Balancing Related Model Order Reduction Applied to Linear Controlled Evolution Equations with Lévy Noise

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von **Martin Redmann, M.Sc.**

geb. am **20.06.1987** in Halle (Saale)

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: **Prof. Dr. Peter Benner**

Prof. Dr. Dr. h.c. Wilfried Grecksch

eingereicht am: **23.05.2016**

Verteidigung am: **22.07.2016**

List of Publications

This thesis is partially based on the papers that are listed below.

Prepints

P. Benner and M. Redmann, *Singular Perturbation Approximation for Linear Systems with Lévy Noise,* MPIMD/15-22, Max Planck Institute Magdeburg, 2015 (submitted).

Section 6.4 is based on this preprint. However, the results in Section 6.4 are more general.

P. Benner and M. Redmann, *An $\mathcal{H}_2$-Type Error Bound for Balancing-Related Model Order Reduction of Linear Systems with Lévy Noise,* MPIMD/15-20, Max Planck Institute Magdeburg, 2015 (submitted).

This preprint contains the results of Subsection 6.3.2, where a more general framework is covered.

Publications

P. Benner, T. Damm, M. Redmann, and Y. R. Rodriguez Cruz, *Positive Operators and Stable Truncation,* Linear Algebra and its Applications, 498:74–87, 2016.

Parts of Section 6.2 are based on this paper.

P. Benner and M. Redmann, *Approximation and Model Order Reduction for Second Order Systems with Lévy-Noise,* AIMS Proceedings, 945–953, 2015.

Results of this paper enter in Sections 5.2 and 6.2.

P. Benner and M. Redmann, *Model Reduction for Stochastic Systems,* Stochastic Partial Differential Equations: Analysis and Computations, 3(3):291–338, 2015.

Sections 5.1, 6.1 and 6.2 are based on this paper. However, in this thesis the results are explained more detailed.

P. Benner and M. Redmann, *Reachability and Observability Concepts for Stochastic Systems,* Proc. Appl. Math. Mech., 13:381–382, 2013.

Section 6.1 covers this article in a more general framework.

Acknowledgements

An dieser Stelle möchte ich mich bei allen, die mich während der Promotion und darüber hinaus begleitet haben, bedanken. Das Verfassen dieser Dissertation wäre ohne die Unterstützung von Familie, Freunden, Kollegen und akademischen Mentoren nicht möglich gewesen.

Zuallererst möchte ich mich bei meinem Betreuer, Professor Peter Benner, bedanken, für dessen Glauben an dieses Dissertationsthema an der Schnittstelle von Stochastik und Numerik, welcher mir erst das Promovieren ermöglicht hat. Ich bin ihm sehr dankbar dafür, dass ich an dieser interessanten Thematik arbeiten durfte und für die Gelegenheit ein Teil seiner Arbeitsgruppe am Max-Planck-Institut in Magdeburg zu sein. Unter seiner Betreuung habe ich viel Neues dazu gelernt und ich habe die Freiheiten, die mir von ihm beim Forschen gewährt worden, sehr geschätzt. Dies trug maßgeblich zum Gelingen dieser Arbeit bei.

Ebenso großer Dank gebührt Professor Wilfried Grecksch. Es war ein großes Privileg so viele Jahre sein Schüler sein zu dürfen. Er ist mir seit dem Grundstudium stets ein Ansprechpartner für jegliche Art von Problemen gewesen, und hat mir fachlich und auch in Krisenzeiten immer mit viel Engagement weiter geholfen. Sein Einsatz und sein Glauben an mich haben sehr zu meiner Motivation und zur erfolgreichen Erstellung dieser Arbeit beigetragen.

Ich bin dankbar für sehr fruchtbare wissenschaftliche Diskussionen mit Kollegen, insbesondere mit Professor Tobias Damm, der mir immer mit sehr viel Interesse und Hilfsbereitschaft zur Seite stand und großen Anteil am erfolgreichen Verfassen dieser Arbeit hat.

Des Weiteren möchte ich all meinen Kollegen an der Martin-Luther-Universität in Halle und in der Arbeitsgruppe “Computational Methods in Systems and Control Theory” am MPI in Magdeburg danken. Sie haben dazu beigetragen, dass ich mich nicht nur fachlich, sondern auch menschlich weiterentwickelt habe. Im Besonderen danke ich Frank Wusterhausen für viele interessante und Erfolg bringende Diskussionen. Tobias Breiten, Sara Grundel und Martin Stoll danke ich dafür, dass sie mir zu einem guten Start in Magdeburg verholfen haben und für ihren Beitrag zu einer positiven Arbeitsatmosphäre. Danke auch an Martin Köhler und Jens Saak, die mir beide stets unkompliziert und schnell bei IT Problemen ausgeholfen haben. Ich danke ebenso meinen

Bürokollegen Pawan Goyal, Hamdullah Yücel und Jessica Bosch. Mit ihnen hat das Arbeiten stets Freude bereitet. Ich habe in vielerlei Hinsicht von ihnen gelernt. Ebenso bin ich dankbar für die vielen tollen Momente abseits der Arbeit.

Mein abschließender Dank gilt meiner Familie und sehr guten Freunden. Ich danke Nils Demant und Alexander Oettel dafür immer als treue Freunde für mich da zu sein. Für die gemeinsame Zeit danke ich Franziska Deutschmann, die mich nachhaltig als Mensch geprägt hat und die mich immer zu Höchstleistungen in Studium und Promotion motiviert hat. Ich danke Melina Freitag für all ihre Unterstützung, welche mir auch bezüglich der Dissertation und beim weiteren Arbeiten in der Forschung viel Kraft gegeben hat. Meinen Großeltern, meinen Eltern und meinen Brüdern danke ich, dass sie immer und bedingungslos für mich da gewesen sind. Liebe Familie, ihr seid die Besten! Diese Dissertation widme ich euch.

Abstract

In this thesis, we study numerical solutions to linear controlled stochastic (partial) differential equations (S(P)DEs) with Lévy noise. We summarize well-known results in the theory of infinite dimensional Lévy processes and introduce the corresponding stochastic calculus. Then, linear controlled SPDEs in an evolution equation framework are discussed and analyzed. We emphasize the stochastic heat and damped wave equation and are particularly interested in numerical solutions of these SPDEs. To this end, we investigate a Galerkin scheme which is a useful tool to discretize an SPDE in the spatial component. Using the already existing results in the deterministic and Wiener noise setting, a Galerkin method for linear heat and damped wave equations with Lévy noise is established. Since the resulting semi-discretized SPDEs might be of large order, we study model order reduction techniques of ordinary systems with Lévy noise in order to reduce the large dimensions with the goal of saving computational time in mind. Based on the theory for deterministic linear systems, the main focus in this thesis is on generalizing balancing related model order reduction schemes to a stochastic setting. Methods like balanced truncation and the singular perturbation approximation are discussed and properties of the corresponding reduced order models, such as error bounds and preservation of stability, are investigated. The efficiency of the two approaches is demonstrated by conducting several numerical experiments.

Zusammenfassung

In der vorliegenden Arbeit wird die numerische Lösung von gesteuerten stochastischen (partiellen) Differentialgleichungen (S(P)DEs) mit Lévy-Rauschen behandelt. Wir fassen bekannte Resultate auf dem Gebiet der unendlichdimenionalen Lévy-Prozesse zusammen und liefern die Grundlagen zur dazugehörigen stochastischen Integrationstheorie. Darauf aufbauend führen wir lineare gesteuerte SPDEs in Form abstrakter Evolutionsgleichungen ein und analysieren diese. Dabei legen wir insbesondere den Fokus auf stochastische Wärmeleitungs- und gedämpfte Wellengleichungen, welche wir numerisch approximieren. Wir untersuchen die Galerkin Methode, welche genutzt werden kann um SPDEs im Ort zu diskretisieren. Auf der Grundlage von bereits bestehenden Resultaten für deterministische Gleichungen und für SPDEs mit Wiener-Rauschen, findet die Galerkin Methode hier Anwendung bei linearen Wärmeleitungs- und linearen gedämpften Wellengleichungen mit Lévy-Rauschen. Da die resultierenden semi-diskretisierten SPDEs von großer Ordnung sein können, betrachten wir Modellreduktionsverfahren für gewöhnliche Systeme mit Lévy-Rauschen um die große Dimension zu reduzieren mit dem Ziel Rechenzeit zu sparen. Auf der Theorie für deterministische linear Systeme aufbauend werden in dieser Arbeit Modellreduktionsverfahren für SDEs verallgemeinert, bei denen die zugrunde liegenden Systeme balanciert werden. Methoden wie das balancierte Abschneiden und die singulär gestörte Approximation werden diskutiert und die Eigenschaften der dazugehörigen reduzierten Modelle, wie Fehlerschranken und die Stabilitätserhaltung, untersucht. Die Effizienz der beiden Ansätze demonstrieren wir durch das Durchführen von numerischen Experimenten.

Contents

Nomenclature **x**

List of Figures **xiii**

1. Introduction **1**

1.1. Motivation for stochastic systems 1
1.2. Numerical approximations of stochastic systems and balancing related model order reduction 2
1.3. Outline of the thesis 4

2. Lévy Processes in Hilbert Spaces **10**

2.1. Definition and basic properties of Lévy processes in Hilbert spaces 10
2.2. Examples 12
2.2.1. Poisson process 13
2.2.2. Compound Poisson process 14
2.2.3. Wiener processes in Hilbert spaces 18
2.3. Jumps and jump measures of Lévy processes 23
2.4. Lévy-Khinchin decomposition 26
2.5. Square integrable Lévy processes 35
2.6. Conclusions 38

3. Stochastic Integration **39**

3.1. Stochastic integrals with respect to Lévy martingales 39
3.2. Stochastic integrals with respect to general Lévy processes 52
3.3. Conclusions 54

4. Linear Controlled SPDEs with Lévy Noise **55**

4.1. C_0-semigroups 55

4.2. Linear controlled evolution equations with Lévy noise . . . 57
4.3. Examples . . . 63
4.3.1. Stochastic heat equation . . . 64
4.3.2. Stochastic damped wave equation . . . 66
4.4. Conclusions . . . 69

5. Numerical Approximations for Linear SPDEs with Lévy Noise 70
5.1. Approximation of stochastic heat equations . . . 71
5.2. Approximation of stochastic damped wave equations . . . 79
5.3. Conclusions . . . 87

6. Linear Ordinary SDEs with Lévy Noise and Balancing Related Model Order Reduction 88
6.1. Linear control with Lévy noise . . . 88
6.1.1. Reachability concept . . . 89
6.1.2. Observability concept . . . 104
6.2. Type 1 balanced truncation for stochastic systems . . . 107
6.2.1. Procedure . . . 107
6.2.2. Stability analysis of the ROM . . . 110
6.2.3. Further properties of the ROM . . . 119
6.2.4. Error bound for type 1 balanced truncation . . . 120
6.2.5. Numerical experiments . . . 129
6.3. Type 2 balanced truncation for stochastic systems . . . 137
6.3.1. Comparison with type 1 BT, procedure and properties of the ROM . . . 138
6.3.2. $\mathscr{H}_2$ error bound for type 2 balanced truncation . . . 141
6.4. Singular perturbation approximation for stochastic systems . . . 145
6.4.1. Procedure . . . 146
6.4.2. Preservation of (asymptotic) mean square stability . . . 150
6.4.3. Other properties in the ROM . . . 152
6.4.4. Error bound for the SPA . . . 156
6.4.5. Numerical experiments . . . 161
6.5. Conclusions . . . 168

7. Conclusions 169
7.1. Summary . . . 169

7.2. Outlook 170

A. Appendix **172**

A.1. Semimartingales and Ito's formula in $\mathbb{R}^d$ 172

A.2. Lévy-type integrals 174

Bibliography **177**

Nomenclature

List of Acronyms

BT	balanced truncation
LTS	long-term support
MATLAB	software from The MathWorks Inc.
MOR	model order reduction
ODE	ordinary differential equation
PDE	partial differential equation
Pentium	processor series from Intel
RAM	random-access memory
ROM	reduced order model
SDE	stochastic differential equation
SPA	singular perturbation approximation
SPDE	stochastic partial differential equation

List of Symbols

$\mathbb{N}$	natural numbers $\{1,2,3,\dots\}$
$\mathbb{Z}_+$	$\mathbb{N}\cup\{0\}$
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	non-negative real numbers
$\mathbb{R}^{m\times n}$	vector space of real matrices with m rows and n columns
$\mathbb{R}^n$	$\mathbb{R}^{n\times 1}$
$\mathbb{C}$	complex numbers
$\mathbb{C}_-$	open left half of the complex plane
A^T	transpose of a matrix A
A^*	adjoint operator of A
A^{-1}	inverse of a matrix or an operator A
$A^\#$	Moore-Penrose pseudoinverse of a matrix A

$A^{\frac{1}{2}}$	square root of a matrix or an operator A
$A > 0$	symmetric positive definite matrix A
$\operatorname{diag}(a_1, a_2, \dots, a_n)$	diagonal matrix with the diagonal entries $a_1, a_2, \dots, a_n$
$D(\cdot)$	domain of an operator
$\operatorname{im}(\cdot)$	image of a matrix or an operator; space spanned by the columns (matrix case)
I_n, I	n-dimensional identity matrix or identity matrix of suitable dimension
$\ker(\cdot)$	kernel of a matrix or an operator; inverse image of zero
$\tilde{A} \otimes \hat{A}$	Kronecker product of matrices $\tilde{A}$ and $\hat{A}$
$\operatorname{span}\{x_1, x_2, \dots, x_n\}$	space spanned by the vectors $x_1, x_2, \dots, x_n$
$\sigma(A)$	spectrum of a matrix or an operator A
$\operatorname{tr}(\cdot)$	trace of a matrix or an operator; sum of the diagonal entries (matrix case)
$\operatorname{vec}(\cdot)$	vectorization of a matrix
$\mathscr{B}(\cdot)$	Borel σ-algebra
$\sigma(S)$	the smallest σ-algebra containing the subset S of a power set
$\mathscr{P}_T$	predictable σ-algebra
$\mathscr{A}_1 \otimes \mathscr{A}_2$	product σ-algebra consisting of the σ-algebras $\mathscr{A}_1$, $\mathscr{A}_2$; $\sigma(\{A_1 \times A_2 : A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2\})$
$\mu_1 * \mu_2$	convolution of the measures μ_1 and μ_2
$\mu_1 \otimes \mu_2$	product measure consisting of the measures μ_1 and μ_2
$(\Omega, \mathscr{F}, \mathbb{P})$	probability space; measure space with $\mathbb{P}(\Omega) = 1$
$(\mathscr{F}_t)_{t \geq 0}$	filtration; increasing family of sub-σ-algebras of $\mathscr{F}$; $\mathscr{F}_{t_1} \subseteq \mathscr{F}_{t_2}$ for $t_1 \leq t_2$
$\mathbb{E}[\cdot]$	expected value
$[\cdot, \cdot]_t$	quadratic covariation process
$\operatorname{rand}(\cdot, \cdot), \operatorname{randn}(\cdot, \cdot)$	matrix of uniformly or normally distributed random numbers
$L(U, H)$	space of linear and bounded operators from U to H
$L(H)$	$L(U, H)$ with $U = H$
$L_{(HS)}(U, H)$	space of Hilbert-Schmidt operators from U to H
$L_1^+(U)$	space of symmetric non-negative trace class operators on U
$L^2(X, \mathscr{A}, \mu)$	space of square integrable functions with respect to the measure μ defined on the measurable space $(X, \mathscr{A})$
$L^2(K)$	$L^2(K, \mathscr{B}(K), dt)$ with $K \subseteq \mathbb{R}^n$
L^2_T	space of $(\mathscr{F}_t)_{t \geq 0}$-adapted processes in $L^2(\Omega \times [0, T], \mathscr{F} \otimes \mathscr{B}([0, T]), \mathbb{P} \otimes dt)$
L^2	L^2_T, where $[0, T]$ is replaced by $\mathbb{R}_+$

$\mathscr{L}_T^2$	$L^2(\Omega \times [0,T], \mathscr{P}_T, \mathbb{P} \otimes dt)$, space of integrands with respect to a square integral Lévy process with mean zero
$\mathscr{M}^2$	space of square integrable martingales
$\mathscr{M}_L^2$	subspace of Lévy processes in $\mathscr{M}^2$
$R^\perp$	orthogonal complement of a subspace R
$\mathscr{S}$	space of simple processes
$\overline{S}$	closure of a set S
$1_S(\cdot), \chi_S(\cdot)$	indicator function of a set S
$\lVert\cdot\rVert_B$	norm in a Banach space B
$\langle\cdot,\cdot\rangle_H$	inner product in a Hilbert space H
$\lVert\cdot\rVert_2, \langle\cdot,\cdot\rangle_2$	Euclidean norm and inner product
$\lVert\cdot\rVert_F, \langle\cdot,\cdot\rangle_F$	Frobenius norm and inner product

List of Figures

2.1. Trajectories of scalar Lévy processes 13

5.1. The Galerkin solution to the stochastic heat equation in Example 4.14 80
5.2. The output of the stochastic heat equation in Example 4.14 81
5.3. The Galerkin solution to the stochastic damped wave equation in Example 4.16 . 85
5.4. The output of the stochastic damped wave equation in Example 4.16 86

6.1. Average temperature non-heated area compared with reduced order output by BT 133
6.2. Output of the original model compared with the output of the ROM by BT 136
6.3. Output of the original model compared with the output of the ROM by SPA on the time interval $[0,\pi]$ 165
6.4. Comparison of BT and SPA using a particular trajectory 167

1. Introduction

1.1. Motivation for stochastic systems

Many phenomena in real life can be described by ordinary differential equations (ODEs), partial differential equations (PDEs), or both. Famous examples are the motion of viscous fluids, the description of water or sound waves and the distribution of heat. For an accurate mathematical modeling of these real world applications it is often required to take random effects into account. An example is the financial market, where stock prices and interest rates are randomly impacted. Furthermore, there are many phenomena in natural sciences containing uncertainties such as the growth of a population or the movement of particles. Systems that are influenced by wind such as the surface of a lake or a river might also not just follow deterministic laws. The expansion of heat delivers another example, where stochastic components could enter, compare Examples 4.14 and 5.3.

Uncertainties in an ODE or PDE model can for example be represented by an additional noise term. This leads to stochastic differential equations (SDEs) or stochastic PDEs (SPDEs). A possible way is to consider equations driven by Wiener noise. We refer to Arnold [6]; Kloeden, Platen [41] and Kuo [47], where Wiener processes, stochastic integrals and SDEs with Wiener noise are studied. In Da Prato, Zabczyk [20]; Gawarecki, Mandrekar [26] and in Prévôt and Röckner [56] they treat infinite dimensional Wiener processes as well as Wiener driven SPDEs. Dealing with Wiener noise yields just continuous systems. This has the disadvantage of not covering models with jump. Therefore, many financial products cannot be described since jumps are required to model risky stocks, see Madan et al. [49] or Madan, Seneta [50]. Moreover, the prices of electricity have discontinuities which is shown in A. Veraart, L. Veraart [69]. Another example from natural sciences is the surface of a river with waterfalls. The same holds true for phenomena including effects of wind since jumps appear there as well, compare Examples 4.16 and 5.5. Lévy processes, which in general are not continuous, provide a possible solution to this problem. One can find detailed information regarding Lévy processes in finite dimensional spaces in Bertoin [15] and Sato [65]. SDEs with Lévy noise with the corresponding integration theory are well investigated in Applebaum [5]. Peszat, Zabczyk [55] extend the work of [20, 26, 56]. They provide a comprehensive book containing the stochastic analysis of infinite dimensional Lévy processes and the theory of Lévy driven SPDEs with various examples.

1.2. Numerical approximations of stochastic systems and balancing related model order reduction

It is necessary to discretize a time-dependent PDE in space and time in order to solve it numerically. As a possible strategy discretizing in space can be considered as a first step. This can be done by finite element methods that are for example explained in Thomée [68] for the parabolic case. To ensure a high accuracy of the approximation this can lead to highly complex systems of ODEs in terms of the large state space dimension. Solving such complex ODE systems causes large computational costs which are aimed to be reduced. In this regard, model order reduction (MOR) becomes a key ingredient. MOR originates in the field of deterministic control theory and is used to save computational time by replacing large scale systems by systems of low order in which the main information of the original system should be captured. A particular class of MOR schemes is called balancing related MOR. They are based on reachability and observability concepts and corresponding energy functionals. Now the idea of balancing a system is to create a system, where the dominant reachable and observable states are the same. Then, the difficult to observe and difficult to reach states (states producing the least observation energy and causing the most energy to reach, respectively) are neglected. A famous representative of this class is balanced truncation (BT) which assumes asymptotic stability of the original system. This was considered first in Moore [53] for linear deterministic system; see Antoulas [2] or Obinata, Anderson [54] for a thorough treatment of the topic. BT was also established for deterministic bilinear systems in Benner, Damm [9] and Zhang et al. [72]. An alternative method to obtain a reduced order model (ROM) is the singular perturbation approximation (SPA), see Liu, Anderson [48] and Fernando, Nicholson [24] for deterministic linear systems. Rather than setting all truncated states to zero as in BT (compare equation (6.24), where the truncated states are represented by x_1), they are assumed constant which allows to solve for them and thus include this information in the differential equation for the remaining states. This has the advantage of a zero steady-state error, a property often important in applications. The SPA also exists for bilinear systems. For that framework, we refer to Hartmann et al. [28].

As mentioned above, in many situations a deterministic perspective on a problem is not satisfactory since this approach might neglect random disturbances. So, rather than studying PDE models it can be more meaningful to consider SPDEs instead to obtain a more accurate model. By numerical approximations, an SPDE can be reduced to a finite dimensional system as well. A possibility to do that is the spectral Galerkin method which is for example investigated in Grecksch, Kloeden [27]; Hausenblas [29]; Jentzen, Kloeden [40]; Blömker, Jentzen [17] for Wiener driven systems and Redmann, Benner [60] for equations with Lévy noise. Alternatively, finite element methods can be applied. Based on [68], Kruse investigates this scheme in [45, 46] for SPDEs with Wiener noise. Barth [7] and Barth, Lang [8] consider finite element approaches for equations with more general noise processes such as Lévy processes. By semi-discretizing we reduce an SPDE to an SDE which, similar as in the deterministic framework, might be large dimensional. For that reason, generalizing MOR techniques to stochastic systems can easily be motivated and is taken into account. Inspired by the application in the field of numerical solutions to SPDEs, two types of BT and the SPA are extended to stochastic systems. To this end, balanced truncation is considered first for SDEs with Wiener noise in Benner, Damm [9] and for systems with Lévy noise by Benner, Redmann in [14]. Additionally, Benner and Redmann pointed out the benefit of BT in detail by applying it to Lévy driven damped wave and heat equations, see [14] and [60]. A second way to generalize BT to stochastic systems is discussed in Benner et al. [10]; Benner, Damm [22] and Redmann, Benner [59]. This new approach, the so-called type 2 BT, is motivated by the aim of achieving a certain error bound which cannot be proven in the ansatz used in [14]. Moreover, Redmann and Benner [61] studied the SPA for SDEs with Lévy noise and successfully applied it to stochastic damped wave equations. From numerical experiments it could be concluded that SPA can be better than BT when using small controls or when fixing a large time interval for the underlying SDE.

These generalized MOR techniques also play a role in this thesis. In particular, this dissertation contains results from [14, 59, 60, 61]. Furthermore, approaches studied in [9, 10, 22] are frequently used here as well.

After reducing the state space dimension of a spatially-discretized SPDE, we desire to discretize the resulting low order SDE in time. The most common method is the so-called Euler-Maruyama scheme. Explicit and (semi-)implicit approaches for SDEs with Wiener noise are discussed in [16, 33, 34, 35, 36, 37, 41, 51] and methods with additional Poisson noise are considered in [25, 31, 32]. As in the deterministic case, the choice of the time step size for explicit schemes strongly depends on the dimension of the underlying SDE. So, applying explicit methods to large scale stochastic systems is not efficient in terms of the computational costs. Since the order of the system is already reduced to a low dimension by MOR, explicit Euler-Maruyama methods can be considered as well without causing too much computational cost.

1.3. Outline of the thesis

In order to render this thesis as self-contained as possible, we summarize the most important results for Lévy processes and the corresponding integration theory in Chapters 2 and 3. Chapter 2 contains definitions and results mainly from Albeverio, Rüdiger [1]; [5, 55] and Redmann [58]. We start with basic definitions and properties of Lévy processes taking values in Hilbert spaces U before we discuss particular examples as Poisson, (compensated) compound Poisson and Wiener processes. They are also the main ingredients to describe a general Lévy process $L = (L(t))_{t\geq 0}$ which, as will be detailed in Section 2.4, is represented by the sum of independent processes

$$L(t) = at + W(t) + J_1(t) + J_2(t), \quad t \geq 0, \tag{1.1}$$

where $a \in U$. W denotes a Wiener process which, apart from a linear drift, is the only continuous Lévy process. Since W is a Gaussian process, it is square integrable as well. $J_1 + J_2$ represents the discontinuous part of L. J_1 covers the jumps of L smaller than a constant $r_0 > 0$ (with respect to the norm in U), where there might be infinitely many. It is defined as the limit of a sequence of compensated compound Poisson processes. Moreover, J_1 has existing moments of arbitrary order since it is a Lévy process with bounded jumps. The remaining part J_2 turns out to be a compound Poisson process which is a process with finite jump activity and piece-wise constant paths. It contains the large jumps of L which roughly speaking are the ones greater or equal than r_0. Because the jumps of J_2 are not bounded, the corresponding moments need not exist. So, the existence of moments of L depends on the existence of moments corresponding to J_2.

Studying SPDEs with Lévy noise requires the definition of a stochastic integral of the form $\int_0^t \Psi(s)dL(s)$, $t \geq 0$, where Ψ is an operator-valued stochastic process. In order to define this kind of integral, the Lévy-Khinchin decomposition in (1.1) becomes a key tool. In Chapter 3, we define an integral with respect to L by introducing an integral for every summand of the the decomposition. For the linear drift, the definition of a (path-wise) Bochner integral is needed, see Denk [23]. Since W and J_1 are square integrable mean zero processes, we can cover this part by the definition of

$$\int_0^t \Psi(s)dM(s), \quad t \geq 0, \tag{1.2}$$

in Section 3.1. Here, the process M is a square integrable Lévy process with mean zero. Due to the independent, homogeneous increments and the existing second moment there is a symmetric, non negative definite trace class operator $\mathscr{Q}$ such that

$$\mathbb{E}\left\langle M(t),x\right\rangle_U \left\langle M(t),y\right\rangle_U = t\left\langle \mathscr{Q}x,y\right\rangle_U, \quad x,y \in U,$$

which is for example proven in [55]. Having such a covariance operator $\mathscr{Q}$, the definition of (1.2) becomes similar to the Wiener case. We discuss properties of this integral such as the mean, the Ito isometry and the martingale property. For the results regarding (1.2) we intensively make use of the following references [3, 20, 26, 55, 56, 58]. Chapter 3 is concluded by the missing definition of a stochastic integral with respect to J_2. Since J_2 has piece-wise constant paths and just finitely many jumps the corresponding integral is a finite random sum, compare Applebaum [4].

In Chapter 4, we introduce an abstract controlled linear SPDE that we solve numerically in the following chapters

$$\begin{aligned} d\mathscr{X}(t) &= [\mathscr{A}\mathscr{X}(t) + \mathscr{B}u(t)]\,dt + \mathscr{N}(\mathscr{X}(t-))dM(t), \quad \mathscr{X}(0) = X_0 \in H, \\ \mathscr{Y}(t) &= \mathscr{C}\mathscr{X}(t), \quad t \geq 0. \end{aligned} \tag{1.3}$$

Above, $\mathscr{A} : D(\mathscr{A}) \to H$ is a generator of a contraction semigroup with H being a separable Hilbert space. The SPDE, driven by a square integrable Lévy process M with mean zero, is controlled by $u(t) \in \mathbb{R}^m$ and equipped with a finite dimensional output $\mathscr{Y}(t) \in \mathbb{R}^p$. The input operator $\mathscr{B} \in L(\mathbb{R}^m, H)$ and the output operator $\mathscr{C} \in L(H, \mathbb{R}^p)$ are linear and bounded operators. The same holds for the operator $\mathscr{N}$ that is defined on H and takes values in a suitable operator space. We first show that the cadlag mild solution of the abstract SPDE is well-defined before we emphasize two examples that are covered by our framework. In particular, we consider cases, where $\mathscr{A} = \Delta$ and $\mathscr{A} = \left[\begin{smallmatrix} 0 & I \\ -\Delta & -\alpha I \end{smallmatrix}\right]$ for $\alpha > 0$ are included, i.e. the stochastic heat and the stochastic damped wave equation.

In Chapter 5, we approximate the SPDE in (1.3) with a Galerkin method for the two special cases mentioned above. For the stochastic heat equation we apply the techniques used in [27, 40, 29] and generalize them to systems with Lévy noise. We moreover contribute a Galerkin scheme based on similar ideas to a second order systems, i.e. in particular to damped wave equations with Lévy noise. So, we can reduce (1.3) to a finite dimensional system of the form

$$dx(t) = [Ax(t) + Bu(t)]dt + \sum_{k=1}^{q} N^k x(t-) dM_k(t), \quad t \geq 0, \quad x(0) = x_0 \in \mathbb{R}^n, \tag{1.4}$$
$$y_n(t) = Cx(t),$$

where A, $N^k \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and M_k ($k = 1,\dots,q$) are uncorrelated square integrable scalar Lévy processes with mean zero. Note that N^k is just a notation of a matrix with index k and not the k-th matrix power. We show the convergence of the corresponding output y_n, $n \in \mathbb{N}$, to the SPDE output $\mathscr{Y}$

$$\mathbb{E}\left\|y_n(t) - \mathscr{Y}(t)\right\|_2^2 \to 0 \quad \text{ for } n \to \infty, \quad t \geq 0.$$

In order to ensure a small error in $\mathbb{E}\left\|y_n(t) - \mathscr{Y}(t)\right\|_2^2$, we fix a large n which is the state space dimension of equation (1.4). To support the theory in Chapter 5, we plot paths of the Galerkin solution of stochastic damped wave and heat equations and the corresponding output.

Choosing a large state space dimension n in (1.4) motivates the work done in Chapter 6, where we consider balancing related MOR techniques for Lévy driven systems. This kind of MOR requires the mean square asymptotic stability of (1.4), that is

$$\sigma\left(I_n \otimes A + A \otimes I_n + \sum_{k=1}^{q} N^k \otimes N^k \cdot \mathbb{E}\left[M_k^2(1)\right]\right) \subset \mathbb{C}_-.$$

The idea is to find a system

$$d\tilde{x}(t) = [A_R\tilde{x}(t) + B_R u(t)]dt + \sum_{k=1}^{q} N_R^k \tilde{x}(t-) dM_k(t),$$
$$\hat{y}(t) = C_R\tilde{x}(t)$$

with A_R, $N_R^k \in \mathbb{R}^{r\times r}$, $B_R \in \mathbb{R}^{r\times m}$, $C_R \in \mathbb{R}^{p\times r}$ and $r \ll n$ such that $\hat{y} \approx y_n$.

We start with generalized reachability and observability concepts that roughly speaking are used to characterize the importance of (average) states in system (1.4). The unimportant ones are those producing the least observation energy and causing the most energy to reach. For this characterization, we introduce Gramians $P := \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds$ and $Q := \mathbb{E}\left[\int_0^\infty \Phi^T(s)C^TC\Phi(s)ds\right]$, where the matrix-valued process Φ is the fundamental solution to (1.4). We prove that these Gramians satisfy

$$AP + PA^T + \sum_{k=1}^{q} N^k P (N^k)^T \,\mathbb{E}\left[M_k^2(1)\right] = -BB^T, \tag{1.5}$$

$$A^TQ + QA + \sum_{k=1}^{q} (N^k)^T Q N^k \,\mathbb{E}\left[M_k^2(1)\right] = -C^TC, \tag{1.6}$$

compare [9] for the Wiener case, where $\mathbb{E}\left[M_k^2(1)\right] = 1$ for $k = 1, \dots, q$. Assuming a minimal system (every state is reachable and observable), we could further show that the difficult to reach (observe, respectively) states are contained in the space spanned by the eigenvectors corresponding to the small eigenvalues of P (Q, respectively). As the next step we describe the procedure of balancing systems with Lévy noise which is similar to the one of the deterministic case [2]. Using a suitable balancing state space invertible transformation matrix $\hat{T} \in \mathbb{R}^{n\times n}$, we obtain a system with Gramians $P = Q = \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$ which guarantees that the dominant reachable and observable states are the same. We call the numbers $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ Hankel singular values. In order to then neglect the unimportant states in the balanced system, we investigate two methods namely BT and SPA for linear controlled systems with Lévy noise. Working with the partitions

$$TAT^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \; TN^kT^{-1} = \begin{bmatrix} N_{11}^k & N_{12}^k \\ N_{21}^k & N_{22}^k \end{bmatrix}, \; TB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \; CT^{-1} = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \tag{1.7}$$

the reduced order coefficients for BT are

$$(A_R, N_R^k, B_R, C_R) = (A_{11}, N_{11}^k, B_1, C_1)$$

and the matrices corresponding to the SPA look as follows:

$$(A_R, N_R^k, B_R, C_R) = (A_{11} - A_{12}A_{22}^{-1}A_{21}, N_{11}^k - N_{12}^kA_{22}^{-1}A_{21}, B_1, C_1 - C_2A_{22}^{-1}A_{21}) =: (\bar{A}, \bar{N}^k, \bar{B}, \bar{C}).$$

The properties of both ROMs are quite similar. We show that starting with a completely reachable and observable original system, this property can be lost in the ROM obtained by BT or SPA, respectively. Moreover, for both schemes the ROM is neither balanced nor the Hankel singular values coincide with the ones of the original model in general. We further prove that BT preserves mean square asymptotic stability, i.e.

$$\sigma\left(I_r\otimes A_{11}+A_{11}\otimes I_r+\sum_{k=1}^{q}N_{11}^k\otimes N_{11}^k\cdot\mathbb{E}\left[M_k^2(1)\right]\right)\subset\mathbb{C}_-,$$

where we use a more system theoretical approach compared to the proof in Benner et al. [11]. The same we conjecture for the ROM by SPA but so far we can only show that it is mean square stable, that is

$$\sigma\left(I_r\otimes\bar{A}+\bar{A}\otimes I_r+\sum_{k=1}^{q}\bar{N}^k\otimes\bar{N}^k\cdot\mathbb{E}\left[M_k^2(1)\right]\right)\subset\overline{\mathbb{C}_-},$$

where this property is slightly more general than the one proven in [61]. We also establish an error bound for both methods, where $\hat{y}_i$ $(i=BT,SPA)$ is either the reduced order output by BT or SPA below. The bound is of the form

$$\sup_{t\in[0,T]}\mathbb{E}\left\|y_n(t)-\hat{y}_i(t)\right\|_2\leq\left(\operatorname{tr}(\Sigma_2K_i)\right)^{\frac{1}{2}}\left\|u\right\|_{L_T^2},\tag{1.8}$$

where $\Sigma_2=\operatorname{diag}(\sigma_{r+1},\ldots,\sigma_n)$ with $\sigma_{r+1},\ldots,\sigma_n$ being the $n-r$ smallest Hankel singular values corresponding to the difficult to reach and observe states. Depending on the method used, we have different weighting matrices K_i $(i=BT,SPA)$. K_i mainly consists of matrices from the partition in (1.7). The less important the neglected states in terms of the concepts we use, the smaller the values $\sigma_{r+1},\ldots,\sigma_n$, the smaller the error bound in (1.8) becomes.

To provide a complete summary of the current status in balancing related MOR for stochastic systems, we moreover discuss the so-called type 2 BT for Lévy driven systems which was first considered in [10, 22] for the Wiener case. We state the results achieved therein including the error bound and stability analysis for this ansatz and contribute an $\mathscr{H}_2$-type error bound for a more general framework than in [59]. We there replace the solution of (1.5) by a new Gramian P_2 fulfilling

$$A^TP_2^{-1}+P_2^{-1}A+\sum_{k=1}^{q}(N^k)^TP_2^{-1}N^k\cdot\mathbb{E}\left[M_k^2(1)\right]\leq-P_2^{-1}BB^TP_2^{-1}.$$

In contrast to the first approach, a system is constructed, where the solution of (1.6) and now P_2 are diagonalized simultaneously, i.e. we ensure that $P_2=Q={}_2\Sigma=\operatorname{diag}({}_2\sigma_1,\ldots,{}_2\sigma_n)$. The diagonal entries of ${}_2\Sigma$ are the Hankel singular values of the type 2 BT. With this new reachability

Gramian we can still guarantee an error bound of the form (1.8), where the weighting matrix is a different one and where Σ_2 is substituted by ${}_2\Sigma_2 = \operatorname{diag}({}_2\sigma_{r+1}, \dots, {}_2\sigma_n)$.

To illustrate the efficiency of the generalized balancing related MOR schemes we run several numerical simulations in Chapter 6. We therefore apply both BT and the SPA to large scale systems which for example result from semi-discretizing SPDEs with Lévy noise. In particular, we compare trajectories of outputs corresponding to large scale systems with the ones of ROMs obtained by using the proposed MOR techniques. Furthermore, we compute the exact errors of these numerical approximations in order to show the sharpness of the error bounds which we derive as well. A numerical comparison between BT and the SPA is also provided.

In the appendix, we first briefly state information about finite dimensional semi-martingales and corresponding Ito formulas, before we discuss scalar Lévy-type integrals and their quadratic covariation.

2. Lévy Processes in Hilbert Spaces

In this chapter, we give an overview about basic properties of Lévy processes which we deal with in the following chapters. This chapter is mainly based on Peszat, Zabczyk [55] but results of Albeverio, Rüdiger [1]; Applebaum [5] and Redmann [58] also enter here.

2.1. Definition and basic properties of Lévy processes in Hilbert spaces

This section is based on Section 4.1 in the book of Peszat and Zabczyk [55], where one also finds a more detailed overview. Below, let U denote a separable Hilbert space and $\mathscr{B}(U)$ be the corresponding Borel σ-algebra.

Definition 2.1. *An U-valued stochastic process $L = (L(t))_{t\geq 0}$ has* independent increments *if the $(U, \mathscr{B}(U))$-valued random variables $L(t_1) - L(t_0), L(t_2) - L(t_1), \dots, L(t_n) - L(t_{n-1})$ are independent for arbitrary times $0 \leq t_0 < t_1 < \dots < t_n$.*
If the random variables $L(t+h) - L(t)$ and $L(s+h) - L(s)$ have the same distribution for $s, t \in \mathbb{R}_+$ and $h > 0$, then the process L has homogeneous increments.

Definition 2.2. *Suppose $L = (L(t))_{t\geq 0}$ to be a stochastic process in U with independent and homogeneous increments. If furthermore $L(0) = 0$ holds and L is continuous in probability, i.e.*

$$\lim_{s \to t} \mathbb{P}\{\|L(s) - L(t)\|_U > \varepsilon\} = 0$$

for all $t \in \mathbb{R}_+$ and $\varepsilon > 0$, then L is called Lévy process.

Definition 2.3. *Let $\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}\right)$[1] be a filtered probability space. Moreover, if L is an $(\mathscr{F}_t)_{t\geq 0}$-adapted Lévy process taking values in U and for all $t, h \geq 0$ the random variable*

[1] We assume that $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and $\mathscr{F}_0$ contains all sets A with $\mathbb{P}(A) = 0$.

$L(t+h)-L(t)$ is independent of $\mathscr{F}_t$, then L is called Lévy process with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$.

Remark. *If one considers a Lévy process in the sense of Definition 2.2 equipped with the natural filtration, then one automatically obtains the property in Definition 2.3.*

Important representatives of the class of Lévy processes are Poisson, Wiener and compound Poisson processes which we investigate in the Subsections 2.2.1, 2.2.2 and 2.2.3.
Now, let L be a U-valued Lévy process and μ_t be the distribution of $L(t)$. Furthermore, by $\mu * \nu$ we denote the convolution of the measures μ and ν.

Proposition 2.4. *The family of probability measures $(\mu_t)_{t\geq 0}$ has the following property:*

$$\mu_{s+t} = \mu_s * \mu_t.$$

Proof. The property in Proposition 2.4 is a direct consequence of the decomposition

$$L(s+t) = [L(s+t)-L(t)] + [L(t)-L(0)], \quad s,t \geq 0,$$

and the independent and homogeneous increments of L. □

For that reason, $(\mu_t)_{t\geq 0}$ is called convolution semigroup with the neutral element μ_0. Here, we interpret μ_0 as a distribution of a random variable which equals 0 with probability 1.
In fact, μ_t is an infinitely divisible measure since $L(t)$ can be represented by

$$L(t) = \sum_{n=1}^{\frac{t}{dt}} L(n\cdot dt) - L((n-1)\cdot dt),$$

where we choose the number dt such that $\frac{t}{dt} \in \mathbb{N}$. Since every summand has the same distribution as $L(dt)$, we have $\mu_t = (\mu_{dt})^{*\frac{t}{dt}}$, where $(\mu_{dt})^{*\frac{t}{dt}}$ denotes the $\frac{t}{dt}$-th convolution of the measure μ_{dt}.

Remark. *The finite dimensional distributions of a stochastic process with independent increments are characterized by the distributions of the increments of the process. Hence, a Lévy process L is determined by the family of distributions $(\mu_t)_{t\geq 0}$.*

In many cases, we need a certain property for the paths of a Lévy process. Therefore, we introduce the definition below.

Definition 2.5. *A stochastic process $L = (L(t))_{t\geq 0}$ is* cadlag *if the following conditions are fulfilled:*

- *L has right-continuous trajectories with probability 1, i.e.*

$$\mathbb{P}\left\{\lim_{s\to t+}\|L(s)-L(t)\|_U = 0, \forall t\geq 0\right\} = 1$$

- *and left limits $L(t-)$ exist, i.e.*

$$\mathbb{P}\left\{\lim_{s\to t-}\|L(s)-L(t-)\|_U = 0, \forall t\geq 0\right\} = 1.$$

We now formulate a result which contains the vital property for trajectories of Lévy processes.

Theorem 2.6. *Every Lévy process $\hat{L} = \left(\hat{L}(t)\right)_{t\geq 0}$ has a cadlag modification. Hence, there is a cadlag Lévy process $\hat{L}$ such that $\mathbb{P}\left\{L(t) = \hat{L}(t)\right\} = 1$ for all $t\geq 0$.*

Proof. This result and its proof can be found in Theorem 4.3 in [55]. □

Below, we always assume to have Lévy processes with the cadlag property, i.e. we work with the cadlag modification from now on. For that reason, the expression $\Delta L(t) := L(t) - L(t-)$ is well-defined for the Lévy process L.

Theorem 2.7. *Suppose $L = (L(t))_{t\geq 0}$ is a cadlag Lévy process with values in U and bounded jumps, i.e. $\|L(t)-L(t-)\|_U < c$ $\mathbb{P}$-a.s. for $c > 0$ and all $t\geq 0$, then for arbitrary $\beta > 0$ and $t\geq 0$ it holds that*

$$\mathbb{E}\left[\mathrm{e}^{\beta\|L(t)\|_U}\right] < \infty.$$

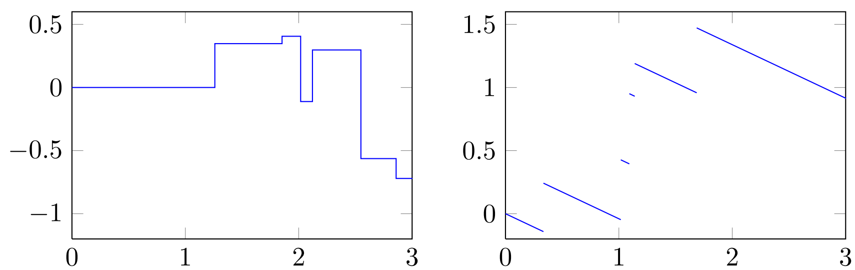
Proof. One finds this result and its proof in Theorem 4.4 in [55]. □

Thus, the existence of the moments $\mathbb{E}\|L(t)\|_U^p$, $p\in\mathbb{N}$, is guaranteed for Lévy processes with bounded jumps.

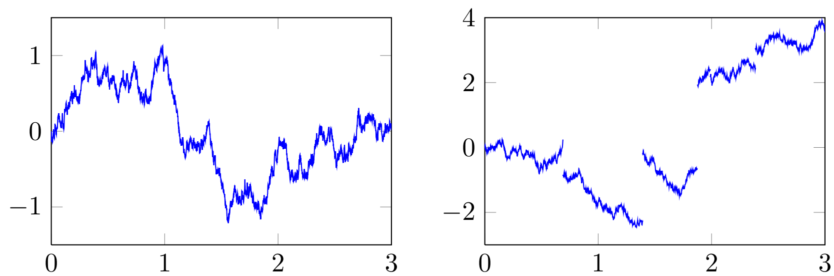
2.2. Examples

In this section, we state three important examples for the class of Lévy processes, namely Poisson, compound Poisson and Wiener processes. The first two examples are jump processes and

the Wiener process is the only continuous representative.
One can see a trajectory of a compound Poisson process with values in $\mathbb{R}$ in the first picture of Figure 2.1a. It is characterized by random jump times and random jump sizes. The times between the jumps are represented by independent and exponentially distributed random variables. The second picture of Figure 2.1a shows a scalar compensated compound Poisson process which is a compound Poisson process subtracted by its mean function. In this plot we see that the mean function is linear in case it exists. Figure 2.1b shows paths of scalar Lévy processes. The first one is a Wiener process and the second picture illustrates a process with a Wiener and a compound Poisson part such that it is piece-wise continuous with random jumps in between.



(a) A trajectory of a compound Poisson and a compensated compound Poisson process



(b) A trajectory of a Wiener process and a combination of a Wiener process and a compound Poisson process being independent

Figure 2.1.: Trajectories of scalar Lévy processes

2.2.1. Poisson process

We start with discussing the fist representative of the class of Lévy processes which is the so-called Poisson process. It is a pure jump process which takes values in $\mathbb{Z}_+$. The Poisson process plays a fundamental role in the theory of Lévy processes. It is of importance in the analysis of the jumps of Lévy processes in Section 2.3 and takes a central role in the Lévy-Khinchin decomposition which we study in Section 2.4. In this subsection, we mainly focus on basic properties of the Poisson process.

Definition 2.8. *A* Poisson process $N = (N(t))_{t\geq 0}$ with intensity $a < \infty$ *is a Lévy process for which the random variable $N(t)$ has a Poisson distribution with parameter at for $t \geq 0$, i.e.*

$$\mathbb{P}\{N(t) = k\} = \frac{(at)^k}{k!} \mathrm{e}^{-at} \quad \text{for } a < \infty,\ k \in \mathbb{Z}_+.$$

We take all the propositions of this subsection from Proposition 4.9 in [55]. We provide a representation of a Poisson process first.

Proposition 2.9. *A Poisson process $N = (N(t))_{t\geq 0}$ with intensity $a < \infty$ has the representation*

$$N(t) = \sum_{k=1}^{\infty} \chi\{\tau_k \leq t\} \text{ and } \tau_k = T_1 + T_2 + \ldots + T_k, \tag{2.1}$$

where $(T_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables which are exponentially distributed with parameter a.

Hence, N is a $\mathbb{Z}_+$-valued process which has a finite number of jumps on a finite time interval. Every jump of N is of size 1.

Proposition 2.10. *Given a Poisson process $N = (N(t))_{t\geq 0}$ with intensity $a < \infty$ and furthermore let $z \in \mathbb{C}$, then for $t \geq 0$*

$$\mathbb{E}\left[\mathrm{e}^{zN(t)}\right] = \exp\{at(\mathrm{e}^z - 1)\}. \tag{2.2}$$

The following proposition shows that a Poisson process is characterized by its jump sizes.

Proposition 2.11. *A $\mathbb{Z}_+$-valued Lévy process N with*

$$\mathbb{P}\{\Delta N(t) := N(t) - N(t-) \in \{0,1\}\} = 1 \tag{2.3}$$

is a Poisson process.

2.2.2. Compound Poisson process with values in Hilbert spaces

Compound Poisson processes represent a possible extension of Poisson processes. They have random jump sizes and can be considered U-valued, where we assume U to be a separable

Hilbert space with the corresponding Borel σ-algebra $\mathscr{B}(U)$.

Definition 2.12. *Let ν be a finite measure on U with $\nu(\{0\}) = 0$, then a* compound Poisson process with jump intensity measure ν *is a cadlag Lévy process which has the following distribution:*

$$\mathbb{P}\{L(t) \in \Gamma\} = \mathrm{e}^{-\nu(U)t} \sum_{k=0}^{\infty} \frac{t^k}{k!} \nu^{*k}(\Gamma), \quad t \geq 0, \quad \Gamma \in \mathscr{B}(U). \tag{2.4}$$

Above, ν^{*k} denotes the k-th convolution of the measure ν and we assume that $\nu^0 = \delta_0$ holds. Based on the distribution, we now state a representation for a compound Poisson process. This is a consequence of the following two lemmas which are proven in Theorem 4.15 in [55]. There, let ν be a finite measure on U with $\nu(\{0\}) = 0$ and $a = \nu(U)$.

Lemma 2.13. *Let $Z_1, Z_2, \ldots$ be independent, identically distributed random variables which take values in $U \setminus \{0\}$ with the distribution $a^{-1}\nu$. Additionally, let $(N(t))_{t\geq 0}$ be a Poisson process with intensity a which is independent of the random variables $Z_1, Z_2, \ldots$, then*

$$L(t) = \sum_{i=1}^{N(t)} Z_i, \quad t \geq 0, \tag{2.5}$$

is a compound Poisson process with jump intensity measure ν.

Lemma 2.14. *Suppose that L is a compound Poisson process with jump intensity measure ν, then one finds independent, identically distributed random variables $Z_1, Z_2, \ldots$ with distribution $a^{-1}\nu$ and a Poisson process $(N(t))_{t\geq 0}$ with intensity a which is independent of $Z_1, Z_2, \ldots$ such that $L(t) = \sum_{i=1}^{N(t)} Z_i$.*

Remark.
- *Hence, the random variables $N(t)$ provide the number of jumps of the compound Poisson process until time t. Consequently, $\nu(U) = \mathbb{E}[N(1)]$ equals the mean number of jumps in U in the time interval $[0,1]$.*
- *Moreover, $\nu(\Gamma)$ gives the mean number of jumps of L in $\Gamma \in \mathscr{B}(U \setminus \{0\})$ in the time interval $[0,1]$, since*

$$\mathbb{E}\left[\sum_{n=1}^{N(t)} \chi_\Gamma(Z_n)\right] = \sum_{k=0}^{\infty} \mathbb{P}\{N(t) = k\}\, \mathbb{E}\left[\sum_{n=1}^{k} \chi_\Gamma(Z_n)\right]$$

$$= \sum_{k=0}^{\infty} \mathbb{P}\{N(t)=k\} \sum_{n=1}^{k} \mathbb{P}\{Z_n \in \Gamma\} = \sum_{k=0}^{\infty} k\mathbb{P}\{N(t)=k\} \frac{\nu(\Gamma)}{\nu(U)} = t\nu(\Gamma).$$

Below, we consider important properties of compound Poisson processes L with jump intensity measure ν. We start with an equivalent condition for integrability, see Proposition 4.18 (i) in [55].

Proposition 2.15. *L is integrable if*

$$\int_U \|y\|_U \, \nu(dy) < \infty. \tag{2.6}$$

If (2.6) holds, then

$$\mathbb{E}[L(t)] = t\int_U y\nu(dy). \tag{2.7}$$

Here, the last term is interpreted as a Bochner integral.

We introduce the compensated compound Poisson process $\hat{L}(t) = L(t) - \mathbb{E}[L(t)]$, $t \geq 0$, and consider, besides properties of L, the ones for $\hat{L}$ as well.

Proposition 2.16. *Let $\left(\hat{L}(t)\right)_{t\geq 0}$ be a compensated compound Poisson process with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$, then $\hat{L}$ is a $(\mathscr{F}_t)_{t\geq 0}$-martingale.*

Proof. Let $0 \leq s \leq t$, then

$$\left\langle \mathbb{E}\left[\chi_A\left(\hat{L}(t) - \hat{L}(s)\right)\right], x\right\rangle_U = \mathbb{E}\left[\chi_A \left\langle \hat{L}(t) - \hat{L}(s), x\right\rangle_U\right]$$

for all $x \in U$ and every $A \in \mathscr{F}_s$ due to the linearity of the Bochner integral. Since the increments $\hat{L}(t) - \hat{L}(s)$ are independent of $\mathscr{F}_s$, it holds that

$$\left\langle \mathbb{E}\left[\chi_A\left(\hat{L}(t) - \hat{L}(s)\right)\right], x\right\rangle_U = \mathbb{P}(A)\mathbb{E}\left[\left\langle \hat{L}(t) - \hat{L}(s), x\right\rangle_U\right] = 0,$$

because $\hat{L}$ has mean zero. Thus,

$$\mathbb{E}\left[\chi_A \hat{L}(t)\right] = \mathbb{E}\left[\chi_A \hat{L}(s)\right]$$

for all $A \in \mathscr{F}_s$ due to the linearity of the Bochner integral. □

We now characterize the second moments of (compensated) compound Poisson processes. The following results can be found in Proposition 4.18 (iii) in [55].

Proposition 2.17. *L is square integrable if and only if*

$$\int_U \|y\|_U^2 \, \nu(dy) < \infty. \tag{2.8}$$

If (2.8) holds, then

$$\mathbb{E}\left\|\hat{L}(t)\right\|_U^2 = t \int_U \|y\|_U^2 \, \nu(dy) \tag{2.9}$$

and

$$\mathbb{E}\|L(t)\|_U^2 = t \int_U \|y\|_U^2 \, \nu(dy) + t^2 \left\| \int_U y \nu(dy) \right\|_U^2. \tag{2.10}$$

Furthermore, for all $x, \tilde{x} \in U$ and $t \geq 0$, we have

$$\mathbb{E}\left\langle \hat{L}(t), x \right\rangle_U \left\langle \hat{L}(t), \tilde{x} \right\rangle_U = t \int_U \langle x, y \rangle_U \langle \tilde{x}, y \rangle_U \, \nu(dy). \tag{2.11}$$

Finally, we state characteristic functions of $L(t)$ and $\hat{L}(t)$, $t \geq 0$.

Proposition 2.18. *For all $z \in \mathbb{C}$, $t \geq 0$ and $x \in U$, we have*

$$\mathbb{E}\left[\mathrm{e}^{z \langle x, \hat{L}(t) \rangle_U} \right] = \exp\left\{ -t \int_U \left(1 - \mathrm{e}^{z \langle x, y \rangle_U} + z \langle x, y \rangle_U \right) \nu(dy) \right\} \tag{2.12}$$

and

$$\mathbb{E}\left[\mathrm{e}^{z \langle x, L(t) \rangle_U} \right] = \exp\left\{ -t \int_U \left(1 - \mathrm{e}^{z \langle x, y \rangle_U} \right) \nu(dy) \right\}. \tag{2.13}$$

Proof. For simplicity we just compute the characteristic function of $L(t)$. We can easily conclude the one of $\hat{L}(t)$ from the characteristic function of $L(t)$. It holds that

$$\mathbb{E}\left[\mathrm{e}^{z \langle x, L(t) \rangle_U} \right] = \sum_{k=0}^{\infty} \mathrm{e}^{-at} \frac{(at)^k}{k!} \mathbb{E}\left[\mathrm{e}^{z \left\langle x, \sum_{n=1}^{k} Z_n \right\rangle_U} \right].$$

Due to the independence of the random variables Z_n, $n \in \mathbb{N}$, it follows

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{z\langle x,L(t)\rangle_U}\right] &= \sum_{k=0}^{\infty} \mathrm{e}^{-at}\,\frac{(at)^k}{k!}\prod_{n=1}^{k}\mathbb{E}\left[\mathrm{e}^{z\langle x,Z_n\rangle_U}\right] = \sum_{k=0}^{\infty}\mathrm{e}^{-at}\,\frac{(at)^k}{k!}\left(\int_U \mathrm{e}^{z\langle x,y\rangle_U}\,a^{-1}\nu(dy)\right)^k \\ &= \exp\left\{-at + at\int_U \mathrm{e}^{z\langle x,y\rangle_U}\,a^{-1}\nu(dy)\right\}.\end{aligned}$$

Since $a = \int_U 1\nu(dy)$, we obtain

$$\mathbb{E}\left[\mathrm{e}^{z\langle x,L(t)\rangle_U}\right] = \exp\left\{-t\int_U\left(1-\mathrm{e}^{z\langle x,y\rangle_U}\right)\nu(dy)\right\}.$$

□

2.2.3. Wiener processes in Hilbert spaces

Before dealing with Wiener processes, it is necessary to discuss Gaussian random variables first which take values in Hilbert spaces. We start with the well-known finite dimensional case to show the analogy to Hilbert space-valued Gaussian random variables.

We call a random vector $Y \in \mathbb{R}^n$ *Gaussian* with mean $\mu = \mathbb{E}[Y]$ and covariance matrix $\tilde{\mathscr{Q}} = \mathbb{E}[(Y-\mu)\cdot(Y-\mu)^T]$ if it has the density

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}}\det(\tilde{\mathscr{Q}})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y-\mu)^T\,\tilde{\mathscr{Q}}^{-1}(y-\mu)\right\},\quad \text{for all } y\in\mathbb{R}^n.$$

We write $Y \sim \mathscr{N}\left(\mu,\tilde{\mathscr{Q}}\right)$. The expression $\mathbb{E}\langle Y-\mu,x\rangle_{\mathbb{R}^n}\langle Y-\mu,y\rangle_{\mathbb{R}^n}$ is given by

$$\mathbb{E}\langle Y-\mu,x\rangle_{\mathbb{R}^n}\langle Y-\mu,y\rangle_{\mathbb{R}^n} = \left\langle\tilde{\mathscr{Q}}x,y\right\rangle_{\mathbb{R}^n} \tag{2.14}$$

for all $x,y\in\mathbb{R}^n$.

We assume U to be a Hilbert space below. The next definition is based on Definition 3.30 in [55]. For a Gaussian with values in U we extend the definition as follows:

Definition 2.19. *A U-valued random variable X is called* (centered) Gaussian *if the scalar random variable $\langle X,x\rangle_U$ is (centered) Gaussian for all $x\in U$.*
A U-valued stochastic process $(X(t))_{t\geq 0}$ is called (centered) Gaussian *if the scalar stochastic process $(\langle X(t),x\rangle_U)_{t\geq 0}$ is (centered) Gaussian for all $x\in U$.*

With the help of the next theorem we can formulate property (2.14) for the infinite-dimensional case as well.

Theorem 2.20. *Let the random variable X be centered Gaussian taking values in U, then $\mathbb{E}\,\|X\|_U^2 < \infty$ holds.*

Proof. The proof is done in Theorem 3.31 in [55]. $\square$

Suppose X is a Gaussian random variable in U with $\mathbb{E}[X] = \mu$. With the above Theorem we conclude that the mapping

$$(x,y) \mapsto \mathbb{E}\,\langle X-\mu, x\rangle_U \,\langle X-\mu, y\rangle_U$$

with $(x,y) \in U \times U$ is a symmetric, non negative definite and bounded bilinear form on U. Therefore, there exists a symmetric, non negative definite, linear and bounded operator $\mathcal{Q} : U \to U$, such that

$$\mathbb{E}\,\langle X-\mu, x\rangle_U \,\langle X-\mu, y\rangle_U = \langle \mathcal{Q}x, y\rangle_U .$$

In addition, $\mathcal{Q}$ is a trace class operator, since

$$\operatorname{tr}(\mathcal{Q}) = \sum_{n=1}^{\infty} \langle \mathcal{Q}e_n, e_n\rangle_U = \sum_{i=1}^{\infty} \mathbb{E}\,\langle X-\mu, e_n\rangle_U^2 = \mathbb{E}\,\|X-\mu\|_U^2 < \infty$$

for an arbitrary orthonormal basis $(e_n)_{n\in\mathbb{N}}$ of U. We briefly write $\mathcal{Q} \in L_1^+(U)$, where $L_1^+(U)$ is the space of all non negative definite and symmetric trace class operators on U.
We now introduce the Wiener process which is another representative of the class of Lévy processes. We always denote a Wiener process by W. Da Prato, Zabczyk [20] and Prévôt and Röckner [56] assume to have Gaussian increments of W in their definitions. This is actually not necessary. We will see below that the definition we use from [55] implies the Gaussian distribution.

Definition 2.21. *A U-valued Lévy process $W = (W(t))_{t\geq 0}$ with mean zero and $\mathbb{P}$-a.s. continuous trajectories is called* Wiener process.

Below, we state important properties of W.

Theorem 2.22. *W is a Gaussian process. Moreover, $\mathbb{E}\,\|W(t)\|_U^2 < \infty$ holds for all $t \geq 0$.*

Proof. From the definition given above, W is a Lévy process with almost surely continuous trajectories and mean zero. Hence, the scalar process $\langle W(t),x\rangle_U$, $t \geq 0$, is a real-valued Lévy process with almost surely continuous trajectories and mean zero for all $x \in U$. So, $\hat{W} = (\langle W(t),x\rangle_U)_{t\geq 0}$ is a real-valued Wiener process[2]. Therefore, $\hat{W}$ is a Gaussian process in particular, such that W is Gaussian by Definition 2.19. Consequently, by Theorem 2.20, W is square integrable. □

From Theorem 2.22, there exists an operators $\mathscr{Q}(t) \in L_1^+(U)$, such that

$$\mathbb{E}\,\langle W(t),x\rangle_U\,\langle W(t),y\rangle_U = \langle \mathscr{Q}(t)x,y\rangle_U \tag{2.15}$$

for all $t \geq 0$ and $x,y \in U$.
Furthermore, $\mathscr{Q}(t) = t\mathscr{Q}_0$ holds, where $\mathscr{Q}_0 \in L_1^+(U)$. We obtain this property by Theorem 2.39. Rewriting equation (2.15), we have

$$\mathbb{E}\,\langle W(t),x\rangle_U\,\langle W(t),y\rangle_U = t\,\langle \mathscr{Q}_0x,y\rangle_U \tag{2.16}$$

for all $t \geq 0$ and $x,y \in U$. We call the operator $\mathscr{Q}_0$ in equation (2.16) covariance operator of W. The following theorem provides a representation of a Wiener process W. The here given proof contains more details compared to the one stated in Section 4.4 in [55].

Theorem 2.23. *Let $\mathscr{Q}_0$ be the covariance operator of W and $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of the Hilbert space U which contains the eigenvectors of $\mathscr{Q}_0$. Additionally, let $(\gamma_n)_{n\in\mathbb{N}}$ be the corresponding eigenvalues, so that $\mathscr{Q}_0e_n = \gamma_n e_n$ for $n \in \mathbb{N}$, then*

$$W(t) = \sum_{n=1}^{\infty} W_n(t)e_n, \quad t \geq 0, \tag{2.17}$$

where the real-valued Wiener processes

$$W_n(t) = \langle W(t),e_n\rangle_U\,, \quad n \in \mathbb{N},$$

are independent with covariances

$$\mathbb{E}[W_n(s)W_n(t)] = \min\{s,t\}\ \gamma_n$$

[2]This property is a consequence of Lévy's martingale characterization of a standard Wiener process.

for $s,t \geq 0$. The series in (2.17) converges $\mathbb{P}$-a.s. and in $L^2(\Omega,\mathscr{F},\mathbb{P};U)$. Moreover, it can be shown that the series $\sum_{n=1}^{\infty}\gamma_n$ converges.

Proof. The processes $(\langle W(t),e_n\rangle_U)_{t\geq 0}$, $n\in\mathbb{N}$, are real-valued Lévy processes with mean zero and almost surely continuous trajectories. Hence, they are real-valued Wiener processes due to Lévy's martingale characterization for standard Wiener processes (Theorem 4.12 in [55]). Let $s\geq t$, then

$$\mathbb{E}\,\langle W(s)-W(t),e_n\rangle_U\,\langle W(t),e_m\rangle_U = \mathbb{E}\,\langle W(s)-W(t),e_n\rangle_U\,\mathbb{E}\,\langle W(t),e_m\rangle_U = 0,$$

since the increments are independent. By (2.16), we obtain

$$\begin{aligned}\mathbb{E}\,\langle W(s),e_n\rangle_U\,\langle W(t),e_m\rangle_U &= \mathbb{E}\,\langle W(t),e_n\rangle_U\,\langle W(t),e_m\rangle_U = t\,\langle Q_0e_n,e_m\rangle_U = t\,\gamma_n\,\langle e_n,e_m\rangle_U\\ &= \begin{cases} t\,\gamma_n & \text{if } n=m,\\ 0 & \text{if } n\neq m.\end{cases}\end{aligned}$$

The convergence with probability 1 is a consequence of the Fourier series representation $u=\sum_{n=1}^{\infty}\langle u,e_n\rangle_U\,e_n$ of an arbitrary element $u\in U$. Let us now consider the mean square convergence. We already know that $W(t)$, $t\geq 0$, can be interpreted as the following limit

$$\lim_{k\to\infty} X_k(t) = W(t)\quad \mathbb{P}\text{-a.s.}$$

Here, we set $X_k=\sum_{n=1}^{k}W_n(t)e_n$. It follows

$$\lim_{k\to\infty}\|X_k(t)\|_U^2 = \|W(t)\|_U^2\quad \mathbb{P}\text{-a.s.}$$

Since $\|X_k(t)\|_U^2=\sum_{n=1}^{k}|W_n(t)|^2$, the sequence $\left(\|X_k(t)\|_U^2\right)_{k\in\mathbb{N}}$ is increasing and hence, by the monotone convergence theorem, we obtain

$$\mathbb{E}\,\|X_k(t)\|_U^2 \to \mathbb{E}\,\|W(t)\|_U^2$$

for $k\to\infty$. For all $k\in\mathbb{N}$ it holds that

$$\mathbb{E}\,\|X_k(t)\|_U^2 = \sum_{n=1}^{k}\mathbb{E}|W_n(t)|^2 = t\sum_{n=1}^{k}\gamma_n.$$

Thus, we have

$$t\sum_{n=1}^{\infty}\gamma_n = \mathbb{E}\,\|W(t)\|_U^2 < \infty.$$

From the following equation

$$\mathbb{E}\left\|\sum_{n=k}^{m} W_n(t)e_n\right\|_U^2 = t\sum_{n=k}^{m}\gamma_n$$

with $k<m$ the convergence of series (2.17) in $L^2(\Omega,\mathscr{F},\mathbb{P};U)$ follows, since the right-hand side tends to zero for $k,m\to\infty$ and since $L^2(\Omega,\mathscr{F},\mathbb{P};U)$ is complete. □

Because the distribution of $\langle W(t),x\rangle_U$ is given by $\mathscr{N}(0,t\,\langle \mathscr{Q}_0x,x\rangle_U)$, we have the following characteristic function for $W(t)$:

$$\mathbb{E}\left[\mathrm{e}^{i\langle W(t),x\rangle_U}\right] = \exp\left\{-\frac{t}{2}\,\langle \mathscr{Q}_0x,x\rangle_U\right\},\quad t\geq 0,\quad x\in U. \tag{2.18}$$

Proposition 2.24. *Suppose that $(W(t))_{t\geq 0}$ is a U-valued Wiener process with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$, then W is an $(\mathscr{F}_t)_{t\geq 0}$-martingale.*

Proof. Let $0\leq s\leq t$ and $A\in\mathscr{F}_s$. We derive

$$\langle \mathbb{E}\,[\chi_A\,(W(t)-W(s))]\,,x\rangle_U$$

for $x\in U$. Due to the linearity of the Bochner integral, we have

$$\langle \mathbb{E}\,[\chi_A\,(W(t)-W(s))]\,,x\rangle_U = \mathbb{E}\,[\chi_A\,\langle W(t)-W(s),x\rangle_U]\,.$$

Since the increments $W(t)-W(s)$ are independent of $\mathscr{F}_s$, we obtain

$$\langle \mathbb{E}\,[\chi_A\,(W(t)-W(s))]\,,x\rangle_U = \mathbb{P}(A)\mathbb{E}\,[\langle W(t)-W(s),x\rangle_U] = 0,$$

because W has mean zero. Again, by the linearity of the Bochner integral it follows that

$$\mathbb{E}\,[\chi_A W(t)] = \mathbb{E}\,[\chi_A W(s)]$$

for all $A\in\mathscr{F}_s$. □

2.3. Jumps and jump measures of Lévy processes

In this section, by $L = (L(t))_{t\geq 0}$ we denote an arbitrary Lévy process. Further, we define $\Delta L(t) := L(t) - L(t-)$ as a jump of L at time t. First of all, we deal with the jumps of L. Below, we always assume $0 \leq t < \infty$. Now we introduce the following object:

$$N(t,A) = \#\{0 \leq s \leq t, \Delta L(s) \in A\} = \sum_{0\leq s\leq t} \chi_A(\Delta L(s)), \quad A \in \mathscr{B}(U \setminus \{0\}), \tag{2.19}$$

which counts the jumps of L which are in the set A. We only sum over the s in the right hand side of (2.19), where the corresponding summand is non-zero such that this expression is well-defined as a sum over a countable set. N we call the *jump counting measure* corresponding to L. This definition is helpful for the following.

Definition 2.25. *We call the set function ν with $\nu(A) = \mathbb{E}[N(1,A)]$ for $A \in \mathscr{B}(U \setminus \{0\})$* jump intensity measure *of L on $U \setminus \{0\}$.*

Remark. *The set function ν is not necessarily finite on $U \setminus \{0\}$. There are also Lévy processes which have infinitely many jumps on average on the time interval $[0,1]$, compare also [1].*

Moreover, we can find sets, in which we have finitely many jumps $\mathbb{P}$-a.s. as well as finitely many jumps on average on a finite time interval. For that reason, we proceed with the next definition.

Definition 2.26. *A set $A \in \mathscr{B}(U \setminus \{0\})$ is* separated from zero *if 0 is no element of the closure of A, i.e. $0 \notin \bar{A}$.*

Lemma 2.27. *If $A \in \mathscr{B}(U \setminus \{0\})$ is separated from zero, then the random function N is finite, i.e. $N(t,A) < \infty$ $\mathbb{P}$-a.s. for all $t \geq 0$.*

Proof. One finds this result and the corresponding proof in Proposition 2.8 in [1]. □

We still assume that the set $A \in \mathscr{B}(U \setminus \{0\})$ is separated from zero. Hence, $N = (N(t,A))_{t\geq 0}$ is a $\mathbb{Z}_+$-valued Lévy process which obviously just has jumps of size 1. Consequently, from Proposition 2.11 N is a Poisson process with parameter $\nu(A)$. This we summarize in the theorem below.

Theorem 2.28. *Let $A \in \mathscr{B}(U \setminus \{0\})$ be separated from zero, i.e. $0 \notin \bar{A}$, then*

$$\mathbb{P}\{N(t,A) = n\} = \frac{(\nu(A)t)^n}{n!} \mathrm{e}^{-\nu(A)t}, \; n \in \mathbb{N}, \quad t \geq 0. \tag{2.20}$$

Proof. This result is proven in [5] (Theorem 2.3.5) or in [42] (Proposition 2.15), respectively. □

Remark. *Since $(N(t,A))_{t\geq 0}$ is a Lévy process with bounded jumps, $t\nu(A) = \mathbb{E}[N(t,A)] < \infty$ holds by Theorem 2.7. Hence, $\nu(A) < \infty$ for sets $A \in \mathscr{B}(U \setminus \{0\})$ which are separated from zero.*

Theorem 2.29. *For pairwise disjoint Borel sets $A_1,\dots,A_m$ which are separated from zero and for distinct times $t_1,\dots,t_m \in \mathbb{R}_+$, the random variables $N(t_1,A_1),\dots,N(t_m,A_m)$ are independent.*

Proof. This statement is proven analogously to Theorem 2.3.5 in [5]. □

By previous considerations, we obtain the next statement.

Proposition 2.30. *(i) For almost all $\omega \in \Omega$ and for all $t \geq 0$, the set function $A \to N(t,A)(\omega)$ is a σ-finite measure on $\mathscr{B}(U \setminus \{0\})$.*

(ii) The set function $A \to \nu(A)$, where $\nu(A) = \mathbb{E}[N(1,A)]$, is a σ-finite measure on $\mathscr{B}(U \setminus \{0\})$.

Proof. See Theorem 2.13, Corollary 2.14, Theorem 2.17 and Corollary 2.18 in [1]. □

Remark. *We can extend the set functions that we consider in Proposition 2.30 to set functions on the σ-algebra $\mathscr{B}(U)$ by setting $N(t,\{0\})(\omega) = \nu(\{0\}) = 0$ for almost all $\omega \in \Omega$.*

As a next step, we define an integral of the form $\int_A f(x)N(t,dx)(\omega)$ for $\mathscr{B}(U \setminus \{0\})/\mathscr{B}(U)$-measurable functions $f : U \setminus \{0\} \to U$. Here, the set $A \in \mathscr{B}(U \setminus \{0\})$ is separated from zero, i.e. $0 \notin \bar{A}$. If we fix ω, then $N(t,\cdot)(\omega)$, $t \geq 0$, is a measure on $\mathscr{B}(U \setminus \{0\})$ by Proposition 2.30. Therefore, $\int_A f(x)N(t,dx)$ can be introduced as a random Bochner integral.
Everything that follows is based on Subsection 2.3.2 in the book of Applebaum [5]. Applebaum considers the same integral there but in the finite dimensional case, where $U = \mathbb{R}^d$.

From Lemma 2.27, we know that $N(t,A)$ is $\mathbb{P}$-a.s. finite for sets $A \in \mathscr{B}(U \setminus \{0\})$ that are separated from zero. In particular, this means that there are just finitely many $x \in A$, so that $N(t,\{x\}) \neq 0$ holds. Hence, we have

$$\int_A f(x)N(t,dx)(\omega) = \sum_{x\in A} f(x)N(t,\{x\})(\omega) \tag{2.21}$$

as a finite sum for each $t \geq 0$ and almost all $\omega \in \Omega$. As in (2.19), we do not consider the $x \in A$ in the summation, where the summand is zero. Since $N(t,\{x\}) \neq 0$ if and only if $\Delta L(s) = x$ for at least one time $0 \le s \le t$, we obtain

$$\int_A f(x)N(t,dx)(\omega) = \sum_{0\le s\le t} f(\Delta L(s))\chi_A(\Delta L(s)) \tag{2.22}$$

for almost every $\omega \in \Omega$. Because $(N(t,A))_{t\ge 0}$ represents a Poisson process for fixed A, we call the integral introduced above a *Poisson integral*. Below, we consider the characteristic function of $\int_A f(x)N(t,dx)$, $t \ge 0$. The following theorem provides that the distribution of the Poisson integral is characterized by the jump intensity measure ν. It is already proven for the finite dimensional case in [5] and here extended to a Hilbert space setting.

Theorem 2.31. *Let A be a Borel set separated from zero and $y \in U$, then we have*

$$\mathbb{E}\left[\exp\left\{i\left\langle y, \int_A f(x)N(t,dx)\right\rangle_U\right\}\right] = \exp\left\{t\int_A \left(\mathrm{e}^{i\langle y,f(x)\rangle_U} - 1\right)\nu(dx)\right\}. \tag{2.23}$$

Proof. We prove this theorem for simple functions first. For that reason, let $f(x) = \sum_{i=1}^n u_i\chi_{A_i}(x)$, where $u_i \in U$ and A_i are pairwise disjoint Borel subsets of A for every $i = 1,\dots,n$. It follows that $\int_A f(x)N(t,dx) = \sum_{i=1}^n u_iN(t,A_i)$. Thus,

$$\mathbb{E}\left[\exp\left\{i\left\langle y, \int_A f(x)N(t,dx)\right\rangle_U\right\}\right] = \mathbb{E}\left[\exp\left\{i\left\langle y, \sum_{i=1}^n u_iN(t,A_i)\right\rangle_U\right\}\right].$$

Since the random variables $N(t,A_i)$ are independent, it holds that

$$\mathbb{E}\left[\exp\left\{i\left\langle y, \int_A f(x)N(t,dx)\right\rangle_U\right\}\right] = \prod_{i=1}^n \mathbb{E}\left[\exp\left\{i\langle y,u_i\rangle_U N(t,A_i)\right\}\right].$$

Due to Proposition 2.10, we obtain

$$\begin{aligned}\mathbb{E}\left[\exp\left\{i\left\langle y, \int_A f(x)N(t,dx)\right\rangle_U\right\}\right] &= \prod_{i=1}^n \exp\left\{t\left(\mathrm{e}^{i\langle y,u_i\rangle_U} - 1\right)\nu(A_i)\right\}\\ &= \exp\left\{t\sum_{i=1}^n \left(\mathrm{e}^{i\langle y,u_i\rangle_U} - 1\right)\nu(A_i)\right\}\\ &= \exp\left\{t\int_A \left(\mathrm{e}^{i\langle y,f(x)\rangle_U} - 1\right)\nu(dx)\right\}.\end{aligned}$$

We can transfer the Poisson integral for simple functions to the representation in (2.22) since

$$\begin{aligned}\sum_{i=1}^{n} u_i N(t,A_i) &= \sum_{i=1}^{n} u_i \sum_{0\leq s\leq t} \chi_{A_i}(\Delta L(s)) = \sum_{0\leq s\leq t} \sum_{i=1}^{n} u_i \chi_{A_i}(\Delta L(s))\, \chi_A(\Delta L(s)) \\ &= \sum_{0\leq s\leq t} f(\Delta L(s))\, \chi_A(\Delta L(s)) .\end{aligned}$$

For general measurable functions $\tilde{f}$ we know about the existence of a sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ such that $f_n \to \tilde{f}$ point-wise for all $x \in U \setminus \{0\}$. This yields

$$\left\| \sum_{0\leq s\leq t} f_n(\Delta L(s))\, \chi_A(\Delta L(s)) - \sum_{0\leq s\leq t} \tilde{f}(\Delta L(s))\, \chi_A(\Delta L(s)) \right\|_U \to 0$$

for $n \to \infty$ and for almost every $\omega \in \Omega$. All in all, this means

$$\int_A f_n(x) N(t,dx) \to \int_A \tilde{f}(x) N(t,dx) \quad \mathbb{P}\text{-a.s.}$$

Lebesgue's theorem provides

$$\begin{aligned}\mathbb{E}\left[\exp\left\{ i \left\langle y, \int_A \tilde{f}(x) N(t,dx) \right\rangle_U \right\}\right] &= \lim_{n\to\infty} \mathbb{E}\left[\exp\left\{ i \left\langle y, \int_A f_n(x) N(t,dx) \right\rangle_U \right\}\right] \\ &= \lim_{n\to\infty} \exp\left\{ t \int_A \left(\mathrm{e}^{i\langle y, f_n(x)\rangle_U} - 1 \right) \nu(dx) \right\}.\end{aligned}$$

Again, by Lebesgue's theorem, we obtain

$$\mathbb{E}\left[\exp\left\{ i \left\langle y, \int_A \tilde{f}(x) N(t,dx) \right\rangle_U \right\}\right] = \exp\left\{ t \int_A \left(\mathrm{e}^{i\langle y, \tilde{f}(x)\rangle_U} - 1 \right) \nu(dx) \right\}.$$

□

2.4. Lévy-Khinchin decomposition

Throughout this section, we intensively use arguments and results from Section 4.5 in [55] in which the Lévy-Khinchin decomposition in Hilbert spaces is proven. Here, we do not prove new aspects but we contribute a more detailed discussion, e.g. in the proofs, and add more interpretations compared to [55] for a better understanding of this theory. For a finite dimensional

version of this result we refer to Section 2.4 in [5].
Let U be a separable Hilbert space and L be an arbitrary U-valued cadlag Lévy process defined on a filtered probability space $\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}\right)$[3]. Furthermore, let L be an $(\mathscr{F}_t)_{t\geq 0}$-adapted process and the increments $L(t+h)-L(t)$ be independent of $\mathscr{F}_t$ for $t,h\geq 0$.
The Lévy-Khinchin decomposition, we consider here, is one of the central results in the theory of Lévy processes and is helpful to introduce a stochastic integral with respect to L as we do in Section 3.2. The idea of this decomposition is to express $L(t)$ as the sum of $L_C(t)$ and $L_J(t)$, $t\geq 0$. Here, L_C is a process with continuous trajectories ($\mathbb{P}$-a.s.) and L_J is the jump part of L.
First of all, we consider the jump part L_J of L and construct a Lévy process which contains all jumps of L. We assume to have a monotonically decreasing null sequence $(r_k)_{k\in\mathbb{Z}_+}$. L has finitely many jumps in the set $A_0 = \{x : \|x\|_U \geq r_0\}$ on a finite time interval since A_0 is separated from zero. Hence, the expression

$$L_{A_0}(t) := \sum_{0\leq s\leq t} \chi_{A_0}(\Delta L(s))\,\Delta L(s), \quad t\geq 0, \tag{2.24}$$

is well-defined as a finite sum ($\mathbb{P}$-a.s.). So, we already have a process which contains jumps lying in A_0. To identify a process including the jumps which are smaller than r_0 with respect to the norm, turns out to be more complicated since we may have infinitely many jumps in Borel sets which are not separated from zero. We now consider the sets $A_k = \{x : r_k \leq \|x\|_U < r_{k-1}\}$, $k\in\mathbb{N}$. Obviously, they are separated from zero such that the sums

$$L_{A_k}(t) := \sum_{0\leq s\leq t} \chi_{A_k}(\Delta L(s))\,\Delta L(s), \quad t\geq 0,$$

are finite ($\mathbb{P}$-a.s.). These sums can represent the jumps of the Lévy process in A_k, $k\in\mathbb{N}$. So, we are able to cover the jumps in $\{x : r_n \leq \|x\|_U \leq r_0\}$, $n\in\mathbb{N}$, by $\sum_{k=1}^{n} L_{A_k}(t)$. The “small” jumps are the ones that cause problems since the series $\sum_{k=1}^{\infty} L_{A_k}(t)$, $t\geq 0$, can be divergent. This ansatz to represent the jumps in $\{x : \|x\|_U < r_0\}$ is not possible. Before we state a well-defined process that contains the “small” jumps, we consider Lemmas first which are important for the proof of the Lévy-Khinchin decomposition.
Below, let $L_A(t) := \sum_{0\leq s\leq t} \chi_A(\Delta L(s))\,\Delta L(s)$, $t\geq 0$, for Borel sets A separated from zero. With the following Lemma we can classify the processes L_{A_k}, $k\in\mathbb{Z}_+$.

[3] $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and complete.

Lemma 2.32. *Let $A \in \mathscr{B}(U \setminus \{0\})$ be separated from zero and $x \in U$, then*

$$\mathbb{E}\left[\mathrm{e}^{i\langle x, L_A(t)\rangle_U}\right] = \exp\left\{-t\int_A \left(1-\mathrm{e}^{i\langle x,y\rangle_U}\right)\nu(dy)\right\},$$

where ν is the jump intensity measure that corresponds to L.

Proof. See Theorem 2.31 with $f(x) = x$. □

Hence, the processes L_{A_k}, $k \in \mathbb{Z}_+$, are compound Poisson processes with jump intensity measure $\chi_{A_k}(y)\,\nu(dy)$.

Lemma 2.33. *Let L be a Lévy process and A separated from zero, then L_A and $L - L_A$ are independent Lévy processes.*

Proof. One finds the result in Appendix F of [55]. □

From the two Lemmas above, we conclude the following:

Lemma 2.34. *Let A and $\tilde{A}$ be disjoint Borel sets that are separated from zero, then L_A and $L_{\tilde{A}}$ are independent Lévy processes.*

Proof. Since $A \cup \tilde{A}$ is separated from zero, by Lemma 2.32 the process $L_{A\cup\tilde{A}}$ is a compound Poisson process and hence a Lévy process. Because A and $\tilde{A}$ are disjoint, we obtain

$$L_{A\cup\tilde{A}} = L_A + L_{\tilde{A}}.$$

With Lemma 2.33 the independence of L_A and $L_{\tilde{A}}$ follows. □

As already mentioned, the series $\sum_{k=1}^{\infty} L_{A_k}(t)$, $t \geq 0$, does not converge in general but we can compensate every summand, i.e. we subtract the mean. The mean exists by Theorem 2.7 due to the bounded jumps of the Lévy processes L_{A_k}, $k \in \mathbb{N}$, and is given by $\mathbb{E}\left[L_{A_k}(t)\right] = t\int_{A_k} y\,\nu(dy)$, see equation (2.7). So, we can formulate the following, which can also be found in Lemma 4.26 in [55].

Theorem 2.35. *The series $\sum_{k=1}^{\infty}\left(L_{A_k}(t) - t\int_{A_k} y\,\nu(dy)\right)$ converges with probability 1 uniformly with respect to t on every compact interval $[0,T]$ as well as in mean square (in $L^2(\Omega,\mathscr{F},\mathbb{P};U)$).*

With the processes from the above theorem and from (2.24) we now completely characterized the jump part L_J. Before we prove Theorem 2.35, we first of all need another property of the jump intensity measure ν. We already know that ν is finite on $A_0 = \{x : \|x\|_U \geq r_0\}$. Next, we formulate a result that shows the behaviour of ν on $\{x : \|x\|_U < r_0\}$. We take this result from Theorem 4.23 in [55], where it is proven with the same techniques.

Lemma 2.36. *If ν is the jump intensity measure corresponding to a Lévy process L, then*

$$\int_{\{y:\|y\|_U<r_0\}} \|y\|_U^2 \,\nu(dy) < \infty.$$

Proof. Due to Lemma 2.33, $\tilde{L} = L - L_{A_0}$ is a Lévy process. This process has jumps bounded by r_0. Hence, by Theorem 2.7, the second moment of $\tilde{L}$ exists. We set

$$\tilde{L}_n(t) := \left(\tilde{L}(t) - \sum_{k=1}^n L_{A_k}(t)\right) - \mathbb{E}\left(\tilde{L}(t) - \sum_{k=1}^n L_{A_k}(t)\right)$$

with $n \in \mathbb{N}$ and $t \geq 0$. This yields

$$\mathbb{E}\left\|\tilde{L}(t) - \mathbb{E}\tilde{L}(t)\right\|_U^2 = \mathbb{E}\left\|\tilde{L}_n(t) + \left(\sum_{k=1}^n L_{A_k}(t) - \mathbb{E}\sum_{k=1}^n L_{A_k}(t)\right)\right\|_U^2.$$

By Lemma 2.33, $\tilde{L} - \sum_{k=1}^n L_{A_k}$ and $\sum_{k=1}^n L_{A_k}$ are independent Lévy processes. Thus,

$$\mathbb{E}\left\|\tilde{L}(t) - \mathbb{E}\tilde{L}(t)\right\|_U^2 = \mathbb{E}\left\|\tilde{L}_n(t)\right\|_U^2 + \mathbb{E}\left\|\sum_{k=1}^n L_{A_k}(t) - \mathbb{E}\sum_{k=1}^n L_{A_k}(t)\right\|_U^2.$$

So, by equation (2.9), we obtain

$$t\int_{\{y:r_n\leq\|y\|_U<r_0\}} \|y\|_U^2 \,\nu(dy) = \mathbb{E}\left\|\sum_{k=1}^n L_{A_k}(t) - \mathbb{E}\sum_{k=1}^n L_{A_k}(t)\right\|_U^2 \leq \mathbb{E}\left\|\tilde{L}(t) - \mathbb{E}\tilde{L}(t)\right\|_U^2 < \infty,$$

$n \in \mathbb{N}$, and hence the result follows for $t = 1$ and $n \to \infty$. ☐

Remark. *(i) If we set $r_0 = 1$, then $\int_U \min\{\|y\|_U^2, 1\}\,\nu(dy) < \infty$.*

(ii) The measure ν is also called Lévy measure.

In the following proof the same arguments as in the proof of Lemma 4.26 in [55] are used but it contains much more details for a better understanding of every step of the proof.

Proof of Theorem 2.35. We start with showing the mean square convergence. By Lemma 2.34 it is known that $L_{A_1}, \ldots, L_{A_n}$, $n \in \mathbb{N}$, are independent Lévy processes. Thus,

$$\mathbb{E}\left\|\sum_{k=1}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2 = \sum_{k=1}^{n}\mathbb{E}\left\|L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right\|_U^2. \tag{2.25}$$

So, $F = \left(\mathbb{E}\left\|\sum_{k=1}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2\right)_{n\in\mathbb{N}}$ is a monotonically increasing sequence. Moreover, using (2.9) and Lemma 2.36, we know:

$$\begin{aligned}\mathbb{E}\left\|\sum_{k=1}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2 &= t\int_{\{y: r_n \leq \|y\|_U < r_0\}} \|y\|_U^2\, \nu(dy) \\ &\leq t\int_{\{y: \|y\|_U < r_0\}} \|y\|_U^2\, \nu(dy) < \infty\end{aligned}$$

for all $n \in \mathbb{N}$ and $t \geq 0$ such that F is bounded as well. For that reason, F converges. Furthermore, by equation (2.25), we obtain

$$\begin{aligned}&\mathbb{E}\left\|\sum_{k=m+1}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2 \\ &= \sum_{k=1}^{n}\mathbb{E}\left\|L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right\|_U^2 - \sum_{k=1}^{m}\mathbb{E}\left\|L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right\|_U^2 \\ &= \mathbb{E}\left\|\sum_{k=1}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2 - \mathbb{E}\left\|\sum_{k=1}^{m}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right\|_U^2\end{aligned} \tag{2.26}$$

for $m < n$. The term in (2.26) tends to zero for $m, n \to \infty$ since F converges. Hence, the convergence of the series $\sum_{k=1}^{\infty}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)$, $t \geq 0$, in $L^2(\Omega, \mathscr{F}, \mathbb{P}; U)$ follows due to the completeness of the space.

Now, we fix a finite time interval $[0,T]$ and set $Z_{m,n} := \left(\sum_{k=m}^{n}\left(L_{A_k}(t)-\mathbb{E}L_{A_k}(t)\right)\right)_{t\in[0,T]}$. $Z_{m,n}$ is a martingale (with respect to $(\mathscr{F}_t)_{t\in[0,T]}$) such that $\|Z_{m,n}\|_U^2$ is a submartingale. By Doob's submartingale inequality (Theorem 3.41 in [55]), we have the following:

$$\mathbb{P}\left\{\sup_{0\le t\le T}\|Z_{m,n}(t)\|_U^2 \ge \varepsilon\right\} \le \frac{\mathbb{E}\|Z_{m,n}(T)\|_U^2}{\varepsilon}.$$

The right-hand side tends to zero for $m,n\to\infty$. So, the series converges uniformly in probability on $[0,T]$ since every Cauchy sequence with respect to a measure, converges with respect to the measure. The desired convergence with probability 1 is a consequence of Proposition 3.11 and Corollary 3.12 in [55]. □

Remark. $\hat{L}=\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k} y\nu\,(dy)\right)$, *$t\ge 0$, is a Lévy process because it is represented by a sum of independent Lévy processes. Every summand of* $\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k} y\nu\,(dy)\right)$, *$t\ge 0$, is a square integrable martingale with respect to $(\mathscr{F}_t)_{t\ge 0}$. So, $\hat{L}$ is a square integrable martingale with respect to $(\mathscr{F}_t)_{t\ge 0}$ as well.*

Below, we state the Lévy-Khinchin decomposition, a result that is also stated in Section 4.5 in the book of Peszat and Zabczyk [55]. In the corresponding proof, we characterize the continuous part L_C of L.

Theorem 2.37 (Lévy-Khinchin decomposition)**.** *Let $(r_k)_{k\in\mathbb{Z}_+}$ be a monotonically decreasing null sequence, $A_0:=\{x:\|x\|_U\ge r_0\}$ and $A_k:=\{x:r_k\le\|x\|_U<r_{k-1}\}$ for $k\in\mathbb{N}$. Moreover, let ν be the jump intensity measure of L and $a\in U$, then a Lévy process can be decomposed as follows:*

$$L(t)=at+W(t)+\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k} y\nu\,(dy)\right)+L_{A_0}(t),\quad t\ge 0. \tag{2.27}$$

Here, W is a Wiener process, L_{A_k} is a compound Poisson process with jump intensity measure $\chi_{\{y:r_k\le\|y\|_U<r_{k-1}\}}(y)\,\nu\,(dy)$ for every fixed $k\ge 1$ and L_{A_0} is a compound Poisson process with jump intensity measure $\chi_{\{y:\|y\|_U\ge r_0\}}(y)\,\nu\,(dy)$. Additionally, the summands of this decomposition are independent processes and the series in (2.27) converges in mean square as well as uniformly with probability 1 on every compact time interval.

Proof. By the above considerations, we already characterized the jump part of L. We define

$$L_C(t):=L(t)-\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k} y\nu\,(dy)\right)-L_{A_0}(t),\quad t\ge 0.$$

Hence, L_C is Lévy process with $\mathbb{P}$-almost surely continuous trajectories. So, $a := \mathbb{E}[L_C(1)]$ exists and $W(t) := L_C(t) - at$, $t \geq 0$, is a Lévy process with mean zero and $\mathbb{P}$-almost surely continuous trajectories. By Definition 2.21, W is a Wiener process. We conclude the independence of the processes occurring in the decomposition from the Lemmas 2.33 and 2.34. □

Many authors choose an alternative representation of the jumps part in the Lévy-Khinchin decomposition. For example Applebaum in [3] defines the processes L_{A_k}, $k \in \mathbb{Z}_+$, as Poisson integrals (see Section 2.3), i.e.

$$L_{A_k}(t) = \int_{A_k} xN(t,dx), \quad t \geq 0.$$

Applebaum further sets $\tilde{N}(t,dx) := N(t,dx) - t\nu(dx)$ and rewrites the series in (2.27) as

$$\int_{\{x:\|x\|_U < r_0\}} x\tilde{N}(t,dx) := \lim_{n\to\infty} \int_{\{x:r_n<\|x\|_U<r_0\}} x\tilde{N}(t,dx) \tag{2.28}$$

for $t \geq 0$, where

$$\int_A x\tilde{N}(t,dx) := \int_A xN(t,dx) - t\int_A x\nu(dx)$$

for sets $A \in \mathscr{B}(U \setminus \{0\})$ separated from zero. The limit in definition (2.28) is meant to be in the $L^2(\Omega, \mathscr{F}, \mathbb{P}; U)$ sense or with probability 1, respectively. There are other approaches to define the integral in (2.28) such as in [1]. This integral representation has the advantage that an Ito integral with respect to the series in (2.27) can be defined by introducing an integral for the product measure $\tilde{N}(dt,dx)$.
The article of Applebaum [4] contains the Lévy-Khinchin decomposition for Lévy processes with values in separable Banach spaces and an even more general result is obtained by Riedle and van Gaans [63] for Lévy processes taking values in arbitrary Banach spaces.

Remark. *We can transfer the processes L_{A_k} with $k = 0,1,\dots$ into the form in (2.5). Therefore, we take the object introduced in (2.19):*

$$N(t,A_k) = \#\{0 \leq s \leq t, \Delta L(s) \in A_k\}.$$

Furthermore, we define a sequence $\left(\tau_i^k\right)_{i\in\mathbb{N}}$ of stopping times $k \in \mathbb{Z}_+$ as follows:

$$\tau_1^k := \inf\{t > 0 : \Delta L(t) \in A_k\} \text{ and } \tau_i^k := \inf\left\{t > \tau_{i-1}^k : \Delta L(t) \in A_k\right\}$$

($i = 2, 3, \ldots$). We define

$$Z_i^k = \Delta L(\tau_i^k),$$

then $\left(Z_i^k\right)_{i\in\mathbb{N}}$, $k \in \mathbb{Z}_+$, is the sequence of independent and identically distributed jump sizes. The processes L_{A_k}, $k \in \mathbb{Z}_+$ are then given by

$$L_{A_k}(t) = \sum_{i=1}^{N(t,A_k)} Z_i^k, \quad t \geq 0.$$

The jump sizes Z_i^k, $k \in \mathbb{Z}_+$ and $i \in \mathbb{N}$, have the distribution

$$\mathbb{P}\left\{Z_i^k \in \Gamma\right\} = \frac{\nu(\Gamma \cap A_k)}{\nu(A_k)}, \; \Gamma \in \mathscr{B}(U \setminus \{0\}).$$

Below, we provide a lemma including a detailed proof. The same result can be found in Section 4.6 in [55] but without proof.

Lemma 2.38. *Let U be a Hilbert space and $L = (L(t))_{t\geq 0}$ be a Lévy process with values in U. From its Lévy-Khinchin decomposition, we obtain*

$$\mathbb{E}\left[\mathrm{e}^{i\langle x, L(t)\rangle_U}\right] = \int_U \mathrm{e}^{i\langle x,y\rangle_U}\,\mu_t(dy) = \mathrm{e}^{-t\psi(x)}, \quad x \in U,$$

where

$$\begin{aligned}\psi(x) = &-i\,\langle a, x\rangle_U + \frac{1}{2}\langle \mathscr{Q}_0 x, x\rangle_U \\ &+ \int_U \left(1 - \mathrm{e}^{i\langle x,y\rangle_U} + \chi_{\{y:\|y\|_U<1\}}(y)\, i\,\langle x, y\rangle_U\right)\nu(dy). \end{aligned} \tag{2.29}$$

Above, we use the notation of Theorem 2.37 which means that $a \in U$, $\mathscr{Q}_0$ is the covariance operator of the Wiener part, ν is the jump intensity measure of L and μ_t denotes the distribution of $L(t)$.

Proof. From the Lévy-Khinchin decomposition, we obtain

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{i\langle x,L(t)\rangle_U}\right] = \mathbb{E}\left[\mathrm{e}^{i\langle x,at\rangle_U}\right]\cdot\mathbb{E}\left[\mathrm{e}^{i\langle x,W(t)\rangle_U}\right]\cdot\mathbb{E}\left[\mathrm{e}^{i\left\langle x,\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k} y\nu(dy)\right)\right\rangle_U}\right] \\ \cdot\,\mathbb{E}\left[\mathrm{e}^{i\left\langle x,L_{A_0}(t)\right\rangle_U}\right]\end{aligned}$$

by the independence of the processes. Below, we derive the characteristic functions of every component of the decomposition.

$$\mathbb{E}\left[\mathrm{e}^{i\langle x,W(t)\rangle_U}\right] = \mathrm{e}^{\frac{-t}{2}\langle \mathscr{Q}_0 x,x\rangle_U}$$

due to (2.18) and

$$\mathbb{E}\left[\mathrm{e}^{i\left\langle x,L_{A_0}(t)\right\rangle_U}\right] = \mathrm{e}^{-t\int_{\{y:\|y\|\geq r_0=1\}}\left(1-\mathrm{e}^{i\langle x,y\rangle_U}\right)\nu(dy)}$$

because of Lemma 2.32. The series in the Lévy-Khinchin converges $\mathbb{P}$-almost surely such that the corresponding characteristic function is

$$\mathbb{E}\left[\mathrm{e}^{i\left\langle x,\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k}y\nu(dy)\right)\right\rangle_U}\right] = \lim_{n\to\infty}\mathbb{E}\left[\mathrm{e}^{i\left\langle x,\sum_{k=1}^{n}\left(L_{A_k}(t)-t\int_{A_k}y\nu(dy)\right)\right\rangle_U}\right]$$

by Lebesgue's Theorem. Using Lemma 2.32 yields

$$\mathbb{E}\left[\mathrm{e}^{i\left\langle x,\sum_{k=1}^{n}\left(L_{A_k}(t)-t\int_{A_k}y\nu(dy)\right)\right\rangle_U}\right] = \mathrm{e}^{-t\int_{\{y:r_n\leq\|y\|_U<r_0=1\}}\left(1-\mathrm{e}^{i\langle x,y\rangle_U}+i\langle x,y\rangle_U\right)\nu(dy)}$$

for $n\in\mathbb{N}$. Hence, it follows

$$\mathbb{E}\left[\mathrm{e}^{i\left\langle x,\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k}y\nu(dy)\right)\right\rangle_U}\right] = \mathrm{e}^{-t\int_{\{y:\|y\|_U<r_0=1\}}\left(1-\mathrm{e}^{i\langle x,y\rangle_U}+i\langle x,y\rangle_U\right)\nu(dy)}.$$

Finally, we show that the integral in (2.29) is well-defined:

$$\begin{aligned}&\int_{\{y:\|y\|_U<r_0=1\}}\left|1-\mathrm{e}^{i\langle x,y\rangle_U}+i\langle x,y\rangle_U\right|\nu(dy)+\int_{\{y:\|y\|_U\geq r_0=1\}}\left|1-\mathrm{e}^{i\langle x,y\rangle_U}\right|\nu(dy)\\&\leq\tilde{k}\int_{\{y:\|y\|_U<r_0=1\}}\|y\|_U^2\,\nu(dy)+\int_{\{y:\|y\|_U\geq r_0=1\}}2\nu(dy)<\infty,\end{aligned}$$

where $\tilde{k}(x)$ is a suitable constant. The upper bound for the first summand follows by its Taylor series representation since

$$\left|1-\mathrm{e}^{i\langle x,y\rangle_U}+i\langle x,y\rangle_U\right| = \left|\sum_{k=2}^{\infty}\frac{\left(i\langle x,y\rangle_U\right)^k}{k!}\right|.$$

By the triangle and Cauchy-Schwarz inequalities, we have

$$\left|\sum_{k=2}^{\infty} \frac{(i\langle x,y\rangle_U)^k}{k!}\right| \le \sum_{k=2}^{\infty} \frac{\|x\|_U^k \|y\|_U^k}{k!} \le \sum_{k=2}^{\infty} \frac{\|x\|_U^k}{k!} \|y\|_U^2 = \tilde{k}(x) \|y\|_U^2 .$$

The last estimate is a consequence of $\|y\|_U < 1$. □

From Lemma 2.38 we conclude that every Lévy process L is characterized by $(a, \mathscr{Q}_0, \nu)$ which, for that reason, is called characteristic triple of L.

2.5. Square integrable Lévy processes

Below, we follow the remarks of Peszat and Zabczyk in Section 4.9 in [55] and summarize the most important results for Lévy processes with existing second moments.

In this section, L denotes an adapted and square integrable Lévy process which is defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, \mathbb{P})$[4] and which takes values in the Hilbert space U. Furthermore, we assume that the increments $L(t+h) - L(t)$ are independent of $\mathscr{F}_t$ for $t,h > 0$. In addition, by $L_1^+(U)$ we denote the space of all symmetric, non negative definite trace class operators on U.

Theorem 2.39. *There are $m \in U$ and an operator $\mathscr{Q} \in L_1^+(U)$ such that*

$$\begin{aligned} &\mathbb{E}\langle L(t), x\rangle_U = t\ \langle m, x\rangle_U , \\ &\mathbb{E}\langle L(t) - mt, x\rangle_U \langle L(s) - ms, y\rangle_U = \min\{s,t\} \langle \mathscr{Q}x, y\rangle_U , \\ &\mathbb{E}\|L(t) - mt\|_U^2 = t\ \mathrm{tr}(\mathscr{Q}). \end{aligned} \tag{2.30}$$

for all $t,s \ge 0$ and $x,y \in U$.

Proof. This result is proven in Theorem 4.44 in [55]. □

We call the vector m and the operator $\mathscr{Q}$ in Theorem 2.39 mean and covariance operator of L, respectively. Next, we state a result which is already proven in Theorem 4.47 (i) in [55]. Since the proof is quite short, we included it for reasons of completeness.

[4] We assume that $(\mathscr{F}_t)_{t\ge 0}$ is right-continuous and $\mathscr{F}_0$ contains all sets A with $\mathbb{P}(A) = 0$.

Theorem 2.40. *Suppose L is a Lévy process taking values in U with the corresponding jump intensity measure ν, then L is square integrable if and only if*

$$\int_U \|y\|_U^2 \, \nu(dy) < \infty. \tag{2.31}$$

Proof. We use the decomposition in (2.27). W is square integrable by Theorem 2.22. Moreover, the series $\sum_{k=1}^{\infty} \left(L_{A_k}(t) - t \int_{A_k} y \nu(dy) \right)$ converges by Theorem 2.35 in mean square. Hence, it follows

$$\begin{aligned}
&\mathbb{E} \left\| \sum_{k=1}^{\infty} \left(L_{A_k}(t) - t \int_{A_k} y \nu(dy) \right) \right\|_U^2 = \lim_{n \to \infty} \mathbb{E} \left\| \sum_{k=1}^{n} \left(L_{A_k}(t) - t \int_{A_k} y \nu(dy) \right) \right\|_U^2 \\
&= t \int_{\{y : \|y\|_U < r_0\}} \|y\|_U^2 \, \nu(dy) < \infty.
\end{aligned}$$

Finally, the compound Poisson process L_{A_0} with jump intensity measure $\chi_{\{y:\|y\|_U \geq r_0\}}(y)\, \nu(dy)$ is square integrable by Proposition 2.17 if and only if $\int_{\{y:\|y\|_U \geq r_0\}} \|y\|_U^2 \, \nu(dy) < \infty$. □

If condition (2.31) holds, then the process L_{A_0} is especially integrable. Hence, the mean value exists and is given by

$$\mathbb{E}[L_{A_0}(t)] = t \int_{\{y:\|y\|_U \geq r_0\}} y \nu(dy) = t \int_{A_0} y \nu(dy), \quad t \geq 0.$$

Using the decomposition in (2.27), we now state a representation for square integrable Lévy processes L.

Theorem 2.41. *Let L be a square integrable Lévy process, then*

$$L(t) = bt + W(t) + M_J(t), \quad \mathbb{P}\text{-a.s.}, \quad t \geq 0 \tag{2.32}$$

Here, $b \in U$, W is a Wiener process and M_J is martingale with respect to $(\mathscr{F}_t)_{t \geq 0}$ which contains all jumps of L. Further, we obtain that M_J and W are independent Lévy processes and that $\mathbb{E}[L(t)] = tb$, $t \geq 0$.

Proof. Below, we choose the notation of Theorem 2.37. We set $b = a + \int_{A_0} y \nu(dy)$ and $M_J(t) = \sum_{k=1}^{\infty} \left(L_{A_k}(t) - t \int_{A_k} y \nu(dy) \right) + \left(L_{A_0}(t) - t \int_{A_0} y \nu(dy) \right)$, $t \geq 0$, and representation (2.32) follows. The martingale property of the process M_J with respect to the filtration $(\mathscr{F}_t)_{t \geq 0}$ follows by the fact that it is the sum of martingales. The independence of the processes W and M_J is a consequence of Lemma 2.33. The Wiener process W has mean zero. The same is true for M_J since it is a process that consists of the sum of compensated processes. $\square$

Below, L is a square integrable Lévy process. With the above theorem we can state the structure of the covariance operator $\mathscr{Q}$ of L. For that reason, we consider the following expression and also use the independence of W and M_J:

$$\begin{aligned} &\mathbb{E}\langle W(t)+M_J(t),x\rangle_U\langle W(t)+M_J(t),y\rangle_U \\ &=\mathbb{E}\langle W(t),x\rangle_U\langle W(t),y\rangle_U+\mathbb{E}\langle M_J(t),x\rangle_U\langle M_J(t),y\rangle_U \\ &=t\langle(\mathscr{Q}_0+\mathscr{Q}_1)x,y\rangle_U \end{aligned} \tag{2.33}$$

for $x,y\in U$ and $t\geq 0$. So, we obtain $\mathscr{Q}=\mathscr{Q}_0+\mathscr{Q}_1$, where $\mathscr{Q}_0$ is the covariance operator of the Wiener process and $\mathscr{Q}_1$ the one of the jump part. In the next theorem, we provide an equation that characterizes the operator $\mathscr{Q}_1$. We take this characterization from Theorem 4.47 (ii) in [55]. We also state a slightly modified proof compared to [55] for reasons of completeness.

Theorem 2.42. *Let $\mathscr{Q}_1$ be the covariance operator of the jump process M_J, then $\mathscr{Q}_1$ is given by:*

$$\langle\mathscr{Q}_1x,y\rangle_U=\int_U\langle x,z\rangle_U\langle y,z\rangle\,\nu(dz) \tag{2.34}$$

with $x,y\in U$.

Proof. From the proof of Theorem 2.41 we know that M_J has the following representation:

$$M_J(t)=\sum_{k=1}^{\infty}\left(L_{A_k}(t)-t\int_{A_k}y\nu(dy)\right)+\left(L_{A_0}(t)-t\int_{A_0}y\nu(dy)\right),t\geq 0.$$

Due to Lemma 2.34, the processes $L_{A_0},L_{A_1},L_{A_2},\dots$ are independent. Therefore, we have

$$\begin{aligned} &\mathbb{E}\langle M_J(t),x\rangle_U\langle M_J(t),y\rangle_U \\ &=\mathbb{E}\left\langle L_{A_0}(t)-t\int_{A_0}y\nu(dy),x\right\rangle_U\left\langle L_{A_0}(t)-t\int_{A_0}y\nu(dy),y\right\rangle_U \\ &\quad+\sum_{k=1}^{\infty}\mathbb{E}\left\langle L_{A_k}(t)-t\int_{A_k}y\nu(dy),x\right\rangle_U\left\langle L_{A_k}(t)-t\int_{A_k}y\nu(dy),y\right\rangle_U \end{aligned}$$

for $x, y \in U$. Applying equation (2.11) yields

$$\mathbb{E} \langle M_J(t), x \rangle_U \langle M_J(t), y \rangle_U = t \int_U \langle x, z \rangle_U \langle y, z \rangle \nu(dz)$$

and the result follows. $\square$

2.6. Conclusions

In this chapter, we gave a detailed overview on Lévy processes with values in Hilbert spaces. We started with stating basic properties. Afterwards, important representatives, such as the Wiener process and the compound Poisson process were discussed. Furthermore, we analyzed the jump behaviour of general Lévy processes, for example, the number of jumps within certain sets. Based on these preliminary considerations, we provided the Lévy-Khinchin decomposition of a general Lévy process, a representation that is often used to introduce a corresponding stochastic integral, see Section 3.2. We concluded this chapter by discussing square integrable Lévy processes.

3. Stochastic Integration

In this chapter, we introduce stochastic integrals with respect to Hilbert space-valued Lévy processes. The main focus is on the stochastic analysis of square integrable Lévy martingales M which are square integrable martingales and Lévy processes at the same time. The definition of an integral with respect to M, which we introduce in Section 3.1, is the key to establish integrals for arbitrary Lévy processes. We deal with the more general setting in Section 3.2.
Integrals based on M are similar to integrals of the form $\int_0^t \Phi(s)dW(s)$, $t \in [0,T]$. Here, W is a Wiener process with values in separable Hilbert space and Φ, the integrand, is operator-valued. This Wiener-type integral is investigated in Section 4.2 in the book of Da Prato and Zabczyk [20], in Section 2.2 in the work of Gawarecki and Mandrekar [26] as well as in Section 2.3 of the book of Prévôt and Röckner [56]. Peszat and Zabczyk [55] extend this integral for square integrable martingales, see Sections 8.2 and 8.3. As an alternative to integrals with respect to Lévy martingales we refer to the article of Applebaum [3]. Applebaum states an integral which is based on martingale-valued measures which are much more general. All results of the following section are based on the already existing theory in [3, 20, 26, 55, 56] and for all proofs, ideas from these references are used as well. Additionally, we make use of material and approaches from Redmann [58].

3.1. Stochastic integrals with respect to Lévy martingales

In this section, let U be a separable Hilbert space and $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$[1] a filtered probability space. We assume that M, which is defined on the filtered probability space, is a square integrable martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$ and takes values in U. At the same time, we assume that M is a Lévy process with respect to $(\mathscr{F}_t)_{t\geq 0}$ (see Definition 2.3). We call the process M *Lévy martingale*. We write $M \in \mathscr{M}_L^2(U)$, where $\mathscr{M}_L^2(U)$ is the space of all square integrable Lévy martingales in U.

[1] $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and $\mathscr{F}_0$ contains all sets A with $\mathbb{P}(A) = 0$.

Below, we introduce a stochastic integral of the form $\int_0^t \Psi(s)dM(s)$, $t \in [0,T]$, where $\Psi(s,\omega)$ is an operator from U to the separable Hilbert space H.
Throughout this section, $L(U,H)$ denotes the Banach space of all linear and bounded operators from U to H. First, we characterize the class of simple processes.

Definition 3.1. *The $L(U,H)$-valued process $\Psi = (\Psi(t))_{t\in[0,T]}$ is called* simple *if for $0 = t_0 < t_1 < \ldots < t_{m+1} = T$ it has the representation*

$$\Psi(s) = \sum_{i=0}^{m} \chi_{(t_i,t_{i+1}]}(s)\Psi_i, \quad s \in [0,T]. \tag{3.1}$$

Here, the random variable $\Psi_i : \Omega \to L(U,H)$ is $\mathscr{F}_{t_i}/\mathscr{B}(L(U,H))$-measurable, $i \in \{0,1,\ldots,m\}$, and just takes finitely many values in $L(U,H)$.

Remark. *We can rearrange the representation of simple functions in (3.1) as follows*

$$\sum_{i=0}^{m} \chi_{(t_i,t_{i+1}]}(s)\Psi_i = \sum_{i=0}^{m} \chi_{(t_i,t_{i+1}]}(s) \sum_{j=1}^{N(i)} \chi_{A_j^i}(\omega)\Phi_j^i = \sum_{i=0}^{m}\sum_{j=1}^{N(i)} \chi_{(t_i,t_{i+1}]\times A_j^i}(s,\omega)\Phi_j^i,$$

where for $i = 1,\ldots,m$, we have

- *$N(i) \in \mathbb{N}$,*
- *pairwise disjoint sets $A_1^i,\ldots,A_{N(i)}^i \in \mathscr{F}_{t_i}$ and*
- *$\Phi_1^i,\ldots,\Phi_{N(i)}^i \in L(U,H)$.*

By $\mathscr{S}$ we denote the class of simple processes with values in $L(U,H)$ from now on. For processes $\Psi \in \mathscr{S}$, we define

$$I_T^M(\Psi) := \int_0^T \Psi(s)dM(s) := \sum_{i=0}^{m} \Psi_i\left(M(t_{i+1}) - M(t_i)\right)$$

and for $0 \le t_0 \le t \le T$, we set

$$I_{t_0,t}^M(\Psi) := \int_{t_0}^t \Psi(s)dM(s) := \int_0^T \chi_{[t_0,t]}(s)\Psi(s)dM(s).$$

The next theorem deals with the martingale property of the stochastic integral for integrands in $\mathscr{S}$. This result is well-known in the stochastic analysis and is proven similarly to the Wiener case which which can be found in [20, 26, 56]. We also refer to a more general frame work in [3] and [55]. Below, we use arguments of the Wiener case in the proof and show that they can also be applied in a more general setting.

Theorem 3.2. *For simple integrands $\Psi \in \mathscr{S}$ the stochastic process $\left(\int_0^t \Psi(s)dM(s)\right)_{t\in[0,T]}$ is a martingale with respect to $(\mathscr{F}_t)_{t\in[0,T]}$.*

Proof. Let $0 \leq s < t \leq T$ and $A \in \mathscr{F}_s$. We easily see that

$$\mathbb{E}\left[\chi_A I^M_{0,t}(\Psi)\right] = \mathbb{E}\left[\chi_A\left(I^M_{0,s}(\Psi) + I^M_{s,t}(\Psi)\right)\right].$$

We suppose that $s \in (t_k, t_{k+1}]$, $t \in (t_l, t_{l+1}]$ and $(t_k, t_{k+1}] \cap (t_l, t_{l+1}] = \emptyset$ w.l.o.g. It holds

$$\begin{aligned}&\mathbb{E}[\chi_A I^M_{s,t}(\Psi)]\\ &\quad= \mathbb{E}\left[\chi_A(\Psi_k(M(t_{k+1}) - M(s)) + \sum_{i=k+1}^{l-1} \Psi_i(M(t_{i+1}) - M(t_i)) + \Psi_l(M(t) - M(t_l)))\right].\end{aligned}$$

Below, we use the representation of simple functions given in the remark above and analyze every summand separately:

$$\begin{aligned}\mathbb{E}[\chi_A(\Psi_k(M(t_{k+1}) - M(s)))] &= \mathbb{E}\left[\chi_A(\sum_{j=1}^{N(k)} \chi_{A^k_j}\Phi^k_j(M(t_{k+1}) - M(s)))\right]\\ &= \sum_{j=1}^{N(k)} \Phi^k_j \mathbb{E}\left[\chi_{A\cap A^k_j}(M(t_{k+1}) - M(s))\right] = 0,\end{aligned}$$

because $A \cap A^k_j \in \mathscr{F}_s$. Furthermore, we obtain

$$\begin{aligned}\mathbb{E}\left[\chi_A(\sum_{i=k+1}^{l-1} \Psi_i(M(t_{i+1}) - M(t_i)))\right] &= \mathbb{E}\left[\chi_A(\sum_{i=k+1}^{l-1}\sum_{j=1}^{N(i)} \chi_{A^i_j}\Phi^i_j(M(t_{i+1}) - M(t_i)))\right]\\ &= \sum_{i=k+1}^{l-1}\sum_{j=1}^{N(i)} \Phi^i_j \mathbb{E}\left[\chi_{A\cap A^i_j}(M(t_{i+1}) - M(t_i))\right] = 0,\end{aligned}$$

since $A \cap A^i_j \in \mathscr{F}_{t_i}$. Finally, for the last term it follows

$$\mathbb{E}[\chi_A(\Psi_l(M(t) - M(t_l)))] = \mathbb{E}\left[\chi_A(\sum_{j=1}^{N(l)} \chi_{A^l_j}\Phi^l_j(M(t) - M(t_l)))\right]$$

$$= \sum_{j=1}^{N(l)} \Phi_j^l \mathbb{E}\left[\chi_{A\cap A_j^l}\left(M(t)-M(t_l)\right)\right] = 0,$$

due to $A \cap A_j^l \in \mathscr{F}_{t_l}$. Summarizing the previous steps yields

$$\mathbb{E}\left[\chi_A I_{0,t}^M(\Psi)\right] = \mathbb{E}\left[\chi_A I_{0,s}^M(\Psi)\right]$$

for all $A \in \mathscr{F}_s$. □

Theorem 2.41 implies that every square integrable Lévy process is a martingale if and only if it has mean zero. Due to equation (2.30) we know that the square integrable Lévy martingale M has a covariance operator $\mathscr{Q} \in L_1^+(U)$ which does not depend on time. In fact, the operator $\mathscr{Q}$ is given by:

$$\mathbb{E}\left\langle M(t),x\right\rangle_U \left\langle M(t),y\right\rangle_U = t\ \left\langle \mathscr{Q}x,y\right\rangle_U \tag{3.2}$$

for all $t \geq 0$ and $x,y \in U$. By the following proposition, the square root of a covariance operator can be defined.

Proposition 3.3. *Let $\hat{\mathscr{Q}} \in L(U)$ be non negative definite and symmetric, then there exists a unique non negative definite and symmetric operator $\hat{\mathscr{Q}}^{\frac{1}{2}} \in L(U)$ such that $\hat{\mathscr{Q}}^{\frac{1}{2}}\hat{\mathscr{Q}}^{\frac{1}{2}} = \hat{\mathscr{Q}}$.*

Proof. For the proof we refer to Proposition 2.3.4 in [56]. □

We introduce the Hilbert space of Hilbert-Schmidt operators from U to H. We denote this space by $L_{(HS)}(U,H)$ and the corresponding norm is $\|\cdot\|_{L_{(HS)}(U,H)}$. $L_{(HS)}(U,H)$ contains all operators $R \in L(U,H)$ with $\|R\|^2_{L_{(HS)}(U,H)} := \mathrm{tr}(RR^*) < \infty$, where R^* is the adjoint operator of R. The inner product in $L_{(HS)}(U,H)$ is given by $\langle R,S\rangle_{L_{(HS)}(U,H)} := \mathrm{tr}(SR^*)$ for $S,R \in L_{(HS)}(U,H)$.

Remark. *By Proposition 3.3 there is an operator $\mathscr{Q}^{\frac{1}{2}}$ corresponding to the covariance operator $\mathscr{Q}$ of M. This square root $\mathscr{Q}^{\frac{1}{2}}$ is a Hilbert-Schmidt operator since $\left\|\mathscr{Q}^{\frac{1}{2}}\right\|^2_{L_{(HS)}(U)} = \mathrm{tr}(\mathscr{Q}) < \infty$.*

Before we resume with the properties of the stochastic integral, we provide the following proposition which contains a well-known result.

Proposition 3.4. *Let K be a separable Hilbert space, $S \in L(K,H)$ and $T_o \in L_{(HS)}(U,K)$, then $ST_o \in L_{(HS)}(U,H)$ and $\|ST_o\|_{L_{(HS)}(U,H)} \leq \|S\|_{L(K,H)}\,\|T_o\|_{L_{(HS)}(U,K)}$.*

Proof. Let $(u_k)_{k\in\mathbb{N}}$ be an orthonormal basis of U. This yields

$$\begin{aligned}\|ST_o\|^2_{L_{(HS)}(U,H)} &= \operatorname{tr}(ST_o(ST_o)^*) = \sum_{k=1}^{\infty} \|ST_o u_k\|_H^2 \\ &\leq \|S\|^2_{L(K,H)} \sum_{k=1}^{\infty} \|T_o u_k\|_K^2 = \|S\|^2_{L(K,H)} \|T_o\|^2_{L_{(HS)}(U,K)} < \infty.\end{aligned}$$

□

In the next theorem, we state important properties for the integral $I_T^M(\Psi)$, where $\Psi \in \mathscr{S}$ is simple. These results are proven as in [3, 20, 26, 55, 56]. In particular, we apply techniques from the Wiener case in the corresponding proof. Again, they can be used for more general integrals with respect to M.

Theorem 3.5. *Let Ψ have the representation (3.1), then*

$$\mathbb{E}\left[I_T^M(\Psi)\right] = 0 \tag{3.3}$$

and

$$\mathbb{E}\left\|I_T^M(\Psi)\right\|_H^2 = \mathbb{E}\int_0^T \left\|\Psi(s)\mathscr{Q}^{\frac{1}{2}}\right\|^2_{L_{(HS)}(U,H)} ds. \tag{3.4}$$

Proof. Let $(h_k)_{k\in\mathbb{N}}$ be an orthonormal basis of H and $(u_k)_{k\in\mathbb{N}}$ be an orthonormal basis of U. We obtain equation (3.3) by

$$\begin{aligned}\mathbb{E}[I_T(\Psi)] &= \mathbb{E}[\sum_{k=1}^{\infty}\sum_{i=0}^{m} \langle h_k, \Psi_i(M(t_{i+1}) - M(t_i))\rangle_H h_k] \\ &= \sum_{k=1}^{\infty}\sum_{i=0}^{m} \mathbb{E}[\langle \Psi_i^* h_k, M(t_{i+1}) - M(t_i)\rangle_U] h_k \\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{i=0}^{m} \mathbb{E}[\langle h_k, \Psi_i u_l\rangle_H \langle u_l, M(t_{i+1}) - M(t_i)\rangle_U] h_k \\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{i=0}^{m} \mathbb{E}[\mathbb{E}\{\langle h_k, \Psi_i u_l\rangle_H \langle u_l, M(t_{i+1}) - M(t_i)\rangle_U | \mathscr{F}_{t_i}\}] h_k.\end{aligned}$$

The random variable $\langle h_k, \Psi_i u_l\rangle_H$ is $\mathscr{F}_{t_i}/\mathscr{B}(\mathbb{R})$-measurable. Additionally, $(\langle u_l, M(t)\rangle_U)_{t\geq 0}$ is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$ such that the result follows. We now derive the second moment:

$$\begin{aligned}\mathbb{E}\left\|I_T^M(\Psi)\right\|_H^2 &= \mathbb{E}\left\|\sum_{i=0}^{m}\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2\\ &= \sum_{i=0}^{m}\mathbb{E}\left\|\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2\\ &\quad+\sum_{\substack{i,j=0\\ i\neq j}}^{m}\mathbb{E}\left\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),\Psi_j\left(M(t_{j+1})-M(t_j)\right)\right\rangle_H.\end{aligned}$$

Below, we consider the mixed terms first. Let $i<j$ w.l.o.g. Hence, it holds that

$$\begin{aligned}&\mathbb{E}\left\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),\Psi_j\left(M(t_{j+1})-M(t_j)\right)\right\rangle_H\\ &=\mathbb{E}\left\langle\Psi_j^*\Psi_i\left(M(t_{i+1})-M(t_i)\right),M(t_{j+1})-M(t_j)\right\rangle_U\\ &=\sum_{l=1}^{\infty}\mathbb{E}\left\langle\Psi_j^*\Psi_i\left(M(t_{i+1})-M(t_i)\right),u_l\right\rangle_U\left\langle u_l,M(t_{j+1})-M(t_j)\right\rangle_U\\ &=\sum_{l=1}^{\infty}\mathbb{E}\left[\mathbb{E}\left\{\left\langle\Psi_j^*\Psi_i\left(M(t_{i+1})-M(t_i)\right),u_l\right\rangle_U\left\langle u_l,M(t_{j+1})-M(t_j)\right\rangle_U|\mathscr{F}_{t_j}\right\}\right]=0.\end{aligned}$$

We get the last equation since $\left\langle\Psi_j^*\Psi_i\left(M(t_{i+1})-M(t_i)\right),u_l\right\rangle_U$ is $\mathscr{F}_{t_j}/\mathscr{B}(\mathbb{R})$-measurable. This holds because

$$\begin{aligned}\left\langle\Psi_j^*\Psi_i\left(M(t_{i+1})-M(t_i)\right),u_l\right\rangle_U &= \left\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),\Psi_j u_l\right\rangle_H\\ &=\sum_{k=1}^{\infty}\left\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),h_k\right\rangle_H\left\langle h_k,\Psi_j u_l\right\rangle_H,\end{aligned}$$

where the terms $\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),h_k\rangle_H$ and $\left\langle h_k,\Psi_j u_l\right\rangle_H$ are$\mathscr{F}_{t_j}/\mathscr{B}(\mathbb{R})$-measurable. Therefore, we obtain

$$\mathbb{E}\left\|I_T^M(\Psi)\right\|_H^2=\sum_{i=0}^{m}\mathbb{E}\left\|\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2.$$

We have

$$\mathbb{E}\left\|\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2$$

$$
\begin{aligned}
&= \mathbb{E}\sum_{k=1}^{\infty}\langle\Psi_i\left(M(t_{i+1})-M(t_i)\right),h_k\rangle_H^2 \\
&= \mathbb{E}\sum_{k=1}^{\infty}\langle M(t_{i+1})-M(t_i),\Psi_i^*h_k\rangle_U^2 \\
&= \mathbb{E}\sum_{k=1}^{\infty}\left(\sum_{l=1}^{\infty}\langle M(t_{i+1})-M(t_i),u_l\rangle_U\langle\Psi_i u_l,h_k\rangle_H\right)^2 \\
&= \sum_{k=1}^{\infty}\sum_{s,l=1}^{\infty}\mathbb{E}\left[\langle M(t_{i+1})-M(t_i),u_l\rangle_U\langle\Psi_i u_l,h_k\rangle_H\ \langle M(t_{i+1})-M(t_i),u_s\rangle_U\langle\Psi_i u_s,h_k\rangle_H\right] \\
&= \sum_{k=1}^{\infty}\sum_{s,l=1}^{\infty}\mathbb{E}\left[\mathbb{E}\left\{\langle M(t_{i+1})-M(t_i),u_l\rangle_U\langle\Psi_i u_l,h_k\rangle_H\langle M(t_{i+1})-M(t_i),u_s\rangle_U\langle\Psi_i u_s,h_k\rangle_H\right.\right. \\
&\qquad\qquad \left.\left.|\mathscr{F}_{t_i}\right\}\right],
\end{aligned}
$$

Since the terms $\langle\Psi_i u_l,h_k\rangle_H$ and $\langle\Psi_i u_s,h_k\rangle_H$ are $\mathscr{F}_{t_i}/\mathscr{B}(\mathbb{R})$-measurable and the process $\langle M(t_{i+1})-M(t_i),u_l\rangle_U\langle M(t_{i+1})-M(t_i),u_s\rangle_U$ is independent of $\mathscr{F}_{t_i}$, we obtain

$$
\begin{aligned}
&\mathbb{E}\left\|\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2 \\
&= \mathbb{E}\left[\sum_{k=1}^{\infty}\sum_{s,l=1}^{\infty}\langle\Psi_i u_l,h_k\rangle_H\langle\Psi_i u_s,h_k\rangle_H\,\mathbb{E}\left[\langle M(t_{i+1})-M(t_i),u_l\rangle_U\langle M(t_{i+1})-M(t_i),u_s\rangle_U\right]\right].
\end{aligned}
$$

Applying equation (3.2) yields

$$
\begin{aligned}
\mathbb{E}\left\|\Psi_i\left(M(t_{i+1})-M(t_i)\right)\right\|_H^2 &= \mathbb{E}\sum_{k=1}^{\infty}\sum_{s,l=1}^{\infty}\langle\Psi_i u_l,h_k\rangle_H\langle\Psi_i u_s,h_k\rangle_H\langle\mathscr{Q}u_l,u_s\rangle_U\,(t_{i+1}-t_i) \\
&= (t_{i+1}-t_i)\,\mathbb{E}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\langle\Psi_i u_l,h_k\rangle_H\sum_{s=1}^{\infty}\langle u_s,\Psi_i^*h_k\rangle_U\langle\mathscr{Q}u_l,u_s\rangle_U \\
&= (t_{i+1}-t_i)\,\mathbb{E}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\langle\Psi_i u_l,h_k\rangle_H\langle\Psi_i^*h_k,\mathscr{Q}u_l\rangle_U \\
&= (t_{i+1}-t_i)\,\mathbb{E}\sum_{k=1}^{\infty}\langle\Psi_i^*h_k,\mathscr{Q}\Psi_i^*h_k\rangle_U \\
&= (t_{i+1}-t_i)\,\mathbb{E}\sum_{k=1}^{\infty}\langle\Psi_i\mathscr{Q}\Psi_i^*h_k,h_k\rangle_H \\
&= (t_{i+1}-t_i)\,\mathbb{E}\left[\operatorname{tr}\left(\Psi_i\mathscr{Q}\Psi_i^*\right)\right] \\
&= (t_{i+1}-t_i)\,\mathbb{E}\left\|\Psi_i\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2.
\end{aligned}
$$

This means

$$\mathbb{E}\left\|I_T^M(\Psi)\right\|_H^2 = \sum_{i=0}^{m}(t_{i+1}-t_i)\,\mathbb{E}\left\|\Psi_i \mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2$$

and result (3.4) follows. □

We now extend the space of integrable processes. Therefore, we define $\mathscr{P}_T$ as the smallest sub-σ-algebra of $\mathscr{B}([0,T])\otimes\mathscr{F}$ such that all mappings $F:[0,T]\times\Omega\to\mathbb{R}$, which are $(\mathscr{F}_t)_{t\in[0,T]}$-adapted with left-continuous trajectories, are measurable with respect to $\mathscr{P}_T$. The σ-algebra $\mathscr{P}_T$ is generated as follows:

$$\mathscr{P}_T = \sigma\left(\{(s,t]\times A\colon 0\le s\le t\le T,\, A\in\mathscr{F}_s\}\cup\{\{0\}\times A\colon A\in\mathscr{F}_0\}\right). \tag{3.5}$$

We call a $\mathscr{P}_T/\mathscr{B}(U)$-measurable mapping $F:[0,T]\times\Omega\to U$ *predictable* or *U-predictable*. A mapping Ψ on $[0,T]\times\Omega$, which takes values in the set of linear operators from U to H is called *predictable* if $[0,T]\times\Omega\ni(t,\omega)\mapsto\Psi(t,\omega)x$ is $\mathscr{P}_T/\mathscr{B}(H)$-measurable for all $x\in U$.

By $\mathscr{L}_T^2$ we introduce the space of all predictable mappings Ψ on $[0,T]\times\Omega$, taking values in the set of linear operators from U to H, that satisfy $\|\Psi\|_T<\infty$, where

$$\|\Psi\|_T^2 := \mathbb{E}\int_0^T \left\|\Psi(s)\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds.$$

We say that two mappings $\Psi_1,\Psi_2\in\mathscr{L}_T^2$ are equal in $\mathscr{L}_T^2$ if $\left\|(\Psi_1(s)-\Psi_2(s))\,\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}=0$ holds $\mathbb{P}\otimes dt$-almost surely. Hence, $\left(\mathscr{L}_T^2,\|\cdot\|_T\right)$ is a Hilbert space with inner product

$$\langle\Psi_1,\Psi_2\rangle_T := \mathbb{E}\int_0^T \left\langle\Psi_1(s)\mathscr{Q}^{\frac{1}{2}},\Psi_2(s)\mathscr{Q}^{\frac{1}{2}}\right\rangle_{L_{(HS)}(U,H)} ds.$$

The next remark illustrates that it is important to use predictable integrands if one desires to have an integral with the martingale property and with mean zero.

Remark. *Suppose that $N(t)$, $t\ge 0$, is a Poisson process, then $M(t)=N(t)-\mathbb{E}[N(1)]t$ is square integrable Lévy process with mean zero. $\Psi(t):=\Delta N(t)$, $t\ge 0$, is an adapted process with respect to the natural filtration but it is not predictable. We then have*

$$\int_0^t \Psi(s)dM(s) = \sum_{0\le s\le t}\Delta N(s) = N(t).$$

So, this integral is neither a martingale nor has mean zero. This example can also be found in [66].

The next Proposition shows that every function in $\mathscr{L}_T^2$ can be approximated by a sequence of simple functions with respect to the norm $\|\cdot\|_T$. For the proof we use the ideas and techniques of Lemma 3.1.2 in [67].

Proposition 3.6. *The class of simple functions $\mathscr{S}$ is dense in $\mathscr{L}_T^2$.*

Proof. Let $(h_k)_{k\in\mathbb{N}}$ and $(u_k)_{k\in\mathbb{N}}$ be an orthonormal bases of H and U, respectively. For all $i,j\in\mathbb{N}$ we define the operators $S_{ij}\in L(U,H)$ by

$$S_{ij}u_k := \begin{cases} h_j & \text{for } k=i \\ 0 & \text{for } k\neq i. \end{cases}$$

The adjoint operator of S_{ij} is given by

$$S_{ij}^*h_k := \begin{cases} u_i & \text{for } k=j \\ 0 & \text{for } k\neq j. \end{cases}$$

Since the following holds:

$$\left\|S_{ij}\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}}^2 = \sum_{k=1}^{\infty}\left\langle S_{ij}\mathscr{Q}S_{ij}^*h_k,h_k\right\rangle_H = \langle\mathscr{Q}u_i,u_i\rangle_U \leq \operatorname{tr}(\mathscr{Q}),$$

a mapping, which only takes the value S_{ij}, is an element of $\mathscr{L}_T^2$. We choose a simple function in $S\in\mathscr{S}$ which is defined the following way

$$S(s,\omega) = \chi_{(t_0,t_1]}(s)\chi_A(\omega)S_{ij},$$

where $0\leq t_0\leq t_1\leq T$ and $A\in\mathscr{F}_{t_0}$. By $\mathscr{S}^\perp$ we denote the orthogonal complement of $\mathscr{S}$ in $\mathscr{L}_T^2$. As the next step, we take an arbitrary mapping $R\in\mathscr{S}^\perp$. We obtain

$$\begin{aligned} 0 &= \langle R,S\rangle_T \\ &= \mathbb{E}\int_0^T \left\langle R(s)\mathscr{Q}^{\frac{1}{2}}, S(s)\mathscr{Q}^{\frac{1}{2}}\right\rangle_{L_{(HS)}(U,H)} ds \end{aligned}$$

$$= \mathbb{E}\left(\chi_A \int_{t_0}^{t_1} \sum_{k=1}^{\infty} \left\langle R(s)\mathscr{Q}S_{ij}^* h_k, h_k \right\rangle_H ds\right)$$
$$= \mathbb{E}\left(\chi_A \int_{t_0}^{t_1} \left\langle R(s)\mathscr{Q}u_i, h_j \right\rangle_H ds\right).$$

For $G \in \mathscr{P}_T$ we set

$$\mu(G) := \int_G \left\langle R(s)\mathscr{Q}u_i, h_j \right\rangle_H ds\mathbb{P}(d\omega),$$

such that μ is a signed measure on $\mathscr{P}_T$. For all sets G of type $(t_0, t_1] \times A$, we have $\mu(G) = 0$. Since the sets of the form $(t_0, t_1] \times A$ represent a π-system (non-empty family of subsets that is closed under finite intersections, see Section 3.1 in [55]) which generates $\mathscr{P}_T$, $\mu(G) = 0$ follows for all $G \in \mathscr{P}_T$. Hence, we have $\left\langle R(s)\mathscr{Q}u_i, h_j \right\rangle_H = 0$ $\mathbb{P} \otimes ds$-almost surely for all $i, j \in \mathbb{N}$. This yields $R(s)\mathscr{Q} = 0$ $\mathbb{P} \otimes ds$-almost surely. Then,

$$\|R\|_T^2 = \mathbb{E}\int_0^T \left\| R(s)\mathscr{Q}^{\frac{1}{2}} \right\|_{L_{(HS)}(U,H)}^2 ds = \mathbb{E}\int_0^T \mathrm{tr}\left(R(s)\mathscr{Q}R(s)^*\right) ds = 0$$

and consequently $\mathscr{S}^{\perp} = \{0\}$. So, we obtain $\bar{\mathscr{S}} = \mathscr{L}_T^2$. □

Let $\Psi \in \mathscr{L}_T^2$ and $(\Psi_n)_{n\in\mathbb{N}} \subset \mathscr{L}_T^2$ be a sequence of simple processes such that

$$\|\Psi_n - \Psi\|_T \to 0$$

for $n \to \infty$. The existence of this approximating sequence is ensured by Proposition 3.6. With equation (3.4) we obtain

$$\mathbb{E}\left\| \int_0^T \Psi_n(s)dM(s) - \int_0^T \Psi_m(s)dM(s) \right\|_H^2 = \mathbb{E}\int_0^T \left\| \left(\Psi_n(s) - \Psi_m(s)\right)\mathscr{Q}^{\frac{1}{2}} \right\|_{L_{(HS)}(U,H)}^2 ds. \quad (3.6)$$

The right-hand side of equation (3.6) tends to zero for $m, n \to \infty$ since $(\Psi_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{L}_T^2$. $\left(\int_0^T \Psi_n(s)dM(s)\right)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2\left(\Omega, \mathscr{F}, \mathbb{P}; H\right)$ and hence convergent. Therefore, we can define

$$\int_0^T \Psi(s)dM(s) := L^2 - \lim_{n\to\infty} \int_0^T \Psi_n(s)dM(s)$$

and for $0 \leq t_0 \leq t \leq T$ we set

$$\int_{t_0}^{t} \Psi(s)dM(s) := L^2 - \lim_{n\to\infty} \int_{t_0}^{t} \Psi_n(s)dM(s).$$

Here, "$L^2 - \lim_{n\to\infty}$" symbolizes the limit in $L^2(\Omega,\mathscr{F},\mathbb{P};H)$.

Below, we transfer properties of integrals for simple processes to $\int_0^T \Psi(s)dM(s)$, where $\Psi \in \mathscr{L}_T^2$. For that reason, we choose a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n := \int_0^T \Psi_n(s)dM(s)$. By the definition given above it converges to $x := \int_0^T \Psi(s)dM(s)$ in mean square. By using elementary estimations, for $n \to \infty$ we have

$$\|\mathbb{E}[x_n] - \mathbb{E}[x]\|_H = \|\mathbb{E}[x_n - x]\|_H \leq \mathbb{E}\,\|x_n - x\|_H \leq \left(\mathbb{E}\,\|x_n - x\|_H^2\right)^{\frac{1}{2}} \to 0.$$

$L^2(\Omega,\mathscr{F},\mathbb{P};H)$ is a Hilbert space and hence from mean square convergence, the convergence with respect to the norm follows, i.e.

$$\mathbb{E}\,\|x_n\|_H^2 \to \mathbb{E}\,\|x\|_H^2$$

for $n \to \infty$. We obtain

$$\begin{aligned}\mathbb{E}\left\|\int_0^T \Psi(s)dM(s)\right\|_H^2 &= \lim_{n\to\infty} \mathbb{E}\int_0^T \left\|\Psi_n(s)\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds \\ &= \mathbb{E}\int_0^T \left\|\Psi(s)\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds\end{aligned}$$

by using the fact that $\mathscr{L}_T^2$ is a Hilbert space as well which for $n \to \infty$ implies

$$\|\Psi_n\|_T^2 \to \|\Psi\|_T^2 .$$

The martingale property can also be shown for the more general integrands. So, let $0 \leq v < t$ and $A \in \mathscr{F}_v$, then

$$\begin{aligned}&\left\|\mathbb{E}\chi_A\left[\int_0^t \Psi_n(s)dM(s) - \int_0^t \Psi(s)dM(s)\right]\right\|_H \\ &\leq \mathbb{E}\left[\chi_A\left\|\int_0^t \Psi_n(s)dM(s) - \int_0^t \Psi(s)dM(s)\right\|_H\right]\end{aligned}$$

$$\leq \left(\mathbb{E} \left\| \int_0^t \Psi_n(s) dM(s) - \int_0^t \Psi(s) dM(s) \right\|_H^2 \right)^{\frac{1}{2}} \to 0$$

for $n \to \infty$. On the other hand, due to the martingale property for the stochastic integral with simple integrands, we have

$$\begin{aligned}
&\left\| \mathbb{E} \chi_A \left[\int_0^t \Psi_n(s) dM(s) - \int_0^v \Psi(s) dM(s) \right] \right\|_H \\
&= \left\| \mathbb{E} \chi_A \left[\int_0^v \Psi_n(s) dM(s) - \int_0^v \Psi(s) dM(s) \right] \right\|_H \\
&\leq \mathbb{E} \left[\chi_A \left\| \int_0^v \Psi_n(s) dM(s) - \int_0^v \Psi(s) dM(s) \right\|_H \right] \\
&\leq \left(\mathbb{E} \left\| \int_0^v \Psi_n(s) dM(s) - \int_0^v \Psi(s) dM(s) \right\|_H^2 \right)^{\frac{1}{2}} \to 0
\end{aligned}$$

for $n \to \infty$. Thus,

$$\mathbb{E} \left[\chi_A \int_0^t \Psi(s) dM(s) \right] = \mathbb{E} \left[\chi_A \int_0^v \Psi(s) dM(s) \right]$$

holds for all $A \in \mathscr{F}_v$.

In some cases, like in Section 3.2, we restrict ourselves to an important subset of $\mathscr{L}_T^2$. So, we introduce $\tilde{\mathscr{L}}_T^2$ as the set of all predictable mappings $\tilde{\Psi} : \Omega \times [0,T] \to L(U,H)$ that satisfy

$$\mathbb{E} \int_0^T \|\tilde{\Psi}(s)\|_{L(U,H)}^2 \, ds < \infty. \tag{3.7}$$

It is easy to see that $\tilde{\mathscr{L}}_T^2 \subset \mathscr{L}_T^2$ since for $\tilde{\Psi} \in \tilde{\mathscr{L}}_T^2$, we obtain

$$\int_0^T \left\| \tilde{\Psi}(s) \mathscr{Q}^{\frac{1}{2}} \right\|_{L_{(HS)}(U,H)}^2 ds \leq \operatorname{tr}(\mathscr{Q}) \mathbb{E} \int_0^T \|\tilde{\Psi}(s)\|_{L(U,H)}^2 \, ds < \infty$$

by using Proposition 3.4.

We conclude this section with two further important properties of a stochastic integral which are frequently used. The first property is already considered by Applebaum in [3] (Theorem 3) for integrals based on more general martingale-valued measures. Below, we assume that $\tilde{H}$ is another separable Hilbert space equipped with the norm $\|\cdot\|_{\tilde{H}}$.

Theorem 3.7. *Let $C \in L(H,\tilde{H})$ and $\Psi \in \mathscr{L}_T^2$, then*

$$C\int_0^T \Psi(s)dM(s) = \int_0^T C\Psi(s)dM(s) \quad \mathbb{P}\text{-a.s.}$$

Proof. First of all, we ensure that the right-hand side integral is well-defined. This is the case because by Proposition 3.4 it follows

$$\mathbb{E}\int_0^T \left\| C\Psi(s)\mathscr{Q}^{\frac{1}{2}} \right\|^2_{L_{(HS)}(U,\tilde{H})} ds \leq \|C\|^2_{L(H,\tilde{H})}\, \mathbb{E}\int_0^T \left\| \Psi(s)\mathscr{Q}^{\frac{1}{2}} \right\|^2_{L_{(HS)}(U,H)} ds < \infty.$$

Let $\tilde{\Psi} \in \mathscr{S}$. So, we have

$$\begin{aligned} C\int_0^T \tilde{\Psi}(s)dM(s) &= C\sum_{i=0}^m \Psi_i\left(M(t_{i+1}) - M(t_i)\right) \\ &= \sum_{i=0}^m C\Psi_i\left(M(t_{i+1}) - M(t_i)\right) \\ &= \int_0^T C\tilde{\Psi}(s)dM(s). \end{aligned}$$

We choose a sequence $(\Psi_n)_{n\in\mathbb{N}}$ of simple functions that approximates Ψ in $\mathscr{L}_T^2$. For $n \to \infty$

$$\begin{aligned} &\mathbb{E}\left\| C\int_0^T \Psi_n(s)dM(s) - C\int_0^T \Psi(s)dM(s) \right\|^2_{\tilde{H}} \\ &\leq \|C\|^2_{L(H,\tilde{H})}\, \mathbb{E}\left\| \int_0^T \Psi_n(s)dM(s) - \int_0^T \Psi(s)dM(s) \right\|^2_H \to 0 \end{aligned}$$

by the definition of the stochastic integral. Since

$$\begin{aligned} &\mathbb{E}\left\| C\int_0^T \Psi_n(s)dM(s) - \int_0^T C\Psi(s)dM(s) \right\|^2_{\tilde{H}} \\ &= \mathbb{E}\left\| \int_0^T C\Psi_n(s)dM(s) - \int_0^T C\Psi(s)dM(s) \right\|^2_{\tilde{H}} \\ &= \mathbb{E}\left\| \int_0^T C\left(\Psi_n(s) - \Psi(s)\right)dM(s) \right\|^2_{\tilde{H}} \\ &= \mathbb{E}\int_0^T \left\| C\left(\Psi(s) - \Psi_n(s)\right)\mathscr{Q}^{\frac{1}{2}} \right\|^2_{L_{(HS)}(U,\tilde{H})} ds \end{aligned}$$

$$= \|C\|^2_{L(H,\tilde{H})} \mathbb{E} \int_0^T \left\| (\Psi(s) - \Psi_n(s)) \mathscr{Q}^{\frac{1}{2}} \right\|^2_{L_{(HS)}(U,H)} ds \to 0$$

for $n \to \infty$, the result follows. □

The second property is the mean square continuity of the stochastic integral, compare also with Theorem 8.7 in [55] for a different proof.

Theorem 3.8. *Let $\Psi \in \mathscr{L}^2_T$, then the process*

$$I^M_t(\Psi) = \int_0^t \Psi(s) dM(s), \quad t \in [0,T],$$

is mean square continuous.

Proof. Let $0 \leq s \leq t \leq T$. This result directly follows from the Ito isometry

$$\mathbb{E} \left\| I^M_t(\Psi) - I^M_s(\Psi) \right\|^2_H = \mathbb{E} \left\| I^M_{s,t}(\Psi) \right\|^2_H = \int_s^t \mathbb{E} \left\| \Psi(\tau) \mathscr{Q}^{\frac{1}{2}} \right\|^2_{L_{(HS)}(U,H)} d\tau$$

and the continuity of the ordinary integral. □

3.2. Stochastic integrals with respect to general Lévy processes

Stochastic integrals based on Hilbert space-valued Lévy processes are treated in Section 8.6 in [55]. In [3], such integrals are discussed as well but mainly for deterministic integrands. In both works the Lévy-Khinchin decomposition of a Lévy process is used in order to define stochastic integrals. In [4], ideas are stated how to define an integral with respect to Lévy processes taking values in a separable Banach space.

Below, let U, H be separable Hilbert spaces and $\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}\right)$[2] be a filtered probability space. All processes that occur in this section are defined on this probability space. Further, let L be an arbitrary Lévy process with respect to the filtration $(\mathscr{F}_t)_{t\geq 0}$ (see Definition 2.3) that takes values in U. We recall Theorem 2.37. This tells us that L can be decomposed as follows:

$$L(t) = at + M(t) + P(t), \quad t \geq 0.$$

[2] $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and complete.

Here, $a \in U$ and M is a square integrable martingale and a Lévy process with respect to $(\mathscr{F}_t)_{t\geq 0}$ as well (We call M square integrable Lévy martingale.). Moreover, P is a compound Poisson process with respect to $(\mathscr{F}_t)_{t\geq 0}$. In addition, we know that M contains the continuous part and the part with jumps bounded by $r_0 > 0$. P is a process with piece-wise constant trajectories which just has jumps larger than r_0 with respect to the norm.

It remains to introduce an integral with respect to P to be able to define one for L. A possible way to define it is stated in Peszat, Zabczyk [55] but we prefer the ansatz which is sketched in the article of Applebaum [4].

Below, let N be the jump counting measure corresponding to L that we define in (2.19). By Proposition 2.30 we know that for all $t \geq 0$ and $\omega \in \Omega$, the function $N(t,\cdot)(\omega)$ is a measure on $\mathscr{B}(U\setminus\{0\})$. N can be extended to a measure on $\mathscr{B}(\mathbb{R}_+ \times (U\setminus\{0\}))$. So, let $\mathscr{S}(\mathbb{R}_+)\times \mathscr{B}(U\setminus\{0\})$ be the semiring of sets of the form $(t_1,t_2]\times A$, $A \in \mathscr{B}(U\setminus\{0\})$ and $0 \leq t_1 < t_2$. We set

$$N((t_1,t_2]\times A)(\omega) := N(t_2,A)(\omega) - N(t_1,A)(\omega).$$

Hence, $N(dt,dx)$ is a σ-finite pre-measure for all $\omega \in \Omega$ on $\mathscr{S}(\mathbb{R}_+)\times\mathscr{B}(U\setminus\{0\})$. By $N(dt,dx)$ we also denote the measure which is the unique extension of the pre-measure on the σ-algebra $\mathscr{B}(\mathbb{R}_+\times(U\setminus\{0\}))$, see Section 2 in [64] for more details. From Section 2.3 we already know that P can be interpreted as a random Bochner integral:

$$P(t) = \int_{\{x:\|x\|_U \geq r_0\}} xN(t,dx), \quad t \geq 0.$$

We define an integral with respect to P as a random Bochner integral as well

$$\int_0^T \Phi(s)dP(s) := \int_0^T \int_{\{x:\|x\|_U\geq r_0\}} \Phi(s)xN(ds,dx) \tag{3.8}$$

for predictable mappings $\Phi : \Omega \times [0,T] \to L(U,H)$ and $T > 0$. Since N is a finite counting measure on $[0,T]\times\{x : \|x\|_U \geq r_0\}$, we identify the integral defined in (3.8) with a finite sum

$$\int_0^T \Phi(s)dP(s) = \sum_{0\leq s\leq T} \Phi(s)\Delta L(s)\chi_{\{x:\|x\|_U\geq r_0\}}(\Delta L(s)).$$

for $\mathbb{P}$-almost all $\omega \in \Omega$.

Definition 3.9. *We fix a finite time interval* $[0,T]$. *We then call the stochastic process*

$$Y(t) := \int_0^t f(s)ds + \int_0^t \tilde{\Psi}(s)dM(s) + \int_0^t \Phi(s)dP(s), \quad t \in [0,T], \tag{3.9}$$

Lévy-type integral. *Here,* $f : \Omega \times [0,T] \to H$ *is* $\mathscr{B}([0,T]) \otimes \mathscr{F}$*-measurable and*

$$\int_0^T \|f(s)\|_H ds < \infty \ \ \mathbb{P}\text{-a.s.}$$

holds. Moreover, we choose $\tilde{\Psi} \in \tilde{\mathscr{L}}_T^2$ *defined in (3.7) and a predictable process* $\Phi : \Omega \times [0,T] \to L(U,H)$. *The first term in (3.9) is a path-wise Bochner integral and the second term is an integral with respect to the square integrable Lévy martingale* M *which we focus on in Section 3.1.*

The expression (3.9) contains a definition of a stochastic integral of the form $\int_0^t \Psi(s)dL(s)$, $t \in [0,T]$, which we obtain by setting $f = \Psi a$ and $\tilde{\Psi} = \Phi = \Psi$ for $\Psi \in \tilde{\mathscr{L}}_T^2$. If the Lévy process L is square integrable, then by Section 2.5 it has the representation

$$L(t) = bt + \tilde{M}(t), \quad t \geq 0,$$

where $b \in U$ and $\tilde{M}$ is a square integrable Lévy martingale. Consequently, in this case the structure of the Lévy-type integral is simpler.

3.3. Conclusions

In this chapter, we introduced a stochastic integral with respect to a square integrable Lévy process taking values in a Hilbert space. Based on the results for the Wiener case, we analyzed properties such as the mean, the Ito isometry and the martingale property of this stochastic integral. This chapter was concluded by the definition of an integral with respect to general Lévy processes using the Lévy-Khinchin decomposition which we studied in Section 2.4.

4. Linear Controlled SPDEs with Lévy Noise

Many works deal with stochastic partial differential equations (SPDEs) with Lévy noise. In [55] equations with additive as well as with multiplicative noise are considered. There, the concepts of weak and mild solutions are investigated and cases are stated in which both concepts coincide. Furthermore, in the thesis of Stolze [67], the existence and uniqueness of weak solutions for equations with additive noise under certain conditions is shown. The basics for the thesis of Stolze are provided by Applebaum in [3]. In the works of F. Knäble [42] and K. Knäble [43] more SPDEs with multiplicative noise can be found. All the above references deal with SPDEs in an abstract Hilbert space setting, but there are also articles, where equations in Banach spaces are studied, see for example [63].
In this chapter, we discuss linear SPDEs that are driven by Lévy noise which are controlled in addition. Below, we use an abstract evolution equation approach in order to represent SPDEs. Therefore, we define C_0-semigroups first to be able to characterize mild solutions. Afterwards, we introduce the model we focus on and show that it is well-defined. We conclude this chapter with two examples.

4.1. C_0-semigroups

We need the concept of C_0-semigroups and the corresponding generators which is vital in order to be able to introduce evolution equations. We take the following definitions and theorems from the book of Vrabie [70]. He introduces C_0-semigroups on Banach spaces but for us the Hilbert space setting is sufficient. Below, let H denote a separable Hilbert space.

Definition 4.1. *A family $(S(t))_{t\geq 0}$ of bounded and linear operators $S(t) \in L(H)$ is called C_0-*semigroup *on H if*

(i) $S(0) = I$, where I is the identity on H,

(ii) $S(t+s) = S(t)S(s)$ *for all* $t, s \geq 0$,

(iii) $\lim_{t\downarrow 0} S(t)x = x$ *for all* $x \in H$.

Remark. *Condition (iii) in Definition 4.1 is equivalent to* $t \mapsto S(t)x$ *is continuous for every* $x \in H$. *This strong continuity condition of* $(S(t))_{t\geq 0}$ *is also often used in the literature.*

Theorem 4.2. *Let* $(S(t))_{t\geq 0}$ *be a* C_0*-semigroup, then there exist constants* $\hat{\alpha} \in \mathbb{R}$ *and* $K \geq 1$ *such that for all* $t \geq 0$

$$\|S(t)\|_{L(H)} \leq K\,\mathrm{e}^{\hat{\alpha}t}. \tag{4.1}$$

Proof. See Theorem 2.3.1 in [70]. □

Definition 4.3. *An* infinitesimal generator *or simply* generator of a C_0-semigroup $(S(t))_{t\geq 0}$ *is an operator* $\mathscr{A} : D(\mathscr{A}) \subset H \to H$ *with*

- $D(\mathscr{A}) = \left\{x \in H, \quad \lim_{t\downarrow 0} \frac{1}{t}(S(t)x - x) \text{ exists}\right\}$ *and*
- $\mathscr{A}x = \lim_{t\downarrow 0} \frac{1}{t}(S(t)x - x)$ *for* $x \in D(\mathscr{A})$.

Remark.
- *The generator of a* C_0*-semigroup is a linear operator but it is not necessarily bounded.*
- *A* C_0*-semigroup is called* generalized contraction semigroup *if* $K = 1$ *in (4.1). It is called* contraction semigroup *if furthermore* $\hat{\alpha} = 0$ *holds.*

Next, we state basic properties of C_0-semigroups.

Theorem 4.4. *Let* $\mathscr{A} : D(\mathscr{A}) \subset H \to H$ *be the generator of a* C_0*-semigroup* $(S(t))_{t\geq 0}$, *then*

(i) for all $x \in H$ *and* $t \geq 0$, *we have*

$$\lim_{h\downarrow 0} \frac{1}{h} \int_t^{t+h} S(s)x\, ds = S(t)x,$$

(ii) for all $x \in H$ *and* $t > 0$, *it holds*

$$\int_0^t S(s)x\, ds \in D(\mathscr{A}) \quad \text{and} \quad \mathscr{A}\left(\int_0^t S(s)x\, ds\right) = S(t)x - x,$$

(iii) for all $x \in D(\mathscr{A})$ and $t \geq 0$, we obtain $S(t)x \in D(\mathscr{A})$, the mapping $t \mapsto S(t)x$ is once continuously differentiable and

$$\frac{d}{dt}(S(t)x) = \mathscr{A}S(t)x = S(t)\mathscr{A}x,$$

(iv) for all $x \in D(\mathscr{A})$ and $0 \leq t_0 \leq t < \infty$, we have

$$\int_{t_0}^{t} \mathscr{A}S(s)x\,ds = \int_{t_0}^{t} S(s)\mathscr{A}x\,ds = S(t)x - S(t_0)x.$$

Proof. See Theorem 2.3.2 in [70]. □

The next theorem contains a very important result for contraction semigroups.

Theorem 4.5. *If $(S(t))_{t\geq 0}$ is a contraction semigroup on H, then there is a Hilbert space $\hat{H}$ that contains H and a strongly continuous group $\hat{S}$ on $\hat{H}$ such that $\mathrm{Pr}_H \hat{S}(t) = S(t)$, $t \geq 0$, where Pr_H is the orthogonal projection of $\hat{H}$ onto H.*

Proof. This result is a consequence of the Theorems 9.22 and 9.23 in [55]. □

We summarize important properties of the generator of a C_0-semigroup in the following two theorems.

Theorem 4.6. *Let $\mathscr{A} : D(\mathscr{A}) \subset H \to H$ be the generator of a C_0-semigroup $(S(t))_{t\geq 0}$, then $D(\mathscr{A})$ is dense in H and $\mathscr{A}$ is a closed operator.*

Proof. See Theorem 2.4.1 in [70]. □

Theorem 4.7. *If $\mathscr{A} : D(\mathscr{A}) \subset H \to H$ is the generator of two C_0-semigroups $(S(t))_{t\geq 0}$ and $(T(t))_{t\geq 0}$, then $S(t) = T(t)$ holds for all $t \geq 0$.*

Proof. See Theorem 2.4.2 in [70]. □

4.2. Linear controlled evolution equations with Lévy noise

In this section, we deal with an infinite dimensional system, where the noise process is denoted by M. M takes values in a separable Hilbert space U and is defined on a probability space

$(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$[1]. In addition, we assume that M is a Lévy process with respect to $(\mathscr{F}_t)_{t\geq 0}$ which is square integrable with mean zero (see Definition 2.3). The most important properties regarding this process and the definition of an integral with respect to M can be found in the Chapters 2 and 3.

By the assumptions we make on M, with Theorem 2.39 we know about the existence of a covariance operator $\mathscr{Q}$ which is symmetric, positive definite and of trace class. For $x, y \in U$ and $s, t \geq 0$, it is characterized by

$$\mathbb{E} \langle M(t), x\rangle_U \langle M(s), y\rangle_U = \min\{t, s\} \langle \mathscr{Q}x, y\rangle_U . \tag{4.2}$$

Below, by H we denote another separable Hilbert space. We then consider the following stochastic differential equation that is equipped with an output equation:

$$\begin{aligned} d\mathscr{X}(t) &= [\mathscr{A}\mathscr{X}(t) + \mathscr{B}u(t)]\,dt + \mathscr{N}(\mathscr{X}(t-))dM(t), \quad \mathscr{X}(0) = X_0 \in H, \\ \mathscr{Y}(t) &= \mathscr{C}\mathscr{X}(t), \quad t \geq 0. \end{aligned} \tag{4.3}$$

We make the following assumptions:

- $\mathscr{A} : D(\mathscr{A}) \to H$ is a generator of a contraction semigroup $(S(t))_{t\geq 0}$.
- $\mathscr{N}$ is a linear mapping on H with values in the set of all linear operators from U to H such that $\mathscr{N}(h)\mathscr{Q}^{\frac{1}{2}}$ is a Hilbert-Schmidt operator for every $h \in H$. In addition,

  $$\left\| \mathscr{N}(h)\mathscr{Q}^{\frac{1}{2}} \right\|_{L_{HS}(U,H)} \leq \tilde{M} \|h\|_H \tag{4.4}$$

  holds for some constant $\tilde{M} > 0$, where L_{HS} indicates the Hilbert-Schmidt norm.
- The process $u : \mathbb{R}_+ \times \Omega \to \mathbb{R}^m$ is $(\mathscr{F}_t)_{t\geq 0}$-adapted with

  $$\int_0^T \mathbb{E} \|u(s)\|_2^2 \, ds < \infty \tag{4.5}$$

  for each $T > 0$, where $\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^m$.
- $\mathscr{B}$ is a linear and bounded operator on $\mathbb{R}^m$ with values in H and $\mathscr{C} \in L(H, \mathbb{R}^p)$.

[1] $(\mathscr{F}_t)_{t\geq 0}$ is right-continuous and complete.

Remark. *Theoretically, the control u can be chosen infinite dimensional and everything that follows below can be shown as well. There are numerical reasons to assume u to be finite dimensional. In the context of (balancing related) model order reduction it is often required to just have a low number of inputs to keep the computational costs low.*

Definition 4.8. *An $(\mathscr{F}_t)_{t\geq 0}$-adapted cadlag process $(\mathscr{X}(t))_{t\geq 0}$ with values in H is called* mild solution *of (4.3) if $\mathbb{P}$-almost surely*

$$\mathscr{X}(t) = S(t)X_0 + \int_0^t S(t-s)\mathscr{B}u(s)ds + \int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s) \tag{4.6}$$

holds for all $t \geq 0$.

We now show that the concept of mild solutions for equation (4.3) is well-defined. Furthermore, we prove the existence and uniqueness of a cadlag mild solution for a fixed control. Therefore, we state the following lemmas.

Lemma 4.9. *Let $\Psi : [0,T]\times\Omega \to H$ be a $\mathscr{P}_T/\mathscr{B}(H)$-measurable process, then*

$$[0,T]\times\Omega \ni (s,\omega) \mapsto \chi_{[0,t)}(s)S(t-s)\Psi(s,\omega)$$

is a $\mathscr{P}_T/\mathscr{B}(H)$-measurable mapping for all $t\in[0,T]$.

Proof. Knoche, Frieler [44] state the proof in Lemma 3.5. □

Applying Lemma 4.9, the process

$$\int_0^t S(t-s)\Psi(s)dM(s), \quad t\in[0,T],$$

is well-defined for processes $\Psi \in \mathscr{L}_T^2$ since

$$\mathbb{E}\int_0^t \left\|S(t-s)\Psi(s)Q^{\frac{1}{2}}\right\|^2_{L_{(HS)}(U,H)} ds \leq K^2\,\mathrm{e}^{2\hat{\alpha}T}\,\mathbb{E}\int_0^T \left\|\Psi(s)Q^{\frac{1}{2}}\right\|^2_{L_{(HS)}(U,H)} ds < \infty$$

by using inequality (4.1). In particular, $K=1$ and $\hat{\alpha}=0$ because $(S(t))_{t\geq 0}$ is a contraction.

Lemma 4.10. *Let Ψ be an $(\mathscr{F}_t)_{t\in[0,T]}$-adapted process with values in H which is $\mathbb{P}$-almost surely Bochner integrable, then the process*

$$\int_0^t S(t-s)\Psi(s)ds, \quad t\in[0,T],$$

is well-defined, $\mathbb{P}$*-almost surely continuous and* $(\mathscr{F}_t)_{t\in[0,T]}$*-adapted.*

Proof. The proof is similar to the one stated in Lemma 3.9 in [44]. $\square$

Lemma 4.11. *Let* $(\Psi(t)_{t\in[0,T]})$ *be an* H*-valued square integrable martingale with respect to a filtration* $(\mathscr{F}_t)_{t\in[0,T]}$ *which is continuous in probability, then* Ψ *has a cadlag modification.*

Proof. For a proof, see [55] (Theorem 3.41). $\square$

Theorem 4.12. *For a fixed adapted control* u *fulfilling (4.5), there is a unique mild solution* $(\mathscr{X}(t))_{t\geq 0}$ *to equation (4.3) satisfying*

$$\sup_{t\in[0,T]} \mathbb{E}\,\|\mathscr{X}(t)\|_H^2 < \infty.$$

Proof. Let $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$ be the Banach space of $(\mathscr{F}_t)_{t\in[0,T]}$-adapted cadlag processes $\mathscr{X}$ satisfying $\|\mathscr{X}\|_{\mathbb{B}}^2 := \sup_{t\in[0,T]} \mathbb{E}\,\|\mathscr{X}(t)\|_H^2 < \infty$. We assume that two processes are equal in $\mathbb{B}$ if they coincide $dt\otimes\mathbb{P}$-almost surely. We define a mapping

$$\rho(\mathscr{X})(t) = S(t)X_0 + \int_0^t S(t-s)\mathscr{B}u(s)ds + \int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s)$$

for $\mathscr{X}\in\mathbb{B}$ and $t\in[0,T]$ show that it is well-defined. The mapping

$$[0,T]\ni t\mapsto S(t)X_0$$

is continuous and hence cadlag. Since it is deterministic, obviously it is $(\mathscr{F}_t)_{t\in[0,T]}$-adapted. Due to the contraction property of $(S(t))_{t\geq 0}$, we have

$$\sup_{t\in[0,T]} \mathbb{E}\,\|S(t)X_0\|_H^2 \leq \|X_0\|_H^2 < \infty.$$

The process $\mathscr{B}u(s)$, $s\in[0,T]$, is $(\mathscr{F}_t)_{t\in[0,T]}$-adapted because the control u is adapted. This process is $\mathbb{P}$-almost surely Bochner integrable since

$$\begin{aligned}\mathbb{E}\left[\int_0^t \|\mathscr{B}u(s)\|_H\,ds\right] &\leq \|\mathscr{B}\|_{L(\mathbb{R}^m,H)} \int_0^T \mathbb{E}\,\|u(s)\|_2\,ds \\ &\leq \|\mathscr{B}\|_{L(\mathbb{R}^m,H)}\, T^{\frac{1}{2}} \left(\int_0^T \mathbb{E}\,\|u(s)\|_2^2\,ds\right)^{\frac{1}{2}} < \infty\end{aligned}$$

by using the Cauchy-Schwarz inequality. Hence, the process

$$\int_0^t S(t-s)\mathscr{B}u(s)ds, \quad t \in [0,T], \tag{4.7}$$

well-defined and cadlag adapted by Lemma 4.10. We proceed with the process in (4.7). It is in $\mathbb{B}$ because by the Cauchy-Schwarz inequality, we obtain

$$\sup_{t\in[0,T]} \mathbb{E}\left\|\int_0^t S(t-s)\mathscr{B}u(s)ds\right\|_H^2 \leq \sup_{t\in[0,T]} t\int_0^t \mathbb{E}\,\|S(t-s)\mathscr{B}u(s)\|_H^2\,ds$$

and furthermore, we have

$$\sup_{t\in[0,T]} \mathbb{E}\left\|\int_0^t S(t-s)\mathscr{B}u(s)ds\right\|_H^2 \leq \|\mathscr{B}\|_{L(\mathbb{R}^m,H)}^2\, T\int_0^T \mathbb{E}\,\|u(s)\|_2^2\,ds < \infty.$$

We now focus on the stochastic integral. By the remarks below Lemma 4.9, the process

$$\left(\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s)\right)_{t\in[0,T]} \tag{4.8}$$

is well-defined if the process

$$\mathscr{N}(\mathscr{X}(t-)), \quad t \in [0,T], \tag{4.9}$$

is in the space $\mathscr{L}_T^2$. Since X is cadlag, the process of the left limits $(\mathscr{X}(t-))_{t\in[0,T]}$ is obviously H-predictable and hence the process in (4.9) is H-predictable. Due to inequality (4.4) and the fact that $\mathscr{X}$ and the process of left limits coincide almost everywhere with respect to the Lebesgue measure, we have

$$\mathbb{E}\int_0^t \left\|\mathscr{N}(\mathscr{X}(s-))Q^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds \leq \tilde{M}^2\mathbb{E}\int_0^T \|X(s)\|_H^2\,ds \leq \tilde{M}^2 T\sup_{t\in[0,T]} \mathbb{E}\,\|X(s)\|_H^2 < \infty.$$

In order to show that the stochastic convolution in (4.8) has a cadlag modification we argue as in Theorem 9.24 in [55]. By Theorem 4.5 the contraction semigroup S can be written in the form $S(t) = \mathrm{Pr}_H\,\hat{S}(t)$, $t \geq 0$, where $\hat{S}$ is a strongly continuous group on some larger Hilbert space $\hat{H}$

and Pr_H is the orthogonal projection of $\hat{H}$ onto H. Thus, for $t \in [0,T]$ we have

$$\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s) = \mathrm{Pr}_H \hat{S}(t) \int_0^t \hat{S}(-s)\mathscr{N}(\mathscr{X}(s-))dM(s).$$

Since $S(-s)\mathscr{N}(\mathscr{X}(s-))$, $s \in [0,T]$, belongs to $\mathscr{L}_T^2$, the process

$$\int_0^t \hat{S}(-s)\mathscr{N}(\mathscr{X}(s-))dM(s), \quad t \in [0,T], \tag{4.10}$$

is a square integrable martingale with respect to $(\mathscr{F}_t)_{t\in[0,T]}$ by the results we have obtained in Section 3.1. It is also continuous in probability by Theorem 3.8 such that with Lemma 4.11 we conclude that (4.10) has a cadlag modification. Because $\hat{S}$ is strongly continuous, the process (4.8) has a cadlag modification as well.

We now show that the well-defined mapping ρ has exactly one fixed point in $\mathbb{B}$. This is a consequence of the Banach fixed point theorem[2] if we can prove that ρ is a contraction, i.e. the mapping ρ Lipschitz continuous with a Lipschitz constant less than 1. To ensure this, we use arguments of the proofs of Theorem 3.2 in [44] and Theorem 9.29 in [55]. Following this references, we introduce a family of norms $\|\cdot\|_{\mathbb{B},\beta}$, $\beta > 0$, which for $\mathscr{X} \in \mathbb{B}$ is given by

$$\|\mathscr{X}\|_{\mathbb{B},\beta}^2 := \sup_{t\in[0,T]} \mathrm{e}^{-\beta t}\,\mathbb{E}\,\|\mathscr{X}(t)\|_H^2 .$$

This norms are equivalent to $\|\cdot\|_{\mathbb{B}}$ since

$$\|\mathscr{X}\|_{\mathbb{B},\beta} \leq \|\mathscr{X}\|_{\mathbb{B}} \leq \mathrm{e}^{\frac{\beta T}{2}}\,\|\mathscr{X}\|_{\mathbb{B},\beta} .$$

Below, let $\mathscr{X}_1, \mathscr{X}_2 \in \mathbb{B}$, then it holds

$$\begin{aligned}
&\mathbb{E}\left\|\int_0^t S(t-s)\left(\mathscr{N}(\mathscr{X}_1(s-)) - \mathscr{N}(\mathscr{X}_2(s-))\right)dM(s)\right\|_H^2 \\
&= \int_0^t \mathbb{E}\left\|S(t-s)\mathscr{N}(\mathscr{X}_1(s-) - \mathscr{X}_2(s-))Q^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds \\
&\leq \int_0^t \mathbb{E}\left\|\mathscr{N}(\mathscr{X}_1(s) - \mathscr{X}_2(s))Q^{\frac{1}{2}}\right\|_{L_{(HS)}(U,H)}^2 ds \\
&\leq \tilde{M}^2 \int_0^t \mathbb{E}\,\|\mathscr{X}_1(s) - \mathscr{X}_2(s)\|_H^2\, ds
\end{aligned}$$

[2] We refer to Section IV.7 in [71].

$$\leq \tilde{M}^2 \int_0^t \mathrm{e}^{\beta s}\, ds\, \|\mathscr{X}_1 - \mathscr{X}_2\|^2_{\mathbb{B},\beta}$$
$$\leq \tilde{M}^2 \frac{1}{\beta} \mathrm{e}^{\beta t}\, \|\mathscr{X}_1 - \mathscr{X}_2\|^2_{\mathbb{B},\beta}$$

and thus

$$\|\rho(\mathscr{X}_1) - \rho(\mathscr{X}_2)\|_{\mathbb{B},\beta} = \left(\sup_{t\in[0,T]} \mathrm{e}^{-\beta t}\, \mathbb{E} \left\| \int_0^t S(t-s)\,(\mathscr{N}(\mathscr{X}_1(s-)) - \mathscr{N}(\mathscr{X}_2(s-)))\, dM(s) \right\|_H^2 \right)^{\frac{1}{2}}$$
$$\leq \tilde{M} \left(\frac{1}{\beta}\right)^{\frac{1}{2}} \|\mathscr{X}_1 - \mathscr{X}_2\|_{B,\beta}\,.$$

Choosing $\beta > \tilde{M}^2$ yields the desired contraction property such that we have unique fixed point of ρ in $\mathbb{B}$. □

Remark.

- *System (4.3) can be easily extended by adding a term in the drift part that is Lipschitz continuous in $\mathscr{X}$ and by considering more general Lipschitz continuous functions $\mathscr{N}$. For existence and uniqueness results we refer to [44] and [55], where these assumptions are used in an uncontrolled setting with a predictable mild solution concept.*

- *The linear operator $\mathscr{A}$ in equation (4.3) can be replaced by a generator of a generalized contraction semigroup.*

- *Even arbitrary generators $\mathscr{A}$ can be used, see Theorem 9.29 in [55], where the required techniques for a proof are used in a different framework. In this case, one has to modify the solution concept. Instead of constructing cadlag adapted mild solutions one has to switch to predictable solutions since the stochastic convolution in equation (4.6) has no cadlag modification for general $\mathscr{A}$. So, the left limit in the diffusion part of (4.3) has to be replaced by $\mathscr{X}(s)$.*

4.3. Examples

In this section, we provide two examples that are covered by the system (4.3). In the first subsection, we introduce a framework that includes the stochastic heat equation and in the second subsection we discuss the stochastic damped wave equation. Here, we keep the notation that is used in Section 4.2. For more examples with general generator $\mathscr{A}$ we refer to [55].

4.3.1. Stochastic heat equation

We suppose that $\mathscr{A} : D(\mathscr{A}) \to H$ is a densely defined linear operator being self adjoint and negative definite such that we have an orthonormal basis $(h_k)_{k\in\mathbb{N}}$ of H consisting of eigenvectors of $\mathscr{A}$:

$$\mathscr{A} h_k = -\lambda_k h_k, \tag{4.11}$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots$ are the corresponding eigenvalues. We then know that the linear operator $\mathscr{A}$ generates a contraction C_0-semigroup $(S(t))_{t\geq 0}$ defined by

$$S(t)x = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t} \langle x, h_k \rangle h_k \tag{4.12}$$

for $x \in H$. It is exponentially stable ($\hat{\alpha} < 0$ in (4.1)) for the case $0 < \lambda_1$. An important example that satisfies the assumptions on the generator is $\mathscr{A} = \Delta$ which is the heat equation case. Below, we state two examples in order to demonstrate what is covered by the abstract setting in Section 4.2.

Example 4.13. *We consider a bar of length π which is heated on $[0, \frac{\pi}{2}]$. Moreover, there is ice at the boundary. So, the temperature of the bar is described by the following stochastic partial differential equation, where the noise can be interpreted as a random heat source or wind that affects the bar:*

$$\begin{aligned} \frac{\partial \mathscr{X}(t,\zeta)}{\partial t} &= \frac{\partial^2}{\partial \zeta^2} \mathscr{X}(t,\zeta) + 1_{[0,\frac{\pi}{2}]}(\zeta) u(t) + a \mathscr{X}(t-,\zeta) \frac{\partial M(t)}{\partial t}, \qquad (4.13)\\ \mathscr{X}(t,0) &= 0 = \mathscr{X}(t,\pi), \\ \mathscr{X}(0,\zeta) &= X_0(\zeta) \end{aligned}$$

for $a \in \mathbb{R}$, $t \geq 0$ and $\zeta \in [0,\pi]$. Here, we assume that

- *M is a scalar square integrable Lévy process with zero mean,*
- *$H = L^2([0,\pi])$, $U = \mathbb{R}$, $m = 1$,*
- *$\mathscr{A} = \frac{\partial^2}{\partial \zeta^2}$, $\mathscr{B} = 1_{[0,\frac{\pi}{2}]}(\cdot)$ and $\mathscr{N}(x) = ax$ for $x \in L^2([0,\pi])$.*

It is a well-known fact that here the eigenvalues of the second derivative are given by the sequence

$-\lambda_k = -k^2$, $k \in \mathbb{N}$, and the corresponding eigenvectors which represent an orthonormal basis are $h_k = \sqrt{\frac{2}{\pi}}\sin(k\cdot)$.

If furthermore $\mathbb{E}\left[M(1)^2\right] a^2 < 2$, then the solution $\mathscr{X}^h$ of the uncontrolled SPDE (4.13) is exponentially mean square stable, i.e.

$$\mathbb{E}\left\|\mathscr{X}^h(t,\cdot)\right\|_H^2 \leq c\,\mathrm{e}^{-\tilde{\alpha}t}\left\|X_0(\cdot)\right\|_H^2 \tag{4.14}$$

for $c, \tilde{\alpha} > 0$. This is a consequence of Theorem 3.1 in Ichikawa [38] and Theorem 5 in Haussmann [30]. For further information regarding the exponential mean square stability condition (4.14), see Section 5 in Curtain [18].[3]

We are interested in the average temperature of the bar on $[\frac{\pi}{2}, \pi]$ such that the scalar output of the system is

$$\mathscr{Y}(t) = \frac{2}{\pi}\int_{\frac{\pi}{2}}^{\pi}\mathscr{X}(t,\zeta)d\zeta, \tag{4.15}$$

where $\mathscr{C}x = \frac{2}{\pi}\int_{\frac{\pi}{2}}^{\pi}x(\zeta)d\zeta$ for $x \in L^2([0,\pi])$.

Next, we consider a more complex example with a two dimensional spatial variable and Neumann boundary conditions.

Example 4.14. *We consider a two dimensional surface with perfect insulation at the boundary which is heated in the middle. This can be modeled by the following controlled stochastic partial differential equation with $t \geq 0$ and $\zeta \in [0,\pi]^2$:*

$$\begin{aligned}
\frac{\partial \mathscr{X}(t,\zeta)}{\partial t} &= \Delta\mathscr{X}(t,\zeta) + 1_{[\frac{\pi}{4},\frac{3\pi}{4}]^2}(\zeta)u(t) + \mathrm{e}^{-\left|\zeta_1-\frac{\pi}{2}\right|-\zeta_2}\mathscr{X}(t-,\zeta)\frac{\partial M(t)}{\partial t}, \\
\frac{\partial \mathscr{X}(t,\zeta)}{\partial \mathbf{n}} &= 0, \quad t \geq 0,\ \zeta \in \partial[0,\pi]^2, \\
\mathscr{X}(0,\zeta) &\equiv 0.
\end{aligned} \tag{4.16}$$

Again, M is a scalar square integrable Lévy process with zero mean which can model a random heat source or the impact of wind.

Further, we set

[3] Curtain, Ichikawa and Haussmann stated these conditions for exponential mean square stability for the Wiener case which can be easily generalized for the case of square integrable Lévy processes with mean zero.

- $H = L^2([0,\pi]^2)$, $U = \mathbb{R}$, $m = 1$,
- $\mathscr{A}$ *is the Laplace operator,* $\mathscr{B} = 1_{[\frac{\pi}{4},\frac{3\pi}{4}]^2}(\cdot)$ *and*
- $\mathscr{N}(x) = \mathrm{e}^{-|\cdot-\frac{\pi}{2}|-\cdot}\, x$ *for* $x \in L^2([0,\pi]^2)$.

The eigenvalues of the Laplacian on $[0,\pi]^2$ *are given by* $-\lambda_{ij} = -(i^2+j^2)$, $i,j \in \mathbb{Z}_+$, *and the corresponding eigenvectors which represent an orthonormal basis are* $h_{ij} = \frac{f_{ij}}{\|f_{ij}\|_H}$, *where* $f_{ij} = \cos(i\cdot)\cos(j\cdot)$. *For simplicity, to ensure the form given in (4.11), we write* $-\lambda_k$, $k \in \mathbb{N}$, *for the k-th largest eigenvalue and we denote the corresponding eigenvector by* h_k.

The scalar output of the system is the average temperature on the non heated area

$$\mathscr{Y}(t) = \frac{4}{3\pi^2}\int_{[0,\pi]^2\setminus[\frac{\pi}{4},\frac{3\pi}{4}]^2} \mathscr{X}(t,\zeta)d\zeta, \tag{4.17}$$

where $\mathscr{C}x = \frac{4}{3\pi^2}\int_{[0,\pi]^2\setminus[\frac{\pi}{4},\frac{3\pi}{4}]^2} x(\zeta)d\zeta$ *for* $x \in L^2([0,\pi]^2)$.

Remark.

- *The SPDEs in Examples 4.13 and 4.14 are only symbolic equations. They have to be interpreted in the sense of a mild solution defined in (4.6).*
- *In the Examples 4.13 and 4.14, the noise process* M *can be infinite dimensional as well. So, the linear mapping* $\mathscr{N}$ *can for example have the form*

  $$\mathscr{N}(x) = \sum_{k=1}^{\infty} \nu_k \langle h_k, x\rangle_H h_k \langle \cdot, u_k\rangle_U ,$$

  where $x \in H$, $(u_k)_{k\in\mathbb{N}}$ *is an orthonormal basis of* U *and* $\sum_{k=1}^{\infty}\nu_k^2 < \infty$. *Since* $\mathscr{N}(x) \in L(U,H)$, *the corresponding SPDE would be well defined.*

4.3.2. Stochastic damped wave equation

In this section, we introduce a damped wave equation with Lévy noise. It is possible to transform this SPDE into a first order system (4.3) following the approach in [19].

Let M_1 and M_2 be independent scalar square integrable Lévy processes with zero mean being defined on a complete probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq0},\mathbb{P})$.[4] In addition, we assume M_k ($k = 1,2$) to be $(\mathscr{F}_t)_{t\geq0}$-adapted and the increments $M_k(t+h)-M_k(t)$ to be independent of $\mathscr{F}_t$ for

[4] We assume that $(\mathscr{F}_t)_{t\geq0}$ is right-continuous and that $\mathscr{F}_0$ contains all $\mathbb{P}$ null sets.

$t,h \geq 0$.
Suppose $\tilde{\mathscr{A}} : D(\tilde{\mathscr{A}}) \to \tilde{H}$ is a self adjoint and positive definite operator such that we can choose an orthonormal basis $\left(\tilde{h}_k\right)_{k\in\mathbb{N}}$ of the separable Hilbert space $\tilde{H}$ consisting of eigenvectors of $\tilde{\mathscr{A}}$:

$$\tilde{\mathscr{A}}\tilde{h}_k = \tilde{\lambda}_k \tilde{h}_k,$$

where $0 < \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \ldots$ are the corresponding eigenvalues. We denote the well defined square root of $\tilde{\mathscr{A}}$ by $\tilde{\mathscr{A}}^{\frac{1}{2}}$. $D(\tilde{\mathscr{A}}^{\frac{1}{2}})$ equipped with the inner product $\langle x,y\rangle_{D(\tilde{\mathscr{A}}^{\frac{1}{2}})} = \left\langle \tilde{\mathscr{A}}^{\frac{1}{2}}x, \tilde{\mathscr{A}}^{\frac{1}{2}}y\right\rangle_{\tilde{H}}$ represents a Hilbert space. In this case, the norm $\|\cdot\|_{D(\tilde{\mathscr{A}}^{\frac{1}{2}})}$ is equivalent to the graph norm of the linear operator $\tilde{\mathscr{A}}^{\frac{1}{2}}$.

The equation that we consider next is also studied by Curtain in [19] for M_1, M_2 being Wiener processes and $u \equiv 0$. There, the exponential mean square stability is analyzed for example. The system we focus on is the following (symbolic) second order stochastic differential equation:

$$\ddot{\mathscr{Z}}(t) + \alpha\dot{\mathscr{Z}}(t) + \tilde{\mathscr{A}}\mathscr{Z}(t) + \tilde{\mathscr{B}}u(t) + \tilde{\mathscr{D}}_1\mathscr{Z}(t-)\dot{M}_1(t) + \tilde{\mathscr{D}}_2\dot{\mathscr{Z}}(t-)\dot{M}_2(t) = 0 \tag{4.18}$$

with initial conditions $\mathscr{Z}(0) = z_0$, $\dot{\mathscr{Z}}(0) = z_1$ and output equation

$$\mathscr{Y}(t) = \mathscr{C}\begin{pmatrix}\mathscr{Z}(t)\\ \dot{\mathscr{Z}}(t)\end{pmatrix}, \quad t \geq 0.$$

Above, we make the following assumptions:

- The constant α is positive, $\tilde{\mathscr{D}}_1 \in L(D(\tilde{\mathscr{A}}^{\frac{1}{2}}), \tilde{H})$ and $\tilde{\mathscr{D}}_2 \in L(\tilde{H})$.
- $\tilde{\mathscr{B}}$ is a linear and bounded operator on $\mathbb{R}^m$ with values in $\tilde{H}$ and $\mathscr{C} \in L(D(\tilde{\mathscr{A}}^{\frac{1}{2}}) \times \tilde{H}, \mathbb{R}^p)$.

We introduce the Hilbert space $H = D(\tilde{\mathscr{A}}^{\frac{1}{2}}) \times \tilde{H}$ equipped with the inner product

$$\left\langle \begin{pmatrix}\tilde{Z}_1\\ \tilde{Z}_2\end{pmatrix}, \begin{pmatrix}\bar{Z}_1\\ \bar{Z}_2\end{pmatrix}\right\rangle_H = \left\langle \tilde{\mathscr{A}}^{\frac{1}{2}}\tilde{Z}_1, \tilde{\mathscr{A}}^{\frac{1}{2}}\bar{Z}_1\right\rangle_{\tilde{H}} + \langle \tilde{Z}_2, \bar{Z}_2\rangle_{\tilde{H}}.$$

An orthonormal basis of H is given by $(h_k)_{k\in\mathbb{N}}$ defined by

$$h_{2i-1} = \tilde{\lambda}_i^{-\frac{1}{2}}\begin{pmatrix}\tilde{h}_i\\ 0\end{pmatrix} \text{ and } h_{2i} = \begin{pmatrix}0\\ \tilde{h}_i\end{pmatrix} \tag{4.19}$$

for $i \in \mathbb{N}$. The second order equation (4.18) can be expressed by the following first order system:

$$\begin{aligned} d\mathscr{X}(t) &= \mathscr{A}\mathscr{X}(t) + \mathscr{B}u(t)dt + \mathscr{D}_1\mathscr{X}(t-)dM_1(t) + \mathscr{D}_2\mathscr{X}(t-)dM_2(t), \\ \mathscr{Y}(t) &= \mathscr{C}\mathscr{X}(t), \quad t \geq 0, \quad \mathscr{X}(0) = X_0 = \left(\begin{smallmatrix} z_0 \\ z_1 \end{smallmatrix}\right), \end{aligned} \tag{4.20}$$

where

$$\mathscr{X}(t) = \begin{pmatrix} \mathscr{Z}(t) \\ \dot{\mathscr{Z}}(t) \end{pmatrix}, \mathscr{A} = \begin{bmatrix} 0 & I \\ -\tilde{\mathscr{A}} & -\alpha I \end{bmatrix}, \mathscr{B} = \begin{bmatrix} 0 \\ -\tilde{\mathscr{B}} \end{bmatrix}, \mathscr{D}_1 = \begin{bmatrix} 0 & 0 \\ -\tilde{\mathscr{D}}_1 & 0 \end{bmatrix} \text{ and } \mathscr{D}_2 = \begin{bmatrix} 0 & 0 \\ 0 & -\tilde{\mathscr{D}}_2 \end{bmatrix}.$$

Regarding this transformation we follow [19], where it is used as well. The next lemma from [57] provides a stability result and is also needed to define a cadlag mild solution of (4.20).

Lemma 4.15. *For every $\alpha > 0$ the linear operator $\mathscr{A}$ with domain $D(\tilde{\mathscr{A}}) \times D(\tilde{\mathscr{A}}^{\frac{1}{2}})$ generates an exponential stable contraction semigroup $(S(t))_{t\geq 0}$ with*

$$\|S(t)\|_{L(H)} \leq \mathrm{e}^{-ct},$$

where

$$c \geq \frac{2\alpha\tilde{\lambda}_1}{4\tilde{\lambda}_1 + \alpha(\alpha + \sqrt{\alpha^2 + 4\tilde{\lambda}_1})}.$$

Of course, system (4.20) can be transformed into the form we have in (4.3) by setting

$$M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \quad \text{and} \quad \mathscr{N}(x) = \begin{bmatrix} \mathscr{D}_1 x & \mathscr{D}_2 x \end{bmatrix},$$

where $x \in H$ and $U = \mathbb{R}^2$. Sometimes we use the alternative representation for the $(\mathscr{F}_t)_{t\geq 0}$-adapted cadlag mild solution $(\mathscr{X}(t))_{t\geq 0}$ of (4.20) which for all $t \geq 0$ is

$$\mathscr{X}(t) = S(t)X_0 + \int_0^t S(t-s)\mathscr{B}u(s)ds + \sum_{i=1}^{2} \int_0^t S(t-s)\mathscr{D}_i\mathscr{X}(s-)dM_i(s). \tag{4.21}$$

We present an example for system (4.18) to conclude this subsection. This example is just a symbolic SPDE which is defined in the sense of (4.21)

Example 4.16. *The lateral displacement of an electricity cable impacted by wind can be modeled* *by the following second order SPDE:*

$$\frac{\partial^2}{\partial t^2}\mathscr{X}(t,\zeta)+\alpha\frac{\partial}{\partial t}\mathscr{X}(t,\zeta)=\frac{\partial^2}{\partial\zeta^2}\mathscr{X}(t,\zeta)+\mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}u(t)+2\,\mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}\,\mathscr{X}(t-,\zeta)\frac{\partial}{\partial t}M_1(t)$$

for $t\in[0,T]$ *and* $\zeta\in[0,\pi]$. *Here, we have*

- $\tilde{\mathscr{A}}=-\frac{\partial^2}{\partial\zeta^2}$, *the operator* $\tilde{\mathscr{B}}$ *is represented by the function* $-\,\mathrm{e}^{-(\cdot-\frac{\pi}{2})^2}$,
- $\tilde{\mathscr{D}}_2=0$, $\tilde{\mathscr{D}}_1$ *is characterized by* $-2\,\mathrm{e}^{-(\cdot-\frac{\pi}{2})^2}$ *and*
- $\tilde{H}=L^2([0,\pi])$, $D(\tilde{\mathscr{A}}^{\frac{1}{2}})=H_0^1([0,\pi])$, $m=1$.

The boundary and initial conditions are

$$\mathscr{X}(0,t)=0=\mathscr{X}(\pi,t)\quad\text{and}\quad\mathscr{X}(0,\zeta),\,\frac{\partial}{\partial t}\mathscr{X}(t,\zeta)\bigg|_{t=0}\equiv 0.$$

The output is an approximation for the position of the middle of the string

$$\mathscr{Y}(t)=\frac{1}{2\varepsilon}\int_{\frac{\pi}{2}-\varepsilon}^{\frac{\pi}{2}+\varepsilon}\mathscr{X}(t,\zeta)d\zeta,$$

where $\varepsilon>0$ *is small. Here, we set* $\mathscr{C}=\begin{bmatrix}\hat{\mathscr{C}} & 0\end{bmatrix}$ *with* $\hat{\mathscr{C}}x=\frac{1}{2\varepsilon}\int_{\frac{\pi}{2}-\varepsilon}^{\frac{\pi}{2}+\varepsilon}x(\zeta)d\zeta$ $(x\in D(\tilde{\mathscr{A}}^{\frac{1}{2}}))$, *such that* $p=1$.

4.4. Conclusions

In this chapter, we gave an insight into the C_0-semigroup theory before we introduced linear controlled SPDEs with Lévy noise in an abstract Hilbert space setting. We discussed the existence and uniqueness of cadlag mild solutions for these SPDEs and considered two specific examples, namely the heat and the damped wave equation with Lévy noise.

5. Numerical Approximations for Linear SPDEs with Lévy Noise

In this chapter, we investigate a Galerkin scheme that is for example studied by Grecksch, Kloeden [27] and Jentzen, Kloeden [40] for particular SPDEs with Wiener noise. One finds a more general setting in Hausenblas [29]. This chapter is manly based on Benner, Redmann [14] (Section 5.1) and [60] (Section 3).
In [27] and [40] the stochastic heat equation driven by Wiener noise is covered. We make use of these techniques in proving the convergence of the numerical solutions in the case of stochastic heat equations with Lévy noise in Section 5.1. Furthermore, we derive the Galerkin solution for particular examples. In Section 5.2, we apply similar ideas to a completely different problem, i.e. we investigate Galerkin methods for second order systems and apply it to particular examples.
Finally, we illustrate the numerical solutions in a plot, where all the numerical experiments are run on a desktop computer with a dual-core Intel Pentium processor E5400 and 3GB RAM. All algorithms are implemented and executed in MATLAB 7.14.0.739 (R2012a) running on Ubuntu 10.04.1 LTS.

In this chapter, we semi-discretize the type of equation we introduce in Section 4.2. Below, we assume the existence of an orthonormal basis $(h_k)_{k\in\mathbb{N}}$ of H which belongs to $D(\mathscr{A})$. For the separable Hilbert space U, we choose an orthonormal basis $(u_k)_{k\in\mathbb{N}}$ which consists of eigenvectors of the covariance operator $\mathscr{Q}$[1] of M, see equation (4.2). We denote the corresponding eigenvalues by $(\mu_k)_{k\in\mathbb{N}}$ such that

$$\mathscr{Q}u_k = \mu_k u_k.$$

We then approximate the mild solution of the infinite dimensional equation (4.3) for special cases. Therefore, we construct a sequence $(X_n)_{n\in\mathbb{N}}$ of finite dimensional adapted cadlag pro-

[1]By Theorem VI.21 in Reed, Simon [62], $\mathscr{Q}$ is a compact operator such that this property follows by the spectral theorem.

cesses with values in $H_n = \operatorname{span}\{h_1, \ldots, h_n\}$ given by

$$\begin{aligned} dX_n(t) &= [\mathscr{A}_n X_n(t) + \mathscr{B}_n u(t)]\,dt + \mathscr{N}_n(X_n(t-))dM_n(t), \quad t \geq 0, \\ X_n(0) &= X_{0,n}, \end{aligned} \tag{5.1}$$

where we define

- $M_n(t) = \sum_{k=1}^n \langle M(t), u_k \rangle_U u_k$, $t \geq 0$, is a $\operatorname{span}\{u_1, \ldots, u_n\}$-valued Lévy process,
- $\mathscr{A}_n x = \sum_{k=1}^n \langle \mathscr{A} x, h_k \rangle_H h_k \in H_n$ holds for all $x \in D(\mathscr{A})$,
- $\mathscr{B}_n x = \sum_{k=1}^n \langle \mathscr{B} x, h_k \rangle_H h_k \in H_n$ holds for all $x \in \mathbb{R}^m$,
- $\mathscr{N}_n(x)y = \sum_{k=1}^n \langle \mathscr{N}(x)y, h_k \rangle_H h_k \in H_n$ holds for all $y \in U$ and $x \in H$,
- $X_{0,n} = \sum_{k=1}^n \langle X_0, h_k \rangle_H h_k \in H_n$.

5.1. Approximation of stochastic heat equations

In this section, we discuss a Galerkin scheme for the special framework in Subsection 4.3.1 which covers the heat equation. So, we choose an orthonormal basis $(h_k)_{k\in\mathbb{N}}$ of eigenvectors of $\mathscr{A}$, see equation (4.11). The advantage of this case is that we can use the special structure of the C_0-semigroup in (4.12).

Since $\mathscr{A}_n$ is a bounded operator for every $n \in \mathbb{N}$, we know that $\mathscr{A}_n$ generates a C_0-semigroup on H_n of the form $S_n(t) = \mathrm{e}^{A_n t}$, $t \geq 0$. For all $x \in H_n$, it furthermore has the representation $S_n(t)x = \sum_{k=1}^n \mathrm{e}^{-\lambda_k t} \langle x, h_k \rangle_H h_k$. Hence, $\|(S_n(t) - S(t))x\|_H \to 0$ for $n \to \infty$.

The mild solution of equation (5.1) is given by

$$X_n(t) = S_n(t)X_{0,n} + \int_0^t S_n(t-s)\mathscr{B}_n u(s)ds + \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s)$$

for $t \geq 0$. Furthermore, we consider the p-dimensional approximating output

$$y_n(t) = \mathscr{C} X_n(t), \quad t \geq 0.$$

Below, one can proof Theorem 5.1 with similar arguments as in [27] and [40]. This shows that the following is true for $n \to \infty$ and $t \geq 0$:

$$\mathbb{E}\,\|y_n(t) - \mathscr{Y}(t)\|_2^2 \to 0.$$

We furthermore provide the proof of the following theorem. This proof is missing in Theorem 5.2 of [14].

Theorem 5.1. *Under the assumptions on $\mathscr{A}$ we made in Subsection 4.3.1,*

$$\mathbb{E}\,\|X_n(t) - \mathscr{X}(t)\|_H^2 \to 0$$

holds for $n \to \infty$ and $t \geq 0$.

Proof.

$$\begin{aligned}\mathbb{E}\,\|X_n(t) - \mathscr{X}(t)\|_H^2 \leq{}& 3\mathbb{E}\left\|S(t)X_0 - S_n(t)X_{0,n}\right\|_H^2 \\ &+ 3\mathbb{E}\left\|\int_0^t (S(t-s)\mathscr{B} - S_n(t-s)\mathscr{B}_n)u(s)ds\right\|_H^2 \\ &+ 3\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s) - \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s)\right\|_H^2.\end{aligned}$$

Using $S(t)x = \sum_{i=1}^{\infty} \mathrm{e}^{-\lambda_i t}\langle x, h_i\rangle_H h_i$ $(x \in H)$ yields

$$\begin{aligned}3\mathbb{E}\left\|S(t)X_0 - S_n(t)X_{0,n}\right\|_H^2 &= 3\mathbb{E}\left\|\sum_{i=n+1}^{\infty} \mathrm{e}^{-\lambda_i t}\langle X_0, h_i\rangle_H h_i\right\|_H^2 = 3\mathbb{E}\sum_{i=n+1}^{\infty} \mathrm{e}^{-2\lambda_i t}\langle X_0, h_i\rangle_H^2 \\ &\leq 3\mathbb{E}\left\|X_0 - X_{0,n}\right\|_H^2. \end{aligned} \tag{5.2}$$

The Hölder inequality delivers

$$\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}_n u(s)ds\right\|_H^2 \leq t\mathbb{E}\int_0^t \|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}_n u(s)\|_H^2\,ds.$$

So, we have

$$\|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}_n u(s)\|_H^2 = \sum_{i=n+1}^{\infty} \mathrm{e}^{-2\lambda_i t}\langle \mathscr{B}u(s), h_i\rangle_H^2$$

$$\leq \|\mathscr{B}u(s) - \mathscr{B}_n u(s)\|_H^2 \to 0$$

$\mathbb{P}$- almost surely for $n \to \infty$ and

$$3t\mathbb{E}\int_0^t \|\mathscr{B}u(s) - \mathscr{B}_n u(s)\|_H^2 \, ds \to 0 \tag{5.3}$$

for $n \to \infty$ by Lebesgue's theorem. Furthermore, we obtain

$$\begin{aligned}
&3\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-))dM(s) - \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s)\right\|_H^2 \\
&\leq 6\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-)) - S_n(t-s)\mathscr{N}_n(X_n(s-))dM(s)\right\|_H^2 && (5.4)\\
&\quad + 6\mathbb{E}\left\|\int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM(s) - \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s)\right\|_H^2. && (5.5)
\end{aligned}$$

By applying the Ito isometry (see Section 3.1) to term (5.4), we have

$$\begin{aligned}
&6\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-)) - S_n(t-s)\mathscr{N}_n(X_n(s-))dM(s)\right\|_H^2 \\
&= 6\int_0^t \mathbb{E}\left\|(S(t-s)\mathscr{N}(\mathscr{X}(s-)) - S_n(t-s)\mathscr{N}_n(X_n(s-)))\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{HS}(U,H)}^2 ds.
\end{aligned}$$

Since a cadlag process has just finitely many jumps, we can replace the left limit processes by the original ones. Furthermore, we insert the definition of the Hilbert-Schmidt norm and representation of the C_0-semigroup and obtain

$$\begin{aligned}
&6\mathbb{E}\left\|\int_0^t S(t-s)\mathscr{N}(\mathscr{X}(s-)) - S_n(t-s)\mathscr{N}_n(X_n(s-))dM(s)\right\|_H^2 \\
&= 6\int_0^t \mathbb{E}\sum_{k=1}^\infty \left\|(S(t-s)\mathscr{N}(\mathscr{X}(s)) - S_n(t-s)\mathscr{N}_n(X_n(s)))\mathscr{Q}^{\frac{1}{2}}u_k\right\|_H^2 ds \\
&= 6\int_0^t \mathbb{E}\sum_{k=1}^\infty \left\|\sum_{i=1}^\infty \mathrm{e}^{-\lambda_i(t-s)}\langle\mathscr{N}(\mathscr{X}(s))\mathscr{Q}^{\frac{1}{2}}u_k, h_i\rangle_H h_i - \sum_{i=1}^n \mathrm{e}^{-\lambda_i(t-s)}\langle\mathscr{N}(X_n(s))\mathscr{Q}^{\frac{1}{2}}u_k, h_i\rangle_H h_i\right\|_H^2 ds \\
&\leq 12\int_0^t \mathbb{E}\sum_{k=1}^\infty\sum_{i=1}^n \mathrm{e}^{-2\lambda_i(t-s)}\langle\mathscr{N}(\mathscr{X}(s) - X_n(s))\mathscr{Q}^{\frac{1}{2}}u_k, h_i\rangle_H^2 ds \\
&\quad + 12\int_0^t \mathbb{E}\sum_{k=1}^\infty\sum_{i=n+1}^\infty \mathrm{e}^{-2\lambda_i(t-s)}\langle\mathscr{N}(\mathscr{X}(s))\mathscr{Q}^{\frac{1}{2}}u_k, h_i\rangle_H^2 ds
\end{aligned}$$

$$\leq 12 \int_0^t \mathbb{E} \left\| \mathcal{N}(\mathscr{X}(s) - X_n(s)) \mathscr{Q}^{\frac{1}{2}} \right\|_{L_{HS}(U,H)}^2 ds \tag{5.6}$$

$$+ 12 \int_0^t \mathbb{E} \left\| (\mathcal{N}(\mathscr{X}(s)) - \mathcal{N}_n(\mathscr{X}(s))) \mathscr{Q}^{\frac{1}{2}} \right\|_{L_{HS}(U,H)}^2 ds. \tag{5.7}$$

The term (5.7) converges to zero for $n \to \infty$ by Lebesgue's theorem. Moreover, for (5.6) we obtain

$$12 \int_0^t \mathbb{E} \left\| \mathcal{N}(\mathscr{X}(s) - X_n(s)) \mathscr{Q}^{\frac{1}{2}} \right\|_{L_{HS}(U,H)}^2 ds \leq 12 \tilde{M}^2 \int_0^t \mathbb{E} \left\| \mathscr{X}(s) - X_n(s) \right\|_H^2 ds \tag{5.8}$$

applying inequality (4.4). Using the Fourier series representation for M and M_n in (5.5), we have

$$6\mathbb{E} \left\| \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM(s) - \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM_n(s) \right\|_H^2 =$$
$$6\mathbb{E} \left\| \sum_{k=n+1}^{\infty} \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) u_k d \left\langle M(s), u_k \right\rangle_U \right\|_H^2 .$$

We define $M^k(s) := \langle M(s), u_k \rangle_U$, $s \geq 0$, and since M^i and M^j are uncorrelated processes ($i \neq j$), it follows

$$6\mathbb{E} \left\| \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM(s) - \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM_n(s) \right\|_H^2$$
$$= 6 \sum_{k=n+1}^{\infty} \mathbb{E} \left\| \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) u_k dM^k(s) \right\|_H^2 .$$

By the Ito isometry and by replacing the left limits, we obtain

$$6\mathbb{E} \left\| \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM(s) - \int_0^t S_n(t-s) \mathcal{N}_n(X_n(s-)) dM_n(s) \right\|_H^2$$
$$\leq 6 \sum_{k=n+1}^{\infty} \mu_k \int_0^t \mathbb{E} \left\| S_n(t-s) \mathcal{N}_n(X_n(s)) u_k \right\|_H^2 ds$$
$$\leq 6 \sum_{k=n+1}^{\infty} \mu_k \int_0^t \mathbb{E} \left\| S(t-s) \mathcal{N}(X_n(s)) u_k \right\|_H^2 ds$$
$$\leq 12 \sum_{k=1}^{\infty} \mu_k \int_0^t \mathbb{E} \left\| S(t-s) \mathcal{N}(\mathscr{X}(s) - X_n(s)) u_k \right\|_H^2 ds$$
$$+ 12 \sum_{k=n+1}^{\infty} \mu_k \int_0^t \mathbb{E} \left\| S(t-s) \mathcal{N}(\mathscr{X}(s)) u_k \right\|_H^2 ds$$

$$\begin{aligned}
&= 12\sum_{k=1}^{\infty}\int_0^t \mathbb{E}\left\|S(t-s)\mathcal{N}(\mathcal{X}(s)-X_n(s))\mathcal{Q}^{\frac{1}{2}}u_k\right\|_H^2 ds \\
&\quad + 12\sum_{k=n+1}^{\infty}\int_0^t \mathbb{E}\left\|S(t-s)\mathcal{N}(\mathcal{X}(s))\mathcal{Q}^{\frac{1}{2}}u_k\right\|_H^2 ds \\
&\le 12\int_0^t \mathbb{E}\left\|\mathcal{N}(\mathcal{X}(s)-X_n(s))\mathcal{Q}^{\frac{1}{2}}\right\|_{L_{HS}(U,H)}^2 ds \\
&\quad + 12\int_0^t \mathbb{E}\sum_{k=n+1}^{\infty}\left\|\mathcal{N}(\mathcal{X}(s))\mathcal{Q}^{\frac{1}{2}}u_k\right\|_H^2 ds.
\end{aligned}$$

By (4.4), we finally have

$$\begin{aligned}
&6\mathbb{E}\left\|\int_0^t S_n(t-s)\mathcal{N}_n(X_n(s-))dM(s) - \int_0^t S_n(t-s)\mathcal{N}_n(X_n(s-))dM_n(s)\right\|_H^2 \\
&\le 12\tilde{M}^2\int_0^t \mathbb{E}\left\|\mathcal{X}(s)-X_n(s)\right\|_H^2 ds \\
&\quad + 12\int_0^t \mathbb{E}\sum_{k=n+1}^{\infty}\left\|\mathcal{N}(\mathcal{X}(s))\mathcal{Q}^{\frac{1}{2}}u_k\right\|_H^2 ds. \qquad (5.9)
\end{aligned}$$

The term (5.9) converges to zero for $n\to\infty$ by Lebesgue's theorem. If $\mathcal{N}$ takes values in $L(U,H)$ the expression (5.9) is less or equal to

$$\mathbb{E}\left\|M_n(1)-M(1)\right\|_U^2 \, 12\int_0^t \mathbb{E}\left\|\mathcal{N}(\mathcal{X}(s))\right\|_{L(U,H)}^2 ds.$$

We define $f_n(t)$ as the sum of the terms in (5.2), (5.3), (5.7) and (5.9). Summarizing everything, we obtain

$$\mathbb{E}\left\|X_n(t)-\mathcal{X}(t)\right\|_H^2 \le f_n(t) + 24\tilde{M}^2\int_0^t \mathbb{E}\left\|\mathcal{X}(s)-X_n(s)\right\|_H^2 ds.$$

Hence,

$$\mathbb{E}\left\|X_n(t)-\mathcal{X}(t)\right\|_H^2 \le f_n(t)\,\mathrm{e}^{24\tilde{M}^2 t} \qquad (5.10)$$

by Gronwall's inequality. The right hand side of inequality (5.10) converges to zero since $f_n(t)$ tends to zero for $n\to\infty$. □

Remark. • *If $U=\mathbb{R}^q$, then one has to replace M_n by M in equation (5.1) since it is not necessary to discretize the noise. For this finite dimensional case, Theorem 5.1 holds as* *well. In detail, we obtain*

$$\mathbb{E}\left\|X_n(t)-\mathscr{X}(t)\right\|_H^2 \leq \left(6\int_0^t \mathbb{E}\left\|\left(\mathscr{N}(\mathscr{X}(s))-\mathscr{N}_n(\mathscr{X}(s))\right)\mathscr{Q}^{\frac{1}{2}}\right\|_{L_{HS}(\mathbb{R}^q,H)}^2 ds + 3t\mathbb{E}\int_0^t \left\|\mathscr{B}u(s)-\mathscr{B}_n u(s)\right\|_H^2 ds + 3\mathbb{E}\left\|X_0-X_{0,n}\right\|_H^2\right) \mathrm{e}^{6\tilde{M}^2 t}.$$

- *Since the function f is increasing in inequality (5.10), we have a uniform convergence with respect to time on a compact interval $[0,T]$, i.e. for $n \to 0$*

$$\sup_{t\in[0,T]} \mathbb{E}\left\|X_n(t)-\mathscr{X}(t)\right\|_H^2 \to 0.$$

We now characterize the vector of Fourier coefficients of the Galerkin solution X_n and set

$$x(t) = \left(\langle X_n(t),h_1\rangle_H,\ldots,\langle X_n(t),h_n\rangle_H\right)^T.$$

The components of x fulfill the following:

$$\langle X_n(t),h_k\rangle_H = \left\langle S_n(t)X_{0,n},h_k\right\rangle_H + \int_0^t \langle S_n(t-s)\mathscr{B}_n u(s),h_k\rangle_H\,ds + \left\langle \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s),h_k\right\rangle_H.$$

Using the representation $S_n(t)x = \sum_{i=1}^n \mathrm{e}^{-\lambda_i t}\langle x,h_i\rangle_H h_i$ ($x \in H_n$), we have

$$\left\langle S_n(t)X_{0,n},h_k\right\rangle_H = \mathrm{e}^{-\lambda_k t}\left\langle X_{0,n},h_k\right\rangle_H = \mathrm{e}^{-\lambda_k t}\langle X_0,h_k\rangle_H$$

and

$$\langle S_n(t-s)\mathscr{B}_n u(s),h_k\rangle_H = \mathrm{e}^{-\lambda_k(t-s)}\langle \mathscr{B}_n u(s),h_k\rangle_H = \sum_{l=1}^m \mathrm{e}^{-\lambda_k(t-s)}\langle \mathscr{B}e_l,h_k\rangle_H \langle u(s),e_l\rangle_{\mathbb{R}^m}$$

for $k = 1,\ldots,n$, where e_l is the l-th unit vector in $\mathbb{R}^m$. Furthermore, it holds

$$\left\langle \int_0^t S_n(t-s)\mathscr{N}_n(X_n(s-))dM_n(s),h_k\right\rangle_H = \sum_{j=1}^n \int_0^t \left\langle S_n(t-s)\mathscr{N}_n(X_n(s-))u_j,h_k\right\rangle_H d\left\langle M(s),u_j\right\rangle_U$$
$$= \sum_{j=1}^n\sum_{i=1}^n \int_0^t \left\langle S_n(t-s)\mathscr{N}_n(h_i)u_j,h_k\right\rangle_H \langle X_n(s-),h_i\rangle_H\, d\left\langle M(s),u_j\right\rangle_U$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \int_0^t \mathrm{e}^{-\lambda_k(t-s)} \left\langle \mathscr{N}(h_i)u_j, h_k \right\rangle_H \langle X_n(s-), h_i \rangle_H \, d\left\langle M(s), u_j \right\rangle_U .$$

Hence, in compact form x is given by

$$x(t) = \mathrm{e}^{At} x_0 + \int_0^t \mathrm{e}^{A(t-s)} Bu(s)ds + \sum_{j=1}^{n} \int_0^t \mathrm{e}^{A(t-s)} N^j x(s-)dM^j(s), \tag{5.11}$$

where

- $A = \mathrm{diag}(-\lambda_1, \ldots, -\lambda_n)$, $B = (\langle \mathscr{B}e_i, h_k \rangle_H)_{\substack{k=1,\ldots,n \\ i=1,\ldots,m}}$, $N^j = \left(\left\langle \mathscr{N}(h_i)u_j, h_k \right\rangle_H\right)_{k,i=1,\ldots,n}$,
- $x_0 = (\langle X_0, h_1 \rangle_H, \ldots, \langle X_0, h_n \rangle_H)^T$ and $M^j(s) = \left\langle M(s), u_j \right\rangle_U$.

The processes M^j are uncorrelated real-valued Lévy processes with $\mathbb{E}\left|M^j(t)\right|^2 = t\mu_j$, $t \geq 0$, and zero mean. Below, we show that the solution of equation (5.11) fulfills the strong solution equation as well. We set

$$g(t) := x_0 + \int_0^t \mathrm{e}^{-As} Bu(s)ds + \sum_{j=1}^{n} \int_0^t \mathrm{e}^{-As} N^j x(s-)dM^j(s), \quad t \geq 0,$$

and determine the stochastic differential of $\mathrm{e}^{At} g(t)$ via the Ito product formula in Corollary A.4:

$$\begin{aligned} e_i^T x(t) = e_i^T \mathrm{e}^{At} g(t) &= e_i^T g(0) + \int_0^t d\left(e_i^T \mathrm{e}^{As}\right) g(s-) + \int_0^t e_i^T \mathrm{e}^{As} dg(s) \\ &= e_i^T \left(x_0 + \int_0^t A \mathrm{e}^{As} g(s)ds + \int_0^t Bu(s)ds + \sum_{j=1}^{n} \int_0^t N^j x(s-)dM^j(s) \right), \end{aligned}$$

where e_i is the i-th unit vector of $\mathbb{R}^n$ and the quadratic covariation terms are zero, since $t \mapsto e_i^T \mathrm{e}^{At}$ is a continuous semimartingale with a martingale part of zero, see (A.2). Hence,

$$x(t) = x_0 + \int_0^t [Ax(s) + Bu(s)]ds + \sum_{j=1}^{n} \int_0^t N^j x(s-)dM^j(s), \quad t \geq 0. \tag{5.12}$$

The corresponding output of the Galerkin solution is

$$y_n(t) = Cx(t), \quad t \geq 0,$$

where $C = (\langle \mathscr{C} h_k, e_l \rangle_{\mathbb{R}^p})_{\substack{l=1,\dots,p \\ k=1,\dots,n}}$ since

$$\langle y_n(t), e_l \rangle_{\mathbb{R}^p} = \langle \mathscr{C} X_n(t), e_l \rangle_{\mathbb{R}^p} = \sum_{k=1}^{n} \langle \mathscr{C} h_k, e_l \rangle_{\mathbb{R}^p} \langle X_n(t), h_k \rangle_H$$

for $l = 1, \dots, p$, where e_l is the l-th unit vector in $\mathbb{R}^p$.

We conclude this section by stating the matrices of the approximate system (5.12) corresponding to the examples we give in Subsection 4.3.1.

Example 5.2. *The Fourier coefficients of the Galerkin solution corresponding to the SPDE in Example 4.13 is given by (5.12) with*

- $A = \operatorname{diag}\left(-1, -4, \dots, -n^2\right)$,
- $N = N^1 = (\langle \mathscr{N}(h_i), h_k \rangle_H)_{k,i=1,\dots,n} = (\langle a h_i, h_k \rangle_H)_{k,i=1,\dots,n} = a I_n$,
- $B = (\langle \mathscr{B}, h_k \rangle_H)_{k=1,\dots,n} = \left(\left\langle 1_{[0,\frac{\pi}{2}]}(\cdot), h_k \right\rangle_H\right)_{k=1,\dots,n} = \left(\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{k} \left[1 - \cos(\frac{k\pi}{2})\right]\right)_{k=1,\dots,n}$,

where we use $H = L^2([0,\pi])$, $\lambda_k = k^2$ and $h_k = \sqrt{\frac{2}{\pi}} \sin(k\cdot)$. Since here $U = \mathbb{R}$ and $\mathbb{R}^m = \mathbb{R}$, we just have a trivial basis $\{1\}$ for both spaces. The corresponding approximate scalar output y_n is characterized by the vector

$$C^T = (\mathscr{C} h_k)_{k=1,\dots,n} = \left(\left(\frac{2}{\pi}\right)^{\frac{3}{2}} \frac{1}{k} \left[\cos(\frac{k\pi}{2}) - \cos(k\pi)\right]\right)_{k=1,\dots,n},$$

where the integral operator $\mathscr{C}$ is defined in (4.15).

Example 5.3. *The Fourier coefficients of the Galerkin solution corresponding to the SPDE in Example 4.14 are given by (5.12) with*

- $A = \operatorname{diag}(0, -1, -1, -2, \dots)$,
- $N = N^1 = \left(\left\langle \mathrm{e}^{-|\cdot - \frac{\pi}{2}| - \cdot} h_i, h_k \right\rangle_H\right)_{k,i=1,\dots,n}$,
- $B = \left(\left\langle 1_{[\frac{\pi}{4}, \frac{3\pi}{4}]^2}(\cdot), h_k \right\rangle_H\right)_{k=1,\dots,n}$,

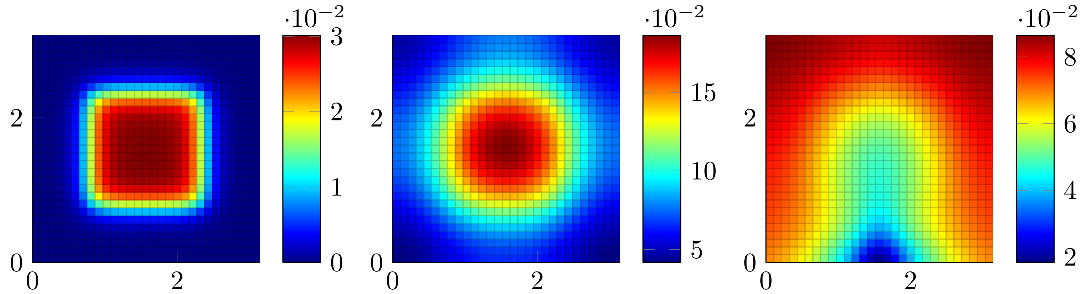
where $H = L^2([0,\pi]^2)$ and λ_k, h_k are chosen as in Example 4.14. Again, we have a trivial basis for the spaces U and $\mathbb{R}^m$. The initial state is $x_0 = 0$. Since y_n is scalar, C becomes a vector which is given by $C^T = (\mathscr{C}h_k)_{k=1,\dots,n}$, where $\mathscr{C}$ is given in (4.17). In Figure 5.1 we illustrate the Galerkin solution of the stochastic heat equation of Example 4.14 on the time interval $[0,\pi]$, where we set $M(t) = w(t) - (N(t) - t)$. The processes $(w(t))_{t\in[0,\pi]}$ and $(N(t))_{t\in[0,\pi]}$ are independent processes, where w is a standard Wiener process and $(N(t))_{t\in[0,\pi]}$ is a Poisson process with parameter 1. In the first picture in Figure 5.1a it can be seen that the middle of the square is heated by the control that we fix as follows $u(t) = -2e_1^T x(t) + \mathrm{e}^{w(t)}$, $t \in [0,\pi]$. Here, the control term $\mathrm{e}^{w(t)}$ is of the form $\tilde{u}(t,\omega) = \tilde{u}_D(t)r(t,\omega)$, $t \in [0,\pi]$ and $\omega \in \Omega$, where $r(t,\omega) = \mathrm{e}^{w(t,\omega)}$ is interpreted as a multiplicative positive random perturbation of $\tilde{u}_D \equiv 1$.
The noise acts mainly in a neighborhood of the middle of the x-axis as one can see in the third graphic in Figure 5.1a. The first jump occurs there which cools down the system. In the picture after (first picture in Figure 5.1b) the jump also affects the center. The square then regains the heat from the input until the next jump appears, see third picture in Figure 5.1b and first picture in Figure 5.1c. Finally, the heat is mainly concentrated in the upper part of the square while the lower region is colder due to the noise which randomly causes a loss of heat there. The corresponding output, which is the mean temperature on the non-heated area, is stated in Figure 5.2. There, the interaction of Wiener and Poisson noise can be seen. The trajectory is piecewise impacted by the Wiener part and has jumps in between, where the temperature is randomly reduced.

5.2. Approximation of stochastic damped wave equations

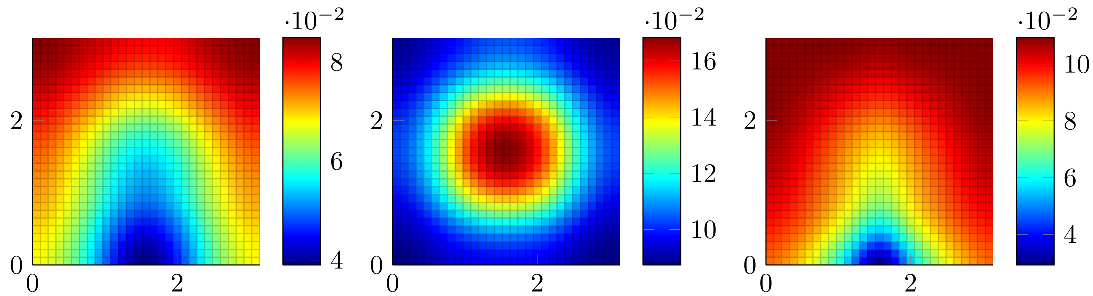
In this section, we discuss an approximation scheme for the first order system that we obtain in Subsection 4.3.2. Since the noise process is finite dimensional, i.e. $U = \mathbb{R}^2$, we do not need to discretize the noise here. That means that M_n is replaced by $M = \left(\begin{smallmatrix} M_1 \\ M_2 \end{smallmatrix}\right)$ in the Galerkin system (5.1), where M_1, M_2 are independent, square integrable and scalar Lévy processes with mean zero. Furthermore, we have the following generator of a contraction semigroup

$$\mathscr{A} = \begin{bmatrix} 0 & I \\ -\tilde{\mathscr{A}} & -\alpha I \end{bmatrix},$$

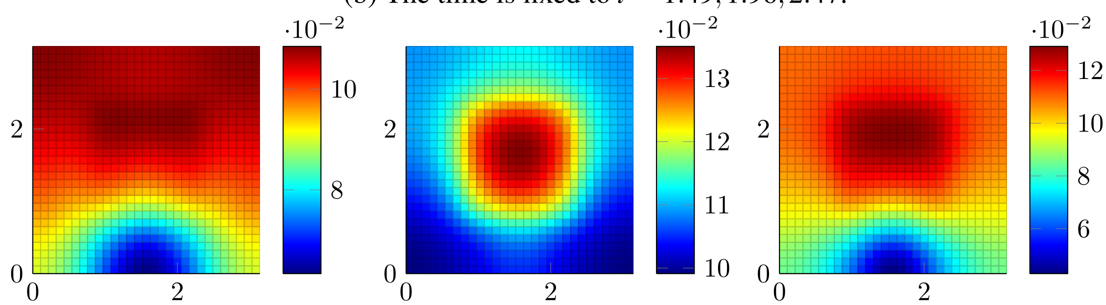
where $\tilde{\mathscr{A}}$ is a self adjoint, positive definite operator, compare Lemma 4.15. The orthonormal basis $(h_k)_{k\in\mathbb{N}}$ of H, which is characterized by the eigenvectors and eigenvalues of $\tilde{\mathscr{A}}$, is stated



(a) The time is fixed to $t = 0.03, 0.63, 1.43$.



(b) The time is fixed to $t = 1.49, 1.96, 2.47$.



(c) The time is fixed to $t = 2.59, 2.83, 2.95$.

Figure 5.1.: The Galerkin solution to the stochastic heat equation in Example 4.14

in (4.19). Below, we deal with the alternative representation of the mild solution in (4.21). The operator $\mathscr{N}_n$ in the Galerkin equation (5.1) is then determined by the operators $\mathscr{D}_{i,n}$ $(i = 1, 2)$ which are given by

$$\mathscr{D}_{i,n}x = \sum_{k=1}^{n} \langle \mathscr{D}_i x, h_k \rangle_H \, h_k$$

for all $x \in H$.

In contrast to the heat equation case, we do not have an explicit representation for the C_0-semigroup $(S(t))_{t \geq 0}$. So, the proof of convergence of the Galerkin solution turns out to be

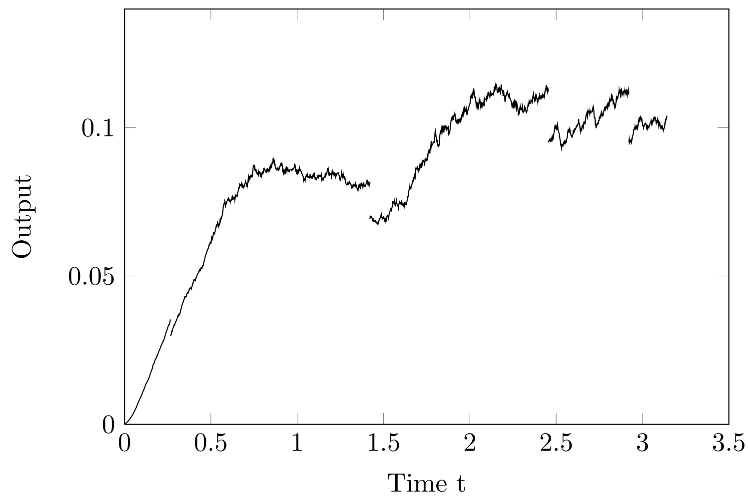


Figure 5.2.: The output of the stochastic heat equation in Example 4.14

different. We define a C_0-semigroup $(S_n(t))_{t\geq 0}$ on H_n which is defined by

$$S_n(t)x = \sum_{k=1}^{n} \langle S(t)x, h_k \rangle_H h_k$$

for all $x \in H$ and generated by $\mathscr{A}_n$ such that the mild solution of equation (5.1) is given by

$$X_n(t) = S_n(t)X_{0,n} + \int_0^t S_n(t-s)\mathscr{B}_n u(s)ds + \sum_{i=1}^{2} \int_0^t S_n(t-s)\mathscr{D}_{i,n} X_n(s-)dM_i(s).$$

Below, we formulate the central result of this section, where the ideas in [27, 29, 40, 60] are used.

Theorem 5.4. *Under the assumptions we made in Subsection 4.3.2, the solution X_n of equation (5.1) approximates the solution $\mathscr{X}$ of equation (4.20), i.e.*

$$\mathbb{E}\,\|X_n(t) - \mathscr{X}(t)\|_H^2 \to 0$$

for $n \to \infty$ and $t \geq 0$. Of course, this implies the convergence of the corresponding outputs $\mathbb{E}\,\|y_n(t) - \mathscr{Y}(t)\|_2^2 \to 0$.

Proof.

$$\mathbb{E}\,\|\mathscr{X}(t) - X_n(t)\|_H^2 \leq 4\mathbb{E}\left\|S(t)X_0 - S_n(t)X_{0,n}\right\|_H^2$$

$$
\begin{aligned}
&+4\mathbb{E}\left\| \int_0^t (S(t-s)\mathscr{B} - S_n(t-s)\mathscr{B}_n)u(s)ds \right\|_H^2 \\
&+4\mathbb{E}\left\| \int_0^t S(t-s)\mathscr{D}_1\mathscr{X}(s-) - S_n(t-s)\mathscr{D}_{1,n}X_n(s-)dM_1(s) \right\|_H^2 \\
&+4\mathbb{E}\left\| \int_0^t S(t-s)\mathscr{D}_2\mathscr{X}(s-) - S_n(t-s)\mathscr{D}_{2,n}X_n(s-)dM_2(s) \right\|_H^2.
\end{aligned}
$$

Since $(S(t))_{t\geq 0}$ is a contraction semigroup, we obtain

$$
\begin{aligned}
4\mathbb{E}\left\|S(t)X_0 - S_n(t)X_{0,n}\right\|_H^2 &\leq 8\mathbb{E}\left\|S(t)X_0 - S_n(t)X_0\right\|_H^2 + 8\mathbb{E}\left\|S_n(t)X_0 - S_n(t)X_{0,n}\right\|_H^2 \\
&\leq 8\mathbb{E}\left\|S(t)X_0 - S_n(t)X_0\right\|_H^2 + 8\mathbb{E}\left\|X_0 - X_{0,n}\right\|_H^2.
\end{aligned} \tag{5.13}
$$

By the representation $S_n(t)x = \sum_{i=1}^n \langle S(t)x, h_i\rangle_H h_i$ $(x\in H)$ and Lebesgue's theorem, (5.13) tends to zero for $n\to\infty$. The Hölder inequality yields

$$
\begin{aligned}
&\mathbb{E}\left\| \int_0^t (S(t-s)\mathscr{B} - S_n(t-s)\mathscr{B}_n)u(s)ds \right\|_H^2 \\
&\leq t\mathbb{E}\int_0^t \left\|(S(t-s)\mathscr{B} - S_n(t-s)\mathscr{B}_n)u(s)\right\|_H^2 ds.
\end{aligned}
$$

We obtain

$$
\begin{aligned}
&\left\|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}_n u(s)\right\|_H^2 \\
&\leq 2\left\|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}u(s)\right\|_H^2 + 2\left\|S_n(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}_n u(s)\right\|_H^2 \\
&\leq 2\left\|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}u(s)\right\|_H^2 + 2\left\|\mathscr{B}u(s) - \mathscr{B}_n u(s)\right\|_H^2 \to 0
\end{aligned}
$$

$\mathbb{P}$- almost surely for $n\to\infty$ and

$$
8t\mathbb{E}\int_0^t \left\|\mathscr{B}u(s) - \mathscr{B}_n u(s)\right\|_H^2 + \left\|S(t-s)\mathscr{B}u(s) - S_n(t-s)\mathscr{B}u(s)\right\|_H^2 ds \to 0 \tag{5.14}
$$

for $n\to\infty$ by Lebesgue's theorem. By applying the Ito isometry from Section 3.1, we have

$$
\begin{aligned}
&4\mathbb{E}\left\| \int_0^t S(t-s)\mathscr{D}_i\mathscr{X}(s-) - S_n(t-s)\mathscr{D}_{i,n}X_n(s-)dM_i(s) \right\|_H^2 \\
&= 4\int_0^t \mathbb{E}\left\|S(t-s)\mathscr{D}_i\mathscr{X}(s-) - S_n(t-s)\mathscr{D}_{i,n}X_n(s-)\right\|_H^2 ds\, \mathbb{E}[M_i^2(1)]
\end{aligned}
$$

$$\leq 8\mathbb{E}\left[\int_0^t \|S(t-s)\mathscr{D}_i\mathscr{X}(s) - S_n(t-s)\mathscr{D}_i\mathscr{X}(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)] \tag{5.15}$$
$$+8\mathbb{E}\left[\int_0^t \|S_n(t-s)\mathscr{D}_i\mathscr{X}(s) - S_n(t-s)\mathscr{D}_{i,n}X_n(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)].$$

The term (5.15) converges to zero for $n \to \infty$ by Lebesgue's theorem. Moreover, it holds

$$8\mathbb{E}\left[\int_0^t \|S_n(t-s)\mathscr{D}_i\mathscr{X}(s) - S_n(t-s)\mathscr{D}_{i,n}X_n(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)]$$
$$\leq 16\mathbb{E}\left[\int_0^t \|\mathscr{D}_i\mathscr{X}(s) - \mathscr{D}_{i,n}\mathscr{X}(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)] \tag{5.16}$$
$$+16\mathbb{E}\left[\int_0^t \|\mathscr{D}_{i,n}\mathscr{X}(s) - \mathscr{D}_{i,n}X_n(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)].$$

Again, by Lebesgue's theorem the term (5.16) tends to zero for $n \to \infty$ and

$$16\mathbb{E}\left[\int_0^t \|\mathscr{D}_{i,n}\mathscr{X}(s) - \mathscr{D}_{i,n}X_n(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)]$$
$$\leq 16\,\|\mathscr{D}_i\|_{L(H)}^2\,\mathbb{E}\left[\int_0^t \|\mathscr{X}(s) - X_n(s)\|_H^2\,ds\right]\,\mathbb{E}[M_i^2(1)].$$

Summarizing everything, we obtain

$$\mathbb{E}\,\|\mathscr{X}(t) - X_n(t)\|_H^2 \leq f_n(t) + k_1\int_0^t \mathbb{E}\,\|\mathscr{X}(s) - X_n(s)\|_H^2\,ds,$$

where $k_1 := 16\left(\|\mathscr{D}_1\|_{L(H)}^2\,\mathbb{E}[M_1^2(1)] + \|\mathscr{D}_2\|_{L(H)}^2\,\mathbb{E}[M_2^2(1)]\right)$ and f_n is a sequence of functions consisting of the terms (5.13), (5.14), (5.15) and (5.16). Hence,

$$\mathbb{E}\,\|\mathscr{X}(t) - X_n(t)\|_H^2 \leq f_n(t) + k_1\int_0^t f_n(s)\,\mathrm{e}^{k_1(t-s)}\,ds \tag{5.17}$$

by Gronwall's inequality. The first term of the right hand side of inequality (5.17) converges to zero since $f_n(t)$ converges to zero for $n \to \infty$. In addition, f_n is bounded by the increasing function $\tilde{f}$ defined by

$$\tilde{f}(t) := k_2\left(\mathbb{E}\,\|X_0\|_H^2 + t\int_0^t \mathbb{E}\,\|u(s)\|_{\mathbb{R}^m}^2\,ds + \int_0^t \mathbb{E}\,\|\mathscr{X}(s)\|_H^2\,ds\right)$$

with a suitable constant $k_2 > 0$. So, $f_n(s) \leq \tilde{f}(t)$ for all $0 \leq s \leq t$ and every $n \in \mathbb{N}$. Hence, the second term of the right hand side of inequality (5.17) converges to zero by Lebesgue's theorem. □

Moreover, notice that the mild and the strong solution of equation (5.1) coincide which we use below. This, we can prove by using $S_n(t) = \mathrm{e}^{\mathscr{A}_n t}$ and by applying the Ito product formula from Corollary A.4 as in Section 5.1.

First, we determine the components of y_n. They are given by

$$y_n^\ell(t) = \langle y_n(t), e_\ell \rangle_{\mathbb{R}^p} = \langle \mathscr{C} X_n(t), e_\ell \rangle_{\mathbb{R}^p} = \sum_{k=1}^n \langle \mathscr{C} h_k, e_\ell \rangle_{\mathbb{R}^p} \langle X_n(t), h_k \rangle_H$$

for $\ell = 1, \ldots, p$, where e_ℓ is the ℓ-th unit vector in $\mathbb{R}^p$. We set

$$x(t) = (\langle X_n(t), h_1 \rangle_H, \ldots, \langle X_n(t), h_n \rangle_H)^T \text{ and } C = (\langle \mathscr{C} h_k, e_\ell \rangle_{\mathbb{R}^p})_{\substack{\ell=1,\ldots,p \\ k=1,\ldots,n}}$$

and obtain

$$y_n(t) = Cx(t), \quad t \geq 0. \tag{5.18}$$

The components $x^k(t) := \langle X_n(t), h_k \rangle_H$ of $x(t)$ fulfill the following:

$$dx^k(t) = [\langle \mathscr{A}_n X_n(t), h_k \rangle_H + \langle \mathscr{B}_n u(t), h_k \rangle_H] \, dt + \sum_{i=1}^2 \langle \mathscr{D}_{i,n} X_n(t-), h_k \rangle_H \, dM_i(t).$$

By using the Fourier series representation of X_n, we obtain

$$\begin{aligned} dx^k(t) &= \left[\sum_{j=1}^n \left\langle \mathscr{A}_n h_j, h_k \right\rangle_H x^j(t) + \sum_{j=1}^m \left\langle \mathscr{B}_n e_j, h_k \right\rangle_H \left\langle u(t), e_j \right\rangle_{\mathbb{R}^m} \right] dt \\ &\quad + \sum_{i=1}^2 \sum_{j=1}^n \left\langle \mathscr{D}_{i,n} h_j, h_k \right\rangle_H x^j(t-) dM_i(t) \\ &= \left[\sum_{j=1}^n \left\langle \mathscr{A} h_j, h_k \right\rangle_H x^j(t) + \sum_{j=1}^m \left\langle \mathscr{B} e_j, h_k \right\rangle_H \left\langle u(t), e_j \right\rangle_{\mathbb{R}^m} \right] dt \\ &\quad + \sum_{i=1}^2 \sum_{j=1}^n \left\langle \mathscr{D}_i h_j, h_k \right\rangle_H x^j(t-) dM_i(t), \end{aligned}$$

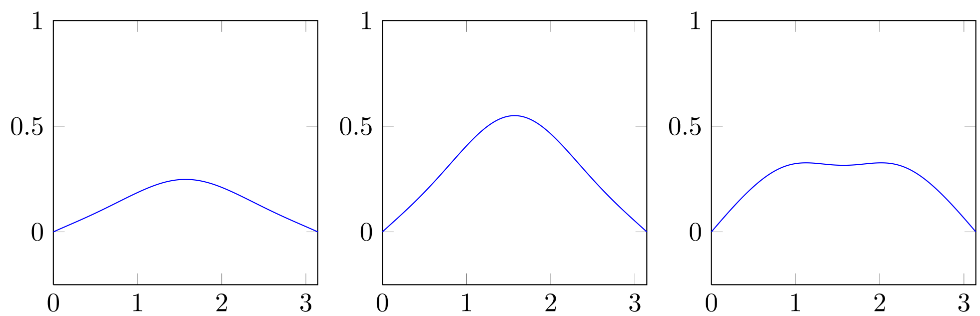
where e_j is the j-th unit vector in $\mathbb{R}^m$. Hence, in compact form, x is given by

$$dx(t) = [Ax(t) + Bu(t)]\,dt + \sum_{i=1}^{2} N^i x(t-)dM_i(t), \tag{5.19}$$

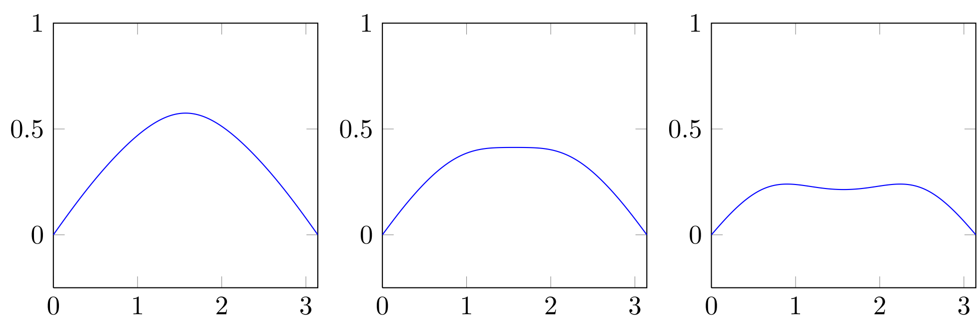
where

- $A = \left(\langle \mathscr{A} h_j, h_k \rangle_H\right)_{k,j=1,\dots,n} = \operatorname{diag}(E_1, \dots, E_{\frac{n}{2}})$ with $E_\ell = \begin{pmatrix} 0 & \sqrt{\tilde{\lambda}_\ell} \\ -\sqrt{\tilde{\lambda}_\ell} & -\alpha \end{pmatrix}$ $(\ell = 1, \dots, \frac{n}{2})$, where $\tilde{\lambda}_\ell$ are the eigenvalues of $\tilde{\mathscr{A}}$,
- $B = \left(\langle \mathscr{B} e_j, h_k \rangle_H\right)_{\substack{k=1,\dots,n \\ j=1,\dots,m}}$ and $N^i = \left(\langle \mathscr{D}_i h_j, h_k \rangle_H\right)_{k,j=1,\dots,n}$.

Next, we study the Galerkin solution of Example 4.16.



(a) The time is fixed to $t = 0.94, 1.41, 1.73$.



(b) The time is fixed to $t = 2.36, 2.67, 3.06$.

Figure 5.3.: The Galerkin solution to the stochastic damped wave equation in Example 4.16

Example 5.5. *In Example 4.16, the eigenvectors of $\tilde{\mathscr{A}} = -\frac{\partial^2}{\partial \zeta^2}$ characterizing the orthonormal basis $(h_k)_{k\in\mathbb{N}}$ in (4.19) are given by $\tilde{h}_k = \sqrt{\frac{2}{\pi}} \sin(k\cdot)$ and the corresponding eigenvalues are $\tilde{\lambda}_k = k^2$ for $k \in \mathbb{N}$. The matrices of the Galerkin system (5.19) are*

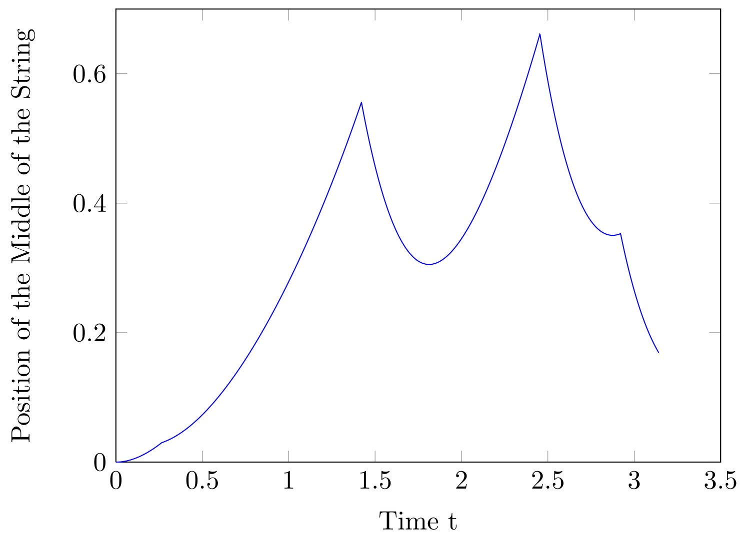


Figure 5.4.: The output of the stochastic damped wave equation in Example 4.16

- $A = \operatorname{diag}\left(E_1, \dots, E_{\frac{n}{2}}\right)$ *with* $E_\ell = \left(\begin{smallmatrix} 0 & \ell \\ -\ell & -\alpha \end{smallmatrix}\right)$,
- $B = (\langle \mathscr{B}, h_k \rangle_H)_{k=1,\dots,n}$ *with*

$$\langle \mathscr{B}, h_{2\ell-1} \rangle_H = 0, \quad \langle \mathscr{B}, h_{2\ell} \rangle_H = \sqrt{\frac{2}{\pi}} \left\langle \mathrm{e}^{-(\cdot - \frac{\pi}{2})^2}, \sin(\ell \cdot) \right\rangle_{L^2([0,\pi])},$$

- $N^2 = 0$ *and* $N^1 = \left(\left\langle \mathscr{D}_1 h_j, h_k \right\rangle_H\right)_{k,j=1,\dots,n} = \left(n_{kj}\right)_{k,j=1,\dots,n}$ *with*

$$n_{(2\ell-1)j} = 0, \quad n_{(2\ell)j} = \begin{cases} 0, & \text{if } j = 2v, \\ \frac{4}{\pi v} \left\langle \sin(\ell \cdot), \mathrm{e}^{-(\cdot - \frac{\pi}{2})^2} \sin(v \cdot) \right\rangle_{L^2([0,\pi])}, & \text{if } j = 2v-1, \end{cases}$$

  for $j = 1, \dots, n$ *and* $v = 1, \dots, \frac{n}{2}$,
- *and the output matrix* C *in (5.18) is given by* $C^T = (\mathscr{C} h_k)_{k=1,\dots,n}$ *with*

$$\mathscr{C} h_{2\ell} = 0 \quad \text{and} \quad \mathscr{C} h_{2\ell-1} = \frac{1}{\sqrt{2\pi}\ell^2 \varepsilon} \left[\cos\left(\ell\left(\frac{\pi}{2} - \varepsilon\right)\right) - \cos\left(\ell\left(\frac{\pi}{2} + \varepsilon\right)\right)\right],$$

  where we assume n *to be even and* $\ell = 1, \dots, \frac{n}{2}$.

In Figure 5.3 we plot the Galerkin solution of the stochastic damped wave equation of Example 4.16 on the time interval $[0, \pi]$ *with* $\alpha = 2$, *where we set* $M_1(t) = -2(N(t) - t)$ *with* $(N(t))_{t \in [0,\pi]}$

being a Poisson process with parameter 1*. Due to the input* $u(t) = 1$*,* $t \in [0, \pi]$*, we give to the system the string moves up as we can see in the first two pictures in Figure 5.3a. The Poisson noise causes jumps at random times. This uncertainty here models the effect of wind on the string. In the third picture in Figure 5.3a the wind appears and pushes back the string. After the wind stops blowing, the the cable moves up again, see first picture in Figure 5.3b, until it is impacted twice by wind in the following two pictures. The corresponding output, which is the position of the middle of the cable, is stated in Figure 5.4. In this graph we have peaks at random positions characterized by the jumps of the underlying Poisson process. These peaks mark the appearance of the wind which forces the cable to move in the opposite direction. Whenever the wind blows, we have a tiny jump in the system which is not visible in Figure 5.4 due to its small size.*

5.3. Conclusions

In this chapter, we introduced a spectral Galerkin scheme for linear controlled SPDEs with Lévy noise. In particular, we discussed this method for stochastic heat and damped wave equations. We showed that the Galerkin solution converges to the mild solutions of the SPDEs considered here. To conclude each of the two sections, we ran simulations on a particular heat and a particular damped wave equation with scalar Lévy noise.

6. Linear Ordinary SDEs with Lévy Noise and Balancing Related Model Order Reduction

6.1. Linear control with Lévy noise

Before describing type 1 balanced truncation (BT) and the singular perturbation approximation (SPA) for the stochastic case, we define observability and reachability concepts to motivate both schemes. We introduce observability and reachability Gramians for our Lévy driven system like Benner, Damm [9] do (Section 2.2). Additionally, we show that the sets of observable and reachable states are characterized by these Gramians. This is analogous to deterministic systems, where observability and reachability concepts are described in Subsections 4.2.1 and 4.2.2 in Antoulas [2].

In this section, we only stress the finite dimensional case, since we apply model order reduction schemes like BT or the SPA just to large scale ordinary stochastic differential equations which might come from a spatial discretization of a stochastic partial differential equation (SPDE), see Chapter 5. One might also think about applying model order reducing techniques to SPDEs directly but this kind of idea would require different approaches. So, the Gramians of the finite dimensional case considered here are based on a generalized fundamental solution Φ which is a matrix-valued stochastic process. Unfortunately, such a fundamental solution does not exist for an infinite dimensional system of the form (4.3) such that the corresponding Gramians would have to be defined in a different way.

This section is based on Sections 3.1 and 3.2 in Benner, Redmann [14] and extends [13] by providing more details and by considering a more general framework.

6.1.1. Reachability concept

Let $M_1, \dots, M_q$ be real-valued uncorrelated and square integrable Lévy processes with mean zero defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$.[1] In addition, we assume M_k $(k = 1, \dots, q)$ to be $(\mathscr{F}_t)_{t\geq 0}$-adapted and the increments $M_k(t+h) - M_k(t)$ to be independent of $\mathscr{F}_t$ for $t, h \geq 0$.

We consider linear controlled systems of the type that we obtain in Sections 5.1 and 5.2, i.e

$$dx(t) = [Ax(t) + Bu(t)]dt + \sum_{k=1}^{q} N^k x(t-)dM_k(t), \quad t \geq 0, \quad x(0) = x_0 \in \mathbb{R}^n, \tag{6.1}$$

where A, $N^k \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Note that N^k is just a notation of a matrix with index k and not the k-th matrix power. With L^2_T we denote the space of all $(\mathscr{F}_t)_{t\geq 0}$-adapted stochastic processes v with values in $\mathbb{R}^m$, which are square integrable with respect to $\mathbb{P} \otimes dt$. We call the norm in L^2_T *energy norm*. It is given by

$$\|v\|^2_{L^2_T} := \mathbb{E}\int_0^T v^T(t)v(t)dt = \mathbb{E}\int_0^T \|v(t)\|_2^2 dt,$$

where we define the processes v_1 and v_2 to be equal in L^2_T if they coincide almost surely with respect to $\mathbb{P} \otimes dt$. For the case $T = \infty$, we denote the space by L^2. Further, we assume controls $u \in L^2_T$ for every $T > 0$. We start with the definition of a solution of (6.1).

Definition 6.1. *An $\mathbb{R}^n$-valued and $(\mathscr{F}_t)_{t\geq 0}$-adapted cadlag process $(x(t))_{t\geq 0}$ is called solution of (6.1) if*

$$x(t) = x_0 + \int_0^t [Ax(s) + Bu(s)]ds + \sum_{k=1}^{q} \int_0^t N^k x(s-)dM_k(s) \tag{6.2}$$

$\mathbb{P}$-almost surely holds for all $t \geq 0$.

Below, the solution of (6.1) at time $t \geq 0$ with initial condition $x_0 \in \mathbb{R}^n$ and given control u is always denoted by $x(t, x_0, u)$. For the solution of (6.1) in the uncontrolled case ($u \equiv 0$), we briefly write $x_{h,x_0} := x(t, x_0, 0)$, where x_{h,x_0} is called homogeneous solution. Furthermore, $\|\cdot\|_2$ denotes the Euclidean norm. We assume the homogeneous solution to be *asymptotically mean square*

[1] We assume that $(\mathscr{F}_t)_{t\geq 0}$ is right continuous and that $\mathscr{F}_0$ contains all $\mathbb{P}$ null sets.

stable, which means that

$$\mathbb{E}\left\|x_{h,x_0}(t)\right\|_2^2 \to 0$$

for $t \to \infty$ and $x_0 \in \mathbb{R}^n$. This concept of stability is also used in Benner, Damm [9] and is necessary for defining (infinite) Gramians, which are introduced later.

Proposition 6.2. *Let x_{h,x_0} be the solution of (6.1) in the uncontrolled case with any initial value $x_0 \in \mathbb{R}^n$, then $\mathbb{E}\left[x_{h,x_0}(t)x_{h,x_0}^T(t)\right]$ is the solution of the matrix integral equation*

$$\mathbb{Y}(t) = x_0x_0^T + \int_0^t \mathbb{Y}(s)ds\,A^T + A\int_0^t \mathbb{Y}(s)ds + \sum_{k=1}^q N^k \int_0^t \mathbb{Y}(s)ds\,\left(N^k\right)^T\,\mathbb{E}\left[M_k(1)^2\right] \tag{6.3}$$

for $t \geq 0$.

Proof. We determine the stochastic differential of the matrix-valued process $x_{h,x_0}x_{h,x_0}^T$ via using the Ito formula in Corollary A.5. This yields

$$x_{h,x_0}(t)x_{h,x_0}^T(t) = x_0x_0^T + \int_0^t x_{h,x_0}(s-)dx_{h,x_0}^T(s) + \int_0^t dx_{h,x_0}(s)x_{h,x_0}^T(s-) + \left([e_i^T x_{h,x_0}, x_{h,x_0}^T e_j]_t\right)_{i,j=1,\dots,n},$$

where e_i is the i-th unit vector. We obtain

$$\int_0^t x_{h,x_0}(s-)dx_{h,x_0}^T(s) = \int_0^t x_{h,x_0}(s-)x_{h,x_0}^T(s)A^T ds + \sum_{k=1}^q \int_0^t x_{h,x_0}(s-)x_{h,x_0}^T(s-)(N^k)^T dM_k(s) \text{ and}$$
$$\int_0^t dx_{h,x_0}(s)x_{h,x_0}^T(s-) = \int_0^t Ax_{h,x_0}(s)x_{h,x_0}^T(s-)ds + \sum_{k=1}^q \int_0^t N^k x_{h,x_0}(s-)x_{h,x_0}^T(s-)dM_k(s)$$

by inserting the stochastic differential of x_{h,x_0}. Thus, by taking the expectation, we obtain

$$\begin{aligned}\mathbb{E}\left[x_{h,x_0}(t)x_{h,x_0}^T(t)\right] = x_0x_0^T + \int_0^t \mathbb{E}\left[x_{h,x_0}(s-)x_{h,x_0}^T(s)\right]A^T ds + \int_0^t A\mathbb{E}\left[x_{h,x_0}(s)x_{h,x_0}^T(s-)\right]ds \\ + \left(\mathbb{E}[e_i^T x_{h,x_0}, x_{h,x_0}^T e_j]_t\right)_{i,j=1,\dots,n}\end{aligned}$$

applying that an Ito integral has mean zero, see Section 3.1. Considering equation (A.5), we have

$$\mathbb{E}[e_i^T x_{h,x_0}, x_{h,x_0}^T e_j]_t = e_i^T \sum_{k=1}^q \int_0^t \mathbb{E}\left[N^k x_{h,x_0}(s)x_{h,x_0}^T(s)\left(N^k\right)^T\right]ds \cdot c_k e_j,$$

where $c_k := \mathbb{E}\left[M_k(1)^2\right]$. In addition, we use the property that a cadlag process has at most countably many jumps on a finite time interval (see Theorem 2.7.1 in Applebaum [5]), such that we can replace the left limit by the function value itself. Thus,

$$\begin{aligned}\mathbb{E}\left[x_{h,x_0}(t)x^T_{h,x_0}(t)\right] = x_0x_0^T + \int_0^t \mathbb{E}\left[x_{h,x_0}(s)x^T_{h,x_0}(s)\right] ds\, A^T + A\int_0^t \mathbb{E}\left[x_{h,x_0}(s)x^T_{h,x_0}(s)\right] ds \\ + \sum_{k=1}^{q} N^k \int_0^t \mathbb{E}\left[x_{h,x_0}(s)x^T_{h,x_0}(s)\right] ds\, \left(N^k\right)^T \cdot c_k\end{aligned} \tag{6.4}$$

which gives the result. □

We introduce an additional concept of stability for the homogeneous system ($u \equiv 0$) corresponding to equation (6.1). We call x_{h,x_0} *exponentially mean square stable* if there exist $c, \beta > 0$ such that

$$\mathbb{E}\left\|x_{h,x_0}(t)\right\|_2^2 \leq \|x_0\|_2^2\, c\, \mathrm{e}^{-\beta t}$$

for $t \geq 0$. This stability concept turns out to be equivalent to asymptotic mean square stability which is stated in the next theorem.

Theorem 6.3. *The following are equivalent:*

(i) The uncontrolled equation (6.1) is asymptotically mean square stable.

(ii) The uncontrolled equation (6.1) is exponentially mean square stable.

(iii) The eigenvalues of $\left(I_n \otimes A + A \otimes I_n + \sum_{k=1}^q N^k \otimes N^k \cdot \mathbb{E}\left[M_k(1)^2\right]\right)$ *have negative real parts.*

(iv) There exists a matrix $X > 0$, *such that*

$$A^T X + XA + \sum_{k=1}^{q} (N^k)^T X N^k \mathbb{E}\left[M_k^2(1)\right] < 0.$$

(v) For all $Y > 0$, *there exists a matrix* $X > 0$, *such that*

$$A^T X + XA + \sum_{k=1}^{q} (N^k)^T X N^k \mathbb{E}\left[M_k^2(1)\right] = -Y.$$

Especially, mean square asymptotic stability implies the stability of A, that is $\sigma(A) \subset \mathbb{C}_-$.

Proof. With Proposition 6.2 the proof is similar to the Wiener case which is considered in Theorem 3.6.1 in [21]. We make use of the techniques applied in [21] to prove the more general Lévy noise case below.

From Proposition 6.2 it is known that $\mathbb{E}\left[x_{h,x_0}(t)x_{h,x_0}^T(t)\right]$ is the solution of the matrix differential equation

$$\dot{\mathbb{Y}}(t) = \mathbb{Y}(t)\, A^T + A\, \mathbb{Y}(t) + \sum_{k=1}^{m} N^k\, \mathbb{Y}(t)\, (N^k)^T\, \mathbb{E}\left[M_k(1)^2\right]. \tag{6.5}$$

Equation (6.5) is equivalent to

$$\frac{d}{dt}\operatorname{vec}(\mathbb{Y}(t)) = \left(I \otimes A + A \otimes I + \sum_{k=1}^{m} N^k \otimes N^k \cdot \mathbb{E}\left[M_k(1)^2\right]\right)\operatorname{vec}(\mathbb{Y}(t)). \tag{6.6}$$

We first show $(iii) \Rightarrow (ii)$. From (iii) the asymptotic stability of (6.6) follows. Asymptotic stability of (6.6) implies exponential stability, such that

$$\left\|\operatorname{vec}(\mathbb{Y}(t))\right\|_2^2 \leq \left\|\operatorname{vec}\left(x_0 x_0^T\right)\right\|_2^2 K_1\, \mathrm{e}^{-\beta_1 t} = \left\|x_0 x_0^T\right\|_F^2 K_1\, \mathrm{e}^{-\beta_1 t} \leq \left\|x_0 x_0^T\right\|_{2,ind}^2 \tilde{c} K_1\, \mathrm{e}^{-\beta_1 t}$$

for $K_1, \beta_1, \tilde{c} > 0$, where $\|\cdot\|_{2,ind}$ is the matrix norm that is induced by $\|\cdot\|_2$. Since

$$\|\mathbb{Y}(t)\|_{2,ind}^2 \leq \|\mathbb{Y}(t)\|_F^2 = \|\operatorname{vec}(\mathbb{Y}(t))\|_2^2$$

holds, equation (6.5) is exponentially stable and hence (ii) follows. It is obvious that (ii) implies (i). We now focus on $(i) \Rightarrow (iii)$. From (i) we conclude that equation (6.5) is asymptotically stable. The asymptotic stability of (6.6) follows by

$$\|\operatorname{vec}(\mathbb{Y}(t))\|_2^2 = \|\mathbb{Y}(t)\|_F^2 \leq \tilde{c}\, \|\mathbb{Y}(t)\|_{2,ind}^2$$

and asymptotic stability of (6.6) implies (iii). We continue with the proof of $(iii) \Rightarrow (v)$. Obviously, condition (iii) is equivalent to

$$\sigma\left(I_n \otimes A^T + A^T \otimes I_n + \sum_{k=1}^{q} \left(N^k\right)^T \otimes \left(N^k\right)^T \cdot \mathbb{E}\left[M_k(1)^2\right]\right) \subset \mathbb{C}_-$$

which, by the considerations above, is again equivalent to the exponentially mean square stability of the following equation

$$dx_{h,x_0}(t) = A^T x_{h,x_0}(t)dt + \sum_{k=1}^{q} \left(N^k\right)^T x_{h,x_0}(t-)dM_k(t), \quad t \geq 0. \tag{6.7}$$

Let Φ_D be the fundamental solution to the dual system (6.7), i.e. Φ_D satisfies

$$\Phi_D(t) = I_n + \int_0^t A^T \Phi_D(s)ds + \sum_{k=1}^{q} \int_0^t (N^k)^T \Phi_D(s-)dM_k(s).$$

For an arbitrary matrix $Y > 0$ the integral

$$\mathbb{E} \int_0^\infty \Phi_D(t) Y \Phi_D^T(t) dt = X > 0$$

exists by the exponentially mean square stability of (6.7). We set $\mathbb{Y}(t) := \Phi_D(t) Y \Phi_D^T(t)$ and as in Proposition 6.2, we obtain

$$\mathbb{Y}(t) = Y + \int_0^t \mathbb{Y}(s)ds\, A + A^T \int_0^t \mathbb{Y}(s)ds + \sum_{k=1}^{q} \left(N^k\right)^T \int_0^t \mathbb{Y}(s)ds\, N^k\, \mathbb{E}\left[M_k(1)^2\right]$$

for $t \geq 0$. Letting $t \to \infty$ and using the exponentially mean square stability of the dual system, we find

$$-Y = X\,A + A^T\,X + \sum_{k=1}^{q} \left(N^k\right)^T X\, N^k\, \mathbb{E}\left[M_k(1)^2\right]$$

which is the desired result. Since, (v) obviously implies (iv), it remains to show that $(iv) \to (ii)$. Let $X > 0$ such that

$$A^T X + XA + \sum_{k=1}^{q} (N^k)^T X N^k \mathbb{E}\left[M_k^2(1)\right] = -Y < 0. \tag{6.8}$$

So, due to Proposition 6.2, we have

$$\begin{aligned}
\mathbb{E}\left[x_{h,x_0}^T(t) X x_{h,x_0}(t)\right] &= \mathbb{E}\left[\operatorname{tr}\left(X x_{h,x_0}(t) x_{h,x_0}^T(t)\right)\right] = \operatorname{tr}\left(X \mathbb{E}\left[x_{h,x_0}(t) x_{h,x_0}^T(t)\right]\right) \\
&= \operatorname{tr}\Bigg(X \Bigg(x_0 x_0^T + \int_0^t \mathbb{E}\left[x_{h,x_0}(s) x_{h,x_0}^T(s)\right] ds\, A^T + A \int_0^t \mathbb{E}\left[x_{h,x_0}(s) x_{h,x_0}^T(s)\right] ds
\end{aligned}$$

$$+\sum_{k=1}^{q} N^k \int_0^t \mathbb{E}\left[x_{h,x_0}(s)x_{h,x_0}^T(s)\right] ds \left(N^k\right)^T \cdot c_k\Bigg)\Bigg)$$
$$= x_0^T X x_0 + \mathbb{E}\left[\int_0^t x_{h,x_0}^T(s)A^T X x_{h,x_0}(s)ds + \int_0^t x_{h,x_0}^T(s)XAx_{h,x_0}(s)ds\right]$$
$$+\mathbb{E}\left[\int_0^t \sum_{k=1}^{q} x_{h,x_0}^T(s)\left(N^k\right)^T XN^k \cdot c_k x_{h,x_0}(s)ds\right].$$

Inserting equation (6.8) yields

$$\mathbb{E}\left[x_{h,x_0}^T(t)Xx_{h,x_0}(t)\right] = x_0^T X x_0 - \mathbb{E}\left[\int_0^t x_{h,x_0}^T(s)Yx_{h,x_0}(s)ds\right].$$

and hence

$$\dot{g}(t) = -\mathbb{E}\left[x_{h,x_0}^T(t)Yx_{h,x_0}(t)\right],$$

where $g(t) := \mathbb{E}\left[x_{h,x_0}^T(t)Xx_{h,x_0}(t)\right]$. Now, let k_1 be the smallest and k_2 be the largest eigenvalue of X such that $k_1 v^T v \leq v^T X v \leq k_2 v^T v$. Furthermore, we assume k_3 to be the smallest eigenvalue of Y, then we obtain

$$\dot{g}(t) \leq -k_3\mathbb{E}\left[x_{h,x_0}^T(t)x_{h,x_0}(t)\right] \leq -\frac{k_3}{k_2}\mathbb{E}\left[\int_0^t x_{h,x_0}^T(s)Xx_{h,x_0}(s)ds\right] = -\frac{k_3}{k_2}g(t).$$

By Gronwall's inequality, we have

$$\mathbb{E}\left[x_{h,x_0}^T(t)x_{h,x_0}(t)\right] = \frac{1}{k_1}\mathbb{E}\left[x_{h,x_0}^T(t)Xx_{h,x_0}(t)\right] \leq \frac{1}{k_1}x_0^T X x_0\,\mathrm{e}^{-\frac{k_3}{k_2}t} \leq \frac{k_2}{k_1}x_0^T x_0\,\mathrm{e}^{-\frac{k_3}{k_2}t}$$

which yields the required result and concludes the proof. □

As in the deterministic case, there exists a fundamental solution, which we define by

$$\Phi(t) := \left[x_{h,e_1}(t), x_{h,e_2}(t), \ldots, x_{h,e_n}(t)\right]$$

for $t \geq 0$, where e_i is the i-th unit vector ($i = 1, \ldots, n$). Thus, Φ fulfills the following integral equation:

$$\Phi(t) = I_n + \int_0^t A\Phi(s)ds + \sum_{k=1}^{q}\int_0^t N^k\Phi(s-)dM_k(s).$$

The columns of Φ represent a minimal generating set such that we have $x_{h,x_0}(t) = \Phi(t)x_0$. With $B = [b_1, b_2, \ldots, b_m]$ one can see that

$$\Phi(t)B = [\Phi(t)b_1, \Phi(t)b_2, \ldots, \Phi(t)b_m] = \left[x_{h,b_1}(t), x_{h,b_2}(t) \ldots, x_{h,b_m}(t)\right].$$

Hence, we have

$$\Phi(t)BB^T\Phi^T(t) = x_{h,b_1}(t)x_{h,b_1}^T(t) + x_{h,b_2}(t)x_{h,b_2}^T(t) + \ldots + x_{h,b_m}(t)x_{h,b_m}^T(t),$$

such that

$$\begin{aligned}\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right] = BB^T + \int_0^t \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds\, A^T + A\int_0^t \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds \\ + \sum_{k=1}^q N^k \int_0^t \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds\, (N^k)^T\, \mathbb{E}\left[M_k(1)^2\right]\end{aligned} \tag{6.9}$$

holds for every $t \geq 0$. Due to the assumption that the homogeneous solution x_{h,x_0} is asymptotically mean square stable for an arbitrary initial value x_0, yielding $\mathbb{E}\left[x_{h,x_0}^T(t)x_{h,x_0}(t)\right] \to 0$ for $t \to \infty$, we obtain

$$\begin{aligned}0 = BB^T + \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds\, A^T + A\int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds \\ + \sum_{k=1}^q N^k \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds\, (N^k)^T\, \mathbb{E}\left[M_k(1)^2\right]\end{aligned}$$

by taking the limit $t \to \infty$ in equation (6.9). Therefore, we can conclude that the matrix $P := \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds$, which exists by the asymptotic mean square stability assumption, is the solution to a generalized Lyapunov equation

$$AP + PA^T + \sum_{k=1}^q N^k P \left(N^k\right)^T \mathbb{E}\left[M_k(1)^2\right] = -BB^T.$$

P is the reachability Gramian of system (6.1), where this definition of the Gramian is also used in Benner, Damm [9] for stochastic systems driven by Wiener noise. Note that in this case $\mathbb{E}\left[M_k(1)^2\right] = 1$.

Remark. *The solution to the matrix equation*

$$0 = BB^T + AP + PA^T + \sum_{k=1}^{q} N^k P (N^k)^T \cdot \mathbb{E}\left[M_k(1)^2\right] \tag{6.10}$$

is unique if and only if the solution to

$$-\operatorname{vec}(BB^T) = \left(I_n \otimes A + A \otimes I_n + \sum_{k=1}^{q} N^k \otimes N^k \cdot \mathbb{E}\left[M_k(1)^2\right]\right) \operatorname{vec}(P)$$

is unique. By the assumption of mean square asymptotic stability the eigenvalues of the matrix $I \otimes A + A \otimes I + \sum_{k=1}^{q} N^k \otimes N^k \cdot \mathbb{E}\left[M_k(1)^2\right]$ are non-zero, hence the matrix equation (6.10) is uniquely solvable.

More general, we consider stochastic processes $(\Phi(t,\tau))_{t\geq\tau}$ with starting time $\tau \geq 0$ and initial condition $\Phi(\tau,\tau) = I_n$ satisfying

$$\Phi(t,\tau) = I_n + \int_\tau^t A\Phi(s,\tau)ds + \sum_{k=1}^{q} \int_\tau^t N^k \Phi(s-,\tau) dM_k(s) \tag{6.11}$$

for $t \geq \tau \geq 0$. We have $\Phi(t,0) = \Phi(t)$. Analogous to equation (6.9), we can show that

$$\begin{aligned} \mathbb{E}\left[\Phi(t,\tau)BB^T\Phi^T(t,\tau)\right] = BB^T &+ \int_\tau^t \mathbb{E}\left[\Phi(s,\tau)BB^T\Phi^T(s,\tau)\right] ds\, A^T \\ &+ A \int_\tau^t \mathbb{E}\left[\Phi(s,\tau)BB^T\Phi^T(s,\tau)\right] ds \\ &+ \sum_{k=1}^{q} N^k \int_\tau^t \mathbb{E}\left[\Phi(s,\tau)BB^T\Phi^T(s,\tau)\right] ds\, (N^k)^T\, \mathbb{E}\left[M_k(1)^2\right]. \end{aligned} \tag{6.12}$$

This yields that $\mathbb{Y}(t) := \mathbb{E}\left[\Phi(t,\tau)BB^T\Phi^T(t,\tau)\right]$ is the solution to the differential equation

$$\dot{\mathbb{Y}}(t) = A\mathbb{Y}(t) + \mathbb{Y}(t)A^T + \sum_{k=1}^{q} N^k \mathbb{Y}(t)(N^k)^T\, \mathbb{E}\left[M_k(1)^2\right] \tag{6.13}$$

for $t \geq \tau$ with initial condition $\mathbb{Y}(\tau) = BB^T$.

Remark. *For $t \geq \tau \geq 0$, we have $\Phi(t,\tau) = \Phi(t)\Phi^{-1}(\tau)$, since $\Phi(t)\Phi^{-1}(\tau)$ fulfills equation (6.11).*
Compared to the deterministic case ($N^k = 0$) we do not have the semigroup property for the

fundamental solution. So, it is not true that $\Phi(t,\tau) = \Phi(t-\tau)$ $\mathbb{P}$-almost surely holds, because the trajectories of the noise processes on $[0,t-\tau]$ and $[\tau,t]$ are different in general. We can however conclude that $\mathbb{E}\left[\Phi(t,\tau)BB^T\Phi^T(t,\tau)\right] = \mathbb{E}\left[\Phi(t-\tau)BB^T\Phi^T(t-\tau)\right]$, since both terms solve equation (6.13) as can be seen employing (6.9).

Now, we derive the solution representation of the system (6.1) via using the stochastic variation of constants method. For the Wiener case, this result is stated in Theorem 1.4.1 in Damm [21].

Proposition 6.4. *$(\Phi(t)z(t))_{t\geq 0}$ is a solution to equation (6.1), where z satisfies the differential equation*

$$dz(t) = \Phi^{-1}(t)Bu(t)dt, \quad z(0) = x_0.$$

Proof. We want to determine the stochastic differential of $\Phi(t)z(t)$, $t \geq 0$, where its i-th component is given by $e_i^T\Phi(t)z(t)$. Applying the Ito product formula from Corollary A.4 yields

$$e_i^T\Phi(t)z(t) = e_i^T + \int_0^t e_i^T\Phi(s-)d(z(s)) + \int_0^t z^T(s)d(\Phi^T(s)e_i).$$

Above, the quadratic covariation terms are zero, since z is a continuous semimartingale with a martingale part of zero (see equation (A.2)). Applying that $s \mapsto \Phi(\omega,s)$ and $s \mapsto \Phi(\omega,s-)$ coincide almost everywhere with respect to the Lebesgue measure for $\mathbb{P}$-almost all fixed $\omega \in \Omega$, we have

$$\begin{aligned}
e_i^T\Phi(t)z(t) &= e_i^T + \int_0^t e_i^T\Phi(s)\Phi^{-1}(s)Bu(s)ds + \int_0^t z^T(s)\Phi^T(s)A^Te_ids \\
&\quad + \sum_{k=1}^q \int_0^t z^T(s)\Phi^T(s-)(N^k)^Te_idM_k(s) \\
&= e_i^Tx_0 + e_i^T\int_0^t Bu(s)ds + e_i^T\int_0^t A\Phi(s)z(s)ds + e_i^T\sum_{k=1}^q\int_0^t N^k\Phi(s-)z(s)dM_k(s).
\end{aligned}$$

This yields

$$\Phi(t)z(t) = x_0 + \int_0^t A\Phi(s)z(s)ds + \sum_{k=1}^q \int_0^t N^k\Phi(s-)z(s)dM_k(s) + \int_0^t Bu(s)ds$$

and hence the required result. $\square$

Below, we set $P_t := \int_0^t \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right]ds$ and call P_t finite *reachability Gramian* at time $t \geq$

0. Furthermore, we define the so-called finite deterministic Gramian $P_{D,t} := \int_0^t \mathrm{e}^{As} BB^T \mathrm{e}^{A^T s} ds$. P_t and $P_{D,t}$, $t \geq 0$, coincide in the case $N^k = 0$. By $x(T,0,u)$ we denote the solution to the inhomogeneous system (6.1) at time T with initial condition zero for a given input u. From Proposition 6.4, we already know that

$$x(T,0,u) = \int_0^T \Phi(T)\Phi^{-1}(t)Bu(t)dt = \int_0^T \Phi(T,t)Bu(t)dt.$$

Now, we have the goal to steer the average state of the system (6.1) from zero to any given $x \in \mathbb{R}^n$ via the control u with minimal energy. First of all we need the following definition, which is motivated by the remarks above Theorem 2.3 in [9].

Definition 6.5. *A state $x \in \mathbb{R}^n$ is called* reachable on average (from zero) *if there is a time $T > 0$ and a control function $u \in L_T^2$, such that we have*

$$\mathbb{E}[x(T,0,u)] = x.$$

We say that the stochastic system is *completely reachable* if every average vector $x \in \mathbb{R}^n$ is reachable. Next, we characterize the set of all reachable average states. First of all, we need the following proposition, where we define $P := \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds$ in analogy to the deterministic case.

Proposition 6.6. *The finite reachability Gramians P_t, $t > 0$, have the same image as the infinite reachability Gramian P, that is,*

$$\operatorname{im} P_t = \operatorname{im} P$$

for all $t > 0$.

Proof. Since P and P_t are positive semidefinite and symmetric by definition it is sufficient to show that their kernels are equal. First, we assume $v \in \ker P$. Thus,

$$0 \leq v^T P_t v \leq v^T P v = 0,$$

since $t \mapsto v^T P_t v$ is increasing such that $v \in \ker P_t$ follows. On the other hand, if $v \in \ker P_t$ we have

$$0 = v^T P_t v = \int_0^t v^T \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] v ds.$$

Hence, we can conclude that $v^T \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] v = 0$ for almost all $s \in [0,t]$. Additionally, we know that $t \mapsto \mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right]$ is the solution to the linear matrix differential equation

$$\dot{\mathbb{Y}}(t) = A\mathbb{Y}(t) + \mathbb{Y}(t)A^T + \sum_{k=1}^{q} N^k \mathbb{Y}(t)(N^k)^T \,\mathbb{E}\left[M_k(1)^2\right]$$

with initial condition $\mathbb{Y}(0) = BB^T$ for $t \geq 0$. The vectorized form $\mathrm{vec}(\mathbb{Y})$ satisfies

$$\mathrm{vec}(\dot{\mathbb{Y}}(t)) = \left(I_n \otimes A + A \otimes I_n + \sum_{k=1}^{q} N^k \otimes N^k \cdot \mathbb{E}\left[M_k(1)^2\right]\right) \mathrm{vec}(\mathbb{Y}(t)), \quad \mathrm{vec}(\mathbb{Y}(0)) = \mathrm{vec}(BB^T).$$

Thus, the entries of $\mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right]$ are analytic functions. This implies that the function $f(t) := v^T \mathbb{E}\left[\Phi(t)BB^T\Phi^T(t)\right] v$ is analytic, such that $f \equiv 0$ on $[0,\infty)$. Thus,

$$0 = \int_0^\infty v^T \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] v ds = v^T P v$$

and the result follows. □

The next proposition shows that the reachable average states are characterized by the deterministic Gramian $P_D := \int_0^\infty \mathrm{e}^{As} BB^T \mathrm{e}^{A^T s} ds$, which exists due to the asymptotic stability of the matrix A, which is a necessary condition for asymptotic mean square stability of system (6.1).

Proposition 6.7. *An average state $x \in \mathbb{R}^n$ is reachable (from zero) if and only if $x \in \operatorname{im} P_D$, where $P_D := \int_0^\infty \mathrm{e}^{As} BB^T \mathrm{e}^{A^T s} ds$.*

Proof. Provided $x \in \operatorname{im} P_D$, we will show that this average state can be reached with the following input function:

$$[0,T] \ni t \mapsto u(t) = B^T \mathrm{e}^{A^T(T-t)} P_{D,T}^{\#} x, \tag{6.14}$$

where $P_{D,T}^{\#}$ denotes the Moore-Penrose pseudoinverse of $P_{D,T}$. Thus, we obtain

$$\mathbb{E}\left[x(T,0,u)\right] = \mathbb{E}\left[\int_0^T \Phi(T,t)BB^T \mathrm{e}^{A^T(T-t)} P_{D,T}^{\#} x dt\right]$$

by inserting the function u. Applying the expectation to both sides of equation (6.11) yields

$$\mathbb{E}\left[\Phi(t,\tau)\right] = \mathrm{e}^{A(t-\tau)}.$$

Using this fact, we obtain

$$\mathbb{E}\left[x(T,0,u)\right] = \int_0^T \mathrm{e}^{A(T-t)} BB^T \mathrm{e}^{A^T(T-t)} P_{D,T}^{\#} x dt.$$

We substitute $s = T - t$ and since $x \in \operatorname{im} P_{D,T}$ by Proposition 6.6, we get

$$\mathbb{E}\left[x(T,0,u)\right] = \int_0^T \mathrm{e}^{As} BB^T \mathrm{e}^{A^T s} ds P_{D,T}^{\#} x = P_{D,T} P_{D,T}^{\#} x = x.$$

The energy of the input function $u(t) = B^T \mathrm{e}^{A^T(T-t)} P_{D,T}^{\#} x$ is

$$\|u\|_{L_T^2}^2 = x^T P_{D,T}^{\#} x < \infty.$$

On the other hand, if $x \in \mathbb{R}^n$ is reachable, then there exists an input function u and a time $t > 0$ such that, by definition,

$$x = \mathbb{E}\left[x(t,0,u)\right] = \mathbb{E}\left[\int_0^t \Phi(t,s) B u(s) ds\right] = \int_0^t \mathrm{e}^{A(t-s)} B \mathbb{E}\left[u(s)\right] ds.$$

We get the last equation by applying the expectation to both sides of equation (6.1). We assume that $v \in \ker P_D$. Hence,

$$|\langle x, v\rangle_2| = \left|\int_0^t \left\langle \mathrm{e}^{A(t-s)} B \mathbb{E}\left[u(s)\right], v\right\rangle_2 ds\right| = \left|\int_0^t \left\langle \mathbb{E}\left[u(s)\right], B^T \mathrm{e}^{A^T(t-s)} v\right\rangle_2 ds\right|.$$

Employing the Cauchy-Schwarz inequality, we get

$$|\langle x, v\rangle_2| \leq \int_0^t \|\mathbb{E}\left[u(s)\right]\|_2 \left\|B^T \mathrm{e}^{A^T(t-s)} v\right\|_2 ds \leq \int_0^t \left(\mathbb{E}\|u(s)\|_2^2\right)^{\frac{1}{2}} \left\|B^T \mathrm{e}^{A^T(t-s)} v\right\|_2 ds$$

and by the Hölder inequality, we have

$$\begin{aligned} |\langle x, v\rangle_2| &\leq \|u\|_{L_t^2} \left(\int_0^t \left\|B^T \mathrm{e}^{A^T(t-s)} v\right\|_2^2 ds\right)^{\frac{1}{2}} \\ &= \|u\|_{L_t^2} \left(v^T \int_0^t \mathrm{e}^{A(t-s)} BB^T \mathrm{e}^{A^T(t-s)} ds\, v\right)^{\frac{1}{2}} = \|u\|_{L_t^2} \left(v^T P_{D,t} v\right)^{\frac{1}{2}}. \end{aligned}$$

Since $t \mapsto v^T P_{D,t} v$ is increasing, we obtain

$$|\langle x, v\rangle_2| \leq \|u\|_{L_t^2} \left(v^T P_D v\right)^{\frac{1}{2}} = 0.$$

Thus, $\langle x, v\rangle_2 = 0$, such that we can conclude that $x \in \operatorname{im} P_D$ due to $\operatorname{im} P_D = (\ker P_D)^{\perp}$. □

Below, we point out the relation between the reachable set and the reachability Gramian $P := \int_0^\infty \mathbb{E}\left[\Phi(s) B B^T \Phi^T(s)\right] ds$.

Proposition 6.8. *If an average state $x \in \mathbb{R}^n$ is reachable (from zero), then $x \in \operatorname{im} P$.*

Proof. By definition, there exists an input function u and a time $t > 0$ such that

$$x = \mathbb{E}\left[x(t, 0, u)\right] = \mathbb{E}\left[\int_0^t \Phi(t,s) B u(s) ds\right]$$

for reachable $x \in \mathbb{R}^n$. We assume that $v \in \ker P$. So, we have

$$|\langle x, v\rangle_2| = \left|\mathbb{E}\left[\int_0^t \langle \Phi(t,s) B u(s), v\rangle_2 \, ds\right]\right| = \left|\mathbb{E}\left[\int_0^t \left\langle u(s), B^T \Phi^T(t,s) v\right\rangle_2 ds\right]\right|.$$

Employing the Cauchy-Schwarz inequality, we obtain

$$|\langle x, v\rangle_2| \leq \mathbb{E}\left[\int_0^t \|u(s)\|_2 \left\|B^T \Phi^T(t,s) v\right\|_2 ds\right].$$

By the Hölder inequality, we have

$$|\langle x, v\rangle_2| \leq \|u\|_{L_t^2} \left(\mathbb{E}\left[\int_0^t \left\|B^T \Phi^T(t,s) v\right\|_2^2 ds\right]\right)^{\frac{1}{2}} = \|u\|_{L_t^2} \left(v^T \mathbb{E}\left[\int_0^t \Phi(t,s) B B^T \Phi^T(t,s) ds\right] v\right)^{\frac{1}{2}}.$$

With the remarks above Proposition 6.4, we obtain

$$\mathbb{E}\left[\Phi(t-s) B B^T \Phi^T(t-s)\right] = \mathbb{E}\left[\Phi(t,s) B B^T \Phi^T(t,s)\right],$$

such that

$$|\langle x, v\rangle_2| \leq \|u\|_{L_t^2} \left(v^T P_t v\right)^{\frac{1}{2}}.$$

Since $t \mapsto v^T P_t v$ is increasing, it follows

$$|\langle x, v\rangle_2| \leq \|u\|_{L_t^2} \left(v^T P v\right)^{\frac{1}{2}} = 0.$$

Thus, $\langle x, v\rangle_2 = 0$, such that we can conclude that $x \in \operatorname{im} P$ due to $\operatorname{im} P = (\ker P)^\perp$. □

Consequently, by Propositions 6.7 and 6.8, we have $\operatorname{im} P_D \subseteq \operatorname{im} P$. Now, we state the minimal energy to steer the system to a desired average state.

Proposition 6.9. *Let $x \in \mathbb{R}^n$ be reachable, then the input function given by (6.14) is the one with the minimal energy to reach x at any time $T > 0$. This minimal energy is given by $x^T P_{D,T}^{\#} x$, where $P_{D,T}^{\#}$ is the Moore-Penrose pseudoinverse of $P_{D,T}$.*

Proof. We use the following representation from the proof of Proposition 6.7:

$$\mathbb{E}\left[x(T,0,u)\right] = \int_0^T \mathrm{e}^{A(T-t)} B \mathbb{E}\left[u(t)\right] dt.$$

Let $u(t)$ be as in (6.14) and $\tilde{u}(t)$, $t \in [0,T]$, an additional function for which we can reach the average state x at time T, then

$$\int_0^T \mathrm{e}^{A(T-t)} B \left(\mathbb{E}\left[\tilde{u}(t)\right] - u(t)\right) dt = 0,$$

such that

$$\mathbb{E}\left[\int_0^T u(t)^T \left(\tilde{u}(t) - u(t)\right) dt\right] = \int_0^T u(t)^T \left(\mathbb{E}\left[\tilde{u}(t)\right] - u(t)\right) dt = 0$$

follows. Hence, with (6.14), we have

$$\|\tilde{u}\|_{L_T^2}^2 = \|u + (\tilde{u} - u)\|_{L_T^2}^2 = \|u\|_{L_T^2}^2 + \|\tilde{u} - u\|_{L_T^2}^2 \geq \|u\|_{L_T^2}^2.$$

From the proof of Proposition 6.7, we know that the energy of u is given by $x^T P_{D,T}^{\#} x$. □

The following result shows that the finite reachability Gramian P_T provides information about the degree of reachability of an average state as well.

Proposition 6.10. *Let $x \in \mathbb{R}^n$ be reachable, then*

$$x^T P_T^{\#} x \leq x^T P_{D,T}^{\#} x$$

for every time $T > 0$*.*

Proof. Since x is reachable, $x \in \operatorname{im} P_T$ by Proposition 6.6 and Proposition 6.8. Hence, we can write $x = P_T P_T^{\#} x$, where $P_T^{\#}$ denotes the Moore-Penrose pseudoinverse of P_T. From its definition, the finite reachability Gramian is represented by $P_T = \mathbb{E}\left[\int_0^T \Phi(T-t)BB^T\Phi^T(T-t)dt\right]$ and since

$$\mathbb{E}\left[\Phi(T-t)BB^T\Phi^T(T-t)\right] = \mathbb{E}\left[\Phi(T,t)BB^T\Phi^T(T,t)\right],$$

we have

$$x = \mathbb{E}\left[\int_0^T \Phi(T,t)BB^T\Phi^T(T,t)P_T^{\#}x dt\right].$$

Now, we choose the control $u(t) = B^T \mathrm{e}^{A^T(T-t)} P_{D,T}^{\#} x$, $t \in [0,T]$, of minimal energy to reach x, then

$$\mathbb{E}\left[\int_0^T \Phi(T,t)B\left(B^T\Phi^T(T,t)P_T^{\#}x - u(t)\right)dt\right] = 0.$$

Setting $v(t) = B^T\Phi^T(T,t)P_T^{\#}x$ for $t \in [0,T]$ yields

$$\mathbb{E}\left[\int_0^T v^T(t)\left(v(t) - u(t)\right)dt\right] = 0.$$

We therefore obtain

$$x^T P_{D,T}^{\#} x = \|u\|_{L_T^2}^2 = \|v + (u-v)\|_{L_T^2}^2 = \|v\|_{L_T^2}^2 + \|u-v\|_{L_T^2}^2 \geq \|v\|_{L_T^2}^2 = x^T P_T^{\#} x$$

which gives the result. □

Consequently, the expression $x^T P_T^{\#} x$ yields a lower bound for the energy to reach x and is the L_T^2-norm squared of the function $v(t) = B^T\Phi^T(T,t)P_T^{\#}x$, $t \in [0,T]$. With the control v we would also be able to steer the system to x in case it would be a admissible control. Unfortunately, unavailable future information enters in v which means that it is not $(\mathscr{F}_t)_{t\in[0,T]}$- adapted. So, one can interpret the energy $x^T\left(P_{T,D}^{\#} - P_T^{\#}\right)x$ as the benefit of knowing the future until time T.
By Proposition 6.9, the minimal energy that is needed to steer the system to x is given by $\inf_{T>0} x^T P_{D,T}^{\#} x$. By definition of the family of matrices $P_{D,T}$ we know that it is increasing in

time such that the pseudoinverse $P_{D,T}^{\#}$ is decreasing in time. Hence, it is clear that the minimal energy is given by $x^T P_D^{\#} x$, where $P_D^{\#}$ is the pseudoinverse of the deterministic Gramian P_D. The result in Proposition 6.10 provides a lower bound for the minimal energy to reach x:

$$x^T P^{\#} x \leq x^T P_D^{\#} x \tag{6.15}$$

with $P^{\#}$ being the pseudoinverse of the reachability Gramian P. Using inequality (6.15), we get only partial information about the degree of reachability of an average state x from $P^{\#}$. So, it remains an open question whether an alternative reachability concept would be more suitable to motivate the Gramian P.
Similar results are obtained by Benner and Damm [9] in Theorem 2.3 for stochastic differential equations driven by Wiener processes. For the deterministic case we refer to Subsection 4.3.1 in Antoulas [2].

6.1.2. Observability concept

Below, we introduce the concept of observability for the output equation

$$y(t) = Cx(t) \tag{6.16}$$

corresponding to the stochastic linear system (6.1), where $C \in \mathbb{R}^{p\times n}$. Therefore, we need the following proposition.

Proposition 6.11. *Let $\hat{Q}$ be a symmetric positive semidefinite matrix and let $x_{h,a} := x(\cdot, a, 0)$, $x_{h,b} := x(\cdot, b, 0)$ be the homogeneous solutions to (6.1) with initial conditions $a, b \in \mathbb{R}^n$, then*

$$\mathbb{E}\left[x_{h,a}(t)^T \hat{Q} x_{h,b}(t)\right] = a^T \hat{Q} b + \mathbb{E}\left[\int_0^t x_{h,a}^T(s)\hat{Q}Ax_{h,b}(s)ds\right] + \mathbb{E}\left[\int_0^t x_{h,a}^T(s)A^T\hat{Q}x_{h,b}(s)ds\right]$$
$$+\mathbb{E}\left[\int_0^t x_{h,a}(s)^T \sum_{k=1}^{q} (N^k)^T \hat{Q} N^k \mathbb{E}\left[M_k(1)^2\right] x_{h,b}(s)ds\right]. \tag{6.17}$$

Proof. By applying the Ito product formula from Corollary A.4, we have

$$x_{h,a}^T(t)\hat{Q}x_{h,b}(t) = a^T\hat{Q}b + \int_0^t x_{h,a}^T(s-)d(\hat{Q}x_{h,a}(s)) + \int_0^t x_{h,b}^T(s-)\hat{Q}d(x_{h,a}(s)) + \sum_{i=1}^{n}[e_i^T x_{h,a}, e_i^T\hat{Q}x_{h,b}]_t,$$

where e_i is the i-th unit vector ($i = 1, \dots, n$). We get

$$\int_0^t x_{h,a}^T(s-) d(\hat{Q} x_{h,b}(s)) = \int_0^t x_{h,a}^T(s-) \hat{Q} A x_{h,b}(s) ds + \sum_{k=1}^q \int_0^t x_{h,a}^T(s-) \hat{Q} N^k x_{h,b}(s-) dM_k(s)$$

and

$$\int_0^t x_{h,b}^T(s-) \hat{Q} d(x_{h,a}(s)) = \int_0^t x_{h,b}(s-)^T \hat{Q} A x_{h,a}(s) ds + \sum_{k=1}^q \int_0^t x_{h,b}(s-)^T \hat{Q} N^k x_{h,a}(s-) dM_k(s).$$

By equation (A.5), the mean of the quadratic covariations is given by

$$\mathbb{E}[e_i^T x_{h,a}, e_i^T \hat{Q} x_{h,b}]_t = \sum_{k=1}^q \mathbb{E} \int_0^t e_i^T N^k x_{h,a}(s) e_i^T \hat{Q} N^k x_{h,b}(s) ds \, \mathbb{E}\left[M_k(1)^2\right].$$

Since by Section 3.1 the Ito integrals have mean zero, we obtain

$$\begin{aligned}\mathbb{E}\left[x_{h,a}(t)^T \hat{Q} x_{h,b}(t)\right] = a^T \hat{Q} b + \mathbb{E}\left[\int_0^t x_{h,a}^T(s) \hat{Q} A x_{h,b}(s) ds\right] + \mathbb{E}\left[\int_0^t x_{h,a}^T(s) A^T \hat{Q} x_{h,b}(s) ds\right] \\ + \sum_{k=1}^q \mathbb{E}\left[\int_0^t x_{h,a}(s)^T (N^k)^T \hat{Q} N^k x_{h,b}(s) ds\right] \mathbb{E}\left[M_k(1)^2\right]\end{aligned}$$

using that the trajectories of $x_{h,a}$ and $x_{h,b}$ only have jumps on Lebesgue zero sets. □

If we set $a = e_i$ and $b = e_j$ in Proposition 6.11, we obtain

$$\begin{aligned}\mathbb{E}\left[e_i^T \Phi(t)^T \hat{Q} \Phi(t) e_j\right] = e_i^T \hat{Q} e_j + \mathbb{E}\left[\int_0^t e_i^T \Phi^T(s) \hat{Q} A \Phi(s) e_j ds\right] + \mathbb{E}\left[\int_0^t e_i^T \Phi^T(s) A^T \hat{Q} \Phi(s) e_j ds\right] \\ + \mathbb{E}\left[\int_0^t e_i^T \Phi(s)^T \sum_{k=1}^q (N^k)^T \hat{Q} N^k \mathbb{E}\left[M_k(1)^2\right] \Phi(s) e_j ds\right].\end{aligned}$$

This yields (letting $i, j = 1, \dots n$)

$$\begin{aligned}\mathbb{E}\left[\Phi(t)^T \hat{Q} \Phi(t)\right] = \hat{Q} + \mathbb{E}\left[\int_0^t \Phi^T(s) \hat{Q} A \Phi(s) ds\right] + \mathbb{E}\left[\int_0^t \Phi^T(s) A^T \hat{Q} \Phi(s) ds\right] \\ + \mathbb{E}\left[\int_0^t \Phi(s)^T \sum_{k=1}^q (N^k)^T \hat{Q} N^k \mathbb{E}\left[M_k(1)^2\right] \Phi(s) ds\right].\end{aligned}$$

Let Q be the solution of the generalized Lyapunov equation

$$A^T Q + QA + \sum_{k=1}^{q} (N^k)^T Q N^k \, \mathbb{E}\left[M_k(1)^2\right] = -C^T C. \tag{6.18}$$

Then,

$$\mathbb{E}\left[\Phi(t)^T Q \Phi(t)\right] = Q - \mathbb{E}\left[\int_0^t \Phi^T(s) C^T C \Phi(s) ds\right]$$

and by taking the limit $t \to \infty$, we have

$$Q = \mathbb{E}\left[\int_0^\infty \Phi^T(s) C^T C \Phi(s) ds\right], \tag{6.19}$$

due to the asymptotic mean square stability of the homogeneous equation ($u \equiv 0$), which provides the existence of the integral in equation (6.19) as well.

Remark. *The matrix equation (6.18) is uniquely solvable, since*

$$L := \left(A^T \otimes I_n + I_n \otimes A^T + \sum_{k=1}^{q} (N^k)^T \otimes (N^k)^T \cdot \mathbb{E}\left[M_k(1)^2\right]\right)$$

has non zero eigenvalues and hence the solution of $L \cdot \operatorname{vec}(Q) = -\operatorname{vec}(C^T C)$ *is unique.*

Next, we assume that the system (6.1) is uncontrolled, that means $u \equiv 0$. By using our knowledge concerning the homogeneous system, $x(t, x_0, 0)$ is given by $\Phi(t)x_0$, where here, $x_0 \in \mathbb{R}^n$ denotes the initial value of the system. Hence, we obtain $y(t) = C\Phi(t)x_0$.
We observe y on a time interval $[0, \infty)$. The problem is to find x_0 from given observations. The energy produced by the initial value x_0 is

$$\|y\|_{L^2}^2 := \mathbb{E}\int_0^\infty y^T(t) y(t) dt = x_0^T \mathbb{E}\int_0^\infty \Phi^T(t) C^T C \Phi(t) dt \, x_0 = x_0^T Q x_0, \tag{6.20}$$

where we set $Q := \mathbb{E}\int_0^\infty \Phi^T(s) C^T C \Phi(s) ds$. As in Benner, Damm [9], Q takes the part of the observability Gramian of the stochastic system with output equation (6.16). We call a state x_0 unobservable if it is in the kernel of Q. Otherwise it is said to be observable. We say that a system is completely observable if the kernel of Q is trivial.

6.2. Type 1 balanced truncation for stochastic systems

For obtaining a reduced order model for a deterministic linear time-invariant system, balanced truncation is a method of major importance. For the procedure of balanced truncation in the deterministic case, see Antoulas [2], Benner et al. [12] and Obinata, Anderson [54]. In this section, we want to generalize this method for stochastic linear systems, which are influenced by Lévy noise.

6.2.1. Procedure

Below, we summarize the results of Section 4.1 in [14]. We assume $A, N^k \in \mathbb{R}^{n\times n}$ $(k=1,\dots,q)$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$, and consider the following stochastic system that may occur after a space discretization of a stochastic PDE, see Sections 5.1 and 5.2:

$$
\begin{aligned}
dx(t) &= [Ax(t)+Bu(t)]dt + \sum_{k=1}^{q} N^k x(t-)dM_k(t), \quad t\geq 0, \quad x(0)=x_0, && (6.21)\\
y(t) &= Cx(t),
\end{aligned}
$$

where the noise processes M_k $(k=1,\dots,q)$ are uncorrelated real-valued and square integrable Lévy processes with mean zero. We assume the homogeneous solution x_{h,x_0}, which satisfies

$$
dx_h(t) = Ax_h(t)dt + \sum_{k=1}^{q} N^k x_h(t-)dM_k(t), \quad t\geq 0, \quad x_h(0)=x_0,
$$

to be mean square asymptotically stable. In addition, we require that the system (6.21) is completely reachable and observable, which is equivalent to P_D and Q being positive definite. Hence, the reachability Gramian P is also positive definite using Propositions 6.7 and 6.8.

Let $\hat{T} \in \mathbb{R}^{n\times n}$ be a regular matrix. If we transform the states using

$$
\hat{x}(t) = \hat{T}x(t),
$$

we obtain the following system:

$$
\begin{aligned}
d\hat{x}(t) &= [\tilde{A}\hat{x}(t)+\tilde{B}u(t)]dt + \sum_{k=1}^{q} \tilde{N}^k \hat{x}(t-)dM_k(t), \quad \hat{x}(0)=\hat{T}x_0, && (6.22)\\
y(t) &= \tilde{C}\hat{x}(t), \quad t\geq 0,
\end{aligned}
$$

where $\tilde{A} = \hat{T}A\hat{T}^{-1}$, $\tilde{N}^k = \hat{T}N^k\hat{T}^{-1}$, $\tilde{B} = \hat{T}B$ and $\tilde{C} = C\hat{T}^{-1}$. For an arbitrary fixed input u, the transformed system (6.22) has always the same output as the system (6.21).
The reachability Gramian $P := \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds$ of system (6.21) fulfills

$$-BB^T = AP + PA^T + \sum_{k=1}^{q} N^k P(N^k)^T \cdot c_k,$$

where $c_k = \mathbb{E}\left[M_k(1)^2\right]$. By multiplying with $\hat{T}$ from the left and $\hat{T}^T$ from the right, we obtain

$$\begin{aligned}-\tilde{B}\tilde{B}^T &= \hat{T}AP\hat{T}^T + \hat{T}PA^T\hat{T}^T + \sum_{k=1}^{q} \hat{T}N^k P(N^k)^T\hat{T}^T \cdot c_k \\ &= \tilde{A}\hat{T}P\hat{T}^T + \hat{T}P\hat{T}^T\tilde{A}^T + \sum_{k=1}^{q} \tilde{N}^k\hat{T}P\hat{T}^T(\tilde{N}^k)^T \cdot c_k.\end{aligned}$$

Hence, the reachability Gramian of the transformed system (6.22) is given by $\tilde{P} = \hat{T}P\hat{T}^T$. For the observability Gramian of the transformed system $\tilde{Q} = \hat{T}^{-T}Q\hat{T}^{-1}$ holds, where the matrix $Q := \int_0^\infty \mathbb{E}\left[\Phi^T(s)C^TC\Phi(s)\right] ds$ is the observability Gramian of the original system. Hence,

$$-\tilde{C}^T\tilde{C} = \tilde{A}^T\tilde{Q} + \tilde{Q}\tilde{A} + \sum_{k=1}^{q} (\tilde{N}^k)^T\tilde{Q}\tilde{N}^k \cdot c_k.$$

In addition, it is easy to verify that the generalized Hankel singular values $\sigma_1 \geq \ldots \geq \sigma_n > 0$ of (6.21), where $\sigma_i = \sqrt{\text{eig}_i(PQ)}$ $(i = 1, \ldots, n)$, are equal to those of (6.22).

As in the deterministic case (see [2] and [54]), we choose $\hat{T}$ such that $\tilde{Q}$ and $\tilde{P}$ are equal and diagonal. A system with equal and diagonal Gramians is called balanced. The corresponding balancing $\hat{T}$ is given by

$$\hat{T} = \Sigma^{\frac{1}{2}}K^TU^{-1} \text{ and } \hat{T}^{-1} = UK\Sigma^{-\frac{1}{2}}, \tag{6.23}$$

where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$, U comes from the Cholesky decomposition of $P = UU^T$ and K is an orthogonal matrix corresponding to the eigenvalue decomposition (singular value decomposition, respectively) of $U^TQU = K\Sigma^2K^T$. Therefore, we obtain

$$\tilde{Q} = \tilde{P} = \Sigma.$$

Our aim is to truncate the average states that are difficult to observe and difficult to reach, which are those producing least observation energy and causing the most energy to reach, respectively. By equation (6.20), we can say that the states which are difficult to observe are contained in the space spanned by the eigenvectors corresponding to the small eigenvalues of Q. Using (6.15), an average state x is particularly difficult to reach if the expression $x^T P^{-1} x$ is large. Those states are contained in the space spanned by the eigenvectors corresponding to the small eigenvalues of P (or to the large eigenvalues of P^{-1}, respectively). The eigenspaces that correspond to the small eigenvalues of P contain all difficult-to-reach states if we would know the future completely, see the remarks below Proposition 6.10. In a balanced system, the dominant reachable and observable states are the same.

We consider the following partitions:

$$\hat{T} = \begin{bmatrix} W^T \\ T_2^T \end{bmatrix}, \hat{T}^{-1} = \begin{bmatrix} V & T_1 \end{bmatrix} \text{ and } \hat{x} = \begin{pmatrix} \tilde{x} \\ x_1 \end{pmatrix},$$

where $W^T \in \mathbb{R}^{r\times n}, V \in \mathbb{R}^{n\times r}$ and $\tilde{x}$ takes values in $\mathbb{R}^r$ $(r < n)$. Hence, we have

$$\begin{aligned} \begin{pmatrix} d\tilde{x}(t) \\ dx_1(t) \end{pmatrix} &= \left(\begin{bmatrix} W^T AV & W^T AT_1 \\ T_2^T AV & T_2^T AT_1 \end{bmatrix} \begin{pmatrix} \tilde{x}(t) \\ x_1(t) \end{pmatrix} + \begin{bmatrix} W^T B \\ T_2^T B \end{bmatrix} u(t) \right) dt \\ &+ \sum_{k=1}^{q} \begin{bmatrix} W^T N^k V & W^T N^k T_1 \\ T_2^T N^k V & T_2^T N^k T_1 \end{bmatrix} \begin{pmatrix} \tilde{x}(t-) \\ x_1(t-) \end{pmatrix} dM_k(t) \end{aligned} \tag{6.24}$$

and

$$y(t) = \begin{bmatrix} CV & CT_1 \end{bmatrix} \begin{pmatrix} \tilde{x}(t) \\ x_1(t) \end{pmatrix}.$$

By truncating the system and neglecting the x_1 terms, the approximating reduced order model is given by

$$\begin{aligned} d\tilde{x}(t) &= [W^T AV\tilde{x}(t) + W^T Bu(t)]dt + \sum_{k=1}^{q} W^T N^k V\tilde{x}(t-)dM_k(t), \\ \hat{y}(t) &= CV\tilde{x}(t). \end{aligned} \tag{6.25}$$

6.2.2. Stability analysis of the ROM

In this subsection, we prove that balanced truncation (BT) for systems with Lévy noise preserves mean square asymptotic stability. This is already proven in [11], where the proof is based on the theory of generalized Lyapunov operators. In contrast to the already existing result we use a system theoretical ansatz to prove the lemmas that are needed to show the central result of this subsection.

We will now prove that the ROM by type 1 BT is also mean square asymptotically stable. For simplicity of the notation, we assume to already have a balanced realization (A,B,C,N^k) of system (6.21). Below, we use the following partitions

$$A := \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\ N^k := \begin{bmatrix} N^k_{11} & N^k_{12} \\ N^k_{21} & N^k_{22} \end{bmatrix},\ B := \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \text{ and } C := \begin{bmatrix} C_1 & C_2 \end{bmatrix} \tag{6.26}$$

such that $(W^TAV, W^TB, CV, W^TN^kV) = (A_{11}, B_1, C_1, N^k_{11})$ in system (6.25). Since the system is balanced, for the reachability equation, we know

$$\begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}\begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} + \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}\begin{bmatrix} A^T_{11} & A^T_{12} \\ A^T_{21} & A^T_{22} \end{bmatrix} + \sum_{k=1}^{q} \begin{bmatrix} N^k_{11} & N^k_{21} \\ N^k_{12} & N^k_{22} \end{bmatrix}\begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}\begin{bmatrix} (N^k_{11})^T & (N^k_{12})^T \\ (N^k_{21})^T & (N^k_{22})^T \end{bmatrix} \cdot c_k$$
$$= - \begin{bmatrix} B_1B_1^T & B_1B_2^T \\ B_2B_1^T & B_2B_2^T \end{bmatrix} \tag{6.27}$$

and for the observability equation, we have

$$\begin{bmatrix} A^T_{11} & A^T_{21} \\ A^T_{12} & A^T_{22} \end{bmatrix}\begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} + \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \sum_{k=1}^{q} \begin{bmatrix} (N^k_{11})^T & (N^k_{21})^T \\ (N^k_{12})^T & (N^k_{22})^T \end{bmatrix}\begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}\begin{bmatrix} N^k_{11} & N^k_{12} \\ N^k_{21} & N^k_{22} \end{bmatrix} \cdot c_k$$
$$= - \begin{bmatrix} C_1^TC_1 & C_1^TC_2 \\ C_2^TC_1 & C_2^TC_2 \end{bmatrix}, \tag{6.28}$$

where $\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, $\Sigma_2 = \operatorname{diag}(\sigma_{r+1}, \ldots, \sigma_n)$ and $c_k = \mathbb{E}[M_k(1)^2]$. We will now show that the homogeneous solution $\tilde{x}_{h,x_0}$ of the reduced system (6.25) fulfilling

$$d\tilde{x}_h(t) = A_{11}\tilde{x}_h(t)dt + \sum_{k=1}^{q} N^k_{11}\tilde{x}_h(t-)dM_k(t), \quad \tilde{x}_h(0) = x_0, \tag{6.29}$$

is mean square stable which means that it is bounded in mean square.

Lemma 6.12. *Let $\hat{A}, \hat{N}^1, \dots, \hat{N}^q \in \mathbb{R}^{d\times d}$. If there exists a positive definite matrix $X > 0$ such that*

$$\hat{A}X + X\hat{A}^T + \sum_{k=1}^{q} \hat{N}^k X (\hat{N}^k)^T \cdot c_k = -Y,$$

where $Y \geq 0$, then the solution x_{h,x_0} of

$$dx_h(t) = \hat{A}x_h(t)dt + \sum_{k=1}^{q} \hat{N}^k x_h(t-)dM_k(t), \quad t \geq 0, \quad x_h(0) = x_0$$

is mean square stable, i.e. there exist $k_1, k_2 > 0$ such that

$$\mathbb{E}\left\|x_{h,x_0}(t)\right\|_2^2 \leq \frac{k_2}{k_1}\left\|x_0\right\|_2^2, \quad t \geq 0. \tag{6.30}$$

Furthermore, (6.30) is equivalent to

$$\sigma\left(I_d \otimes \hat{A} + \hat{A} \otimes I_d + \sum_{k=1}^{q} \hat{N}^k \otimes \hat{N}^k \cdot \mathbb{E}\left[M_k(1)^2\right]\right) \subset \overline{\mathbb{C}_-} \tag{6.31}$$

Proof. From equation (6.17), we conclude that

$$\begin{aligned}\mathbb{E}\left[x_{h,x_0}(t)^T X x_{h,x_0}(t)\right] = x_0^T X x_0 + \mathbb{E}\left[\int_0^t x_{h,x_0}^T(s) X \hat{A} x_{h,x_0}(s)ds\right] + \mathbb{E}\left[\int_0^t x_{h,x_0}^T(s)\hat{A}^T X x_{h,x_0}(s)ds\right]\\ + \mathbb{E}\left[\int_0^t x_{h,x_0}(s)^T \sum_{k=1}^{q} (\hat{N}^k)^T X \hat{N}^k c_k x_{h,x_0}(s)ds\right].\end{aligned}$$

Thus,

$$\mathbb{E}\left[x_{h,x_0}(t)^T X x_{h,x_0}(t)\right] = x_0^T X x_0 - \mathbb{E}\left[\int_0^t x_{h,x_0}^T(s) Y x_{h,x_0}(s)ds\right] \leq x_0^T X x_0.$$

Using $k_1 v^T v \leq v^T X v \leq k_2 v^T v$, where k_1 is the smallest and k_2 the largest eigenvalue of X, we obtain

$$k_1 \mathbb{E}\left[x_{h,x_0}(t)^T x_{h,x_0}(t)\right] \leq k_2 x_0^T x_0.$$

Now, $\mathbb{E}\left\|x_{h,x_0}(t)\right\|_2^2$ being bounded is equivalent to the solution $\mathbb{E}\left[x_{h,x_0}(t)x_{h,x_0}^T(t)\right]$ of

$$\dot{\mathbb{Y}}(t) = \hat{A}\mathbb{Y}(t) + \mathbb{Y}(t)\hat{A}^T + \sum_{k=1}^{q} \hat{N}^k\mathbb{Y}(t)(\hat{N}^k)^T\, \mathbb{E}\left[M_k(1)^2\right]$$

being bounded, compare Proposition 6.2. This again is equivalent to (6.31) and hence gives the required result. $\square$

We select the left upper block of equation (6.27) and obtain

$$A_{11}\Sigma_1 + \Sigma_1 A_{11}^T + \sum_{k=1}^{q} N_{11}^k\Sigma_1(N_{11}^k)^T \cdot c_k = -\left(\sum_{k=1}^{q} N_{12}^k\Sigma_2(N_{12}^k)^T \cdot c_k + B_1B_1^T\right).$$

By Lemma 6.12 we conclude that

$$\sigma\left(K\right) \subset \overline{\mathbb{C}_-}, \tag{6.32}$$

where $K = I_r \otimes A_{11} + A_{11} \otimes I_r + \sum_{k=1}^{q} N_{11}^k \otimes N_{11}^k \cdot \mathbb{E}\left[M_k(1)^2\right]$. It remains to show that K has no eigenvalues on the imaginary axis. Before doing this, we need another result which is proven in the next lemma.

Lemma 6.13. *Let $A_i, N_i^k \in \mathbb{R}^{d_i \times d_i}$ ($i = 1,2$), where $k = 1,\dots,q$, with*

$$\sigma\left(I_{d_1} \otimes A_1 + A_1 \otimes I_{d_1} + \sum_{k=1}^{q} N_1^k \otimes N_1^k\right) \subset \overline{\mathbb{C}_-} \text{ and } \sigma\left(I_{d_2} \otimes A_2 + A_2 \otimes I_{d_2} + \sum_{k=1}^{q} N_2^k \otimes N_2^k\right) \subset \mathbb{C}_-,$$

then it holds

$$\sigma\left(I_{d_1} \otimes A_2 + A_1 \otimes I_{d_2} + \sum_{k=1}^{q} N_1^k \otimes N_2^k\right) \subset \mathbb{C}_-. \tag{6.33}$$

Proof. Let Φ_i ($i = 1,2$) be the fundamental solution of

$$dx_h(t) = A_i x_h(t)dt + \sum_{k=1}^{q} N_i^k x_h(t)dw_k(t), \quad t \geq 0, \tag{6.34}$$

where $w_1,\dots,w_q$ are independent Wiener processes. Moreover, let $K_i \in \mathbb{R}^{d_i \times m}$, then by Proposition 6.17 the function $\mathbb{E}\left[\Phi_1(t)K_1K_2^T\Phi_2^T(t)\right]$, $t \geq 0$, is the solution of the following differential equation:

$$\dot{\mathbb{Y}}(t) = \mathbb{Y}(t)A_2^T + A_1\mathbb{Y}(t) + \sum_{k=1}^{q} N_1^k \mathbb{Y}(t)(N_2^k)^T, \quad \mathbb{Y}(0) = K_1 K_2^T. \tag{6.35}$$

This differential equation is asymptotically stable since

$$\left\|\mathbb{E}\left[\Phi_1(t)K_1K_2^T\Phi_2^T(t)\right]\right\|_F \leq \mathbb{E}\left\|\Phi_1(t)K_1K_2^T\Phi_2^T(t)\right\|_F \leq \sqrt{\mathbb{E}\left\|\Phi_1(t)K_1\right\|_F^2}\sqrt{\mathbb{E}\left\|\Phi_2(t)K_2\right\|_F^2}$$

using the inequality of Cauchy-Schwarz. The term $\mathbb{E}\left\|\Phi_1(t)K_1\right\|_F^2$ is bounded by Lemma 6.12 and the $\mathbb{E}\left\|\Phi_2(t)K_2\right\|_F^2$ tends to zero for $t \to \infty$ by the assumption of mean square asymptotic stability. Hence,

$$\left\|\mathbb{E}\left[\Phi_1(t)K_1K_2^T\Phi_2^T(t)\right]\right\|_F \to 0$$

for $t \to \infty$. The asymptotic stability of the differential equation (6.35) is equivalent to (6.33). □

We are now ready to state the central result of this subsection. There, the key ideas of the proof of Theorem 4.1 in [11] are transfered.

Theorem 6.14. *Let (A, B, C, N^k) be an asymptotically mean square stable and balanced realization of (6.21) with partitions defined in (6.26). For the diagonal Gramian*

$$\Sigma = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}$$

we assume that $\sigma(\Sigma_1) \cap \sigma(\Sigma_2) = \emptyset$ holds, then the reduced order model (6.25) is mean square asymptotically stable, i.e.

$$\sigma(K) \subset \mathbb{C}_-, \tag{6.36}$$

where $K = I_r \otimes A_{11} + A_{11} \otimes I_r + \sum_{k=1}^{q} N_{11}^k \otimes N_{11}^k \cdot \mathbb{E}\left[M_k(1)^2\right]$.

Proof. To simplify the presentation we will restrict our attention to the case, where $q = 1$. We further set $N = N^1\sqrt{\mathbb{E}\left[M_1(1)^2\right]}$. From the proof it will be easy to see that this is no loss of generality. Using Theorem 3.1 in [11], we obtain

$$\alpha(K) := \max\left\{\Re(\lambda) : \lambda \in \sigma(K)\right\} \in \sigma(K). \tag{6.37}$$

Assume that (6.36) does not hold, then by (6.37) and (6.32), we have $0 \in \sigma(K)$. Hence, there exists a nonzero matrix $V_1 \geq 0$ such that

$$A_{11}^T V_1 + V_1 A_{11} + N_{11}^T V_1 N_{11} = 0 \,. \tag{6.38}$$

The left upper block of (6.27) is

$$A_{11}\Sigma_1 + \Sigma_1 A_{11}^T + N_{11}\Sigma_1 N_{11}^T = -B_1 B_1^T - N_{12}\Sigma_2 N_{12}^T \,. \tag{6.39}$$

We have

$$\begin{aligned} 0 &= \langle A_{11}^T V_1 + V_1 A_{11} + N_{11}^T V_1 N_{11}, \Sigma_1 \rangle_F = \langle (V_1, A_{11}\Sigma_1 + \Sigma_1 A_{11}^T + N_{11}\Sigma_1 N_{11}^T \rangle_F \\ &= -\langle V_1, B_1 B_1^T + N_{12}\Sigma_2 N_{12}^T \rangle_F \end{aligned}$$

and hence $B_1^T V_1 = 0$ since

$$0 = \langle V_1, B_1 B_1^T \rangle_F = \operatorname{tr}(V_1 B_1 B_1^T) = \operatorname{tr}(V_1^{\frac{1}{2}} B_1 B_1^T V_1^{\frac{1}{2}}) = \left\| V_1^{\frac{1}{2}} B_1 \right\|_F^2$$

Analogously, we obtain $N_{12}^T V_1 = 0$.

Below, we discuss the invariance of $\ker V_1$ and $\operatorname{im} V_1$. Without loss of generality, we can assume that V_1 has maximal rank, i.e.

$$\left(\tilde{V}_1 \geq 0 \text{ and } A_{11}^T \tilde{V}_1 + \tilde{V}_1 A_{11} + N_{11}^T \tilde{V}_1 N_{11} = 0\right) \Rightarrow \operatorname{rank} \tilde{V}_1 \leq \operatorname{rank} V_1 \,. \tag{6.40}$$

We now observe that $\ker V_1$ is invariant under A_{11} and N_{11} and $\operatorname{im} V_1$ is invariant under A_{11}^T and N_{11}^T. To see this, let $V_1 z = 0$, then

$$0 = z^T \left(A_{11}^T V_1 + V_1 A_{11} + N_{11}^T V_1 N_{11}\right) z = z^T N_{11}^T V_1 N_{11} z \,,$$

whence also $V_1 N_{11} z = 0$, i.e. $N_{11} z \in \ker V_1$. From this, we have

$$0 = \left(A_{11}^T V_1 + V_1 A_{11} + N_{11}^T V_1 N_{11}\right) z = V_1 A_{11} z \,,$$

implying $A_{11} z \in \ker V_1$. Thus, $A_{11} \ker V_1 \subset \ker V_1$ and $N_{11} \ker V_1 \subset \ker V_1$.
Since $\ker V_1 = (\operatorname{im} V_1)^\perp$, it follows further that $\operatorname{im} V_1$ is invariant under A_{11}^T and N_{11}^T.

Let $\operatorname{im} V_1 = \operatorname{im} V_{11}$, with $V_{11}^T V_{11} = I$, $V_1 = V_{11} D_{11} V_{11}^T$, for some $D_{11} > 0$ and $\ker V_1 = \operatorname{im} V_{12}$ with $V_{12}^T V_{12} = I$, so that in particular $V_{11}^T V_{12} = 0$. By the invariance properties, we know that

$$A_{11}^T V_{11} = V_{11} \tilde{A}_{11}^T \quad \text{and} \quad A_{11} V_{12} = V_{12} \tilde{A}_{22} \tag{6.41}$$

for suitable matrices $\tilde{A}_{11}$ and $\tilde{A}_{22}$. Analogously,

$$N_{11}^T V_{11} = V_{11} \tilde{N}_{11}^T \quad \text{and} \quad N_{11} V_{12} = V_{12} \tilde{N}_{22} \tag{6.42}$$

for suitable matrices $\tilde{N}_{11}$ and $\tilde{N}_{22}$. Note that

$$\begin{aligned} 0 &= A_{11}^T V_1 + V_1 A_{11} + N_{11}^T V_1 N_{11} \\ &= A_{11}^T V_{11} D_{11} V_{11}^T + V_{11} D_{11} V_{11}^T A_{11} + N_{11}^T V_{11} D_{11} V_{11}^T N_{11} \\ &= V_{11} \left(\tilde{A}_{11}^T D_{11} + D_{11} \tilde{A}_{11} + \tilde{N}_{11}^T D_{11} \tilde{N}_{11} \right) V_{11}^T , \end{aligned}$$

whence $\tilde{A}_{11}^T D_{11} + D_{11} \tilde{A}_{11} + \tilde{N}_{11}^T D_{11} \tilde{N}_{11} = 0$ implying that $\sigma(I \otimes \tilde{A}_{11} + \tilde{A}_{11} \otimes I + \tilde{N}_{11} \otimes \tilde{N}_{11}) \subset \overline{\mathbb{C}_-}$ by Lemma 6.12. Moreover, $N_{12}^T V_{11} = 0$ and $B_1^T V_{11} = 0$, because $N_{12}^T V_1 = 0$ and $B_1^T V_1 = 0$.

As a next step, we us a unitary similarity transformation of the Kronecker matrix $I_n \otimes A + A \otimes I_n + N \otimes N$ corresponding to the original model characterizing the mean square asymptotic stability of the system. Considering the unitary transformation matrix $U = \left[\begin{array}{cc|c} V_{11} & V_{12} & 0 \\ \hline 0 & 0 & I \end{array} \right]$ yields

$$\begin{aligned} U^T A U &= \begin{bmatrix} V_{11}^T A_{11} V_{11} & V_{11}^T A_{11} V_{12} & V_{11}^T A_{12} \\ V_{12}^T A_{11} V_{11} & V_{12}^T A_{11} V_{12} & V_{12}^T A_{12} \\ A_{21} V_{11} & A_{21} V_{12} & A_{22} \end{bmatrix} \\ &= \begin{bmatrix} \tilde{A}_{11} & 0 & V_{11}^T A_{12} \\ V_{12}^T A_{11} V_{11} & \tilde{A}_{22} & V_{12}^T A_{12} \\ A_{21} V_{11} & A_{21} V_{12} & A_{22} \end{bmatrix} =: \begin{bmatrix} \tilde{A}_{11} & 0 & \tilde{A}_{13} \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} \end{bmatrix} = \tilde{A} , \end{aligned}$$

$$U^T N U = \begin{bmatrix} \tilde{N}_{11} & 0 & V_{11}^T N_{12} \\ V_{12}^T N_{11} V_{11} & \tilde{N}_{22} & V_{12}^T N_{12} \\ N_{21} V_{11} & N_{21} V_{12} & N_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{N}_{11} & 0 & 0 \\ V_{12}^T N_{11} V_{11} & \tilde{N}_{22} & V_{12}^T N_{12} \\ N_{21} V_{11} & N_{21} V_{12} & N_{22} \end{bmatrix} =: \begin{bmatrix} \tilde{N}_{11} & 0 & 0 \\ \tilde{N}_{21} & \tilde{N}_{22} & \tilde{N}_{23} \\ \tilde{N}_{31} & \tilde{N}_{32} & \tilde{N}_{33} \end{bmatrix} = \tilde{N} \,,$$

$$U^T B = \begin{bmatrix} V_{11}^T & 0 \\ V_{12}^T & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} V_{11}^T B_1 \\ V_{12}^T B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} 0 \\ V_{12}^T B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \tilde{B}_2 \\ \tilde{B}_3 \end{bmatrix} = \tilde{B} \,,$$

$$CU = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 & \tilde{C}_3 \end{bmatrix} = \tilde{C} \,, \quad U^T \Sigma U = \begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{21}^T & 0 \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} & 0 \\ 0 & 0 & \tilde{\Sigma}_{33} \end{bmatrix} = \tilde{\Sigma}$$

with $\tilde{\Sigma}_{33} = \Sigma_2$ and $\sigma\left(\begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{21}^T \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{bmatrix}\right) = \sigma(\Sigma_1)$.

Let us write $\tilde{A}_1 = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}$, $\tilde{N}_1 = \begin{bmatrix} \tilde{N}_{11} & 0 \\ \tilde{N}_{21} & \tilde{N}_{22} \end{bmatrix}$, $\tilde{\Sigma}_1 = \begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{21}^T \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{bmatrix}$ and define

$$T(X) = \tilde{A}_1 X + X \tilde{A}_1^T + \tilde{N}_1 X \tilde{N}_1^T \,.$$

As seen above, $T^*(D_1) = 0$ for $D_1 = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix}$, where T^* denotes the adjoint operator of T with respect to the Frobenius inner product.

By construction $T(\tilde{\Sigma}_1) \leq 0$ and $T^*(\tilde{\Sigma}_1) \leq 0$, whence also $(T + T^*)(\tilde{\Sigma}_1) \leq 0$ implying

$$\sigma(T + T^*) \subset \overline{\mathbb{C}_-} \cap \mathbb{R} \tag{6.43}$$

using Lemma 6.12. Considering the left upper blocks of

$$\tilde{A}\tilde{\Sigma} + \tilde{\Sigma}\tilde{A}^T + \tilde{N}\tilde{\Sigma}\tilde{N}^T = -\tilde{B}\tilde{B}^T \quad \text{and} \tag{6.44}$$

$$\tilde{A}^T\tilde{\Sigma} + \tilde{\Sigma}\tilde{A} + \tilde{N}^T\tilde{\Sigma}\tilde{N} = -\tilde{C}^T\tilde{C} \,, \tag{6.45}$$

we obtain

$$\tilde{\Sigma}_{11}\tilde{A}_{11}^T + \tilde{A}_{11}\tilde{\Sigma}_{11} + \tilde{N}_{11}\tilde{\Sigma}_{11}\tilde{N}_{11}^T = 0 \quad \text{and} \tag{6.46}$$

$$\begin{aligned}\tilde{A}_{11}^T\tilde{\Sigma}_{11} + \tilde{\Sigma}_{11}\tilde{A}_{11} + \tilde{N}_{11}^T\tilde{\Sigma}_{11}\tilde{N}_{11} = &-\tilde{C}_1^T\tilde{C}_1 - \tilde{N}_{21}^T\tilde{\Sigma}_{22}\tilde{N}_{21} - \tilde{N}_{31}^T\tilde{\Sigma}_{33}\tilde{N}_{31}\\ &- \tilde{\Sigma}_{21}^T\tilde{A}_{21} - \tilde{A}_{21}^T\tilde{\Sigma}_{21} - \tilde{N}_{21}^T\tilde{\Sigma}_{21}\tilde{N}_{11} - \tilde{N}_{11}^T\tilde{\Sigma}_{21}^T\tilde{N}_{21}\,.\end{aligned} \tag{6.47}$$

Taking the inner product of (6.47) with $\tilde{\Sigma}_{11}$, we get

$$\begin{aligned}&\langle \tilde{C}_1^T\tilde{C}_1 + \tilde{N}_{21}^T\tilde{\Sigma}_{22}\tilde{N}_{21} + \tilde{N}_{31}^T\tilde{\Sigma}_{33}\tilde{N}_{31}, \tilde{\Sigma}_{11}\rangle_F\\ &\quad = -\langle \tilde{\Sigma}_{21}^T\tilde{A}_{21} + \tilde{A}_{21}^T\tilde{\Sigma}_{21} + \tilde{N}_{21}^T\tilde{\Sigma}_{21}\tilde{N}_{11} + \tilde{N}_{11}^T\tilde{\Sigma}_{21}^T\tilde{N}_{21}, \tilde{\Sigma}_{11}\rangle_F \qquad (6.48)\\ &\quad = -2\langle \tilde{\Sigma}_{21}^T\tilde{A}_{21} + \tilde{N}_{11}^T\tilde{\Sigma}_{21}^T\tilde{N}_{21}, \tilde{\Sigma}_{11}\rangle_F\\ &\quad = -2\operatorname{tr}\left((\tilde{\Sigma}_{21}^T\tilde{A}_{21} + \tilde{N}_{11}^T\tilde{\Sigma}_{21}^T\tilde{N}_{21})\tilde{\Sigma}_{11}\right)\\ &\quad = -2\operatorname{tr}\left(\tilde{\Sigma}_{21}^T(\tilde{A}_{21}\tilde{\Sigma}_{11} + \tilde{N}_{21}\tilde{\Sigma}_{11}\tilde{N}_{11}^T)\right)\\ &\quad = 2\langle -\tilde{A}_{21}\tilde{\Sigma}_{11} - \tilde{N}_{21}\tilde{\Sigma}_{11}\tilde{N}_{11}^T, \tilde{\Sigma}_{21}\rangle_F\,. \qquad (6.49)\end{aligned}$$

The second block in the first column of $\tilde{A}\tilde{\Sigma} + \tilde{\Sigma}\tilde{A}^T + \tilde{N}\Sigma\tilde{N}^T = -\tilde{B}\tilde{B}^T$ is

$$0 = \tilde{\Sigma}_{21}\tilde{A}_{11}^T + \tilde{N}_{21}\tilde{\Sigma}_{11}\tilde{N}_{11}^T + \tilde{N}_{22}\tilde{\Sigma}_{21}\tilde{N}_{11}^T + \tilde{A}_{21}\tilde{\Sigma}_{11} + \tilde{A}_{22}\tilde{\Sigma}_{21}$$

and hence

$$-\tilde{A}_{21}\tilde{\Sigma}_{11} - \tilde{N}_{21}\tilde{\Sigma}_{11}\tilde{N}_{11}^T = \tilde{\Sigma}_{21}\tilde{A}_{11}^T + \tilde{N}_{22}\tilde{\Sigma}_{21}\tilde{N}_{11}^T + \tilde{A}_{22}\tilde{\Sigma}_{21} =: T_{21}(\tilde{\Sigma}_{21})\,. \tag{6.50}$$

We choose an arbitrary matrix Y of suitable dimension and introduce X as the following block matrix $X = \begin{bmatrix} 0 & 0 \\ Y & 0 \end{bmatrix}$ with $\langle X, X\rangle_F = \langle Y, Y\rangle_F$. Then, $(T + T^*)(X) = \begin{bmatrix} \star & \star \\ (T_{21} + T_{21}^*)(Y) & \star \end{bmatrix}$, so that by $T + T^*$ being self-adjoint and negative semidefinite, we have

$$0 \geq \langle X, (T + T^*)(X)\rangle_F = \langle Y, (T_{21} + T_{21}^*)(Y)\rangle_F.$$

Thus, $T_{21} + T_{21}^*$ is self-adjoint and negative semidefinite as well. Using this fact with $Y = \tilde{\Sigma}_{21}$ and inserting (6.50) into (6.49), we obtain

$$0 \leq 2\langle T_{21}(\tilde{\Sigma}_{21}), \tilde{\Sigma}_{21}\rangle_F = \langle (T_{21} + T_{21}^*)(\tilde{\Sigma}_{21}), \tilde{\Sigma}_{21}\rangle_F \leq 0\,. \tag{6.51}$$

From (6.51) and (6.50) it follows that the right hand side of (6.49) vanishes and consequently

$$\tilde{C}_1 = 0\,, \quad \tilde{N}_{21} = 0\,, \quad \tilde{N}_{31} = 0\,. \tag{6.52}$$

Moreover, from (6.51), we obtain

$$(T_{21} + T_{21}^*)(\tilde{\Sigma}_{21}) = 0, \tag{6.53}$$

because the quadratic form defined by $-(T_{21} + T_{21}^*)$ is positive semidefinite.

Exploiting (6.52), we find that the second blocks of the first column of (6.44) and (6.45), respectively take the forms

$$\begin{aligned} 0 &= \tilde{\Sigma}_{21}\tilde{A}_{11}^T + \tilde{N}_{22}\tilde{\Sigma}_{21}\tilde{N}_{11}^T + \tilde{A}_{21}\tilde{\Sigma}_{11} + \tilde{A}_{22}\tilde{\Sigma}_{21} = T_{21}(\tilde{\Sigma}_{21}) + \tilde{A}_{21}\tilde{\Sigma}_{11}, \\ 0 &= \tilde{A}_{22}^T\tilde{\Sigma}_{21} + \tilde{\Sigma}_{21}\tilde{A}_{11} + \tilde{\Sigma}_{22}\tilde{A}_{21} + \tilde{N}_{22}^T\tilde{\Sigma}_{21}\tilde{N}_{11} = T_{21}^*(\tilde{\Sigma}_{21}) + \tilde{\Sigma}_{22}\tilde{A}_{21}\,. \end{aligned}$$

Adding these and using (6.53), we get the homogeneous Sylvester equation

$$0 = \tilde{A}_{21}\tilde{\Sigma}_{11} + \tilde{\Sigma}_{22}\tilde{A}_{21}\,.$$

It follows that $\tilde{A}_{21} = 0$, since all eigenvalues of $\tilde{\Sigma}_{11}$ and $\tilde{\Sigma}_{22}$ are strictly positive. Inserting $\tilde{A}_{21} = \tilde{N}_{21} = 0$ in (6.50) we see that $T_{21}(\tilde{\Sigma}_{21}) = 0$.
Moreover, the mapping $X \mapsto \tilde{A}_{22}^T X + X\tilde{A}_{22} + \tilde{N}_{22}^T X \tilde{N}_{22}$ has all eigenvalues in $\mathbb{C}_-$ or equivalently $\sigma\left(I \otimes \tilde{A}_{22} + \tilde{A}_{22} \otimes I + \tilde{N}_{22} \otimes \tilde{N}_{22}\right) \subset \mathbb{C}_-$. Otherwise, there would exist a non-zero matrix $D_{22} \geq 0$ with

$$\tilde{A}_{22}^T D_{22} + D_{22}\tilde{A}_{22} + \tilde{N}_{22}^T D_{22}\tilde{N}_{22} = 0\,.$$

But then, with $D = \operatorname{diag}(D_{11}, D_{22})$, we would have

$$\begin{aligned} T^*(D) = &\begin{bmatrix} \tilde{A}_{11} & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix}^T \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} + \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} \tilde{A}_{11} & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix} \\ &+ \begin{bmatrix} \tilde{N}_{11} & 0 \\ 0 & \tilde{N}_{22} \end{bmatrix}^T \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} \tilde{N}_{11} & 0 \\ 0 & \tilde{N}_{22} \end{bmatrix} = 0\,. \end{aligned}$$

Thus, $\tilde{V}_1 = [V_{11}, V_{12}] D [V_{11}, V_{12}]^T$ would contradict (6.40).

Hence, $\sigma\left(I \otimes \tilde{A}_{22} + \tilde{A}_{11} \otimes I + \tilde{N}_{11} \otimes \tilde{N}_{22}\right) \subset \mathbb{C}_-$ by Lemma 6.13 which implies that T_{21} is invertible. Consequently, we have $\tilde{\Sigma}_{21} = 0$ from $T_{21}(\tilde{\Sigma}_{21}) = 0$.
Now the third blocks of the first column of (6.44) and (6.45), respectively simplify to

$$0 = \tilde{\Sigma}_{33}\tilde{A}_{13}^T + \tilde{A}_{31}\tilde{\Sigma}_{11} \quad \text{and} \quad 0 = \tilde{A}_{13}^T\tilde{\Sigma}_{11} + \tilde{\Sigma}_{33}\tilde{A}_{31} \, .$$

By suitable multiplication with $\tilde{\Sigma}_{11}$ and $\tilde{\Sigma}_{33}$ we can eliminate either $\tilde{A}_{31}$ or $\tilde{A}_{13}$ to obtain the equations

$$0 = \tilde{\Sigma}_{33}^2\tilde{A}_{13}^T - \tilde{A}_{13}^T\tilde{\Sigma}_{11}^2 \quad \text{and} \quad 0 = \tilde{\Sigma}_{33}^2\tilde{A}_{31} - \tilde{A}_{31}\tilde{\Sigma}_{11}^2 \, .$$

By construction $\sigma(\tilde{\Sigma}_{11}) \subset \sigma(\Sigma_1)$ and $\sigma(\Sigma_1) \cap \sigma(\tilde{\Sigma}_{33}) = \emptyset$. Hence, both Sylvester equations are uniquely solvable and $\tilde{A}_{13}^T = \tilde{A}_{31} = 0$. Finally, with $\tilde{\Sigma}_0 = \operatorname{diag}(\tilde{\Sigma}_{11}, 0, 0)$ we get

$$\tilde{A}\tilde{\Sigma}_0 + \tilde{\Sigma}_0\tilde{A}^T + \tilde{N}\tilde{\Sigma}_0\tilde{N}^T = 0 \, ,$$

contradicting asymptotic stability of the full system. □

Theorem 6.14 has shown that asymptotic mean square stability is preserved in the reduced order model from balanced truncation. This property is vital in the error bound analysis in Subsection 6.2.4 because it ensures the existence of the Gramians corresponding to the reduced order model.

6.2.3. Further properties of the ROM

Below, we transfer results from Section 4.1 in [14] into this subsection.
One persisting problem is to find an explicit structure of the Gramians of the reduced order model. As we will see in an example below, in contrast to the deterministic case the reduced order model is not balanced, that means the Gramians are neither diagonal nor equal. In addition, the Hankel singular values are different from those of the original system.

Example 6.15. *We consider the case, where $q = 1$ and the noise process is a Wiener process w. So, the system we focus on is*

$$\begin{aligned} dx(t) &= [Ax(t) + Bu(t)]dt + Nx(t)dw(t), \\ y(t) &= Cx(t). \end{aligned} \tag{6.54}$$

The following matrices (up to the digits shown) provide a balanced and asymptotically mean square stable system:

$$A = \begin{pmatrix} -4.4353 & 3.9992 & -0.3287 \\ 2.9337 & -11.0285 & -0.4319 \\ -0.0591 & -0.1303 & -11.5362 \end{pmatrix}, \qquad B = \begin{pmatrix} -3.4648 & -1.9391 & -3.6790 \\ 5.7925 & 4.1379 & 2.3036 \\ -0.3258 & 1.1359 & 2.8972 \end{pmatrix},$$

$$N = \begin{pmatrix} -1.4886 & 2.8510 & -0.2429 \\ 0.4720 & 0.5803 & 3.1152 \\ -1.6123 & -0.8082 & -0.0917 \end{pmatrix}, \qquad C = \begin{pmatrix} -3.0588 & 0.4275 & 0.2630 \\ -4.8686 & 1.2886 & 1.0769 \\ -4.3349 & 0.6747 & -0.1734 \end{pmatrix}.$$

The Gramians are given by

$$P = Q = \Sigma = \begin{pmatrix} 8.4788 & 0 & 0 \\ 0 & 3.3232 & 0 \\ 0 & 0 & 1.4726 \end{pmatrix}.$$

The reduced order model $(r = 2)$ is asymptotically mean square stable and has the following Gramians:

$$P_R = \begin{pmatrix} 7.7470 & -0.3562 \\ -0.3562 & 2.5496 \end{pmatrix} \text{ and } Q_R = \begin{pmatrix} 7.7495 & -0.2074 \\ -0.2074 & 2.8980 \end{pmatrix}.$$

The Hankel singular values of the reduced order model are 7.6633 *and* 2.7001.

At the end of this subsection, we provide a short example that shows that the reduced order model need not be completely observable and reachable even if the original system is completely observable and reachable:

Example 6.16. *We consider the equations (6.54) with the matrices*

$$(A,B,N,C) = \left(\begin{pmatrix} -0.25 & 1 \\ 1 & -9 \end{pmatrix}, \begin{pmatrix} 0 \\ \sqrt{7} \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & -3 \end{pmatrix}, \begin{pmatrix} 0 & \sqrt{7} \end{pmatrix} \right)$$

and obtain a balanced and asymptotically mean square stable system being completely reachable and observable. The Hankel singular values are 2 *and* 1. *Truncating yields a system with coefficients* $(A_{11}, B_1, N_{11}, C_1) = (-0.25, 0, 0, 0)$ *having Gramians* $P_R = Q_R = 0$.

6.2.4. Error bound for type 1 balanced truncation

This subsection consists of the result that are obtained in Section 4.2 of [14].
So, let $\left(A, N^k, B, C\right)$ $(k = 1, \ldots, q)$ be a realization of system (6.21). Furthermore, we assume the

initial condition of the system to be zero. We introduce the following partitions:

$$\hat{T}A\hat{T}^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \hat{T}N^k\hat{T}^{-1} = \begin{bmatrix} N_{11}^k & N_{12}^k \\ N_{21}^k & N_{22}^k \end{bmatrix}, \hat{T}B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \text{ and } C\hat{T}^{-1} = \begin{bmatrix} C_1 & C_2 \end{bmatrix}, \quad (6.55)$$

where $\hat{T}$ is the balancing transformation defined in (6.23) and $\left(A_{11}, N_{11}^k, B_1, C_1\right)$ are the coefficients of the reduced order model. The output of the reduced (truncated) system is given by

$$\hat{y}(t) = C_1\tilde{x}(t) = C_1\int_0^t \tilde{\Phi}(t,s)B_1u(s)ds,$$

where $\tilde{\Phi}$ is the fundamental matrix of the truncated system. In addition, we use a result from [11]. Therein it is proven that the homogeneous equation ($u \equiv 0$) of the reduced system is still asymptotically mean square stable. This is vital for the error bound we provide below since the existence of the Gramians of the reduced order model is ensured. Moreover, we know

$$y(t) = Cx(t) = C\int_0^t \Phi(t,s)Bu(s)ds.$$

It is our goal to steer the average state via the control u and to truncate the average states that are difficult to reach for obtaining a reduced order model. Therefore, it is a meaningful criterion to consider the worst case mean error of $\hat{y}(t)$ and $y(t)$. Below, we give a bound for that kind of error:

$$\begin{aligned} \mathbb{E}\left\|\hat{y}(t) - y(t)\right\|_2 &= \mathbb{E}\left\|C\int_0^t \Phi(t,s)Bu(s)ds - C_1\int_0^t \tilde{\Phi}(t,s)B_1u(s)ds\right\|_2 \\ &\leq \mathbb{E}\int_0^t \left\|\left(C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right)u(s)\right\|_2 ds \\ &\leq \mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right\|_F \left\|u(s)\right\|_2 ds, \end{aligned}$$

and by the Cauchy-Schwarz inequality

$$\mathbb{E}\left\|\hat{y}(t) - y(t)\right\|_2 \leq \left(\mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right\|_F^2 ds\right)^{\frac{1}{2}} \left(\mathbb{E}\int_0^t \left\|u(s)\right\|_2^2 ds\right)^{\frac{1}{2}}.$$

holds. Now,

$$\mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right\|_F^2 ds$$

$$
\begin{aligned}
&= \mathbb{E}\int_0^t \left\|C\Phi(t,s)B\right\|_F^2 + \left\|C_1\tilde{\Phi}(t,s)B_1\right\|_F^2 - 2\left\langle C\Phi(t,s)B, C_1\tilde{\Phi}(t,s)B_1\right\rangle_F ds\\
&= \mathbb{E}\int_0^t \operatorname{tr}\left(C\Phi(t,s)BB^T\Phi^T(t,s)C^T\right) ds + \mathbb{E}\int_0^t \operatorname{tr}\left(C_1\tilde{\Phi}(t,s)B_1B_1^T\tilde{\Phi}^T(t,s)C_1^T\right) ds\\
&\quad - 2\mathbb{E}\int_0^t \operatorname{tr}\left(C\Phi(t,s)BB_1^T\tilde{\Phi}^T(t,s)C_1^T\right) ds\\
&= \operatorname{tr}\left(C\int_0^t \mathbb{E}\left[\Phi(t,s)BB^T\Phi^T(t,s)\right] ds\, C^T\right) + \operatorname{tr}\left(C_1\int_0^t \mathbb{E}\left[\tilde{\Phi}(t,s)B_1B_1^T\tilde{\Phi}^T(t,s)\right] ds\, C_1^T\right)\\
&\quad - 2\operatorname{tr}\left(C\int_0^t \mathbb{E}\left[\Phi(t,s)BB_1^T\tilde{\Phi}^T(t,s)\right] ds\, C_1^T\right). \qquad (6.56)
\end{aligned}
$$

Due to the remarks before Proposition 6.4, we have

$$
\begin{aligned}
\mathbb{E}\left[\Phi(t,s)BB^T\Phi^T(t,s)\right] &= \mathbb{E}\left[\Phi(t-s)BB^T\Phi^T(t-s)\right] \text{ and}\\
\mathbb{E}\left[\tilde{\Phi}(t,s)B_1B_1^T\tilde{\Phi}^T(t,s)\right] &= \mathbb{E}\left[\tilde{\Phi}(t-s)B_1B_1^T\tilde{\Phi}^T(t-s)\right]
\end{aligned}
$$

for $0 \leq s \leq t$. Furthermore, we need to analyze the term in (6.56) in order to find an error bound which is practically useful. For that reason, we need the following proposition:

Proposition 6.17. *The $\mathbb{R}^{n\times r}$-valued function $\mathbb{E}\left[\Phi(t)BB_1^T\tilde{\Phi}^T(t)\right]$, $t \geq 0$, is the solution to the following differential equation:*

$$
\dot{\mathbb{Y}}(t) = \mathbb{Y}(t)A_{11}^T + A\mathbb{Y}(t) + \sum_{k=1}^q N^k\mathbb{Y}(t)(N_{11}^k)^T\,\mathbb{E}\left[M_k(1)^2\right], \quad \mathbb{Y}(0) = BB_1^T. \qquad (6.57)
$$

Proof. With $B = [b_1, \ldots, b_m]$ and $B_1 = \left[\tilde{b}_1, \ldots, \tilde{b}_m\right]$, we obtain

$$
\Phi(t)BB_1^T\tilde{\Phi}^T(t) = \Phi(t)b_1\tilde{b}_1^T\tilde{\Phi}^T(t) + \ldots + \Phi(t)b_m\tilde{b}_m^T\tilde{\Phi}^T(t). \qquad (6.58)
$$

By applying the Ito product formula from Corollary A.5, we have

$$
\begin{aligned}
\Phi(t)b_l\tilde{b}_l^T\tilde{\Phi}^T(t) =& b_l\tilde{b}_l^T + \int_0^t d(\Phi(s)b_l)\tilde{b}_l^T\tilde{\Phi}^T(s-) + \int_0^t \Phi(s-)b_l d(\tilde{b}_l^T\tilde{\Phi}^T(s))\\
&+ \left(\left[e_i^T\Phi b_l, e_j^T\tilde{\Phi}\tilde{b}_l\right]_t\right)_{\substack{i=1,\ldots,n\\ j=1,\ldots,r}}.
\end{aligned}
$$

From (A.5), we know that

$$\mathbb{E}\left[e_i^T \Phi b_l, e_j^T \tilde{\Phi} \tilde{b}_l\right]_t = \sum_{k=1}^{q} \mathbb{E}\left[\int_0^t e_i^T N^k \Phi(s) b_l \tilde{b}_1^T \tilde{\Phi}^T(s) (N_{11}^k)^T e_j ds\right] \mathbb{E}\left[M_k(1)^2\right].$$

The Ito integrals have mean zero by Section 3.1. Hence,

$$\begin{aligned}\mathbb{E}\left[\Phi(t) b_l \tilde{b}_l^T \tilde{\Phi}^T(t)\right] = b_l \tilde{b}_l^T + \mathbb{E}\left[\int_0^t \Phi(s) b_l \tilde{b}_l^T \tilde{\Phi}^T(s) ds\right] A_{11}^T + A\mathbb{E}\left[\int_0^t \Phi(s) b_l \tilde{b}_l^T \tilde{\Phi}^T(s) ds\right] \\ + \sum_{k=1}^{q} N^k \mathbb{E}\left[\int_0^t \Phi(s) b_l \tilde{b}_l^T \tilde{\Phi}^T(s) ds\right] (N_{11}^k)^T \mathbb{E}\left[M_k(1)^2\right]\end{aligned}$$

using that the trajectories of Φ and $\tilde{\Phi}$ only have jumps on Lebesgue zero sets. Using equation (6.58), we obtain

$$\begin{aligned}\mathbb{E}\left[\Phi(t) B B_1^T \tilde{\Phi}^T(t)\right] = B B_1^T + \mathbb{E}\left[\int_0^t \Phi(s) B B_1^T \tilde{\Phi}^T(s) ds\right] A_{11}^T \\ + A\mathbb{E}\left[\int_0^t \Phi(s) B B_1^T \tilde{\Phi}^T(s) ds\right] \\ + \sum_{k=1}^{q} N^k \mathbb{E}\left[\int_0^t \Phi(s) B B_1^T \tilde{\Phi}^T(s) ds\right] (N_{11}^k)^T \mathbb{E}\left[M_k(1)^2\right]\end{aligned} \tag{6.59}$$

which proves the result. □

By Proposition 6.17, we can conclude that the function $\mathbb{E}\left[\Phi(t-\tau) B B_1^T \tilde{\Phi}^T(t-\tau)\right]$, $t \geq \tau \geq 0$, is the solution to the equation

$$\dot{\mathbb{Y}}(t) = \mathbb{Y}(t) A_{11}^T + A\mathbb{Y}(t) + \sum_{k=1}^{q} N^k \mathbb{Y}(t) (N_{11}^k)^T \, \mathbb{E}\left[M_k(1)^2\right], \quad \mathbb{Y}(\tau) = B B_1^T, \tag{6.60}$$

for all $t \geq \tau \geq 0$. Analogous to Proposition 6.17 we can conclude that $\mathbb{E}\left[\Phi(t,\tau) B B_1^T \tilde{\Phi}^T(t,\tau)\right]$ is also a solution to equation (6.60), which yields

$$\mathbb{E}\left[\Phi(t,\tau) B B_1^T \tilde{\Phi}^T(t,\tau)\right] = \mathbb{E}\left[\Phi(t-\tau) B B_1^T \tilde{\Phi}^T(t-\tau)\right] \tag{6.61}$$

for all $t \geq \tau \geq 0$. Using equality (6.61), we have

$$\mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right\|_F^2 ds = \operatorname{tr}\left(C\int_0^t \mathbb{E}\left[\Phi(t-s) B B^T \Phi^T(t-s)\right] ds\, C^T\right)$$

$$+\operatorname{tr}\left(C_1 \int_0^t \mathbb{E}\left[\tilde{\Phi}(t-s)B_1B_1^T\tilde{\Phi}^T(t-s)\right] ds\, C_1^T\right)$$
$$-2\operatorname{tr}\left(C \int_0^t \mathbb{E}\left[\Phi(t-s)BB_1^T\tilde{\Phi}^T(t-s)\right] ds\, C_1^T\right).$$

By substitution, we obtain

$$\begin{aligned}\mathbb{E}\int_0^t \left\|C\Phi(t,s)B - C_1\tilde{\Phi}(t,s)B_1\right\|_F^2 ds &= \operatorname{tr}\left(C \int_0^t \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds\, C^T\right)\\ &\quad+\operatorname{tr}\left(C_1 \int_0^t \mathbb{E}\left[\tilde{\Phi}(s)B_1B_1^T\tilde{\Phi}^T(s)\right] ds\, C_1^T\right)\\ &\quad-2\operatorname{tr}\left(C \int_0^t \mathbb{E}\left[\Phi(s)BB_1^T\tilde{\Phi}^T(s)\right] ds\, C_1^T\right)\\ &= \mathbb{E}\int_0^t \left\|C\Phi(s)B - C_1\tilde{\Phi}(s)B_1\right\|_F^2 ds.\end{aligned}$$

The homogeneous equation of the truncated system is still asymptotically mean square stable due to [11]. Hence, the existence of the matrices $P_R = \mathbb{E}\int_0^\infty \tilde{\Phi}(\tau)B_1B_1^T\tilde{\Phi}^T(\tau)d\tau \in \mathbb{R}^{r\times r}$ and $P_M = \mathbb{E}\int_0^\infty \Phi(\tau)BB_1^T\tilde{\Phi}^T(\tau)d\tau \in \mathbb{R}^{n\times r}$ is guaranteed. Therefore,

$$\begin{aligned}\mathbb{E}\left\|\hat{y}(t)-y(t)\right\|_2 &\leq \left(\mathbb{E}\int_0^\infty \left\|C\Phi(s)B - C_1\tilde{\Phi}(s)B_1\right\|_F^2 ds\right)^{\frac{1}{2}}\left(\mathbb{E}\int_0^t \left\|u(s)\right\|_2^2 ds\right)^{\frac{1}{2}}\\ &= \left(\operatorname{tr}\left(CPC^T\right) + \operatorname{tr}\left(C_1P_RC_1^T\right) - 2\operatorname{tr}\left(CP_MC_1^T\right)\right)^{\frac{1}{2}}\left\|u\right\|_{L_t^2}\end{aligned}$$

holds, where $P = \mathbb{E}\int_0^\infty \Phi(\tau)BB^T\Phi^T(\tau)d\tau$ is the reachability Gramian of the original system, P_R the reachability Gramian of the approximating system and P_M a matrix that satisfies the following equation:

$$0 = BB_1^T + P_MA_{11}^T + AP_M + \sum_{k=1}^q N^kP_M(N_{11}^k)^T\,\mathbb{E}\left[M_k(1)^2\right], \tag{6.62}$$

which we get by taking the limit $t\to\infty$ on both sides of equation (6.59). We summarize these results in the following theorem:

Theorem 6.18. *Let* (A, N^k, B, C) *be a realization of system (6.21) and* $(A_{11}, N_{11}^k, B_1, C_1)$ *the co-*

efficients of the reduced order model defined in (6.55), then

$$\sup_{t\in[0,T]} \mathbb{E}\left\|\hat{y}(t)-y(t)\right\|_2 \leq \left(\operatorname{tr}\left(CPC^T\right)+\operatorname{tr}\left(C_1P_RC_1^T\right)-2\ \operatorname{tr}\left(CP_MC_1^T\right)\right)^{\frac{1}{2}}\left\|u\right\|_{L^2_T} \tag{6.63}$$

for every $T>0$, where y and $\hat{y}$ are the outputs of the original and the reduced system, respectively. Here, P denotes the reachability Gramian of system (6.21), P_R denotes the reachability Gramian of the reduced system and P_M satisfies equation (6.62).

Remark. *If $u\in L^2$ we can replace $\left\|\cdot\right\|_{L^2_T}$ by $\left\|\cdot\right\|_{L^2}$ and $[0,T]$ by $\mathbb{R}_+$ in inequality (6.63).*

Now, we are ready to specify the error bound from (6.63) in the following proposition.

Proposition 6.19. *If the realization (A,N^k,B,C) is balanced, then*

$$\begin{aligned}&\operatorname{tr}\left(CPC^T+C_1P_RC_1^T-2CP_MC_1^T\right)\\&=\operatorname{tr}(\Sigma_2(B_2B_2^T+2P_{M,2}A_{21}^T))+\sum_{k=1}^{q}\operatorname{tr}(\Sigma_2(2N_{22}^kP_{M,2}(N_{21}^k)^T+2N_{21}^kP_{M,1}(N_{21}^k)^T-N_{21}^kP_R(N_{21}^k)^T))c_k,\end{aligned}$$

where $P_{M,1}$ are the first r rows of P_M and $P_{M,2}$ are the last $n-r$ rows of P_M, $c_k=\mathbb{E}\left[M_k(1)^2\right]$ and $\Sigma_2=\operatorname{diag}(\sigma_{r+1},\ldots,\sigma_n)$.

Proof. For simplicity of notation, we prove this result just for the case $q=1$ but of course it is straightforward to generalize the proof for an arbitrary q. Here, we additionally set $N:=N^1$ and $c:=c_1$. Then, we have

$$\begin{aligned}&\begin{bmatrix}A_{11}^T & A_{21}^T\\ A_{12}^T & A_{22}^T\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}+\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}+\begin{bmatrix}N_{11}^T & N_{21}^T\\ N_{12}^T & N_{22}^T\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}N_{11} & N_{12}\\ N_{21} & N_{22}\end{bmatrix}c\\&=-\begin{bmatrix}C_1^TC_1 & C_1^TC_2\\ C_2^TC_1 & C_2^TC_2\end{bmatrix}.\end{aligned}$$

Hence,

$$A_{11}^T\Sigma_1+\Sigma_1A_{11}+N_{11}^T\Sigma_1N_{11}c+N_{21}^T\Sigma_2N_{21}c=-C_1^TC_1, \tag{6.64}$$

$$A_{22}^T\Sigma_2+\Sigma_2A_{22}+N_{22}^T\Sigma_2N_{22}c+N_{12}^T\Sigma_1N_{12}c=-C_2^TC_2 \tag{6.65}$$

and

$$A_{21}^T\Sigma_2+\Sigma_1 A_{12}+N_{11}^T\Sigma_1 N_{12}c+N_{21}^T\Sigma_2 N_{22}c=-C_1^T C_2. \tag{6.66}$$

Furthermore,

$$\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}+\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}A_{11}^T & A_{21}^T\\ A_{12}^T & A_{22}^T\end{bmatrix}+\begin{bmatrix}N_{11} & N_{12}\\ N_{21} & N_{22}\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}N_{11}^T & N_{21}^T\\ N_{12}^T & N_{22}^T\end{bmatrix}c$$
$$=-\begin{bmatrix}B_1B_1^T & B_1B_2^T\\ B_2B_1^T & B_2B_2^T\end{bmatrix},$$

such that one can conclude

$$A_{11}\Sigma_1+\Sigma_1 A_{11}^T+N_{11}\Sigma_1 N_{11}^T c+N_{12}\Sigma_2 N_{12}^T c=-B_1B_1^T \tag{6.67}$$

and

$$A_{22}\Sigma_2+\Sigma_2 A_{22}^T+N_{22}\Sigma_2 N_{22}^T c+N_{21}\Sigma_1 N_{21}^T c=-B_2B_2^T. \tag{6.68}$$

From

$$\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}P_{M,1}\\ P_{M,2}\end{bmatrix}+\begin{bmatrix}P_{M,1}\\ P_{M,2}\end{bmatrix}A_{11}^T+\begin{bmatrix}N_{11} & N_{12}\\ N_{21} & N_{22}\end{bmatrix}\begin{bmatrix}P_{M,1}\\ P_{M,2}\end{bmatrix}N_{11}^T c=-\begin{bmatrix}B_1B_1^T\\ B_2B_1^T\end{bmatrix}$$

we also know that

$$A_{11}P_{M,1}+A_{12}P_{M,2}+P_{M,1}A_{11}^T+N_{11}P_{M,1}N_{11}^T c+N_{12}P_{M,2}N_{11}^T c=-B_1B_1^T. \tag{6.69}$$

We define $\mathscr{E}:=\left(\operatorname{tr}\left(C\Sigma C^T\right)+\operatorname{tr}\left(C_1P_RC_1^T\right)-2\ \operatorname{tr}\left(CP_MC_1^T\right)\right)^{\frac{1}{2}}$ and obtain

$$\mathscr{E}^2=\operatorname{tr}\left(\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}C_1^T\\ C_2^T\end{bmatrix}\right)+\operatorname{tr}\left(C_1P_RC_1^T\right)-2\ \operatorname{tr}\left(\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}P_{M,1}\\ P_{M,2}\end{bmatrix}C_1^T\right)$$
$$=\operatorname{tr}(C_2\Sigma_2C_2^T+C_1\Sigma_1C_1^T+C_1P_RC_1^T-2C_1P_{M,1}C_1^T-2C_2P_{M,2}C_1^T).$$

Using equation (6.66) yields

$$\begin{aligned}\operatorname{tr}(-C_2P_{M,2}C_1^T) &= \operatorname{tr}(-C_1^TC_2P_{M,2})\\ &= \operatorname{tr}(A_{21}^T\Sigma_2P_{M,2}+\Sigma_1A_{12}P_{M,2}+N_{11}^T\Sigma_1N_{12}P_{M,2}c+N_{21}^T\Sigma_2N_{22}P_{M,2}c)\\ &= \operatorname{tr}(A_{21}^T\Sigma_2P_{M,2}+A_{12}P_{M,2}\Sigma_1+N_{12}P_{M,2}N_{11}^T\Sigma_1c+N_{21}^T\Sigma_2N_{22}P_{M,2}c).\end{aligned}$$

By equation (6.69), we obtain

$$\begin{aligned}&\operatorname{tr}(-C_2P_{M,2}C_1^T)\\ &= \operatorname{tr}(A_{21}^T\Sigma_2P_{M,2}+N_{21}^T\Sigma_2N_{22}P_{M,2}c)-\operatorname{tr}((B_1B_1^T+P_{M,1}A_{11}^T+A_{11}P_{M,1}+N_{11}P_{M,1}N_{11}^Tc)\Sigma_1).\end{aligned}$$

Using equation (6.64), we have

$$\begin{aligned}\operatorname{tr}(P_{M,1}A_{11}^T+A_{11}P_{M,1}+N_{11}P_{M,1}N_{11}^Tc)\Sigma_1) &= \operatorname{tr}(A_{11}^T\Sigma_1+\Sigma_1A_{11}+N_{11}^T\Sigma_1N_{11}c)P_{M,1})\\ &= -\operatorname{tr}(C_1^TC_1P_{M,1}+N_{21}^T\Sigma_2N_{21}P_{M,1}c),\end{aligned}$$

and hence

$$\begin{aligned}\mathscr{E}^2 = &\operatorname{tr}(C_2\Sigma_2C_2^T+C_1\Sigma_1C_1^T+C_1P_RC_1^T)+2\operatorname{tr}(A_{21}^T\Sigma_2P_{M,2}+N_{21}^T\Sigma_2N_{22}P_{M,2}c)\\ &-2\operatorname{tr}(B_1B_1^T\Sigma_1)+2\operatorname{tr}(N_{21}^T\Sigma_2N_{21}P_{M,1}c).\end{aligned}$$

Thus,

$$\begin{aligned}\mathscr{E}^2 = &\operatorname{tr}(\Sigma_2(C_2^TC_2+2P_{M,2}A_{21}^T+2N_{22}P_{M,2}N_{21}^Tc+2N_{21}P_{M,1}N_{21}^Tc))\\ &+\operatorname{tr}(C_1\Sigma_1C_1^T-B_1B_1^T\Sigma_1+B_1^T(Q_R-\Sigma_1)B_1)\end{aligned}$$

using the identity $\operatorname{tr}(C_1P_RC_1^T)=\operatorname{tr}(B_1^TQ_RB_1)$. Inserting equation (6.67) provides

$$\begin{aligned}\operatorname{tr}(-B_1B_1^T\Sigma_1) &= \operatorname{tr}(A_{11}\Sigma_1\Sigma_1+\Sigma_1A_{11}^T\Sigma_1+N_{11}\Sigma_1N_{11}^T\Sigma_1c+N_{12}\Sigma_2N_{12}^T\Sigma_1c)\\ &= \operatorname{tr}(\Sigma_1\Sigma_1A_{11}+\Sigma_1A_{11}^T\Sigma_1+\Sigma_1N_{11}^T\Sigma_1N_{11}c+N_{12}\Sigma_2N_{12}^T\Sigma_1c)\\ &= -\operatorname{tr}(\Sigma_1C_1^TC_1)-\operatorname{tr}(\Sigma_1N_{21}^T\Sigma_2N_{21}c)+\operatorname{tr}(N_{12}\Sigma_2N_{12}^T\Sigma_1c).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathscr{E}^2 =& \operatorname{tr}(\Sigma_2(C_2^T C_2 + 2P_{M,2}A_{21}^T + 2N_{22}P_{M,2}N_{21}^T c + 2N_{21}P_{M,1}N_{21}^T c)) \\ &+ \operatorname{tr}(\Sigma_2(N_{12}^T\Sigma_1 N_{12}c - N_{21}\Sigma_1 N_{21}^T c)) \\ &+ \operatorname{tr}(B_1^T(Q_R - \Sigma_1)B_1)\end{aligned}$$

holds and from (6.65), it follows that

$$\begin{aligned}\operatorname{tr}(\Sigma_2 N_{12}^T\Sigma_1 N_{12}c) &= \operatorname{tr}(-\Sigma_2(A_{22}^T\Sigma_2 + \Sigma_2 A_{22} + N_{22}^T\Sigma_2 N_{22}c + C_2^T C_2)) \\ &= \operatorname{tr}(-\Sigma_2(\Sigma_2 A_{22}^T + A_{22}\Sigma_2 + N_{22}\Sigma_2 N_{22}^T c + C_2^T C_2)).\end{aligned}$$

Using (6.68) yields

$$\operatorname{tr}(\Sigma_2 N_{12}^T\Sigma_1 N_{12}c) = \operatorname{tr}(\Sigma_2(N_{21}\Sigma_1 N_{21}^T c + B_2 B_2^T - C_2^T C_2)),$$

such that

$$\operatorname{tr}(\Sigma_2(N_{12}^T\Sigma_1 N_{12}c - N_{21}\Sigma_1 N_{21}^T c)) = \operatorname{tr}(\Sigma_2(B_2 B_2^T - C_2^T C_2)),$$

and hence,

$$\begin{aligned}\mathscr{E}^2 =& \operatorname{tr}(\Sigma_2(B_2 B_2^T + 2P_{M,2}A_{21}^T + 2N_{22}P_{M,2}N_{21}^T c + 2N_{21}P_{M,1}N_{21}^T c)) \\ &+ \operatorname{tr}(B_1^T(Q_R - \Sigma_1)B_1).\end{aligned}$$

By definition, the Gramians P_R and Q_R satisfy

$$A_{11}^T Q_R + Q_R A_{11} + N_{11}^T Q_R N_{11}c = -C_1^T C_1$$

and

$$A_{11}P_R + P_R A_{11}^T + N_{11}P_R N_{11}^T c = -B_1 B_1^T.$$

Thus,

$$\begin{aligned}\operatorname{tr}(B_1B_1^T(Q_R-\Sigma_1)) &= \operatorname{tr}(-(A_{11}P_R+P_RA_{11}^T+N_{11}P_RN_{11}^Tc)(Q_R-\Sigma_1))\\ &= \operatorname{tr}(-P_R(A_{11}^T(Q_R-\Sigma_1)+(Q_R-\Sigma_1)A_{11}+N_{11}^T(Q_R-\Sigma_1)N_{11}c))\\ &= \operatorname{tr}(-P_RN_{21}^T\Sigma_2N_{21}c).\end{aligned}$$

Finally, we have

$$\mathscr{E}^2 = \operatorname{tr}(\Sigma_2(B_2B_2^T+2P_{M,2}A_{21}^T+2N_{22}P_{M,2}N_{21}^Tc+2N_{21}P_{M,1}N_{21}^Tc-N_{21}P_RN_{21}^Tc)).$$

□

If we set $N^k=0$ ($k=1,\dots,q$) in Proposition 6.19, then we obtain the $\mathscr{H}_2$ error bound of the deterministic case which can be found in Subsection 7.2.2 of Antoulas [2]. So, the result in Proposition 6.19 can be interpreted as a generalized $\mathscr{H}_2$ error bound. Furthermore, with this representation of the error bound we are able to emphasize the cases in which balanced truncation is a good approximation. In Proposition 6.19 the bound depends on Σ_2 which contains the $n-r$ smallest Hankel singular values $\sigma_{r+1},\dots,\sigma_n$ of the original system. In case these values are small, the error bound indicates that the reduced order model computed by balanced truncation is of good quality.

6.2.5. Numerical experiments

In this subsection, we consider a stochastic heat and a stochastic damped wave equation as examples to support the theory.
For the stochastic heat equation case we conduct parts of the numerical experiments from Section 5 in [14] with a different noise process and we furthermore extend them by plots in which outputs of large order systems are compared with outputs corresponding to ROMs obtained via BT. For the numerical experiments related to the stochastic damped wave equation we follow [60] in order to compare outputs of semi-discretized SPDEs with outputs of ROMs by BT.
The numerical experiments in this subsection are run on a desktop computer with a dual-core Intel Pentium processor E5400 and 3GB RAM. All algorithms are implemented and executed in MATLAB 7.14.0.739 (R2012a) running on Ubuntu 10.04.1 LTS.

Balanced truncation applied to a stochastic heat equation

We recall the stochastic heat equation from Example 4.14 ($t \in [0,\pi]$, $\zeta \in [0,\pi]^2$):

$$\begin{aligned}
\frac{\partial \mathscr{X}(t,\zeta)}{\partial t} &= \Delta \mathscr{X}(t,\zeta) + 1_{[\frac{\pi}{4},\frac{3\pi}{4}]^2}(\zeta)u(t) + \mathrm{e}^{-|\zeta_1-\frac{\pi}{2}|-\zeta_2}\mathscr{X}(t-,\zeta)\frac{\partial M(t)}{\partial t},\\
\frac{\partial \mathscr{X}(t,\zeta)}{\partial \mathbf{n}} &= 0, \quad t\in[0,\pi],\ \zeta \in \partial[0,\pi]^2,\\
\mathscr{X}(0,\zeta) &\equiv 0,
\end{aligned}$$

where $M(t) = w(t) - (N(t) - t)$. We assume here that $w(t)$ and $N(t)$, $t \in [0,\pi]$, are independent processes, where w is a standard Wiener process and N is a Poisson process with parameter 1. Instead of the full state, we are only interested in finitely many observations which here is the average temperature on the non-heated area

$$\mathscr{Y}(t) = \frac{4}{3\pi^2}\int_{[0,\pi]^2\setminus[\frac{\pi}{4},\frac{3\pi}{4}]^2} \mathscr{X}(t,\zeta)d\zeta. \tag{6.70}$$

We approximate $\mathscr{Y}$ by the output of the Galerkin solution discusses in Chapter 5. The Galerkin approximation output for the particular example above is specified in Example 5.3 and looks as follows

$$y(t) = Cx(t),$$

where $C^T = (\mathscr{C}h_k)_{k=1,\dots,n}$. Here, $\mathscr{C}$ is the integral operator defined at the right hand side in (6.70) and $(h_k)_{k=1,\dots,n}$ are the eigenvectors of the Laplace operator with homogeneous Neumann boundary conditions. The corresponding state x is given by (see Example 5.3 again)

$$x(t) = \int_0^t Ax(s) + Bu(s)ds + \int_0^t Nx(s-)dM(s), \tag{6.71}$$

where

- $A = \operatorname{diag}(0,-1,-1,-2,\dots)$,
- $N = \left(\left\langle \mathrm{e}^{-|\cdot-\frac{\pi}{2}|-\cdot} h_i, h_k\right\rangle_{L^2([0,\pi])}\right)_{k,i=1,\dots,n}$,
- $B = \left(\left\langle 1_{[\frac{\pi}{4},\frac{3\pi}{4}]^2}(\cdot), h_k\right\rangle_{L^2([0,\pi])}\right)_{k=1,\dots,n}$.

We fix the state dimension to $n = 1000$. We now reduce the Galerkin solution by applying BT. Before doing this, we have to ensure that the system is mean square asymptotically stable. This is not satisfied since A has a zero eigenvalue. For that reason, we add a stabilizing feedback control $u^S(t) = -2e_1^T x(t)$, $t \in [0, \pi]$, where e_1 is the first unit vector in $\mathbb{R}^n$, such that the control we use is of the form $u(t) = u^S(t) + \tilde{u}(t)$ with arbitrary $\tilde{u} \in L^2_T$ for every $T > 0$. Consequently, we obtain an equation (6.71), where A and u are replaced by $A_S = A - 2Be_1^T$ and $\tilde{u}$, respectively. To now ensure mean square asymptotically stability of the modified system, we have to show that

$$\sigma\left(I_n \otimes A + A \otimes I_n + N_M \otimes N_M\right) \subset \mathbb{C}_- \tag{6.72}$$

by Theorem 6.3, where $N_M = N\sqrt{\mathbb{E}\left[M(1)^2\right]}$. Property (6.72) now holds since the following sufficient condition is satisfied (see Corollary 3.6.3 in [21] and Theorem 5 in [30]):

$$\left\| \int_0^\infty \mathrm{e}^{A_S^T t}\, N_M^T N_M \,\mathrm{e}^{A_S t}\, dt \right\| = 0.1316 < 1.$$

The error between the output y^s of the stabilized system and the output $\hat{y}$ of the ROM by BT from (6.25) is

$$\sup_{t\in[0,\pi]} \mathbb{E}\left\|y^s(t) - \hat{y}(t)\right\|_2 \leq \mathscr{E}\left\|\tilde{u}\right\|_{L^2_T},$$

where $\mathscr{E}$ is the bound in (6.63). For fixed control functions $u_1(t) = \frac{\sqrt{2}}{\pi} w(t)$, $u_2(t) = \sqrt{\frac{2}{1-\mathrm{e}^{-2\pi}}}\,\mathrm{e}^{-t}$, $t \in [0, \pi]$, and the reduced order dimensions $r = 1, 2, 4, 8$ we compute the exact time domain errors. We compare these exact errors with the corresponding error bounds in the following table.

Dim. ROM	Exact Error ($\tilde{u} = u_1$)	Exact Error ($\tilde{u} = u_2$)	Error Bound $\mathscr{E}$
8	$3.6163 \cdot 10^{-6}$	$4.4354 \cdot 10^{-6}$	$3.8971 \cdot 10^{-5}$
4	$1.0025 \cdot 10^{-4}$	$1.2570 \cdot 10^{-4}$	$7.2362 \cdot 10^{-4}$
2	$1.0792 \cdot 10^{-3}$	$1.1681 \cdot 10^{-3}$	$3.8652 \cdot 10^{-3}$
1	$3.8439 \cdot 10^{-3}$	0.0101	0.0335

In order to obtain the exact errors, we discretize the equations with the Euler-Maruyama scheme. The theory regarding this method can be found in [41] for the Wiener case and equations with additional Poisson noise are studied in [31, 32].

These results show that the balanced truncation error bound, which is a worst case bound holding for all feasible input functions, also provides a good prediction of the true time domain error. In particular, it quite well predicts the decrease of the true error for increased dimension of the reduced order model.

To conclude this subsection, we compare trajectories of the output corresponding to the semi-discretized stochastic heat equation with the ones of the reduced order model output. We choose the reduced order dimension $r = 1, 2, 3$ and a control of the form $\tilde{u}(t,\omega) = \tilde{u}_D(t)r(t,\omega) = \mathrm{e}^{w(t,\omega)}$, $t \in [0,\pi]$ and $\omega \in \Omega$, where $r(t,\omega) = \mathrm{e}^{w(t,\omega)}$ is interpreted as a multiplicative positive random perturbation of $\tilde{u}_D \equiv 1$. Corresponding trajectories are given in Figure 6.1. The outputs shown there illustrate how the temperature in the non-heated area is increased by the input. Furthermore, they are continuously impacted by Wiener noise and the Poisson noise causes jumps which reduce the temperature. In Figure 6.1a we fix the reduced order dimension to $r = 1$ and obtain an output which is already not far from the original one. Setting $r = 2$ provides a trajectory in Figure 6.1a which only differs slightly from the output of the original system. Finally, in Figure 6.1c, where $r = 3$, it is not possible to distinguish between the graphs. Consequently, in this case the SPDE in Example 4.14 is well represented by a ROM of order three.

Balanced truncation applied to a stochastic damped wave equation

We apply balancing BT to the discretized version of the following SPDE from Example 4.16. It models the lateral displacement of a damped stretched string with random excitation:

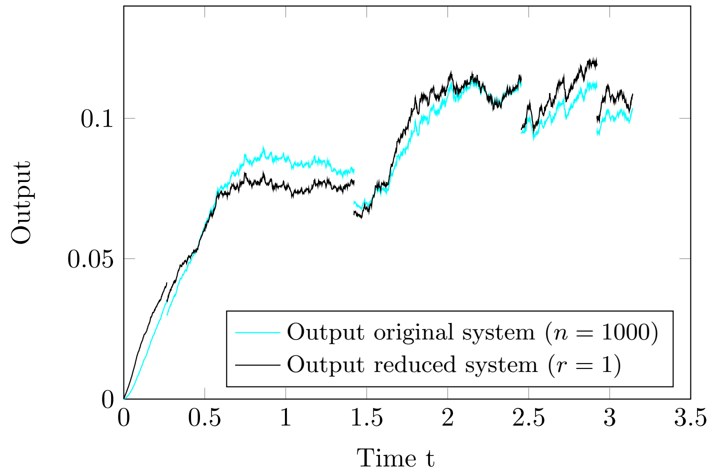
$$\frac{\partial^2}{\partial t^2}\mathscr{X}(t,\zeta) + 2\frac{\partial}{\partial t}\mathscr{X}(t,\zeta) = \frac{\partial^2}{\partial \zeta^2}\mathscr{X}(t,\zeta) + \mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}u(t) + 2\,\mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}\,\mathscr{X}(t-,\zeta)\frac{\partial}{\partial t}(t - N(t))$$

for $t, \zeta \in [0,\pi]$, $\alpha = 2$ and with $(N(t))_{t\geq 0}$ being a Poisson process with parameter 1. We have Dirichlet boundary conditions and $\mathscr{X}(0,\zeta)$, $\frac{\partial}{\partial t}\mathscr{X}(t,\zeta)\Big|_{t=0} \equiv 0$.

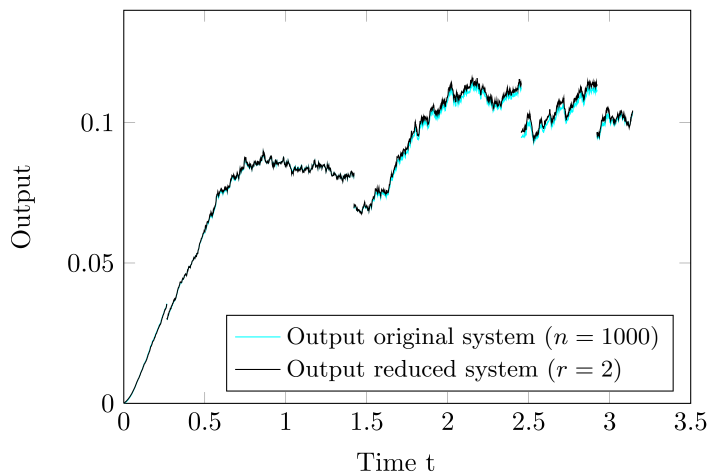
The output, we are interested in, is an approximation for the position of the middle of the string

$$\mathscr{Y}(t) = \frac{1}{2\varepsilon}\int_{\frac{\pi}{2}-\varepsilon}^{\frac{\pi}{2}+\varepsilon}\mathscr{X}(t,\zeta)d\zeta,$$

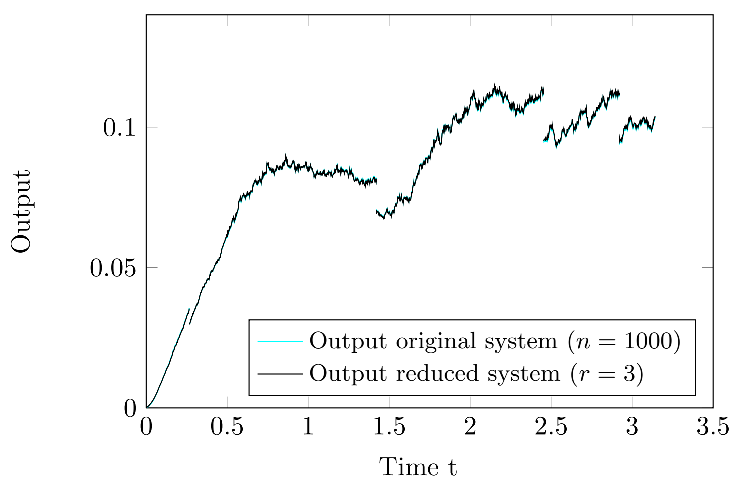
where $\varepsilon > 0$. In order to approximate this output, we use the Galerkin scheme outlined in Chapter



(a) Reduced order dimension $r = 1$ and $\tilde{u}(t) = e^{w(t)}$



(b) Reduced order dimension $r = 2$ and $\tilde{u}(t) = e^{w(t)}$



(c) Reduced order dimension $r = 3$ and $\tilde{u}(t) = e^{w(t)}$

Figure 6.1.: Average temperature non-heated area compared with reduced order output by BT

5. From Example 5.5 we know that the Galerkin system looks as follows:

$$x(t) = \int_0^t Ax(s) + Bu(s)ds + \int_0^t Nx(s-)dM(s), \qquad (6.73)$$
$$y(t) = Cx(t), \quad t \geq 0,$$

where the coefficients are represented by

- $A = \operatorname{diag}\left(E_1, \ldots, E_{\frac{n}{2}}\right)$ with $E_l = \left(\begin{smallmatrix} 0 & l \\ -l & -2 \end{smallmatrix}\right)$ $(l = 1, \ldots, \frac{n}{2})$,
- $B = (b_k)_{k=1,\ldots,n}$ with

$$b_{2l-1} = 0, \quad b_{2l} = \sqrt{\frac{2}{\pi}} \left\langle \mathrm{e}^{-(\cdot - \frac{\pi}{2})^2}, \sin(l\cdot) \right\rangle_H,$$

- $N = \left(n_{kj}\right)_{k,j=1,\ldots,n}$ with

$$n_{(2l-1),j} = 0, \quad n_{(2l),j} = \begin{cases} 0, & \text{if } j = 2v, \\ \frac{4}{\pi v} \left\langle \sin(l\cdot), \mathrm{e}^{-(\cdot - \frac{\pi}{2})^2} \sin(v\cdot) \right\rangle_H, & \text{if } j = 2v-1, \end{cases}$$

  for $j = 1, \ldots, n$ and $v = 1, \ldots, \frac{n}{2}$,
- the output matrix is given by $C^T = (c_k)_{k=1,\ldots,n}$ with

$$c_{2l} = 0 \text{ and } c_{2l-1} = \frac{1}{\sqrt{2\pi} l^2} \left[\cos\left(l\left(\frac{\pi}{2} - \varepsilon\right)\right) - \cos\left(l\left(\frac{\pi}{2} + \varepsilon\right)\right)\right].$$

We fix the dimension of the Galerkin solution to $n = 1000$ and reduce (6.73) by BT. This model order reduction method only works if the system is mean square asymptotically stable, which means that

$$\mathbb{E} \left\| x(t, x_0, 0) \right\|_{\mathbb{R}^n}^2 \to 0 \qquad (6.74)$$

for $t \to \infty$ and an arbitrary initial state $x_0 \in \mathbb{R}^n$. Below, we can ensure that property since the matrix equation

$$A^T X + XA + N^T XN = -I$$

has a positive definite solution $X > 0$, which we check by solving this equation with a numerical algorithm. Due to Theorem 6.3 this is equivalent to condition (6.74), hence the desired model order reduction technique can be used. Let $\hat{y}$ be the ROM output by BT from (6.25). We now compute the exact error

$$\sup_{t\in[0,\pi]} \mathbb{E}\left\|y(t)-\hat{y}(t)\right\|_2,$$

and the corresponding error bound in (6.63) for different dimensions r of the reduced order model. Below, we choose particular normalized control functions $u_1(t) = \sqrt{\frac{2}{\pi}}1_{[0,\frac{\pi}{2}]}(t)$ and $u_2(t) = \frac{\sqrt{8}}{\pi}1_{[0,\frac{\pi}{2}]}(t)w(t)$ $(t \in [0,\pi])$. Moreover, w is a standard Wiener process and the error bound is denoted by $\mathscr{E}_1 := \left(\operatorname{tr}\left(CPC^T\right) + \operatorname{tr}\left(C_1P_RC_1^T\right) - 2\operatorname{tr}\left(CP_MC_1^T\right)\right)^{\frac{1}{2}}$.

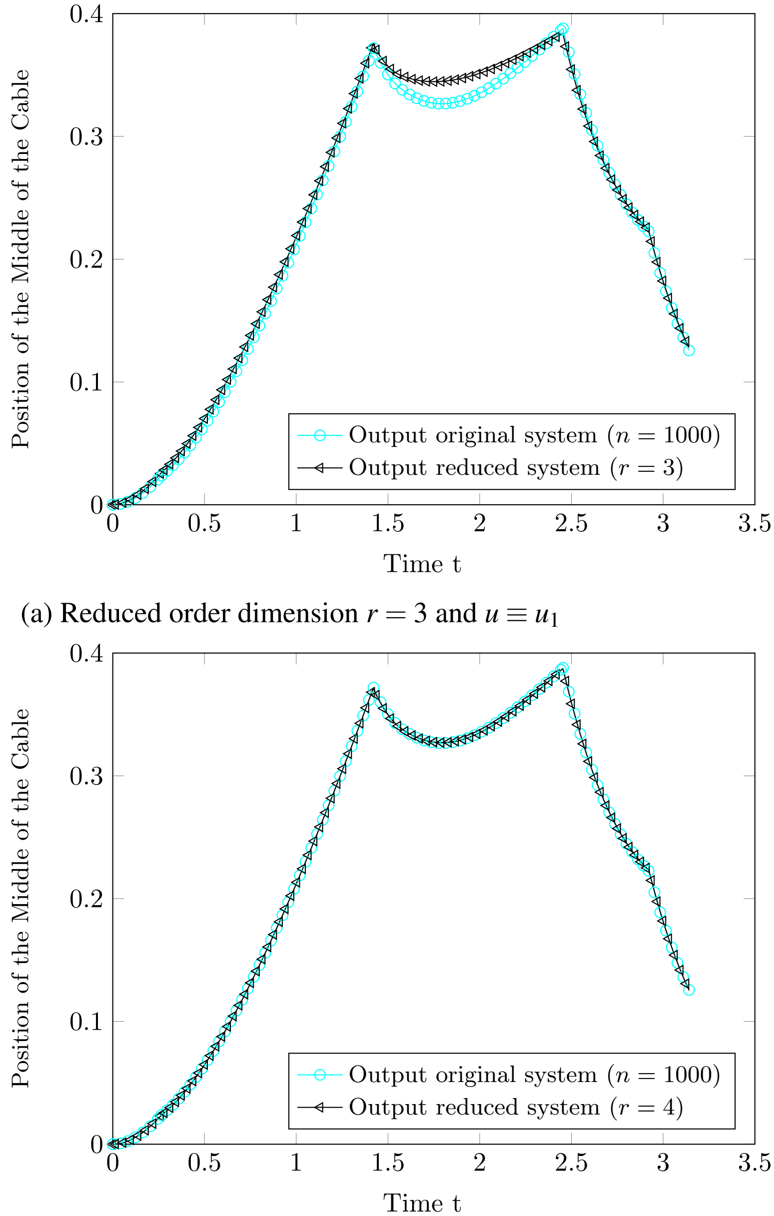
Dim. ROM	BT Exact Error $(u=u_1)$	BT Exact Error $(u=u_2)$	Error Bound $\mathscr{E}_1$
40	$1.9992\cdot 10^{-7}$	$2.0024\cdot 10^{-7}$	$4.0103\cdot 10^{-5}$
20	$4.4660\cdot 10^{-6}$	$4.1435\cdot 10^{-6}$	$1.2695\cdot 10^{-4}$
10	$4.3081\cdot 10^{-5}$	$3.2512\cdot 10^{-5}$	$3.6395\cdot 10^{-4}$
5	$5.1180\cdot 10^{-4}$	$4.2176\cdot 10^{-4}$	$2.3446\cdot 10^{-3}$
3	0.0114	$8.2309\cdot 10^{-3}$	0.0380

(6.75)

In order to obtain the exact errors, we discretize the equations with the Euler-Maruyama scheme, compare [31, 32, 41].

In Figure 6.2a we plot the output y of the Galerkin system with state space dimension $n = 1000$ and the output $\hat{y}$ of the ROM with state space dimension $r = 3$, where we choose $u \equiv u_1$. Due to the input, which can be interpreted as electricity flowing through the cable, the curves are increasing first. Additionally, the cable is randomly hit by wind which is marked by the peaks in this picture. This effect pushes the cable in the opposite direction. After the current completely passed the cable, the graphs decrease to zero due to the stability of the system. It is also obvious that even after such a large reduction of the dimension, the accuracy is quite good. In Figure 6.2b we increase the dimension of the reduced order model by one so that it is difficult to distinguish between the output of the ROM and the output of the Galerkin system. Hence, one can conclude that the output of the SPDE in Example 4.16 can be described by a system of ordinary SDEs of order four.

We conclude this subsection with pointing out the benefit of model order reduction compared to



(a) Reduced order dimension $r = 3$ and $u \equiv u_1$

(b) Reduced order dimension $r = 4$ and $u \equiv u_1$

Figure 6.2.: Output of the original model compared with the output of the ROM by BT

a less fine discretization. Below, we denote the output of the Galerkin system (6.73) with state space dimension n by y_n. We consider the errors

$$\sup_{t\in[0,\pi]} \mathbb{E}\left\|y_{1000}(t)-y_{n_0}(t)\right\|_2, \tag{6.76}$$

where $n_0 = 5, 10, 20$ and state this errors in the following table.

n_0	Exact error (6.76) with $u \equiv u_1$	Exact error (6.76) with $u \equiv u_2$
20	$4.2207 \cdot 10^{-4}$	$3.8454 \cdot 10^{-4}$
10	$4.8938 \cdot 10^{-3}$	$2.8325 \cdot 10^{-3}$
5	0.0277	0.0197

Comparing the exact errors in the table above with the exact error stated in table (6.75), we can see that BT gives low order systems which are better than low order systems resulting from a less fine discretization. Here, BT yields ROMs which are more accurate by a factor of almost 100.

6.3. Type 2 balanced truncation for stochastic systems

In Benner et al. [10] and Damm, Benner [22] an example is presented which clarifies that the $\mathcal{H}_\infty$-error bound from the deterministic case does not hold for the type 1 BT approach that we discuss in Section 6.2. For that reason, a new ansatz to extend BT to SDEs, which we call type 2 BT, is considered by Benner et al. [10] or Damm, Benner [22]. There a new reachability Gramian is used which is not based on the concept introduced in Subsection 6.1.1 and does not allow an energy interpretation. The Gramian is motivated by the missing $\mathcal{H}_\infty$-error bound which can be achieved in the approach here. Moreover, type 2 BT preserves mean square asymptotic stability as type 1 BT does, see Subsection 6.2.2. In this section, we focus on giving an overview on the most important results that have already been proven in [10] and [22]. In Subsection 6.3.1, we briefly discuss the procedure and emphasize results on error bounds and the stability analysis of the methods. In Subsection 6.3.2, we contribute an $\mathcal{H}_2$-type error bound for the new ansatz in [10] and [22] to close the gap in the error bound analysis. This error bound is already established in Redmann, Benner [59] for only one noise term. We consider a more general setting here for a general number $q \in \mathbb{N}$ of noise terms.

6.3.1. Comparison with type 1 BT, procedure and properties of the ROM

We consider the same system as in Subsection 6.2.1, that is

$$dx(t) = [Ax(t) + Bu(t)]dt + \sum_{k=1}^{q} N^k x(t-)dM_k(t), \quad t \geq 0, \quad x(0) = x_0, \tag{6.77}$$
$$y(t) = Cx(t)$$

for A, $N^k \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $u \in L^2_T$ for every $T > 0$. $M_1,\ldots,M_q$ are scalar uncorrelated and square integrable Lévy processes with mean zero. These processes are defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, where $(\mathscr{F}_t)_{t\geq 0}$ satisfies the usual conditions. In addition, we assume M_k $(k = 1,\ldots,q)$ to be $(\mathscr{F}_t)_{t\geq 0}$-adapted and the increments $M_k(t+h) - M_k(t)$ to be independent of $\mathscr{F}_t$ for $t, h \geq 0$. We denote the solution of equation (6.77) by $x(t, x_0, u)$ and the corresponding output by $y(t, x_0, u)$. The SDE in (6.77) is mean square asymptotic stability which means

$$\mathbb{E}\left\|x(t, x_0, 0)\right\|_2^2 \to 0 \quad \text{for } t \to \infty. \tag{6.78}$$

Following the arguments in Sections 6.1.1 and 6.1.2 the Gramians of type 1 BT are based on reachability and observability concepts. If we suppose to have a completely observable and reachable system, which implies P, $Q > 0$, we obtain the following result:

(i) The minimal energy to steer the average state to $x \in \mathbb{R}^n$ is bounded from below as follows:

$$x^T P^{-1} x \leq \inf_{\substack{u\in L^2_T, T>0,\\ \mathbb{E}[x(T,0,u)]=x}} \|u\|^2_{L^2_T}. \tag{6.79}$$

(ii) The energy that is caused by the observation of an initial state $x_0 \in \mathbb{R}^n$ is

$$\|y(\cdot, x_0, 0)\|^2_{L^2} = x_0^T Q x_0,$$

where the Gramians P and Q are solutions to generalized Lyapunov equations:

$$AP + PA^T + \sum_{k=1}^{q} N^k P (N^k)^T \cdot c_k = -BB^T \text{ and } A^T Q + QA + \sum_{k=1}^{q} (N^k)^T Q N^k \cdot c_k = -C^T C \tag{6.80}$$

with $c_k := \mathbb{E}\left[M_k(1)^2\right]$. For an $\mathscr{H}_\infty$-type error bound for type 1 BT we desire to find an arbitrary constant $a > 0$ such that

$$\|y - y_R\|_{L^2} \leq a\,\mathrm{tr}(\Sigma_2)\,\|u\|_{L^2}$$

holds for Σ_2 containing the Hankel singular values corresponding to the truncated part and y_R being the reduced order output by type 1 BT in (6.25). For the deterministic case ($N = 0$), this constant is $a = 2$, see [2]. Example I.3 in [10] or Example II.2 in [22], respectively, show that such a number $a > 0$ does not exist for general matrices N.

The idea behind type 2 BT for stochastic systems, introduced in [22], is to replace the reachability Gramian P satisfying the first equation in (6.80) by a matrix P_2 which solves

$$A^T P_2^{-1} + P_2^{-1} A + \sum_{k=1}^{q} (N^k)^T P_2^{-1} N^k \cdot c_k = -P_2^{-1} B B^T P_2^{-1}. \tag{6.81}$$

This new choice has the disadvantage that P_2 does not allow an energy interpretation as in (6.79). If we set $N^k = 0$ in the first equation of (6.80) and in (6.81), we obtain the same Lyapunov equation that is used for deterministic BT. Hence, both types of introducing BT for SDEs are possible generalizations. Unfortunately, there exist no criteria for the existence of a positive definite solution to (6.81). There are matrices A, B and N^k such that the reachability Gramian P of type 1 BT is positive definite, but there is no positive definite solution to (6.81), see Example II.5 in [10]. For that reason, we turn to a more general matrix *inequality*

$$A^T P_2^{-1} + P_2^{-1} A + \sum_{k=1}^{q} (N^k)^T P_2^{-1} N^k \cdot c_k \leq -P_2^{-1} B B^T P_2^{-1}. \tag{6.82}$$

We have an existence result for this inequality from [10].

Lemma 6.20. *Suppose that (6.78) holds, then there is a solution $P_2 > 0$ to inequality (6.82).*

Balancing for the new method means that we diagonalize the positive definite solutions to the second equation in (6.80) and inequality (6.82) simultaneously. Analogous to the type 1 approach, there is an invertible transformation matrix $\tilde{T}$, so that

$$\left(A, B, C, N^k\right) \mapsto \left(\tilde{T} A \tilde{T}^{-1}, \tilde{T} B, C \tilde{T}^{-1}, \tilde{T} N^k \tilde{T}^{-1}\right)$$

leads to a system with transformed Gramians

$$\tilde{T}P_2\tilde{T}^T = \tilde{T}^{-T}Q\tilde{T}^{-1} = {}_2\Sigma = \operatorname{diag}({}_2\sigma_1, \dots, {}_2\sigma_n) > 0.$$

The fact that $\tilde{P}_2 := \tilde{T}P_2\tilde{T}^T$ is the reachability Gramian of the transformed system is a consequence of simple rearrangements of the inequality (6.82):

$$\begin{aligned}-P_2^{-1}\tilde{T}^{-1}\tilde{T}BB^T\tilde{T}^T\tilde{T}^{-T}P_2^{-1} &= -P_2^{-1}BB^TP_2^{-1}\\ &\geq A^TP_2^{-1} + P_2^{-1}A + \sum_{k=1}^{q}(N^k)^TP_2^{-1}N^k \cdot c_k\\ &= A^T\tilde{T}^T\tilde{T}^{-T}P_2^{-1} + P_2^{-1}\tilde{T}^{-1}\tilde{T}A + \sum_{k=1}^{q}(N^k)^T\tilde{T}^T\tilde{T}^{-T}P_2^{-1}\tilde{T}^{-1}\tilde{T}N^k \cdot c_k.\end{aligned}$$

Pre- and post-multiplying $\tilde{T}^{-T}$ and $\tilde{T}^{-1}$ the above matrix inequality yields

$$-\tilde{P}_2^{-1}(\tilde{T}B)(\tilde{T}B)^T\tilde{P}_2^{-1} \geq (\tilde{T}A\tilde{T}^{-1})^T\tilde{P}_2^{-1} + \tilde{P}_2^{-1}(\tilde{T}A\tilde{T}^{-1}) + \sum_{k=1}^{q}(\tilde{T}N^k\tilde{T}^{-1})^T\tilde{P}_2^{-1}(\tilde{T}N^k\tilde{T}^{-1}) \cdot c_k.$$

The matrix $\tilde{T}$ has the same structure as in the type 1 approach in Subsection 6.2.1. With the partition

$$\tilde{T} = \begin{bmatrix}\tilde{W}^T\\ \tilde{T}_2^T\end{bmatrix} \text{ and } \tilde{T}^{-1} = \begin{bmatrix}\tilde{V} & \tilde{T}_1\end{bmatrix},$$

where $\tilde{W}^T \in \mathbb{R}^{r\times n}, \tilde{V} \in \mathbb{R}^{n\times r}$, we obtain the ROM coefficients

$$\left({}_2A_{11}, {}_2B_1, {}_2C_1, {}_2N_{11}^k\right) = \left(\tilde{W}^TA\tilde{V}, \tilde{W}^TB, C\tilde{V}, \tilde{W}^TN^k\tilde{V}\right).$$

Type 2 balanced truncation preserves mean square asymptotic stability as can be shown like in Theorem II.2 in [10]. So, in the proof of the general Lévy noise case one has to replace the matrices N^k, which occur in the proof of the Wiener case, by $N_M^k := N^k\sqrt{\mathbb{E}\left[M_k^2(1)\right]}$.

Theorem 6.21. *Let ${}_2\sigma_r \neq {}_2\sigma_{r+1}$, then the ROM*

$$d\tilde{x}_R(t) = {}_2A_{11}\tilde{x}_R(t)dt + \sum_{k=1}^{q} {}_2N_{11}^k\tilde{x}_R(t-)dM_k(t), \quad t \geq 0, \quad \tilde{x}_R(0) = \tilde{x}_{R,0}$$

is mean square asymptotically stable if

$$dx(t) = Ax(t)dt + \sum_{k=1}^{q} N^k x(t-)dM_k(t), \quad t \geq 0, \quad x(0) = x_0$$

is mean square asymptotically stable.

The advantage of type 2 BT is the existence of an $\mathscr{H}_\infty$-type error bound which is in contrast to the type 1 method. Below, a result from [10] and [22], respectively is stated in a more general framework. This generalization can be shown by substituting N^k with N^k_M in the proof.

Theorem 6.22. *If $x_0 = 0$ and $\tilde{x}_{R,0} = 0$, then for all $T > 0$, we have*

$$\|y - \tilde{y}_R\|_{L^2_T} \leq 2({}_2\sigma_{r+1} + \ldots + {}_2\sigma_\nu) \|u\|_{L^2_T},$$

where $\tilde{y}_R$ is the output of the type 2 approach and ${}_2\sigma_{r+1}, \ldots, {}_2\sigma_\nu$ are the distinct diagonal entries of ${}_2\Sigma_2 = \operatorname{diag}({}_2\sigma_{r+1}, \ldots, {}_2\sigma_n) = \operatorname{diag}({}_2\sigma_{r+1} I, \ldots, {}_2\sigma_\nu I)$.

As mentioned in [10], the existence of an $\mathscr{H}_2$-type error bound is an open question which we answer in the next section.

6.3.2. $\mathscr{H}_2$ error bound for type 2 balanced truncation

For simplicity of notation, we assume to have a balanced realization of system (6.77) in terms of the type 2 approach. We denote this balanced realization by $({}_2A, {}_2B, {}_2C, {}_2N)$ in order to distinguish between the coefficients of the type 1 and the type 2 ansatz. Since we are in a balanced situation, $P_2 = Q = {}_2\Sigma$ such that

$${}_2A^T {}_2\Sigma + {}_2\Sigma {}_2A + \sum_{k=1}^{q} ({}_2N^k)^T {}_2\Sigma {}_2N^k \cdot c_k = -{}_2C^T {}_2C, \tag{6.83}$$

$${}_2A^T {}_2\Sigma^{-1} + {}_2\Sigma^{-1} {}_2A + \sum_{k=1}^{q} ({}_2N^k)^T {}_2\Sigma^{-1} {}_2N^k \cdot c_k \leq -{}_2\Sigma^{-1} {}_2B {}_2B^T {}_2\Sigma^{-1}, \tag{6.84}$$

where $c_k := \mathbb{E}\left[M_k^2(1)\right]$. Below, we use the following suitable partitions:

$${}_2A = \begin{bmatrix} {}_2A_{11} & {}_2A_{12} \\ {}_2A_{21} & {}_2A_{22} \end{bmatrix}, \; {}_2B = \begin{bmatrix} {}_2B_1 \\ {}_2B_2 \end{bmatrix}, \; {}_2C = \begin{bmatrix} {}_2C_1 & {}_2C_2 \end{bmatrix}, \; {}_2N^k = \begin{bmatrix} {}_2N^k_{11} & {}_2N^k_{12} \\ {}_2N^k_{21} & {}_2N^k_{22} \end{bmatrix}, \; {}_2\Sigma = \begin{bmatrix} {}_2\Sigma_1 & \\ & {}_2\Sigma_2 \end{bmatrix}.$$

By assuming $x_0 = 0$ and $\tilde{x}_{R,0} = 0$, we obtain representations for the outputs as

$$y(t) = {}_2Cx(t) = {}_2C\int_0^t \Phi(t,s){}_2Bu(s)ds \quad \text{and} \quad \tilde{y}_R(t) = {}_2C_1\tilde{x}_R(t) = {}_2C_1\int_0^t \tilde{\Phi}_R(t,s){}_2B_1u(s)ds,$$

where $\Phi(t,s) = \Phi(t)\Phi^{-1}(s)$, $t \geq s \geq 0$ and $\tilde{\Phi}_R$ is the fundamental matrix of the reduced order system. These representations have been proven in Proposition 6.4. Some straightforward rearrangements yield a first error estimate

$$\begin{aligned}
\mathbb{E}\left\|y(t)-\tilde{y}_R(t)\right\|_2 &= \mathbb{E}\left\|{}_2C\int_0^t \Phi(t,s){}_2Bu(s)ds - {}_2C_1\int_0^t \tilde{\Phi}_R(t,s){}_2B_1u(s)ds\right\|_2 \\
&\leq \mathbb{E}\int_0^t \left\|\left({}_2C\Phi(t,s){}_2B - {}_2C_1\tilde{\Phi}_R(t,s){}_2B_1\right)u(s)\right\|_2 ds \\
&\leq \mathbb{E}\int_0^t \left\|{}_2C\Phi(t,s){}_2B - {}_2C_1\tilde{\Phi}_R(t,s){}_2B_1\right\|_F \left\|u(s)\right\|_2 ds.
\end{aligned}$$

Using the Cauchy inequality, we have that

$$\mathbb{E}\left\|y(t)-\tilde{y}_R(t)\right\|_2 \leq \left(\mathbb{E}\int_0^t \left\|{}_2C\Phi(t,s){}_2B - {}_2C_1\tilde{\Phi}_R(t,s){}_2B_1\right\|_F^2 ds\right)^{\frac{1}{2}} \left(\mathbb{E}\int_0^t \left\|u(s)\right\|_2^2 ds\right)^{\frac{1}{2}}.$$

Following the arguments in Section 6.2.4, we obtain

$$\begin{aligned}
\mathbb{E}\int_0^t \left\|{}_2C\Phi(t,s){}_2B - {}_2C_1\tilde{\Phi}_R(t,s){}_2B_1\right\|_F^2 ds &= \mathbb{E}\int_0^t \left\|{}_2C\Phi(s){}_2B - {}_2C_1\tilde{\Phi}_R(s){}_2B_1\right\|_F^2 ds \\
&\leq \mathbb{E}\int_0^\infty \left\|{}_2C\Phi(s){}_2B - {}_2C_1\tilde{\Phi}_R(s){}_2B_1\right\|_F^2 ds \\
&= \operatorname{tr}\left({}_2CP\,{}_2C^T\right) + \operatorname{tr}\left({}_2C_1\,{}_2P_R\,{}_2C_1^T\right) - 2\operatorname{tr}\left({}_2C\,{}_2P_M\,{}_2C_1^T\right),
\end{aligned}$$

where the matrices P, ${}_2P_R$ and ${}_2P_M$ exist by assumption (6.78) and Theorem 6.21. They are the unique solutions to

$$ {}_2AP + P\,{}_2A^T + \sum_{k=1}^q {}_2N^kP({}_2N^k)^T \cdot c_k = -{}_2B\,{}_2B^T \tag{6.85}$$

$$ {}_2A_{11}\,{}_2P_R + {}_2P_R\,{}_2A_{11}^T + \sum_{k=1}^q {}_2N_{11}^k\,{}_2P_R({}_2N_{11}^k)^T \cdot c_k = -{}_2B_1\,{}_2B_1^T, \tag{6.86}$$

$$ {}_2A\,{}_2P_M + {}_2P_M\,{}_2A_{11}^T + \sum_{k=1}^q {}_2N^k\,{}_2P_M({}_2N_{11}^k)^T \cdot c_k = -{}_2B\,{}_2B_1^T. \tag{6.87}$$

Thus, we obtain an error bound

$$\sup_{t\in[0,T]} \mathbb{E}\left\|y(t)-\tilde{y}_R(t)\right\|_2 \leq \left(\operatorname{tr}\left({}_2C P\, {}_2C^T\right) + \operatorname{tr}\left({}_2C_1\, {}_2P_R\, {}_2C_1^T\right) - 2\, \operatorname{tr}\left({}_2C\, {}_2P_M\, {}_2C_1^T\right)\right)^{\frac{1}{2}} \|u\|_{L^2_T},$$

which we specify in the next theorem.

Theorem 6.23. *Let the realization $({}_2A, {}_2B, {}_2C, {}_2N)$ be balanced in terms of the type 2 approach, then*

$$\begin{aligned}
&\operatorname{tr}\left({}_2CP\, {}_2C^T + {}_2C_1\, {}_2P_R\, {}_2C_1^T - 2 {}_2C\, {}_2P_M\, {}_2C_1^T\right)\\
&= \operatorname{tr}({}_2\Sigma_2({}_2B_2\, {}_2B_2^T + 2 {}_2P_{M,2}\, {}_2A_{21}^T))\\
&\quad + \sum_{k=1}^{q} \operatorname{tr}({}_2\Sigma_2(2\, {}_2N_{22}^k\, {}_2P_{M,2}\, ({}_2N_{21}^k)^T + 2\, {}_2N_{21}^k\, {}_2P_{M,1}\, ({}_2N_{21}^k)^T - {}_2N_{21}^k\, {}_2P_R\, ({}_2N_{21}^k)^T))c_k,
\end{aligned}$$

where ${}_2P_{M,1}$ are the first r and ${}_2P_{M,2}$ are the last $n-r$ rows of ${}_2P_M$.

Proof. To simplify the notation in this proof, we set $q=1$, $N=N^1$ and $c=c_1$. We further replace the left subscript 2, marking matrices related to the type 2 approach, by a tilde ~. This means that we substitute ${}_2A$ with $\tilde{A}$, for example.

By selecting the left and right upper block of (6.83), we have

$$\tilde{A}_{11}^T\tilde{\Sigma}_1 + \tilde{\Sigma}_1\tilde{A}_{11} + \tilde{N}_{11}^T\tilde{\Sigma}_1\tilde{N}_{11}\cdot c + \tilde{N}_{21}^T\tilde{\Sigma}_2\tilde{N}_{21}\cdot c = -\tilde{C}_1^T\tilde{C}_1, \tag{6.88}$$

$$\tilde{A}_{21}^T\tilde{\Sigma}_2 + \tilde{\Sigma}_1\tilde{A}_{12} + \tilde{N}_{11}^T\tilde{\Sigma}_1\tilde{N}_{12}\cdot c + \tilde{N}_{21}^T\tilde{\Sigma}_2\tilde{N}_{22}\cdot c = -\tilde{C}_1^T\tilde{C}_2. \tag{6.89}$$

We introduce the reduced order system observability Gramian which exists by Theorem 6.21

$$\tilde{A}_{11}^T\tilde{Q}_R + \tilde{Q}_R\tilde{A}_{11} + \tilde{N}_{11}^T\tilde{Q}_R\tilde{N}_{11}\cdot c = -\tilde{C}_1^T\tilde{C}_1. \tag{6.90}$$

Furthermore, we define

$$\tilde{\varepsilon} := \sqrt{\operatorname{tr}(\tilde{C}P\tilde{C}^T) + \operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T) - 2\operatorname{tr}(\tilde{C}\tilde{P}_M\tilde{C}_1^T)}.$$

Due to the duality between the equations (6.83) and (6.85), the identity $\operatorname{tr}(\tilde{C}P\tilde{C}^T) = \operatorname{tr}(\tilde{B}^T\tilde{\Sigma}\tilde{B})$ holds such that

$$\tilde{\varepsilon}^2 = \operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1) + \operatorname{tr}(\tilde{B}_2^T\tilde{\Sigma}_2\tilde{B}_2) + \operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T) - 2\operatorname{tr}(\tilde{C}_1\tilde{P}_{M,1}\tilde{C}_1^T) - 2\operatorname{tr}(\tilde{C}_2\tilde{P}_{M,2}\tilde{C}_1^T), \tag{6.91}$$

where we use the partition $\tilde{P}_M = \begin{bmatrix} \tilde{P}_{M,1} \\ \tilde{P}_{M,2} \end{bmatrix}$. We insert equation (6.89) which yields

$$\begin{aligned} -\operatorname{tr}(\tilde{C}_2\tilde{P}_{M,2}\tilde{C}_1^T) &= -\operatorname{tr}(\tilde{P}_{M,2}\tilde{C}_1^T\tilde{C}_2) = \operatorname{tr}(\tilde{P}_{M,2}(\tilde{A}_{21}^T\tilde{\Sigma}_2 + \tilde{\Sigma}_1\tilde{A}_{12} + \tilde{N}_{11}^T\tilde{\Sigma}_1\tilde{N}_{12}\cdot c + \tilde{N}_{21}^T\tilde{\Sigma}_2\tilde{N}_{22}\cdot c)) \\ &= \operatorname{tr}(\tilde{\Sigma}_2(\tilde{P}_{M,2}\tilde{A}_{21}^T + \tilde{N}_{22}\tilde{P}_{M,2}\tilde{N}_{21}^T\cdot c)) + \operatorname{tr}(\tilde{\Sigma}_1(\tilde{A}_{12}\tilde{P}_{M,2} + \tilde{N}_{12}\tilde{P}_{M,2}\tilde{N}_{11}^T\cdot c)). \end{aligned}$$

By the upper block of equation (6.87), given by

$$\tilde{A}_{11}\tilde{P}_{M,1} + \tilde{A}_{12}\tilde{P}_{M,2} + \tilde{P}_{M,1}\tilde{A}_{11}^T + \tilde{N}_{11}\tilde{P}_{M,1}\tilde{N}_{11}^T\cdot c + \tilde{N}_{12}\tilde{P}_{M,2}\tilde{N}_{11}^T\cdot c = -\tilde{B}_1^T\tilde{B}_1,$$

we obtain

$$\begin{aligned} -\operatorname{tr}(\tilde{C}_2\tilde{P}_{M,2}\tilde{C}_1^T) = &-\operatorname{tr}(\tilde{\Sigma}_1(\tilde{B}_1\tilde{B}_1^T + \tilde{A}_{11}\tilde{P}_{M,1} + \tilde{P}_{M,1}\tilde{A}_{11}^T + \tilde{N}_{11}\tilde{P}_{M,1}\tilde{N}_{11}^T\cdot c)) \\ &+ \operatorname{tr}(\tilde{\Sigma}_2(\tilde{P}_{M,2}\tilde{A}_{21}^T + \tilde{N}_{22}\tilde{P}_{M,2}\tilde{N}_{21}^T\cdot c)). \end{aligned}$$

With equation (6.88), we have

$$\begin{aligned} \operatorname{tr}(\tilde{\Sigma}_1(\tilde{A}_{11}\tilde{P}_{M,1} + \tilde{P}_{M,1}\tilde{A}_{11}^T + \tilde{N}_{11}\tilde{P}_{M,1}\tilde{N}_{11}^T\cdot c)) &= \operatorname{tr}(\tilde{P}_{M,1}(\tilde{\Sigma}_1\tilde{A}_{11} + \tilde{A}_{11}^T\tilde{\Sigma}_1 + \tilde{N}_{11}^T\tilde{\Sigma}_1\tilde{N}_{11}\cdot c)) \\ &= -\operatorname{tr}(\tilde{P}_{M,1}(\tilde{N}_{21}^T\tilde{\Sigma}_2\tilde{N}_{21}\cdot c + \tilde{C}_1^T\tilde{C}_1)), \end{aligned}$$

so that

$$\begin{aligned} -\operatorname{tr}(\tilde{C}_2\tilde{P}_{M,2}\tilde{C}_1^T) = &\operatorname{tr}(\tilde{\Sigma}_2(\tilde{P}_{M,2}\tilde{A}_{21}^T + \tilde{N}_{22}\tilde{P}_{M,2}\tilde{N}_{21}^T\cdot c + \tilde{N}_{21}\tilde{P}_{M,1}\tilde{N}_{21}^T)) \\ &- \operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1) + \operatorname{tr}(\tilde{C}_1\tilde{P}_{M,1}\tilde{C}_1^T). \end{aligned}$$

Inserting this result into equation (6.91) gives

$$\begin{aligned} \tilde{\varepsilon}^2 = &\operatorname{tr}(\tilde{\Sigma}_2(\tilde{B}_2\tilde{B}_2^T + 2\tilde{P}_{M,2}\tilde{A}_{21}^T + 2\tilde{N}_{22}\tilde{P}_{M,2}\tilde{N}_{21}^T\cdot c + 2\tilde{N}_{21}\tilde{P}_{M,1}\tilde{N}_{21}^T\cdot c)) \\ &+ \operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T) - \operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1). \end{aligned}$$

With equations (6.90) and (6.86) we obtain $\operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T) = \operatorname{tr}(\tilde{B}_1^T\tilde{Q}_R\tilde{B}_1)$. So, we have

$$\operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T) - \operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1) = \operatorname{tr}(\tilde{B}_1\tilde{B}_1^T(\tilde{Q}_R - \tilde{\Sigma}_1)).$$

Equation (6.86) again yields

$$\begin{aligned}\operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T)-\operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1) &= -\operatorname{tr}((\tilde{A}_{11}\tilde{P}_R+\tilde{P}_R\tilde{A}_{11}^T+\tilde{N}_{11}\tilde{P}_R\tilde{N}_{11}^T\cdot c)(\tilde{Q}_R-\tilde{\Sigma}_1))\\ &= -\operatorname{tr}(\tilde{P}_R((\tilde{Q}_R-\tilde{\Sigma}_1)\tilde{A}_{11}+\tilde{A}_{11}^T(\tilde{Q}_R-\tilde{\Sigma}_1)+\tilde{N}_{11}^T(\tilde{Q}_R-\tilde{\Sigma}_1)\tilde{N}_{11}\cdot c)).\end{aligned}$$

Below, we subtract equation (6.88) from equation (6.90) and obtain

$$\operatorname{tr}(\tilde{C}_1\tilde{P}_R\tilde{C}_1^T)-\operatorname{tr}(\tilde{B}_1^T\tilde{\Sigma}_1\tilde{B}_1) = -\operatorname{tr}(\tilde{P}_R\tilde{N}_{21}^T\tilde{\Sigma}_2\tilde{N}_{21}\cdot c).$$

Summarizing the result, we have

$$\tilde{\varepsilon}^2 = \operatorname{tr}(\tilde{\Sigma}_2(\tilde{B}_2\tilde{B}_2^T+2\tilde{P}_{M,2}\tilde{A}_{21}^T+2\tilde{N}_{22}\tilde{P}_{M,2}\tilde{N}_{21}^T\cdot c+2\tilde{N}_{21}\tilde{P}_{M,1}\tilde{N}_{21}^T\cdot c-\tilde{N}_{21}\tilde{P}_R\tilde{N}_{21}^T\cdot c)).$$

□

Remark. *At this point, we would like to apply type 2 BT to the heat and the damped wave equation with Lévy noise, presented in Subsection 6.2.5, in order to compare this approach with type 1 BT. In contrast to type 1 BT, we encounter several issues with type 2 BT in terms of the availability of the reachability Gramian.*
As mentioned after equation (6.81), the existence of a positive definite solution to (6.81) is not guaranteed. For that reason, the reachability Gramian for type 2 BT is defined as a positive definite solution to the matrix inequality (6.82).
So far, no algorithm is known to solve the matrix inequality (6.82) or the matrix equality (6.81) (in case a positive definite solution exists). Therefore, we do not provide numerical examples for type 2 BT in this section.

6.4. Singular perturbation approximation for stochastic systems

In Sections 6.2 and 6.3, we discussed two types of balanced truncation (BT) assuming asymptotic stability of the original system. The idea of these reduction concepts is to balance the system such that one creates a system where the dominant reachable and observable states are the same. Afterwards, the difficult to observe and difficult to reach states are truncated, see [2, 53, 54] for the deterministic case.

An alternative method to obtain a reduced order model (ROM) is the singular perturbation approximation (SPA), see Liu, Anderson [48] and Fernando, Nicholson [24] for deterministic linear systems. Again, starting with a balanced system in the sense of type 1 BT, the coefficients of the ROM are modified with respect to type 1 BT. The SPA also exists for bilinear system. For that framework, we refer to Hartmann et al. [28].

In this section, we generalize the work of Liu and Anderson to linear systems with Lévy noise. In Subsection 6.4.1, we motivate the SPA for stochastic systems and derive the ROM which coincides with the deterministic case ROM if $N^k = 0$. Next, in Subsections 6.4.2 and 6.4.3, we analyze the properties of the ROM. First, we consider the stability of the reduced system. We show that it is mean square stable and discuss why the ideas from Benner et al. [11] (see also Subsection 6.2.2) cannot be adopted in order to prove the preservation of mean square asymptotic stability. Additionally, we state the remaining part to complete the proof of mean square asymptotic stability for the ROM. Besides the stability analysis of the ROM, we investigate the reachability and observability in the reduced model resulting from the SPA. With an example we show that one can lose the complete reachability and observability in the ROM even if one starts with an entirely reachable and observable original model which is in contrast to the deterministic case. In Subsection 6.4.4, we assume to have a ROM that preserves the mean square asymptotic stability which is vital for the existence of the error bound we provide in that section. We obtain this error bound by modifying the coefficients of the ROM in order to have the same structure as in the original system. The modified matrices coincide with the ones that are used in the bilinear case by Hartmann et al. [28]. Furthermore, from that error bound, we can point out the cases in which we have a good approximation by the SPA. Finally, in Subsection 6.4.5, we compare BT and the SPA by reducing a large scale system we get from a special discretization of a second order SPDE with Poisson noise. There, we will see that SPA can be better if one considers the underlying equations on a larger time interval. We present a second example, which we generate randomly, to illustrate further advantages of the SPA compared to BT.

6.4.1. Procedure

This subsection delivers an extension of Section 2 in [61], where the case $q = 1$ is studied.

Let $M_1, \dots, M_q$ be uncorrelated, scalar and square integrable Lévy process with mean zero defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$.[2] In addition, we assume M_i $(i = 1, \dots, q)$ to be $(\mathscr{F}_t)_{t\geq 0}$-adapted and the increments $M_i(t+h) - M_i(t)$ to be independent of $\mathscr{F}_t$ for $t, h \geq 0$.

[2] We assume that $(\mathscr{F}_t)_{t\geq 0}$ is right continuous and that $\mathscr{F}_0$ contains all $\mathbb{P}$ null sets.

We consider the following equations that occur in Sections 5.1 and 5.2:

$$\begin{aligned} dx(t) &= [Ax(t)+Bu(t)]dt + \sum_{k=1}^{q} N^k x(t-)dM_k(t), \quad x(0)=x_0\in\mathbb{R}^n, \\ y(t) &= Cx(t), \quad t\geq 0, \end{aligned} \tag{6.92}$$

where A, $N^k\in\mathbb{R}^{n\times n}$, $C\in\mathbb{R}^{p\times n}$, $B\in\mathbb{R}^{n\times m}$ and $x(t-):=\lim_{s\uparrow t}x(s)$. We assume the control $u\in L^2_T$ for every $T>0$. We recall that the solution of (6.92) at time $t\geq 0$ with initial condition $x_0\in\mathbb{R}^n$ and given control u is always denoted by $x(t,x_0,u)$. As in the case of BT, we assume

$$\mathbb{E}\left\|x(t,x_0,0)\right\|_2^2\to 0 \tag{6.93}$$

for $t\to\infty$ and $x_0\in\mathbb{R}^n$ with equivalent conditions stated in Theorem 6.3. This concept of stability is useful to define the (infinite) Gramians P and Q which we assume to be diagonal and equal below, i.e., we consider a balanced system (6.92). In general, a system is of course not of that form but the procedure of how to generate a balanced system is provided in Subsection 6.2.1. The Gramians $P=\Sigma=Q$ are solutions of the generalized Lyapunov equations (6.94) and (6.95)

$$A^T\Sigma+\Sigma A+\sum_{k=1}^{q}(N^k)^T\Sigma N^k\,\mathbb{E}\left[M_k^2(1)\right] = -C^TC, \tag{6.94}$$

$$A\Sigma+\Sigma A^T+\sum_{k=1}^{q}N^k\Sigma(N^k)^T\,\mathbb{E}\left[M_k^2(1)\right] = -BB^T, \tag{6.95}$$

where $\Sigma=\operatorname{diag}(\sigma_1,\ldots,\sigma_n)$ with $\sigma_1\geq\ldots\geq\sigma_n>0$. We introduce the following partitions

$$\Sigma=\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix},\ A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},\ N^k=\begin{bmatrix}N^k_{11} & N^k_{12}\\ N^k_{21} & N^k_{22}\end{bmatrix},\ C=\begin{bmatrix}C_1 & C_2\end{bmatrix} \text{ and } B=\begin{bmatrix}B_1\\ B_2\end{bmatrix},$$

where Σ_1, A_{11}, $N^k_{11}\in\mathbb{R}^{r\times r}$, $C_1\in\mathbb{R}^{p\times r}$ and $B_1\in\mathbb{R}^{r\times m}$. Using the partition $x=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$, the idea of balanced truncation is to select the first r rows in equation (6.92) and to neglect x_2 which means that we set $x_2=0$. This yields a reduced order model with coefficients $(A_{11},N^k_{11},C_1,B_1)$. One can find a detailed motivation regarding BT in the stochastic case in [9] and Section 6.2. From Subsection 6.2.2 we know that balanced truncation preserves asymptotic stability also in the stochastic case if $\sigma_r \neq \sigma_{r+1}$, that is,

$$\sigma\left(I_r \otimes A_{11} + A_{11} \otimes I_r + \sum_{k=1}^{q} N_{11}^k \otimes N_{11}^k \cdot \mathbb{E}\left[M_k^2(1)\right]\right) \subset \mathbb{C}_-. \tag{6.96}$$

The same is true for the truncated part meaning

$$\sigma\left(I_{n-r} \otimes A_{22} + A_{22} \otimes I_{n-r} + \sum_{k=1}^{q} N_{22}^k \otimes N_{22}^k \cdot \mathbb{E}\left[M_k^2(1)\right]\right) \subset \mathbb{C}_-. \tag{6.97}$$

From the properties (6.96) and (6.97) we can also conclude that A_{11} and A_{22} are invertible, see Theorem 6.3.

The method we introduce below is called singular perturbation approximation (SPA) with a more general idea of setting the symbolic derivative $\frac{dx_2}{dt}$ equal to zero instead. We obtain a system

$$\begin{pmatrix} dx_1(t) \\ 0 \end{pmatrix} = \left(\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u(t)\right) dt + \sum_{k=1}^{q} \begin{bmatrix} N_{11}^k & N_{12}^k \\ N_{21}^k & N_{22}^k \end{bmatrix} \begin{pmatrix} x_1(t-) \\ x_2(t-) \end{pmatrix} dM_k(t), \tag{6.98}$$

$$y(t) = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \quad t \geq 0,$$

where we assume $x_0 = 0$ below. From the second line in (6.98), we obtain

$$0 = \int_0^t A_{21}x_1(s) + A_{22}x_2(s) + B_2u(s)ds + \sum_{k=1}^{q} \int_0^t N_{21}^k x_1(s-) + N_{22}^k x_2(s-) dM_k(s), \tag{6.99}$$

such that an Ito integral equals an ordinary integral which is a strange situation, since the ordinary integral is differentiable and the Ito integral is not, in general. We define the process $F(t) = \int_0^t a(s)ds + \sum_{k=1}^q \int_0^t b_k(s)dM_k(s)$, where $a(s) := A_{21}x_1(s) + A_{22}x_2(s) + B_2u(s)$ and $b_k(s) := N_{21}^k x_1(s-) + N_{22}^k x_2(s-)$ and determine the mean of the stochastic differential of $F^T(t)F(t)$, $t \geq 0$. For that reason, we use an Ito product formula stated in Corollary A.4:

$$F^T(t)F(t) = \int_0^t dF^T(s)F(s) + \int_0^t F^T(s)dF(s) + \sum_{i=1}^{n-r} [F_i, F_i]_t,$$

with $[F_i, F_i]_t$ being the quadratic variation part of the i-th component of F. Inserting the differential of F and using the property

$$\mathbb{E}\left[\sum_{i=1}^{n-r}[F_i, F_i]_t\right] = \sum_{k=1}^{q}\int_0^t \mathbb{E}\left[b_k^T(s)b_k(s)\right] ds\, \mathbb{E}\left[M_k^2(1)\right]$$

from Section A.2 yields

$$\mathbb{E}\left[F^T(t)F(t)\right] = \mathbb{E}\left[\int_0^t a^T(s)F(s)ds\right] + \mathbb{E}\left[\int_0^t F^T(s)a(s)ds\right] + \sum_{k=1}^{q}\int_0^t \mathbb{E}\left[b_k^T(s)b_k(s)\right] ds\, \mathbb{E}\left[M_k^2(1)\right].$$

Setting $S \equiv 0$ provides

$$0 = \sum_{k=1}^{q}\int_0^t \mathbb{E}\left[b_k^T(s)b_k(s)\right] ds\, \mathbb{E}\left[M_k^2(1)\right] = \sum_{k=1}^{q}\mathbb{E}\left\|\int_0^t b_k(s)dM_k(s)\right\|_2^2,$$

since the processes $M_1, \dots, M_q$ are uncorrelated. This implies $\int_0^t b_k(s)dM_k(s) = 0$ $\mathbb{P}$-a.s. for every $k = 1, \dots, q$. If we apply this to equation (6.99), we get

$$x_2(t) = -(A_{22}^{-1}A_{21}x_1(t) + A_{22}^{-1}B_2u(t)). \tag{6.100}$$

By inserting this in the first line in equation (6.98) we have

$$x_1(t) = \int_0^t \bar{A}x_1(s) + \bar{B}u(s)ds + \sum_{k=1}^{q}\int_0^t \bar{N}^k x_1(s-) + \bar{B}_0^k u(s-)dM_k(s) \tag{6.101}$$

and

$$\bar{y}(t) = \bar{C}x_1(t) + \bar{D}u(t),$$

where

$\bar{A} = A_{11} - A_{12}A_{22}^{-1}A_{21}$, $\bar{B} = B_1 - A_{12}A_{22}^{-1}B_2$, $\bar{N}^k = N_{11}^k - N_{12}^k A_{22}^{-1}A_{21}$, $\bar{C} = C_1 - C_2A_{22}^{-1}A_{21}$, $\bar{B}_0^k = -N_{12}^k A_{22}^{-1}B_2$ and $\bar{D} = -C_2A_{22}^{-1}B_2$.

Remark. *(i) The SPA yields a reduced order model (6.101) which has a different structure than the original model (6.92), meaning that we obtained a system in which the output equation is controlled and the control in the state equation is disturbed by Lévy noise. If we* *use this ROM, we have to restrict ourselves to controls with existing left limits $u(t-)$, $t \geq 0$, in order for equation (6.101) to be well defined. Since we prefer for a ROM to have the same shape as the original model we will often emphasize the case $(\bar{B}, \bar{B}_0^k, \bar{D}) = (B_1, 0, 0)$ which we get by setting $B_2 = 0$ in equation (6.100).*

(ii) If we set $(\bar{B}, \bar{B}_0^k, \bar{D}) = (B_1, 0, 0)$, we precisely obtain the matrices that are recommended for the SPA in the bilinear case in [28].

6.4.2. Preservation of (asymptotic) mean square stability

Below, we discuss the results of Section 3.1 in [61] for more than one noise term.
In this subsection, we state the first steps how to prove the asymptotic mean square stability of the ROM by the SPA. Unfortunately, this proof is not complete but our conjecture is that this property is preserved.
We multiply A^{-T} from the left and A^{-1} from the right in equation (6.94) and get

$$\Sigma\tilde{A} + \tilde{A}^T\Sigma + \sum_{k=1}^{q} (\tilde{N}^k)^T \Sigma \tilde{N}^k \mathbb{E}\left[M_k^2(1)\right] = -\tilde{C}^T\tilde{C}, \tag{6.102}$$

where $\tilde{A} = A^{-1}$, $\tilde{N}^k = N^k A^{-1}$ and $\tilde{C} = CA^{-1}$. It can be shown in a straightforward manner that using these transformed coefficients $\tilde{A}$ and $\tilde{N}^k$ instead of A and N^k does not effect the asymptotic mean square stability. By equation (6.95), the corresponding dual equation is

$$(A\Sigma A^T)\tilde{A}^T + \tilde{A}(A\Sigma A^T) + \sum_{k=1}^{q} \tilde{N}^k (A\Sigma A^T) (\tilde{N}^k)^T \mathbb{E}\left[M_k^2(1)\right] = -BB^T. \tag{6.103}$$

The reason to consider the matrices $\tilde{A}$ and $\tilde{N}^k$ is the following equivalence between its left upper blocks and the reduced order model coefficients:

$$\sigma\left(I_r \otimes \tilde{A}_{11} + \tilde{A}_{11} \otimes I_r + \sum_{k=1}^{q} \tilde{N}_{11}^k \otimes \tilde{N}_{11}^k \cdot c_k\right) \subset \mathbb{C}_- \Leftrightarrow \sigma\left(I_r \otimes \bar{A} + \bar{A} \otimes I_r + \sum_{k=1}^{q} \bar{N}^k \otimes \bar{N}^k \cdot c_k\right) \subset \mathbb{C}_-$$

with $c_k = \mathbb{E}\left[M_k^2(1)\right]$. Since one can show that

$$\tilde{A} = \begin{bmatrix} \bar{A}^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -A_{22}^{-1}A_{21}\bar{A}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix},$$

we have $\tilde{A}_{11} = \bar{A}^{-1}$, $\tilde{N}_{11}^k = \bar{N}^k \bar{A}^{-1}$. So, proving asymptotic mean square stability in the ROM is now transformed into the following problem:
Starting with a system with coefficients $\tilde{A}$ and $\tilde{N}^k$, show that this property is preserved if one truncates the system, i.e. one chooses the reduced order coefficients $\tilde{A}_{11}$ and $\tilde{N}_{11}^k$.
The main difficulty is the fact that this system is not balanced since the solution of equation (6.103) is neither diagonal nor does it coincide with the one from equation (6.102). For that reason, the ideas that are used for the stability analysis of BT in [11] (see Sections 4.3 – 4.5) cannot be applied. In the deterministic case, where $N^k = 0$, the dual equation (6.103) is obtained by pre- and post-multiplying equation (6.95) with A^{-1} and A^{-T} which in that case yields a balanced system, see [48]. Unfortunately, this does not work in the more general framework $N^k \neq 0$ because we would get $A^{-1}N^k$ instead of the desired matrix $\tilde{N}^k = N^k A^{-1}$. We could state the desired result then under the assumption that A, N^k commute, which would at least partially prove the conjecture.

Since the solution of equation (6.102) is in diagonal form, we can adopt at least a few arguments from [11] which we state in the proof of the lemma below.

Lemma 6.24. *The reduced order models with the coefficients* $(\tilde{A}_{11}, \tilde{N}_{11}^k)$ *or* $(\bar{A}, \bar{N}^k)$ *are mean square stable, i.e.*

$$\sigma\left(I_r \otimes \tilde{A}_{11} + \tilde{A}_{11} \otimes I_r + \sum_{k=1}^{q} \tilde{N}_{11}^k \otimes \tilde{N}_{11}^k \cdot c_k\right) \subset \overline{\mathbb{C}_-} \tag{6.104}$$

and

$$\sigma\left(I_r \otimes \bar{A} + \bar{A} \otimes I_r + \sum_{k=1}^{q} \bar{N}^k \otimes \bar{N}^k \cdot c_k\right) \subset \overline{\mathbb{C}_-},$$

where $c_k = \mathbb{E}\left[M_k^2(1)\right]$.

Proof. We use a suitable partition of $\tilde{A}$, $\tilde{N}^k$, $\tilde{C}$, Σ and obtain the following equation for the left upper block of (6.102):

$$\Sigma_1 \tilde{A}_{11} + \tilde{A}_{11}^T \Sigma_1 + \sum_{k=1}^{q} (\tilde{N}_{11}^k)^T \Sigma_1 \tilde{N}_{11}^k \cdot c_k = -\tilde{C}_1^T \tilde{C}_1 - \sum_{k=1}^{q} (\tilde{N}_{21}^k)^T \Sigma_2 \tilde{N}_{21}^k \cdot c_k, \tag{6.105}$$

with $\tilde{N}_{21}^k = (N_{21}^k - N_{22}^k A_{22}^{-1} A_{21})\bar{A}^{-1}$ and $\tilde{C}_1 = \bar{C}\bar{A}^{-1}$. Consequently, by Corollary 3.2 in [11], we obtain property (6.104). With the same argument, it also holds that

$$\sigma\left(I_r\otimes\bar{A}+\bar{A}\otimes I_r+\sum_{k=1}^{q}\bar{N}^k\otimes\bar{N}^k\cdot\mathbb{E}\left[M_k^2(1)\right]\right)\subset\overline{\mathbb{C}_-},$$

since by pre- and post-multiplying (6.105) with $\bar{A}^T$ and $\bar{A}$, we get

$$\bar{A}^T\Sigma_1+\Sigma_1\bar{A}+\sum_{k=1}^{q}(\bar{N}^k)^T\Sigma_1\bar{N}^k\cdot c_k=-\bar{C}^T\bar{C}-\sum_{k=1}^{q}(N_{21}^k-N_{22}^kA_{22}^{-1}A_{21})^T\Sigma_2(N_{21}^k-N_{22}^kA_{22}^{-1}A_{21})\cdot c_k$$

and hence the result holds. □

Using Theorem 3.1 in [11], we obtain

$$\alpha(\bar{K}):=\max\left\{\Re(\lambda):\ \lambda\in\sigma(\bar{K})\right\}\in\sigma(\bar{K})$$

with $\bar{K}=I_r\otimes\tilde{A}_{11}+\tilde{A}_{11}\otimes I_r+\sum_{k=1}^{q}\tilde{N}_{11}^k\otimes\tilde{N}_{11}^kc_k$. By (6.104) it remains to show that $0\notin\sigma(\bar{K})$ to get the desired asymptotic mean square stability. We summarize this as follows:

Conjecture 6.25. *The reduced order model with coefficients $(\bar{A},\bar{N}^k)$ is asymptotically mean square stable, i.e. $0\notin\sigma(\bar{K})$.*

The result in Conjecture 6.25 is theoretically important for the existence of the error bound we state in Subsection 6.4.4. Practically, it is easy to check if zero is an eigenvalue of $\bar{K}$ or not since the reduced order dimension r is usually small.

6.4.3. Other properties in the ROM

In this subsection, we point out that starting with a completely observable and reachable original system one can lose these properties in the ROM. Furthermore, with the help of an example, we discuss that the ROM with modified coefficients is not balanced. In addition, we briefly show that the reduced system is not balanced which is in contrast to the deterministic case. Here, we just emphasize the case, where $q=1$. Moreover, we set $M=M_1$, $N=N^1$, $\bar{N}=\bar{N}^1$ and $\bar{B}_0=\bar{B}_0^1$. This part of the thesis uses results of Section 3.2 in [61] and contributes a further discussion about the structure of the ROM Gramians and Hankel singular values in Example 6.29.

We recall that the fundamental solution of the state equation (6.92) is an $\mathbb{R}^{n\times n}$-valued process Φ

satisfying

$$d\Phi(t) = A\Phi(t)dt + N\Phi(t-)dM(t), \quad \Phi(0) = I_n \quad t \geq 0.$$

Summarizing the facts that are shown in Section 6.1, we recall the observability Gramian $Q := \int_0^\infty \mathbb{E}\left[\Phi^T(s)C^T C\Phi(s)\right] ds$ and the reachability Gramian $P = \int_0^\infty \mathbb{E}\left[\Phi(s)BB^T\Phi^T(s)\right] ds$ which exist by assumption (6.93). Q and P solve equations (6.94) and (6.95), respectively, which is proven in Section 6.1. Here, we are in a balanced situation which means that

$$P = Q = \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n).$$

We know that system (6.92) is completely observable if and only if the Gramian Q is positive definite. Since the reachability concept for system (6.92), used in Section 6.1, neglects the information that is contained in N, it is not surprising that P can only provide partial information about the reachability of a state $x \in \mathbb{R}^n$. To be more precise, if x is reachable, then $x \in \operatorname{im} P$ but the other direction is not true. So, it is necessary to introduce the deterministic Gramian $P_D = \int_0^\infty \mathrm{e}^{At} BB^T \mathrm{e}^{A^T t} dt$. Again, using the results in Section 6.1, we know that system (6.92) is completely reachable if and only if $P_D > 0$. This is analogous to the deterministic case, for which the results are stated in [2].

Since the ROM (6.101) has a different structure than the original model, one might think that the Gramian of the ROM has to be defined differently in order to characterize observability and reachability of the system. We will see soon that the additional matrices $\bar{B}_0$ and $\bar{D}$ have no impact in that context. In order to discuss this property we repeat the concepts of observability and reachability of the ROM:

$$dx_1(t) = [\bar{A}x_1(t) + \bar{B}u(t)]dt + [\bar{N}x_1(t-) + \bar{B}_0 u(t-)]dM(t), \quad x_1(0) = \bar{x}_0, \tag{6.106}$$

$$\bar{y}(t) = \bar{C}x_1(t) + \bar{D}u(t). \tag{6.107}$$

Since the observability concept is considered in the uncontrolled case ($u \equiv 0$), the matrix $\bar{D}$ does not enter in the following definition.

Definition 6.26. *An initial state $\bar{x}_0$ is called* observable *if the corresponding observation energy is positive:*

$$\left\|\bar{C}x_1(\cdot, \bar{x}_0, 0)\right\|_{L^2}^2 := \mathbb{E}\int_0^\infty \left\|\bar{C}x_1(t, \bar{x}_0, 0)\right\|_2^2 dt > 0.$$

Since we have $\bar{C}x_1(t,\bar{x}_0,0) = \bar{C}\bar{\Phi}(t)\bar{x}_0$, $t \geq 0$, it follows that

$$\left\|\bar{C}x_1(\cdot,\bar{x}_0,0)\right\|_{L^2}^2 = \bar{x}_0^T \bar{Q}_R x_0$$

with $\bar{Q}_R := \mathbb{E}\int_0^\infty \bar{\Phi}^T(t)\bar{C}^T\bar{C}\bar{\Phi}(t)dt$. Here, $\bar{\Phi}$ denotes the fundamental solution of the ROM. Hence, the ROM is completely reachable if and only if $\bar{Q}_R > 0$. Below, we distinguish between the solution of (6.106) for general $\bar{B}_0$ which we denote by $x_1(t,\bar{x}_0,u)$ and the solution of (6.106) in case $\bar{B}_0 = 0$ which we denote by $x_1^0(t,\bar{x}_0,u)$, $t \geq 0$. Now, we define reachable average states.

Definition 6.27. *A state $\bar{x}$ is called* reachable on average (from zero) *if there is a time $T > 0$ and a control function $u \in L_T^2$, such that we have*

$$\mathbb{E}\left[x_1(T,0,u)\right] = \bar{x}.$$

Applying the expectation on both sides of equation (6.106) and using the property that the Ito integral has mean zero yields that the functions $\mathbb{E}\left[x_1(t,\bar{x}_0,u)\right]$ and $\mathbb{E}\left[x_1^0(t,\bar{x}_0,u)\right]$, $t \geq 0$, are both solutions of the ODE

$$\dot{x}(t) = \bar{A}x(t) + \bar{B}\mathbb{E}[u(t)], \quad x(0) = \bar{x}_0, \quad t \geq 0.$$

Hence, both expected values coincide, such that the matrix $\bar{B}_0$ can be completely neglected in the reachability concept. Setting $\bar{B}_0 = 0$ provides a system of the same form as the original model (6.92). Consequently, we know that the ROM (6.106) is completely reachable if and only if $\bar{P}_{D,R} := \int_0^\infty \mathrm{e}^{\bar{A}t}\bar{B}\bar{B}^T\mathrm{e}^{\bar{A}^T t}dt > 0$. The next example shows that starting with a completely observable and completely reachable system does not mean that the ROM has these properties as well.

Example 6.28. *We define a system (6.92) with $\mathbb{E}\left[M^2(1)\right] = 1$ and coefficients*

$$(A,B,C,N) = \left(\begin{pmatrix} -\frac{17}{2} & 8 & 8 \\ -8 & -20 & -20 \\ -8 & -20 & -\frac{41}{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 4 & 4 \\ -4 & 2 & 2 \\ -4 & 2 & 2 \end{pmatrix}\right),$$

which is asymptotically mean square stable. In addition, we have a balanced system since for the solution of the equations (6.94) and (6.95) it holds that $P = Q = \operatorname{diag}(2,1,1)$. Consequently, it is *also completely observable. The complete reachability follows from $P_D > 0$. The corresponding one dimensional ROM has the coefficients*

$$(\bar{A}, \bar{B}, \bar{B}_0, \bar{C}, \bar{D}, \bar{N}) = (-\frac{117}{10}, 0, 0, 0, 2, -\frac{3}{5}).$$

Since there is no control in the state equation of the ROM and the output of the uncontrolled ROM is identically zero, the reduced order system is neither completely reachable nor completely observable. Of course, this also holds for the modified ROM, where one sets $(\bar{B}, \bar{B}_0, \bar{D}) := (B_1, 0, 0) = (0, 0, 0)$.

The fact that reachability and observability are not necessarily preserved by the SPA is not surprising since analogous observations are made for BT in Subsection 6.2.3 or in [14], respectively.

We conclude this subsection by stating an example which shows that the reduced order model is not balanced for the case $(\bar{B}, \bar{B}_0, \bar{D}) := (B_1, 0, 0)$. Moreover, the Hankel singular values of the reduced system do not coincide with those of the original system which is in contrast to the deterministic case, see [48]. This fact is again not surprising since we made the same observation in Subsection 6.2.3 for BT.
We illustrate these properties with an example below. Here, $\bar{P}_R := \mathbb{E} \int_0^\infty \bar{\Phi}(t) \bar{B}_1 B_1^T \bar{\Phi}^T(t) dt$ is considered to be the reachability Gramian of the reduced system.

Example 6.29. *We consider the case, where M is a Wiener process which we denote by w:*

$$\begin{aligned} dx(t) &= [Ax(t) + Bu(t)]dt + Nx(t)dw(t), \\ y(t) &= Cx(t). \end{aligned}$$

The following matrices (up to the digits shown) provide a balanced and asymptotically mean square stable system:

$$A = \begin{pmatrix} -5 & 2 & 2 \\ 2 & -10 & -4 \\ -2 & -4 & -20 \end{pmatrix}, \quad B = C^T = \begin{pmatrix} 4.89898 & 0 & 0 \\ -3.26599 & 4.83046 & 0 \\ 0.40825 & 1.51814 & 5.61503 \end{pmatrix}, \quad N = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The Gramians are given by

$$P = Q = \Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The reduced order model $(r = 2)$ *has the coefficients*

$$\bar{A} = \left(\begin{smallmatrix} -4.8 & 2.4 \\ 2.4 & -9.2 \end{smallmatrix}\right), \quad B_1 = \left(\begin{smallmatrix} 4.89898 & 0 & 0 \\ -3.26599 & 4.83046 & 0 \end{smallmatrix}\right), \quad \bar{C} = \begin{pmatrix} 4.85815 & -3.34764 \\ -0.15181 & 4.52683 \\ -0.56150 & -1.12301 \end{pmatrix}, \quad \bar{N} = \left(\begin{smallmatrix} 0.9 & 0.8 \\ 0.9 & 0.8 \end{smallmatrix}\right)$$

and is asymptotically mean square stable. It has the following Gramians:

$$\bar{P}_R = \begin{pmatrix} 2.835288 & -0.059676 \\ -0.059676 & 2.022759 \end{pmatrix} \quad \textit{and} \quad \bar{Q}_R = \begin{pmatrix} 2.80743 & -0.12660 \\ -0.12660 & 1.91368 \end{pmatrix}.$$

The Hankel singular values of the reduced order model are 2.8312 *and* 1.9570 *which are different from the Hankel singular values of the original system.*

6.4.4. Error bound for the SPA

We transfer results from Section 4 in [61] to this subsection which we state for a general setting. In this subsection, we provide an error bound for the case $(\bar{B}, \bar{B}_0, \bar{D}) = (B_1, 0, 0)$ and $x_0 = 0$. In the error bound below, the matrix $\bar{P}_R := \mathbb{E} \int_0^\infty \bar{\Phi}(t) B_1 B_1^T \bar{\Phi}^T(t) dt$ enters. For its existence we assume that the mean square asymptotic stability is preserved in the ROM. This means that

$$0 \notin \sigma \left(I_r \otimes \bar{A} + \bar{A} \otimes I_r + \sum_{k=1}^{q} \bar{N}^k \otimes \bar{N}^k \cdot \mathbb{E}\left[M_k^2(1)\right] \right), \tag{6.108}$$

which is a consequence of the discussion in Subsection 6.4.2. Condition (6.108) is usually easy to check since the reduced order dimension r is small.

Following the arguments in Section 4.2 in [14] or Subsection 6.2.4 in this work, respectively, the error of the SPA is bounded as follows:

$$\sup_{t \in [0,T]} \mathbb{E} \left\| y(t) - \bar{y}(t) \right\|_2 \leq \left(\operatorname{tr}\left(C \Sigma C^T \right) + \operatorname{tr}\left(\bar{C} \bar{P}_R \bar{C}^T \right) - 2 \operatorname{tr}\left(C \bar{P}_G \bar{C}^T \right) \right)^{\frac{1}{2}} \|u\|_{L_T^2}^2, \tag{6.109}$$

where

$$\begin{aligned} AP_G + P_G\bar{A}^T + \sum_{k=1}^{q} N^k P_G(\bar{N}^k)^T \mathbb{E}\left[M_k^2(1)\right] &= -BB_1^T, \\ \bar{A}P_R + P_R\bar{A}^T + \sum_{k=1}^{q} \bar{N}^k P_R(\bar{N}^k)^T \mathbb{E}\left[M_k^2(1)\right] &= -B_1B_1^T. \end{aligned} \tag{6.110}$$

Below, we specify this bound to emphasize the cases in which the SPA performs well.

Theorem 6.30. *If the ROM is asymptotically mean square stable, then*

$$\begin{aligned} &\operatorname{tr}\left(C\Sigma C^T\right) + \operatorname{tr}\left(\bar{C}\bar{P}_R\bar{C}^T\right) - 2\ \operatorname{tr}\left(C\bar{P}_G\bar{C}^T\right) \\ &= \operatorname{tr}(2\Sigma_2(\sum_{k=1}^{q}(N_{22}^k\bar{P}_{G,2} + N_{21}^k\bar{P}_{G,1})(N_{21}^k - N_{22}^kA_{22}^{-1}A_{21})^T c_k - (A_{22}\bar{P}_{G,2} + A_{21}\bar{P}_{G,1})(A_{22}^{-1}A_{21})^T)) \\ &\quad + \operatorname{tr}(\Sigma_2(B_2B_2^T - \sum_{k=1}^{q}(N_{21}^k - N_{22}^kA_{22}^{-1}A_{21})\bar{P}_R(N_{21}^k - N_{22}^kA_{22}^{-1}A_{21})^T c_k)), \end{aligned}$$

where $\bar{P}_{G,1}$ are the first r rows of $\bar{P}_G$, $\bar{P}_{G,2}$ are the last $n-r$ rows of $\bar{P}_G$, $c_k = \mathbb{E}\left[M_k^2(1)\right]$ and $\Sigma_2 = \operatorname{diag}(\sigma_{r+1}, \ldots, \sigma_n)$ is the matrix of truncated Hankel singular values.

Proof. To simplify the notation in the proof, we set $q = 1$, $c_1 = c$ and $N^1 = N$. Below, it is easy to see that this simplification does not cause a loss of generality. The right lower block of (6.94) satisfies

$$A_{22}^T\Sigma_2 + \Sigma_2A_{22} + N_{22}^T\Sigma_2N_{22}c + N_{12}^T\Sigma_1N_{12}c = -C_2^TC_2. \tag{6.111}$$

If we multiply (6.94) with A^{-T} from the left and select the left and right upper block of this equation, we obtain

$$\begin{aligned} \Sigma_1 + \bar{A}^{-T}(\Sigma_1A_{11} - A_{21}^TA_{22}^{-T}\Sigma_2A_{21} + \bar{N}^T\Sigma_1N_{11}c + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{21}c) &= -\bar{A}^{-T}\bar{C}^TC_1, \\ \bar{A}^{-T}(\Sigma_1A_{12} - A_{21}^TA_{22}^{-T}\Sigma_2A_{22} + \bar{N}^T\Sigma_1N_{12}c + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{22}c) &= -\bar{A}^{-T}\bar{C}^TC_2 \end{aligned}$$

and thus

$$\bar{A}^T\Sigma_1 + \Sigma_1A_{11} - A_{21}^TA_{22}^{-T}\Sigma_2A_{21} + \bar{N}^T\Sigma_1N_{11}c + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{21}c = -\bar{C}^TC_1, \tag{6.112}$$

$$\Sigma_1A_{12} - A_{21}^TA_{22}^{-T}\Sigma_2A_{22} + \bar{N}^T\Sigma_1N_{12}c + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{22}c = -\bar{C}^TC_2. \tag{6.113}$$

Furthermore, using (6.95) one can conclude

$$A_{11}\Sigma_1+\Sigma_1 A_{11}^T+N_{11}\Sigma_1 N_{11}^T c+N_{12}\Sigma_2 N_{12}^T c=-B_1B_1^T \tag{6.114}$$

and

$$A_{22}\Sigma_2+\Sigma_2 A_{22}^T+N_{22}\Sigma_2 N_{22}^T c+N_{21}\Sigma_1 N_{21}^T c=-B_2B_2^T. \tag{6.115}$$

From the partition of (6.110)

$$\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\bar{P}_{G,1}\\ \bar{P}_{G,2}\end{bmatrix}+\begin{bmatrix}\bar{P}_{G,1}\\ \bar{P}_{G,2}\end{bmatrix}\bar{A}^T+\begin{bmatrix}N_{11} & N_{12}\\ N_{21} & N_{22}\end{bmatrix}\begin{bmatrix}\bar{P}_{G,1}\\ \bar{P}_{G,2}\end{bmatrix}\bar{N}^T c=-\begin{bmatrix}B_1B_1^T\\ B_2B_1^T\end{bmatrix}$$

we also know that

$$A_{11}\bar{P}_{G,1}+A_{12}\bar{P}_{G,2}+\bar{P}_{G,1}\bar{A}^T+N_{11}\bar{P}_{G,1}\bar{N}^T c+N_{12}\bar{P}_{G,2}\bar{N}^T c=-B_1B_1^T, \tag{6.116}$$

$$A_{21}\bar{P}_{G,1}+A_{22}\bar{P}_{G,2}+\bar{P}_{G,2}\bar{A}^T+N_{22}\bar{P}_{G,2}\bar{N}^T c+N_{21}\bar{P}_{G,1}\bar{N}^T c=-B_2B_1^T. \tag{6.117}$$

We define $\mathscr{E}:=\left(\operatorname{tr}\left(C\Sigma C^T\right)+\operatorname{tr}\left(\bar{C}\bar{P}_R\bar{C}^T\right)-2\operatorname{tr}\left(C\bar{P}_G\bar{C}^T\right)\right)^{\frac{1}{2}}$ and obtain

$$\begin{aligned}\mathscr{E}^2&=\operatorname{tr}\left(\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix}\begin{bmatrix}C_1^T\\ C_2^T\end{bmatrix}\right)+\operatorname{tr}\left(\bar{C}\bar{P}_R\bar{C}^T\right)-2\operatorname{tr}\left(\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}\bar{P}_{G,1}\\ \bar{P}_{G,2}\end{bmatrix}\bar{C}^T\right)\\ &=\operatorname{tr}(C_2\Sigma_2C_2^T+C_1\Sigma_1C_1^T+\bar{C}\bar{P}_R\bar{C}^T-2C_1\bar{P}_{G,1}\bar{C}^T-2C_2\bar{P}_{G,2}\bar{C}^T).\end{aligned}$$

Using equation (6.113) yields

$$\begin{aligned}&\operatorname{tr}(-C_2\bar{P}_{G,2}\bar{C}^T)=\operatorname{tr}(-\bar{C}^TC_2\bar{P}_{G,2})\\ &=\operatorname{tr}(\Sigma_1A_{12}\bar{P}_{G,2}-A_{21}^TA_{22}^{-T}\Sigma_2A_{22}\bar{P}_{G,2}+\bar{N}^T\Sigma_1N_{12}\bar{P}_{G,2}c+(N_{21}-N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{22}\bar{P}_{G,2}c)\\ &=\operatorname{tr}(A_{12}\bar{P}_{G,2}\Sigma_1-A_{21}^TA_{22}^{-T}\Sigma_2A_{22}\bar{P}_{G,2}+N_{12}\bar{P}_{G,2}\bar{N}^T\Sigma_1c+(N_{21}-N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{22}\bar{P}_{G,2}c).\end{aligned}$$

By equation (6.116) we obtain

$$\begin{aligned}\operatorname{tr}(-C_2\bar{P}_{G,2}\bar{C}^T)=&\operatorname{tr}(-A_{21}^TA_{22}^{-T}\Sigma_2A_{22}\bar{P}_{G,2}+(N_{21}-N_{22}A_{22}^{-1}A_{21})^T\Sigma_2N_{22}\bar{P}_{G,2}c)\\ &-\operatorname{tr}((B_1B_1^T+\bar{P}_{G,1}\bar{A}^T+A_{11}\bar{P}_{G,1}+N_{11}\bar{P}_{G,1}\bar{N}^Tc)\Sigma_1).\end{aligned}$$

Using equation (6.112), we have

$$\begin{aligned}
&\operatorname{tr}(\bar{P}_{G,1}\bar{A}^T + A_{11}\bar{P}_{G,1} + N_{11}\bar{P}_{G,1}\bar{N}^T c)\Sigma_1) = \operatorname{tr}(\bar{A}^T\Sigma_1 + \Sigma_1 A_{11} + \bar{N}^T\Sigma_1 N_{11}c)\bar{P}_{G,1}) \\
&= -\operatorname{tr}(\bar{C}^T C_1\bar{P}_{G,1} + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2 N_{21}\bar{P}_{G,1}c - (A_{22}^{-1}A_{21})^T\Sigma_2 A_{21}\bar{P}_{G,1})
\end{aligned}$$

and hence,

$$\begin{aligned}
\mathscr{E}^2 = &\operatorname{tr}(C_2\Sigma_2 C_2^T + C_1\Sigma_1 C_1^T + \bar{C}\bar{P}_R\bar{C}^T) - 2\operatorname{tr}(B_1B_1^T\Sigma_1) \\
&+ 2\operatorname{tr}(-(A_{22}^{-1}A_{21})^T\Sigma_2 A_{22}\bar{P}_{G,2} + (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2 N_{22}\bar{P}_{G,2}c) \\
&+ 2\operatorname{tr}((N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2 N_{21}\bar{P}_{G,1}c - (A_{22}^{-1}A_{21})^T\Sigma_2 A_{21}\bar{P}_{G,1}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathscr{E}^2 = &\operatorname{tr}(\Sigma_2(C_2^T C_2 - 2(A_{22}\bar{P}_{G,2} + A_{21}\bar{P}_{G,1})(A_{22}^{-1}A_{21})^T) \\
&+ \operatorname{tr}(2\Sigma_2(N_{22}\bar{P}_{G,2} + N_{21}\bar{P}_{G,1})(N_{21} - N_{22}A_{22}^{-1}A_{21})^T c)) \\
&+ \operatorname{tr}(C_1\Sigma_1 C_1^T + \bar{C}\bar{P}_R\bar{C}^T - 2B_1B_1^T\Sigma_1).
\end{aligned} \tag{6.118}$$

By definition, the Gramians $\bar{P}_R$ and $\bar{Q}_R$ satisfy

$$\bar{A}^T\bar{Q}_R + \bar{Q}_R\bar{A} + \bar{N}^T\bar{Q}_R\bar{N}c = -\bar{C}^T\bar{C} \tag{6.119}$$

and

$$\bar{A}\bar{P}_R + \bar{P}_R\bar{A}^T + \bar{N}\bar{P}_R\bar{N}^T c = -B_1B_1^T, \tag{6.120}$$

such that one can conclude $\operatorname{tr}(\bar{C}\bar{P}_R\bar{C}^T) = \operatorname{tr}(B_1^T\bar{Q}_R B_1)$ from inserting (6.119) into $\operatorname{tr}(\bar{C}\bar{P}_R\bar{C}^T)$. Consequently,

$$\operatorname{tr}(C_1\Sigma_1 C_1^T + \bar{C}\bar{P}_R\bar{C}^T - 2B_1B_1^T\Sigma_1) = \operatorname{tr}(C_1\Sigma_1 C_1^T - B_1B_1^T\Sigma_1) + \operatorname{tr}(B_1^T(\bar{Q}_R - \Sigma_1)B_1).$$

Inserting equation (6.114) gives

$$\begin{aligned}\operatorname{tr}(-B_1B_1^T\Sigma_1) &= \operatorname{tr}(A_{11}\Sigma_1\Sigma_1 + \Sigma_1A_{11}^T\Sigma_1 + N_{11}\Sigma_1N_{11}^Tc\Sigma_1 + N_{12}\Sigma_2N_{12}^Tc\Sigma_1)\\ &= \operatorname{tr}(\Sigma_1\Sigma_1A_{11} + \Sigma_1A_{11}^T\Sigma_1 + \Sigma_1N_{11}^T\Sigma_1N_{11}c + N_{12}\Sigma_2N_{12}^T\Sigma_1c)\\ &= -\operatorname{tr}(\Sigma_1C_1^TC_1) - \operatorname{tr}(\Sigma_1N_{21}^T\Sigma_2N_{21}c) + \operatorname{tr}(N_{12}\Sigma_2N_{12}^T\Sigma_1c)\end{aligned}$$

and therefore

$$\operatorname{tr}(C_1\Sigma_1C_1^T - B_1B_1^T\Sigma_1) = \operatorname{tr}(N_{12}\Sigma_2N_{12}^T\Sigma_1c) - \operatorname{tr}(\Sigma_1N_{21}^T\Sigma_2N_{21}c)$$

holds. From (6.111) it follows that

$$\begin{aligned}\operatorname{tr}(\Sigma_2N_{12}^T\Sigma_1N_{12}c) &= \operatorname{tr}(-\Sigma_2(A_{22}^T\Sigma_2 + \Sigma_2A_{22} + N_{22}^T\Sigma_2N_{22}c + C_2^TC_2))\\ &= \operatorname{tr}(-\Sigma_2(\Sigma_2A_{22}^T + A_{22}\Sigma_2 + N_{22}\Sigma_2N_{22}^Tc + C_2^TC_2)).\end{aligned}$$

Using (6.115) then yields

$$\operatorname{tr}(\Sigma_2N_{12}^T\Sigma_1N_{12}c) = \operatorname{tr}(\Sigma_2(N_{21}\Sigma_1N_{21}^Tc + B_2B_2^T - C_2^TC_2)),$$

such that

$$\operatorname{tr}(C_1\Sigma_1C_1^T - B_1B_1^T\Sigma_1) = \operatorname{tr}(\Sigma_2(B_2B_2^T - C_2^TC_2)).$$

Below, we analyze the term $\operatorname{tr}(B_1^T(\bar{Q}_R - \Sigma_1)B_1)$. First, notice that the following holds:

$$\bar{A}^T\Sigma_1 + \Sigma_1\bar{A} + \bar{N}^T\Sigma_1\bar{N}c = -\bar{C}^T\bar{C} - (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2(N_{21} - N_{22}A_{22}^{-1}A_{21})c.$$

With (6.119) we thus know that

$$\begin{aligned}&\bar{A}^T(\bar{Q}_R - \Sigma_1) + (\bar{Q}_R - \Sigma_1)\bar{A} + \bar{N}^T(\bar{Q}_R - \Sigma_1)\bar{N}c\\ &= (N_{21} - N_{22}A_{22}^{-1}A_{21})^T\Sigma_2(N_{21} - N_{22}A_{22}^{-1}A_{21})c.\end{aligned} \tag{6.121}$$

Applying the equations (6.120) and (6.121) yields

$$\begin{aligned}\operatorname{tr}(B_1^T(\bar{Q}_R-\Sigma_1)B_1) &= -\operatorname{tr}((\bar{A}\bar{P}_R+\bar{P}_R\bar{A}^T+\bar{N}\bar{P}_R\bar{N}^T c)(\bar{Q}_R-\Sigma_1)) \\ &= -\operatorname{tr}(\bar{P}_R((\bar{Q}_R-\Sigma_1)\bar{A}+\bar{A}^T(\bar{Q}_R-\Sigma_1)+\bar{N}^T(\bar{Q}_R-\Sigma_1)\bar{N}c)) \\ &= -\operatorname{tr}(\bar{P}_R(N_{21}-N_{22}A_{22}^{-1}A_{21})^T\Sigma_2(N_{21}-N_{22}A_{22}^{-1}A_{21})c).\end{aligned}$$

We apply these results to (6.118) and obtain

$$\begin{aligned}\mathscr{E}^2 =& \operatorname{tr}(2\Sigma_2((N_{22}\bar{P}_{G,2}+N_{21}\bar{P}_{G,1})(N_{21}-N_{22}A_{22}^{-1}A_{21})^T c-(A_{22}\bar{P}_{G,2}+A_{21}\bar{P}_{G,1})(A_{22}^{-1}A_{21})^T)) \\ &+\operatorname{tr}(\Sigma_2(B_2B_2^T-(N_{21}-N_{22}A_{22}^{-1}A_{21})\bar{P}_R(N_{21}-N_{22}A_{22}^{-1}A_{21})^T c))\end{aligned}$$

which gives the result of this theorem and concludes the proof. □

The error bound representation in Theorem 6.30 depends on Σ_2 which contains the $n-r$ smallest Hankel singular values $\sigma_{r+1},\dots,\sigma_n$ of the original system. In case these values are small, the error bound indicates that the reduced order model obtained by the SPA is of good quality.

6.4.5. Numerical experiments

In this subsection, we compare BT which is discussed in Section 6.2 (see also [9, 14]) and the SPA which we consider above. The aim is to point out the cases, when the SPA is better in order to motivate the practical relevance of this method. We start with an example which we obtain by discretizing an SPDE in the spatial component and afterwards we state a randomly generated example to illustrate further effects. Both examples are not in the balanced form already but balancing these systems can be done easily by the procedure stated in Subsection 6.2.1 and [14]. The numerical experiments are run on a desktop computer with a dual-core Intel Pentium processor E5400 and 3GB RAM. All algorithms are implemented and executed in MATLAB 7.14.0.739 (R2012a) running on Ubuntu 10.04.1 LTS.
The numerical results are taken from Section 5.1 in [61] and extended by output plots showing the performance of the SPA.

SPDE example

To compare BT and the SPA we use an example created in [60]. There, a second order SPDE with Lévy noise is considered and approximated by a large scale system of ordinary SDEs. The same example can also be found in Subsection 4.3.2 with the corresponding approximation in Section 5.2.

We apply balancing related model order reduction to the discretized version of the following SPDE we introduced in Example 4.16. As we mention there, it models the lateral displacement of an electricity cable impacted by wind:

$$\frac{\partial^2}{\partial t^2}\mathscr{X}(t,\zeta)+2\frac{\partial}{\partial t}\mathscr{X}(t,\zeta)=\frac{\partial^2}{\partial \zeta^2}\mathscr{X}(t,\zeta)+\mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}u(t)+2\,\mathrm{e}^{-(\zeta-\frac{\pi}{2})^2}\mathscr{X}(t-,\zeta)\frac{\partial}{\partial t}M(t)$$

for $t,\zeta\in[0,\pi]$ and the damping factor $\alpha=2$. Here, we choose a particular $M(t)=-(N(t)-t)$ with $(N(t))_{t\geq 0}$ being a Poisson process with parameter 1. The boundary and initial conditions are

$$\mathscr{X}(0,t)=0=\mathscr{X}(\pi,t)\quad\text{and}\quad \mathscr{X}(0,\zeta),\,\frac{\partial}{\partial t}\mathscr{X}(t,\zeta)\bigg|_{t=0}\equiv 0.$$

The output is an approximation for the position of the middle of the string, that is,

$$\mathscr{Y}(t)=\frac{1}{2\varepsilon}\int_{\frac{\pi}{2}-\varepsilon}^{\frac{\pi}{2}+\varepsilon}\mathscr{X}(t,\zeta)d\zeta,$$

where $\varepsilon>0$.

The structure of the approximating SDE with state space dimension n is stated at the end of Section 5.2. Below, we specify the matrix of the approximating SDE:

$$dx(t)=[Ax(t)+Bu(t)]\,dt+Nx(s-)dM(s),\quad y_n(t)=Cx(t),\quad t\geq 0,\tag{6.122}$$

where

- the initial condition is $x_0=0$,
- $A=\operatorname{diag}\left(E_1,\dots,E_{\frac{n}{2}}\right)$ with $E_\ell=\left(\begin{smallmatrix}0 & \ell\\ -\ell & -2\end{smallmatrix}\right)$,
- $B=(b_k)_{k=1,\dots,n}$ with

$$b_{2\ell-1}=0,\quad b_{2\ell}=\sqrt{\frac{2}{\pi}}\left\langle \mathrm{e}^{-(\cdot-\frac{\pi}{2})^2},\sin(\ell\cdot)\right\rangle_H,$$

- $N = \left(n_{kj}\right)_{k,j=1,\dots,n}$ with

$$n_{(2\ell-1)j} = 0, \quad n_{(2\ell)j} = \begin{cases} 0, & \text{if } j = 2v, \\ \frac{4}{\pi v} \left\langle \sin(\ell \cdot), \mathrm{e}^{-(\cdot - \frac{\pi}{2})^2} \sin(v \cdot) \right\rangle_H, & \text{if } j = 2v - 1, \end{cases}$$

  for $j = 1, \dots, n$ and $v = 1, \dots, \frac{n}{2}$,

- the output matrix C is given by $C^T = (c_k)_{k=1,\dots,n}$ with

$$c_{2\ell} = 0 \text{ and } c_{2\ell-1} = \frac{1}{\sqrt{2\pi}\ell^2 \varepsilon} \left[\cos\left(\ell\left(\frac{\pi}{2} - \varepsilon\right)\right) - \cos\left(\ell\left(\frac{\pi}{2} + \varepsilon\right)\right)\right],$$

where we assume n to be even, $\ell = 1, \dots, \frac{n}{2}$ and $H = L^2([0, \pi])$. Following the arguments in Section 5.2 this approximation is meaningful since

$$\mathbb{E} \left|y_n(t) - \mathscr{Y}(t)\right|^2 \to 0$$

for $n \to \infty$ and $t \geq 0$. Now, we fix the dimension of (6.122) to $n = 1000$. We check numerically that there is a positive definite solution $X > 0$ to

$$A^T X + XA + N^T XN = -I.$$

This is equivalent to asymptotic mean square stability due to Theorem 6.3 (see also [60]), that is

$$\mathbb{E} \left\|x(t, x_0, 0)\right\|_2^2 \to 0$$

for $t \to \infty$ and any initial condition. So, we can apply balanced truncation and SPA, respectively below. We consider the deviation between y_n and the outputs of the ROMs via BT and via the SPA in the norm on the left hand side of (6.109). We insert particular normalized control functions $u_1(t) = \sqrt{\frac{2}{\pi}} 1_{[0, \frac{\pi}{2}]}(t)$ and $u_2(t) = \frac{\sqrt{8}}{\pi} 1_{[0, \frac{\pi}{2}]}(t) w(t)$ ($t \in [0, \pi]$), where w is a Wiener process. We take the exact errors and the error bound $\mathscr{E}_1$ (see Theorem 6.18) of BT from [60] and we additionally determine these values for the SPA, where $\mathscr{E}_2$ denotes the corresponding error bound stated in Theorem 6.30. Furthermore, we set $(\bar{B}, \bar{B}_0, \bar{D}) = (B_1, 0, 0)$. We obtain the exact errors by discretizing the equations with the Euler-Maruyama scheme that is discussed in [31, 32, 41].

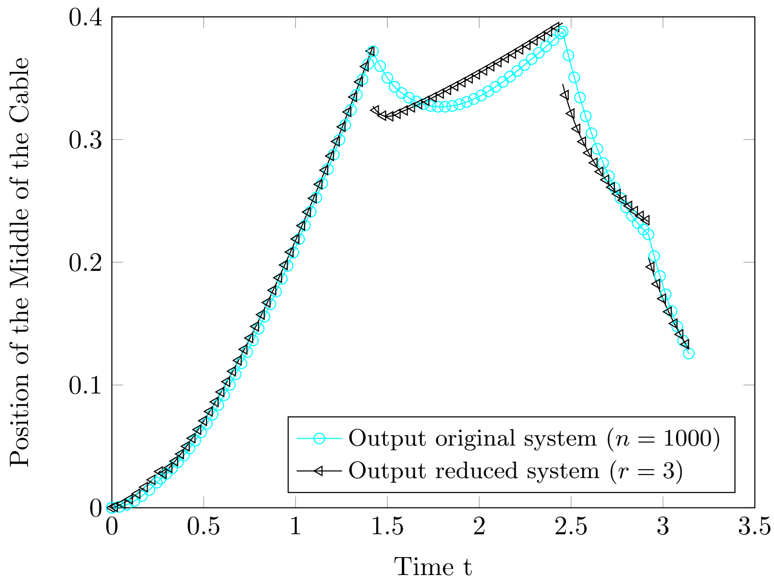
Dim. ROM	BT Exact Error $(u = u_1)$	BT Exact Error $(u = u_2)$	Error Bound $\mathscr{E}_1$
40	$1.9992 \cdot 10^{-7}$	$2.0024 \cdot 10^{-7}$	$4.0103 \cdot 10^{-5}$
20	$4.4660 \cdot 10^{-6}$	$4.1435 \cdot 10^{-6}$	$1.2695 \cdot 10^{-4}$
10	$4.3081 \cdot 10^{-5}$	$3.2512 \cdot 10^{-5}$	$3.6395 \cdot 10^{-4}$
5	$5.1180 \cdot 10^{-4}$	$4.2176 \cdot 10^{-4}$	$2.3446 \cdot 10^{-3}$
3	0.0114	$8.2309 \cdot 10^{-3}$	0.0380

Dim. ROM	SPA Exact Error $(u = u_1)$	SPA Exact Error $(u = u_2)$	Error Bound $\mathscr{E}_2$
40	$2.5543 \cdot 10^{-6}$	$2.3302 \cdot 10^{-6}$	$4.1799 \cdot 10^{-5}$
20	$1.1382 \cdot 10^{-5}$	$7.1896 \cdot 10^{-6}$	$1.2808 \cdot 10^{-4}$
10	$4.7307 \cdot 10^{-5}$	$3.3501 \cdot 10^{-5}$	$3.4039 \cdot 10^{-4}$
5	$6.2538 \cdot 10^{-4}$	$5.0726 \cdot 10^{-4}$	$2.3876 \cdot 10^{-3}$
3	0.0168	0.0124	0.0629

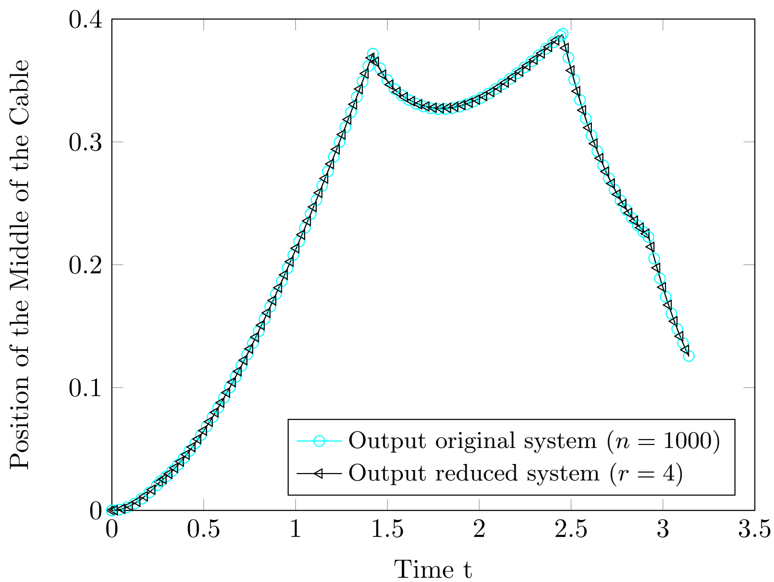
The above tables show the exact errors and the error bounds for BT and SPA ROMs of dimensions $r = 3, 5, 10, 20, 40$. From these numerical results we see that BT is slightly better than the SPA on a time interval $[0, \pi]$ in terms of the actual errors (and, in most cases, also for the error bounds).

When we observe the trajectories for $u \equiv u_1$ in Figure 6.3, we can see that the SPA does not well estimate the jumps if the reduced order dimension is $r = 3$ which is in contrast to BT, compare Figure 6.2a. If $r = 4$, the ROM by SPA is quite accurate what is shown in Figure 6.3b. So, there is no visible difference to BT, see Figure 6.2b.

We are also interested in the long run behavior of the system (6.122). Therefore, we increase the length of the time interval and consider (6.122) on $[0, 8.5\pi]$ next and repeat the procedure. This is done due to the expected zero steady-state error that is known for the deterministic case. Again, we use normalized controls $\tilde{u}_1(t) = \sqrt{\frac{2}{8.5\pi}} 1_{[0, \frac{8.5\pi}{2}]}(t)$ and $\tilde{u}_2(t) = \frac{\sqrt{8}}{8.5\pi} 1_{[0, \frac{8.5\pi}{2}]}(t) w(t)$ $(t \in [0, 8.5\pi])$ and obtain better results for the SPA for growing dimensions of the ROM in the table



(a) Reduced order dimension $r = 3$ and $u \equiv u_1$



(b) Reduced order dimension $r = 4$ and $u \equiv u_1$

Figure 6.3.: Output of the original model compared with the output of the ROM by SPA on the time interval $[0, \pi]$

below. There, we compare the ROMs of dimensions $r = 3, 5, 6, 10, 20$ obtained by BT and SPA.

Dim. ROM	Error SPA $(u = \tilde{u}_1)$	Error BT $(u = \tilde{u}_1)$	Error SPA $(u = \tilde{u}_2)$	Error BT $(u = \tilde{u}_2)$
20	$3.0098 \cdot 10^{-6}$	$4.2581 \cdot 10^{-6}$	$3.1226 \cdot 10^{-6}$	$3.6249 \cdot 10^{-6}$
10	$1.6609 \cdot 10^{-5}$	$2.3312 \cdot 10^{-5}$	$1.5577 \cdot 10^{-5}$	$2.0886 \cdot 10^{-5}$
6	$5.5366 \cdot 10^{-5}$	$7.4563 \cdot 10^{-5}$	$4.9357 \cdot 10^{-5}$	$6.5983 \cdot 10^{-5}$
5	$5.7379 \cdot 10^{-4}$	$5.2218 \cdot 10^{-4}$	$5.1004 \cdot 10^{-4}$	$4.7717 \cdot 10^{-4}$
3	0.0158	0.0101	0.0142	$9.0102 \cdot 10^{-3}$

Below, we would like to compare different outputs visually. Since the reduced order models of BT and the SPA are quite accurate it is not possible to distinguish between the trajectories. For that reason and to conclude this subsection, we create a random example.

Randomly generated example

Here, we consider an example of the form (6.122) with Wiener noise. We generate this by using the MATLAB commands rand$(\cdot,\cdot)$ and randn$(\cdot,\cdot)$ which provide matrices of uniformly and normally distributed random numbers, respectively.

We set the state space dimension of the original model to $n = 500$, the reduced order system dimension to $r = 2$ and

$$A = JDJ^{-1}$$

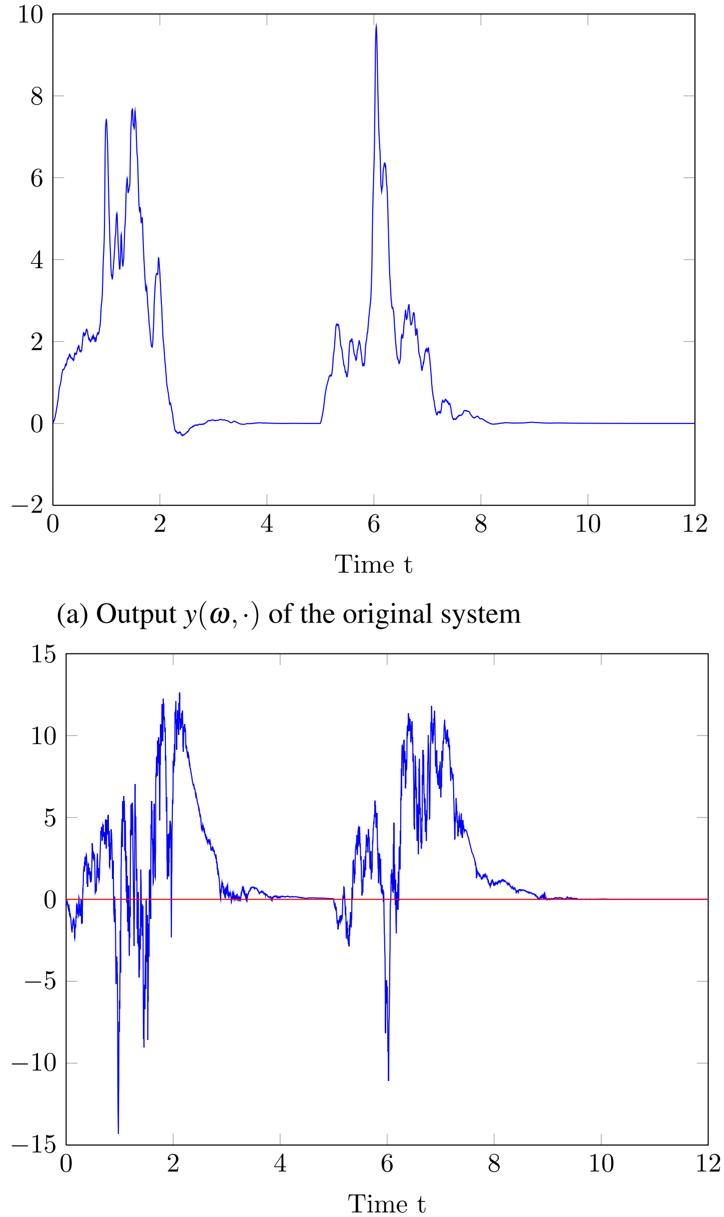
with

$$D = -\operatorname{diag}(10\operatorname{abs}(\operatorname{randn}(n,1))) - 2I_n \text{ and } J = \operatorname{randn}(n),$$

where we use “randn(’state’,1)” for D and ”randn(’state’,2)“ for J. The matrices B, C, N are also random and generated by

$$B = \operatorname{randn}(n,n),\, C = \operatorname{randn}(1,n) \text{ and } N = \operatorname{rand}(n)/100,$$

where we use ”rand(’state’,1)” for N, ”rand(’state’,3)“ for B and “rand(’state’,4)” for C. One can



(a) Output $y(\omega,\cdot)$ of the original system

(b) $|y(\omega,\cdot)-y_{BT}(\omega,\cdot)|-|y(\omega,\cdot)-y_{SPA}(\omega,\cdot)|$ with y_{BT} and y_{SPA} being the outputs of ROMs ($r=2$) by BT and SPA, respectively

Figure 6.4.: Comparison of BT and SPA using a particular trajectory

check numerically that there is a positive definite solution X to

$$A^T X + XA + N^T XN = -I.$$

By Theorem 6.3 this means that the system is asymptotically mean square stable. We insert the controls u_i ($i = 1, \dots, n$) on $[0, 12]$

$$u_i(t) = \begin{cases} k_i & \text{if } t \in [0,2] \cup [5,7] \\ 0 & \text{else,} \end{cases}$$

where the k_i are randomly generated constants.
In Figure 6.4a we visualize a trajectory of the output of the original model and in Figure 6.4b we compare the point-wise error of BT with the point-wise error of the SPA for a particular trajectory and $r = 2$. If the graph in Figure 6.4b is above zero (red line), then the SPA is better. From the two plots we observe that the SPA is a better approximation if the output curve is flat. In this case, it seems to be a good assumption to suppose certain state components to be constant (symbolic derivative $\frac{dX_2}{dt} = 0$, see (6.98)), whereas BT provides a smaller error, where the slope of the output is big.

6.5. Conclusions

In this chapter, we introduced reachability and observability concepts for linear control systems with Lévy noise. Moreover, we used generalized Gramians to characterize the degree of reachability and observability of a state in a stochastic system. Based on certain Gramians, we considered both type 1 and type 2 balanced truncation model order reduction for stochastic systems, provided stability results and theoretical error bounds as well as a comparison between both types of balanced truncation. Furthermore, we introduced the singular perturbation approximation as an alternative model order reduction technique for stochastic systems, derived theoretical error bounds and provided a detailed discussion within the stability analysis. We also discussed numerical examples to support the theory, e.g., we applied balanced truncation and the singular perturbation approximation to a stochastic heat equation and a damped wave equation. To this end, it could be shown that model order reduction using balanced truncation and the singular perturbation approximation performs well in the context of reducing the state space dimension of particular semi-discretized SPDEs, that is, we obtained small exact errors and tight error bounds.

7. Conclusions

7.1. Summary

In this thesis, we stated the most important results regarding the stochastic analysis of Lévy processes. We studied Lévy processes taking values in Hilbert spaces in Chapter 2, discussed important properties and introduced several representatives for this class of processes. Based on the Lévy-Khinchin decomposition, stated in Chapter 2, we defined a stochastic integral with respect to Hilbert space valued Lévy processes and pointed out the mean, the second moment and the martingale property of the integral in Chapter 3. These results form the basis for stochastic partial differential equations (SPDEs) with Lévy noise. Controlled linear SPDEs in an abstract evolution equation setting were introduced in Chapter 4. In particular, we investigated stochastic heat and damped wave equations with Lévy noise which we approximated by a Galerkin scheme in Chapter 5. We also showed the convergence of the Galerkin solution to the mild solution of the corresponding SPDE. The discretization led to numerical solutions of high-dimensional linear ordinary SDEs. These large scale systems provided a motivation to generalize balancing related model order reduction, known from the deterministic setting, for linear controlled stochastic equations which are mean square asymptotically stable. The key idea was to use modified reachability and observability concepts which we investigated for linear systems with Lévy noise. Based on these new concepts we established balancing related model order reduction techniques for linear controlled systems with Lévy noise, that is, balanced truncation (BT) and the singular perturbation approximation (SPA) to stochastic systems, see Chapter 6. In a balanced setting, the reduced order model from BT is obtained by setting all truncated states equal to zero. In the SPA however, truncated states are assumed constant which allows to solve for them and thus include this information in the differential equation for the remaining states. This has the advantage of a zero steady-state error which is often important in applications. We explained the procedure of BT and moreover proved an error bound for BT, showed the preservation of mean square asymptotic stability, discussed further properties such as the loss of reachability and observability in the reduced order model as well as the structure of the reduced order model Gramians and Hankel singular values. In order to illustrate the good quality of BT for stochastic systems we furthermore run simulations on the heat and damped wave equation with Lévy noise. For the SPA approach similar properties can be observed and proven, but the preservation of mean square asymptotic stability could so far not be shown. In contrast to BT we only stated a proof for the mean square stability of the reduced order model. Using numerical experiments on the stochastic damped wave equation, we pointed out the cases in which we have a good approximation by the SPA. It further turned out that BT is more accurate than the SPA for small time intervals and that SPA can perform better if a larger time interval or small controls are considered for the underlying equations.

7.2. Outlook

In this thesis new aspects within the field of model order reduction for linear stochastic control systems were revealed. In particular, we used a Galerkin technique to discretize an SPDE in space and we treated the resulting large scale systems with generalized balancing related model order techniques but there are still several open questions and different approaches that are worth considering and discussing in the context of model order reduction for stochastic systems.

The Galerkin method is restricted to SPDEs with simple domains. In order to make model order reduction applicable to a wider class of SPDEs, it would be interesting to investigate and generalize further approximation schemes for SPDEs, such as finite element methods, in order to be able to handle highly complex domains and even more complex SPDEs.

Besides the already extended techniques such as BT and SPA, our idea is to introduce further model order reduction techniques in a stochastic setting that are for example based on a different choice of Gramians or which are not balancing related. We see the main problem in generalizing other schemes from the deterministic linear framework in the absence of transfer functions. Many methods in the deterministic case are based on these transfer functions which unfortunately do not exist for systems driven by Lévy noise. Hence, the most challenging part would be to find an alternative concept in the stochastic case.

BT and the SPA are methods which are restricted to mean square asymptotically stable systems. This can be a natural assumption but as soon as the noise is too “large”, the stability assumption does not hold anymore. For that reason, alternative methods should be considered which allow the reduction of stochastic systems that are unstable. A first step is to consider systems which are asymptotically stable without a noise component (i.e. the matrix A is Hurwitz) but then become unstable because of an additional noise term which is too "large". As a next step one could think of model order reduction methods, where the underlying SDEs have a coefficient matrix A which is not asymptotically stable.

Model order reduction is a discipline which originates in the field of deterministic control theory and some methods were extended to controlled stochastic settings in this thesis. We think that it is also an interesting question to consider non controlled stochastic settings since many realizations are needed to derive moments of the output such that the output has to be evaluated several times. Replacing the control term by additive noise would be a setting one can think of.

Further issues arise when nonlinear systems are considered. So far our methods are restricted to linear stochastic systems. In order to reduce nonlinear equations such as the semi-discretized stochastic Navier-Stokes equation one has to think about more advanced approaches.

So far every numerical example we considered in this thesis had no space dependent noise. The reason is that large computational costs we encounter to obtain a reduced order model for semi-discretized SPDEs with space dependent noise. To obtain a reduced order model from BT or SPA, Lyapunov equations need to be solved. The more noise terms we have, the more expensive it is to solve these equations. Semi-discretized versions of SPDEs with space dependent noise, would have a number of noise terms that is equal to the state space dimension. It is therefore interesting to develop and analyze methods to approximate infinite dimensional noise by low dimensional noise processes such that we not just reduce systems in the state dimension but also in the dimension of the noise.

A. Appendix

This appendix contains material of Section 2 in Benner, Redmann [14].
Let all stochastic processes appearing in this section be defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$[1]. We denote the set of all cadlag[2] square integrable $\mathbb{R}$-valued martingales with respect to $(\mathscr{F}_t)_{t\geq 0}$ by $\mathscr{M}^2(\mathbb{R})$.

A.1. Semimartingales and Ito's formula in $\mathbb{R}^d$

Below, we introduce the class of semimartingales.

Definition A.1. *(i) An $(\mathscr{F}_t)_{t\geq 0}$-adapted cadlag process X with values in $\mathbb{R}$ is called* semimartingale *if it has the representation $X = X_0 + M + A$. Here, X_0 is an $\mathscr{F}_0$-measurable random variable, $M \in \mathscr{M}^2(\mathbb{R})$ and A is a cadlag process of bounded variation.*[3]

(ii) An $\mathbb{R}^d$-valued process $\vec{X}$ is called semimartingale *if all components are real-valued semimartingales.*

The following is based on Proposition 17.2 in [52].

Proposition A.2. *Let $M, N \in \mathscr{M}^2(\mathbb{R})$, then there exists a unique predictable process*[4] *$\langle M, N\rangle$ of bounded variation such that $MN - \langle M, N\rangle$ is a martingale with respect to $(\mathscr{F}_t)_{t\geq 0}$.*

Next, we consider a decomposition of square integrable martingales (see Theorem 4.18 in [39]).

Theorem A.3. *A process $M \in \mathscr{M}^2(\mathbb{R})$ has the following representation:*

$$M(t) = M_0 + M^c(t) + M^d(t), \quad t \geq 0,$$

[1] $(\mathscr{F}_t)_{t\geq 0}$ shall be right continuous and complete.
[2] Cadlag means that $\mathbb{P}$-almost all paths are right continuous and the left limits exist.
[3] This means that $\mathbb{P}$-almost all paths are of bounded variation.
[4] We define the predictable σ-algebra in (3.5).

where $M^c(0) = M^d(0) = 0$, M_0 is an $\mathscr{F}_0$-measurable random variable, M^c is a continuous process in $\mathscr{M}^2(\mathbb{R})$ and $M^d \in \mathscr{M}^2(\mathbb{R})$.

We need the quadratic covariation $[Z_1, Z_2]$ of two real-valued semimartingales Z_1 and Z_2, which can be introduced by

$$[Z_1, Z_2]_t := Z_1(t)Z_2(t) - Z_1(0)Z_2(0) - \int_0^t Z_1(s-)dZ_2(s) - \int_0^t Z_2(s-)dZ_1(s) \tag{A.1}$$

for $t \geq 0$. By the linearity of the integrals in (A.1) we obtain the property

$$[Z_1, Z_2]_t = \frac{1}{2}\left([Z_1 + Z_2, Z_1 + Z_2]_t - [Z_1, Z_1]_t - [Z_2, Z_2]_t\right), \quad t \geq 0.$$

From Theorem 4.52 in [39], we know that $[Z_1, Z_2]$ is also given by

$$[Z_1, Z_2]_t = \langle M_1^c, M_2^c \rangle_t + \sum_{0 \leq s \leq t} \Delta Z_1(s) \Delta Z_2(s) \tag{A.2}$$

for $t \geq 0$, where M_1^c and M_2^c are the continuous martingale parts of Z_1 and Z_2. Furthermore, we set $\Delta Z(s) := Z(s) - Z(s-)$ with $Z(s-) := \lim_{t \uparrow s} Z(t)$ for a real-valued semimartingale Z. If we rearrange equation (A.1), we obtain the Ito product formula

$$Z_1(t)Z_2(t) = Z_1(0)Z_2(0) + \int_0^t Z_1(s-)dZ_2(s) + \int_0^t Z_2(s-)dZ_1(s) + [Z_1, Z_2]_t \tag{A.3}$$

for $t \geq 0$, which we use for the following corollaries:

Corollary A.4. *Let Y and Z be two $\mathbb{R}^d$-valued semimartingales, then*

$$Y^T(t)Z(t) = Y^T(0)Z(0) + \int_0^t Z^T(s-)dY(s) + \int_0^t Y^T(s-)dZ(s) + \sum_{i=1}^d [Y_i, Z_i]_t$$

for all $t \geq 0$.

Proof. We have

$$\begin{aligned} Y^T(t)Z(t) &= \sum_{i=1}^d Y_i(t)Z_i(t) = \sum_{i=1}^d \left(Y_i(0)Z_i(0) + \int_0^t Z_i(s-)dY_i(s) + \int_0^t Y_i(s-)dZ_i(s) + [Y_i, Z_i]_t \right) \\ &= Y^T(0)Z(0) + \int_0^t Z^T(s-)dY(s) + \int_0^t Y^T(s-)dZ(s) + \sum_{i=1}^d [Y_i, Z_i]_t \end{aligned}$$

by applying the product formula in (A.3). □

Corollary A.5. *Let Y be an $\mathbb{R}^d$-valued and Z be an $\mathbb{R}^n$-valued semimartingale, then*

$$Y(t)Z^T(t) = Y(0)Z^T(0) + \int_0^t dY(s)Z^T(s-) + \int_0^t Y(s-)dZ^T(s) + \left([Y_i,Z_j]_t\right)_{\substack{i=1,\dots,d\\ j=1,\dots,n}}$$

for all $t \geq 0$.

Proof. We consider the stochastic differential of the ij-th component of the matrix-valued process $Y(t)Z^T(t)$, $t \geq 0$, and obtain the following via the product formula in (A.3):

$$\begin{aligned} e_i^T Y(t)Z^T(t)e_j &= e_i^T Y(0)Z^T(0)e_j + \int_0^t Z^T(s-)e_j d(e_i^T Y(s)) + \int_0^t e_i^T Y(s-)d(Z^T(s)e_j) \\ &\quad + [e_i^T Y, Z^T e_j]_t \\ &= e_i^T Y(0)Z^T(0)e_j + e_i^T \int_0^t d(Y(s))Z^T(s-)e_j + e_i^T \int_0^t Y(s-)d(Z^T(s))e_j + [Y_i,Z_j]_t \end{aligned}$$

for all $t \geq 0$, $i \in \{1,\dots,d\}$, and $j \in \{1,\dots,n\}$, where e_i is the i-th unit vector in $\mathbb{R}^d$ or in $\mathbb{R}^n$, respectively. Hence, in compact form we have

$$Y(t)Z^T(t) = Y(0)Z^T(0) + \int_0^t dY(s)Z^T(s-) + \int_0^t Y(s-)dZ^T(s) + \left([Y_i,Z_j]_t\right)_{\substack{i=1,\dots,d\\ j=1,\dots,n}}$$

for all $t \geq 0$. □

A.2. Lévy-type integrals in $\mathbb{R}$

Below, we want to determine the mean of the quadratic covariation of the following Lévy-type integrals:

$$\begin{aligned} \tilde{Z}_1(t) &= \tilde{Z}_1(0) + \int_0^t A_1(s)ds + \sum_{i=1}^q \int_0^t B_1^i(s)dM^i(s), \quad t \geq 0, \\ \tilde{Z}_2(t) &= \tilde{Z}_2(0) + \int_0^t A_2(s)ds + \sum_{i=1}^q \int_0^t B_2^i(s)dM^i(s), \quad t \geq 0, \end{aligned}$$

where the processes M^i ($i = 1,\dots,q$) are uncorrelated scalar square integrable Lévy processes with mean zero. In addition, the processes B_1^i, B_2^i are integrable with respect to M^i ($i = 1,\dots,q$), which by Section 3.1 means that they are predictable with

$$\mathbb{E}\int_0^t \left|B^i(s)\right|^2 ds < \infty, \quad t \geq 0.$$

Furthermore, A_1 and A_2 are $\mathbb{P}$-almost surely Lebesgue integrable and $(\mathscr{F}_t)_{t\geq 0}$-adapted.

We set $b_1(t) := \sum_{i=1}^q \int_0^t B_1^i(s)dM^i(s)$ and $b_2(t) := \sum_{i=1}^q \int_0^t B_2^i(s)dM^i(s)$ and obtain

$$\left[\tilde{Z}_1, \tilde{Z}_2\right]_t = [b_1, b_2]_t$$

for $t \geq 0$ considering equation (A.2), because $\tilde{Z}_i$ has the same jumps and the same martingale part as b_i $(i = 1,2)$. We know that

$$[b_1, b_2]_t = \frac{1}{2}\left([b_1 + b_2, b_1 + b_2]_t - [b_1, b_1]_t - [b_2, b_2]_t\right) \tag{A.4}$$

for $t \geq 0$. Using the definition in (A.1) yields

$$[b_1, b_1]_t = (b_1(t))^2 - 2\int_0^t b_1(s-)db_1(s) = (b_1(t))^2 - 2\sum_{i=1}^q \int_0^t b_1(s-)B_1^i(s)dM^i(s).$$

Thus,

$$\mathbb{E}\,[b_1, b_1]_t = \mathbb{E}\left[(b_1(t))^2\right].$$

Since M^i and M^j are uncorrelated processes for $i \neq j$, we get

$$\mathbb{E}\left[(b_1(t))^2\right] = \sum_{i=1}^q \mathbb{E}\left[\left(\int_0^t B_1^i(s)dM^i(s)\right)^2\right] = \sum_{i=1}^q \int_0^t \mathbb{E}\left[\left(B_1^i(s)\right)^2\right] ds \cdot c_i$$

by applying the Ito isometry proven in Section 3.1, where $c_i := \mathbb{E}\left[\left(M^i(1)\right)^2\right]$. Hence,

$$\mathbb{E}\,[b_1, b_1]_t = \sum_{i=1}^q \int_0^t \mathbb{E}\left[\left(B_1^i(s)\right)^2\right] ds \cdot c_i.$$

Analogously, we can show that

$$\mathbb{E}\left[b_2, b_2\right]_t = \sum_{i=1}^{q} \int_0^t \mathbb{E}\left[\left(B_2^i(s)\right)^2\right] ds \cdot c_i \quad \text{and}$$
$$\mathbb{E}\left[b_1 + b_2, b_1 + b_2\right]_t = \sum_{i=1}^{q} \int_0^t \mathbb{E}\left[\left(B_1^i + B_2^i(s)\right)^2\right] ds \cdot c_i$$

hold for $t \geq 0$. Considering equation (A.4), we obtain

$$\mathbb{E}\left[\tilde{Z}_1, \tilde{Z}_2\right]_t = \mathbb{E}\left[b_1, b_2\right]_t = \sum_{i=1}^{q} \int_0^t \mathbb{E}\left[B_1^i B_2^i(s)\right] ds \cdot c_i. \tag{A.5}$$

At the end of this section, we refer to Subsection 4.4.3 in Applebaum [5]. There, one can find some remarks regarding the quadratic covariation of the Lévy-type integrals defined in that book.

Bibliography

[1] Sergio Albeverio and Barbara Rüdiger. Stochastic Integrals and the Lévy-Itô Decomposition Theorem on Separable Banach Spaces. *Stochastic Anal. Appl.*, 23(2):217–253, 2005.

[2] Athanasios C. Antoulas. *Approximation of large-scale dynamical systems.* Advances in Design and Control 6. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2005.

[3] David Applebaum. Martingale-valued Measures, Ornstein-Uhlenbeck Processes with Jumps and Operator Self-Decomposability in Hilbert Space. Séminaire de probabilités XXXIX. Berlin: Springer. Lecture Notes in Mathematics 1874, 171-196 (2006)., 2006.

[4] David Applebaum. Lévy Processes and Stochastic Integrals in Banach Spaces. 2007.

[5] David Applebaum. *Lévy Processes and Stochastic Calculus. 2nd ed.* Cambridge Studies in Advanced Mathematics 116. Cambridge: Cambridge University Press, 2009.

[6] Ludwig Arnold. *Stochastische Differentialgleichungen. Theorie und Anwendungen.* München-Wien: R. Oldenbourg Verlag, 1973.

[7] Andrea Barth. A finite element method for martingale-driven stochastic partial differential equations. *Commun. Stoch. Anal.*, 4(3):355–375, 2010.

[8] Andrea Barth and Annika Lang. Simulation of stochastic partial differential equations using finite element methods. *Stochastics*, 84(2-3):217–231, 2012.

[9] Peter Benner and Tobias Damm. Lyapunov equations, energy functionals, and model order reduction of bilinear and stochastic systems. *SIAM J. Control Optim.*, 49(2):686–711, 2011.

[10] Peter Benner, Tobias Damm, and Yolanda Rocio Rodriguez Cruz. Dual pairs of generalized Lyapunov inequalities and balanced truncation of stochastic linear systems. *IEEE Trans. Autom. Contr.*, accepted, 2016.

[11] Peter Benner, Tobias Damm, Martin Redmann, and Yolanda Rocio Rodriguez Cruz. Positive Operators and Stable Truncation. *Linear Algebra and its Applications*, 498:74–87, 2016.

[12] Peter Benner, Volker Mehrmann, and Danny C. Sorensen (eds.). *Dimension reduction of large-scale systems. Proceedings of a workshop, Oberwolfach, Germany, October 19–25, 2003.* Lecture Notes in Computational Science and Engineering 45. Berlin: Springer, 2005.

[13] Peter Benner and Martin Redmann. Reachability and Observability Concepts for Stochastic Systems. *Proc. Appl. Math. Mech.*, 13:381–382, 2013.

[14] Peter Benner and Martin Redmann. Model Reduction for Stochastic Systems. *Stoch PDE: Anal Comp*, 3(3):291–338, 2015.

[15] Jean Bertoin. *Lévy Processes.* Cambridge Tracts in Mathematics. 121. Cambridge: Cambridge Univ. Press, 1998.

[16] Wolf-Jürgen Beyn, Elena Isaak, and Raphael Kruse. Stochastic C-Stability and B-Consistency of Explicit and Implicit Euler-Type Schemes. *Journal of Scientific Computing*, pages 1–33, 2015.

[17] Dirk Blömker and Arnulf Jentzen. Galerkin approximations for the stochastic Burgers equation. *SIAM J. Numer. Anal.*, 51(1):694–715, 2013.

[18] Ruth F. Curtain. Linear stochastic Ito equations in Hilbert space. Stochastic control theory and stochastic differential systems, Proc. Workshop, Bad Honnef 1979, Lect. Notes Control Inf. Sci. 16, 61–84 , 1979.

[19] Ruth F. Curtain. Stability of Stochastic Partial Differential Equation. *J. Math. Anal. Appl.*, 79:352–369, 1981.

[20] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions.* Encyclopedia of Mathematics and Its Applications. 44. Cambridge etc.: Cambridge University Press, 1992.

[21] Tobias Damm. *Rational Matrix Equations in Stochastic Control.* Lecture Notes in Control and Information Sciences 297. Berlin: Springer, 2004.

[22] Tobias Damm and Peter Benner. Balanced truncation for stochastic linear systems with guaranteed error bound. *Proceedings of MTNS–2014, Groningen, The Netherlands*, pages 1492–1497, 2014.

[23] Robert Denk. Skript zur Vorlesung Funktionalanalysis II. Universität Konstanz, April 2005. https://cms.uni-konstanz.de/math/denk/home/publikationen-und-skripten/skripten/.

[24] K. V. Fernando and H Nicholson. Singular perturbational model reduction of balanced systems. *IEEE Trans. Automat. Control*, 27:466–468, 1982.

[25] Albert Ferreiro-Castilla, Andreas E. Kyprianou, and Robert Scheichl. An Euler-Poisson Scheme for Lévy driven SDEs. *J. Appl. Probab.*, 53(1):262–278, 2016.

[26] Leszek Gawarecki and Vidyadhar Mandrekar. *Stochastic Differential Equations in Infinite Dimensions with Applications to Stochastic Partial Differential Equations.* Berlin: Springer, 2011.

[27] Wilfried Grecksch and Peter E. Kloeden. Time-discretised Galerkin approximations of parabolic stochastic PDEs. *Bull. Aust. Math. Soc.*, 54(1):79–85, 1996.

[28] Carsten Hartmann, Boris Schäfer-Bung, and Anastasia Thöns-Zueva. Balanced averaging of bilinear systems with applications to stochastic control. *SIAM J. Control Optim.*, 51(3):2356–2378, 2013.

[29] Erika Hausenblas. Approximation for Semilinear Stochastic Evolution Equations. *Potential Anal.*, 18(2):141–186, 2003.

[30] Ulrich G. Haussmann. Asymptotic Stability of the Linear Ito Equation in Infinite Dimensions. *J. Math. Anal. Appl.*, 65:219–235, 1978.

[31] Desmond J. Higham and Peter E. Kloeden. Numerical methods for nonlinear stochastic differential equations with jumps. *Numer. Math.*, 101(1):101–119, 2005.

[32] Desmond J. Higham and Peter E. Kloeden. Strong convergence rates for backward Euler on a class of nonlinear jump-diffusion problems. *J. Comput. Appl. Math.*, 205(2):949–956, 2007.

[33] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations. *SIAM Journal on Numerical Analysis*, 40:1041–1063, 2002.

[34] Yaozhong Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In *Stochastic analysis and related topics, V (Silivri, 1994)*, volume 38 of *Progr. Probab.*, pages 183–202. Birkhäuser Boston, Boston, MA, 1996.

[35] Martin Hutzenthaler and Arnulf Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.*, 236(1112):v+99, 2015.

[36] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467(2130):1563–1576, 2011.

[37] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 2012.

[38] Akira Ichikawa. Dynamic programming approach to stochastic evolution equations. *SIAM J. Control Optimization*, 17:152–174, 1979.

[39] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer, 2003.

[40] Arnulf Jentzen and Peter E. Kloeden. Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise. *Proc. R. Soc. A 2009*, 465:649–667, 2009.

[41] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of Mathematics 23. Berlin: Springer, 2010.

[42] Florian Knäble. Ornstein-Uhlenbeck Equations with time-dependent coefficients and Lévy Noise in finite and infinite dimensions. Master's thesis, Universität Bielefeld, 2008.

[43] Kristian Knäble. Stochastic Evolution Equations with Lévy Noise and Applications to the Heath-Jarrow-Morton Model. Master's thesis, Universität Bielefeld, 2006.

[44] Claudia Knoche and Katja Frieler. Solutions of Stochastic Differential Equations in Infinite Dimensional Hilbert Spaces and their Dependence on Initial Data. Master's thesis, Universität Bielefeld, 2001.

[45] Raphael Kruse. Consistency and stability of a Milstein–Galerkin finite element scheme for semilinear SPDE. *Stochastic Partial Differential Equations: Analysis and Computations*, 2(4):471–516, 2014.

[46] Raphael Kruse. Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise. *IMA Journal of Numerical Analysis*, 34(1):217–251, 2014.

[47] Hui-Hsiung Kuo. *Introduction to Stochastic Integration.* Universitext. New York, NJ: Springer, 2006.

[48] Yi Liu and Brian D.O. Anderson. Singular perturbation approximation of balanced systems. *Int. J. Control*, 50(4):1379–1405, 1989.

[49] Dilip B. Madan, Peter P. Carr, and Eric C. Chang. The Variance Gamma Process and Option Pricing. *European Finance Review*, 2(1):79–105, 1998.

[50] Dilip B. Madan and Eugene Seneta. The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business*, 63(4):511–524, 1990.

[51] Xuerong Mao and Lukasz Szpruch. Strong convergence rates for backward Euler-Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics*, 85(1):144–171, 2013.

[52] Michel Metivier. *Semimartingales: A Course on Stochastic Processes.* De Gruyter Studies in Mathematics, 2. Berlin - New York: de Gruyter, 1982.

[53] Bruce C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Autom. Control*, 26:17–32, 1981.

[54] Goro Obinata and Brian D.O. Anderson. *Model Reduction for Control System Design.* Communications and Control Engineering Series. London: Springer, 2001.

[55] Szymon Peszat and Jerzy Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise. An evolution equation approach.* Encyclopedia of Mathematics and Its Applications 113. Cambridge: Cambridge University Press, 2007.

[56] Claudia Prévôt and Michael Röckner. *A Concise Course on Stochastic Partial Differential Equations.* Lecture Notes in Mathematics 1905. Berlin: Springer, 2007.

[57] Anthony J. Pritchard and Jerzy Zabczyk. Stability and Stabilizability of Infinite-Dimensional Systems. *SIAM Rev.*, 23:25–52, 1981.

[58] Martin Redmann. Zur Stochastischen Analysis Hilbertraum-wertiger Lévy-Prozesse. Master's thesis, Martin-Luther Universität Halle-Wittenberg, 2012.

[59] Martin Redmann and Peter Benner. An H_2-Type Error Bound for Balancing-Related Model Order Reduction of Linear Systems with Lévy Noise. Preprint MPIMD/15-20, Max Planck Institute Magdeburg, December 2015.

[60] Martin Redmann and Peter Benner. Approximation and Model Order Reduction for Second Order Systems with Lévy-Noise. *AIMS Proceedings*, pages 945–953, 2015.

[61] Martin Redmann and Peter Benner. Singular Perturbation Approximation for Linear Systems with Lévy Noise. Preprint MPIMD/15-22, Max Planck Institute Magdeburg, December 2015.

[62] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I: Functional analysis.* New York etc.: Academic Press, A Subsidiary of Harcourt Brace Jovanovich, Publishers, 1980.

[63] Markus Riedle and Onno van Gaans. Stochastic Integration for Lévy Processes with Values in Banach Spaces. *Stochastic Processes Appl.*, 119(6):1952–1974, 2009.

[64] Barbara Rüdiger. Stochastic integration with respect to compensated Poisson random measures on separable Banach spaces. *Stochastics Stochastics Rep.*, 76(3):213–242, 2004.

[65] Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge Studies in Advanced Mathematics. 68. Cambridge: Cambridge University Press, 1999.

[66] Steven E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models.* New York, NY: Springer, 2004.

[67] Stefan Stolze. Stochastic Equations in Hilbert Space with Lévy Noise and their Applications in Finance. Master's thesis, Universität Bielefeld, 2005.

[68] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[69] Almut E. D. Veraart and Luitgard A. M. Veraart. *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*, chapter Modelling Electricity Day-Ahead Prices by Multivariate Lévy Semistationary Processes, pages 157–188. Springer New York, 2014.

[70] Ioan I. Vrabie. *C_0-Semigroups and Applications.* North-Holland Mathematics Studies 191. Amsterdam: North-Holland., 2003.

[71] Dirk Werner. *Functional Analysis. (Funktionalanalysis.) 7th revised and expanded ed.* Springer-Lehrbuch. Berlin: Springer., 2011.

[72] Liqian Zhang, James Lam, Biao Huang, and Guang-Hong Yang. On gramians and balanced truncation of discrete-time bilinear systems. *Int. J. Control*, 76(4):414–427, 2003.

Schriftliche Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert oder verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadenersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, 23.05.16

Martin Redmann