , with only one of them stochastic, namely,  $A$ . Denote by  $\phi_{t,t_0} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  the solution mapping, i.e. for each  $x_0$ ,  $\phi_{t,t_0}(x_0)$  is the value at time  $t$  of the solution with initial condition  $x(t_0) = x_0$ . Assume that the split systems

$$dx = f_a(x)dt + f_A(x) \circ d\mathcal{B}(t), \quad \frac{d}{dt}x = f_b(x),$$

may be integrated analytically and denote by  $\phi_{t,t_0}^{(1)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\phi_{t,t_0}^{(2)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  their respective solution mappings.<sup>1</sup> The simplest splitting integrator advances the numerical solution from  $t_0$  to  $t_0 + h$ ,  $h > 0$  by means of the mapping

$$\tilde{\phi}_{t_0+h,t_0} = \phi_{t_0+h,t_0}^{(2)} \circ \phi_{t_0+h,t_0}^{(1)}.$$

### Step 1. Words series for SDE's expansions

For every letter  $l \in \mathcal{A}$  take  $\lambda_l = 1$  in order to construct the *iterated integrals*  $h \mapsto J(t_0 + h; t_0) \in \mathcal{G}_{sh}$ , see (1.42). Now, for every  $h \geq 0$ , we can write the following expansion thanks to Theorem 2.4:

$$\begin{aligned} \phi_{t_0+h;t_0}(x_0) = & W_{J(t_0+h;t_0)}(x_0) = x_0 + J_A f_A(x_0) + J_a f_a(x_0) + J_b f_b(x_0) \\ & + J_{A^2} f_{A^2}(x_0) + J_{aA} f_{aA}(x_0) + J_{Aa} f_{Aa}(x_0) + J_{Ab} f_{Ab}(x_0) + \cdots, \end{aligned}$$

where we just have dropped the  $t_0$  and  $h$  dependence of the iterated integrals. Let us display those featuring in the above equation:

- $J_a = J_b = h$ .
- $J_A = \mathcal{B}(t_0 + h) - \mathcal{B}(t_0)$ .
- $J_{A^2} = \frac{(\mathcal{B}(t_0+h) - \mathcal{B}(t_0))^2}{2}$ .
- $J_{Aa} = J_{Ab} = \int_{t_0}^{t_0+h} \mathcal{B}(t_0 + s) - \mathcal{B}(t_0) ds$ .
- $J_{aA} = J_{bA} = \int_{t_0}^{t_0+h} s d\mathcal{B}(s)$ .

Then, let us tackle the splitting word series expansion. Define first

$$J_w^{(1)}(t; t_0) = \begin{cases} J_w(t; t_0) & \text{if } b(w) = 0 \\ 0 & \text{if } b(w) \neq 0 \end{cases} \quad (3.2)$$

and

$$J_w^{(2)}(t; t_0) = \begin{cases} J_w(t; t_0) & \text{if } a(w) + A(w) = 0 \\ 0 & \text{if } a(w) + A(w) \neq 0. \end{cases} \quad (3.3)$$

where  $l(w)$  denotes the length of the word  $w$ . The coefficients  $J_w^{(1)}(t; t_0)$  are the iterated integrals obtained by taking  $\lambda_a = \lambda_A = 1, \lambda_b = 0$ . Similarly, to obtain  $J_w^{(2)}(t; t_0)$  just set  $\lambda_a = \lambda_A = 0, \lambda_b = 1$ . As iterated integrals, the random series  $J^{(1)}(t_0 + h; t_0), J^{(2)}(t_0 + h; t_0)$  belong to  $\mathcal{G}_{sh}$  for each  $h > 0$  and for every event  $\omega$ . In this way

$$\phi_{t_0+h;t_0}^{(1)}(x_0) = W_{J^{(1)}(t_0+h;t_0)}(x_0)$$

<sup>1</sup>Observe that second solution operator is a flow indeed so that it only depends on  $t - t_0$ .

and

$$\phi_{t_0+h;t_0}^{(2)}(x_0) = W_{J^{(2)}(t_0+h;t_0)}(x_0).$$

Let us display the first terms of their expansions. For the former:<sup>2</sup>

$$\begin{aligned} \phi_{t_0+h;t_0}^{(1)}(x_0) = & x_0 + J_A f_A(x_0) + J_a f_a(x_0) + J_{A^2} f_{A^2}(x_0) + J_{A^3} f_{A^3}(x_0) \\ & + J_{aA} f_{aA}(x_0) + J_{Aa} f_{Aa}(x_0) + J_{A^4} f_{A^4}(x_0) + J_{a^2} f_{a^2}(x_0) + \dots \end{aligned} \quad (3.4)$$

and for the latter:

$$\phi_{t_0+h;t_0}^{(2)}(x_0) = x_0 + J_b f_b(x_0) + J_{b^2} f_{b^2}(x_0) + \dots \quad (3.5)$$

### Step 2. Word series expansion of numerical integrator

Invoking Proposition 2.2 we find

$$\tilde{\phi}_{t_0+h;t_0}(x_0) = W_{\tilde{J}(t_0+h;t_0)}(x_0), \quad (3.6)$$

with

$$h \mapsto \tilde{J}(t_0+h;t_0) = J^{(1)}(t_0+h;t_0)J^{(2)}(t_0+h;t_0) \in \mathcal{G}_{sh}.$$

Bearing in mind the correspondence between the product of series and the *convolution* (1.16) of their coefficients, the family  $(\tilde{J}_w(t_0+h;t_0))_{w \in \mathcal{W}} \in \mathbb{R}^{\mathcal{W}}$  is computed by

$$(\tilde{J}_w(t_0+h;t_0))_{w \in \mathcal{W}} = (J_w^{(1)}(t_0+h;t_0))_{w \in \mathcal{W}} \star (J_w^{(2)}(t_0+h;t_0))_{w \in \mathcal{W}}.$$

It is quite easy to find the coefficients in the last expansion. If  $w$  is a concatenation  $w'w''$ , where the (possibly empty) word  $w'$  does not include the letter  $c$  and the (possibly empty) word  $w''$  does not include the letters  $a$  or  $b$ , then  $\tilde{J}_w = J_{w'}J_{w''}$ ; if  $w$  is not a concatenation of that form, then the coefficient is 0, and so on. In particular, using underlines to indicate the stochastic weight, we find

$$\begin{aligned} \tilde{J}(t_0+h;t_0) = & \underbrace{1}_0 + \underbrace{J_A A}_{0.5} + \underbrace{h(a+b) + J_{A^2} A^2}_{1} + \\ & + \underbrace{J_{A^3} A^3 + J_A J_b A b + J_{Aa} A a + J_{aA} a A}_{1.5} + \dots \end{aligned}$$

On the other hand,

$$\begin{aligned} J(t_0+h;t_0) = & \underbrace{1}_0 + \underbrace{J_A A}_{0.5} + \underbrace{h(a+b) + J_{A^2} A^2}_{1} + \\ & + \underbrace{J_{A^3} A^3 + J_{Ab} A b + J_{bA} b A + J_{Aa} A a + J_{aA} a A}_{1.5} + \dots \end{aligned}$$

<sup>2</sup>Again, for simplicity we drop the  $t_0$  and  $h$  dependence from the iterated integrals.



If we had chosen the Strang-like integrator

$$\tilde{\phi}_{t_0 h; t_0} = \phi_{t_0+h/2; t_0}^{(1)} \circ \phi_{t_0+h/2; t_0}^{(2)} \circ \phi_{t_0+h; t_0+h/2}^{(1)}$$

instead, we would have found

$$h \mapsto \tilde{J}(t_0 + h; t_0) = J^{(1)}(t_0 + h/2; t_0) J^{(2)}(t_0 + h; t_0) J^{(1)}(t_0 + h; t_0 + h/2) \in \mathcal{G}_{sh}.$$

In that case,

$$\begin{aligned} (\tilde{J}_w(t_0 + h; t_0))_{w \in \mathcal{W}} &= (J_w^{(1)}(t_0 + h/2; t_0))_{w \in \mathcal{W}} \star (J_w^{(2)}(t_0 + h; t_0))_{w \in \mathcal{W}} \\ &\quad \star (J_w^{(1)}(t_0 + h; t_0 + h/2))_{w \in \mathcal{W}}. \end{aligned}$$

### Step 3. Word series of the local error

The local error may be expanded as a word series:

$$\tilde{\phi}_{t_0+h, t_0}(x_0) - \phi_{t_0+h, t_0}(x_0) = W_{\tilde{J}(t_0+h; t_0)(x_0)} - W_{J(t_0+h; t_0)}(x_0) = W_{\delta(t_0+h; t_0)}(x_0)$$

with

$$\delta(t_0 + h; t_0) = J^{(1)}(t_0 + h; t_0) J^{(2)}(t_0 + h; t_0) - J(t_0 + h; t_0).$$

Observe that  $(\delta, w) = 0$  for words  $\|w\| \leq 1$ . If  $\|w\| = 1.5$ ,  $(\delta, w) = 0$  unless we take  $bA$  and  $Ab$ . In this case:

- $(\delta, bA) = -J_{bA}(t_0 + h; t_0),$
- $(\delta, Ab) = hJ_A(t_0 + h; t_0) - J_{Ab}(t_0 + h; t_0).$

If  $f_{bA}, f_{Ab}$  do not vanish, this implies that the strong local error  $\epsilon$  is such that  $\mathbb{E}|\epsilon| = O(h^{1.5})$ , even though we are not still in a position to prove it.

### Composition of observables

In some circumstances (for instance when studying conservation of energy or other invariants of motion) it is of interest to look at the error in an observable  $\chi$  after a single step:

$$\chi\left(\tilde{\phi}_{t_0+h, t_0}(x_0)\right) - \chi\left(\phi_{t_0+h, t_0}(x_0)\right),$$

Expansions of errors of this kind are easily derived with the help of (2.1). In our example, we may write, *without any additional computation*,

$$\begin{aligned} & \chi\left(\tilde{\phi}_{t_0+h,t_0}(x_0)\right) - \chi\left(\phi_{t_0+h,t_0}(x_0)\right) \\ &= D_{\delta(t_0+h;t_0)}\chi(x_0) \\ &= (hJ_A(t_0+h;t_0) - J_{Ab}(t_0+h;t_0))D_{Ab}\chi(x_0) \\ &\quad - J_{bA}(t_0+h;t_0)D_{bA}\chi(x_0) + \dots \end{aligned}$$

It is now immediate to study the averages of the former quantities:

$$\boxed{\mathbb{E}^{x_0}\chi\left(\tilde{\phi}_{t_0+h,t_0}(x_0)\right) - \mathbb{E}^{x_0}\chi\left(\phi_{t_0+h,t_0}(x_0)\right) = D_{\mathbb{E}\delta(t_0+h;t_0)}\chi(x_0),}$$

where, as usual,  $\mathbb{E}^{x_0}$  stands for the expectation conditional  $x(t_0) = x_0$ .

For this simple example the results presented here could have been found easily by elementary means. However, as pointed out above, the word series technique works for arbitrary splitting coefficients leading to high-order algorithms and arbitrary ways of splitting the right-hand side of (3.1) into two or more parts.

### 3.1.2 Solution Operator approach

Composition of word series and Theorem 2.2 are not necessary when the expansions of both strong and weak local errors are to be computed. Let  $\mathbb{X}_{t;t_0}$  denote the pullback of solution operator of the system (3.1), set  $\mathbb{X}_{t;t_0}^{(1)}$  for the pullback solution operator of the sub-system (1) and  $\mathbb{X}_{t;t_0}^{(2)}$  for that of the second one. Let  $\tilde{\mathbb{X}}_{t;t_0} \in \mathcal{L}(\mathcal{O})$  be denote the pullback of the numerical integrator. This means that  $\tilde{\mathbb{X}}_{t;t_0}$  is nothing but the linear operator mapping observables into observables such that  $\tilde{\mathbb{X}}_{t;t_0}(\chi)(x_0) = \chi(\tilde{\phi}_{t;t_0}(x_0))$ . Observe that  $\tilde{\mathbb{X}}_{t;t_0}$  is just the numerical analogue to  $\mathbb{X}_{t;t_0}$ . In this case,

$$\mathbb{X}_{t;t_0}^{(1)}\mathbb{X}_{t;t_0}^{(2)}\chi(x_0) = \chi(\phi_{t;t_0}^{(2)}(\phi_{t;t_0}^{(1)}(x_0))) = \chi(\tilde{\phi}_{t;t_0}(x_0)) = \tilde{\mathbb{X}}_{t;t_0}\chi(x_0).$$

Following Theorem 2.4 we obtain

$$\mathbb{X}_{t;t_0} = D_{J(t;t_0)}, \quad \mathbb{X}_{t;t_0}^{(1)} = D_{J^{(1)}(t;t_0)}, \quad \mathbb{X}_{t;t_0}^{(2)} = D_{J^{(2)}(t;t_0)}.$$

Thus,

$$\boxed{\tilde{\mathbb{X}}_{t_0+h;t_0} = D_{J^{(1)}(t_0+h;t_0)}D_{J^{(2)}(t_0+h;t_0)} = D_{J^{(1)}(t_0+h;t_0)J^{(2)}(t_0+h;t_0)},}$$

which summarizes the two identities we have already obtained:

$$\chi(\tilde{\phi}_{t_0+h;t_0}(x_0)) = D_{J^{(1)}(t_0+h;t_0)J^{(2)}(t_0+h;t_0)}\chi(x_0),$$

and

$$\tilde{\phi}_{t_0+h;t_0}(x_0) = D_{J^{(1)}(t_0+h;t_0)J^{(2)}(t_0+h;t_0)}(Id)(x_0) = W_{J^{(1)}(t_0+h;t_0)J^{(2)}(t_0+h;t_0)}(x_0).$$

## 3.2 Word series for split-step integrators theorem, Stratonovich case

The previous sections were devoted to presenting two different approaches to the following theorem, which is one the main results of the present work. Split-step integrators perform the advance  $t_0 \rightarrow t_0 + h$ ,  $h > 0$  by applying a composition of several solution mappings

$$\phi_{t_0+c_i h, t_0+d_i h}^{(i)}, \quad i = 1, \dots, I,$$

corresponding to SDEs resulting from splitting the right-hand side of (2.3). The  $c_i$  and  $d_i$  are real constants associated with the particular integrator. By proceeding as in the previous sections case, the multiplication of series leads to a word-series representation:

$$\tilde{\phi}_{t_0+h, t_0}(x_0) = W_{\tilde{J}(t_0+h; t_0)}(x_0), \quad i = 1, \dots, I,$$

where, for each nonempty  $w \in \mathcal{W}$ , the iterated integral  $\tilde{J}_w(t_0 + h; t_0)$  is either zero or a sum of products of iterated Stratonovich integrals corresponding to words whose concatenation is  $w$ . Therefore, in each product, the iterated integrals being multiplied correspond to words whose weights add up to  $\|w\|$ .

**Theorem 3.1 (Word series for split-step integrators, Stratonovich case)** *Consider a split-step integrator  $\tilde{\phi}_{t_0+h; t_0}$  for the SDE (2.3). Then, for each event  $\omega$ , there exists a point  $\tilde{J}(t_0 + h; t_0) \in \mathcal{G}_{sh}$  such that*

$$\tilde{\phi}_{t_0+h; t_0}(x_0) = W_{\tilde{J}(t_0+h; t_0)}(x_0).$$

Hence, for any observable  $\chi \in \mathcal{O}$ :

$$\chi(\tilde{\phi}_{t_0+h; t_0}(x_0)) = D_{\tilde{J}(t_0+h; t_0)}\chi(x_0), \quad \mathbb{E}^{x_0}\chi(\tilde{\phi}_{t_0+h; t_0}(x_0)) = D_{\mathbb{E}\tilde{J}(t_0+h; t_0)}\chi(x_0).$$

Moreover,  $\tilde{J}(t_0 + h; t_0)$  is explicitly computable.

### 3.2.1 Properties of $\tilde{J}(t_0 + h; t_0)$ and $\mathbb{E}\tilde{J}(t_0 + h; t_0)$

**Proposition 3.1** *The coefficients  $\tilde{J}_w(t_0 + h; t_0)$ ,  $w \in \mathcal{W}$ , associated with a splitting integrator possess the properties of the exact values  $J_w(t_0 + h; t_0)$  listed in Proposition 2.3.*

**Proof:** The first four items of this proposition are consequences of Proposition 2.3 and the representation of each  $\tilde{J}_w(t_0 + h; t_0)$ ,  $w \neq \emptyset$ , as a sum of products of iterated integrals. For the last item, in view of the linearity of the expectation, it is enough to prove that, for any  $t_j < t_j^*$ ,

$$\sum_j \|w_j\| \notin \mathbb{N} \quad \Rightarrow \quad \mathbb{E}\left(\prod_j J_{w_j}(t_j^*; t_j)\right) = 0.$$

Furthermore, we may restrict the attention to the particular case where any two intervals  $(t_j, t_j^*) \subset \mathbb{R}$  are either disjoint or coincident. Let us group together the iterated integrals sharing the same  $(t_j, t_j^*)$  and write

$$\prod_j J_{w_j}(t_j^*; t_j) = \prod_k \prod_{j \in I_k} J_{w_j}(t_k^*; t_k);$$

here, as  $k$  varies, any two intervals  $(t_k, t_k^*) \subset \mathbb{R}$  are disjoint, and, for each value of  $k$ , the set  $I_k$  comprises the indices  $j$  for which  $(t_j^*; t_j)$  coincides with  $(t_k^*; t_k)$ . Now, by independence,

$$\mathbb{E}\left(\prod_j J_{w_j}(t_j^*; t_j)\right) = \prod_k \mathbb{E}\left(\prod_{j \in I_k} J_{w_j}(t_k^*; t_k)\right),$$

and the proof will be completed if we show that there is at least a value of  $k$  for which

$$\mathbb{E}\left(\prod_{j \in I_k} J_{w_j}(t_k^*; t_k)\right) = 0.$$

Since

$$\sum_k \sum_{j \in I_k} \|w_j\| = \sum_j \|w_j\| \notin \mathbb{N},$$

at least one of the inner sums is not an integer and we may apply the next Lemma:

**Lemma 3.1** *Assume that  $w_1, \dots, w_\ell$ , are words with  $\sum_j \|w_j\| \notin \mathbb{N}$ . Then, for each  $t_0 \geq 0$  and  $h > 0$ ,*

$$\mathbb{E}\left(\prod_j J_{w_j}(t_0 + h; t_0)\right) = 0.$$

**Proof:** By using repeatedly the shuffle relations, the product of iterated integrals may be rewritten as a sum of iterated integrals corresponding to the words  $w'_i$  resulting from shuffling the  $w_j$ ,  $j = 1, \dots, \ell$ . As noted above each  $w'_i$  has the non-integer weight  $\sum_j \|w_j\|$  and we may use the last item of Proposition 2.3.  $\square$

$\square$

The following result will provide, among other things, an easier proof of the last point of Proposition 3.1.

**Proposition 3.2 (Structure of  $\mathbb{E}\tilde{J}(t_0 + h; t_0)$ )** *There exist numbers  $\nu_1, \nu_2, \dots, \nu_n$  such that*

$$\mathbb{E}\tilde{J}(t_0 + h; t_0) = \prod_{i=1}^n \exp(h\nu_i \mathfrak{f}_i),$$

where the  $\mathfrak{f}_i$ 's are infinitesimal generators of the subsystems involved in the splitting, i.e.

$$\mathfrak{f}_i = a_{i_1} + a_{i_2} + \dots + a_{i_j} + \frac{A_{i_1}^2}{2} + \dots + \frac{A_{i_{j'}}^2}{2}$$

with  $a_l \in \mathcal{A}_{det}$ ,  $A_l \in \mathcal{A}_{sto}$ .

**Proof:** Assume that

$$\tilde{J}(t_0 + h; t_0) = \prod_{i=1}^N J^{(i)}(t_0 + c_i h; t_0 + b_i h).$$

Then

$$\mathbb{E}\tilde{J}(t_0 + h; t_0) = \prod_{i=1}^N \mathbb{E}J^{(i)}(t_0 + c_i h; t_0 + b_i h)$$

because either the iterated integrals  $J^{(i)}(t_0 + c_i h; t_0 + b_i h)$  are related to disjoint intervals, in which case they are independent, or they share the same interval so that the Brownian motions involved must be different, hence they are also independent. Kolmogorov's backward equation (2.11) allows to conclude.  $\square$

For example, consider the problem (3.1)

$$dx(t) = \underbrace{f_a(x(t))dt + f_A(x(t)) \circ d\mathcal{B}_A(t)}_{(1)} + \underbrace{f_B(x(t)) \circ d\mathcal{B}_B(t)}_{(2)},$$

and the split step integrator  $\tilde{\phi}_{t_0+h;t_0} = \phi_{t_0+h;t_0}^{(2)} \circ \phi_{t_0+h;t_0}^{(1)}$  that we studied before. Thus,

$$\begin{aligned} \mathbb{E}\tilde{J}(t_0 + h; t_0) &= \mathbb{E}\left(J^{(1)}(t_0 + h; t_0)J^{(2)}(t_0 + h; t_0)\right) \\ &= \mathbb{E}J^{(1)}(t_0 + h; t_0)\mathbb{E}J^{(2)}(t_0 + h; t_0). \end{aligned}$$

Recalling Kolmogorov's backward equation (2.11),

$$\mathbb{E}\tilde{J}(t_0 + h; t_0) = \exp\left(h\left(a + \frac{A^2}{2}\right)\right) \exp\left(h\frac{B^2}{2}\right).$$

□

Observe that Proposition 3.2 proves easily the last item of Proposition 3.1: given a word  $w$ ,

$$\mathbb{E}\tilde{J}(t_0 + h; t_0)_w = \left(\prod_{i=1}^n \exp(h\nu_i f_i), w\right),$$

but  $\prod_{i=1}^n \exp(h\nu_i f_i)$  comprises no term with semiinteger degree.

### 3.3 Strong and weak order conditions

Our purpose is now to obtain the strong and weak order conditions for split-step integrators. Let  $\tilde{\phi}_{t_0+h;t_0}$  denote a split-step integrator for the SDE (2.3). According to the preceding constructions, for each event  $\omega$  there exists a point  $\tilde{J}(t_0 + h; t_0) \in \mathcal{G}_{sh}$  such that the random formal series

$$\delta(t_0 + h; t_0) := \tilde{J}(t_0 + h; t_0) - J(t_0 + h; t_0) \in \mathbb{R}\langle\langle \mathcal{A} \rangle\rangle \quad (3.7)$$

encodes the Taylor expansion of both strong and local errors. Focus on the *strong error*

$$\mathbb{E}|\tilde{\phi}_{t_0+h;t_0}(x_0) - \phi_{t_0+h;t_0}(x_0)|.$$

Plugging  $\tilde{\phi}_{t_0+h;t_0}(x_0) - \phi_{t_0+h;t_0}(x_0) = W_{\delta(t_0+h;t_0)}(x_0)$  above, the strong error now reads

$$\mathbb{E}|W_{\delta(t_0+h;t_0)}(x_0)| = \mathbb{E}\left|\sum_w \delta_w(t_0 + h; t_0) f_w(x_0)\right|. \quad (3.8)$$

On the other hand, the *weak error*

$$|\mathbb{E}^{x_0} \chi(\phi_{t_0+h;t_0})(x_0) - \mathbb{E}^{x_0} \chi(\tilde{\phi}_{t_0+h;t_0})(x_0)|$$

may also be rewritten in terms of word series operators by inserting

$$\chi(\phi_{t_0+h;t_0})(x_0) - \chi(\tilde{\phi}_{t_0+h;t_0})(x_0) = D_{\delta(t_0+h;t_0)}\chi(x_0)$$

in it. In such a way, the weak error reads

$$\begin{aligned} |\mathbb{E} \sum_w \delta_w(t_0 + h; t_0) D_w \chi(x_0)| &= \left| \sum_w \mathbb{E} \delta_w(t_0 + h; t_0) D_w \chi(x_0) \right| \\ &= |D_{\mathbb{E}\delta(t_0+h;t_0)}\chi(x_0)|. \end{aligned}$$

The preparations above prove another main result:

**Theorem 3.2** *For a splitting integrator as above, the local error possesses a word series expansion*

$$\tilde{\phi}_{t,t_0}(x_0) - \phi_{t,t_0}(x_0) = W_{\delta(t_0,h)}(x_0) = \sum_{n \in (1/2)\mathbb{N}} \sum_{\|w\|=n} \delta_w(t_0, h) f_w(x_0), \quad (3.9)$$

with coefficients

$$\delta_w(t_0, h) = \tilde{J}_w(t_0 + h; t_0) - J_w(t_0 + h; t_0), \quad w \in \mathcal{W},$$

that, in any  $L^p$  norm,  $1 \leq p < \infty$ , satisfy, uniformly in  $t_0 \geq 0$ ,

$$\|\delta_w(t_0, h)\|_p = \mathcal{O}(h^{\|w\|}), \quad h \mapsto 0^+.$$

In addition, for each observable  $\chi$ , conditional on  $x_0$ ,

$$\mathbb{E}\chi(\tilde{\phi}_{t,t_0}(x_0)) - \mathbb{E}\chi(\phi_{t,t_0}(x_0)) = \sum_{n \in \mathbb{N}} \sum_{\|w\|=n} (\mathbb{E}\delta_w(t_0, h)) D_w \chi(x_0), \quad (3.10)$$

Observe that the theorem implies that the *strong order conditions*

$$\boxed{\tilde{J}_w(t_0 + h; t_0) = J_w(t_0 + h; t_0), \quad \|w\| = 0, 1/2, 1, \dots, \mu, \quad \mu \in (1/2)\mathbb{N},} \quad (3.11)$$

ensure that the series in (3.9) only comprises terms of size  $\mathcal{O}(h^{\mu+1/2})$ . If, for a given alphabet  $\mathcal{A}$  and given coefficients  $\tilde{J}_w(t_0 + h; t_0)$ , one demands errors  $\mathcal{O}(h^{\mu+1/2})$  for all possible choices of the vector fields  $f_a, f_A$ , then the conditions (3.11) are not only sufficient but also *necessary*. This happens because, as it is easy to show, in such a scenario, the word basis functions are mutually independent. However this consideration is not of much practical value; splitting integrators are useful because they are adapted to the specific structure of the problem being solved and therefore one is interested in the behavior for individual problems not in catering for all possible choices of  $f_a, f_A$ . The best way to deal with specific problems is to write down, up to the desired order, the word series expansions of the true and numerical solutions and compare them after taking into account the shuffle relations and the specific expressions of the word basis functions. For instance, if, for the problem at hand, a word basis function  $f_w$  vanishes identically, then it is clearly not necessary to impose the associated order condition in (3.11).

Similar considerations apply to the *weak order conditions*

$$\mathbb{E}\tilde{J}_w(t_0 + h; t_0) = \mathbb{E}J_w(t_0 + h; t_0), \quad \|w\| = 0, 1, 2, \dots, \nu, \quad \nu \in \mathbb{N}, \quad (3.12)$$

which ensure that the series in (6.10) only comprises terms of size  $\mathcal{O}(h^{\nu+1})$ .

The conditions (3.11)–(3.12) are similar to those found in [12] for stochastic Runge-Kutta integrators (however [12] only shows that a condition corresponding to (3.12) implies that the expectation of the local error is  $\mathcal{O}(h^{\nu+1})$ ; arbitrary observables  $\chi$  are not considered there).

### 3.3.1 Independent order conditions

It should be pointed out that, since both  $J(t_0 + h; t_0)$  and  $\tilde{J}(t_0 + h; t_0)$  satisfy the shuffle relations, the conditions in (3.11) corresponding to different words are not independent from one another. For instance, from the shuffle  $a \sqcup a = 2aa$ ,  $a \in \mathcal{A}$ , we may write

$$J_a(t_0 + h; t_0)^2 = 2J_{aa}(t_0 + h; t_0), \quad \tilde{J}_a(t_0 + h; t_0)^2 = 2\tilde{J}_{aa}(t_0 + h; t_0),$$

and therefore the order condition for the word  $aa$  is fulfilled if and only if the same happens for  $a$ . This problem is easily tackled with the help of the so-called Lyndon words [48].

#### Lyndon words

Assume that we give a total order  $<$  in the alphabet  $\mathcal{A}$ . Then, the set  $\mathcal{W}$  may be ordered with the so-called lexicographic order (cf. [48]). This is simply the order present in any dictionary, once the letters have been alphabetically ordered.

**Definition 3.1 (Lyndon words)** *A Lyndon word is a non-empty word such that whenever it is split into two non-empty subwords, the left word is always lexicographically less than the right word: if  $w, u, v \neq 1$  and  $w = uv$ , then  $u < v$ .*

In other words, a Lyndon word is a non-empty word which is smaller than its non-trivial proper right factors: if  $w, u, v \neq 1$  and  $w = uv$ ,  $w < v$ . Observe that all letters are Lyndon words but for example  $aa$  is not Lyndon word.

Let  $(\mathbb{R}\langle \mathcal{A} \rangle, +, \sqcup)$  denote the *shuffle algebra*, i.e. the additive group  $(\mathbb{R}\langle \mathcal{A} \rangle, +)$  with the shuffle product. In [46], Radford proves that the Lyndon words *generate the shuffle algebra and are algebraically independent*. As  $J(t_0 + h; t_0), \tilde{J}(t_0 + h; t_0) \in \mathcal{G}_{sh}$ , they both satisfy the shuffle relations so that we must only check the conditions 3.11 on Lyndon words. As  $aa$  is not a Lyndon word, the requirement

$$J_{aa}(t_0 + h; t_0) = \tilde{J}_{aa}(t_0 + h; t_0)$$

is redundant once

$$J_a(t_0 + h; t_0) = \tilde{J}_a(t_0 + h; t_0)$$

is fulfilled. Lyndon words are of the particular instance of Hall sets, see [48]. Early works on the use of Hall basis to identify subsets of independent order conditions may be credited to Murua and Sanz-Serna, see for example [40].



### 3.4 Error bounds

In what follows the deterministic vector fields  $f_a$ ,  $a \in \mathcal{A}_{det}$ , and the stochastic vector fields  $f_A$ ,  $A \in \mathcal{A}_{sto}$ , in (2.3) are assumed to be globally Lipschitz, thus guaranteeing existence and uniqueness of the initial value problem for (2.3) itself and for the split systems. The theorems below provide bounds for the weak local error and the mean square local error.

We begin with weak approximations. The third hypotheses used below is the same as inequality (2.17) in [36] which is key in establishing Theorem 2.5 in that reference. The first and second hypotheses just make explicit the differentiability requirements on  $f_a$ ,  $f_A$ , and  $\chi$  that have to be imposed to guarantee that  $D_w\chi$  makes sense when  $w$  has weight  $\nu + 1$ .

**Theorem 3.3 (Weak bounds. Stratonovich case)** *Let  $\nu$  be a positive integer. Assume that:*

- *The deterministic vector fields  $f_a$ ,  $a \in \mathcal{A}_{det}$ , are of class  $C^{2\nu}$ , while the stochastic vector fields  $f_A$ ,  $A \in \mathcal{A}_{sto}$ , are of class  $C^{2\nu+1}$ .*
- *The observable  $\chi$  is of class  $C^{2\nu+2}$  in  $\mathbb{R}^d$ .*
- *There is a constant  $C > 0$  such that for each  $x \in \mathbb{R}^d$  and each word  $w$  of weight  $\nu + 1$ :*

$$|D_w\chi(x)| \leq C(1 + |x|^2)^{1/2}.$$

- *The weak error conditions (3.12) hold.*

*Then there exists a constant  $K > 0$  such that for each  $x_0$ , each  $t_0 \geq 0$  and each  $h > 0$ :*

$$|\mathbb{E}\chi(\tilde{\phi}_{t_0+h,t_0}(x_0)) - \mathbb{E}\chi(\phi_{t_0+h,t_0}(x_0))| \leq K(1 + |x_0|^2)^{1/2}h^{\nu+1}$$

*(the expectation is conditional on  $x_0$ ).*

**Proof:** Define the residuals

$$R_{t_0,h}(x_0) = \chi(\phi_{t_0+h,t_0}(x_0)) - \sum_{\substack{n \in \mathbb{N}/2, \\ n \leq \nu}} \sum_{\|w\|=n} J_w(t_0, h) D_w\chi(x_0)$$

and

$$\tilde{R}_{t_0,h}(x_0) = \chi(\tilde{\phi}_{t_0+h,t_0}(x_0)) - \sum_{\substack{n \in \mathbb{N}/2, \\ n \leq \nu}} \sum_{\|w\|=n} \tilde{J}_w(t_0, h) D_w\chi(x_0)$$

associated with the true and numerical solution respectively. If the weak order conditions hold, we have, after using the fifth item in Proposition 2.3 and its counterpart in Proposition 3.1,

$$\mathbb{E}\chi(\tilde{\phi}_{t_0+h,t_0}(x_0)) - \mathbb{E}\chi(\phi_{t_0+h,t_0}(x_0)) = \mathbb{E}\tilde{R}_{t_0,h}(x_0) - \mathbb{E}R_{t_0,h}(x_0)$$

and our task is to successively bound the two terms the right hand-side.

For the theoretical solution, the standard stochastic Taylor expansion (see e.g. [31, Section 5.6] or [36, Section 1.2]) provides the following representation as an iterated Stratonovich integral

$$R_{t_0,h}(x_0) = \sum_w \int_{t_0}^{t_0+h} \circ d\mathcal{B}_{\ell_r}(s_r) \int_{t_0}^{s_r} \circ d\mathcal{B}_{\ell_{r-1}}(s_{r-1}) \cdots \int_{t_0}^{s_2} \circ d\mathcal{B}_{\ell_1}(s_1) D_w \chi(\phi_{s_1,t_0}(x_0));$$

here the  $\ell_i$  are deterministic or stochastic letters, the sum is extended to all words of the form  $w = \ell_1 \dots \ell_r$ , where  $\|\ell_2 \dots \ell_r\| = \nu$  and it is understood that, for a deterministic letter  $\ell_i$ ,  $d\mathcal{B}_{\ell}(s_i)$  means  $ds$ . We next rewrite the iterated Stratonovich integrals as combinations of iterated Ito integrals as in [31, Remark 5.2.8]; in each resulting iterated integral the sum of the weights of the letters of the Brownian motions that appear is  $\nu + 1$ . An application of [36, Lemma 2.2] then shows that, for a suitable constant  $L$ ,  $\mathbb{E}|R_{t_0,h}(x_0)|^2 \leq L^2(1 + |x_0|^2)h^{2\nu+2}$ , which yields  $\mathbb{E}|R_{t_0,h}(x_0)| \leq L(1 + |x_0|^2)^{1/2}h^{\nu+1}$ .

We now turn to the residual in the numerical solution. As in the proof of Theorem 4 in [41], we observe that, given an initial condition  $x_0 = x(t_0)$  and any splitting algorithm, the numerical solution after one step  $t_0 \rightarrow t_0 + h$  is the same as the value of true solution at  $t_0 + h$  of a time-dependent SDE in which the originally given vector fields are switched on and off as time evolves. For instance, in the simplest case where the SDE is  $dx = f_a(x)dt + f_A(x) \circ d\mathcal{B}_A(t)$  and the (Lie-Trotter) numerical scheme consists of advancing with  $dx = f_A(x) \circ d\mathcal{B}_A(x)$  and then with  $dx = f_a(x)dt$ , the time-dependent SDE is

$$dx = 1_{\{t_0+h/2 < t \leq t_0+h\}} f_a(x)dt + 1_{\{t_0 \leq t \leq t_0+h/2\}} f_A(x) \circ d\mathcal{B}_A(t_0 + 2(t - t_0)),$$

where  $t_0 \leq t \leq t_0 + h$  and  $1_{\{\cdot\}}$  denotes an indicator function. Using this observation the numerical residual may be bounded by reproducing the steps taken above to bound the residual of the true solution.  $\square$

The last result refers to the mean square error. The proof is parallel to that we have just presented and will be omitted.

**Theorem 3.4 (Strong bounds: Stratonovich case)** *Let  $\mu$  be a positive integer multiple of  $1/2$ . Assume that:*

- The deterministic vector fields  $f_a$ ,  $a \in \mathcal{A}_{det}$ , are of class  $C^{2\mu}$ , while the stochastic vector fields  $f_A$ ,  $A \in \mathcal{A}_{sto}$ , are of class  $C^{2\mu+1}$ .
- There is a constant  $C > 0$  such that for each  $x \in \mathbb{R}^d$  and each word  $w$  of weight  $\mu + 1$ :

$$|f_w(x)| \leq C(1 + |x|^2)^{1/2}.$$

- The strong error conditions (3.11) hold.

Then there exists a constant  $K > 0$  such that for each  $x_0$ , each  $t_0 \geq 0$  and each  $h > 0$ :

$$\left( \mathbb{E} |(\tilde{\phi}_{t_0+h, t_0}(x_0) - \phi_{t_0+h, t_0}(x_0))|^2 \right)^{1/2} \leq K(1 + |x_0|^2)^{1/2} h^{\mu+1/2}$$

(the expectation is conditional on  $x_0$ ).

## 3.5 Modified equations for Stratonovich split-step integrators

The abstract boundary value problems we dealt with in the first chapter can also be used in a similar fashion to that of Section 2.4 to express a split-step integrator as solution of a modified equation. The approach is the same: consider Theorem 1.10 and now set  $X := \tilde{J}(t_0 + h; t_0)$ . The conclusions are just the same. Thus we have many degrees of freedom, which enables us to perform derandomizations of the split-step integrator or modified equations without a drift term or even only perturbing the drift term, etc.

The fact that these modified equations *are not adapted* with respect to the natural filtration of the Brownian motions  $(\mathcal{F}_t)_{t \geq t_0}$  is worth remarking. The modified vector field  $\hat{f}$  will be in particular  $\mathcal{F}_{t_0+h}$ -measurable.

### 3.5.1 Derandomization

**Theorem 3.5 (Derandomization)** *For any split-step integrator for the SDE (2.3), there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the random ordinary differential equation*

$$y'(t) = \hat{f}(y(t))$$

with  $y(t_0) = x_0$  satisfies

$$y(t_0 + h) = \tilde{\phi}_{t_0+h, t_0}(x_0).$$

In particular

$$R = \frac{1}{h} \log(\tilde{J}(t_0 + h; t_0)).$$

**Proof:** It is enough to set  $X := \tilde{J}(t_0 + h; t_0)$  in Theorem 1.8.  $\square$

### 3.5.2 Drift-modified SDE's

**Theorem 3.6 (Drift-modified SDE's)** *For any split-step integrator for the SDE (2.3), there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the stochastic differential equation*

$$dy(t) = \hat{f}(y(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(y(t)) \circ d\mathcal{B}_A(t)$$

with  $y(t_0) = x_0$  is such that

$$y(t_0 + h) = \tilde{\phi}_{t_0+h;t_0}(x_0).$$

In particular the random series  $R$  is explicitly computable.

**Proof:** Simply apply Theorem 1.10 with  $X := \tilde{J}(t_0 + h; t_0)$ .  $\square$

### 3.5.3 Drift-free modified SDE

Again, Theorem 1.10 can be used to prove the subsequent results.

**Theorem 3.7 (Drift-free-modified SDE's, first version)** *For any split-step integrator for the SDE (2.3) and for any stochastic letter  $A' \in \mathcal{A}_{sto}$ , there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the stochastic differential equation*

$$dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A'}(t) + \sum_{A \in \mathcal{A}_{sto}, A \neq A'} f_A(y(t)) \circ d\mathcal{B}_A(t)$$

with  $y(t_0) = x_0$  is such that

$$y(t_0 + h) = \tilde{\phi}_{t_0+h;t_0}(x_0).$$

In particular the random series  $R$  is explicitly computable.

**Theorem 3.8 (Drift-free-modified SDE's, second version)** *For any split-step integrator for the SDE (2.3) and for any stochastic letter  $A' \in \mathcal{A}_{sto}$ , there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the stochastic differential equation*

$$dy(t) = \hat{f}(y(t)) \circ d\mathcal{B}_{A'}(t)$$

*with  $y(t_0) = x_0$  yields*

$$\boxed{y(t_0 + h) = \tilde{\phi}_{t_0+h;t_0}(x_0)}.$$

*In particular the random series  $R$  is explicitly computable.*

## **Part II**

### **Word series for Ito calculus**



# Chapter 4

## Combinatorial algebra 2

One of the most striking features of the Ito calculus is the so-called Ito's rule (cf. [35], [31], [45]); for  $\chi$  smooth and for any Brownian motion  $\mathcal{B}(t)$ :

$$d\chi(\mathcal{B}(t)) = \chi'(\mathcal{B}(t))d\mathcal{B}(t) + \frac{1}{2}\chi''(\mathcal{B}(t))dt. \quad (4.1)$$

Heuristically, the Ito interpretation assumes that, unlikely the Stratonovich calculus,  $(d\mathcal{B}(t))^2$  behaves like  $dt$  so that expanding  $d\chi(\mathcal{B}(t))$  (4.1) holds. This is a marked difference with both the Stratonovich and deterministic calculus, where the rules

$$d\chi(\mathcal{B}(t)) = \chi'(\mathcal{B}(t)) \circ d\mathcal{B}(t), \quad (4.2)$$

and

$$d\chi(f(t)) = \chi'(f(t))df(t) \quad (4.3)$$

for  $f$  at least Stieltjes integrable, are among the most familiar mathematical formulae. The relations (4.2) and (4.3) are the basis of the key role played by the shuffle product in the precedings chapters. In order to cater for the needs of the Ito calculus, we must consider a new type of combinatorics tailored to (4.1). Early works on the subject are credited to [26]. These combinatorics are based on a product similar to the shuffle product which is called *quasishuffle* product, see [48]. We present a construction of the quasishuffle product which goes hand in hand with stochastic calculus. Moreover, Ito's rule leads to an isomorphism between shuffle and quasishuffle, which was introduced in a completely different way by Hoffman [29].

### 4.1 The Ito extension of an alphabet

For convenience, we single out the two main peculiarities of Ito calculus (4.1):



- I1 When expanding  $\chi(\mathcal{B}(t))$  to first order, quantities such  $d\mathcal{B}^2(t)$  are no longer neglected, but amount to a  $dt$  contribution. In this way, additional terms appear with respect to the Stratonovich case.
- I2 Such additional terms involve second order derivatives.

The following discussion involving a simple Ito stochastic differential equation is useful to grasp the main ideas of the following constructions. Consider the following Ito SDE:

$$dx(t) = f_A(x(t))d\mathcal{B}(t), \quad x(0) = x_0 \in \mathbb{R}^d \quad (4.4)$$

for a smooth vector field  $f_A$ . Take now an observable  $\chi$ . Then,

$$\begin{aligned} \chi(x(t)) - \chi(x_0) &= \int_{t_0}^t \chi'(dx(s)) + \frac{1}{2} \int_{t_0}^t \chi''(dx(s), dx(s)) \\ &= \int_{t_0}^t \chi'(f_A(x(s)))d\mathcal{B}(s) + \frac{1}{2} \int_{t_0}^t \chi''(f_A(x(s)), f_A(x(s)))ds. \end{aligned} \quad (4.5)$$

From the Stratonovich viewpoint, the alphabet associated with (4.4) comprises only one letter, which is also stochastic,  $A$ . Look next at items *I1* and *I2* above. The most natural approach to *I1* appears to be the addition of one additional letter, closely related to  $A$ . For this reason, we shall denote it by  $\bar{A}$ . Secondly, to cater for the *I2* condition we associate  $\bar{A}$  with the second order differential operator

$$D_{\bar{A}}\chi = \frac{1}{2}\chi''(f_A, f_A).$$

Equation (4.5) is then recast as

$$\chi(x(t)) - \chi(x_0) = \int_{t_0}^t D_A\chi(x(s))d\mathcal{B}(s) + \frac{1}{2} \int_{t_0}^t D_{\bar{A}}\chi(x(s))ds. \quad (4.6)$$

For two independent Brownian motions, the convention  $d\mathcal{B}_A d\mathcal{B}_B = 0$  holds when computing the Taylor expansions, so that there is no need to add extra letters for products of Brownian motions. Note that  $D_{\bar{A}}$  corresponds to a  $dt$  differential in (4.6). Hence, such extra letters are all deterministic.

**Definition 4.1 (Ito extension of alphabets)** Consider an SDE-based alphabet  $\mathcal{A}$ , i.e.  $\mathcal{A}$  is finite and contains both stochastic and deterministic letters. Its Ito completion, to be denoted by  $\bar{\mathcal{A}}$ , is the alphabet comprising stochastic and deterministic letters such that:

- $\bar{\mathcal{A}}_{sto} = \mathcal{A}_{sto}$ , i.e. its stochastic letters are those stochastic of  $\mathcal{A}$ .

- $\bar{\mathcal{A}}_{det} = \mathcal{A}_{det} \cup \{\bar{A}\}_{A \in \mathcal{A}_{sto}}$ , i.e. its deterministic letters are split into two types: on the one hand, the deterministic letters of  $\mathcal{A}$ , on the other, those of the form  $\bar{A}$  for each stochastic letter  $A$ .

Notice that  $\mathcal{A} \subset \bar{\mathcal{A}}$ . The deterministic letters which are not of the form  $\bar{A}$  are called *purely deterministic*. Words in alphabets  $\bar{\mathcal{A}}$  inherit the grading: the weight of a word  $||w||$  is the sum of the weights of its letters with the convention that  $||1|| = 0$ . In the same way, the set  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  has a (finite local) graded algebra structure.

From now on we shall assume that an SDE-based alphabet  $\mathcal{A}$  and its completion  $\bar{\mathcal{A}}$  are given. In addition, a family  $(\mathcal{B}_A(t))_{A \in \mathcal{A}_{sto}}$  of *independent*<sup>1</sup> *Brownian motions* is also assumed to be given.

#### 4.1.1 Iterated integrals

We shall call *family of weights* to any family of continuous functions  $\lambda_x : \mathbb{R} \rightarrow \mathbb{R}$  indexed by letters  $x \in \mathcal{A}$ . In order to extend it to  $\bar{\mathcal{A}}$ , for a letter of the form  $\bar{A}$  we define

$$\lambda_{\bar{A}}(t) = \lambda_A^2(t).$$

We are in a position to introduce the so-called *Stratonovich iterated integrals*  $J(t; t_0; \lambda) \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  and the *Ito iterated integrals*  $I(t; t_0; \lambda) \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  with respect to a family of weights  $\lambda$ . The construction is already familiar. For example, for the Ito case,

$$\begin{aligned} I_1(t; t_0; \lambda) &= 1, \\ I_l(t; t_0; \lambda) &= \int_{t_0}^t \lambda_l(s) ds, \quad l \in \bar{\mathcal{A}}_{det}, \\ I_A(t; t_0; \lambda) &= \int_{t_0}^t \lambda_A(s) d\mathcal{B}_A(s), \quad A \in \bar{\mathcal{A}}_{sto}, \end{aligned}$$

and if  $wl$  and  $wA$  are the words obtained by appending the letter  $l$  or  $A$  at the end of  $w$ ,

$$\begin{aligned} I_{wl}(t, t_0) &= \int_{t_0}^t I_w(s; t_0) \lambda_l(s) ds, \quad l \in \bar{\mathcal{A}}_{det}, \\ I_{wA}(t, t_0) &= \int_{t_0}^t I_w(s; t_0) \lambda_A(s) d\mathcal{B}_A(s), \quad A \in \bar{\mathcal{A}}_{sto}. \end{aligned}$$

---

<sup>1</sup>They are all defined on the same filtered probability space.

In the particular case of  $\mathcal{A} = \{a, A\}$  and  $\bar{\mathcal{A}} = \{a, A, \bar{A}\}$ , taking  $\lambda_l(t) = 1$  for all letters and  $t = h, t_0 = 0$ , the series of  $\mathbb{R}\langle\langle\mathcal{A}\rangle\rangle$  whose coefficients are the iterated integrals are

$$\begin{aligned} J(h; 0) &= 1 + \mathcal{B}_A(h)A + ha + h\bar{A} + \frac{\mathcal{B}_A^2(h)}{2}A^2 + O\left(\frac{3}{2}\right), \\ I(h; 0) &= 1 + \mathcal{B}_A(h)A + ha + h\bar{A} + \frac{\mathcal{B}_A^2(h) - h}{2}A^2 + O\left(\frac{3}{2}\right), \end{aligned}$$

where  $O(\frac{3}{2})$  stands for terms with weight higher than or equal to  $\frac{3}{2}$ .

### 4.1.2 The $\theta$ mapping

The following paragraphs are devoted to introducing an essential tool for the following constructions.

**Theorem 4.1** *For any pair of times  $t > t_0 \geq 0$ , there exists a unique graded algebra automorphism  $\theta : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  such that*

$$\theta J(t; t_0; \lambda) = I(t; t_0; \lambda)$$

for each family of weights  $\lambda$  and for every event  $\omega$ .

**Proof:**

#### Step 1: Uniqueness of $\theta$

Fix a stochastic letter  $A$  and consider the family of weights  $\lambda_A(t) = \lambda_{\bar{A}}(t) = 1$  and  $\lambda_x(t) = 0$  for any other letter  $x$ . Then (for notational convenience we are taking  $t := h, t_0 := 0$ ),

$$\begin{aligned} J(h; 0; \lambda) &= 1 + \mathcal{B}_A(h)A + h\bar{A} + \frac{\mathcal{B}_A^2(h)}{2}AA + O\left(\frac{3}{2}\right), \\ \theta J(h; 0; \lambda) &= 1 + \mathcal{B}_A(h)\theta A + h\theta\bar{A} + \frac{\mathcal{B}_A^2(h)}{2}\theta(AA) + O\left(\frac{3}{2}\right), \\ I(h; 0; \lambda) &= 1 + \mathcal{B}_A(h)A + h\bar{A} + \frac{\mathcal{B}_A^2(h) - h}{2}AA + O\left(\frac{3}{2}\right). \end{aligned}$$

As  $\theta$  is a graded morphism,  $\theta A = A$  so that  $\theta(AA) = AA$ . In particular

$$h\theta\bar{A} = h\left(\bar{A} - \frac{1}{2}AA\right),$$

thus

$$\theta\bar{A} = \bar{A} - \frac{1}{2}AA = \bar{A} - \frac{A^2}{2}.$$

A similar argument shows that for every purely deterministic letter  $a$

$$\theta a = a.$$

### Step 2: Defining $\theta$

We define  $\theta : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  by steps. Firstly, for letters:  $\theta l = l$  for each stochastic or purely deterministic letter  $l$ , i.e. for  $l \in \mathcal{A}$ , and  $\theta\bar{A} = \bar{A} - \frac{1}{2}AA$ . Next, for words: given a word  $w = l_1 l_2 \dots l_N$ , then  $\theta w = \theta l_1 \theta l_2 \dots \theta l_N$  and  $\theta 1 = 1$ . Last, for formal series: if  $S = \sum_w S_w w \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ ,  $\theta S = \sum_w S_w \theta w$ . Now, let us focus on the mapping  $\rho : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  defined by the adjoint relation,

$$\theta S = \sum_w S_{\rho(w)} w \quad (4.7)$$

for every  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ . Observe that

$$(\theta S, w) = S_{\rho(w)} = (S, \rho(w)),$$

which means that  $\rho^T = \theta$ . By construction of  $\theta$ , we observe that for any given word  $w$ ,  $\rho(w)$  is the sum of the words  $u$  obtained by substitution of disjoint pairs  $AA$  in  $w$  by  $\bar{A}$ , divided by  $(-\frac{1}{2})^m$ , where  $m$  is the number of such substitutions. This is nothing but the fact  $(\rho(w), u) = (w, \theta u)$ . The case in which no substitution is made is also taken into consideration. For example:  $\rho(1) = 1$ ,  $\rho(\bar{A}) = \bar{A}$ ,  $\rho(AA) = AA - \frac{1}{2}\bar{A}$ ,  $\rho(AAA) = AAA - \frac{1}{2}\bar{A}A - \frac{1}{2}A\bar{A}$ .

### Step 3: Recursion on $\rho$

Let  $x, y$  be letters. By construction ( $w$  is any word):

$$(R1) \text{ If } x = y = A, A \text{ stochastic, } \rho(wxy) = \rho(wAA) = \rho(wA)A - \frac{1}{2}\rho(w)\bar{A},$$

$$(R2) \text{ In other case } \rho(wxy) = \rho(wx)y,$$

If we introduce a bracket in  $\mathcal{A}$  as  $[A, A] = \bar{A}$  for  $A$  stochastic and  $[x, y] = 0$  in other case, the above recursions become

$$\rho(wxy) = \rho(wx)y - \frac{1}{2}\rho(w)[x, y] \quad (4.8)$$

### Step 4: $\theta J = I$

Fix times  $t > t_0 \geq 0$  and a family of weights  $\lambda$ . For convenience we shall omit the dependence of the iterated integrals on these input data. Suppose  $I_w = J_{\rho(w)}$  for each  $N$ -graded word. Let  $u$  be a word of degree  $N + \frac{1}{2}$ . Then;

Case 1:  $u = wAA$  with  $A$  stochastic.

$$\begin{aligned} I_u &= \int I_{wA}\lambda_A(t)d\mathcal{B}_A = \int I_{wA}\lambda_A(t) \circ d\mathcal{B}_A - \frac{1}{2} \int I_w\lambda_A^2(t)dt = \\ &= \int J_{\rho(wA)}\lambda_A(t) \circ d\mathcal{B}_A - \frac{1}{2} \int J_{\rho(w)}\lambda_{\bar{A}}(t)dt = J_{\rho(wA)A} - \frac{1}{2}J_{\rho(w)\bar{A}} = \\ &= J_{\rho(wAA)}. \end{aligned}$$

Case 2:  $u = wxy \neq wAA$  with  $A$  stochastic. Set  $\mathcal{B}_x(t) := t$ ,  $d\mathcal{B}_x(t) = dt$  if  $x$  is deterministic (the same for  $y$ ). Under these assumptions  $\langle \mathcal{B}_x(t), \mathcal{B}_y(t) \rangle = 0$ , i.e. its quadratic variation is null.

$$\begin{aligned} I_u &= \int I_{wx}\lambda_y(t)d\mathcal{B}_y = \int I_{wx}\lambda_y(t) \circ d\mathcal{B}_y(t) = \int J_{\rho(wx)}\lambda_y(t) \circ d\mathcal{B}_y \\ &= J_{\rho(wx)y} = J_{\rho(wxy)}. \end{aligned}$$

Thus, the claim holds.  $\square$

Theorem 4.1 implies that for every polynomial  $f \in \mathbb{R}\langle \bar{\mathcal{A}} \rangle$ ,

$$J_{\rho(f)} = (J, \rho(f)) = (\theta J, f) = (I, f) = I_f. \quad (4.9)$$

Note that  $\theta^{-1} : \mathbb{R}\langle \bar{\mathcal{A}} \rangle \rightarrow \mathbb{R}\langle \bar{\mathcal{A}} \rangle$  is characterized by

- $\theta^{-1}\bar{A} = \bar{A} + \frac{1}{2}AA$  for  $A$  stochastic,
- $\theta^{-1}x = x$  in other case.

Given two words  $w, w'$ ,

$$(w, w') = (\rho\rho^{-1}w, w') = (\rho^{-1}w, \theta w') = (w, (\rho^{-1})^T \theta w'),$$

where  $^T$  indicates the transposition. In particular  $(\rho^{-1})^T = \theta^{-1}$  so that by using the same combinatorial argument that for  $\theta$  and  $\rho$ , we check that for any word  $w$ ,  $\rho^{-1}(w)$  is the sum of the words obtained by substitution of disjoint pairs  $AA$  in  $w$  by  $\bar{A}$ , divided by  $(\frac{1}{2})^m$ , where  $m$  is the number of such substitutions. As before, the case where no substitution is made is also considered. For example:  $\rho^{-1}(1) = 1$ ,  $\rho^{-1}(\bar{A}) = \bar{A}$ ,  $\rho^{-1}(AA) = AA + \frac{1}{2}\bar{A}$ ,  $\rho^{-1}(AAA) = AAA + \frac{1}{2}\bar{A}\bar{A} + \frac{1}{2}A\bar{A}$ .

### 4.1.3 Completeness

**Definition 4.2** We call a family  $(S_i)_{i \in I} \subset \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  complete if its orthogonal in  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  is the null element, i.e. if  $f \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$  is such that

$$(S_i, f) = 0$$

for every  $i \in I$ , necessarily  $f = 0$ .

The following theorem often provides shortcuts when proving algebraic relations. Sometimes it is easy to check them first on the iterated integrals, which is enough because of their completeness.

**Theorem 4.2** For any interval  $[t_0, t_1]$ ,  $t_1 > t_0$ , the families  $(J(t; t_0))_{t_0 \leq t \leq t_1}$ ,  $(I(t; t_0))_{t_0 \leq t \leq t_1}$  are complete on  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  with probability 1.

**Proof:** We will only check that  $I$  is complete: the Stratonovich case follows from  $J_f = I_{\rho^{-1}(f)}$ . Take a non-zero  $f \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ , almost surely orthogonal to  $I$ , and suppose that there exists no non-zero polynomial in  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  with a degree lower than that of  $f$  which is orthogonal to  $I(t; t_0)$  a.s.. Split the stochastic process

$$t \mapsto (I(t; t_0), f) = \sum_{ul: l \in \mathcal{A}_{det}} (f, ua)I(t; t_0)_{ua} + \sum_{uA: A \in \mathcal{A}_{sto}} (f, uA)I(t; t_0)_{uA},$$

into the sum of a smooth function and a term which is either nowhere differentiable or equal to zero. As  $(I(t; t_0), f) = 0$  a.s., for all  $t$

$$0 = \sum_{ul: l \in \mathcal{A}_{det}} (f, ua)I(t; t_0)_{ua} = \int \sum_{ua: a \in \mathcal{A}_{det}} (f, ul)I(t; t_0)_u dt, \quad (4.10)$$

$$0 = \sum_{uA: A \in \mathcal{A}_{sto}} (f, uA)I(t; t_0)_{uA} = \int \sum_{uA: A \in \mathcal{A}_{sto}} (f, uA)I(t; t_0)_u d\mathcal{B}_A(t). \quad (4.11)$$

By the induction argument we deduce that  $\sum_{ua: a \in \mathcal{A}_{det}} (f, ua)u = 0$ , hence  $f_{ua} = 0$  for  $a \in \mathcal{A}_{det}$ . Now, observe that, for each stochastic letter  $A$ , there exists a polynomial  $\hat{f}_A$  of degree lower than that of  $f$  such that  $f = \sum_{A \in \mathcal{A}_{sto}} \hat{f}_A A$ . Hence,  $(I(t; t_0), f) = \sum_{A \in \mathcal{A}_{sto}} (I(t; t_0), \hat{f}_A A)$ . Set  $\varphi_A(t) := (I(t; t_0), \hat{f}_A A)$ . The second equation in (4.10) is then rewritten as

$$0 = \sum_{A \in \mathcal{A}_{sto}} \varphi_A(t).$$

But  $\mathbb{E}(\varphi_A(t)) = 0$  and

$$\mathbb{E}(\varphi_A^2(t)) = - \sum_{B \in \mathcal{A}_{sto}, B \neq A} \mathbb{E}(\varphi_A(t)\varphi_B(t)) = 0,$$

because  $\mathbb{E}(\varphi_A(t)\varphi_B(t)) = 0$  as  $\mathcal{B}_A$  and  $\mathcal{B}_B$  are independent Wiener processes if  $A \neq B$ . Well-known properties of the stochastic integral ensure that  $\varphi_A(t) = 0$  for each letter  $A$ , thus  $(I(t; t_0), \hat{f}) = 0$ , for all times  $t > t_0 \geq 0$ . To conclude,  $f_w = 0$  for every word  $w \in \bar{\mathcal{A}}$ , as desired.  $\square$

## 4.2 Products and coproducts on $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$

### 4.2.1 Shuffle product

Consider defined on  $\bar{\mathcal{A}}$  the shuffle product and let us denote by  $\delta$  denote its co-product. Recall that this means nothing but

$$(w \sqcup w', u) = (w \otimes w', \delta u)$$

for arbitrary words  $w, w', u$ . Theorem 1.15, which is one of the main results of the first chapter, states that, when choosing  $\lambda_l(t) = 1$ , the Stratonovich iterated integrals satisfy the shuffle relations. As pointed out before, this was already known by Chen, see [20]. It is quite easy to check that the result also holds for arbitrary families of weights.

**Theorem 4.3 (Chen relations)** *For any two given  $p, q \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ , times  $t > t_0 \geq 0$  and for any family of weights  $\lambda$ , equality holds in*

$$J_p(\lambda)J_q(\lambda) = J_{p \sqcup q}(\lambda). \quad (4.12)$$

*for every event. The classical Chen relations are the particular case in which  $p, q$  are words.*

### 4.2.2 Quasishuffle product

Equation (4.12) is of extreme importance when analyzing a Stratonovich SDE solution operator. Recall that the shuffle property was intimately connected with multiplicative operators, which were key in proving results such as Theorems 2.2 and 2.1. The current situation is slightly different because of the new letters  $\bar{\mathcal{A}}$ . In order to replicate the Stratonovich/shuffle analysis, we ought to see whether there exists any product  $\bowtie$  on  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  such that

$$I_p(t; t_0; \lambda)I_q(t; t_0; \lambda) = I_{p \bowtie q}(t; t_0; \lambda) \quad (4.13)$$

for all polynomials  $p, q$  and events  $\omega$ . The best bet appears to be the use of the mapping  $\rho$ . By dropping the time dependence of the iterated integrals, we readily check that

$$I_p I_q = J_{\rho(p)} J_{\rho(q)} = J_{\rho(p) \sqcup \rho(q)} = I_{\rho^{-1}(\rho(p) \sqcup \rho(q))}.$$

This is the cue for us to define the quasishuffle product:

**Definition 4.3 (Quasishuffle product)** *Given two words  $u, v$  on  $\bar{\mathcal{A}}$ , we define its quasishuffle product as*

$$u \bowtie v = \rho^{-1}(\rho(u) \sqcup \rho(v)) \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle. \quad (4.14)$$

For example,  $A \bowtie A = \rho^{-1}(\rho(A) \sqcup \rho(A)) = \rho^{-1}(2AA) = 2AA + \bar{A}$ . We shall briefly see that this product coincides with the quasishuffle product introduced by Hoffman. Moreover,  $\rho^{-1}$  is the so-called *Hoffman exponential*, [29]. As usual, infinite bilinearity extends the product to  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ .

**Theorem 4.4 (Gaines relations)** *For each family of weights  $\lambda$ , times  $t > t_0 \geq 0$  and events  $\omega$ ,*

$$I_p(t; t_0; \lambda)I_q(t; t_0; \lambda) = I_{p \bowtie q}(t; t_0; \lambda),$$

*for any pair of  $p, q \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ . When  $p, q$  are words, the above equality just amounts to the so-called Gaines relations, see [26].*

The density of iterated integrals provides a short path to prove the following recurrence, similar to (1.20).

**Lemma 4.1 (Quasishuffle recursion)** *Let  $u, v$  be words and  $x, y$  letters. Then,*

$$ux \bowtie vy = (ux \bowtie v)y + (u \bowtie vy)x + (u \bowtie v)[x, y]. \quad (4.15)$$

**Proof:** Let  $u, v$  be words and  $x, y$  letters. Once again, set  $\mathcal{B}_l(t) := t$  for  $l$  deterministic and  $\mathcal{B}_0(t) = 0$ . The Gaines relations and Ito formula give

$$\begin{aligned} I_{ux \bowtie vy} &= I_{ux}I_{vy} = \int I_{ux}I_v d\mathcal{B}_x(t) + \int I_{vy}I_u d\mathcal{B}_y(t) + \int I_u I_v d\mathcal{B}_{[x,y]}(t) \\ &= I_{(ux \bowtie v)y} + I_{(u \bowtie vy)x} + I_{(u \bowtie v)[x,y]}. \end{aligned}$$

The completeness of  $I$ , see Theorem (4.2), provides the recursion.  $\square$

**Corollary 4.1** *The quasishuffle product is of the form of a quasishuffle product in the sense of Hoffman, see [29].*

The quasishuffle coproduct is the infinite-  $\mathbb{R}$ - linear algebra homomorphism<sup>2</sup>  $\kappa : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \otimes \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  satisfying:

---

<sup>2</sup>It is sometimes referred as *de-quasishuffle*



(D1)  $\kappa\bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A$ , for each stochastic letter  $A$ .

(D2)  $\kappa x = x \otimes 1 + 1 \otimes x$  in other case.

**Theorem 4.5 (Quasishuffle product duality)** For any words  $w, w', u$ :

$$(w \bowtie w', u) = (w \otimes w', \kappa u),$$

i.e.  $\kappa$  is the coproduct of the quasishuffle product  $\bowtie$ .

**Proof:** Let  $w$  be a word and suppose

$$\kappa w = \sum_{u,v} (w, u \bowtie v) u \otimes v, \quad (4.16)$$

where the sum ranges over all the pairs of words  $u, v$ . If  $A$  is a stochastic letter, the multiplicative property of  $\delta$  implies that  $\delta(w\bar{A}) = \delta w \delta \bar{A} = (\sum_{u,v} (w, u \bowtie v) u \otimes v) (\bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A)$ , i.e.

$$\begin{aligned} \delta(w\bar{A}) &= \sum_{u,v} (w, u \bowtie v) u \bar{A} \otimes v + \sum_{u,v} (w, u \bowtie v) u \otimes v \bar{A} + \sum_{u,v} (w, u \bowtie v) u A \otimes v A \\ &= \sum_{u,v} (w\bar{A}, u \bowtie v) u \otimes v. \end{aligned}$$

The same argument works for a letter  $l$  which is not of the form  $\bar{A}$ , always bearing in mind that  $\kappa l = l \otimes 1 + 1 \otimes l$ .  $\square$

### 4.3 Algebra and group-like elements

The analogy between the shuffle and quasishuffle product suggests paying attention to the same structures we considered in Section 1.2.4. In this case, as both shuffle and quasishuffle are defined in the same framework, both groups and algebras are mixed. This fact provides many interesting connections.

- The **algebra-like elements** (sh stands for shuffle and qsh for quasishuffle):

$$\mathfrak{g}_{sh} = \{R \in \mathbb{R}\langle\langle\bar{A}\rangle\rangle : \delta R = R \otimes 1 + 1 \otimes R\}. \quad (4.17)$$

$$\mathfrak{g}_{qsh} = \{R \in \mathbb{R}\langle\langle\bar{A}\rangle\rangle : \kappa R = R \otimes 1 + 1 \otimes R\}. \quad (4.18)$$

We shall see that, as it is expected, these elements are closely related to derivations  $Der(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ . The sets  $\mathfrak{g}_{sh}$  and  $\mathfrak{g}_{qsh}$  are both linear spaces that may be endowed with Lie algebra structure by considering the bracket

$$[S, R] = SR - RS.$$

These algebra-like elements are called *shuffle-primitive* and *quasishuffle-primitive* respectively.

- The **group-like elements** (sh stands for shuffle, qsh for quasishuffle):

$$\mathcal{G}_{sh} = \{R \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle : \delta R = R \otimes R\}, \quad (4.19)$$

$$\mathcal{G}_{qsh} = \{R \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle : \kappa R = R \otimes R\}. \quad (4.20)$$

The sets  $\mathcal{G}_{sh}$  and  $\mathcal{G}_{qsh}$  play a similar role to that of  $Aut(\mathcal{O}) \subset \mathcal{L}(\mathcal{O})$ . In particular, they both have a group structure.

The definitions above, especially those of the coproducts, easily provide the following characterizations:

**Theorem 4.6 (Characterization of group-like elements)** *Given  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ :*

- (1)  $S \in \mathcal{G}_{sh}$  if and only if  $S_{u \sqcup v} = S_u S_v$  for every pair of words  $u, v$  and  $S_1 = 1$  (Chen relations, cf. [20]).
- (2)  $S \in \mathcal{G}_{qsh}$  if and only if  $S_{u \bowtie v} = S_u S_v$  for every pair of words  $u, v$  and  $S_1 = 1$  (Gaines relations, cf. [26]).

**Proof:** The same argument of Theorem 1.1 is enough to prove the quasishuffle case.  $\square$

**Theorem 4.7 (Characterization of algebra-like elements)** *Given  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ :*

- (1)  $S \in \mathfrak{g}_{sh}$  if and only if  $S_{u \sqcup v} = 0$  for every pair of words  $u, v$  ( $S$  is ortogonal to shuffles).
- (2)  $S \in \mathfrak{g}_{qsh}$  if and only if  $S_{u \bowtie v} = 0$  for every pair of words  $u, v$  ( $S$  is ortogonal to quasishuffles).

**Proof:** The same argument of Theorem 1.2 is also enough in the quasishuffle case.  $\square$

Similarly to Theorem 1.3 we have the next result:

**Theorem 4.8** *It holds that*

$$(1) \exp(\mathfrak{g}_{sh}) = \mathcal{G}_{sh}, \text{ hence } \log(\mathcal{G}_{sh}) = \mathfrak{g}_{sh}.$$

$$(2) \exp(\mathfrak{g}_{qsh}) = \mathcal{G}_{qsh}, \text{ hence } \log(\mathcal{G}_{qsh}) = \mathfrak{g}_{qsh}.$$

**Proof:** The argument given in Theorem 1.3 also works.  $\square$

**Theorem 4.9 (Lie Group-Lie Algebra connection)** *The elements of  $\mathcal{G}_{sh}$  coincide with the velocities at the identity  $1 \in \mathcal{G}_{sh}$  of curves in  $\mathcal{G}_{sh}$ . The corresponding result holds for the group  $\mathcal{G}_{qsh}$  and the algebra  $\mathfrak{g}_{qsh}$ .*

**Proof:** The same argument we gave in Theorem 1.4 also works.  $\square$

The algebra morphism<sup>3</sup>  $\theta$  commutes with logarithms

$$\begin{aligned} \theta \log(S) &= \theta \sum_{n=0}^{\infty} \frac{(S-1)^n}{n} = \sum_{n=0}^{\infty} \frac{\theta(S-1)^n}{n} \\ &= \sum_{n=0}^{\infty} \frac{(\theta S - 1)^n}{n} = \log(\theta S), \end{aligned}$$

and exponentials:

$$\theta \exp(S) = \theta \sum_{n=0}^{\infty} \frac{S^n}{n!} = \sum_{n=0}^{\infty} \frac{\theta S^n}{n!} = \sum_{n=0}^{\infty} \frac{(\theta S)^n}{n!} = \exp(\theta S).$$

This fact will be useful for the following theorem:

**Theorem 4.10** *It holds that:*

$$(1) \theta \mathcal{G}_{sh} = \mathcal{G}_{qsh}.$$

$$(2) \theta \mathfrak{g}_{sh} = \mathfrak{g}_{qsh}.$$

**Proof:** Choose  $S \in \mathcal{G}_{qsh}$ . Then

$$(\theta^{-1}S)_u(\theta^{-1}S)_v = S_{\rho^{-1}(u)}S_{\rho^{-1}(v)} = S_{\rho^{-1}(u) \bowtie \rho^{-1}(v)} = S_{\rho^{-1}(u \sqcup \sqcup v)} = \theta^{-1}S_{u \sqcup \sqcup v},$$

i.e.  $\theta^{-1}S \in \mathcal{G}_{sh}$ . For  $R \in \mathcal{G}_{sh}$ ,

$$\theta R_u \theta R_v = R_{\rho(u)}R_{\rho(v)} = R_{\rho(u) \sqcup \rho(v)} = R_{\rho(u \bowtie v)} = \theta R_{u \bowtie v},$$

so  $\theta R \in \mathcal{G}_{qsh}$ . For the second statement:

$$\theta \mathfrak{g}_{sh} = \theta \log(\mathcal{G}_{sh}) = \log(\theta \mathcal{G}_{sh}) = \log(\mathcal{G}_{qsh}) = \mathfrak{g}_{qsh}.$$

$\square$

<sup>3</sup>Every algebra morphism  $\varphi : \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle \rightarrow \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$  commutes with logarithms and exponentials when defined.

## 4.4 Linear differential equations in $\mathcal{G}_{qsh}$

All the constructions we presented in Section 1.3 also hold under suitable translations from shuffle to quasishuffle, i.e. replacing  $\mathcal{G}_{sh}$  and  $\mathfrak{g}_{sh}$  by  $\mathcal{G}_{qsh}$  and  $\mathfrak{g}_{qsh}$ ,  $\delta$  by  $\kappa$ , etc. For instance, the quasishuffle version of Theorem 1.8 now reads:

**Theorem 4.11 (Boundary value problems)** *Consider the autonomous two-point boundary value problem*

$$S'(t) = S(t)R \quad (4.21)$$

- i)  $S(t_0) = 1$ ,
- ii)  $S(t_0 + h) = X$ , for some  $h > 0$  and  $X \in \mathcal{G}_{qsh}$ ,

with only  $X$  given. Then:

- There exists a unique choice of  $S(t)$  for all  $t$  and  $R$  such that (4.21) holds. In particular  $S(t) \in \mathcal{G}_{qsh}$  for all  $t$  and  $R \in \mathfrak{g}_{qsh}$ .
- The structure of (4.21) allows us to compute recursively the coefficients of  $S(t)$ .
- The unique solution is given explicitly by

$$S(t) = \exp((t - t_0)R), \quad R = \frac{1}{h} \log(X).$$

The quasishuffle version of Theorem 1.10 is:

**Theorem 4.12 (Several integrand boundary value problems)** *Let  $R^{(1)}, R^{(2)}, \dots, R^{(M-1)} \in \mathfrak{g}_{qsh}$  and  $X \in \mathcal{G}_{qsh}$  given. Then, for any election of  $x_i(t)$   $1 \leq i \leq M$  and  $h > 0$ , there exists a unique  $R^{(M)} \in \mathfrak{g}_{sh}$  and a curve in the group  $t \mapsto S(t)$  such that*

$$dS(t) = \sum_{i=1}^M S(t)R^{(i)}dx_i(t) \quad (4.22)$$

and

- i)  $S(0) = 1$ ,
- ii)  $S(h) = X$ .

Moreover, the structure of (4.22) allows to compute recursively the coefficients of  $S(t)$  and  $R$ .

There is a difficulty coming from the fact that the Ito interpretation of the integral does not satisfy, in general, the product rule (1.30). The next section aims to provide the existence of an auxiliary shuffle algebra narrowly connected with that of quasishuffle, which enables us to use the Stratonovich interpretation instead of Ito's one in order to apply the theorems of Section 1.3.

## 4.5 Stochastic duality

### 4.5.1 The auxiliary alphabet

Recall that in the Stratonovich case, letters of the form  $a$  and  $A$  were associated with the Lie derivatives  $D_a, D_A$  of certain vector fields. In particular, these differential operators are derivations of the algebra  $\mathcal{O}$ . This is consistent with Corollary 1.1, i.e.

$$\delta l = l \otimes 1 + 1 \otimes l \leftrightarrow D_l(\chi\varphi) = \chi D_l(\varphi) + \varphi D_l(\chi).$$

Thus, every letter is shuffle-primitive:

$$\boxed{\mathcal{A} \subset \mathfrak{g}_{sh}.$$

Even more,  $\bar{\mathcal{A}} \subset \mathfrak{g}_{sh}$  also holds. As pointed out before, it is of extreme importance to associate each letter with a vector field. But that is not *a priori* true in the Ito case. Both purely deterministic and stochastic letters are elements of the Lie algebra  $\mathfrak{g}_{qsh}$ , i.e. they are quasishuffle-primitive, but  $\delta \bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A$  for each stochastic letter  $A$  and so

$$\boxed{\bar{\mathcal{A}} \not\subset \mathfrak{g}_{qsh}.$$

Just recall that  $\bar{A}$  is related to a second order differential operator. Note that  $A^* := \theta \bar{A} = \bar{A} - \frac{1}{2}A^2 \in \mathfrak{g}_{qsh}$ <sup>4</sup>, i.e.

$$\boxed{\theta \bar{\mathcal{A}} \subset \mathfrak{g}_{qsh}.$$

This suggest considering  $\theta \bar{\mathcal{A}}$  as a new alphabet, instead of  $\bar{\mathcal{A}}$ .

**Definition 4.4 (Auxiliary alphabet)** *The set  $\mathcal{A}^* := \theta \bar{\mathcal{A}}$  comprising the elements of the form  $\theta l$  with  $l \in \bar{\mathcal{A}}$ , i.e. the elements of the form*

- $a$  with  $a$  purely deterministic ( $\theta a = a$ ).
- $A$  with  $A$  stochastic ( $\theta A = A$ ).
- $A^*$  with  $A \in \bar{\mathcal{A}}_{sto}$ .

*will be called the auxiliary alphabet. In particular, it contains both stochastic letters  $\mathcal{A}_{sto}^* := \theta(\bar{\mathcal{A}}_{sto}) = \mathcal{A}_{sto}$  and deterministic ones  $\mathcal{A}_{det}^* := \theta(\bar{\mathcal{A}}_{det})$ , i.e.*

$$\mathcal{A}_{det}^* := \{a \in \mathcal{A} \text{ purely deterministic}\} \cup \{A^* \in \mathcal{A} : A \text{ stochastic}\}. \quad (4.23)$$

From now on  $\mathcal{W}^*$  stands for the words in the alphabet  $\mathcal{A}^*$ . As  $a = \theta a$ ,  $A^* = \theta \bar{A}$ ,  $A = \theta A$ , the following result is holds:

$${}^4\delta(\bar{A} - \frac{A^2}{2}) = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A - \frac{A^2}{2} \otimes 1 + 1 \otimes -\frac{A^2}{2} - A \otimes A = \bar{A} - \frac{A^2}{2} \otimes 1 + 1 \otimes \bar{A} - \frac{A^2}{2}.$$

**Lemma 4.2** *The mapping*

$$\theta : \mathcal{W} \rightarrow \mathcal{W}^*$$

*is a bijection.*

### 4.5.2 Iterated integrals on $\mathcal{A}^*$ and $\theta$ morphism

Let us now turn our attention to the connection with Brownian motions. As  $A^*$  are all deterministic letters, the stochastic weight  $\|\cdot\|$  may be extended to  $\mathcal{A}^*$  and  $\mathcal{W}^*$ . Set  $\mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$  (respec.  $\mathbb{R}\langle\mathcal{A}^*\rangle$ ) for the algebra of non commutative formal series (respec. polynomials) on the letters of  $\mathcal{A}^*$ . The families of weights are defined on a similar manner: for each letter  $l \in \mathcal{A}$ , define  $\lambda_l(t)$ . Then,  $\lambda_{A^*}(t) = \lambda_A^2(t)$ . As an example, if the family of weights is such that  $\lambda_l = 1$ , the series associated with the Stratonovich iterated integrals, to be denoted by  $J^*(t; t_0) \in \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$ , is

$$\begin{aligned} J^*(t; t_0) &= \sum_{u \in \mathcal{A}^*} J_u^*(t; t_0)u = 1 + \sum_A J_A(t; t_0)A + (t - t_0) \sum_a a \\ &+ (t - t_0) \sum_{A^*} A^* + \sum_A J_{A^2}(t; t_0)A^2 + O\left(\frac{3}{2}\right). \end{aligned} \quad (4.24)$$

For the Ito case:

$$\begin{aligned} I^*(t; t_0) &= \sum_{u \in \mathcal{A}^*} I_u^*(t; t_0)u = 1 + \sum_A I_A(t; t_0)A + (t - t_0) \sum_a a \\ &+ (t - t_0) \sum_{A^*} A^* + \sum_A I_{A^2}(t; t_0)A^2 + O\left(\frac{3}{2}\right). \end{aligned} \quad (4.25)$$

Observe that every polynomial  $f \in \mathbb{R}\langle\mathcal{A}^*\rangle$  can be expanded into a polynomial on the letters of  $\mathcal{A}$  and reciprocally. For example

$$f = 3a + 5\bar{A} + 2A\bar{A} \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle,$$

but

$$\begin{aligned} f &= 3a + 5\bar{A} - 5\frac{A^2}{2} + 5\frac{A^2}{2} + 2A(\bar{A} - \frac{A^2}{2} + \frac{A^2}{2}) \\ &= 3a + 5A^* + \frac{5}{2}A^2 + AA^*A^2 \in \mathbb{R}\langle\mathcal{A}^*\rangle. \end{aligned}$$

Thus  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle = \mathbb{R}\langle\mathcal{A}^*\rangle$  as sets, providing the set equality  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle = \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$ . This allows us to consider the algebra morphism  $\theta : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  as a mapping  $\theta : \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle \rightarrow \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$ . The following theorem is a translation of Theorem 4.1:

**Theorem 4.13** *The morphism  $\theta$  may also be considered as a mapping  $\theta : \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle \rightarrow \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$ . In particular, it is the unique graded algebra morphism satisfying that  $\theta J^*(\lambda) = I^*(\lambda)$  for every family of weights. The morphism  $\rho$  may also be considered as a mapping  $\mathbb{R}\langle\mathcal{A}^*\rangle \rightarrow \mathbb{R}\langle\mathcal{A}^*\rangle$ . In this case, it also holds that  $\rho^T = \theta$ .*

Now, we shall present an important theorem. Let us begin by developing a few terms of  $J^*$  in terms of words  $\mathcal{W}$ . We only have to use the identity  $A^* = \bar{A} - \frac{A^2}{2}$ . For the sake of simplicity, we set  $t_0 = 0$  and  $h = t - t_0$ .

$$\begin{aligned} J^* &= 1 + \sum_A J_A A + h \sum_a a + h \sum_{A^*} A^* + \sum_A J_{A^2} A^2 + O\left(\frac{3}{2}\right) \\ &= 1 + \sum_A J_A A + h \sum_a a + h \sum_A \left(\bar{A} - \frac{A^2}{2}\right) + \sum_A J_{A^2} A^2 + O\left(\frac{3}{2}\right) \\ &= 1 + \sum_A J_A A + h \sum_a a + h \sum_A \bar{A} + \sum_A \left(J_{A^2} - \frac{h}{2}\right) A^2 + O\left(\frac{3}{2}\right), \end{aligned} \quad (4.26)$$

but  $J_A = I_A$  and  $J_{A^2} - \frac{h}{2} = I_{A^2}$ , hence

$$\begin{aligned} J^* &= 1 + \sum_A I_A A + h \sum_a a + h \sum_A \bar{A} + \sum_A (I_{A^2}) A^2 + O\left(\frac{3}{2}\right), \\ &= I + O\left(\frac{3}{2}\right). \end{aligned} \quad (4.27)$$

This is no coincidence:

**Theorem 4.14** *For any family of weights  $\lambda$ ,*

$$\boxed{J^*(\lambda) = I(\lambda)}.$$

**Proof:** According to Lemma 4.2, for every word  $w \in \mathcal{W}$  there exists a unique word  $u \in \mathcal{W}^*$  such that  $w = \theta^{-1}u$ . Hence,

$$I(\lambda) = \sum_{w \in \mathcal{W}} I_w(\lambda) w = \sum_{u \in \mathcal{W}^*} I_{\theta^{-1}u}(\lambda) \theta^{-1}u = \theta^{-1} \left( \sum_{u \in \mathcal{W}^*} I_{\theta^{-1}u}(\lambda) u \right),$$

but  $I_{\theta^{-1}u}(\lambda) = I_u^*(\lambda)$  because a letter  $l \in \mathcal{A}$  is deterministic if and only if  $\theta l$  is deterministic and  $\theta A = A$  for  $A \in \mathcal{A}_{sto}$ . This amounts to saying that

$$I(\lambda) = \theta^{-1} \left( \sum_{u \in \mathcal{W}^*} I_u^*(\lambda) u \right) = \theta^{-1} (I^*(\lambda)) = J^*(\lambda).$$

□

### 4.5.3 Shuffle $\sqcup^*$ and quasishuffle $\bowtie^*$

We may also *endow*  $\mathbb{R}\langle\mathcal{A}^*\rangle$  with both a shuffle and quasishuffle products. Firstly, let us approach the shuffle product. We shall denote by  $\sqcup^*$  the shuffle product on  $\mathbb{R}\langle\mathcal{A}^*\rangle$ , defined by shuffling the monomials  $a, A, A^*$  with  $a$  purely deterministic and  $A$  stochastic. For example

$$A^*A \sqcup^* A = AA^*A + 2A^*AA = A\bar{A}A + 2\bar{A}AA - \frac{3}{2}AAAA$$

but

$$A^*A \sqcup A = \bar{A}A \sqcup A - \frac{1}{2}AAA \sqcup A = A\bar{A}A + 2\bar{A}AA - 2AAAA.$$

Let  $\mathfrak{g}_{sh}(\mathcal{A}^*) \subset \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$  denote the Lie algebra of primitive elements w.r.t. this shuffle product and  $\mathcal{G}_{sh}(\mathcal{A}^*) := \exp(\mathfrak{g}_{sh}(\mathcal{A}^*))$  its Lie group. The quasishuffle product  $\bowtie^*$  on  $\mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$  is defined by means of the bracket  $[A, A]^* = A^*$  and equal to zero in other case. The group will be denoted by  $\mathcal{G}_{qsh}(\mathcal{A}^*)$  and its Lie algebra as  $\mathfrak{g}_{qsh}(\theta\mathcal{A})$ . We shall use the notation

$$\begin{aligned} \mathcal{G}_{sh}(\bar{\mathcal{A}}) &:= \mathcal{G}_{sh}, & \mathcal{G}_{qsh}(\bar{\mathcal{A}}) &:= \mathcal{G}_{qsh}, \\ \mathfrak{g}_{sh}(\bar{\mathcal{A}}) &:= \mathfrak{g}_{sh}, & \mathfrak{g}_{qsh}(\bar{\mathcal{A}}) &:= \mathfrak{g}_{qsh}. \end{aligned}$$

The similarity between

$$\theta A^* = \theta(\bar{A} - \frac{A^2}{2}) = \bar{A} - A^2 = A^* - \frac{A^2}{2},$$

and

$$\theta\bar{A} = \bar{A} - \frac{A^2}{2},$$

and  $\theta a = a$  ( $a$  purely deterministic) and  $\theta A = A$  ( $A$  stochastic), is enough to ensure (using standard arguments) that the following formulas hold

$$\theta\mathcal{G}_{sh}(\mathcal{A}^*) = \mathcal{G}_{qsh}(\mathcal{A}^*), \quad (4.28)$$

$$\theta\mathfrak{g}_{sh}(\mathcal{A}^*) = \mathfrak{g}_{qsh}(\theta\mathcal{A}), \quad (4.29)$$

and that the following theorems are true:

**Theorem 4.15** *For any two words in the alphabet  $\mathcal{A}^*$ ,  $\rho(u \bowtie^* v) = \rho(u) \sqcup^* \rho(v)$ , i.e.  $\rho$  is also a  $\sqcup^*$  shuffle -  $\bowtie^*$  quasishuffle homomorphism.*

**Theorem 4.16** *It holds that  $J^*(\lambda) \in \mathcal{G}_{sh}(\mathcal{A}^*)$  and  $I^*(\lambda) \in \mathcal{G}_{qsh}(\mathcal{A}^*)$  for any family of weights.*



### 4.5.4 Stochastic duality theorem

We got then two different algebras  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  and  $\mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle$  but with the same elements. If we look carefully at Theorem 4.14, we may observe that  $J^* \in \mathcal{G}_{sh}(\mathcal{A}^*)$  is developed into an element of  $\mathcal{G}_{qsh}$ . The following lines provide the ultimate relation between the structures based on the alphabets  $\bar{\mathcal{A}}$  and  $\mathcal{A}^*$ .

**Theorem 4.17 (Stochastic duality)** *The following equalities hold as sets:*

$$\mathcal{G}_{qsh}(\bar{\mathcal{A}}) = \mathcal{G}_{sh}(\mathcal{A}^*), \quad (4.30)$$

$$\mathfrak{g}_{qsh}(\bar{\mathcal{A}}) = \mathfrak{g}_{sh}(\mathcal{A}^*). \quad (4.31)$$

**Proof:** The formal calculation  $\mathcal{G}_{qsh}(\bar{\mathcal{A}}) = \theta\mathcal{G}_{sh}(\bar{\mathcal{A}}) = \mathcal{G}_{sh}(\mathcal{A}^*)$  is justified by the following argument. Take  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ . Then

$$\begin{aligned} S &= \theta^{-1}\theta S = \theta^{-1}\theta \sum_{w \in \mathcal{W}(\mathcal{A})} S_w w = \theta^{-1} \sum_{w \in \mathcal{W}(\mathcal{A})} S_w \theta w = \\ &= \theta^{-1} \sum_{u \in \mathcal{W}^*} S_{\theta^{-1}u} u \in \mathbb{R}\langle\langle\mathcal{A}^*\rangle\rangle. \end{aligned}$$

because  $\theta a = a, \theta A = A, \theta \bar{A} = A^*$  for  $a$  purely deterministic and  $A$  stochastic. By construction,  $S \in \mathcal{G}_{qsh}(\bar{\mathcal{A}})$  if and only if  $\sum_{u \in \mathcal{W}^*} S_{\theta^{-1}u} u \in \mathcal{G}_{qsh}(\mathcal{A}^*)$ , so  $S \in \mathcal{G}_{qsh}(\bar{\mathcal{A}})$  if and only if  $S = \theta^{-1} \sum_{u \in \mathcal{W}^*} S_{\theta^{-1}u} u \in \mathcal{G}_{sh}(\mathcal{A}^*)$ . By taking logarithms we obtain the algebra version.  $\square$

For  $A$  stochastic note that  $\bar{A} \in \mathfrak{g}_{sh}(\bar{\mathcal{A}})$  but  $\bar{A} \notin \mathfrak{g}_{qsh}(\bar{\mathcal{A}})$ . In a similar fashion,  $A^* \in \mathfrak{g}_{qsh}(\bar{\mathcal{A}}) \cap \mathfrak{g}_{sh}(\mathcal{A}^*)$  but  $A^* \notin \mathfrak{g}_{qsh}(\mathcal{A}^*)$ .

# Chapter 5

## Words for Stochastic Differential Equations. The Ito case

### 5.1 Extended alphabets for Ito SDE's

Consider the Ito stochastic initial value problem in  $\mathbb{R}^d$  (cf. [35], [31], [45]):

$$dx(t) = \sum_{a \in \mathcal{A}_{det}} f_a(x(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(x(t))d\mathcal{B}_A(t), \quad x(t_0) = x_0, \quad (5.1)$$

where:

- $\mathcal{A} = \mathcal{A}_{det} \cup \mathcal{A}_{sto}$  are finite sets.
- For each  $A \in \mathcal{A}_{det}$ ,  $\mathcal{B}_A(t)$  is a Brownian motion. All of them are defined on the same filtered probability space  $(\Omega, \mathcal{F}, P)$  and are independent.
- The vector fields  $f_l, l \in \mathcal{A}_{det} \cup \mathcal{A}_{sto}$  are smooth enough.

Consider now the Ito extension  $\bar{\mathcal{A}}$  of the alphabet  $\mathcal{A} = \mathcal{A}_{det} \cup \mathcal{A}_{sto}$ , see Definition 4.1. Recall that  $\bar{\mathcal{A}}_{det} = \mathcal{A}_{det} \cup \{\bar{A}\}_{A \in \mathcal{A}_{sto}}$  and  $\bar{\mathcal{A}}_{sto} = \mathcal{A}_{sto}$ . Needless to say, the constructions of the preceding chapter also hold. Let us make a list with the main things that have already been defined previously:

- $\mathcal{W}$  stands for the set of words in  $\bar{\mathcal{A}}$  and  $\mathbb{R}\langle \bar{\mathcal{A}} \rangle, \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$  represent respectively the non-commuting polynomials and the non-commuting formal series on the alphabet  $\bar{\mathcal{A}}$ .
- The weight  $\|\cdot\|$  of a word  $w = a_1 a_2 \dots a_n$  is the sum of the weights of its letters  $\|a_i\|$ .

- By means of shuffling the letters on  $\bar{\mathcal{A}}$  we define a shuffle product on  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  extended by infinite linearity to  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ . Similarly, there is a quasishuffle in both  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  and  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ , see Definition 4.3.
- The shuffle and quasishuffle products induce coproducts in  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ :  $\delta, \kappa : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \otimes \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ . They are the de-shuffle  $\delta$  and de-quasishuffle  $\kappa$ , respectively.
- There are two linear subspaces in  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  related to the shuffle and quasishuffle product via  $\delta$  or  $\kappa$  respectively: the shuffle Lie algebra  $\mathfrak{g}_{sh}$  and the quasishuffle Lie algebra  $\mathfrak{g}_{qsh}$ . Both are Lie algebras whose formal Lie groups are  $\mathcal{G}_{sh}$  and  $\mathcal{G}_{qsh}$ . The relation between groups and algebras is provided by the exponential/logarithm mappings.
- We may construct both Stratonovich and Ito iterated integrals in  $\bar{\mathcal{A}}$ . In particular, for any family of weights  $\lambda$ , it holds that  $J(\lambda) \in \mathcal{G}_{sh}$  and  $I(\lambda) \in \mathcal{G}_{qsh}$ .
- The mapping  $\theta : \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle \rightarrow \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  provides a connection between the shuffle and quasishuffle products. In particular  $\theta J(\lambda) = I(\lambda)$  and  $\theta$  maps bijectively  $\mathcal{G}_{sh}$  into  $\mathcal{G}_{qsh}$  and  $\mathfrak{g}_{sh}$  into  $\mathfrak{g}_{qsh}$ .
- The concept of  $h$ -coherence referred to the stochastic weight, see Definition 2.2, is similarly defined for curves  $h \mapsto S(h) \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$

## 5.2 Word basis operators

If we try to extend the  $\mathbb{R}$ -algebra morphism (1.34) defined in Section 1.4 to the alphabet  $\bar{\mathcal{A}}$ , we observe that, whereas  $D_a$  and  $D_A$  for  $a$  purely deterministic and  $A$  stochastic are well defined, a priori there is no vector field playing the role of  $f_{\bar{\mathcal{A}}}$ . Moreover, observe that

$$\kappa \bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A,$$

in particular  $\bar{A}$  is not quasishuffle primitive as we checked when studying the stochastic duality. If Theorem 1.11 is still true for the quasishuffle case, there is no point in trying to find a vector field  $f_{\bar{\mathcal{A}}}$  whose Lie Derivative is the linear operator  $D_{\bar{\mathcal{A}}}$ , because  $D_{\bar{\mathcal{A}}}$  is not even a derivation. The only way to go is define  $D_{\bar{\mathcal{A}}}$  as a second order differential operator. In this way, for the alphabet  $\bar{\mathcal{A}}$  we are lead to consider the following  $\mathbb{R}$ -algebra morphism

$$\begin{aligned} D : \mathbb{R}\langle\bar{\mathcal{A}}\rangle &\rightarrow \mathcal{L}(\mathcal{O}), \\ p &\mapsto D_p, \end{aligned}$$

defined for each stochastic or purely deterministic letter  $l$  as the Lie derivative of the vector field  $f_l$ , i.e.

$$D_a\chi = \nabla\chi \cdot f_a = \chi'(f_a) = \sum_{i=1}^d f_a^i \frac{\partial\chi}{\partial x^i},$$

and

$$D_{A\chi} = \nabla\chi \cdot f_A = \chi'(f_A) = \sum_{i=1}^d f_A^i \frac{\partial\chi}{\partial x^i}.$$

Instead for the deterministic letters of the form  $\bar{A}$ ,

$$D_{\bar{A}\chi} = \frac{1}{2} f_A^T \text{Hess}(\chi) f_A = \frac{1}{2} \chi''(f_A, f_A),$$

which is a second order differential operator. Obviously  $\text{Hess}$ ,  $\chi'$  and  $\chi''$  stand for the Hessian matrix, the first and the second differential of  $\chi$  respectively. For any word  $w = a_1 a_2 \dots a_n$  with length greater than 1, the algebra morphism feature of  $D$  provides the familiar composition recursion:

$$D_w = D_{a_1} \circ D_{a_2} \circ \dots \circ D_{a_n}. \quad (5.2)$$

Again,  $D_w$  is also a linear operator, but it is not a derivation. The algebra morphism  $D$  may be extended from  $\mathbb{R}\langle\bar{\mathcal{A}}\rangle$  to  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  by an infinite linearity argument, giving rise to the well-known *formal differential operators*  $D_S \in \mathcal{FO}$ , i.e. formal operators of the form

$$D_S = \sum_w S_w D_w, \quad D_S(\chi) = \sum_w S_w D_w(\chi),$$

where  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  and  $\chi \in \mathcal{O}$ .

### 5.2.1 Multiplicative operators and derivations

This section provides natural extensions from the shuffle case to the quasishuffle case of results obtained in Section 1.4.1. Let us get started with the decomposition of  $D_w(\chi\psi)$ , as a sum of products of operators of the form  $D_u$  acting upon the observables  $\chi, \psi$  themselves. Due to the presence of differential operators  $D_{\bar{A}}$ , which were not present in the shuffle case, Theorem 1.11 fails to hold. It is quite easy to find counterexamples, for instance:

$$\delta\bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A},$$

but

$$D_{\bar{A}}(\chi\psi) = \chi D_{\bar{A}}\psi + \psi D_{\bar{A}}\chi + D_{A\chi}D_A\psi \neq \chi D_{\bar{A}}\psi + \psi D_{\bar{A}}\chi.$$

Observe that

$$\kappa\bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A.$$

Obviously, we must replace the shuffle product/coproduct by the quasishuffle product/coproduct.

**Theorem 5.1 (Leibniz's Rule, quasishuffle case)** *Given any polynomial  $f \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$  and two observables  $\chi, \psi$ :*

$$D_f(\chi\psi) = \sum_{u,v} (\kappa f, u \otimes v) D_u\chi D_v\psi = \sum_{u,v} (f, u \bowtie v) D_u\chi D_v\psi.$$

**Proof:** By linearity, we shall only be concerned with the case in which  $f$  is a word. Let  $l$  be a letter which is not of the form  $\bar{A}$ . Then

$$\kappa l = l \otimes 1 + 1 \otimes l,$$

and

$$D_l(\chi\psi) = \chi D_l\psi + \psi D_l\chi = \sum_{u,v} (\kappa l, u \otimes v) D_u\chi D_v\psi.$$

In other case,

$$\kappa\bar{A} = \bar{A} \otimes 1 + 1 \otimes \bar{A} + A \otimes A,$$

and

$$D_{\bar{A}}(\chi\psi) = \chi D_{\bar{A}}\psi + \psi D_{\bar{A}}\chi + D_{A\chi}D_A\psi = \sum_{u,v} (\kappa\bar{A}, u \otimes v) D_u\chi D_v\psi.$$

The recursion argument in Theorem 1.11 is easily translated to this case by means of the definition of the de-quasishuffle coproduct  $\kappa$ .  $\square$

The next corollary also holds.

**Corollary 5.1** *Given any series  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  and two observables  $\chi, \psi$ :*

$$D_S(\chi\psi) = \sum_{u,v} (\kappa S, u \otimes v) D_u\chi D_v\psi = \sum_{u,v} (S, u \bowtie v) D_u\chi D_v\psi.$$

From this result we may readily deduce the following important result that parallels Theorem 1.12.

**Theorem 5.2** *Given any series  $S \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ :*

- The formal operator  $D_S$  is a formal automorphism of  $\mathcal{L}(\mathcal{O})$  if and only if  $S \in \mathcal{G}_{qsh}$ .
- The formal operator  $D_S$  is a formal derivation of  $\mathcal{L}(\mathcal{O})$  if and only if  $S \in \mathfrak{g}_{qsh}$ .

□

### 5.2.2 Word basis functions and composition rules

**Definition 5.1 (Word basis functions)** For each word  $w \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$  we define its word basis function as the mapping

$$f_w := D_w(Id) : \mathbb{R}^d \rightarrow \mathbb{R}^d,$$

where  $D_w(Id)$  is understood as acting component-wise and  $Id : x \mapsto x$  is the identity map of the space of states  $\mathbb{R}^d$ .

A few remarks:

- There is no restriction in defining  $f_p$  for polynomials  $p \in \mathbb{R}\langle\bar{\mathcal{A}}\rangle$ .
- Observe that, similarly to the shuffle case,  $f_1(x) = x$ .
- For words of the form  $w = u\bar{A}$  with  $u$  a (possibly empty) word, the word basis function  $f_w$  vanishes identically because  $D_{\bar{A}}(Id) = 0$ .

The concatenation recursion (5.2) provides the following recurrence formulas. Let  $w$  be a possibly empty word, Then

$$\text{R1 } f_{aw} = f'_w(f_a) \text{ for } a \text{ purely deterministic.}$$

$$\text{R2 } f_{Aw} = f'_w(f_A) \text{ for } A \text{ stochastic.}$$

$$\text{R3 } f_{\bar{A}w} = \frac{1}{2}f''_w(f_A, f_A) \text{ for } A \text{ stochastic.}$$

Observe that we can now recast the Ito Initial Value Problem (5.4) in the following way, similar to the Stratonovich formulation:

$$\boxed{dx(t) = \sum_{l \in \bar{\mathcal{A}}_{det}} f_l(x(t))dt + \sum_{A \in \bar{\mathcal{A}}_{sto}} f_A(x(t))d\mathcal{B}_A(t), \quad x(t_0) = x_0,} \quad (5.3)$$

because

$$\begin{aligned} \sum_{l \in \bar{\mathcal{A}}_{det}} f_l(x(t))dt &= \sum_{a \text{ purely deterministic}} f_a(x(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_{\bar{A}}(x(t))dt \\ &= \sum_{a \text{ purely deterministic}} f_a(x(t))dt = \sum_{a \in \Lambda} f_a(x(t))dt. \end{aligned}$$

The following important results parallel Theorems 2.1 and 2.2. The ideas are a simple translation of shuffle/quasishuffle concepts in their hypotheses and in their theses.

**Theorem 5.3 (Operator acting on word series)** *Choose  $S = S(h) \in \mathcal{G}_{qsh} \subset \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$   $h$ -coherent w.r.t. the weight. Then, for any  $\chi \in \mathcal{O}$  and  $x_0 \in \mathbb{R}^d$ :*

$$\boxed{D_S(\chi)(x_0) = \chi(W_S(x_0))}.$$

**Proof:** The proof is similar to that of the shuffle case. Observe that  $D_S$  is a multiplicative operator. Firstly, take a polynomial  $p$  and then observe that

$$\pi_j(W_S) = D_S(\pi_j),$$

where  $\pi_j$  is the  $j$ -projection. Now,

$$\pi_j \pi_k(W_S) = \pi_j(W_S) \pi_k(W_S) = D_S(\pi_j) D_S(\pi_k) = D_S(\pi_j \pi_k),$$

because  $D_S$  is multiplicative. The linearity shows that the result holds for every polynomial as desired.  $\square$

The same result also holds when replacing observables  $\chi$  by smooth mappings  $\mathbb{R}^d \rightarrow \mathbb{R}^d$ .

**Theorem 5.4 (Composition of word series)** *Choose  $h \mapsto S(h) \in \mathcal{G}_{sh}$  and  $h \mapsto R(h) \in \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$ ,  $h$ -coherent. Then,*

$$\boxed{W_{R(h)}(W_{S(h)}(x_0)) = W_{S(h)R(h)}(x_0)}.$$

**Proof:** The same idea of Theorem 2.2 works here.  $\square$

### 5.3 Solution operator

We present techniques similar to those of Section 2.2 in order to analyze the solution of the initial value problem (5.4):

$$dx(t) = \sum_a f_a(x(t))dt + \sum_A f_A(x(t))d\mathcal{B}_A(t), \quad x(t_0) = x_0. \quad (5.4)$$

As it is clearly understood that  $a$  ranges in the set of purely deterministic letters and  $A$  in the stochastic ones, we omit any mention to these sets for notational convenience. Let  $\mathbb{X}(t)$  denote the pullback of the solution operator. The main idea is to perform a Picard iteration. As a first step we analyze the expansion provided by the Ito rule, which is remarkably different from that provided by the chain rule. Take any observable  $\chi \in \mathcal{O}$ . Then

$$d\chi(x(t)) = \chi'(dx(t)) + \frac{1}{2}\chi''(dx(t), dx(t)).$$

Recall that this is nothing but a shorthand for

$$\chi(x(t)) = \chi(x_0) + \int_{t_0}^t \chi'(dx(t)) + \frac{1}{2} \int_{t_0}^t \chi''(dx(t), dx(t)). \quad (5.5)$$

Now, plug  $dx(t)$  into (5.5). Recall that the Brownian motions  $\mathcal{B}_A(t)$ ,  $A \in \mathcal{A}_{sto}$  are independent, which may be formally expressed by means of the product rule

$$d\mathcal{B}_A d\mathcal{B}_B = \begin{cases} dt & \text{if } A = B \\ 0 & \text{if } A \neq B \end{cases} \quad (5.6)$$

In this way,

$$\begin{aligned} d\chi(x(t)) &= \chi' \left( \sum_a f_a(x(t))dt + \sum_A f_A(x(t))d\mathcal{B}_A \right) \\ &\quad + \frac{1}{2} \sum_{A,B} \chi'' \left( f_A(x(t))d\mathcal{B}_A, f_B(x(t))d\mathcal{B}_B \right) \\ &= \sum_a \chi' \left( f_a(x(t)) \right) dt + \sum_A \chi' \left( f_A(x(t)) \right) d\mathcal{B}_A \\ &\quad + \frac{1}{2} \sum_A \chi'' \left( f_A(x(t)), f_A(x(t)) \right) dt. \end{aligned} \quad (5.7)$$

Rewrite now (5.7) in terms of the operators  $D_l$  for letters  $l$ :

$$d\chi(x(t)) = \sum_a D_a \chi(x(t))dt + \sum_A D_A \chi(x(t))d\mathcal{B}_A + \sum_A D_{\bar{A}} \chi(x(t))dt, \quad (5.8)$$



or

$$\begin{aligned} \chi(x(t)) = & \chi(x_0) + \sum_a \int_{t_0}^t D_a \chi(x(s)) ds + \sum_A \int_{t_0}^t D_A \chi(x(s)) d\mathcal{B}_A(s) \\ & + \sum_A \int_{t_0}^t D_{\bar{A}} \chi(x(s)) ds. \end{aligned} \quad (5.9)$$

The last equation may be recast in terms of an integral equation for the pullback  $\mathbb{X}(t)$ , leading to a straightforward Picard iteration.

$$\mathbb{X}(t) = Id + \sum_a \int_{t_0}^t \mathbb{X}(s) D_a ds + \sum_A \int_{t_0}^t \mathbb{X}(s) D_A d\mathcal{B}_A(s) + \sum_A \int_{t_0}^t \mathbb{X}(s) D_{\bar{A}} ds. \quad (5.10)$$

There would appear two different approaches.

### First approach: Picard Iteration

We may continue to perform a Picard iteration by inserting

$$\begin{aligned} \mathbb{X}(s) = & Id + \sum_a \int_{t_0}^s \mathbb{X}(r) D_a dr + \sum_A \int_{t_0}^s \mathbb{X}(r) D_A d\mathcal{B}_A(r) \\ & + \sum_A \int_{t_0}^s \mathbb{X}(r) D_{\bar{A}} dr, \end{aligned}$$

in the integrands of (5.10). Then

$$\begin{aligned} \mathbb{X}(t) = & Id + \\ & \sum_a \int_{t_0}^t \left( Id + \sum_a \int_{t_0}^s \mathbb{X}(r) D_a dr + \right. \\ & \left. \sum_A \int_{t_0}^s \mathbb{X}(r) D_A d\mathcal{B}_A(r) + \sum_A \int_{t_0}^s \mathbb{X}(r) D_{\bar{A}} dr \right) D_a ds \\ & + \sum_A \int_{t_0}^t \left( Id + \sum_a \int_{t_0}^s \mathbb{X}(r) D_a dr + \right. \\ & \left. \sum_A \int_{t_0}^s \mathbb{X}(r) D_A d\mathcal{B}_A(r) + \sum_A \int_{t_0}^s \mathbb{X}(r) D_{\bar{A}} dr \right) D_A d\mathcal{B}_A(s) \\ & + \sum_A \int_{t_0}^t \left( Id + \sum_a \int_{t_0}^s \mathbb{X}(r) D_a dr + \right. \\ & \left. \sum_A \int_{t_0}^s \mathbb{X}(r) D_A d\mathcal{B}_A(r) + \sum_A \int_{t_0}^s \mathbb{X}(r) D_{\bar{A}} dr \right) D_{\bar{A}} ds. \end{aligned}$$

Thus,

$$\mathbb{X}(t) = Id + \sum_A I_A(t; t_0)D_A + \sum_A I_a(t; t_0)D_a + \sum_A I_{\bar{A}}(t; t_0)D_{\bar{A}} + \cdots .$$

By iterating the process we obtain an expansion of  $\mathbb{X}(t)$  in terms of the Ito iterated integrals. We omit the calculations. The solution operator may be expanded as follows:

$$\boxed{\mathbb{X}(t) = Id + \sum_{w \neq 1} I_w(t; t_0)D_w = \sum_w I_w(t; t_0)D_w = D_{I(t; t_0)}. \quad (5.11)}$$

In particular,

$$\boxed{x(t) = \mathbb{X}(t)(Id) = D_{I(t; t_0)}(Id) = \sum_w I_w(t; t_0)f_w(x_0) = W_{I(t; t_0)}(x_0). \quad (5.12)}$$

### Second approach: Word series operator ansatz

Prior to performing all the calculations of previous epigraph, it may be useful to assume that there exists  $t \mapsto S(t) \in \mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$  such that

$$\mathbb{X}(t) = \sum_w S_w(t; t_0)D_w = D_{S(t; t_0)}.$$

Take a look at equation (5.10). In this way,

$$S(t; t_0) = 1 + \sum_a \int_{t_0}^t S(s; t_0)ads + \sum_A \int_{t_0}^t S(s; t_0)Ad\mathcal{B}_A + \sum_A \int_{t_0}^t S(s; t_0)\bar{A}ds, \quad (5.13)$$

or

$$dS(t; t_0) = 1 + \sum_a S(t; t_0)adt + \sum_A S(t; t_0)Ad\mathcal{B}_A(t) + \sum_A S(t; t_0)\bar{A}dt. \quad (5.14)$$

As expected,  $S(t; t_0)$  is constructed by means of the recursion:

$$\begin{aligned} S_1(t; t_0) &= 1, \\ S_l(t; t_0) &= \int_{t_0}^t ds, \quad l \in \bar{\mathcal{A}}_{det}, \\ S_A(t; t_0) &= \int_{t_0}^t d\mathcal{B}_A(s), \quad A \in \bar{\mathcal{A}}_{sto}, \end{aligned}$$

and if  $wl$  and  $wA$  are the words obtained by appending the letter  $l$  or  $A$  at the end of  $w$ ,

$$S_{wl}(t, t_0) = \int_{t_0}^t S_w(s; t_0) ds, \quad l \in \bar{\mathcal{A}}_{det},$$

$$S_{wA}(t; t_0) = \int_{t_0}^t S_w(s; t_0) d\mathcal{B}_A(s), \quad A \in \bar{\mathcal{A}}_{sto}.$$

Obviously,  $S(t; t_0) = I(t; t_0)$ .

Summing up the preceding lines, we present the following theorem:

**Theorem 5.5 (Solution of Ito SDE)** *The pullback of the solution operator of (5.4) admits a word basis operator expansion; in particular, for each event  $\omega$  it may be identified with the curve  $t \mapsto I(t; t_0) \in \mathcal{G}_{qsh}$  via*

$$\mathbb{X}(t) = D_{J(t; t_0)}.$$

### 5.3.1 The formal series $I(t; t_0)$

Recall that  $\theta J = I$ , thus  $(I, w) = (\theta J, w) = (J, \rho(w))$  and  $I_w = J_{\rho(w)}$  for every word  $w$ . This fact combined with Proposition 2.3 allows us to prove the following results:

**Proposition 5.1** *The iterated Ito integrals possess the following properties.*

- $I(t; t_0) \in \mathcal{G}_{qsh}$ .
- *The joint distribution of any finite subfamily of the family of random variables  $\{h^{-\|w\|} I_w(t_0 + h; t_0)\}_{w \in \mathcal{W}}$  is independent of  $t_0 \geq 0$  and  $h > 0$ .*
- $\mathbb{E} | I_w(t_0 + h; t_0) |^p < \infty$ , for each  $w \in \mathcal{W}$ ,  $t_0 \geq 0$ ,  $h > 0$  and  $p \in [0, \infty)$ .
- *For each  $w \in \mathcal{W}$  and any finite  $p \geq 1$ , the ( $t_0$ -independent)  $L^p$  norm of the random variable  $I_w(t_0 + h; t_0)$  is  $\mathcal{O}(h^{\|w\|})$ , as  $h \downarrow 0$ .*
- $\mathbb{E} I_w(t_0 + h; t_0) = 0$  whenever  $\|w\|$  is not an integer.

## 5.4 Expectation of the solution

This section is devoted to studying the expectation  $\mathbb{E}^{x_0} \chi(x(t))$  for observables  $\chi$ , the bedrock of the weak properties. Word basis operators prove to be useful in the matter in hand:

$$\mathbb{E}\mathbb{X}(t)\chi(x_0) := \mathbb{E}^{x_0} \chi(x(t)) = \mathbb{E}(\chi(x(t)) | x(t_0) = x_0), \quad (5.15)$$

thus

$$\boxed{\mathbb{E}\mathbb{X}(t) = D_{\mathbb{E}J(t;t_0)}}. \quad (5.16)$$

If  $w$  contains a stochastic letter, the stochastic process  $t \mapsto I_w(t; t_0)$  is a centered martingale, hence  $\mathbb{E}I_w(t; t_0) = 0$ . In other case, it is easily checked that

$$I_w(t; t_0) = \frac{h^{\|w\|}}{\|w\|!},$$

because we are computing the hyper-volume of an  $n$ - dimensional simplex.

In the following lines we study the infinitesimal generator (sometimes called the Kolmogorov operator) of an Ito SDE. We refer the reader to [35], [31] and [45].

**Definition 5.2 (Infinitesimal generator)** *Set*

$$\mathfrak{L} := \sum_{a \in \mathcal{A}_{det}} a + \sum_{A \in \mathcal{A}_{sto}} \bar{A}. \quad (5.17)$$

*Then, the operator  $D_{\mathfrak{L}} \in \mathcal{L}(\mathcal{O})$  is called the infinitesimal generator of the Ito SDE (5.4).*

**Proposition 5.2** *The formal series  $\mathbb{E}I(t; t_0)$  possesses the following properties;*

- *If  $\mathcal{A}_{sto}$  is not empty,  $\mathbb{E}I(t; t_0) \notin \mathcal{G}_{qsh}$ , i.e.  $\mathbb{E}\mathbb{X}_{t;t_0}$  is not a multiplicative operator.*
- *Kolmogorov Backward equation (see [35], [45], [31]):*

$$\boxed{\mathbb{E}I = \exp((t - t_0)\mathfrak{L})}, \quad (5.18)$$

*i.e.*

$$\boxed{\mathbb{E}\mathbb{X}_{t;t_0} = \exp((t - t_0)D_{\mathfrak{L}})}. \quad (5.19)$$

- $\mathbb{E}I_w(t; t_0) = 0$  provided that  $\|w\| \notin \mathbb{Z}$ .
- $h \mapsto \mathbb{E}I(t_0 + h; t_0)$  is  $h$  coherent.

**Proof:**

- For the first item:

$$(\mathbb{E}I_A)^2 = 0 \neq \mathbb{E}I_{2AA} + \mathbb{E}I_{\bar{A}} = h,$$

hence  $\mathbb{E}J(t; t_0) \notin \mathcal{G}_{qsh}$  provided that the stochastic alphabet is not empty. In other case,  $\mathbb{E}I(t; t_0) = I(t; t_0) \in \mathcal{G}_{qsh}$ .

- As  $\mathbb{E}I_1(t; t_0) = 1$ ,  $\mathbb{E}I(t; t_0)$  has a logarithm in  $\mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ , i.e. there exists  $R \in \mathfrak{m} \subset \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$  such that

$$\mathbb{E}I(t; t_0) = \exp(R(t; t_0)).$$

Again, use the ansatz  $R(t; t_0) = (t - t_0)X$ ,  $X \in \mathbb{R}\langle\langle\bar{\mathcal{A}}\rangle\rangle$ ; if  $t = t_0 + h$ ,

$$\partial_h \mathbb{E}J(t_0 + h; t_0) = \partial_h \exp(hX) = X \exp(hX).$$

Set  $h = 0$ ;

$$\partial_{h|h=0} \mathbb{E}I(t; t_0) = X.$$

Now,

$$\partial_{h|h=0} \mathbb{E}I(t; t_0) = \sum_w \partial_{h|h=0} \mathbb{E}I_w(t_0 + h; t_0)w.$$

The  $h$ -coherence ensures that  $\partial_{h|h=0} \mathbb{E}J_w(t_0 + h; t_0) = 0$  if  $\|w\| \geq 2$ , so we only have to compute

$$\partial_{h|h=0} \mathbb{E}I_a(t_0 + h; t_0) = \partial_{h|h=0} \mathbb{E}I_{\bar{A}}(t_0 + h; t_0) = 1.$$

Hence,

$$X = \sum_{a \in \mathcal{A}_{det}} a + \sum_{A \in \mathcal{A}_{sto}} \bar{A} = \mathcal{L}.$$

Observe that

$$\mathbb{E}X_{t; t_0} = \exp((t - t_0)D_{\mathcal{L}}).$$

As we pointed out in the Stratonovich case, this is the so-called backward Kolmogorov's equation or the Feymann-Kac formula, see [35], [45] or [31].

- Although we had already proved that  $\mathbb{E}I_w$  vanishes for words  $w$  containing stochastic letters, Kolmogorov's backward equation provides another approach. Indeed, observe that

$$\boxed{\mathbb{E}I_w(t; t_0) = (\mathbb{E}I, w) = (\exp(h\mathfrak{L}), w),}$$

but  $\mathfrak{L}$  contains no stochastic letters, and so does  $\exp(h\mathfrak{L})$ . In this way  $\mathbb{E}I_w(t; t_0) = 0$  for words containing stochastic letters. In other case, it is also straightforward to check that

$$\mathbb{E}I_w(t_0 + h; t_0) = \left( \sum_{n=0}^{\infty} \frac{(a + \dots + \bar{A} + \dots)^n h^n}{n!}, w \right) = \frac{h^{\|w\|}}{\|w\|!}.$$

□

Again, the equation  $\mathfrak{L} \notin \mathfrak{g}_{qsh}$  allows us to check that  $\mathbb{E}I(t; t_0) \notin \mathcal{G}_{qsh}$ . Observe that  $\mathfrak{L} \in \mathfrak{g}_{sh}$ , so that  $\mathbb{E}I(t; t_0) \in \mathcal{G}_{sh}$ . Of course:  $\mathbb{E}I(t; t_0) = I(t; t_0; \lambda)$  where

$$\lambda_x(t) = \begin{cases} 1 & \text{if } x \in \mathcal{A}_{det} \\ 0 & \text{if } x \in \mathcal{A}_{sto} \end{cases}. \quad (5.20)$$

In particular  $I(t; t_0; \lambda)$  may be seen to correspond to the coefficients of the expansion of a *deterministic problem*, which of course satisfy the shuffle relations. What is not obvious by a simple inspection of the series, is that

$$\theta \mathbb{E}I(t; t_0) = \sum_w (\mathbb{E}I)_{\rho(w)}(t; t_0) w = \sum_w \mathbb{E}(I_{\rho(w)})(t; t_0) w \in \mathcal{G}_{qsh}.$$

This is justified by the previously noted  $\mathbb{E}I(t; t_0) \in \mathcal{G}_{sh}$  and  $\theta : \mathcal{G}_{sh} \rightarrow \mathcal{G}_{qsh}$ . In the particular case in which there are only two deterministic letters,  $a, \bar{A}$ , we shall compute a few terms of the series just to check the quasishuffle relations. Set  $X := \theta \mathbb{E}I(t_0 + h; t_0)$ . Then

$$\begin{aligned} X &= \theta \left( 1 + ha + h\bar{A} + \frac{h^2}{2} (a^2 + a\bar{A} + \bar{A}a + \bar{A}^2) + O(3) \right) \\ &= 1 + ha + h\bar{A} - \frac{h}{2} A^2 + \frac{h^2}{2} (a^2 + a\bar{A} + \bar{A}a + \bar{A}^2) \\ &\quad - \frac{h^2}{4} (aA^2 + A^2a) + \frac{h^2}{8} A^4 + O(3). \end{aligned}$$

Now, for example,

- $X_A = 0$ ,

- $X_{2A^2} = -h$ ,
- $X_{\bar{A}} = h$ .

Hence

$$X_A^2 = 0 = X_{A \boxtimes A} = X_{2A^2 + \bar{A}} = -h + h = 0,$$

which was not obvious at all when examining  $\mathbb{E}I$ . The exponential mapping sheds some more light on this. Firstly, observe that

$$\mathbb{E}I(t_0 + h; t_0) = \exp(h(a + \cdots + \bar{A} + \cdots)).$$

Apply next  $\theta$  to both sides to obtain:

$$\begin{aligned} \theta \mathbb{E}I(t_0 + h; t_0) &= \theta \exp(h(a + \cdots + \bar{A} + \cdots)) = \exp(h(\theta a + \cdots + \theta \bar{A} + \cdots)) = \\ &= \exp(h(a + \cdots + A^* + \cdots)). \end{aligned}$$

Recall that  $A^* = \bar{A} - \frac{A^2}{2}$ , see Section 4.5. Moreover,  $a, A^* \in \mathfrak{g}_{sh}$ , so that  $\theta \mathbb{E}I(t; t_0) \in \mathcal{G}_{sh}$ , and

$$D_{\theta \mathbb{E}I(t; t_0)} \chi(x_0) = \chi(r(t)), \quad (5.21)$$

where  $r(t)$  is the solution of the ordinary differential equation

$$r'(t) = \sum_a f_a(r(t)) + \sum_A f_{A^*}(r(t)). \quad (5.22)$$

Observe that  $f_{A^*} = D_{A^*}(Id) = f_{\bar{A}} - \frac{1}{2}f_{A^2} = -\frac{1}{2}f_{A^2} = -\frac{1}{2}f'_A(f'_A)$ .

## 5.5 Ito-Stratonovich equivalence

In this section we first show an application of the results presented in Section 4.5; Theorem 4.14 provides the well-known equivalence between the Ito and Stratonovich interpretations of stochastic differential equations, namely:

**Theorem 5.6 (Ito-Stratonovich equivalence)** *The solution of the Stochastic Ito differential equation*

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t))dt + \sum_{A \in \Gamma} f_A(x(t))d\mathcal{B}_A(t),$$

*satisfies the Stochastic Stratonovich differential equation*

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t))dt + \sum_{A \in \Gamma} f_A(x(t)) \circ d\mathcal{B}_A(t) + \sum_{A \in \Gamma} f_{A^*}(x(t))dt,$$

**Proof:** Combining Theorems 5.5 and 4.14:

$$\mathbb{X}_{t;t_0} = \sum_{w \in \mathcal{W}} I_w(t; t_0) D_w = D_{I(t;t_0)} = D_{J^*(t;t_0)} = \sum_{u \in \mathcal{W}^*} J_u^*(t; t_0) D_u. \quad (5.23)$$

Recall that  $\mathcal{A}^*$  comprises all purely deterministic letters, i.e.  $\Lambda \subset \mathcal{A}^*$ , all stochastic ones,  $\Gamma \subset \mathcal{A}^*$ , and those of the form  $A^* = \bar{A} - \frac{A^2}{2}$ . Now, Theorem 2.4 ensures that the pullback  $\mathbb{Y}_{t;t_0}$  of the solution operator of

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t)) dt + \sum_{A \in \Gamma} f_A(x(t)) \circ d\mathcal{B}_A(t) + \sum_{A \in \Gamma} f_{A^*}(x(t)) dt$$

has the word basis operator expansion

$$\mathbb{Y}_{t;t_0} = \sum_{u \in \mathcal{W}^*} J_u^*(t; t_0) D_u,$$

so that  $\mathbb{X}_{t;t_0} = \mathbb{Y}_{t;t_0}$ .  $\square$

Observe that the cited theorem also provides the following well-known (and useful) corollary:

**Corollary 5.2 (Additive noise coincidence)** *Assume that the stochastic vector fields of the Stochastic Ito differential equation*

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t)) dt + \sum_{A \in \Gamma} f_A(x(t)) d\mathcal{B}_A(t)$$

*are constant.<sup>1</sup> Then, there is no distinction between interpretations:  $x(t)$  also satisfies the Stochastic Stratonovich differential equation*

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t)) dt + \sum_{A \in \Gamma} f_A(x(t)) \circ d\mathcal{B}_A(t).$$

**Proof:** According to Theorem 5.6,  $x(t)$  also satisfies

$$dx(t) = \sum_{a \in \Lambda} f_a(x(t)) dt + \sum_{A \in \Gamma} f_A(x(t)) \circ d\mathcal{B}_A(t) + \sum_{A \in \Gamma} f_{A^*}(x(t)) dt.$$

But if  $f_A$  is constant,

$$f_{A^*} = -\frac{f'_A(f_A)}{2} = 0,$$

as desired.  $\square$

---

<sup>1</sup>This is sometimes referred as a *additive noise*.



## 5.6 Derandomization of Stochastic Ito Differential Equations

We shall address now the Ito derandomization which goes hand in hand with the Stratonovich case.

**Theorem 5.7 (Ito Derandomization)** *Take  $h > 0$  and consider the SDE (5.4). Then there exists a random formal vector field  $\hat{f}(x) = W_{R_h}(x)$  with  $R_h \in \mathfrak{g}_{qsh}$  such that the random ordinary differential equation*

$$y'(t) = \hat{f}(y(t)),$$

*is a modified equation for (5.4), i.e. whenever the initial condition  $y(t_0) = x_0$  is given, the equation at time  $t_0 + h$  reaches the value*

$$\boxed{y(t_0 + h) = x(t_0 + h)}.$$

*Furthermore, the random algebra point  $R_h$  is readily computed via*

$$R_h = \frac{1}{h} \log(I(t_0 + h; t_0)).$$

**Proof:** Choose  $h > 0$  and consider Theorem 4.11. Set  $X = I(t_0 + h; t_0) \in \mathcal{G}_{qsh}$ . Then, there exist a random  $R \in \mathfrak{g}_{qsh}$  and a random curve in the quasishuffle group  $t \mapsto S(t) \in \mathcal{G}_{qsh}$  such that  $S'(t) = S(t)R$  and whose end-points are  $S(t_0) = 1, S(t_0 + h) = X$ . In addition:

$$\begin{aligned} R &= \frac{1}{h} \log(X) = \frac{1}{h} \log(I(t_0 + h; t_0)) \\ &= \frac{1}{h} (I(t_0 + h; t_0) - 1) - \frac{1}{2h} (I(t_0 + h; t_0) - 1)^2 + \frac{1}{3h} (I(t_0 + h; t_0) - 1)^3 \\ &\quad + \dots \\ &= \sum_{A \in \mathcal{A}_{sto}} \frac{I_A(t_0 + h; t_0)}{h} A + \sum_{l \in \mathcal{A}_{det}} l + O(0.5), \end{aligned}$$

and

$$S(t) = \exp((t - t_0)R).$$

Set  $y(t) := W_{S(t)}(x_0)$  and consider the formal vector field  $\hat{f} := W_R = \sum_w R_w f_w$ . Then,

$$y'(t) = \partial_t \sum_w S(t) f_w(x_0) = \sum_w S'(t) f_w(x_0) = \sum_w S(t) R f_w(x_0) = W_{S(t)R}(x_0).$$

Using (5.4) we write

$$y'(t) = W_{S(t)R}(x_0) = W_R(W_{S(t)}(x_0)) = \hat{f}(y(t)).$$

Observe  $R = R(h)$  is  $h$ -coherent with  $N = -1$ . In this way,

$$\boxed{y'(t) = \hat{f}(y(t))}, \quad (5.24)$$

with  $y(t_0) = x_0$ ,  $\boxed{y(t_0 + h) = x(t_0 + h)}$ .  $\square$

As we did in the Stratonovich case, we stress the following elementary facts:

- In a particular problem, many word basis functions may vanish, reducing the number of coefficients of  $R_w$  that must be computed.
- Again, if the value  $x(t_0 + h)$  is to be computed with a certain accuracy, there is no need for solving the SDE itself. There is a random ordinary differential equation interpolating the SDE at time  $t_0$  whose only randomness stems from the random variables  $I_w(t_0 + h; t_0)$ .
- The vector field  $\hat{f}$  is  $\mathcal{F}_{t_0+h}$ -measurable where  $(\mathcal{F}_t)_{t \geq t_0}$  is the natural filtration of the Brownian motions.

### Examples

- **Brownian Motion.** Consider  $d = 1$  and

$$\boxed{dx(t) = d\mathcal{B}(t), \quad x(0) = 0.}$$

The solution is  $x(t) = \mathcal{B}(t)$ . Set  $f_A(x) = 1$ . Unlike the Stratonovich case, the alphabet  $\mathcal{A}$  comprises two letters:  $A$  and  $\bar{A}$ . It is easily checked that  $f_w = 0$  unless  $w = A, 1$ . The word series for this alphabet are of the form

$$W_S(x_0) = \sum_w S_w f_w(x_0) = S_1 x_0 + S_A.$$

A second major difference from the Stratonovich case is that we can no longer obtain a simple formula for  $\log(I)$ . A priori, this is not a cause for concern because we are only interested in the values:  $\log(I)_1 = 0$  and  $\log(I)_A = I_A$ . In such a way, the formal vector field is  $W_R(x_0) = \mathcal{B}(h)/h$  so that the random differential equation (2.13) simply reads

$$y'(t) = W_R(y(t)) = \frac{\mathcal{B}(h)}{h}.$$

As expected, we obtain the same result as we already obtained in the Stratonovich case. The comparison between solutions reads

$$\boxed{y(t) = \frac{\mathcal{B}(h)}{h}t, \quad x(t) = \mathcal{B}(t).} \quad (5.25)$$

Again, observe that the vector field  $f_R$  is not adapted.

- **Exponential Brownian Motion.** Focus on

$$\boxed{dx(t) = x(t)d\mathcal{B}(t), \quad x(0) = 1.}$$

The solution is  $x(t) = e^{\mathcal{B}(t)-t/2}$ . Notice the correction  $-t/2$  is a characteristic element of the Ito Calculus. Set  $f_A(x) = x$ . Our alphabet is  $\mathcal{A} = \{A, \bar{A}\}$ . It holds that  $f_{A^n}(x) = x$ ,  $f_1(x) = x$ . In other case,  $f_w = 0$ . The word series for this alphabet are of the form

$$W_S(x_0) = \sum_w S_w f_w(x_0) = x_0 \sum_{n=0}^{\infty} S_{A^n}.$$

As we noted previously, we can no longer obtain a simple expression for  $\log(I)$ , but there are some shortcuts for not computing one by one the coefficients  $R_w$ . Observe that we are only interested in the values  $R_{A^n}$ , and that  $R \in \mathfrak{g}_{qsh}$ . This just amounts to saying that  $R$  is ortogonal to quasishuffles. For example,  $R_{\bar{A}} = 1$ , but  $A \bowtie A = 2A^2 + \bar{A}$ . Hence,

$$R_{A \bowtie A} = 0 = 2R_{A^2} + R_{\bar{A}}.$$

Then,  $R_{A^2} = -\frac{1}{2}$ . It is not difficult to check by a recursion argument that  $R_{A^n} = 0$  for  $n \geq 3$ . Then,

$$\hat{f}(x) = x\left(\frac{\mathcal{B}(h)}{h} - \frac{1}{2}\right).$$

The derandomization of the Exponential Brownian Motion is

$$y'(t) = W_R(y(t)) = \left(\frac{\mathcal{B}(h)}{h} - \frac{1}{2}\right)y(t),$$

so that

$$\boxed{y(t) = \exp\left(\left(\frac{\mathcal{B}(h)}{h} - \frac{1}{2}\right)t\right), \quad x(t) = \exp(\mathcal{B}(t) - \frac{t}{2}).} \quad (5.26)$$

- **Geometric Brownian Motion.** Consider

$$\boxed{dx(t) = \mu x(t)dt + \sigma x(t)d\mathcal{B}(t), \quad x(0) = 1,}$$

whose solution is the well-known geometric Brownian motion with drift parameter  $\mu$  and stochastic volatility  $\sigma$  (cf. [35] or [45])

$$x(t) = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma\mathcal{B}(t)\right).$$

Set  $f_a(x) = \mu x$  and  $f_A(x) = \sigma x$ . The alphabet is  $\mathcal{A} = \{a, A, \bar{A}\}$ . Denote by  $a(w)$  the numbers of  $a$ 's in the word  $w$ . Similarly for letters  $A, \bar{A}$ . Then  $f_w(x) = \mu^{a(w)}\sigma^{A(w)}x$  in case  $\bar{A}(w) = 0$ . If  $\bar{A}(w) \neq 0$ ,  $f_w(x) = 0$ . The word series for this alphabet are of the form

$$W_S(x) = x \sum_{w:\bar{A}(w)=0} S_w.$$

The same recursion argument based on the quasishuffle relations leads to  $R_{A^2} = -1/2$  and  $R_{A^n} = R_{a^{n-1}} = 0$  for  $n \geq 3$ . Recall next equation (2.16). The same argument allows us to state

$$\sum_{w:a(w)=j,|w|=n,\bar{A}(w)=0} w = a^j \sqcup A^{n-j}. \quad (5.27)$$

Of course  $a^j \sqcup A^{n-j} = a^j \bowtie A^{n-j}$ , meaning that

$$\sum_{w:a(w)=j,|w|=n,\bar{A}(w)=0} w = a^j \bowtie A^{n-j}. \quad (5.28)$$

Observe that words  $w$  such that  $a(w) = j$ ,  $|w| = n$ ,  $\bar{A}(w) = 0$  share the same word basis function, namely

$$f_w(x) = \mu^j \sigma^{n-j} x.$$

In such a way:

$$\sum_{a(w)=j,|w|=n,\bar{A}(w)=0} R_w f_w = \mu^j \sigma^{n-j} x \sum_{a(w)=j,|w|=n,\bar{A}(w)=0} R_w = R_{a^j \bowtie A^{n-j}} = 0.$$

To conclude:

$$\sum_{|w|=n,\bar{A}(w)=0} R_w f_w = \sum_{j=0}^n \sum_{a(w)=j,|w|=n,\bar{A}(w)=0} R_w f_w = 0.$$

The formal vector field is

$$\hat{f}(x) = R_A f_A(x) + R_a f_a(x) + R_{A^2} f_{A^2}(x) = \left( \mu - \frac{\sigma^2}{2} + \frac{\sigma \mathcal{B}(h)}{h} \right) x.$$

The derandomization of the SDE is

$$y'(t) = \left( \mu - \frac{\sigma^2}{2} + \frac{\sigma \mathcal{B}(h)}{h} \right) y(t).$$

The comparison of the solutions after imposing the initial condition is

$$y(t) = \exp\left(\left(\mu - \frac{\sigma^2}{2} + \frac{\sigma \mathcal{B}(h)}{h}\right)t\right), \quad x(t) = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma \mathcal{B}(t)\right).$$

# Chapter 6

## Splitting integrators for Ito SDE's

This chapter aims to provide both the strong and weak order conditions for split-step integrators of Ito SDE's as well as modified equations. The approach taken here may be seen as patterned after chapter 3.

### 6.1 Numerical solution

We have already checked that any split-step integrator for Stratonovich SDE's *admitted a word series expansion*. In addition, we presented two different approaches to obtaining such an expansion. With the help of notions such as the Ito extension of alphabets or the quasishuffle product, we are in a position to reach the same results. What is more, the two different approaches that we present in this chapter for Ito SDE's parallel those of the Stratonovich case. For notational issues, we shall approach these concepts by means of a very simple example: the initial value problem

$$dx = \underbrace{f_a(x)dt + f_A(x)d\mathcal{B}_A(t)}_{(1)} + \underbrace{f_b(x)dt}_{(2)}, \quad (6.1)$$

where we assume that the split systems (1)

$$dx = \underbrace{f_a(x)dt + f_A(x)d\mathcal{B}_A(t)}_{(1)}, \quad (6.2)$$

and (2)

$$\frac{dx}{dt} = \underbrace{f_b(x)}_{(2)} \quad (6.3)$$

are explicitly integrable.

### 6.1.1 First approach: Quasishuffle Group

#### Step 1. Words series for SDE's expansions

The alphabet associated with the SDE is  $\mathcal{A} = \{a, b, A\}$  so that its Ito extension is  $\bar{\mathcal{A}} = \{a, b, A, \bar{A}\}$ . Denote by  $\phi_{t,t_0} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  the solution mapping of the system (6.1) and set  $\phi_{t,t_0}^{(1)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\phi_{t,t_0}^{(2)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  for the solution mappings of the split systems (1) and (2). Again, for this particular example we could have written  $\phi_{t-t_0}^{(2)}$  because the equation is autonomous. We will focus on the simple splitting integrator that advances the numerical solution from  $t_0$  to  $t_0 + h$ ,  $h > 0$  by means of the mapping

$$\tilde{\phi}_{t_0+h,t_0} = \phi_{t_0+h,t_0}^{(1)} \circ \phi_{t_0+h,t_0}^{(2)}.$$

Define<sup>1</sup>

$$I_w^{(1)}(t; t_0) = \begin{cases} I_w(t; t_0) & \text{if } b(w) = 0 \\ 0 & \text{if } b(w) \neq 0 \end{cases} \quad (6.4)$$

and

$$I_w^{(2)}(t; t_0) = \begin{cases} I_w(t; t_0) & \text{if } b(w) = |w| \\ 0 & \text{if } a(w) \neq |w| \end{cases} \quad (6.5)$$

i.e.  $I^{(1)}(t; t_0)$  denotes the iterated integrals taking  $\lambda_b = 0$  and  $\lambda_l = 1$  if  $l \neq b$ , whereas  $I^{(2)}$  is obtained by setting  $\lambda_b = 1$ ,  $\lambda_l = 0$  in other case. Theorem 4.4 ensures that  $I^{(1)}(t; t_0), I^{(2)}(t; t_0) \in \mathcal{G}_{qsh}$  for every  $t \geq t_0$  and each event  $\omega$  (they are iterated integrals). We are then in a position to obtain the expansions of both solution operators by means of Theorem 5.5. On the one hand

$$\phi_{t_0+h;t_0}^{(1)}(x_0) = W_{I^{(1)}(t_0+h;t_0)}(x_0),$$

and on the other,

$$\phi_{t_0+h;t_0}^{(2)}(x_0) = W_{I^{(2)}(t_0+h;t_0)}(x_0).$$

Let us display a few terms:

$$\begin{aligned} \phi_{t_0+h;t_0}^{(1)}(x_0) &= x_0 + I_A^{(2)} f_A(x_0) + I_a^{(2)} f_a(x_0) + I_{AA}^{(2)} f_{AA}(x_0) + \\ &\quad + I_{Aa}^{(2)} f_{Aa}(x_0) + I_{aA}^{(2)} f_{aA}(x_0) + I_{\bar{A}\bar{A}}^{(2)} f_{\bar{A}\bar{A}}(x_0) + I_{\bar{A}AA}^{(2)} f_{\bar{A}AA}(x_0) \\ &\quad + \dots \\ \phi_{t_0+h;t_0}^{(2)}(x_0) &= x_0 + I_b^{(1)} f_b(x_0) + I_{bb}^{(1)} f_{bb}(x_0) + I_{bbb}^{(1)} f_{bbb}(x_0) + O(h^3). \end{aligned}$$

#### Step 2. Word series expansion of the numerical integrator

<sup>1</sup>Recall that  $|w|$  stands for the length of the word  $w$ , i.e., its total number of letters.

After invoking Proposition 5.4 we find

$$\tilde{\phi}_{t_0+h,t_0}(x_0) = W_{\tilde{I}(t_0+h;t_0)}(x_0), \quad (6.6)$$

with

$$\tilde{I}(t_0+h;t_0) = I^{(1)}(t_0+h;t_0)I^{(2)}(t_0+h;t_0) \in \mathcal{G}_{qsh}. \quad (6.7)$$

The coefficients  $\tilde{I}_w(t_0+h;t_0)$  are computed via the convolution product (1.16):

$$\left(\tilde{I}_w(t_0+h;t_0)\right)_{w \in \mathcal{W}} = \left(I_w^{(1)}(t_0+h;t_0)\right)_{w \in \mathcal{W}} \star \left(I_w^{(2)}(t_0+h;t_0)\right)_{w \in \mathcal{W}}. \quad (6.8)$$

Again, the recipe (6.8) provides a remarkably straightforward way to compute the coefficients: for words  $w = w'w''$ , where the (possibly empty) word  $w'$  does not include the letter  $b$  and the (possibly empty) word  $w''$  does not include the letters  $a$  or  $A$ , then  $\tilde{I}_w = I_{w'}I_{w''}$ ; if  $w$  is not a concatenation of that form, then the coefficient is 0, and so on. For this particular example (as usual, we use underlines to indicate the stochastic weight):

$$\begin{aligned} \tilde{I}(t_0+h;t_0) = & \underbrace{1}_0 + \underbrace{I_AA}_{0.5} + \underbrace{h(a+b+\bar{A}) + I_{A^2}A^2}_1 + \\ & + \underbrace{I_{A^3}A^3 + I_A I_b A b + I_{Aa} A a + I_{aA} a A + I_{\bar{A}\bar{A}} \bar{A}\bar{A} + I_{A\bar{A}} A \bar{A}}_{1.5} + \dots, \end{aligned}$$

whereas

$$\begin{aligned} I(t_0+h;t_0) = & \underbrace{1}_0 + \underbrace{I_AA}_{0.5} + \underbrace{h(a+b+\bar{A}) + I_{A^2}A^2}_1 + \\ & + \underbrace{I_{A^3}A^3 + I_{Ab}Ab + I_{bA}bA + I_{Aa}Aa + I_{aA}aA + I_{\bar{A}\bar{A}}\bar{A}\bar{A} + I_{A\bar{A}}A\bar{A}}_{1.5} \\ & + \dots. \end{aligned}$$

### Step 3. Word series of the local error

Set

$$\delta(t_0+h;t_0) = I^{(1)}(t_0+h;t_0)I^{(2)}(t_0+h;t_0) - I(t_0+h;t_0).$$

Then, the strong local error may be expanded as

$$\tilde{\phi}_{t_0+h,t_0}(x_0) - \phi_{t_0+h,t_0}(x_0) = W_{\delta(t_0,h)}(x_0).$$

Observe that  $(\delta, w) = 0$  for words  $\|w\| \leq 1$ , but, if  $\|w\| = 1.5$ ,  $(\delta, w) = 0$  except for the words



- $(\delta, bA) = -I_{bA}$ ,
- $(\delta, Ab) = hI_A - I_{Ab}$ .

If  $f_{bA}, f_{Ab}$  do not vanish, we will later check<sup>2</sup> that the strong local error  $\epsilon$  is such that  $\mathbb{E}|\epsilon| = O(h^{1.5})$ .

### Composition of observables

Take an observable  $\chi$ . Then, the error in computing  $\chi(x(t))$  after a single step by means of the split-step integrator is

$$\chi\left(\tilde{\phi}_{t_0+h,t_0}(x_0)\right) - \chi\left(\phi_{t_0+h,t_0}(x_0)\right),$$

which may easily be expanded as a word series with the help of (5.3). Again, *without any additional computation*,

$$\begin{aligned} \chi\left(\tilde{\phi}_{t_0+h,t_0}(x_0)\right) - \chi\left(\phi_{t_0+h,t_0}(x_0)\right) &= D_{\delta(t_0+h;t_0)}\chi(x_0) \\ &= (hI_A - I_{Ab})D_{Ab}\chi(x_0) - I_{bA}D_{bA}\chi(x_0) \\ &\quad + \dots \end{aligned}$$

### Average of observables

The value  $\mathbb{E}\chi(x(t))$  is generally of more importance than the realizations  $\chi(x(t))$  themselves. One more time, this expectation may be expanded as word series *without any additional computation*:

$$\boxed{\mathbb{E}^{x_0}\chi\left(\tilde{\phi}_{t_0+h,t_0}(x_0)\right) - \mathbb{E}^{x_0}\chi\left(\phi_{t_0+h,t_0}(x_0)\right) = D_{\mathbb{E}\delta(t_0+h;t_0)}\chi(x_0),}$$

In this case, observe that  $(\mathbb{E}\delta, w) = 0$  if  $\|w\| \leq 1.5$ , but  $\delta_{ba} = -\frac{h^2}{2}$ , hence

$$(\mathbb{E}\delta, ba) = -\frac{h^2}{2} \neq 0.$$

Under regularity assumptions, we shall check that the above condition yields a weak local error of  $O(h^2)$ .

---

<sup>2</sup>by arguing almost identically to the Stratonovich case

### 6.1.2 Second approach: Solution Operators

As in the Stratonovich case, Theorem 5.4 is not necessary to obtain the expansions of both strong and weak local errors. Assume  $\mathbb{X}_{t;t_0}$  denotes the solution operator of the system (6.1) and set  $\mathbb{X}_{t;t_0}^{(1)}, \mathbb{X}_{t;t_0}^{(2)}$  for those of subsystems (1) and (2) respectively. Let  $\tilde{\mathbb{X}}_{t;t_0} \in \mathcal{L}(\mathcal{O})$  denote the pullback of the numerical integrator. Recall that  $\tilde{\mathbb{X}}_{t;t_0}$  maps any observable  $\chi$  into the observable  $\tilde{\mathbb{X}}_{t;t_0}(\chi)(x_0) = \chi(\tilde{\phi}_{t;t_0}(x_0))$ . In our example  $\tilde{\mathbb{X}}_{t;t_0} = \mathbb{X}_{t;t_0}^{(1)} \mathbb{X}_{t;t_0}^{(2)}$ :

$$\mathbb{X}_{t;t_0}^{(1)} \mathbb{X}_{t;t_0}^{(2)} \chi(x_0) = \chi(\phi_{t;t_0}^{(2)}(\phi_{t;t_0}^{(1)}(x_0))) = \chi(\tilde{\phi}_{t;t_0}(x_0)).$$

Following Theorem 5.5:

- $\mathbb{X}_{t;t_0} = D_{I(t_0+h;t_0)}$ ,
- $\mathbb{X}_{t;t_0}^{(1)} = D_{I^{(1)}(t_0+h;t_0)}$ ,
- $\mathbb{X}_{t;t_0}^{(2)} = D_{I^{(2)}(t_0+h;t_0)}$ .

Thus:

$$\tilde{\mathbb{X}}_{t;t_0} = D_{I^{(1)}(t_0+h;t_0)} D_{I^{(2)}(t_0+h;t_0)} = D_{I^{(1)}(t_0+h;t_0) I^{(2)}(t_0+h;t_0)},$$

and

- $\chi(\tilde{\phi}_{t_0+h;t_0}(x_0)) = D_{\tilde{I}(t_0+h;t_0)} \chi(x_0)$ ,
- $\tilde{\phi}_{t_0+h;t_0}(x_0) = D_{\tilde{I}(t_0+h;t_0)}(Id)(x_0) = W_{\tilde{I}(t_0+h;t_0)}(x_0)$ ,

with  $\tilde{I}(t_0 + h; t_0) = I^{(1)}(t_0 + h; t_0) I^{(2)}(t_0 + h; t_0)$ .

## 6.2 Word series for split-step integrators theorem, Ito case

**Theorem 6.1 (Word series for split-step integrators, Ito case)** *Consider a split-step integrator  $\tilde{\phi}_{t_0+h;t_0}$  for the SDE (5.4). Then, there exists a random point  $\tilde{I}(t_0 + h; t_0) \in \mathcal{G}_{qsh}$  such that for each event  $\omega$*

$$\tilde{\phi}_{t_0+h;t_0}(x_0) = W_{\tilde{I}(t_0+h;t_0)}(x_0),$$

and for any observable  $\chi \in \mathcal{O}$

$$\chi(\tilde{\phi}_{t_0+h;t_0}(x_0)) = D_{\tilde{I}(t_0+h;t_0)} \chi(x_0), \quad \mathbb{E}^{x_0} \chi(\tilde{\phi}_{t_0+h;t_0}(x_0)) = D_{\mathbb{E}\tilde{I}(t_0+h;t_0)} \chi(x_0).$$

Moreover,  $\tilde{I}(t_0 + h; t_0)$  is explicitly computable.

### 6.2.1 Properties of $\tilde{I}(t_0 + h; t_0)$ and $\mathbb{E}\tilde{I}(t_0 + h; t_0)$

**Proposition 6.1** *The coefficients  $\tilde{I}_w(t_0 + h; t_0)$ ,  $w \in \mathcal{W}$ , associated with a splitting integrator possess the properties of the exact values  $I_w(t_0 + h; t_0)$  listed in Proposition 5.1.*

**Proof:** The arguments given in Proposition 6.1 may be easily translated into this case.  $\square$

The following result, similar to Proposition 6.2, provides insight into the structure of  $\mathbb{E}\tilde{I}(t_0 + h; t_0)$ .

**Proposition 6.2 (Structure of  $\mathbb{E}\tilde{I}(t_0 + h; t_0)$ )** *There exist numbers  $\nu_1, \nu_2, \dots, \nu_n$  such that*

$$\mathbb{E}\tilde{I}(t_0 + h; t_0) = \prod_{i=1}^n \exp(h\nu_i \mathfrak{f}_i),$$

where the  $\mathfrak{f}_i$ 's are infinitesimal generators of the subsystems involved in the splitting, i.e.

$$\mathfrak{f}_i = a_{i_1} + a_{i_2} + \dots + a_{i_j} + \bar{A}_{i_1} + \dots + \bar{A}_{i_j},$$

with  $a_l \in \mathcal{A}_{det}$ ,  $A_l \in \mathcal{A}_{sto}$ . In particular,  $\mathbb{E}\tilde{I}(t_0 + h; t_0) \in \mathcal{G}_{sh}$ , so that  $\theta^{-1}\mathbb{E}\tilde{I}(t_0 + h; t_0) \in \mathcal{G}_{qsh}$ .

**Proof:** Assume

$$\tilde{I}(t_0 + h; t_0) = \prod_{i=1}^N I^{(i)}(t_0 + c_i h; t_0 + b_i h).$$

Then

$$\mathbb{E}\tilde{I}(t_0 + h; t_0) = \prod_{i=1}^N \mathbb{E}I^{(i)}(t_0 + c_i h; t_0 + b_i h)$$

because either the iterated integrals  $I^{(i)}(t_0 + c_i h; t_0 + b_i h)$  correspond to disjoint intervals, in which case they are independent, or they correspond to the same interval so that the Brownian motions involved must be different: they are hence independent. Kolmogorov's backward equation (5.18) allows to conclude. Observe that

$$\mathfrak{f}_i = a_{i_1} + a_{i_2} + \dots + a_{i_j} + \bar{A}_{i_1} + \dots + \bar{A}_{i_j} \in \mathfrak{g}_{sh},$$

so that

$$\exp(h\nu_i \mathfrak{f}_i) \in \mathcal{G}_{sh}.$$

Then,

$$\mathbb{E}\tilde{I}(t_0 + h; t_0) = \prod_{i=1}^n \exp(h\nu_i \mathfrak{f}_i) \in \mathcal{G}_{sh}.$$

$\square$

In our example

$$dx = \underbrace{f_a(x)dt}_{(1)} + \underbrace{f_b(x)dt + f_A(x)d\mathcal{B}_A(t)}_{(2)},$$

$$\mathbb{E}\tilde{I}(t_0+h; t_0) = \mathbb{E}\left(I^{(1)}(t_0+h; t_0)I^{(2)}(t_0+h; t_0)\right) = \mathbb{E}I^{(1)}(t_0+h; t_0)\mathbb{E}I^{(2)}(t_0+h; t_0),$$

so that

$$\mathbb{E}\tilde{I}(t_0+h; t_0) = e^{ha}e^{h(b+\bar{A})} \in \mathcal{G}_{sh}.$$

□

Proposition 6.2 enables us to prove, just as Proposition 3.2 did, the last item of Proposition 6.1: given a word  $w$ ,

$$\mathbb{E}\tilde{I}(t_0+h; t_0)_w = \left(\prod_{i=1}^n \exp(h\nu_i \mathfrak{f}_i), w\right),$$

but  $\prod_{i=1}^n \exp(h\nu_i \mathfrak{f}_i)$  comprises no term with semiinteger weight.

### 6.3 Strong and weak order conditions

**Theorem 6.2** *For a splitting integrator, the local error possesses a word series expansion*

$$\tilde{\phi}_{t,t_0}(x_0) - \phi_{t,t_0}(x_0) = W_{\delta(t_0,h)}(x_0) = \sum_{n \in (1/2)\mathbb{N}} \sum_{\|w\|=n} \delta_w(t_0, h) f_w(x_0), \quad (6.9)$$

with coefficients

$$\delta_w(t_0, h) = \tilde{I}_w(t_0+h; t_0) - I_w(t_0+h; t_0), \quad w \in \mathcal{W},$$

that, in any  $L^p$  norm,  $1 \leq p < \infty$ , satisfy, uniformly in  $t_0 \geq 0$ ,

$$\|\delta_w(t_0, h)\|_p = \mathcal{O}(h^{\|w\|}), \quad h \downarrow 0.$$

In addition, for each observable  $\chi$ , conditional on  $x_0$ ,

$$\mathbb{E}\chi(\tilde{\phi}_{t,t_0}(x_0)) - \mathbb{E}\chi(\phi_{t,t_0}(x_0)) = \sum_{n \in \mathbb{N}} \sum_{\|w\|=n} (\mathbb{E}\delta_w(t_0, h)) D_w \chi(x_0), \quad (6.10)$$

The theorem implies that the *strong order conditions*

$$\tilde{I}_w(t_0+h; t_0) = I_w(t_0+h; t_0), \quad \|w\| = 0, 1/2, 1, \dots, \mu, \quad \mu \in (1/2)\mathbb{N}, \quad (6.11)$$

ensure that the series in (6.9) only comprises terms of size  $\mathcal{O}(h^{\mu+1/2})$ . If, for a given alphabet of the form  $\tilde{\mathcal{A}}$  and given coefficients  $\tilde{I}_w(t_0 + h; t_0)$ , one demands errors  $\mathcal{O}(h^{\mu+1/2})$  for *all possible choices* of the vector fields  $f_a, f_A$ , then the conditions (6.11) are not only sufficient but also *necessary*.

As in the Stratonovich case, the best way to deal with specific problems is to write down, up to the desired order, the word series expansions of the true and numerical solutions and compare them after taking into account the quasishuffle relations and the specific expressions of the word basis functions. When handling a particular problem, there is typically a considerable saving of time (with respect to the general and abstract case), because many word basis functions generally vanish.

## 6.4 Error bounds

Assume that every vector field  $f_l$  in (5.4) is globally Lipschitz. This guarantees both existence and uniqueness of the initial value problem for (5.4) itself and for the split systems. The following theorems, which are of importance when measuring the local errors by means of word series procedures, also hold in the Ito case. Recall that, when proving Theorem 3.3, we rewrote iterated Stratonovich integrals in terms of Ito iterated integrals because of inequalities like (2.17) in [36] are formulated in terms of iterated Ito integrals. In order to prove the error bound theorems, the ideas behind the Stratonovich case also hold, making the proof even shorter due to the preceding remarks.

**Theorem 6.3 (Weak bounds: Ito case)** *Let  $\nu$  be a positive integer. Assume that:*

- *The deterministic vector fields  $f_a, a \in \mathcal{A}_{det}$ , are of class  $C^{2\nu}$ , while the stochastic vector fields  $f_A, A \in \mathcal{A}_{sto}$ , are of class  $C^{2\nu+1}$ .*
- *The observable  $\chi$  is of class  $C^{2\nu+2}$  in  $\mathbb{R}^d$ .*
- *There is a constant  $C > 0$  such that for each  $x \in \mathbb{R}^d$  and each word  $w$  of weight  $\nu + 1$ :*

$$|D_w \chi(x)| \leq C(1 + |x|^2)^{1/2}.$$

- *The weak error conditions (3.12) hold.*

*Then there exists a constant  $K > 0$  such that for each  $x_0$ , each  $t_0 \geq 0$  and each  $h > 0$ :*

$$|\mathbb{E}\chi(\tilde{\phi}_{t_0+h,t_0}(x_0)) - \mathbb{E}\chi(\phi_{t_0+h,t_0}(x_0))| \leq K(1 + |x_0|^2)^{1/2}h^{\nu+1}$$

*(the expectation is conditional on  $x_0$ ).*

The mean square error may be bounded as follows:

**Theorem 6.4 (Strong bounds: Ito case)** *Let  $\mu$  be a positive integer multiple of  $1/2$ . Assume that:*

- *The deterministic vector fields  $f_a$ ,  $a \in \mathcal{A}_{det}$ , are of class  $C^{2\mu}$ , while the stochastic vector fields  $f_A$ ,  $A \in \mathcal{A}_{sto}$ , are of class  $C^{2\mu+1}$ .*
- *There is a constant  $C > 0$  such that for each  $x \in \mathbb{R}^d$  and each word  $w$  of weight  $\mu + 1$ :*

$$|f_w(x)| \leq C(1 + |x|^2)^{1/2}.$$

- *The strong error conditions (3.11) hold.*

*Then there exists a constant  $K > 0$  such that for each  $x_0$ , each  $t_0 \geq 0$  and each  $h > 0$ :*

$$\left( \mathbb{E} |(\tilde{\phi}_{t_0+h,t_0}(x_0) - \phi_{t_0+h,t_0}(x_0))|^2 \right)^{1/2} \leq K(1 + |x_0|^2)^{1/2} h^{\mu+1/2}$$

*(the expectation is conditional on  $x_0$ ).*

## 6.5 Modified equations for Ito split-step integrators

The theorems of Section 1.3 does not apply to the Ito calculus because of the *Ito correction*. For the same reason, the natural extensions of those theorems to the quasishuffle case also lack useful applications so that if we want to obtain modified equations, we must find an alternative procedure. The solution is to use the Stochastic Duality, see Section 4.5, to turn the Ito SDE into a Stratonovich one and apply then the results of Chapter 3. Thus, the peculiarities we encounter when renormalizing in the Ito case are those of the Stratonovich. For example, it is worth remarking the fact that these modified equations *are not adapted* with respect to the natural filtration of the Brownian motions  $(\mathcal{F}_t)_{t \geq t_0}$ . The modified vector field  $\hat{f}$  will in particular be  $\mathcal{F}_{t_0+h}$ -measurable. Observe that in the particular case of Derandomization, there is no point in resorting to the Stochastic Duality: stochastic calculus is not involved in the renormalization.

### 6.5.1 Derandomization

**Theorem 6.5 (Derandomization)** *For any split-step integrator for the Ito SDE (5.4), there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the random ordinary differential equation*

$$y'(t) = \hat{f}(y(t))$$

with  $y(t_0) = x_0$  satisfies

$$\boxed{y(t_0 + h) = \tilde{\phi}_{t_0+h;t_0}(x_0)}.$$

In particular

$$R = \frac{1}{h} \log(\tilde{I}(t_0 + h; t_0)).$$

**Proof:** We set  $X := \tilde{I}(t_0 + h; t_0)$  in Theorem 4.11.  $\square$

## 6.5.2 Drift-modified SDE's

**Theorem 6.6 (Drift-modified SDE's)** *For any split-step integrator for the Ito SDE (5.4), there exists a random formal vector field  $\hat{f}(x) = W_R(x)$  such that the stochastic Ito differential equation*

$$dy(t) = \hat{f}(y(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(y(t))d\mathcal{B}_A(t)$$

with  $y(t_0) = x_0$  is such that

$$\boxed{y(t_0 + h) = \tilde{\phi}_{t_0+h;t_0}(x_0)}.$$

In particular  $R$  is explicitly computable.

**Proof:** Set  $X := \tilde{I}(t_0 + h; t_0) \in \mathcal{G}_{qsh}$  and rewrite it in terms of letters in  $\mathcal{A}^*$ ; the stochastic duality Theorem 4.17 ensures that  $X \in \mathcal{G}_{sh}(\mathcal{A}^*)$ . We have translated the Ito/quasishuffle problem into a Stratonovich/shuffle one. Apply now Theorem 1.10 in order to prove the existence of a certain (and explicitly computable)  $R \in \mathfrak{g}_{sh}(\mathcal{A}^*)$  such that, when setting  $\hat{f}(x) = W_R(x)$ , the formal stochastic Stratonovich equation

$$dy(t) = \hat{f}(y(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(y(t)) \circ d\mathcal{B}_A(t) + \sum_{A \in \mathcal{A}_{sto}} f_{A^*}(y(t))dt \quad (6.12)$$

is such that  $y(t_0 + h) = x(t_0 + h)$  when  $y(t_0) = x_0$ , i.e. it is a *modified equation* for the splitting integrator. Now, by Theorem 5.6,  $t \mapsto y(t)$  also satisfies the stochastic Ito differential equation

$$dy(t) = \hat{f}(y(t))dt + \sum_{A \in \mathcal{A}_{sto}} f_A(y(t))d\mathcal{B}_A(t). \quad (6.13)$$

Observe that the stochastic duality guarantees that, when expanding  $R$  as an element of  $\mathbb{R}\langle\langle \bar{\mathcal{A}} \rangle\rangle$ ,  $R \in \mathfrak{g}_{qsh}$  so that  $D_R$  is derivation.  $\square$

## **Part III**

# **Application of word series to Langevin Dynamics**





# Chapter 7

## Langevin dynamics and some split-step integrators

### 7.1 Langevin dynamics

In this chapter we shall be concerned with the numerical solution of the so-called Langevin equations (Paul Langevin, 1908):

$$\begin{aligned}dq &= M^{-1}p dt, \\ dp &= F(q) dt - \gamma p dt + \sigma M^{1/2} d\mathcal{B}(t).\end{aligned}\tag{7.1}$$

In this case,  $M$  is the  $d \times d$  diagonal mass matrix with constant diagonal entries  $m_i > 0$ ,  $\gamma > 0$  is the friction coefficient,  $\sigma$  governs the fluctuation due to noise,  $\mathcal{B}$  is a  $d$ -dimensional Brownian motion, and the force  $F$  originates from a potential. Since the noise is additive there is no distinction between the Stratonovich and Ito interpretations. Unless we choose the force to be linear, there is no analytic solution for (7.1), so that we must resort to numerical integration. Langevin dynamics arise when modelling the dynamics of molecular systems. The underlying idea of using Langevin dynamics is *to account for the overwhelming number of degrees of freedom by means of stochastic fluctuations*. Jostling of surrounding molecules provokes changes in momentum. The extra fluctuation of energy is modelled both by the random term  $\sigma M^{1/2} d\mathcal{B}$  and by a friction term,  $-\gamma p$ . Further details may be found in the [[45], Chapter 5] or [[34], Chapter 6].

### 7.2 Split-step algorithms

We present some splitting algorithms for (7.1). The so-called ABOBA and BAOAB schemes, terms coined in [32], are among the most popular integrators for (7.1).

To the contrary, both BUB and UBU<sup>1</sup> (whose names have been coined in the present work), to the best of my knowledge, have not been used so far, despite the fact they have strong order 2 (both ABOBA and BAOAB have strong order 1).

### 7.2.1 ABOBA

Following [32], [34], we split Langevin equations into three pieces as follows:

$$\begin{pmatrix} dq \\ dp \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}p dt \\ 0 \end{pmatrix}}_A + \underbrace{\begin{pmatrix} 0 \\ F(q) dt \end{pmatrix}}_B + \underbrace{\begin{pmatrix} 0 \\ -\gamma p dt + \sigma M^{\frac{1}{2}} d\mathcal{B} \end{pmatrix}}_O.$$

Each split system may be integrated explicitly. The  $O$  split system

$$\begin{aligned} dq &= 0, \\ dp &= -\gamma p dt + \sigma M^{1/2} d\mathcal{B}(t), \end{aligned}$$

leads to an Ornstein-Uhlenbeck process for  $p$ :

$$\begin{aligned} q(t) &= q(t_0), \\ p(t) &= p(t_0)e^{-\gamma(t-t_0)} + \sigma M^{1/2} \int_{t_0}^t e^{-\gamma(t-s)} d\mathcal{B}(s). \end{aligned}$$

The one-time-step formulas for the updates read

$$\begin{aligned} q &\mapsto q, \\ p &\mapsto pe^{-\gamma(t-t_0)} + \sigma M^{1/2} \sqrt{\frac{1 - e^{-2\gamma(t-t_0)}}{2\gamma}} Z, \end{aligned}$$

where  $Z$  is a  $d$ -dimensional standard normal random vector.

With a terminology common in molecular dynamics, the solution of the split system associated with  $A$  is a ‘drift’ in position,  $q \mapsto q + (t - t_0)M^{-1}p$  ( $p$  remains constant). The solution of the split system corresponding to  $B$  is a ‘kick’ in momentum  $p \mapsto p + (t - t_0)F(q)$  ( $q$  remains constant). The symbol  $\phi^l$ ,  $l = A, B, O$ , denotes the solution operator of the split system  $l$ . For instance  $\phi_{t;t_0}^A(q_0, p_0) = (q(t), p(t))$  where  $(q(t), p(t))$  is the value at time  $t$  of the solution of the split system  $A$  with initial condition  $(q(t_0), p(t_0)) = (q_0, p_0)$ . The palindromic algorithm defined by

$$\tilde{\phi}_{t_0+h;t_0}^{ABOBA} = \phi_{t_0+h;t_0+h/2}^A \circ \phi_{t_0+h;t_0+h/2}^B \circ \phi_{t_0+h;t_0}^O \circ \phi_{t_0+h/2;t_0}^B \circ \phi_{t_0+h/2;t_0}^A. \quad (7.2)$$

<sup>1</sup>All of these four integrators use one force evaluation per time step.

is denoted by ABOBA, see [32].

Set  $\delta = e^{-\gamma h}$ . The  $d$  dimensional random vectors  $Z_1, Z_2, \dots$  are (standard) normally distributed and independent. The step  $n \mapsto n + 1$  of ABOBA is carried out by means of the formulas:

- Given  $q_n, p_n \in \mathbb{R}^d$ :

$$\begin{aligned} q_{n+1/2} &= q_n + \frac{h}{2} M^{-1} p_n, \\ F_{n+1/2} &= F(q_{n+1/2}), \\ p_{n+1/3} &= p_n + \frac{h}{2} F_{n+1/2}, \\ p_{n+2/3} &= p_{n+1/3} \delta + \sigma M^{1/2} \sqrt{\frac{1 - \delta^2}{2\gamma}} Z_{n+1}, \\ p_{n+1} &= p_{n+2/3} + \frac{h}{2} F_n, \\ q_{n+1} &= q_{n+1/2} + \frac{h}{2} M^{-1} p_{n+1}. \end{aligned}$$

### 7.2.2 BAOAB

The algorithm BAOAB is defined by

$$\tilde{\phi}^{BAOAB} = \phi_{t_0+h; t_0+h/2}^B \circ \phi_{t_0+h; t_0+h/2}^A \circ \phi_{t_0+h; t_0}^O \circ \phi_{t_0+h/2; t_0}^A \circ \phi_{t_0+h/2; t_0}^B. \quad (7.3)$$

Given  $q_n, p_n, F_n$ , the formulas for the step  $n \mapsto n + 1$  are

$$\begin{aligned} p_{n+1/3} &= p_n + \frac{h}{2} F_n, \\ q_{n+1/2} &= q_n + \frac{h}{2} M^{-1} p_{n+1/3}, \\ p_{n+2/3} &= p_{n+1/3} \delta + \sigma M^{1/2} \sqrt{\frac{1 - \delta^2}{2\gamma}} Z_{n+1}, \\ q_{n+1} &= q_{n+1/2} + \frac{h}{2} M^{-1} p_{n+2/3}, \\ \text{Set } F_{n+1} &= F(q_{n+1}), \\ p_{n+1} &= p_{n+2/3} + \frac{h}{2} F_{n+1}. \end{aligned}$$

Note that BAOAB uses one force evaluation per time step with one extra evaluation at the initialization.

### 7.2.3 BUB

The following splitting first appeared in [1], although it was conceived to cater for the need of an efficient sampling algorithm for the Boltzmann distribution.<sup>2</sup> No reference to its use in the strong sampling of (7.1) has been found. In order to maintain similarities with ABOBA/BAOAB we decided to use the acronyms B and U for its split sections:

$$\begin{pmatrix} dq \\ dp \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}pdt \\ -\gamma pdt + \sigma M^{1/2}d\mathcal{B} \end{pmatrix}}_U + \underbrace{\begin{pmatrix} 0 \\ F(q)dt \end{pmatrix}}_B.$$

The system corresponding to  $U$  can also be solved explicitly:

$$\begin{aligned} q(t) &= q(t_0) + \frac{1 - e^{-\gamma(t-t_0)}}{\gamma} M^{-1}p_0 \\ &\quad + \frac{\sigma M^{-1/2}}{\gamma} \left( \mathcal{B}(t_0 + h) - \mathcal{B}(t_0) - \int_{t_0}^t e^{-\gamma(t-s)} d\mathcal{B}(s) \right), \\ p(t) &= p(t_0)e^{-\gamma(t-t_0)} + \sigma M^{1/2} \int_{t_0}^t e^{-\gamma(t-s)} d\mathcal{B}(s), \end{aligned}$$

and over the interval  $[t_0, t_0 + h]$  it may be simulated by

$$\begin{aligned} q &\mapsto q + \frac{1 - e^{-\gamma h}}{\gamma} M^{-1}p + \frac{\sigma M^{-1/2}}{\gamma} \left( X^{(1)} - X^{(2)} \right), \\ p &\mapsto pe^{-\gamma h} + \sigma M^{1/2} X^{(2)}, \end{aligned}$$

where  $X^{(1)}$  and  $X^{(2)}$  stand for  $d$  dimensional, normally distributed, random vectors

$$\begin{aligned} X^{(1)} &= \sqrt{h}Z^{(1)}, \\ X^{(2)} &= \sigma^{(2)}\rho Z^{(1)} + \sigma^{(2)}\sqrt{1 - \rho^2}Z^{(2)}, \end{aligned}$$

with  $Z^{(1)}, Z^{(2)}$  are  $d$  dimensional independent standard normally distributed random vectors and

$$\begin{aligned} \delta &= e^{-\gamma h}, \\ \rho &= \sqrt{\frac{1 - \delta}{1 + \delta} \cdot \frac{2}{\gamma h}}, \\ \sigma^{(2)} &= \sqrt{\frac{1}{2\gamma}(1 - \delta^2)}. \end{aligned}$$

---

<sup>2</sup>which is the invariant distribution of (7.1), see [34].

Note that  $\rho$  stands for the correlation among  $X_i^{(1)}$  and  $X_i^{(2)}$  and  $\sigma^{(2)} = \text{Var}(X_i^{(2)})$ . In such a way, the distribution of

$$\left( B_1(t_0 + h) - B_1(t_0), \dots, B_d(t_0 + h) - B_d(t_0), \int_{t_0}^{t_0+h} e^{-\gamma(t_0+h-s)} d\mathcal{B}_1(s), \dots, \int_{t_0}^{t_0+h} e^{-\gamma(t_0+h-s)} d\mathcal{B}_d(s) \right)$$

is that of  $(X^{(1)T}, X^{(2)T})$ .

The scheme BUB is defined as

$$\tilde{\phi}_{t_0+h;t_0}^{BUB} = \phi_{t_0+h;t_0+h/2}^B \circ \phi_{t_0+h;t_0}^U \circ \phi_{t_0+h/2;t_0}^B, \quad (7.4)$$

see [1]. In order to implement it, a sequence  $Z_1^{(1)}, Z_1^{(2)}, \dots, Z_n^{(1)}, Z_n^{(2)}, \dots$  of  $d$ -dimensional independent, normally distributed random vectors must be generated. Then, given  $q_n, p_n, F_n$ , the formulas for the  $n \mapsto n+1$  step are:

• Draw  $Z_{n+1}^{(1)}, Z_{n+1}^{(2)}$  standard, normally distributed. Compute then  $X_{n+1}^{(1)}, X_{n+1}^{(2)}$  by

$$\begin{aligned} X_{n+1}^{(1)} &= \sqrt{h} Z_{n+1}^{(1)}, \\ X_{n+1}^{(2)} &= \sigma^{(2)} \rho Z_{n+1}^{(1)} + \sigma^{(2)} \sqrt{1 - \rho^2} Z_{n+1}^{(2)}, \end{aligned}$$

and set

$$\begin{aligned} p_{n+1/3} &= p_n + \frac{h}{2} F_n, \\ q_{n+1} &= q_n + \frac{1 - \delta}{\gamma} M^{-1} p_{n+1/3} + \frac{\sigma}{\gamma} M^{-1/2} (X_{n+1}^{(1)} - X_{n+1}^{(2)}), \\ p_{n+2/3} &= p_{n+1/3} \delta + \sigma M^{1/2} X_{n+1}^{(2)}, \\ \text{Set } F_{n+1} &= F(q_{n+1}), \\ p_{n+1} &= p_{n+2/3} + \frac{h}{2} F_{n+1}. \end{aligned}$$

As BAOAB, BUB uses one force evaluation per time step with one additional force at the initialization.

### 7.2.4 UBU

The scheme UBU is defined as

$$\tilde{\phi}_{t_0+h;t_0}^{UBU} = \phi_{t_0+h;t_0+h/2}^U \circ \phi_{t_0+h;t_0}^B \circ \phi_{t_0+h/2;t_0}^U. \quad (7.5)$$

Set

$$\begin{aligned}\tilde{\delta} &= e^{-\gamma h/2}, \\ \tilde{\rho} &= \sqrt{\frac{1-\tilde{\delta}}{1+\tilde{\delta}} \cdot \frac{4}{\gamma h}}, \\ \tilde{\sigma}^{(2)} &= \sqrt{\frac{1}{2\gamma}(1-\tilde{\delta}^2)}.\end{aligned}$$

In order to simulate UBU we must generate a sequence  $Z_1^{(1)}, Z_1^{(2)}, W_1^{(1)}, W_1^{(2)}, \dots, Z_n^{(1)}, Z_n^{(2)}, W_n^{(1)}, W_n^{(2)}, \dots$  of  $d$ -dimensional independent, normally distributed random vectors.

Given  $q_n, p_n$ , the way to simulate the  $n \mapsto n+1$  iteration step is:

- Compute  $X_{n+1}^{(1)}, X_{n+1}^{(2)}, Y_{n+1}^{(1)}, Y_{n+1}^{(2)}$

$$\begin{aligned}X_{n+1}^{(1)} &= \sqrt{h/2}Z_{n+1}^{(1)}, \\ X_{n+1}^{(2)} &= \tilde{\sigma}^{(2)}\tilde{\rho}Z_{n+1}^{(1)} + \tilde{\sigma}^{(2)}\sqrt{1-\tilde{\rho}^2}Z_{n+1}^{(2)}, \\ Y_{n+1}^{(1)} &= \sqrt{h/2}W_{n+1}^{(1)}, \\ Y_{n+1}^{(2)} &= \tilde{\sigma}^{(2)}\tilde{\rho}W_{n+1}^{(1)} + \tilde{\sigma}^{(2)}\sqrt{1-\tilde{\rho}^2}W_{n+1}^{(2)},\end{aligned}$$

and

$$\begin{aligned}q_{n+1/2} &= q_n + \frac{1-\tilde{\delta}}{\gamma}M^{-1}p_n + \frac{\sigma}{\gamma}M^{-1/2}\left(X_n^{(1)} - X_n^{(2)}\right), \\ p_{n+1/3} &= p_n\tilde{\delta} + \sigma M^{1/2}X_n^{(2)}, \\ p_{n+2/3} &= p_{n+1/3} + hF(q_{n+1/2}), \\ q_{n+1} &= q_{n+1/2} + \frac{1-\tilde{\delta}}{\gamma}M^{-1}p_{n+2/3} + \frac{\sigma}{\gamma}M^{-1/2}\left(W_n^{(1)} - W_n^{(2)}\right), \\ p_{n+1} &= p_{n+2/3}\tilde{\delta} + \sigma M^{1/2}W_n^{(2)}.\end{aligned}$$

Note that the implementation of UBU can be reformulated in order to cut down by half the number of random variables generated at each step: perform blocks  $BU$  with an  $h$  time size.

### 7.3 Letters for the Langevin Dynamics

The additive stochastic noise present in Langevin Dynamics makes both Stratonovich and Ito interpretations coincide so that we are free to choose any of them.

To simplify the analysis there is an obvious election. When considering the Stratonovich interpretation we only have to take care of purely deterministic and stochastic letters. In turn, for the Ito interpretation, we must also take into account the extra  $d$  letters of the form  $\bar{A}$ , which *a priori* increases the complexity of the required algebraic calculations. We shall deal with this point in Section 7.5. After setting  $x = (q, p) \in \mathbb{R}^d \times \mathbb{R}^d$ , we find that the equations are the particular instance of (2.3) given by

$$dx(t) = f_a(x)dt + f_b(x)dt + f_c(x)dt + \sum_{i=1}^d f_{A_i}(x) \circ dB_i(t) \quad (7.6)$$

with

$$f_a(q, p) = (M^{-1}p, 0), \quad f_b(q, p) = (0, F(q)), \quad f_c(q, p) = (0, -\gamma p),$$

and, for  $i = 1, \dots, d$ ,

$$f_{A_i}(q, p) = (0, \sigma \sqrt{m_i} e_i),$$

where  $e_i$  is the  $i$ -th unit vector in  $\mathbb{R}^d$ . The deterministic letters  $a$ ,  $b$  and  $c$  are respectively associated with inertia, potential forces and friction; as it will become apparent below the word basis functions  $f_w$ ,  $w \in \mathcal{W}$  also have clear physical meaning.

### 7.3.1 Word basis functions

The structure of the Langevin equations implies that many word basis functions are identically zero. The vector fields  $f_a$ ,  $f_b$ ,  $f_c$ , and  $f_{A_i}$  have many null components and additional simplifications are due to  $f_{A_i}$  being constant,  $f_a$  and  $f_c$  being linear in  $p$  and independent of  $q$ , and  $f_b$  being independent of  $p$ . In particular, the relation  $f_{ba}(q, p) = (M^{-1}F(q), 0)$  shows that  $f_{ba}$  is a function of  $q$  alone and, since the  $q$  components of  $f_c$  and  $f_{A_i}$  vanish, we have, in view of the word basis recursions of Subsection 5.2.2

$$f_{cba}(q, p) = 0, \quad f_{A_i ba}(q, p) = 0, \quad i = 1, \dots, d, \quad (7.7)$$

for each  $q$  and  $p$ . Physically, (7.7) means that the value  $M^{-1}F(q)$  of the acceleration created by the potential forces would not be affected if noise or friction changed instantaneously the momentum of the system. On the other hand, in general,

$$f_{cab}(q, p) \neq 0, \quad f_{A_i ab}(q, p) \neq 0, \quad i = 1, \dots, d. \quad (7.8)$$

The second block of  $f_{ab}(q, p) = (0, \partial_q F(q) M^{-1}p) = (0, (d/dt)F(q))$  is the contribution to  $(d^2/dt^2)p$  that arises from the potential forces. This contribution is a



$w$	<i>constant</i>	<i>linear</i>	$(\star, 0)$	$(0, \star)$
$A_i a$	✓		✓	
$A_i c$	✓			✓
$ab$				✓
$ba$			✓	
$bc$				✓
$ca$		✓	✓	
$cc$		✓		✓
$A_i ab$				✓
$A_i ca$	✓		✓	
$A_i cc$	✓			✓

Table 7.1: Non-zero word basis functions with stochastic weight  $< 3$ . A checkmark on the column titled constant (linear) means that the corresponding word basis is constant (linear). The symbol  $(\star, 0)$  denotes a mapping  $f : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$  of the form  $f(q, p) = (\hat{f}(q, p), 0)$  with  $\hat{f}(q, p) \in \mathbb{R}^d$ . Similarly,  $(0, \star)$ . Observe that every word basis function is either of the form  $(\star, 0)$  or  $(0, \star)$ .

function of  $q$  and  $p$  and its value would be affected if friction or noise changed instantaneously the momentum. It is also useful to note at this point that, according to the recurrences of Subsection 5.2.2, if  $f_w$  vanishes identically, then the same is true for all words of the form  $w'w$ , i.e. for all words that have  $w$  as a suffix. Table 7.3.1 lists the words  $w$  with weight  $< 3$  and nonvanishing basis function, as well as some additional relevant information.

For completeness, we include a table with the words such that  $f_w \neq 0$  and  $\|w\| = 3$ , see Table 7.3.1.

### 7.3.2 Unidimensional calculations

For the particular case in which  $d = 1$  we are going to provide explicit calculations of the non-vanishing word basis functions. Note that for this case there only exists one stochastic letter.

#### Weight 0

$$f_1(q, p) = \begin{pmatrix} q \\ p \end{pmatrix}.$$

$w$	<i>constant</i>	<i>linear</i>	$(\star, 0)$	$(0, \star)$
$aba$			✓	
$bab$				✓
$aab$				✓
$cab$				✓
$abc$				✓
$cca$		✓	✓	
$bca$			✓	
$bcc$				✓
$ccc$		✓		✓

Table 7.2: Words with stochastic weight  $\|w\| = 3$  such that  $f_w \neq 0$ .**Weight 0.5**

$$f_A(q, p) = \begin{pmatrix} 0 \\ \sigma\sqrt{m} \end{pmatrix}.$$

**Weight 1**

$$f_a(q, p) = \begin{pmatrix} p/m \\ 0 \end{pmatrix}, \quad f_b(q, p) = \begin{pmatrix} 0 \\ F(q) \end{pmatrix}, \quad f_c(q, p) = \begin{pmatrix} 0 \\ -\gamma p \end{pmatrix}.$$

**Weight 1.5**

$$f_{Aa}(q, p) = \begin{pmatrix} \sigma/\sqrt{m} \\ 0 \end{pmatrix}, \quad f_{Ac}(q, p) = \begin{pmatrix} 0 \\ -\gamma\sigma\sqrt{m} \end{pmatrix}.$$

**Weight 2**

$$f_{ba}(q, p) = \begin{pmatrix} F(q)/m \\ 0 \end{pmatrix}, \quad f_{ab}(q, p) = \begin{pmatrix} 0 \\ F'(q)p/m \end{pmatrix}, \quad f_{bc}(q, p) = \begin{pmatrix} 0 \\ -F(q)\gamma \end{pmatrix},$$

$$f_{ca}(q, p) = \begin{pmatrix} -\gamma p/m \\ 0 \end{pmatrix}, \quad f_{cc}(q, p) = \begin{pmatrix} 0 \\ \gamma^2 p \end{pmatrix}.$$

**Weight 2.5**

$$f_{Aab}(q, p) = \begin{pmatrix} 0 \\ F'(q)\sigma/\sqrt{m} \end{pmatrix}, \quad f_{Aca}(q, p) = \begin{pmatrix} -\gamma\sigma/\sqrt{m} \\ 0 \end{pmatrix}, \quad f_{Acc}(q, p) = \begin{pmatrix} 0 \\ \gamma^2\sigma\sqrt{m} \end{pmatrix}.$$

**Weight 3**

$$\begin{aligned}
f_{aba}(q, p) &= \begin{pmatrix} \frac{F'(q)p}{m^2} \\ 0 \end{pmatrix}, & f_{aab}(q, p) &= \begin{pmatrix} 0 \\ \frac{F''(q)p^2}{m^2} \end{pmatrix}, & f_{ccc}(q, p) &= \begin{pmatrix} 0 \\ -\gamma^3 p \end{pmatrix}, \\
f_{bca}(q, p) &= \begin{pmatrix} -\frac{\gamma F(q)}{m} \\ 0 \end{pmatrix}, & f_{cab}(q, p) &= \begin{pmatrix} 0 \\ -\frac{\gamma F'(q)p}{m} \end{pmatrix}, & f_{abc}(q, p) &= \begin{pmatrix} 0 \\ -\frac{\gamma F'(q)p}{m} \end{pmatrix}, \\
f_{bab}(q, p) &= \begin{pmatrix} 0 \\ \frac{F(q)F'(q)}{m} \end{pmatrix}, & f_{cca}(q, p) &= \begin{pmatrix} \frac{\gamma^2 p}{m} \\ 0 \end{pmatrix}, & f_{bcc}(q, p) &= \begin{pmatrix} 0 \\ F(q)\gamma^2 \end{pmatrix},
\end{aligned}$$

**7.4 Word series expansions and comparisons****7.4.1 Word series expansion**

The preparations above enable us to employ the algebraic framework we have developed in previous chapters to analyze the performance of these integrators. We already checked that any split-step integrator for (7.1) may be expanded in terms of word series operators. Each numerical one-step approximation has a word series expansion associated with a random group point  $\tilde{J}(t_0 + h; t_0) \in \mathcal{G}_{sh}$  and the true solution is readily identified with it a random curve  $h \mapsto J(t_0 + h; t_0)$ . Examination of the formal series  $J(t_0 + h; t_0) - \tilde{J}(t_0 + h; t_0)$  and its expectation yields strong and weak convergence orders of split-step integrators.

Taking into account the simplifications provided in Table 7.3.1, after setting  $x := (q, p)$ , we may expand the true solution as

$$\begin{aligned}
x(t_0 + h) &= W_{J(t_0+h;t_0)}(x_0) = \sum_w J_w(t_0 + h; t_0) f_w(x_0) \\
&= x_0 + \sum_{i=1}^d J_{A_i}(t_0 + h; t_0) f_{A_i}(x_0) + h(f_a(x_0) + f_b(x_0) + f_c(x_0)) \\
&\quad + \sum_{i=1}^d J_{A_i a}(t_0 + h; t_0) f_{A_i a}(x_0) + J_{A_i c}(t_0 + h; t_0) f_{A_i c}(x_0) \\
&\quad + \frac{h^2}{2}(f_{ab}(x_0) + f_{ba}(x_0) + f_{bc}(x_0) + f_{ca}(x_0) + f_{cc}(x_0)) \\
&\quad + \sum_{i=1}^d \left( J_{A_i ba}(t_0 + h; t_0) f_{A_i ba}(x_0) + J_{A_i ac}(t_0 + h; t_0) f_{A_i ac}(x_0) \right. \\
&\quad \left. + J_{A_i cc}(t_0 + h; t_0) f_{A_i cc}(x_0) \right) + \mathcal{R}_{true},
\end{aligned}$$

where the remainder  $\mathcal{R}_{true}$  is a  $2d$ -dimensional random vector fulfilling

$$\mathbb{E}\|\mathcal{R}_{true}\|_2 = O(h^3), \quad (7.9)$$

and, of course,

- $J_{A_i}(t_0 + h; t_0) = \mathcal{B}_i(t_0 + h) - \mathcal{B}_i(t_0)$ ,
- $J_{A_i a}(t_0 + h; t_0) = J_{A_i c}(t_0 + h; t_0) = \int_{t_0}^{t_0+h} \mathcal{B}_i(t_0 + s) - \mathcal{B}_i(t_0) ds$ ,
- $J_{A_i ab}(t_0 + h; t_0) = J_{A_i ca}(t_0 + h; t_0) = J_{A_i cc}(t_0 + h; t_0) = \int_{t_0}^{t_0+h} J_{A_i a}(s; t_0) ds$ .

By virtue of Theorem 3.2, for any palindromic split-step integrator<sup>3</sup>  $x^1$ , there exist random variables  $\tilde{J}_{A_i a}$ ,  $\tilde{J}_{A_i c}$ ,  $\tilde{J}_{A_i ab}$ ,  $\tilde{J}_{A_i ca}$ ,  $\tilde{J}_{A_i cc}$  and a remainder  $\tilde{\mathcal{R}}$  satisfying property (7.9) (all of them depending on the method itself) such that

$$\begin{aligned} x^1 &= W_{\tilde{J}(t_0+h;t_0)}(x_0) = \sum_w \tilde{J}_w(t_0 + h; t_0) f_w(x_0) \\ &= x_0 + \sum_{i=1}^d J_{A_i}(t_0 + h; t_0) f_{A_i}(x_0) + h(f_a(x_0) + f_b(x_0) + f_c(x_0)) \\ &\quad + \sum_{i=1}^d \tilde{J}_{A_i a} f_{A_i a}(x_0) + \tilde{J}_{A_i c} f_{A_i c}(x_0) \\ &\quad + \frac{h^2}{2} (f_{ab}(x_0) + f_{ba}(x_0) + f_{bc}(x_0) + f_{ca}(x_0) + f_{cc}(x_0)) \\ &\quad + \sum_{i=1}^d \tilde{J}_{A_i ba} f_{A_i ba}(x_0) + \tilde{J}_{A_i ac} f_{A_i ac}(x_0) + \tilde{J}_{A_i cc} f_{A_i cc}(x_0) \\ &\quad + \tilde{\mathcal{R}}. \end{aligned}$$

The random variable  $\tilde{J}_w(t_0 + h; t_0)$  is easily computed by techniques developed in Chapter 3. The series  $\tilde{J}(t_0 + h; t_0) \in \mathcal{G}_{sh}$  is nothing but a product of as many iterated integrals of the form  $J^{(i)}(t_0 + h; t_0)$  as split systems are gathered in the numerical integrator. These iterated integrals are calculated with families of weights such that  $\lambda_l(t) \in \{0, 1\}$  for every letter. Recall this *Boolean choice* is made as follows

$$\lambda_l(t) = \begin{cases} 1 & \text{if } l \text{ belongs to the split system} \\ 0 & \text{if } l \text{ does not belong to the split system} \end{cases}$$

---

<sup>3</sup>Because of its palindromic structure,  $\tilde{J}_{A_i}(t_0 + h; t_0) = J_{A_i}(t_0 + h; t_0)$ ,  $\tilde{J}_a(t_0 + h; t_0) = \tilde{J}_b(t_0 + h; t_0) = \tilde{J}_c(t_0 + h; t_0) = h$  and  $\tilde{J}_{xy}(t_0 + h; t_0) = h^2/2$  for  $x, y \in \{a, b, c\}$ . This case obviously includes ABOBA, BAOAB, BUB, and UBU integrators.

$w$	$\tilde{J}_w^{ABOBA}$	$\tilde{J}_w^{BAOAB}$	$\tilde{J}_w^{BUB}$	$\tilde{J}_w^{UBU}$
$A_i a$	$J_{A_i} \frac{h}{2}$	$J_{A_i} \frac{h}{2}$	✓	✓
$A_i c$	✓	✓	✓	✓
$A_i ab$	0	$J_{A_i} \frac{h^2}{4}$	$J_{A_i a} \frac{h}{2}$	$J_{A_i a}(t_0 + \frac{h}{2}; t_0)h$
$A_i ca$	$J_{A_i c} \frac{h}{2}$	$J_{A_i c} \frac{h}{2}$	✓	✓
$A_i cc$	✓	✓	✓	✓

Table 7.3: Coefficients of the word series expansion for methods ABOBA, BAOBA, BUB and UBU. A check mark signals agreement with the exact  $J_w$ . Unless otherwise is indicated, the iterated stochastic integrals have domain  $(t_0 + h; t_0)$ .

We refer to Chapter 3. In any case, let us explain in detail the calculations in, for example, the ABOBA case. From the definition of ABOBA, using the convolution product (1.16), it is clear that, if  $w$  is not of the form  $a^k b^\ell w' b^m a^n$ , with  $k, \ell, m, n$  nonnegative integers and  $w'$  a word not including the letters  $a$  or  $b$ , then  $\tilde{J}_w^{ABOBA} = 0$ . For a word that may be written in that form in a unique way (e.g.  $abccba$ ), the value of  $\tilde{J}_w^{ABOBA}$  is

$$J_{a^k}(t_0 + h/2; t_0) J_{b^\ell}(t_0 + h/2; t_0) J_{w'}(t_0 + h; t_0) \\ \times J_{b^m}(t_0 + h; t_0 + h/2) J_{a^n}(t_0 + h; t_0 + h/2),$$

which simply reads

$$= \frac{1}{k! \ell! m! n!} \left( \frac{h}{2} \right)^{k+\ell+m+n} J_{w'}(t_0 + h; t_0).$$

For a word that may be written in the form  $a^k b^\ell w' b^m a^n$  in several ways, we sum over all possible ways (e.g. for  $aa$ , we have  $\ell = m = 0$ ,  $w' = \emptyset$ , and three possibilities,  $(k, n) = (2, 0)$ ,  $(k, n) = (1, 1)$ ,  $(k, n) = (0, 2)$  leading to a coefficient  $(1/2)(h/2)^2 + (h/2)^2 + (1/2)(h/2)^2 = h^2/2$ ). Similar considerations, with the roles of  $a$  and  $b$  interchanged apply to the alternative BAOAB method. BUB and UBU cases are even simpler because of the smaller number of sections involved in them. It now takes no time to find the coefficients  $\tilde{J}_w(t_0 + h; t_0)$ , displayed in Table 7.3.

These coefficients *readily determine both the strong and weak order of convergence*, see table 8.1.

- *Strong order.* Observe that  $\tilde{J}_{Aa}^{ABOBA} = \tilde{J}_{Aa}^{BAOAB} \neq J_{Aa}$ . Hence, they both have a strong local error of the size  $O(h^{1.5})$ . In turn,  $\tilde{J}_w^{BUB} = \tilde{J}_w^{UBU} = J_w$

<i>Error</i>	ABOBA	BAOAB	BUB	UBU
<i>Strong</i>	1	1	2	2
<i>Weak</i>	2	2	2	2

Table 7.4: Strong and weak orders of convergence for integrators considered.

if  $\|w\| \leq 2.5$ . Thus, there exists such a  $C > 0$  that their local strong error is bounded by  $Ch^{2.5}$  if  $h$  is small enough.

- *Weak order.* It is not difficult to check that

$$\mathbb{E}\tilde{J}_w^{ABOBA} = \mathbb{E}\tilde{J}_w^{BAOAB} = \mathbb{E}\tilde{J}_w^{BUB} = \mathbb{E}\tilde{J}_w^{UBU} = \mathbb{E}J_w$$

for words  $\|w\| \leq 3$ . In turn, all integrators fail to satisfy this condition for words with stochastic weight  $\|w\| = 3.5$ . Hence, their weak local error is of the form  $O(h^3)$ .

The celebrated Milstein theorem [37] states that both ABOBA and BAOAB have strong order of convergence 1, whereas BUB and UBU have a second order of convergence. However, they all perform with the same weak order of convergence. Observe that the sparsity patterns  $(\star, 0)$  and  $(0, \star)$  present in the word basis functions make the one step errors satisfy the following unexpected properties:

$$\begin{aligned} \mathbb{E}\|p(t_0 + h) - p^{ABOBA}\| &= O(h^2), \\ \mathbb{E}\|p(t_0 + h) - p^{BAOAB}\| &= O(h^2), \\ \mathbb{E}\|q(t_0 + h) - q^{BUB}\| &= O(h^3), \\ \mathbb{E}\|q(t_0 + h) - q^{UBU}\| &= O(h^3). \end{aligned}$$

### 7.4.2 ABOBA and BAOAB comparison

In several papers Matthews C. and Leimkuhler B. (cf. [32], [33], [34]) observed that BAOAB turns out to be clearly superior to ABOBA. At this point, we are in a position to compare BAOAB and ABOBA. Since at the words  $A_i a$ ,  $i = 1, \dots, d$ , both methods are in error, for both of them, the local error expansion in (6.9) begins with  $\mathcal{O}(h^{3/2})$  terms. Furthermore ABOBA and BAOAB share the same coefficient values  $\tilde{J}_w$  at the leading (i.e.  $\mathcal{O}(h^{3/2})$ ) order and also at the next order (corresponding to words of weight 2). In fact, for the words that feature in Table

7.3, the only difference between both integrators corresponds to the words  $A_i ab$ ,  $i = 1, \dots, d$ . For these, the exact solution has coefficient

$$J_{A_i ab} \sim \mathcal{N}\left(0, \frac{h^5}{20}\right),$$

BAOAB has

$$\tilde{J}_{A_i ab}^{BAOAB} = \frac{h^2}{4} J_{A_i} \sim \mathcal{N}\left(0, \frac{h^5}{16}\right),$$

while, as noted above,

$$\tilde{J}_{A_i ab}^{ABOBA} = 0,$$

due to the pattern  $ab$  after the stochastic letter. The joint distribution of  $J_{A_i ab}$  and  $\tilde{J}_{A_i ab}^{BAOAB}$  is Gaussian with covariance  $h^5/24$  and therefore the correlation between both variables is

$$\frac{h^5/24}{\sqrt{h^5/20} \sqrt{h^5/16}} = \frac{\sqrt{5}}{3} \approx 0.74,$$

while  $J_{A_i ab}$  and  $\tilde{J}_{A_i ab}^{ABOBA}$  are obviously uncorrelated. Thus, for this word, ABOBA provides a very poor approximation to the exact coefficient. Due to the symmetric role played by the letters  $a$  and  $b$  in the algebra of words, for  $A_i ba$ , it is BAOAB that has an identically zero coefficient. However this is irrelevant for the present discussion because, for that word, the basis function vanishes as noted in (7.7).

Cases where  $f_w \neq 0$ ,  $\tilde{J}_w^{ABOBA} = 0$ , but  $\tilde{J}_w^{BAOAB}$  provides a nontrivial approximation to  $J_w$  occur for higher values of the weight. For the deterministic word  $cab$ ,  $\tilde{J}_{cab}^{BAOAB} = h^3/4$  and  $\tilde{J}_{cab}^{ABOBA} = 0$  (the correct value is  $h^3/6$ ). For  $A_i cab$  with weight  $7/2$ , the exact solution has

$$J_{A_i cab} \sim \mathcal{N}\left(0, \frac{h^7}{252}\right),$$

while

$$\tilde{J}_{A_i cab}^{BAOAB} \sim \mathcal{N}\left(0, \frac{h^7}{148}\right),$$

and, again due to the  $ab$  pattern,

$$\tilde{J}_{A_i cab}^{ABOBA} = 0.$$

Now the correlation between the BAOAB coefficient and the true value is  $\sqrt{21}/5 \approx 0.91$ .

Why does ABOBA provide poor approximations for words like  $A_i ba$ ,  $cab$ ,  $A_i cab$ ? By looking at the physical meaning of the corresponding word basis functions (see e.g. the discussion of (7.8) presented above), we see that the *above shortcomings of ABOBA stem from the following algorithmic source*. In any given

time step, ABOBA uses the *same* value of  $F$  in both kicks ( $q$  is not updated between those kicks) and, furthermore, that common value of  $F$  only depends on the values of  $q$  and  $p$  at the beginning of the step. Thus, over the whole step, the momentum increment  $hF$  due to the potential forces does not ‘see’ the presence of friction or noise in the current step. On the contrary, in BAOAB the change in  $p$  at substep O (friction and noise) causes that the kicking force varies from the first kick to the second.<sup>4</sup>

## 7.5 Ito-Stratonovich comparison

If we decide to work with the Ito approach instead of Stratonovich’s, statements such as Corollary 5.2 demonstrate that we eventually must reach the same conclusions but at a price: we should consider  $d$  extra letters, i.e. those of the form  $\bar{A}_i$ . Let  $\bar{\mathcal{A}}$  denote the Ito alphabet associated with Langevin equations. Recall that the Ito extension of the alphabet  $\mathcal{A}$  is the larger alphabet  $\bar{\mathcal{A}} = \mathcal{A} \cup \{\bar{A}_i\}_{1 \leq i \leq d}$ . As soon as we get down to calculate the corresponding word basis functions we realize that, as expected, those containing letters of the form  $\bar{A}_i$  vanish. Table (7.5) displays the word basis functions on words  $w \in \bar{\mathcal{A}}$  such that  $f_w \neq 0$  and  $\|w\| \geq 1.5$ . Now that the word basis functions coincide, at least,<sup>5</sup> up to weight 2.5, the iterant of any palindromic split-step integrator may be rewritten

$$\begin{aligned}
x^1 &= W_{\tilde{I}(t_0+h;t_0)}(x_0) = \sum_w \tilde{I}_w(t_0+h;t_0) f_w(x_0) \\
&= x_0 + \sum_{i=1}^d I_{A_i}(t_0+h;t_0) f_{A_i}(x_0) + h(f_a(x_0) + f_b(x_0) + f_c(x_0)) \\
&\quad + \sum_{i=1}^d \tilde{I}_{A_i a} f_{A_i a}(x_0) + \tilde{I}_{A_i c} f_{A_i c}(x_0) \\
&\quad + \frac{h^2}{2} (f_{ab}(x_0) + f_{ba}(x_0) + f_{bc}(x_0) + f_{ca}(x_0) + f_{cc}(x_0)) \\
&\quad + \sum_{i=1}^d \tilde{I}_{A_i ba} f_{A_i ba}(x_0) + \tilde{I}_{A_i ac} f_{A_i ac}(x_0) + \tilde{I}_{A_i cc} f_{A_i cc}(x_0) + \tilde{\mathcal{R}},
\end{aligned}$$

where the remainder is such that  $\mathbb{E}\|\tilde{\mathcal{R}}\|_2^2 = O(h^5)$  and  $\tilde{I}(t_0+h;t_0) \in \mathcal{G}_{qsh}$ . In order to conclude we only have to check that  $\tilde{I}_w(t_0+h;t_0) = \tilde{J}_w(t_0+h;t_0)$  for

<sup>4</sup>Note that BAOBA reuses in the first kick of the next step the value of  $F(q)$  employed in the second kick of the present step, so that both ABOBA and BAOAB use twice each evaluation of the potential force.

<sup>5</sup>There is no need to go further when examining both strong and weak errors.



the words involved, namely,  $w = A_i a, A_i c, A_i ab, A_i ca, A_i cc$ . We shall prove the following general statement:

**Proposition 7.1** *Take two (possibly empty) words  $u, v$  on the alphabet  $\bar{\mathcal{A}}$  containing only deterministic letters. Then, for each stochastic letter  $A_i$*

$$I_{uA_iv}(t_0 + h; t_0) = J_{uA_iv}(t_0 + h; t_0).$$

**Proof:** Firstly, for each  $s \geq 0$

$$I_u(t_0 + s; t_0) = J_u(t_0 + s; t_0) = \frac{s^{\|u\|}}{\|u\|!}.$$

Hence,

$$I_{uA_i}(t_0 + s; t_0) = J_{uA_i}(t_0 + s; t_0) = \int_{t_0}^{t_0+h} \frac{s^{\|u\|}}{\|u\|!} d\mathcal{B}_{A_i}(s).$$

Integrate next  $\|v\| \geq 0$  times with respect to  $dt$  in the corresponding domains both  $I_{uA_i}(t_0 + s; t_0)$  and  $J_{uA_i}(t_0 + s; t_0)$  in order to obtain  $I_{uA_iv}(t_0 + h; t_0)$  and  $J_{uA_iv}(t_0 + h; t_0)$ . In this way,

$$I_{uA_iv}(t_0 + h; t_0) = J_{uA_iv}(t_0 + h; t_0),$$

and we are done.  $\square$

**Corollary 7.1** *Take two (possibly empty) words  $u, v$  on the alphabet  $\bar{\mathcal{A}}$  containing only deterministic letters. Then, for each stochastic letter  $A_i$*

$$\tilde{I}_{uA_iv}(t_0 + h; t_0) = \tilde{J}_{uA_iv}(t_0 + h; t_0).$$

In addition, for each word  $w$  comprising only deterministic letters

$$\tilde{I}_w(t_0 + h; t_0) = \tilde{J}_w(t_0 + h; t_0).$$

**Proof:** As the sections involved in a split-step integrator  $\tilde{\phi}$  do not vary when the integral interpretation changes, the expressions of  $\tilde{J}$  and  $\tilde{I}$  as sums of products for words  $w$  not containing letters of the form  $\bar{A}_i$  are closely related:  $\tilde{J}_w$  is obtained from  $\tilde{I}_w$  by switching its  $I$  factors by  $J$  factors. In this way, using the former proposition:

$$\tilde{I}_{uA_iv} = \sum_j \prod_{i \in I_j} I_{w_i} = \sum_j \prod_{i \in I_j} J_{w_i} = \tilde{J}_{uA_iv}.$$

$w$	$constant$	$linear$	$(\star, 0)$	$(0, \star)$
$A_i a$	✓		✓	
$A_i c$	✓			✓
$ab$				✓
$ba$			✓	
$bc$				✓
$ca$		✓	✓	
$cc$		✓		✓
$A_i ab$				✓
$A_i ca$	✓		✓	
$A_i cc$	✓			✓

Table 7.5: Words  $w \in \bar{\mathcal{A}}$  of weight  $\|w\| \geq 1.5$  such that its corresponding word basis function does not vanish identically.

Similarly, if  $w$  only comprises deterministic letters,

$$\tilde{I}_w(t_0 + h; t_0) = \sum_j \prod_{i \in I_j} I_{w_i} = \sum_j \prod_{i \in I_j} J_{w_i} = \tilde{J}_w(t_0 + h; t_0).$$

□

In this way, when considering the Ito interpretation, all word basis expansions coincide, as expected, with those of Stratonovich at least up to order 3, providing the same conclusions that we have obtained before.

## 7.6 Modified equations

### 7.6.1 Drift modified equations

Given a split-step integrator  $\tilde{\phi}$ , Section 2.4 demonstrates the existence of a *formal vector field*  $\hat{f}$  with random coefficients that enables the construction of a stochastic differential equation

$$dy(t) = \hat{f}(y)dt + \sum_{i=1}^d f_{A_i}(y) \circ d\mathcal{B}_i(t) \quad (7.10)$$

which turns to be a strong *modified equation* for the split-step integrator: its solution satisfies

$$y(t_0 + h) = \phi_{t_0+h;t_0}(x_0),$$

$w$	ABOBA	BAOAB	BUB	UBU
$aba$	$h^3/4$	0	0	$h^3/4$
$bab$	0	$h^3/4$	$h^3/4$	0
$aab$	$h^3/8$	$h^3/4$	$h^3/4$	$h^3/8$
$cab$	0	$h^3/4$	$h^3/4$	$h^3/8$
$abc$	$h^3/4$	0	0	$h^3/4$
$cca$	$h^3/4$	$h^3/4$	$h^3/6$	$h^3/6$
$bca$	$h^3/4$	$h^3/4$	$h^3/4$	$h^3/8$
$bcc$	$h^3/4$	$h^3/4$	$h^3/4$	$h^3/8$
$ccc$	$h^3/6$	$h^3/6$	$h^3/6$	$h^3/6$

Table 7.6: Coefficients  $\tilde{J}_w(t_0+h; t_0)$  for the 3-weighted deterministic words. Each coefficient is of the form  $N_w h^3$ .

provided  $y(t_0) = x_0$ . Set  $\hat{f}(x) = W_R(x)$ . According to Theorem 3.6,  $R$  satisfies the boundary value problem

$$\boxed{dS(t) = S(t)Rdt + \sum_{i=1}^d S(t)A_i d\mathcal{B}_{A_i}(t).} \quad (7.11)$$

with  $S(0) = 1, S(h) = \tilde{J}(t_0 + h; t_0)$ . Firstly, it is straightforward to check that  $R_a = R_b = R_c = 1$  for any split-step integrator of strong order 1. Secondly, for words of the form  $w = A_i x$  with letters  $x \in \{a, b, c\}$  computations show

$$S_{A_i x}(s) = J_{A_i x}(t_0 + s; t_0) + sR_{A_i x}.$$

As  $S(h) = \tilde{J}(t_0 + h; t_0)$ ,

$$R_{A_i x} = \frac{\tilde{J}_{A_i x}(t_0 + h; t_0) - J_{A_i x}(t_0 + h; t_0)}{h}. \quad (7.12)$$

At first glance, there is no need to calculate the coefficients  $R_{x A_i}, R_{A_i A_j}$  because  $f_{x A_i} = f_{A_i A_j} = 0$  and they all seem not to interfere with any other current calculation. For the ABOBA, BAOAB, BUB and UBU integrators,<sup>6</sup>

$$R_{xy} = 0, S_{xy}(t) = 1,$$

<sup>6</sup>In general, for any palindromic integrator.

$w$	ABOBA	BAOAB	BUB	UBU
$aba$	$1/12$	$-1/6$	$-1/6$	$1/12$
$bab$	$-1/6$	$1/12$	$1/12$	$-1/6$
$aab$	$-1/24$	$1/12$	$1/12$	$-1/24$
$cab$	$-1/6$	$1/12$	$1/12$	$-1/24$
$abc$	$1/12$	$0$	$0$	$1/12$
$cca$	$1/12$	$1/12$	$0$	$0$
$bca$	$1/12$	$1/12$	$1/12$	$-1/24$
$bcc$	$1/12$	$1/12$	$1/12$	$-1/24$
$ccc$	$0$	$0$	$0$	$0$

Table 7.7: Coefficients  $\frac{1}{h^2}R_w = N_w - \frac{1}{6}$  for 3-letter deterministic words, required to construct the drift-modified equations.

for  $x, y \in \{a, b, c\}$ . Plugging the expression of the formal vector field into (7.10),<sup>7</sup> we obtain

$$dq = \left( M^{-1}p + \sigma \sum_{i=1}^d \frac{R_{A_i a}}{\sqrt{m_i}} + \mathcal{R}_{1.5} \right) dt, \quad (7.13)$$

$$dp = \left( F(q) - \gamma\sigma \sum_{i=1}^d \sqrt{m_i} R_{A_i c} + \mathcal{R}'_{1.5} \right) dt - \gamma p dt + \sigma M^{1/2} d\mathcal{B}(t),$$

where the remainder  $\mathcal{R}_{1.5}$  and  $\mathcal{R}'_{1.5}$  are random variables whose  $L^2$  norm is  $O(h^{1.5})$ . Let us cast some more light on the previous system. For words of the form  $w = A_i xy$ ,  $x, y \in \{a, b, c\}$ ,

$$S_{A_i xy}(s) = J_{A_i xy}(t_0 + s; t_0) + \frac{s^2}{2} R_{A_i x} + s R_{A_i xy}.$$

Hence,

$$R_{A_i ab} = \frac{\tilde{J}_{A_i xy}(t_0 + h; t_0) - J_{A_i xy}(t_0 + h; t_0)}{h} - \frac{h}{2} R_{A_i x}. \quad (7.14)$$

In this way, we continue to make explicit terms in (7.13):

$$dq = \left( M^{-1}p + \sigma \sum_{i=1}^d \frac{R_{A_i a}}{\sqrt{m_i}} - \gamma\sigma \sum_{i=1}^d \frac{R_{A_i ca}}{\sqrt{m_i}} + O(h^2) \right) dt, \quad (7.15)$$

$$dp = \left( F(q) - \gamma\sigma \sum_{i=1}^d \sqrt{m_i} R_{A_i c} + \sigma \sum_{i=1}^d \frac{R_{A_i ab}}{\sqrt{m_i}} \frac{\partial F(q)}{\partial q_i} + \gamma^2 \sigma \sum_{i=1}^d \sqrt{m_i} R_{A_i cc} \right) dt + O(h^2) dt - \gamma p dt + \sigma M^{1/2} d\mathcal{B}(t). \quad (7.16)$$

<sup>7</sup>We should pay attention to Table 7.3.1

We aim next to obtain the analytic expression of the  $O(h^2)$  terms. Look at Table 7.6.1. For each word  $w = xyz$  with  $x, y, z \in \{a, b, c\}$ , define the auxiliary coefficient  $N_w \in \mathbb{N}$  by means of the equation

$$\tilde{J}_w(t_0 + h; t_0) = N_w h^3.$$

Solving the equations of Theorem 3.6 we obtain

$$S_w(t) = \frac{t^3}{6} + tR_w.$$

Hence, if  $S_w(h) = \tilde{J}_w(t_0 + h; t_0) = N_w h^3$ , necessarily

$$R_w = (N_w - \frac{1}{6})h^2.$$

Take a look at Table 7.3.1 and observe again the patterns  $(\star, 0)$  and  $(0, \star)$ . We can now write down explicitly the  $O(h^2)$  terms in in equation (7.13):

$$\begin{aligned} dq &= (M^{-1}p + \sigma \sum_{i=1}^d \frac{R_{A_i a}}{\sqrt{m_i}} - \gamma \sigma \sum_{i=1}^d \frac{R_{A_i c a}}{\sqrt{m_i}}) dt, \\ &+ h^2 \left( (N_{aba} - \frac{1}{6}) f_{aba}(x) + (N_{bca} - \frac{1}{6}) f_{bca}(x) + (N_{cca} - \frac{1}{6}) f_{cca}(x) \right) dt \\ &+ \mathcal{R}_{2.5}, \end{aligned} \tag{7.17}$$

$$\begin{aligned} dp &= (F(q) - \gamma \sigma \sum_{i=1}^d \sqrt{m_i} R_{A_i c} + \sigma \sum_{i=1}^d \frac{R_{A_i ab}}{\sqrt{m_i}} \frac{\partial F(q)}{\partial q_i} + \gamma^2 \sigma \sum_{i=1}^d \sqrt{m_i} R_{A_i cc}) dt \\ &+ h^2 \left( (N_{bab} - \frac{1}{6}) f_{bab}(x) + (N_{aab} - \frac{1}{6}) f_{aab}(x) + (N_{cab} - \frac{1}{6}) f_{cab}(x) \right. \\ &+ (N_{abc} - \frac{1}{6}) f_{abc}(x) + (N_{bbc} - \frac{1}{6}) f_{bcc}(x) + (N_{ccc} - \frac{1}{6}) f_{ccc}(x) \left. \right) dt \\ &+ \mathcal{R}'_{2.5} - \gamma p dt + \sigma M^{1/2} d\mathcal{B}(t). \end{aligned} \tag{7.18}$$

This is what a *strong modified equation* for any split-step integrator of strong order at least 1 is like. The coefficients are computed by using formulas (7.12), (7.14) and Table 7.6.1. Observe that  $N_w \in \mathbb{Q}$  whereas  $R_w$  are  $\mathcal{F}_{t_0+h}$ -measurables random variables.<sup>8</sup>

## 7.6.2 Perturbed Hamiltonian dynamics

From now on, we restrict our analysis to the unidimensional case. The following constructions are easily extended to the  $d$ -dimensional case. Look carefully at the

<sup>8</sup>As usually, the filtration  $\mathcal{F}$  is the natural filtration of the Brownian motions.

word basis displayed in subsection 7.3.2. Go to weight 3. There is something special about the first couple of word basis  $f_{aba}, f_{aab}$ : consider

$$H(q, p) = \frac{F'(q)p^2}{m^2}$$

and observe that

$$\begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} = 2f_{aba} - f_{aab} = \begin{pmatrix} \frac{2F'(q)p}{m^2} \\ -\frac{F''(q)p^2}{m^2} \end{pmatrix}.$$

Choose next

$$H(q, p) = -\frac{\gamma F'(q)p}{m},$$

and focus on the second row of word basis functions present in there. Now,

$$\begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} = f_{bca} - \lambda f_{bab} - (1 - \lambda)f_{bcc} = \begin{pmatrix} \frac{F'(q)p}{m^2} \\ -\frac{F''(q)p^2}{m^2} \end{pmatrix},$$

for any  $\lambda \in \mathbb{R}$ . The last row of word basis functions correspond to the following Hamiltonian:

$$H(q, p) = -\gamma^2 V(q) + \frac{F^2(q)}{2m} + \gamma^2 \frac{p^2}{2m}.$$

Observe that

$$\begin{pmatrix} \frac{\partial H}{\partial p} \\ -\frac{\partial H}{\partial q} \end{pmatrix} = \begin{pmatrix} \gamma^2 \frac{p}{m} \\ -\gamma^2 F(q) - \frac{F(q)F'(q)}{m} \end{pmatrix} = f_{cca} - f_{bab} - f_{bcc}.$$

There is no such a similar analysis for the word basis corresponding to  $ccc$  but this is not a cause for concern: for every split-step integrator  $\tilde{J}_{ccc}(t_0 + h; t_0) = h^3/6$ , hence  $R_{ccc} = 0$ . This means that  $f_{ccc}$  does not appear on the word basis expansion of the modified equation.

**Theorem 7.1 (Modified Hamiltonian)** *Consider a split-step integrator for the Langevin equations (7.1) in  $\mathbb{R}^d$  and let  $\tilde{J}(t_0 + h; t_0)$  be its associated random point in the group. Assume that the potential function  $q \in \mathbb{R}^d \mapsto V(q)$  is not constant. Then, there exists a random perturbation  $\tilde{H}(q, p)$  of the Hamiltonian*

$$(q, p) \mapsto H(q, p) = V(q) + \frac{p^T M^{-1} p}{2}$$

such that the system

$$\begin{aligned} dq &= \frac{\partial \tilde{H}}{\partial p} dt, \\ dq &= -\frac{\partial \tilde{H}}{\partial q} dt - \gamma p + \sigma \sqrt{M} d\mathcal{B}(t), \end{aligned} \tag{7.19}$$

is a strong modified equation for the splitting, if and only if

c1  $N_{aba} + 2N_{aab} = 1/2$ , or, equivalently  $R_{aba}/2 = -R_{aab}$  if  $F(q) = -\nabla V(q)$  is not constant. In other case,  $N_{aba} = 1/6$ , i.e.  $R_{aba} = 0$ .

c2  $N_{bca} + 1/6 = N_{cab} + N_{abc}$ , or equivalently,  $R_{bca} = R_{cab} + R_{abc}$ .

In this way, if  $d = 1$  and the split-step has strong order 1, the perturbed Hamiltonian is

$$\begin{aligned} \tilde{H}(q, p) = & V(q) + \frac{p^2}{2m} + \left( \frac{\sigma}{\sqrt{m}} R_{Aa} - \frac{\gamma\sigma}{\sqrt{m}} R_{Aca} \right) p - \frac{\sigma}{\sqrt{m}} F(q) R_{Aab} \\ & + \left( \frac{\gamma\sigma}{\sqrt{m}} R_{Ac} - \gamma^2 \sigma \sqrt{m} R_{Acc} \right) q \\ & + h^2 \left( R_{bab} \frac{F^2(q)}{2m} - R_{bcc} \gamma^2 V(q) + R_{cca} \frac{\gamma^2 p^2}{2m} \right. \\ & \left. - R_{bca} \frac{\gamma F(q) p}{m} + R_{aba} \frac{F'(q) p^2}{2m^2} \right) + \mathcal{R}_{2.5}. \end{aligned} \quad (7.20)$$

**Proof:** The result, at least in the one-dimensional case, is a direct consequence of the preparations above. Other cases follow easily. Given a split-step integrator, compute first  $\tilde{J}(t_0 + h; t_0)$ . Secondly, obtain the random series  $R_h$  and construct  $W_{R_h}(q, p) = \hat{f}(q, p) = W_R(q, p)$ . Suppose

$$\hat{f}(q, p) = \begin{pmatrix} \frac{\partial \tilde{H}}{\partial p} \\ -\frac{\partial \tilde{H}}{\partial q} \end{pmatrix}$$

in order to obtain  $\tilde{H}$  if possible. Then, the conditions stated in the theorem appear naturally.  $\square$

What we have done is to construct a new intrinsic object related to some split step integrators. Obviously, the agreement

$$\mathbb{E}^{x_0} \|\tilde{H}(q, p) - H(q, p)\| = O(h^m)$$

with  $m \in \mathbb{Z}/2$  forces the integrator to have a strong local error of the form  $O(h^{m+1})$ .

The following corollary provides one more reason in favour of BAOAB over ABOBA and BUB over UBU.

**Corollary 7.2 (Modified Hamiltonian dynamics)** *There is a modified Hamiltonian dynamics for BAOAB and for BUB but not for ABOBA nor UBU.*

**Proof:** Just check the conditions of Theorem 7.1 for these four methods. In particular, the modified Hamiltonian for BAOAB is

$$\begin{aligned}\tilde{H}^{BAOAB}(q, p) &= V(q) + \frac{p^2}{2m} + \\ &+ \left( \frac{\sigma}{\sqrt{m}} R_{Aa} - \frac{\gamma\sigma}{\sqrt{m}} R_{Aca} \right) p - \frac{\sigma}{\sqrt{m}} F(q) R_{Aab} \\ &+ h^2 \left( \frac{F^2(q)}{24m} - \frac{\gamma^2}{12} V(q) + \frac{\gamma^2 p^2}{24m} - \gamma \frac{F(q)p}{12m} - \frac{F'(q)p^2}{12m^2} \right) \\ &+ \mathcal{R}_{2.5},\end{aligned}\tag{7.21}$$

with

$$\begin{aligned}R_{Aa} &= \frac{J_a}{2} - \frac{J_{Aa}}{h}, \\ R_{Aca} &= \frac{J_{Ac}}{2} - \frac{J_{Aca}}{h}, \\ R_{Aab} &= \frac{J_{Aa}}{2} - \frac{J_{Aab}}{h}.\end{aligned}$$

However, the Hamiltonian for BUB is:

$$\begin{aligned}\tilde{H}^{BUB}(q, p) &= V(q) + \frac{p^2}{2m} - \frac{\sigma}{\sqrt{m}} F(q) R_{Aab} \\ &+ h^2 \left( \frac{F^2(q)}{24m} - \frac{\gamma^2}{12} V(q) - \gamma \frac{F(q)p}{12m} - \frac{F'(q)p^2}{12m^2} \right) \\ &+ \mathcal{R}_{2.5},\end{aligned}\tag{7.22}$$

with

$$R_{Aab} = \frac{J_{Aa}}{2} - \frac{J_{Aab}}{h}.$$

□

## 7.7 Appendix: distributions of some auxiliary random variables

We now present some useful computations. All them are necessary to carry out the implementation of the methods.



- The distribution of  $\int_a^b e^{-\gamma(b-s)} d\mathcal{B}(s)$  is

$$\mathcal{N}\left(0, \sqrt{\frac{1 - e^{-2\gamma(b-a)}}{2\gamma}}\right),$$

which *only depends on*  $b - a$ .

- In particular,

$$\int_{t_0}^{t_0+h} e^{\gamma s} d\mathcal{B}(s) \sim \mathcal{N}\left(0, \sqrt{\frac{e^{2\gamma(t_0+h)} - e^{2\gamma t_0}}{2\gamma}}\right).$$

- Set  $\Delta\mathcal{B}(t_0) := \mathcal{B}(t_0 + h) - \mathcal{B}(t_0)$  and  $\delta = e^{-h\gamma}$ . If  $\Sigma$  stands for the covariance matrix,

$$\Sigma(\Delta\mathcal{B}(t_0), \int_{t_0}^{t_0+h} e^{-\gamma(t_0+h-s)} d\mathcal{B}(s)) = \begin{pmatrix} \sqrt{h} & \frac{1-\delta}{\gamma} \\ \frac{1-\delta}{\gamma} & \sqrt{\frac{1-\delta^2}{2\gamma}} \end{pmatrix}.$$

The correlation coefficient  $\rho$  of  $\Delta\mathcal{B}(t_0)$  and  $\int_{t_0}^t e^{-\gamma(t_0+h-s)} d\mathcal{B}(s)$  is

$$\rho(\Delta\mathcal{B}(t_0), \int_{t_0}^t e^{-\gamma(t_0+h-s)} d\mathcal{B}(s)) = \sqrt{\frac{1-\delta}{1+\delta} \cdot \frac{2}{h\gamma}}.$$

- Set next  $V = J_{Aa}(t_0 + h; t_0) - \tilde{J}_{Aa}^{ABOBA}(t_0 + h; t_0) = \int_{t_0}^{t_0+h} t_0 + \frac{h}{2} - s d\mathcal{B}(s)$ . Then

$$\Sigma(V, \int_{t_0}^{t_0+h} e^{-\gamma(t_0+h-s)} d\mathcal{B}(s)) = \begin{pmatrix} \sqrt{\frac{h^3}{12}} & \frac{1-\delta}{\gamma^2} - \frac{h}{2\gamma}(1+\delta) \\ \frac{1-\delta}{\gamma^2} - \frac{h}{2\gamma}(1+\delta) & \sqrt{\frac{1-\delta^2}{2\gamma}} \end{pmatrix}$$

and

$$\rho(V(h), \int_{t_0}^{t_0+h} e^{-\gamma(t_0+h-s)} d\mathcal{B}(s)) = 2\sqrt{\frac{6}{h^3}} \left( \sqrt{\frac{1-\delta}{(1+\delta)\gamma^3}} - \frac{h}{2} \sqrt{\frac{1+\delta}{(1-\delta)\gamma}} \right).$$

# Chapter 8

## Tuned methods

Word series turn out to be useful to obtain the expansion in terms of word basis functions and word basis operators of split-step integrators. This allows us to obtain, among other things, the strong and weak order conditions only by comparing such expansion with those of the true solution. But word series expansions of split-step integrators provide much more information. In some cases, including the Langevin equations, they may be used to *tune* certain integrators by replacing those  $\tilde{J}_w(t_0 + h; t_0)$  in error by the correct values  $J_w(t_0 + h; t_0)$ .

### 8.1 Tuning procedure

Look carefully at words of the form  $w = A_i a$ . They are key in the “low” strong error 1 of both BAOAB and ABOBA: both *break* the U section into C and O so that they cannot replicate the true  $J_w(t_0 + h; t_0)$ ; they give instead the value  $hJ_{A_i}(t_0 + h; t_0)$ , which in any case is not a bad approximation:

- $Cov(J_w(t_0 + h; t_0), \tilde{J}_w(t_0 + h; t_0)) = \frac{h^3}{2}$ ,
- $Var(J_w(t_0 + h; t_0)) = h^3$ ,
- $Var(\tilde{J}_w(t_0 + h; t_0)) = \frac{h^3}{3}$ .

If the coefficient  $\tilde{J}_w(t_0 + h; t_0)$  is in error, we can perhaps take advantage of its companion word basis function. Focus next on Table 7.3.1. Fortunately there is a checkmark for words  $A_i a$  ensuring that  $f_w$  is constant. Then, if we simulate the true  $J_w(t_0 + h; t_0)$  jointly with one step iteration of, say, ABOBA, the simple “extract and replace” move

$$\hat{x}_1^{ABOBA} := x_1^{ABOBA} + \sum_{i=1}^d (J_{A_i a} - \tilde{J}_{A_i a}^{ABOBA}) f_{A_i a}(x_0)$$

does not cost anything because of the constant nature of  $f_{A_i a}$  and results in a strong second order integrator using one force evaluation per time step. The same construction can be easily done in the BAOAB case, although the force evaluation can no longer be reused. This “extract and replace” move is critical to the development of many integrators of high strong order for Langevin Dynamics (and for any other stochastic differential equations).

## 8.2 ABOBA and BAOAB tuned

Let  $x^{ABOBA}$  be the result of a step of ABOBA, starting at  $x_0$ . The *correction*

$$\hat{x}_1^{ABOBA} := x_1^{ABOBA} + \sum_{i=1}^d (J_{A_i a} - \tilde{J}_{A_i a}^{ABOBA}) f_{A_i a}(x_0)$$

*tunes* its strong order: the new numerical scheme satisfies

$$\mathbb{E} \|\hat{x}_1^{ABOBA} - x(t_0 + h)\| = O(h^{2.5}),$$

as can be readily computed by checking its word basis expansion. The same argument is valid for BAOAB. We name these two methods ABOBA *tuned* and BAOAB *tuned*. The numerical advantage lies in the *constant nature* of  $f_{A_i a}$ :

$$f_{A_i a}(x_0) = \begin{pmatrix} \frac{\sigma}{\sqrt{m_i}} e_i \\ 0 \end{pmatrix}, \quad (8.1)$$

where  $e_i$  and  $0$  are respectively the  $i$ -coordinate vector and the zero vector of  $\mathbb{R}^d$ . Rewrite compactly

$$J_{A_i a} - \tilde{J}_{A_i a}^{ABOBA} = J_{A_i a} - J_{A_i} \frac{h}{2} = \int_{t_0}^{t_0+h} (t_0 + \frac{h}{2} - s) d\mathcal{B}(s)$$

and observe that  $f_{A_i a}$  is of the form  $(\star, 0)$ , so that the correction *is only performed in the  $q$  component*. In this way, each one-step approximation  $q^{ABOBA}$  of ABOBA may be *corrected* as

$$\boxed{q^{ABOBA} + M^{-1/2} \sigma \int_{t_0}^{t_0+h} (t_0 + \frac{h}{2} - s) d\mathcal{B}(s),}$$

obtaining a two strong order scheme.

**ABOBA-BAOAB tuned implementetation of a time-step**

Set first

$$\delta := e^{-\gamma h}, \quad \rho = 2\sqrt{\frac{6}{h^3}} \left( \sqrt{\frac{1-\delta}{\gamma^3(1+\delta)}} - \frac{h}{2} \sqrt{\frac{1+\delta}{\gamma(1-\delta)}} \right),$$

$$\sigma_2 := \sqrt{\frac{1}{2\gamma}(1-\delta^2)}.$$

Then, given  $q_n, p_n \in \mathbb{R}^d$ :

- Generate two independent  $d$ -dimensional standard normal random vectors  $Z_{n+1}^{(1)}, Z_{n+1}^{(2)}$ .

- Compute  $X_{n+1}^{(1)}, X_{n+1}^{(2)}$ :

$$X_{n+1}^{(1)} = \frac{h\sqrt{h}}{12} Z_{n+1}^{(1)},$$

$$X_{n+1}^{(2)} = \sigma_2 \rho Z_{n+1}^{(1)} + \sigma_2 \sqrt{1-\rho^2} Z_{n+1}^{(2)},$$

- Perform one time step with ABOBA in case of ABOBA tuned (respec. BAOAB for BAOAB tuned):

$$q_{n+1/2} = q_n + \frac{h}{2} M^{-1} p_n,$$

$$p_{n+1/3} = p_n + \frac{h}{2} F(q_{n+1/2}),$$

$$p_{n+2/3} = p_{n+1/3} \delta + \sigma e^{\gamma(t_0+h)} M^{1/2} X_{n+1}^{(2)},$$

$$p_{n+1} = p_{n+2/3} + \frac{h}{2} F(q_{n+1/2}),$$

$$q'_{n+1} = q_{n+1/2} + \frac{h}{2} M^{-1} p_{n+1}.$$

- Perform the correction:

$$q_{n+1} = q'_{n+1} + M^{-1/2} \sigma X_{n+1}^{(1)}. \quad (8.2)$$

□

Observe that the distribution of

$$\left( \int_{t_0}^{t_0+h} (t_0 + \frac{h}{2} - s) d\mathcal{B}_1(s), \dots, \int_{t_0}^{t_0+h} (t_0 + \frac{h}{2} - s) d\mathcal{B}_d(s), \right. \\ \left. \int_{t_0}^{t_0+h} e^{\gamma s} d\mathcal{B}_1(s), \dots, \int_{t_0}^{t_0+h} e^{\gamma s} d\mathcal{B}_d(s) \right)$$

<i>Error</i>	ABOBA	BAOAB	BUB	UBU	ABOBA <i>tuned</i>	BAOAB <i>tuned</i>
<i>Strong</i>	1	1	2	2	2	2
<i>Weak</i>	2	2	2	2	2	2

Table 8.1: Strong and weak orders of convergence.

is that of  $(X_{n+1}^{(1)T}, X_{n+1}^{(2)T})$ .

Note that in spite of the fact that ABOBA tuned only requires one force evaluation per time step, BAOAB tuned no longer can reuse the second force evaluation: BAOAB tuned needs two force evaluations per time step.

## 8.3 Deterministic integrators tuned

### 8.3.1 Tuning damped Hamiltonian

The tuning procedure may also be used to turn deterministic integrators into stochastic ones. In particular, any second order accurate integrator for the deterministic problem,<sup>1</sup>

$$\begin{aligned} q' &= M^{-1}p \\ p' &= F(q) - \gamma p \end{aligned} \quad (8.3)$$

where  $q, p \in \mathbb{R}^d$ , may be tuned to obtain a strong second order integrator for the Langevin dynamics. Indeed, set  $x(t) = (q(t), p(t))$  and let  $x_1$  be the step of such an integrator with step-size  $h$ , starting at  $x_0$  and  $t = t_0$ . Then, according to Theorem 2.4 and Table 7.3.1

$$\begin{aligned} x_1 &= x_0 + h(f_a(x_0) + f_b(x_0) + f_c(x_0)) + \frac{h^2}{2}(f_{ab}(x_0) + f_{ba}(x_0) \\ &\quad + f_{bc}(x_0) + f_{ca}(x_0) + f_{cc}(x_0)) + O(h^3). \end{aligned}$$

If we want to perform a correction  $\hat{x}_1$  on  $x_1$  such that

$$\mathbb{E} \|x(t_0 + h) - \hat{x}_1\|_2^2 = O(h^5),$$

we are forced to consider

$$\hat{x}_1 = x_1 + \sum_{i=1}^d J_{A_i} f_{A_i}(x_0) + \sum_{i=1}^d J_{A_i a} f_{A_i a}(x_0) + \sum_{i=1}^d J_{A_i c} f_{A_i c}(x_0), \quad (8.4)$$

<sup>1</sup>We may simply take  $\sigma = 0$  in Langevin dynamics (7.1)

where of course all iterated integrals have time domain  $(t_0, t_0 + h)$ . But the question is: Is this numerical integrator *competitive when considering its numerical cost*? Take next a look at Table 7.3.1. We find a checkmark in the column “constant” when the words are  $A_i$ ,  $A_{ia}$  and  $A_{ic}$ . This is crucial as it enables us to perform the correction (8.4) without *any additional cost*. After writing down all the word basis involved, the correction turns out to be

$$\begin{aligned}\hat{q} &= q_1 + \frac{\sigma}{\sqrt{m}} J_{Aa}, \\ \hat{p} &= p_1 + \sigma\sqrt{m}(J_A - \gamma J_{Ac}).\end{aligned}\tag{8.5}$$

### Implementetation of a step of a tuned deterministic integrator

Given  $q_n, p_n \in \mathbb{R}^d$ :

- Draw two independent  $d$ -dimensional standard normal random vectors  $Z_{n+1}^{(1)}$ ,  $Z_{n+1}^{(2)}$ .
- Compute  $X_{n+1}^{(1)}, X_{n+1}^{(2)}$ :

$$\begin{aligned}X_{n+1}^{(1)} &= \sqrt{h}Z_{n+1}^{(1)}, \\ X_{n+1}^{(2)} &= \frac{h\sqrt{h}}{2}Z_{n+1}^{(1)} + \frac{h\sqrt{h}}{2\sqrt{3}}Z_{n+1}^{(2)},\end{aligned}$$

In such a way, the distribution of

$$\begin{aligned}(\mathcal{B}_1(t_0 + h) - \mathcal{B}_1(t_0), \dots, \mathcal{B}_d(t_0 + h) - \mathcal{B}_d(t_0), \\ , J_{A_{1a}}(t_0 + h; t_0), \dots, J_{A_{da}}(t_0 + h; t_0))\end{aligned}$$

is that of  $(X_{n+1}^{(1)T}, X_{n+1}^{(2)T})$ .

- Perform one time step with the deterministic integrator  $(q_n, p_n)^T \mapsto (q'_{n+1}, p'_{n+1})^T$ .
- Perform the correction (8.4):

$$\begin{aligned}q_{n+1} &= q'_{n+1} + \sigma M^{-1/2} X_{n+1}^{(2)}, \\ p_{n+1} &= p'_{n+1} + \sigma M^{1/2} (X_{n+1}^{(1)} - \gamma X_{n+1}^{(2)}).\end{aligned}$$

□

**Examples**

- **Strang splitting tuned.** Split the dynamic (8.3) into the pieces

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}p \\ -\gamma p \end{pmatrix}}_{(1)} + \underbrace{\begin{pmatrix} 0 \\ F(q) \end{pmatrix}}_{(2)},$$

and consider then the Strang splitting integrator,

$$\tilde{\phi}_h = \phi_{h/2}^{(1)} \circ \phi_h^{(2)} \circ \phi_{h/2}^{(1)}. \quad (8.6)$$

The previous integrator is second order accurate and only requires one force evaluation per time step. Draw two independent  $d$ -dimensional standard normal random vectors  $Z_{n+1}^{(1)}, Z_{n+1}^{(2)}$  and compute subsequently  $X_{n+1}^{(1)}, X_{n+1}^{(2)}$ :

$$\begin{aligned} X_{n+1}^{(1)} &= \sqrt{h}Z_{n+1}^{(1)}, \\ X_{n+1}^{(2)} &= \frac{h\sqrt{h}}{2}Z_{n+1}^{(1)} + \frac{h\sqrt{h}}{2\sqrt{3}}Z_{n+1}^{(2)}, \end{aligned}$$

Last, setting  $(q', p') = \tilde{\phi}_h(q_0, p_0)$ , we perform

$$\begin{aligned} \hat{q} &= q' + \sigma M^{-1/2}X_{n+1}^{(2)}, \\ \hat{p} &= p' + \sigma M^{1/2}(X_{n+1}^{(1)} - \gamma X_{n+1}^{(2)}) \end{aligned} \quad (8.7)$$

to obtain the *tuned* integrator.

- **RK tuned.** Consider the explicit second order Runge-Kutta scheme

$$x_1 = x_0 + hf\left(x_0 + \frac{h}{2}f(x_0)\right) \quad (8.8)$$

and applied it to the damped Hamiltonian system (8.3). This means that

$$f(q, p) = \begin{pmatrix} M^{-1}p \\ -\gamma p + F(q) \end{pmatrix}.$$

Subsequently draw two independent  $d$ -dimensional standard normally distributed random vectors  $Z_{n+1}^{(1)}, Z_{n+1}^{(2)}$  and compute  $X_{n+1}^{(1)}, X_{n+1}^{(2)}$ :

$$\begin{aligned} X_{n+1}^{(1)} &= \sqrt{h}Z_{n+1}^{(1)}, \\ X_{n+1}^{(2)} &= \frac{h\sqrt{h}}{2}Z_{n+1}^{(1)} + \frac{h\sqrt{h}}{2\sqrt{3}}Z_{n+1}^{(2)}. \end{aligned}$$

Set first  $(q', p') = \phi_h(q_0, p_0)$  and then perform

$$\begin{aligned}\hat{q} &= q' + \sigma M^{-1/2} X_{n+1}^{(2)}, \\ \hat{p} &= p' + \sigma M^{1/2} (X_{n+1}^{(1)} - \gamma X_{n+1}^{(2)})\end{aligned}\quad (8.9)$$

to obtain again the *tuned* integrator.

### 8.3.2 Tuning perturbed Hamiltonian dynamics

More options to tune are available. Consider the perturbed Hamiltonian system (for a given  $h > 0$ , a would-be time step):

$$\begin{aligned}q' &= M^{-1}p, \\ p' &= F(q)\left(1 - \frac{\gamma h}{2}\right).\end{aligned}\quad (8.10)$$

Set first  $\Lambda = 1 - \gamma h/2$ . The Taylor expansion of the flow  $\Phi_t(q_0, p_0)$  rewritten in terms of word series is:<sup>2</sup>

$$\begin{aligned}\Phi_t(q_0, p_0) &= (q_0, p_0) + h(f_a + \Lambda f_b) + \frac{h^2 \Lambda}{2}(f_{ab} + f_{ba}) + O(h^3) \\ &= (q_0, p_0) + h(f_a + f_b) + \frac{h^2}{2}(f_{ab} + f_{ba} - \gamma f_b) + O(h^3).\end{aligned}$$

A direct computation shows that  $-\gamma f_b = f_{bc}$ . Thus,

$$\Phi_t(q_0, p_0) = (q_0, p_0) + h(f_a + f_b) + \frac{h^2}{2}(f_{ab} + f_{ba} + f_{bc}) + O(h^3).$$

For convenience we set  $x := (q, p)$ ,  $x_0 := (q_0, p_0)$ . For the step size  $h$ , a second order integrator for the system (8.10) generates an approximation  $x_1$  such that

$$\Phi_h(q_0, p_0) = x_1 + O(h^3).$$

Hence, the correction to obtain a strong second order method for the Langevin equations is:

$$\begin{aligned}\hat{x}_1 &= x_1 + \sum_{i=1}^d J_{A_i} f_{A_i}(x_0) + h f_c(x_0) + \sum_{i=1}^d J_{A_i a} f_{A_i a}(x_0) \\ &\quad + \sum_{i=1}^d J_{A_i c} f_{A_i c}(x_0) + \frac{h^2}{2}(f_{ca}(x_0) + f_{cc}(x_0)).\end{aligned}\quad (8.11)$$

Explicitly, if the deterministic integrator generates the approximation  $(q, p)$ , then

$$\begin{aligned}\hat{q} &= q - \frac{\gamma h^2}{2m} p_0 + \frac{\sigma}{\sqrt{m}} J_{A_a}, \\ \hat{p} &= p + \left(-\gamma h + \frac{\gamma^2 h^2}{2}\right) p_0 + \sigma \sqrt{m} (J_A - \gamma J_{A_c}).\end{aligned}\quad (8.12)$$

<sup>2</sup>assume the word basis are evaluated at  $(q_0, p_0)$



## 8.4 Further considerations

The tuning procedure can be applied to any kind of stochastic differential equations, although it is useful mainly when some vector fields are constant. It is the constant (or maybe linear) nature of some word basis functions which makes it possible to avoid the computational cost of evaluating intricated derivatives. The tuning procedure is quite flexible. For example, if we have an integrator of strong order three for the system

$$\begin{aligned}q' &= M^{-1}p, \\p' &= F(q) - \gamma p + \tilde{\sigma},\end{aligned}$$

where  $\tilde{\sigma}$  is a certain random variable depending on  $\sigma$ , then we may obtain an integrator of strong order three for Langevyn dynamics only adding three extra random normal variables. However, in this case we face a crucial problem: any deterministic integrator we choose needs *compulsorily* two force evaluations. This makes it useless in many practical cases.

# Chapter 9

## Numerical experiments

We present some numerical experiments that illustrate the performance of several numerical integrators for the Langevin equations. All of them have been studied previously in Chapters 7 and 8. The integrators we have chosen are:

1. ABOBA,
2. ABOBA tuned,
3. BAOAB,
4. BAOAB tuned,
5. BUB,
6. UBU,
7. “Strang tuned”, see (8.9).
8. “RK tuned”, see (8.8).

The main issue with regard to the computational cost is the *number of force evaluations per time step*. Recall that all of them but BAOAB tuned and RK tuned use one force evaluation per time step. Apparently, some of them require two force evaluations per time step but this problem is easily tackled by reusing a former force evaluation of the previous step. This is the case of BUB and BAOAB. The tuning procedure precludes such reuse so that BAOAB tuned compulsorily needs two force evaluations per time step depriving it of usefulness. As the RK method we have chosen uses two force evaluations to integrate the associated deterministic problem, RK tuned also fails to be useful. The other six methods, which have similar computational cost, result to be considerably cheap. Their memory storage is similar. They require the simulation of a number of independent standard normally distributed random variables ranging from  $d$  to  $2d$  per time step.

## 9.1 Strong errors

The following tables show both the *empirical average* and the *empirical standard deviation* of the random variables<sup>1</sup>

$$\boxed{\epsilon_q = |q(1) - \hat{q}| \quad \text{and} \quad \epsilon_p = |p(1) - \hat{p}|.}$$

Obviously, we are interested in the values

$$\boxed{\mathbb{E}\epsilon_q = \mathbb{E}|q(1) - \hat{q}|, \quad \mathbb{E}\epsilon_p = \mathbb{E}|p(1) - \hat{p}|,} \quad (9.1)$$

and

$$\boxed{\text{Var}(\epsilon_q), \quad \text{Var}(\epsilon_p),} \quad (9.2)$$

that we compute via a Monte Carlo method (see [4], [23]). The experiments have been conducted with the following parameter choices:

- Number of trajectories  $N = 10000$ .
- Dimension of the system  $d = 1$ .
- Initial conditions  $(q_0, p_0) = (1, 0)$ .
- $\gamma = \sigma = 1$ .
- $F(q) = -\sin(q)$ .
- Time interval  $[0, 1]$ .

The true solution has been computed with BUB with step size  $1/2^{11} \approx 0.0004882$ . There is no point in considering a higher number of trajectories: the statistical error is mainly present at the fifth figure after the decimal point. Assumably, all the figures up to that place are not in error as we have checked by repeating the experiment several times.

## 9.2 Weak errors

In this section we compare the performance of the methods in estimating averages. We have chosen the force to be linear so that we can compute explicitly the expectation of the random variable. To avoid cancellations, we decided to study the energy at time  $t = 1$ , i.e. we are considering the observable  $\chi(q, p) = q^2 + p^2$ . The election of the equation parameters is

---

<sup>1</sup>Obviously  $\hat{q}$  and  $\hat{p}$  are the numerical approximations.

<i>Step size</i>	ABOBA	ABOBA <i>tuned</i>	BAOAB	BAOAB <i>tuned</i>
1/2	0.105033282	0.008665816	0.107261136	0.009682680
1/4	0.053489018	0.002599530	0.053460003	0.002778036
1/8	0.026657196	0.000694208	0.026553624	0.000738185
1/16	0.013256903	0.000179242	0.013285547	0.000190850
1/32	0.006647715	0.000045654	0.006663691	0.000048492
1/64	0.003294561	0.000011479	0.003300991	0.000012101

Table 9.1: Empirical average of  $\epsilon_q$ .

<i>Step size</i>	BUB	UBU	Strang <i>tuned</i>	RK <i>tuned</i>
1/2	0.007309562	0.007871186	0.009110992	0.027105142
1/4	0.001765289	0.001936524	0.002324443	0.006010250
1/8	0.000438339	0.000482830	0.000594207	0.001407215
1/16	0.000109346	0.000120889	0.000149653	0.000339979
1/32	0.000027279	0.000030169	0.000037600	0.000083312
1/64	0.000006803	0.000007533	0.000009388	0.000020753

Table 9.2: Empirical average of  $\epsilon_q$ .

- Dimension of the system  $d = 1$ .
- Initial conditions  $(q_0, p_0) = (1, 0)$ .
- $\gamma = 1, \sigma = \sqrt{2}$ .
- $F(q) = -q$ .
- Time interval  $[0, 1]$ .

Under these assumptions, it may be checked that

$$\mathbb{E}(q^2(1) + p^2(1)) = 1.699445410\dots$$

For each numerical integrator, the value

$$\mathbb{E}(\hat{q}^2(1) + \hat{p}^2(1))$$

is computed via Monte Carlo simulation with  $N = 10^9$  trajectories. The following tables display the values

$$|\mathbb{E}(\hat{q}^2(1) + \hat{p}^2(1)) - 1.699445410|.$$

<i>Step size</i>	ABOBA	ABOBA <i>tuned</i>	BAOAB	BAOAB <i>tuned</i>
1/2	0.034474267	0.024763006	0.031870062	0.017533604
1/4	0.015333077	0.006092446	0.014909932	0.00416406
1/8	0.007361907	0.001534150	0.007247580	0.001010644
1/16	0.003600140	0.000377405	0.003592407	0.00025088
1/32	0.001793881	0.000093178	0.001803870	0.000063382
1/64	0.000890704	0.000022807	0.000894835	0.000017242

Table 9.3: Empirical average of  $\epsilon_p$ .

<i>Step size</i>	BUB	UBU	Strang <i>tuned</i>	RK <i>tuned</i>
1/2	0.013016669	0.009304839	0.01855844	0.026281795
1/4	0.003249232	0.002296745	0.004613470	0.006705204
1/8	0.000809251	0.000569178	0.001150821	0.001683484
1/16	0.000202077	0.000142744	0.000292704	0.000420755
1/32	0.000050482	0.000035462	0.000077682	0.000109683
1/64	0.000012582	0.000008921	0.000018182	0.000027892

Table 9.4: Empirical average of  $\epsilon_p$ .

The choice  $N = 10^9$  is made to ensure that the first four figures of the values featured in the tables are not particularly affected by the statistical error. However, figures from the fifth place onwards are all in error due to the sampling error.

<i>Step size</i>	ABOBA	ABOBA <i>tuned</i>	BAOAB	BAOAB <i>tuned</i>
1/2	0.080533235	0.010125751	0.080762794	0.009964294
1/4	0.040466527	0.002568672	0.040935924	0.002489252
1/8	0.020129031	0.00062932	0.020046754	0.000620057
1/16	0.009933918	0.000156952	0.010028965	0.000152128
1/32	0.005028880	0.000036689	0.005043769	0.000035704
1/64	0.002507090	0.000010461	0.002509568	0.000010595

Table 9.5: Standard deviation of  $\epsilon_q$ .

<i>Step size</i>	BUB	UBU	Strang <i>tuned</i>	RK <i>tuned</i>
1/2	0.004334722	0.003211787	0.028025632	0.031267760
1/4	0.001047738	0.000809220	0.006687259	0.007277462
1/8	0.000259770	0.000200917	0.001626152	0.001776607
1/16	0.000065066	0.000050504	0.000403758	0.000429562
1/32	0.000016163	0.000012664	0.000103711	0.000110384
1/64	0.000004047	0.000003137	0.000029124	0.000030591

Table 9.6: Standard deviation of  $\epsilon_q$ .

<i>Step size</i>	ABOBA	ABOBA <i>tuned</i>	BAOAB	BAOAB <i>tuned</i>
1/2	0.028777761	0.019560064	0.025432414	0.014075727
1/4	0.012565886	0.004820640	0.012301486	0.003371973
1/8	0.006047904	0.001221532	0.005928355	0.000797321
1/16	0.002947320	0.000299791	0.002952129	0.000196959
1/32	0.001460540	0.000073821	0.001496821	0.000050608
1/64	0.000737329	0.000017300	0.000741493	0.000013807

Table 9.7: Standard deviation of  $\epsilon_p$ .

<i>Step size</i>	BUB	UBU	Strang <i>tuned</i>	RK <i>tuned</i>
1/2	0.009781001	0.007751334	0.015652292	0.019177075
1/4	0.002441522	0.001884217	0.003747171	0.004869095
1/8	0.000608338	0.000470876	0.000932822	0.001232376
1/16	0.000152156	0.000116078	0.000238682	0.000309125
1/32	0.000037844	0.000029228	0.000062328	0.000082155
1/64	0.000009499	0.000007370	0.000020126	0.000025258

Table 9.8: Standard deviation of  $\epsilon_p$ .

<i>Step size</i>	ABOBA	ABOBA <i>tuned</i>	BAOAB	BAOAB <i>tuned</i>
1/2	0.056627501	0.077199282	0.071031765	0.051553592
1/4	0.013453604	0.019233463	0.017594870	0.011891240
1/8	0.003315052	0.004903162	0.004356423	0.002936432
1/16	0.000808485	0.001228953	0.001194216	0.000779547
1/32	0.000170813	0.000344013	0.000339452	0.000171061
1/64	0.000095971	0.000126470	0.000064133	0.000046770

Table 9.9: Weak errors in computing  $\mathbb{E}(p^2(1) + q^2(1))$  with  $N = 10^9$  samples.

<i>Step size</i>	BUB	UBU	Strang <i>tuned</i>	RK <i>tuned</i>
1/2	0.026667008	0.011460287	0.045399112	0.030197539
1/4	0.007133881	0.002751709	0.006842759	0.004761855
1/8	0.001870518	0.000706971	0.001198350	0.001133691
1/16	0.000465679	0.000243288	0.000275714	0.000248311
1/32	0.000178172	0.000006182	0.000086111	0.000020176
1/64	0.000007012	0.000005378	0.000051052	0.000052459

Table 9.10: Weak errors in computing  $\mathbb{E}(p^2(1) + q^2(1))$  with  $N = 10^9$  samples.

## Further developments

- **Modified equations.** In some cases, strong modified equations, in tandem with Theorems 2.4 and 5.5 yield *weak modified equations*. Weak modified equations may be used to obtain invariant densities. This is the case of BUB, whose invariant density may be written down explicitly up to terms of order  $O(h^4)$ .
- **Langevin Dynamics with non constant parameters.** There are many generalizations of Langevin dynamics. One such extension of particular importance assumes that both  $\gamma$  and  $\sigma$  are functions of the position  $q$ . The same analysis may also be carried out there. Moreover, it turns out to be that the tuning procedure also enables the construction of methods with *strong order 2*.
- **Word series also enable to construct split-step integrators of arbitrarily high weak order for Stratonovich and Ito SDE's.** Despite the high order they attain, these integrators are no longer consistent in the sense that there is a lack of pathwise representation of the SDE. This kind of integrators could be described as integrators that generate random variables whose moments coincide up to a desired order all with those of the true solution.





# Bibliography

- [1] M. AINSWORTH, J. LEVESLEY, M. MARLETTA, *The Graduate Student's Guide to Numerical Analysis'98*, Springer, 1999.
- [2] A. ÁLAMO, J.M. SANZ-SERNA A technique for studying strong and weak local errors of splitting stochastic integrators, *SIAM J. Numer. Anal.* 54, (2016), 3239–3257.
- [3] A. L. ARAÚJO, A. MURUA, AND J.M. SANZ-SERNA, *Symplectic methods based on decompositions*, *SIAM J. Numer. Anal.*, 34 (1997), pp. 1926–1947.
- [4] S. ASMUSSEN AND P. W. GLYNN, *Stochastic simulation: Algorithms and analysis*, vol. 57, Springer Science and Business Media, 2007.
- [5] F. BAUDOIN, *Diffusion Processes and Stochastic Calculus*, European Mathematical Society, Textbooks in Mathematics, 2014, Vol. 16.
- [6] F. BAUDOIN, *An Introduction to the Geometry of Stochastic Flows*, Imperial College Press, 2004.
- [7] S. BLANES, F. CASAS, AND A. MURUA, *Splitting and composition methods in the numerical integration of differential equations*, *Bol. Soc. Esp. Mat. Apl. SeMA*, 45 (2008), pp. 89–145.
- [8] S. BLANES, F. CASAS, A. FARRÉS, J. LASKAR, J. MAKAZAGA, AND A. MURUA, *New families of symplectic splitting methods for numerical integration in dynamical astronomy*, *Appl. Numer. Math.*, 68 (2013), pp. 58–72.
- [9] N. BOU-RABEE, A. DONEV AND E. VANDEN- EIJNDEN, *Metropolis integration schemes for self-adjoint diffusions* *SIAM J. Multiscale M. and Simulation*, 12 (2014), pp. 781–831.

- [10] N. BOU-RABEE AND H. OWHADI, *Long-run accuracy of variational integrators in the stochastic context*, SIAM J. Numer. Anal., 48 (2010), pp. 278-297.
- [11] N. BOU-RABEE AND E. VANDEN-EIJNDEN, *Pathwise accuracy and ergodicity of metropolized integrators for SDE's*, Pure and applied mathematics, 63 (2010), pp. 655-696.
- [12] K. BURRAGE AND P. M. BURRAGE, *Order conditions of stochastic Runge-Kutta methods by B-series*, SIAM J. Numer. Anal., 38 (2000), pp. 1626–1646.
- [13] J. C. BUTCHER, *Coefficients for the study of Runge-Kutta integration processes*, J. Austral. Math. Soc., 3 (1963), pp. 185–201.
- [14] M. P. CALVO, A. MURUA, AND J. M. SANZ-SERNA, *Modified equations for ODEs*, in Chaotic Numerics, P. E. Kloeden and K. J. Palmer eds., Contemporary Mathematics, Vol. 172, American Mathematical Society, Providence, (1994), pp. 63–74.
- [15] M. P. CALVO AND J. M. SANZ-SERNA, *Canonical B-series*, Numer. Math., 67, (1994), pp. 161–175.
- [16] P. CHARTIER, A. MURUA, AND J.M. SANZ-SERNA, *Higher-Order averaging, formal series and numerical integration I: B-series*, Found. Comput. Math., 10 (2010), pp. 695–727.
- [17] P. CHARTIER, A. MURUA, AND J.M. SANZ-SERNA, *Higher-Order averaging, formal series and numerical integration II: the quasi-periodic case*, Found. Comput. Math., 12 (2012), pp. 471-508.
- [18] P. CHARTIER, A. MURUA, AND J.M. SANZ-SERNA, *A formal series approach to averaging: exponentially small error estimates*, DCDS A, 32 (2012), pp. 3009-3027.
- [19] P. CHARTIER, A. MURUA, AND J.M. SANZ-SERNA, *Higher-Order averaging, formal series and numerical integration III: Error bounds*, Found. Comput. Math., 15 (2015), pp. 591–612.
- [20] K. T. CHEN, *Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula*, Ann. of Math., 65 (1957), pp. 163–178.

- [21] K. EBRAHIMI-FARD, A. LUNDERVOLD, S. J. A. MALHAM, H. MUNTHE-KAAS, AND A. WIESE, *Algebraic structure of stochastic expansions and efficient simulation*, Proc. R. Soc. A, 468 (2012), pp. 2361–2382.
- [22] J. ÉCALLE, *Les fonctions résurgents*, Publ. Math. d’Orsay. Vol 1, 81-05, Vol 2, 81-06, Vol. 3 85-05, (1981), (1985).
- [23] G. S. FISHMAN, *Monte Carlo: Concepts, Algorithms, and Applications*, Springer, 1995.
- [24] D. FRENKEL AND B. SMIT. *Understanding Molecular Simulation: From Algorithms to Applications*, Second Edition. Academic Press, 2001.
- [25] M. FLIESS, *Fonctionnelles causales non-linéaires et indéterminées non-commutatives*, Bull. Soc. Math. France, 109 (1981), pp. 3–40.
- [26] J. G. GAINES, *The algebra of iterated stochastic integrals*, Stochastics, 49 (1994), pp. 169–179.
- [27] E. HAIRER, CH. LUBICH, AND G. WANNER, *Geometric Numerical Integration, 2nd ed.*, Springer, Berlin, 2006.
- [28] E. HAIRER AND G. WANNER, *On the Butcher group and general multi-value methods*, Computing 13 (1974), pp. 1–15.
- [29] M. HOFFMAN *Quasi-shuffle products*. Journal of Algebraic Combinatorics 11 (2000), pp. 49-68.
- [30] A. JONES AND B. LEIMKUHLE. *Adaptive stochastic methods for sampling driven molecular systems*. The Journal of Chemical Physics, 135(8):084125, (2011).
- [31] P. E. KLOEDEN AND E. PLATEN, *Numerical solution of stochastic differential equations*, Springer, Berlin, 1992.
- [32] B. LEIMKUHLE AND C. MATTHEWS, *Rational Construction of Stochastic Numerical Methods for Molecular Sampling*, App. Math. Res. Express, (2013), pp. 34-56.
- [33] B. LEIMKUHLE AND C. MATTHEWS, *Robust and efficient configurational molecular sampling via Langevin Dynamics*, J. Chem. Phys, 138 (2013), 174102.

- [34] B. LEIMKUHLER AND C. MATTHEWS, *Molecular Dynamics*, Springer, 2015.
- [35] X. MAO, *Stochastic differential equations and applications*, Woodhead Publishing, 2007.
- [36] G. N. MILSTEIN, *Numerical Integration of Stochastic Differential Equations*, Kluwer, Dordrecht, 1995.
- [37] G.N. MILSTEIN, *A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations*. Siam. Theory of Probability and Its Applications. 32(4), pp. 738–741. (1986).
- [38] A. MURUA, *The Hopf algebra of rooted trees, free Lie algebras and Lie series*, Found. Comput. Math., 6 (2006), pp. 387–426.
- [39] A. MURUA AND J. M. SANZ-SERNA, *Computing normal forms and formal invariants of dynamical systems by means of word series*, Nonlinear Analysis 138, (2016), pp. 326-345.
- [40] A. MURUA AND J. M. SANZ-SERNA, *Order conditions for numerical integrators obtained by composing simpler integrators*, Phil. Trans. R. Soc. Lond. A 357 (1999), pp. 1079–1100.
- [41] A. MURUA AND J. M. SANZ-SERNA, *Word series for dynamical systems and their numerical integrators*, Found. Comput. Math. 17, (2017), pp. 675-712. (arXiv1502.05528).
- [42] A. MURUA AND J. M. SANZ-SERNA, *Computing normal forms and formal invariants of dynamical systems by means of word series*, Nonlinear Anal.-Theor. (to appear) (arXiv 1510.00250).
- [43] A. MURUA AND J. M. SANZ-SERNA, *Averaging and computing normal forms with word series algorithms*, submitted (arXiv1512.03601).
- [44] A. MURUA AND J. M. SANZ-SERNA *Hopf algebra techniques to handle dynamical systems and numerical integrators*. arXiv 1702.08354. Submitted.
- [45] B. OKSENDAL, *Stochastic differential equations. an introduction with applications*, Springer, 2013.
- [46] D. E. RADFORD, *A natural ring basis for the shuffle algebra and an application to group schemes*, Journal of Algebra 58 (1979), pp. 432-454.

- [47] R. REE, *Lie elements and an algebra associated with shuffles*, Ann. of Math., 68 (1958), pp. 210–220.
- [48] C. REUTENAUER, *Free Lie Algebras*, Clarendon Press, Oxford, 1993.
- [49] J. M. SANZ-SERNA, *Geometric integration*, in The State of the Art in Numerical Analysis, I. S. Duff and G. A. Watson eds., Clarendon Press, Oxford, 1997, pp. 121–143.
- [50] J. M. SANZ-SERNA AND M. P. CALVO, *Numerical Hamiltonian Problems*, Chapman and Hall, London, 1994.
- [51] J.M. SANZ-SERNA AND A. MURUA, *Formal series and numerical integrators: some history and some new techniques*, in Proceedings of the 8th International Congress on Industrial and Applied Mathematics (ICIAM 2015), Lei Guo and Zhi-Ming eds., Higher Education, Press, Beijing, 2015, pp. 311–331.
- [52] K. ZYGALAKIS, *On the Existence and the Applications of Modified Equations for Stochastic Differential Equations*, SIAM J. Sci. Comput., 33(1), pp. 102-130.