Abstract. In this work, we develop a novel principal component analysis (PCA) for semimartingales by introducing a suitable spectral analysis for the quadratic variation operator. Motivated by high-dimensional complex systems typically found in interest rate markets, we investigate correlation in high-dimensional high-frequency data generated by continuous semimartingales. In contrast to the traditional PCA methodology, the directions of large variations are not deterministic, but rather they are bounded variation adapted processes which maximize quadratic variation almost surely. This allows us to reduce dimensionality from high-dimensional semimartingale systems in terms of covariation rather than the usual covariance concept.

The proposed methodology allows us to investigate space-time data driven by multi-dimensional latent semimartingale state processes. The theory is applied to discretely-observed stochastic PDEs which admit finite-dimensional realizations. In particular, we provide consistent estimators for finite-dimensional invariant manifolds for Heath-Jarrow-Morton models. More importantly, components of the invariant manifold induced by volatility and drift dynamics are consistently estimated and identified.

Contents

1. Introduction 2
1.1. Organization of the paper 5
2. Assumptions and Preliminary Results 5
3. Random directions and principal components 9
4. Bounded variation component and quadratic variation in $M$ 14
4.1. Correlation in $d$-dimensional asset prices 15
4.2. Stochastic PDEs with finite-dimensional realizations 15
4.3. Noise dimension vs Quadratic variation dimension 17
5. Estimation of $(W, D)$ 19
5.1. Identification of the Spaces $(W, D)$ 19
5.2. An illustrative example 21
5.3. Estimation of the spaces $(W, D)$ 22
6. Estimation of Finite-Dimensional Invariant Manifolds 23
6.1. Splitting the invariant manifold 27
6.2. Preliminaries on Factor models 32
6.3. Main Results 35
7. Appendix I: Estimating the dimension of invariant manifolds 43
8. Appendix II: Estimating $\dim Q$ 45
References 53

Date: March 20, 2015.
1. Introduction

Dimension reduction techniques has been intensively studied over the last years due to the advent of high-dimensional data in a variety of applied fields. Nowadays, deep results for a number of methodologies allow us to represent high-dimensional systems in terms of lower dimensional quantities. Generally speaking, either one projects the data onto lower dimensional manifolds through a suitable eigenvector analysis or one looks for reducing dimension mappings $\Phi : \mathcal{V} \to \mathcal{M}$ such that $\mathcal{Y}|X \sim \mathcal{Y}|\Phi(X)$, where $X$ is the predictor and $\mathcal{Y}$ is the response. Other approaches attempt to model the manifold on which the data lies by means of locally linear embeddings and other geometric techniques. The literature is vast, so we refer the reader to the excellent review surveys by Lee and Verleysen [32], Shang [38] and Burgers [12] and other references therein.

Towards an effective reduction dimension, it is crucial to interpret correctly what kind of lower dimensional manifold one has to find in order to represent the data properly. For instance, if the data comes from a Gaussian data generating process, then lower dimensional manifolds which encode most of the underlying covariance structure are the best ones. In fact, if the variance component reasonable describes the dynamics in the data, then the classical Principal Component Analysis (henceforth abbreviated by PCA) and its various extensions are the natural candidates to reduce dimensionality.

There are many cases where correlation in high-dimensional systems may not be accurately described by covariance structures. An important example is the correlation structure typically found in high-frequency data which is better described by the so-called quadratic variation matrix

$$[X]_t := [X^i, X^j]_t; 1 \leq i, j \leq d; 0 \leq t \leq T,$$

where $X = (X^1, \ldots, X^d)$ is a $d$-dimensional semimartingale sampled over the time horizon $[0, T]$.

In a financial context, the process $[X]_t$ is called the volatility (sometimes called integrated volatility). This is by far the most important quantity which needs to be estimated. Volatility is central for asset pricing, asset allocation and risk management. From the statistical point of view, quadratic variation is a property of sample paths, not probability measures. In other words, it can be computed given a single historical path and does not depended upon the probability you assign to that path. More importantly, it is directly linked with theoretical stochastic volatility models which play a central role in Financial Engineering. For this reason, the quadratic variation better describes volatility and correlation in financial data than the usual covariance concept. See e.g Ait-Sahalia and Jacod [2] and other references therein for further details.

The estimation of high-dimensional quadratic variation matrices has been a topic of great interest in the last years. Indeed, one can find deep results on the estimation of high-dimensional quadratic variation matrices for discretely-observed semimartingales with continuous paths. The literature is already rich including cases where high-dimensional semimartingales are sampled in a non-synchronous manner, with measurement errors and under sparsity situations. We refer the reader to the monograph by Ait-Sahalia and Jacod [2], Wang and Zou [43], Zheng and Li
Despite all the recent progress on volatility matrix estimation, there has been remarkably little fundamental theoretical study on dimension reduction techniques based on high-dimensional quadratic variation matrices. One notorious difficulty is the dynamic interpretation of directions and principal components over the time horizon which in typical cases is formulated in a high-frequency domain. Indeed, \{[X]_t; 0 \leq t \leq T\} is fully random which makes the analysis more evolved than the standard PCA. More precisely, all the potentially optimal projections will be stochastic processes rather than deterministic vectors.

In view of the fact many correlation structures in high-dimensional data are fully represented by quadratic variation operators, it is natural and necessary to construct a dimension reduction methodology strictly associated to \([X]\) rather than on classical covariance or conditional distributions. This is the program we start to carry out in this paper.

Let \(M = (M^1, \ldots, M^d)\) be a \(d\)-dimensional semimartingale. The starting point of the analysis is to solve an identification problem related to a possible singularity of the random matrix \([M]_T\) which can be typically found e.g in large portfolios of financial assets (see e.g Ait-Sahalia and Xiu [1] and Fan, Li and Yu [21]). More precisely, if \(\text{rank} [M]_T < d\) a.s, then under mild assumptions, one can split the set \(M = \text{span} \{M^1, \ldots, M^d\}\) into two orthogonal linear spaces \((W, D)\) such that

\[
M = W \oplus D
\]

where \(W\) and \(D\) contains only elements of \(M\) with non-zero and zero quadratic variation, respectively. The space \(W\) fully describes the volatility structure of \(M\) while \(D\) is responsible for the hidden pure drift dynamics of \(M\). At this point, we stress that the potential singularity of the quadratic variation matrix \([M]_T\) induces drift components into \(M\) which cannot be discarded. Both spaces are equally important to explain the dynamics of \(M\). In strong contrast, directions with null variance can be fully discarded in the classical PCA. This is the first major difference between the classical PCA and the theory developed in this article.

As far as the reduction dimension is concerned, we show there exist a set \(\{v^1, \ldots, v^d\}\) of orthogonal stochastic processes such that one can obtain random linear transformations of \(M\), namely

\[
S^i_t = \sum_{j=1}^d v^i_j M^j_t; 0 \leq t \leq T,
\]

such that \(S^i\) has the \(i\)-th largest quadratic variation among all possible choices of suitable standardized process. Equally importantly, under mild condition one can show the set of principal components \(\{S^i; 1 \leq i \leq p\}\) is a \(p\)-dimensional semimartingale \((p \leq d)\) which makes the dimension reduction feasible from the point of view of stochastic analysis and, in particular, continuous-time Finance modelling. In particular, \(\{v^1, \ldots, v^p\}\) is a time-varying set of eigenvectors of \([M]\) over \([0, T]\) which makes the estimation feasible.

Based on the existence of consistent estimators for the quadratic variation matrix \([M]_T\), we are able to propose consistent estimators for \((W, D)\) by means of a simple eigenvalue analysis of \([M]_T\) based on high-frequency observations of \(M\). This allows
us to reduce dimensionality in terms of quadratic variation in a very clear and consistent way.

In the second part of this article, we illustrate the theory by studying the problem of the estimation of the so-called finite-dimensional invariant manifolds w.r.t to a stochastic PDE

\[ dr_t = \left( A(r_t) + F(r_t) \right) dt + \sum_{j=1}^{m} \sigma_j(r_t) dB_t^j; r_0 = h \in E; 0 \leq t \leq T, \]

where \( E \) is a potentially infinite-dimensional Sobolev-type space of continuous functions and \( (A, F, \sigma_i; 1 \leq i \leq m) \) satisfy standard assumptions for the existence of solution.

Many space-time phenomena in natural and social sciences can be described by solutions of stochastic PDEs like (1.3). However, the intrinsic infinite-dimensionality of space-time data generated by models like (1.3) creates a big challenge in the statistical analysis of these models. In particular cases, it is well-known that one can reduce dimensionality and still get a very rich class of space-time data generated by models of type (1.3). For instance, under Lie algebra conditions (see Filipovic and Teichmann [24] and Bjork and Svensson [10]) on the coefficients of (1.3), it is well known that there exists a family of affine manifolds \( \{G_t; 0 \leq t \leq T\} \) of curves and a \( d \)-dimensional semimartingale factor process \( M \) such that

\[ r_t(\cdot) = G_t(\cdot, M_t); r_0 = h; 0 \leq t \leq T, \]

where \( \mathcal{G} = \{G_t(\cdot); x \in \mathcal{X} \subset \mathbb{R}^d; 0 \leq t \leq T\} \subset E \) is a finite-dimensional parameterized family of smooth curves. We shall write it as \( \mathcal{G}_t = \phi_t + V \) where \( V = \text{span} \{\lambda_1, \ldots, \lambda_d\} \) is a \( d \)-dimensional vector space generated by smooth curves and \( \phi \) is an \( E \)-valued smooth parametrization.

Two central unsolved problems in the stochastic PDE modelling are (i) the construction of statistical tests to check existence of \( \mathcal{G} \) and (ii) the development of associated estimation methods. The importance of this research agenda can be mainly understood in applications to interest rate modelling and other term-structure problems in Mathematical Finance. The literature is vast so we refer the reader to the works Bjork and Christensen [9], Bjork and Landn [10], Bjork Svensson [11], Filipovic [22, 25, 26], Filipovic and Teichmann [24], Richter and Teichmann [30], Teichmann and Wuthrich [12], Harms, Stefanovits, Teichmann and Wuthrich [27], Ohashi [33], Angeline [4], Slinko [39] and Malliavin, Mancino and Recchioni [34] and other references therein. In short, under the assumption of existence of \( \mathcal{G} \), the estimation of \( V \) is essential for a consistent daily calibration of potentially infinite-dimensional term-structure models.

In this work, we apply the semimartingale PCA to estimate and identify components of invariant manifolds \( \mathcal{G} \) which depicts volatility and drift dynamics in space. More precisely, let us consider the finite rank random linear operator \( Q_T : E \to V \) defined by

\[ Q_T f := \langle Q_T(\cdot), f \rangle_E; f \in E, \]
where $Q_T(u,v) := [r(u), r(v)]_T; u, v \geq \mathbb{R}_+$ and we set $Q := \text{range } Q_T$. The typical situation is $\dim Q \leq \dim V$ a.s. Let $\mathcal{N}$ be the complementary subspace of $Q$ in $V$. Under mild assumptions, we have the following splitting

$$V = Q \oplus \mathcal{N} \text{ a.s.}$$

In one hand, the pair of subspaces $(Q, \mathcal{N})$ should be considered as the analogous spaces to $(W, D)$ but in the spatial variable. On the other hand, we stress that $M$ is not observed and (1.4) is treated as a factor model.

The present methodology allows us to estimate and identify directions on the invariant manifold which comes from the volatility (represented by $Q$) and the drift (represented by $\mathcal{N}$). More importantly, we are able to identify them separately which allows us to estimate null and non-null quadratic variation factors by projecting space-time data of the form (1.3) onto $(\hat{Q} \oplus \hat{\mathcal{N}})$. In particular, this allows us to identify and estimate hidden volatility and drift components of stochastic PDEs (1.3) based on suitable asymptotic factor representations in the spirit of classical discrete-type factor models in Econometrics (see e.g Stock and Watson [40], Bai [6] and Bai and Ng [8]). We consider this separation feature as the most important aspect of the second part of this article.

1.1. Organization of the paper. The remainder of this article is structured as follows. In Section 2, we introduce notation and some basic definitions. In Section 3, we present a simple spectral analysis on quadratic variation matrices and we interpret random directions in terms of principal components. In Section 4, we discuss typical examples of high-dimensional semimartingale high-frequency data which can be summarized by principal components in terms of quadratic variation. In Section 5, we identify suitable subspaces of semimartingale systems which allows us to estimate null and non-null quadratic variation components. In Section 6, we apply semimartingale PCA to the problem of estimation of finite-dimensional invariant manifolds w.r.t stochastic PDEs. Appendices I and II contain some technical results related to dimension estimation.

Acknowledgement: The first author would like to thank the Mathematics department of ETH Zurich and Forschungsinstitut für Mathematik (FIM) for the very kind hospitality during the first year of this research project. In particular, we would like to thank Josef Teichmann for inspirational discussions about finite dimensional realizations of stochastic PDEs and for his encouragement to study their statistical aspects.

2. Assumptions and Preliminary Results

Throughout this article, we are going to work with a fixed stochastic basis of the form $(\Omega, \mathcal{F}_T, \mathcal{F}, \mathbb{P})$ where $(\Omega, \mathcal{F}_T, \mathbb{P})$ is a probability space equipped with a sample space $\Omega$, a sigma-algebra $\mathcal{F}_T$, probability measure $\mathbb{P}$ and a fixed terminal time $0 < T < \infty$. We equip the interval $[0, T]$ with the Borel sigma algebra $\mathcal{B}_T$ and we assume the filtration $\mathbb{F} := \{\mathcal{F}_t; 0 \leq t \leq T\}$ satisfies the usual conditions.

All the algebraic setup in this article will be based on the real linear space $X^d$ constituted by the set of all $\mathbb{R}^d$-valued $\mathcal{F}_T \times \mathcal{F}_T$-measurable processes. In this article, the most important subclass of $X^d$ will be the subspace $S^d$ constituted by the set of all $\mathbb{R}^d$-valued continuous $\mathbb{F}$-semimartingales on $(\Omega, \mathcal{F}_T, \mathbb{F}, \mathbb{P})$. When $d = 1$, we set
matrices. We equip \( M \in \mathbb{R}^p \) in the following convention: If \( d \) random variables for \( B \) is the following one: Let

\[
\text{Assumption 2.2 is very natural since our theory relies on the study of a realization of null quadratic variation from a non-null quadratic variation process.}
\]

Remark 2.1. We clearly do not loose generality by imposing Assumption 2.1.

Example: One typical example of semimartingale which satisfies Assumption 2.2 is the following one: Let \( B = (B^1, \ldots, B^d) \) and \( \tilde{B} = (\tilde{B}^1, \ldots, \tilde{B}^d) \) be a pair of two \( d \)-dimensional Brownian motions such that \( B^i \) and \( \tilde{B}^i \) are independent Brownian
motions for $i = 1, \ldots, d$. Now consider a system of a $2d$-dimensional Itô process $(M^1, V^1), \ldots, (M^d, V^d)$ written in vectorized form as follows

$$
(2.3) \quad \left( \frac{dM^i_t}{dV^i_t} \right) = \left( \begin{array}{cc} M^i_t (r^i_t - q^i_t) \\ \kappa_i \left( \bar{\nu}_i - \nu^i_t \right) \end{array} \right) dt
$$

$$
(2.4) \quad + \left( \begin{array}{cc} M^i_t \sqrt{V^i_t} & 0 \\ 0 & \eta_i \sqrt{V^i_t} \end{array} \right) \left( \begin{array}{cc} 1 & 0 \\ \rho_i & \sqrt{1 - \rho^2_i} \end{array} \right) \left( \begin{array}{c} dB^i_t \\ dB^i_t \end{array} \right)
$$

where $(r^i, q^i); i = 1, \ldots, d$ are adapted processes, the parameters $(\rho_i, \kappa_i, \bar{\nu}_i, V^i_0, \eta_i) \in [-1, 1] \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R}$ are assumed to satisfy well-known conditions in such way that $(M^i, V^i)$ is a well-defined Markov process (see e.g. [30]) for $i = 1, \ldots, d$. Then, one can easily check that for every $t \in (0, T)$, we have

$$
[M^i_t, M^i_t] = \int_0^t |M^i_s|^2 dB^i_s ds > 0 \text{ a.s for every } i = 1, \ldots, d.
$$

Hence, the classical Heston model described in (2.3) and (2.4) satisfies Assumption 2.2.

Let $\mathcal{M}_t := \text{span}\{M^1, \ldots, M^d\}$ be the linear space spanned by the 1-dimensional measurable processes $M^1, \ldots, M^d$ over $[0, t]$ for $0 \leq t \leq T$. Assumption 2.1 yields

$$
\dim \mathcal{M}_t = d \text{ for every } t \in (0, T].
$$

Let us now split $\mathcal{M}_t$ into two orthogonal subspaces. At first, we set

$$
\mathcal{D}_t := \{X \in \mathcal{M}_t; [X]_t = 0 \text{ a.s.}\}
$$

Observe that $\mathcal{D}_t$ is a well-defined linear subspace of $\mathcal{M}_t$ for every $t \in [0, T]$. More importantly, the following remark holds.

**Remark 2.2.** We recall that any continuous bounded variation local martingale must be constant a.s. Moreover, for every $t \in (0, T]$

$$
\{\omega; [Y, Y]_t(\omega) = 0\} = \{\omega \in \Omega; N(\omega) = 0 \text{ over the interval } [0, t]\}
$$

where $N$ is the local martingale component of the special semimartingale decomposition of $Y \in \mathcal{S}$. Therefore, Assumption 2.2 allows us to state that if $M \in \mathcal{S}^d$ is a truly $d$-dimensional process, then $\mathcal{D}_t$ is a subspace of $\mathcal{M}_t$ only constituted by continuous bounded variation adapted processes over $[0, t]$.

**Definition 2.2.** Let $\mathcal{M}_t$ be the span generated by a truly $d$-dimensional measurable process $M \in \mathcal{X}^d$ over $[0, t]$. If $\dim \mathcal{D}_t > 0$, then we say that $\mathcal{M}_t$ has a **null quadratic variation component** over the interval $[0, t]$. In particular, if $M \in \mathcal{S}^d$ and $\dim \mathcal{D}_t > 0$, then we say that $\mathcal{M}_t$ has a **bounded variation component** over $[0, t]$.

Let us give a toy example showing how a non-trivial dimension induced by bounded variation processes may appear in a very simple context.

**Example:** Let $B$ be a one-dimensional Brownian motion and let $M_t = (B_t, B_t + t); 0 \leq t \leq T$. Of course, $M$ is a truly 2-dimensional semimartingale where $\dim \mathcal{M}_t = 2$ for every $t \in (0, T]$. In particular, we clearly have $\dim \mathcal{D}_t = 1$ for every $t \in (0, T]$. 

For a deeper discussion of bounded variation components in semimartingale systems, we refer the reader to Section [3]. Let us now provide a natural notion of "quadratic variation dimension" in $\mathcal{M}_t$. To do so, let us consider the following quotient space

$$\tilde{\mathcal{M}}_t := \mathcal{M}_t / \mathcal{D}_t; \ 0 \leq t \leq T.$$  

By definition, $\tilde{\mathcal{M}}_t$ can be identified by $(\mathcal{M}_t, \sim)$ where the equivalence relation is given by

$$X \sim Y \iff X - Y \text{ is a null quadratic variation process in } \mathcal{M}_t \text{ over the interval } [0, t].$$

The following simple result connects the rank of $[\mathcal{M}]_t$ with the dimension of $\tilde{\mathcal{M}}_t$.

**Lemma 2.1.** Let $M \in \mathcal{X}^d$ be a truly $d$-dimensional measurable process satisfying Assumption [2.2]. Then,

$$\text{rank} [\mathcal{M}]_t = \dim \tilde{\mathcal{M}}_t \text{ a.s.}$$

**Proof.** The result for $t = 0$ is obvious so we fix $0 < t \leq T$. Let $p_t = \dim \tilde{\mathcal{M}}_t$ and let $\pi_{\mathcal{D}_t} : \mathcal{M}_t \to \tilde{\mathcal{M}}_t$ be the standard projection of $\mathcal{M}_t$ onto $\tilde{\mathcal{M}}_t$. Now, observe that for each $P \in \mathcal{M}_t$, $\pi_{\mathcal{D}_t}(P)$ is a set of continuous measurable processes, each of which differs from each other by a continuous null quadratic variation measurable process over the interval $[0, t]$. Nevertheless, for each process in $\pi_{\mathcal{D}_t}(P)$ its quadratic variation is equal to $[P]_t$. Therefore, we may define its quadratic variation as $[P]_t$.

By the polarization identity, we may define

$$[\pi_{\mathcal{D}_t}(P), \pi_{\mathcal{D}_t}(Q)]_t := [P, Q]_t$$

for any $P, Q \in \mathcal{M}_t$. In particular, this shows that $[N, Z]_t$ is a well-defined random variable for any $N, Z \in \tilde{\mathcal{M}}_t$.

Since $\text{span}\{\pi_{\mathcal{D}_t}(M^1), \ldots, \pi_{\mathcal{D}_t}(M^d)\} = \tilde{\mathcal{M}}_t$, then $\pi_{\mathcal{D}_t}(M^1), \ldots, \pi_{\mathcal{D}_t}(M^d)$ contains $p_t$ linearly independent components in the vector space $\tilde{\mathcal{M}}_t$. Therefore, $\dim \tilde{\mathcal{M}}_t$ equals to the number of linearly independent components in the subset $\{\pi_{\mathcal{D}_t}(M^1), \ldots, \pi_{\mathcal{D}_t}(M^d)\}$.

Let us now consider a subset of $k$ equivalence classes $\pi_{\mathcal{D}_t}(M^{\sigma(1)}), \ldots, \pi_{\mathcal{D}_t}(M^{\sigma(k)})$, where $\sigma : \{1, \ldots, k\} \to \{1, \ldots, d\}$ is a function. Let $c_1, \ldots, c_k \in \mathbb{R}$. In the sequel, we denote by $\vec{0}$ the null element of $\tilde{\mathcal{M}}_t$. With this notation at hand, Cauchy-Schwartz inequality yields

$$\sum_{i=1}^{k} c_i \pi_{\mathcal{D}_t}(M^{\sigma(i)}) = \vec{0} \iff \forall N \in \tilde{\mathcal{M}}_t, \left[ \sum_{i=1}^{k} c_i \pi_{\mathcal{D}_t}(M^{\sigma(i)}), N \right]_t = 0 \text{ a.s.}$$

In particular,

$$\sum_{i=1}^{k} c_i \pi_{\mathcal{D}_t}(M^{\sigma(i)}) = \vec{0} \iff \forall j = 1, \ldots, k, \left[ \sum_{i=1}^{k} c_i \pi_{\mathcal{D}_t}(M^{\sigma(i)}), \pi_{\mathcal{D}_t}(M^{\sigma(j)}) \right]_t = 0 \text{ a.s.}$$
By recalling that \( \{ \pi_{D_i}(M^{\sigma(1)}), \ldots, \pi_{D_i}(M^{\sigma(k)}) \} \) is linearly independent if, and only if,
\[
\sum_{i=1}^{k} c_i \pi_{D_i}(M^{\sigma(i)}) = 0 \Rightarrow c_1 = \cdots = c_k = 0,
\]
then the statement \( \{ \pi_{D_i}(M^{\sigma(1)}), \ldots, \pi_{D_i}(M^{\sigma(k)}) \} \) is a linearly independent set is equivalent that the system of equations
\[
\sum_{i=1}^{k} c_i \left[ \pi_{D_i}(M^{\sigma(i)}), \pi_{D_i}(M^{\sigma(j)}) \right]_t = 0 \quad a.s., \quad j = 1, \ldots, k
\]
has only the trivial solution \( c_1 = \cdots = c_k = 0 \) almost surely. In other words,
\[
(2.9) \quad \det \left( \left[ \pi_{D_i}(M^{\sigma(i)}), \pi_{D_i}(M^{\sigma(j)}) \right]_t : i, j = 1, \ldots, k \right) \neq 0 \quad a.s.
\]
From (2.8), (2.9) and Assumption 2.2, we shall conclude the proof. \( \square \)

**Remark 2.3.** Let \( M \) be a truly \( d \)-dimension semimartingale satisfying Assumption 2.2. In one hand, in typical examples, the dimension of the quotient space \( \tilde{M}_t \) (or equivalently \( [M]_t \)) is constant over \( [0, T] \). The reader can take the classical Heston model as a concrete example. On the other hand, the spectrum of \([M]_t\) may qualitatively change along the interval \([0, T]\). This type of flexibility will be important for applications to e.g. calibration of high-dimensional continuous-time models in Mathematical Finance. For this reason, we maintain the dependence on the time variable for the rest of this section.

Summing up the results of this section, we arrive at the following direct sum
\[
(2.10) \quad \mathcal{M}_t = \mathcal{W}_t \oplus \mathcal{D}_t; 0 \leq t \leq T,
\]
where \( \{ \mathcal{W}_t; 0 \leq t \leq T \} \) is the unique (up to isomorphisms) family of complementary linear subspaces of \( \mathcal{M}_t \) which realizes (2.10). One should notice that \( \mathcal{W}_t \) is formed by the null process in \( \mathcal{X} \) on \([0, t]\) and of elements \( V \) in \( \mathcal{M}_t \) such that \([V, V]_t > 0 \ a.s.\). Of course, \( \mathcal{W}_t \) is isomorphic to \( \tilde{M}_t \) for every \( t \in [0, T] \). To shorten notation, in the remainder of this article, we write \( \mathcal{M} := \mathcal{M}_T, \mathcal{W} := \mathcal{W}_T, \mathcal{M} := \tilde{M}_T \) and \( \mathcal{D} := \mathcal{D}_T \).

### 3. Random directions and principal components

Let us start with some heuristics related to reduction dimension for a high-dimensional vector of semimartingales \( M = (M^1, \ldots, M^d) \in \mathcal{S}^d \) which we suspect there may be some redundancy in the sense of quadratic variation. Perhaps there may be some way to combine \( M^1, \ldots, M^d \) that captures much of the quadratic variation in a few aggregate semimartingales. In particular, we shall seek random variables \( v_t = (v^1_t, \ldots, v^d_t) \in \mathbb{R}^0 \) such that
\[
(3.1) \quad S_t := \sum_{j=1}^{d} v^j_t M^j_t
\]
has the largest possible quadratic variation over \([0,t]\), where \(v_t = (v_t^1, \ldots, v_t^d)\) in (3.1) is interpreted as a *random coefficient* at time \(t \in [0,T]\) rather than a process. In other words, we seek a random linear combination of the form (3.1) such that

\[
\sum_{i,j=1}^{d} v_t^i v_t^j [M^i, M^j]_t
\]

has almost surely the largest possible value over the subset of \(\mathbb{L}^{0,d}_t\) with Euclidean norm 1 for a given \(t \in [0,T]\). Indeed, we do compute the quadratic variation of the linear combination \(S\) at time \(t\) by considering \(v_t\) as a random constant over \([0,t]\) which yields

\[
\sum_{i,j=1}^{d} v_t^i v_t^j [M^i, M^j]_t = \lim_{||\Pi|| \to 0} \sum_{t_i \in \Pi} (S_{t_i} - S_{t_{i-1}})^2
\]

in probability for any net of partitions \(\Pi = \{0 = t_0 < \cdots < t_n = t\}\) whose mesh \(||\Pi|| \to 0\).

The random coefficient

\[
\bar{v}_t = \arg\max_{w_t \in \mathbb{L}^{0,d}_t, ||w_t||_d = 1} \sum_{i,j=1}^{d} w_t^i w_t^j [M^i, M^j]_t
\]

encodes the way to combine \(M^1, \ldots, M^d\) to maximize quadratic variation at time \(t \in [0,T]\). The new variable - the leading principal component - is \(\sum_{i=1}^{d} \bar{v}_t^i M^i\). We shall continue this strategy by seeking a possible lower dimensional pairwise orthogonal sequence of aggregate variables which might explain most of the quadratic variation at each time \(t \in [0,T]\).

In this section, we are going to explain how these heuristic arguments can be made rigorous. For simplicity of exposition, we assume that one observes all trajectories of a given truly \(d\)-dimensional continuous time semimartingale \(M\) satisfying Assumptions 2.1 and 2.2.

Let us start with the following simple result which complements Lemma 2.1

**Proposition 3.1.** Let \(M \in \mathcal{X}^d\) be a truly \(d\)-dimensional process satisfying Assumption 2.2. Then, the set ker([M]_t) is deterministic \(\forall t \in [0,T]\).

**Proof.** For \(t = 0\) the statement is obvious, so let us fix \(t \in (0,T]\) and let \(D_t\) be the subspace of \(\mathcal{M}_t\) given by (2.5). Let \(p_t\) be the dimension of \(\mathcal{M}_t\). Let \(N^1, \ldots, N^{d-p_t}\) be a basis of \(D_t\) and let \(R^1, \ldots, R^{p_t}\) be a complement basis of \(\mathcal{M}_t\) in such a way that \(\{N^1, \ldots, N^{d-p_t}, R^1, \ldots, R^{p_t}\}\) is a basis of \(\mathcal{M}_t\). Let \(A\) be the change of basis from \(\{N^1, \ldots, N^{d-p_t}, R^1, \ldots, R^{p_t}\}\) to \(M = \{M^1, \ldots, M^d\}\) with matrix representation \(A = \{(a_{ij})|_{1 \leq i,j \leq d}\}\). We set \(\Omega^* := \Omega - \Omega\) where \(\Omega := \{\omega; \text{rank} [M]_t(\omega) \neq p_t\} \cup \{N^\ell(\omega) > 0 \text{ for some } \ell \in \{1, \ldots, d-p_t\}\}\). From Lemma 2.1 and the definition of \(D_t\), we know that \(\Omega^*\) has full probability. We pick \(\omega \in \Omega^*\). Of course,

\[
a_1 := (a_{11}, \ldots, a_{d1}), \ldots, a_{d-p_t} := (a_{1(d-p_t)}, \ldots, a_{d(d-p_t)}),
\]

constitutes a set of \(d - p_t\) linearly independent deterministic vectors in \(\mathbb{R}^d\) and by the every definition

---

\(^{1}\)A random set \(A\) is deterministic if there exists a subset \(A \subset \mathbb{R}^d\) such that \(A = B\) a.s.
\[ [M]_t(\omega) a_\ell = \sum_{k=1}^{d} a_{k\ell} [M^k, M^\ell]_t(\omega) = [M^\ell, N^\ell]_t(\omega) = 0 \]

for \( 1 \leq \ell \leq d-p_t, 1 \leq i \leq d \). Since \( \ker[M]_t(\omega) \subset \mathbb{R}^d \) has dimension \( d-p_t \) for every \( \omega \in \Omega^* \), then \( \ker[M]_t(\omega) = \text{span} \{a_1, \ldots, a_{d-p_t}\} \) for every \( \omega \in \Omega^* \).

**Definition 3.1.** Let \( A : \Omega \rightarrow \mathbb{M}^{+}_{d \times d} \) be a random matrix. We say that \( A \) has a deterministic eigenvalue \( \lambda \) if either \( \det(\lambda I - A(\omega)) = 0 \) for almost all \( \omega \in \Omega \) or \( \det(\lambda I - A(\omega)) \neq 0 \) for almost all \( \omega \in \Omega \).

Let us summarize some important consequences of the above results.

**Corollary 3.1.** If Assumptions 2.1 and 2.2 hold for a time \( t \in (0,T] \), then \( \ker[M]_t \) has a deterministic basis; the eigenvalue \( 0 \) of \( [M]_t \) is deterministic; \( \text{rank}[M]_t \) is deterministic; \( \mathbb{R}^d \setminus \ker[M]_t \) is deterministic.

One of the main applications of the classical PCA methodology is the interpretation of the eigenvalues and eigenvectors of the covariance matrix. In the classical case, the eigenvector associated to the largest eigenvalue points to the direction that has the largest variance. We will now proceed to a similar interpretation in terms of quadratic variation rather than variance. The main difference here is that the eigenvalues and eigenvectors are not deterministic but rather they are bounded variation adapted process. Let us start with a simple result.

**Proposition 3.2.** Let \( M \in \mathcal{S}^d \) be a truly \( d \)-dimensional semimartingale satisfying Assumption 2.2. Let us consider the vector of eigenvalues \( \lambda_1(\omega), \ldots, \lambda_d(\omega) \) (ordered in such way that \( \lambda_1(\omega) \geq \lambda_2(\omega) \geq \ldots \geq \lambda_d(\omega) \)) of the matrix \([M]_t(\omega)\) for each \((\omega,t) \in \Omega \times [0,T] \). Then, for each \( i \), \( \{\lambda_i; 0 \leq t \leq T\} \) is an adapted bounded variation process.

**Proof.** By the very definition, any eigenvalue \( \lambda_i(\omega) \) is a root of the characteristic polynomial \( p(\lambda) = \det(\lambda I - [M]_t(\omega)) \) of the random matrix \([M]_t(\omega)\). The degree of this polynomial is \( d \) and its coefficients depend on the entries of \([M]_t(\omega)\), except that its term of degree \( d \) is always \((-1)^d \lambda^d \). This allows us to conclude that the ordered eigenvalues are \( \mathbb{F} \)-adapted. In particular, by the classical Weyl’s perturbation theorem, we know there exists a deterministic constant \( C \) such that

\[
\max_j |\lambda_j^t(\omega) - \lambda_j^s(\omega)| \leq C \| [M]_t(\omega) - [M]_s(\omega) \|_\infty; (\omega,t) \in \Omega \times [0,T]
\]

where \( \| \cdot \|_\infty \) denotes the entrywise \( \infty \)-norm of a symmetric matrix. By writing \( \| [M]_t - [M]_s \|_\infty = \max_{1 \leq j \leq d} \sum_{i=1}^{d} |[M^i, M^j]_t(\omega) - [M^i, M^j]_s(\omega)| \), we clearly see \( t \mapsto \lambda_j^t(\omega) \) has bounded variation for almost all \( \omega \in \Omega \).

Let us now interpret the eigenvalues and eigenvectors of the quadratic variation matrix in a similar manner of what we interpret the eigenvalues and eigenvectors of the covariance matrix in classical PCA. We are going to introduce the brackets which encode quadratic variation of random linear combinations as described at the beginning of this section.

At first, we notice that if \( (A_1^1, A_1^2) \in L^{0,2}_t \) and \( (Y^1, Y^2) \) is a pair of real-valued semimartingales, then we shall define
\( \langle A^1_t Y^1, A^2_t Y^2 \rangle_t := A^1_t A^2_t [Y^1, Y^2]_t = \lim_{||\Pi|| \to 0} \sum_{t_i \in \Pi} (A^1_t Y^1_{t_i} - A^1_{t_i} Y^1_{t_i}) (A^2_t Y^2_{t_i} - A^2_{t_i} Y^2_{t_i}) \)

in probability for any net of partitions \( \Pi = \{0 = t_0 < \cdots < t_n = t\} \) whose mesh \( ||\Pi|| \to 0 \). In general, we define

\[
\langle X^T_t Y, Z^T_t W \rangle_t := \sum_{i,j=1}^d X^T_t Z^T_t [Y^i, W^j]_t; \quad X_t, Z_t \in \mathbb{L}^0_{i,d}, Y, W \in \mathcal{S}^d
\]

so that \( \langle X^T_t Y, Z^T_t W \rangle \) is a well defined \( \mathcal{F}_t \)-random variable for every \( X_t, Z_t \in \mathbb{L}^0_{i,d} \) and \( Y, W \in \mathcal{S}^d \). If \( X_t \in \mathbb{L}^0_{i,d} \) and \( Y \in \mathcal{S}^d \), then we write \( \langle X^T_t Y \rangle := \langle X^T_t Y, X^T_t Y \rangle \).

We stress that \( (X_t, Z_t) \) in (3.2) is viewed as a random vector at time \( t \in [0, T] \) which induces random linear combinations \( X^T_t Y \) and \( Z^T_t W \). The bracket given in (3.2) encodes the quadratic covariation of \( X^T_t Y \) and \( Z^T_t W \) at time \( t \in [0, T] \) where \( X_t \) and \( Z_t \) are treated as random constants over \([0, t]\) in the computation of (3.2). This is perfectly consistent to what happens in practice because at a given time \( t \in [0, T] \), one observes a high-frequency data from \( M \) and one has to decide if there exists linear combinations of the elements of \( M_t \) which summarizes the quadratic variation \([M]_t\).

Let us now recall a notion of order of contact between two real-valued functions. It plays a major role in choosing smooth versions of eigenvectors from matrix curves.

Let us now recall a notion of order of contact between two real-valued functions. It plays a major role in choosing smooth versions of eigenvectors from matrix curves. We refer the reader to e.g. Rutter [37] and Alekseevsky, Kriegl, Losik, and Michor [3] for further details.

**Definition 3.2.** For a continuous real-valued function \( f \) defined in a neighborhood of \( t_0 \), the order of flatness \( m_{t_0}(f) \) at \( t_0 \) is defined by the supremum of all integers \( p \) such that \( f(t) = (t - t_0)^p g(t) \) near \( t_0 \) for a continuous function \( g \). We say that two functions \( f \) and \( h \) meet of order \( \geq p \) at \( t_0 \) when \( m_{t_0}(f - h) \geq p \). Let \( A(t); 0 \leq t \leq T \) be a parameterized family of self-adjoint matrices. We say that the curve \( t \mapsto A(t) \) is generic, if no two of continuously parameterized eigenvalues meet of infinite order at any \( t \in [0, T] \) if they are not equal for all \( t \).

We are now able to summarize our discussion with the following result.

**Theorem 3.1.** Let \( M \) be a semimartingale satisfying Assumptions 2.1 and 2.2. For a given \( t \in [0, T] \), let \( (\lambda^1_t, \ldots, \lambda^d_t) \) be the list of eigenvalue as in Proposition 3.2 and let \( (v^1_t, \ldots, v^d_t) \) be an associated set of eigenvectors. Then, for every \( t \in [0, T] \)

\[
\langle (v^k_t)^T M \rangle_t = \max_{X_t \in \mathcal{Y}^k_t, ||X_t||_{\mathbb{L}^d} = 1} \langle X^T_t M \rangle_t = \lambda^k_t \text{ a.s } \quad k = 2, \ldots, d
\]

where \( \mathcal{Y}^k_t := \text{orthogonal complement of span } \{v^1_t, \ldots, v^{k-1}_t\} \) in \( \mathbb{R}^d \) for \( k = 2, \ldots, d \). In addition, if \( t \mapsto [M]_t \) is a generic smooth curve a.s and \( \dim M_t = p \) is a.s constant over the time interval \([0, T]\), then for each \( \omega \in \Omega \), there exists a choice of eigenvectors \( (v^1(\omega), \ldots, v^d(\omega)) \) over \([0, T]\) such that

\[
S^k_t := (v^k_t)^T M_t; 0 \leq t \leq T,
\]

is a semimartingale for each \( i \in \{1, \ldots, p\} \).
Proof. Fix a realization $\omega \in \Omega$ and $t \in [0, T]$. Let $A = (a_{ij})$ be a $d \times d$ matrix with entries given by

$$a_{ij} = [M^i, M^j]_t(\omega); \quad i, j = 1, \ldots, d.$$ 

It follows from assumptions 2.1 and 2.2 that $A$ is a non-negative definite matrix. Now, let us take $v_t \in L_{\ast}^{0,d}$, and let $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$ be given by $z_i = v_i^t(\omega)$. Then,

$$\langle z, A z \rangle_{\mathbb{R}^d} = \langle v_t^\top M \rangle_t.$$

Now, the result follows by applying standard arguments on the quadratic forms over $\mathbb{R}^n$ for each $(\omega, t) \in \Omega \times [0, T]$. Now, if $t \mapsto [M]_t(\omega)$ is $C^\infty$ then from Theorem 7.6 in Alekseevsky et al [3], one can choose smooth versions for related eigenvectors $v^1(\omega), \ldots, v^d(\omega)$ with bounded variation paths. By Gaussian elimination and Proposition 3.2, one can readily see that one can choose it in such way that $(v^1, \ldots, v^d)$ is a $d$-dimensional adapted process. The usual integration by parts for stochastic integrals allows us to state that $S = (S^1, \ldots, S^p)$ is a semimartingale. □

Similar to the classical PCA methodology based on covariance matrices, Theorem 3.1 yields a dimension reduction based on quadratic variation rather than covariance as follows. Let $M$ be a truly $d$-dimensional semimartingale satisfying assumption 2.2 and let us assume that one observes $[M]_t(\omega)$ for a given $(\omega, t) \in \Omega \times (0, T]$. Summing up the above results, we shall reduce dimensionality as follows

$$S_i^t = \sum_{j=1}^d v_i^j M_j^t; \quad i = 1, \ldots, \dim \tilde{\mathcal{M}}_t, \quad 0 \leq t \leq T.$$

At this point it is pertinent to make some remarks about (3.4). At first, the assumption in Theorem 3.1 that $\dim \tilde{\mathcal{M}}_t = p$ is constant a.s over $[0, T]$ holds in typical cases found in practice. See Remark 2.3

Remark 3.1. In order to get semimartingale principal components, the assumption that $t \mapsto [M]_t$ is generic cannot be avoided. See e.g example 7.7 in Alekseevsky et al [3]. However, one should notice that if two eigenvalues meet at an infinite order at a time $t_0$, then all derivatives at this point must coincide. In other words, distinct eigenvalues may intersect at a given time $t_0$, but at some finite order of differentiability, they must present distinct derivatives at $t_0$. The fact that the principal components $\{S^1, \ldots, S^p\}$ is a $p$-dimensional semimartingale (under finite order of contact) has great importance because one is able to use standard stochastic analysis techniques directly on $S$.

By the very definition, $\lambda^1_t \geq \lambda^2_t \geq \ldots \geq \lambda^d_t \geq 0$ a.s for every $t \in [0, T]$ which means that $S^i$ presents the ith largest quadratic variation among $\{S^1, \ldots, S^p\}$. One should notice that the principal components are orthogonal in the sense $[S^i, S^j]_t = 0$ a.s for $i \neq j$. Moreover, the ith eigenvector

$$v_i^t = \arg\max_{X_t \in \mathcal{V}_i^t, \|X_t\|_{\mathbb{R}^d} = 1} \langle X_t^\top M \rangle_t$$
must be interpreted as the random coefficient which makes instantaneous adjustments to the quadratic variation of the $i$th principal component $S^i$ over $[0,t]$. Lastly,

$$\frac{\lambda^i_t(\omega)}{\sum_{i=1}^d \lambda^i_t(\omega)}$$

represents the percentage explained by the $k$th principal component $S^k$ over $[0,t]$ for a realization $\omega \in \Omega$.

Remark 3.2. We stress that if one is interested to find a linear combination

$$\sum_{j=1}^d \bar{v}_j^t M_j^i; 0 \leq t \leq T,$$

such that $\bar{v}_T = \arg\max_{v \in \mathcal{H}} [v^T M]_T$ a.s over a suitable subspace $\mathcal{H}$ of $\mathbb{R}^d$-valued semimartingale processes $v$, then one has to play with an infinite-dimensional optimization problem which may not be feasible in practical examples by using a single realization $\omega \in \Omega$.

Observe that in classical PCA, eigenvectors associated to largest eigenvalues point to the direction that has the largest variance. The spectral analysis of this section provides random directions which maximizes quadratic variation rather than variance.

Remark 3.3. One should notice that semimartingales which present null quadratic variation cannot be neglected because they may contribute to the dynamics of $M$ as a drift. More precisely, when $\dim \mathcal{W} < d$ then

$$\mathcal{M} = \mathcal{W} \oplus \mathcal{D}$$

where $\dim \mathcal{D} > 0$. In this case, for each $i = 1, \ldots, d$ there exists $\alpha_i \in \mathbb{R}^d$, $(Z^{i,1}, \ldots, Z^{i,m}) \in \mathcal{W}$ and $(X^{i,1}, \ldots, X^{i,(d-m)}) \in \mathcal{D}$ such that

$$M^i = \sum_{j=1}^m \alpha_{i,j} Z^{i,j} + \sum_{\ell=1}^{d-m} \alpha_{i,\ell} X^{i,\ell}; i = 1, \ldots, d.$$  \hspace{1cm} (3.5)

We stress that this phenomena is intrinsic to the principal component analysis of high-dimensional semimartingale systems. Indeed, in the classical PCA, a random vector with zero variance must be constant. In contrast, even if $\dim \mathcal{W} = 0$ then $X^i \in \mathcal{S}^{d-m}$ in (3.5) cannot be discarded.

Let us now briefly discuss the importance of the subspaces $(\mathcal{D}, \mathcal{W})$ in concrete multi-dimensional semimartingale systems.

4. Bounded variation component and quadratic variation in $\mathcal{M}$

In this section, we discuss two concrete examples of models which exemplify the importance of analyzing the principal components of multi-dimensional semimartingale systems in terms of $(\mathcal{W}, \mathcal{D})$ rather than covariance matrices. Let us

---

This remark obviously extends to the more general case when $M \in \mathcal{X}^d$ satisfies Assumptions 2.1 and 2.2. In this case, the null quadratic variation components $(X^{i,1}, \ldots, X^{i,(d-m)})$ in (3.5) may have paths of unbounded variation.
briefly describe two cases in the context of $d$-dimensional Itô processes and stochastic PDEs admitting finite-dimensional realizations related to risk management and interest rate modelling, respectively. The former example treats the case when one has sample data from an arbitrary $d$-dimensional semimartingale process, while the latter illustrates space-time data driven by latent semimartingale factors.

4.1. Correlation in $d$-dimensional asset prices. Correlation among asset prices is a well-known phenomena and it has been studied by many authors in the context of covariance and, more recently, quadratic variation matrices. Let us suppose the asset log-prices form a $d$-dimensional Itô process

$$M_i^t = M_i^0 + \int_0^t b_i^s ds + \sum_{j=1}^d \int_0^t \sigma_{ij}^s dB_j^s; \quad i = 1, \ldots, d; \quad 0 \leq t \leq T,$$

where $b : [0, T] \times \Omega \to \mathbb{R}^d$ and $\sigma : [0, T] \times \Omega \to \mathbb{R}^{d \times d}$ satisfy usual conditions to get a well-defined $d$-dimensional semimartingale. For simplicity of exposition, let us assume that $d$ is known.

One typical example of the existence of bounded variation component in $M$ is the occurrence of correlation among $M_1^t, \ldots, M_d^t$ which can be measured by volatility, i.e., quadratic variation. This type of phenomena has been recently studied by Y. Ait-Sahalia and D. Xiu \[1\] who identify nontrivial correlation among $\{M_i^t; i = 1, \ldots, d\}$ by means of suitable estimators $[\hat{M}_i^t, M_j^t]_T; i, j = 1, \ldots, d$. In the presence of correlation among assets as in \[1\], the subspace $D$ naturally emerges as a nontrivial subspace of $M$ due to the fact that $\text{rank } [M_i^t]_T < d$.

Based on high-frequency data $\{M_i^t; 1 \leq j \leq n, i = 1, \ldots, d\}$ over $[0, T]$, we are going to present a suitable spectral analysis on $[\hat{M}]_T$ which will allow us to provide consistent estimators for the linear contribution of non-null quadratic variation and pure drift components of $W$ and $D$, respectively. See Theorem 5.1, Corollary 5.1, (5.11) and (5.12) for more details.

4.2. Stochastic PDEs with finite-dimensional realizations. Let us now discuss one example of space-time data which can be studied by using the machinery of this article. Let us concentrate the discussion in one major research theme related to interest rate modelling: The calibration problem of Heath-Jarrow-Morton models based on forward rate curves, in particular, the phenomena of the so-called finite-dimensional realizations. We refer the reader to \[9, 10, 11, 22\] and other references therein. Let us briefly recall how $(W, D)$ arises in this context. The classical HJM model can be described by a stochastic PDE of the form

$$dr_t = \left(A(r_t) + \alpha_{HJM}(r_t)\right)dt + \sum_{i=1}^m \sigma^i(r_t)dB_i^t; \quad r_0 \in E,$$

where $A = \frac{d}{dx}$ is the first-order derivative operator acting as an infinitesimal generator of a $C_0$-semigroup on a separable Hilbert space $E$ which we assume to be a space of functions $g : [a, b] \to \mathbb{R}$ where $-\infty < a \leq x \leq b < +\infty$. The drift vector field $\alpha_{HJM}$ has great importance for pricing and hedging derivative products and it is fully determined by $\sigma = \{\sigma^1, \ldots, \sigma^m\}$. See e.g \[23\] for more details.

3$[\hat{M}]_T$ is an estimator for $[M]_T$ satisfying some mild regularity conditions.
One central issue in the literature is the use of the stochastic PDE (4.2) in practice. In this case, it is very important to know when (4.2) admits a finite-dimensional subset \( G \) where the stochastic PDE never leaves as long as the initial forward rate curve \( r_0 \in G \), namely

\[
\mathbb{P}\{ r_t \in G; \forall t \in [0,T]\} = 1 \quad \text{if} \quad r_0 \in G.
\]

The subset \( G \) can be interpreted as a finite-dimensional parameterized family of smooth curves \( G = \{G(\cdot; x); x \in \mathcal{Z} \subset \mathbb{R}^d \} \subset E \) which can be used to estimate the volatility component of the model (4.2) starting with an initial curve \( r_0 \in G \). See e.g [4, 10]. Therefore, one central issue in interest rate modelling is the existence, characterization and estimation of \( G \). See [9, 10, 11, 22, 24, 25, 26, 35, 36, 4, 34] and other references therein.

As far as the existence is concerned, Filipovic and Teichmann [24] have shown that the existence of \( G \) is equivalent to

\[
\dim \{ \mu, \sigma^i; i = 1, \ldots, m \}_{LA} < \infty,
\]

in a neighborhood of \( r_0 \), where \( \mu \) is the Stratonovich drift induced by \( \sigma \) and \( x \mapsto \{ \mu, \sigma^1, \ldots, \sigma^m \}_{LA}(x) \) is the Lie algebra generated by the vector fields \( \mu, \sigma^1, \ldots, \sigma^m \).

In fact, \( G \subset E \) must be an affine submanifold of \( E \). In particular, there exists a parametrization \( \phi : [0,T] \to E \), a truly \( d \)-dimensional Brownian semimartingale \( M \) and a linear subspace \( V = \text{span}\{\lambda_1, \ldots, \lambda_d\} \) spanned by a basis \( \{\lambda_i\}_{i=1}^d \) such that

\[
r_t(x) = \phi_t(x) + \sum_{j=1}^d M^j_t \lambda_j(x) \ a.s; \ 0 \leq t \leq T; \ x \in [a,b].
\]

Under some assumptions (see e.g Duffie and Khan [19]), the semimartingale state process can be generically written as the following affine process

\[
dM_t = (\alpha + \beta M_t)dt + \Lambda D(M_t)dB_t; \ 0 \leq t \leq T,
\]

where \( \alpha \in \mathbb{R}^d \), \( \beta \) and \( \Lambda \) are \( d \times d \)-matrices, and

\[
D(M_t) := \begin{pmatrix}
\sqrt{\gamma^1_t} M_t + \delta_1 \\
\vdots \\
\sqrt{\gamma^d_t} M_t + \delta_d
\end{pmatrix}
\]

for \( 0 \leq t \leq T \), where \( \delta_i \in \mathbb{R} \) and \( \gamma_i = (\gamma_{i1}, \ldots, \gamma_{id})^T \in \mathbb{R}^d \).

The system (4.5) plays a key role in the affine class of term-structure models. In particular, it naturally appears as the state process in HJM models admitting finite-dimensional realizations. In contrast to the previous example of sample data from the \( d \)-dimensional semimartingale (4.1), \( M \) in (4.5) is not observed.

For a given pair \((M, V)\) as above, one can actually show (see Proposition 6.1), there exists a unique splitting \( V = V_1 \oplus V_2 \) which realizes

\[
r_t(x) = \phi_t(x) + \sum_{i=1}^p Y^{i^*}_t \varphi_i(x) + \sum_{j=p+1}^d \tilde{Y}^j_t \varphi_j(x) \ a.s
\]
for $0 \leq t \leq T$; $x \in [a, b]$. Here, $\{Y^i; i = 1, \ldots, p\}$ is a basis for $\mathcal{W}$ and $\{\tilde{Y}^j; j = p+1, \ldots, d\}$ is it is basis for $\mathcal{D}$ such that

$$M = \mathcal{W} \oplus \mathcal{D}.$$  

Moreover, $V_1 = \text{span}\{\varphi_1, \ldots, \varphi_p\}$ and $V_2 = \text{span}\{\varphi_{p+1}, \ldots, \varphi_d\}$.

Under the assumption that a stochastic PDE (one typical example is (4.2)) admits a finite-dimensional realization (4.4), we are going to present consistent estimators for the minimal invariant subspace $V$. More precisely, based on high-frequency data and techniques from factor analysis, we take advantage of the structure induced by $(\mathcal{W}, \mathcal{D})$ in order to provide consistent estimators $(\tilde{V}_1, \tilde{V}_2)$ for $(V_1, V_2)$ related to the minimal invariant subspace $V$.

**Remark 4.1.** We recall that finite dimensional invariant sub-manifolds (as described by $V$ in section 4.2) have fundamental importance since they constitute parameterized families of functions tailor-made for the model to calibrate initial forward rate curves. In order to consistently calibrate $\sigma$ of a generic HJM model (4.2) from the minimal invariant affine submanifold over a time horizon $[0, T]$, one can use $\tilde{V} := \tilde{V}_1 \oplus \tilde{V}_2$ (to be presented in the next sections) jointly with simulation procedures based on (4.4). This allows us to avoid ad-hoc and inconsistent calibration procedures. In particular, our methodology avoids the use of Ricatti-type systems which is problematic in high-dimension.

4.3. Noise dimension vs Quadratic variation dimension. It is convenient to point out that the rank of a quadratic variation matrix is not the maximal rank of the underlying volatility process studied by Jacod and Podolskij [31]. In fact, let $M$ be a $d$-dimensional Itô process of the form

$$M_t = M_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s; 0 \leq t \leq T.$$ 

Let $R_t := \sup_{0 \leq s \leq t} \text{rank} (c_s)$ where $c_s := \sigma_s \sigma_s^\top; 0 \leq t \leq T$.

**Proposition 4.1.** If $\sigma$ has continuous paths, then $R_t \leq \text{rank} [M]_t$ a.s for every $t \in [0, T]$. In particular, the inequality may be strict.

**Proof.** Let us fix a realization $\omega \in \Omega$ in a set of full measure and some $t$ in $[0, T]$. Let, also, $R_t > 0$. Then, since $c_t$ is a continuous matrix-valued function, and the rank is an integer-valued lower-semicontinuous function, there exists $t^* \in [0, t]$ such that $\text{rank} c_{t^*} = R_t$.

Since $c_t$ is a non-negative definite matrix, we can find a set of $R_t$ linearly independent eigenvectors for $c_t$, say, $v_1, \ldots, v_{R_t}$, with respective eigenvalues $\lambda_1, \ldots, \lambda_{R_t}$, such that $\lambda_i > 0$, for $i = 1, \ldots, R_t$.

Now, observe that if $c_1, \ldots, c_{R_t}$ are real numbers such that $c_1^2 + \cdots + c_{R_t}^2 > 0$, then by putting $w = c_1 v_1 + \cdots + c_{R_t} v_{R_t}$ and using the orthogonality of the eigenvectors, we have

$$\langle w, c_{t^*} w \rangle_{\mathbb{R}^d} = c_1^2 \lambda_1 + \cdots + c_{R_t}^2 \lambda_{R_t} > 0.$$  

Note also that, for any such vector $w$, the function $t \mapsto \langle w, c_t w \rangle_{\mathbb{R}^d}$ is continuous, so we can find an open interval $I$ containing $t^*$, with length $|I| = 2\delta$ (for some $\delta > 0$), satisfying

$$\forall s \in I, \quad \langle w, c_s w \rangle_{\mathbb{R}^d} > 1/2 \langle w, c_{t^*} w \rangle_{\mathbb{R}^d}.$$
Furthermore, using the non-negative definiteness of \( c \), we have that
\[
(4.8) \quad \forall u \in [0, T], \quad \langle w, c_u w \rangle_{\mathbb{R}^d} \geq 0.
\]

Now, suppose, by contradiction, that \( \text{rank } [M]_t < R_t \). Then, we can find real numbers \( c_1, \ldots, c_{R_t} \), with \( c_1^2 + \cdots + c_{R_t}^2 > 0 \), such that, for \( w = c_1 v_1 + \cdots + c_{R_t} v_{R_t} \), where \( v_1, \ldots, v_{R_t} \) are the eigenvectors of \( c_t^* \), given above, we have \([M]_t w = 0\), and, in particular,
\[
\langle w, [M]_t w \rangle_{\mathbb{R}^d} = 0.
\]

Then, using (4.6), (4.7) and (4.8), we obtain
\[
0 = \langle w, [M]_t w \rangle_{\mathbb{R}^d} = \int_0^t \langle w, c_s w \rangle_{\mathbb{R}^d} ds \\
\geq \int_I \langle w, c_s w \rangle_{\mathbb{R}^d} ds \\
> \int_I 1/2 \langle w, c_t^* w \rangle_{\mathbb{R}^d} ds \\
= \delta \langle w, c_t^* w \rangle_{\mathbb{R}^d} > 0.
\]

This contradiction shows that \( R_t \leq \text{rank } [M]_t \).

To show that the inequality may be strict, consider the following example: Let us assume that \( T \geq 1 \) and we take
\[
\sigma_s = \begin{pmatrix} f(s) & 0 \\
0 & f(s - 1) \end{pmatrix},
\]
where \( f(t) = t(1 - t)1_{[0,1]} \), with \( 1_A \) is the indicator function of the set \( A \). Then, clearly,
\[
c_s = \begin{pmatrix} (f(s))^2 & 0 \\
0 & (f(s - 1))^2 \end{pmatrix},
\]
and \( R_t = 1 \), for all \( t > 0 \), whereas \( \text{rank } [M]_t = 2 \) for \( t > 1 \).

\[\square\]

Remark 4.2. The main message of the above proposition is that if a direction has a non-null quadratic variation for some time \( t_0 > 0 \), then this direction has non-null quadratic variation for all times \( t \geq t_0 \). This phenomenon does not occur with the volatility matrix \( c_s \), as shown above.

We also stress that Assumptions 2.1 and 2.2 yield the study of a statistical test to check the existence of a null quadratic variation component in \( M \). The full derivation of the statistical test will be further explored in a forthcoming paper.

Corollary 4.1. Let \( M \in \mathcal{X}^d \) be a truly \( d \)-dimensional process satisfying Assumption 2.2. Let \( \lambda_1^T, \ldots, \lambda_d^T \) be the ordered eigenvalues of the associated quadratic variation matrix \([M]_T^\sigma\) such that \( \lambda_1^T \geq \cdots \geq \lambda_d^T \). The test \( H_0 : \lambda_d^T = 0 \) versus \( H_1 : \lambda_d^T > 0 \), is a well-defined statistical test and it is equivalent to \( H_0 : \text{rank } [M]_T = 1 \) versus \( H_1 : \text{rank } [M]_T = d \).

Proof. From Assumption 2.2 \( \mathbb{P}\{\lambda_d^T > 0\} > 0 \) implies \( \mathbb{P}\{\lambda_1^T > 0\} = 1 \). So that, either, for almost all realizations, \( \lambda_d^T = 0 \), or for almost all realizations, \( \lambda_d^T > 0 \). Since the null hypothesis is that \( \lambda_d^T = 0 \), under the null-hypothesis, this eigenvalue is deterministic almost surely, and due to Lemma 2.1, this is equivalent to a test on the rank of the random Hermitian matrix \([M]_T^\sigma\). \(\square\)
5. Estimation of \((\mathcal{W}, \mathcal{D})\)

In this section, we show how to estimate the pair \((\mathcal{W}, \mathcal{D})\) which realizes

\[
\mathcal{M} = \mathcal{W} \oplus \mathcal{D}
\]

for a given \(M \in \mathcal{X}^d\) satisfying Assumptions 2.1 and 2.2. At first, let us identify natural candidates for estimators based on the full knowledge of the quadratic variation matrix \([M]_T\).

5.1. Identification of the Spaces \((\mathcal{W}, \mathcal{D})\). Throughout this section, we are going to fix a truly \(d\)-dimensional process \(M = (M^1, \ldots, M^d) \in \mathcal{X}^d\) satisfying Assumption 2.2. Let \(\mathcal{M} = \mathcal{W} \oplus \mathcal{D}\) be the splitting introduced in (2.10). Throughout this section, we are going to fix a basis \(Y = \{Y^1, \ldots, Y^d\}\) for \(\mathcal{M}\) such that \(Y^1, \ldots, Y^p \in \mathcal{W}\) and \(Y^{p+1}, \ldots, Y^d \in \mathcal{D}\) where \(\text{dim } \mathcal{W} = p\) and \(\text{dim } \mathcal{D} = d - p\).

In the sequel, we describe how to identify the pair os subspaces \((\mathcal{W}, \mathcal{D})\) in such way that one can construct suitable estimators for these objects. The strategy is the construction of a suitable invertible linear operator based on the vector of observed process \(M\). In order to simplify the exposition, we first assume that one is able to observe all trajectories of a given \(M \in \mathcal{X}^d\) in continuous time. In the next section, we consider the case where only high-frequency observations from \(M\) is available.

Let us start with some technical lemmas. In the sequel, if \(T : \mathbb{R}^d \to \mathbb{R}^d\) is a linear operator then its associated matrix will be denoted by \([T] \in \mathbb{M}_{d \times d}\).

Lemma 5.1. Let us assume that a process \(M = (M^1, \ldots, M^d) \in \mathcal{X}^d\) satisfies Assumptions 2.1 and 2.2 span \(\{M^1, \ldots, M^d\}\) = \(\mathcal{M}\) and let \([M]_T\) be the quadratic variation matrix of \(M\). Let \(T : \mathbb{R}^d \to \mathbb{R}^d\) be an invertible deterministic linear operator with matrix \([T]\). Assume that the last \(p - d\) rows of \([T] \circ [M]_T\) are null vectors in \(\mathbb{R}^d\). Then, \((J^1, \ldots, J^p) := TM\) is such that \(J^1, \ldots, J^p\) is a basis for \(\mathcal{W}\) and \(J^{p+1}, \ldots, J^d\) is a basis for \(\mathcal{D}\).

Proof. Let \([T] = \{T_{ij}; 1 \leq i, j \leq d\}\). At first, if \(T\) is a deterministic and invertible linear operator, then \(\{J^1, \ldots, J^d\}\) is a linearly independent subset of \(\mathcal{M}\). We shall assume \(d - p > 0\). By assumption, the last \(p - d\) rows of \([T] \circ [M]_T\) are null vectors in \(\mathbb{R}^d\), and hence

\[
\sum_{j=1}^{d} T_{ij} [M^\ell, M^j]_T = 0 \text{ a.s, } \ell = 1, \ldots, d; i = p + 1, \ldots, d
\]

which by linearity implies that

\[
[M^\ell, \sum_{j=1}^{d} T_{ij} M^j]_T = 0 \text{ a.s; } \ell = 1, \ldots, d; i = p + 1, \ldots, d.
\]

More importantly, (5.1) yields \([\sum_{j=1}^{d} T_{ij} M^j]_T = 0 \text{ a.s; } p + 1 \leq i \leq d\). Since span \(\{J^{p+1}, \ldots, J^d\} \subset \mathcal{M}\), we actually have span \(\{J^{p+1}, \ldots, J^d\} \subset \mathcal{D}\) and the linear independence yields span \(\{J^{p+1}, \ldots, J^d\} = \mathcal{D}\).

Therefore,\footnote{Of course, this can be done by just taking a basis for \(\mathcal{D}\) and a complement basis for \(\mathcal{M}\).}
\[(5.2) \quad \text{span} \{J^1, \ldots, J^d\} = \text{span}\{J^1, \ldots, J^p\} \oplus \mathcal{D} \subset \mathcal{M} = \mathcal{W} \oplus \mathcal{D}.\]

Since \(\{J^1, \ldots, J^p\}\) is a linearly independent subset of \(\mathcal{M}\), than \((5.2)\) yields

\[
\text{span}\{J^1, \ldots, J^p\} = \mathcal{W}.
\]

This concludes the proof. \(\square\)

Lemma 5.1 gives a simple criteria to find a basis for \(\langle \mathcal{W}, \mathcal{D} \rangle\). But, unfortunately, in practice is very difficult to find a purely deterministic linear operator satisfying assumptions Lemma 5.1. The next result shows how to overcome this difficulty by considering a pathwise argument on a suitable random matrix.

**Proposition 5.1.** Let \(M = (M^1, \ldots, M^d)\) be a \(d\)-dimensional process satisfying Assumptions 2.1 and 2.2, \(\text{span} \{M^1, \ldots, M^d\} = \mathcal{M}\) and let \([M]_T\) be the quadratic variation matrix of \(M\). Let \(v_i = (v_{ij})_{1 \leq j \leq d}\) be the \(i\)th random eigenvector associated to the ordered (decreasing order) eigenvalues of \([M]_T\). Let \(\mathcal{V} : \Omega \to \mathbb{M}_{d \times d}\) be the random matrix given by

\[
(5.3) \quad \mathcal{V}(\omega) := \{v_{ij}(\omega); 1 \leq i, j \leq d\}.
\]

Then there exists a set \(\Omega^*\) of full measure such that for each realization \(\omega \in \Omega^*\), \(\{(\mathcal{V}(\omega)M)_i; p + 1 \leq i \leq d\}\) is a basis for \(\mathcal{D}\) and \(\{(\mathcal{V}(\omega)M)_i; 1 \leq i \leq p\}\) is a basis for \(\mathcal{W}\).

**Proof.** By applying the standard spectral theorem on \([M]_T(\omega)\), we see that the set of eigenvectors \(\{v_i(\omega); 1 \leq i \leq d\}\) associated to \([M]_T(\omega)\) constitutes an orthonormal basis for \(\mathbb{R}^d\) so that \(\mathcal{V}(\omega)\) is invertible for every \(\omega \in \Omega\). Let \(p = \dim \mathcal{W}\). If \(d - p > 0\) then Lemma 2.1 yields \(v_i \in \ker \ [M]_T\ a.s.; p + 1 \leq i \leq d\) and Proposition 3.1 allows us to state there exists a set \(\Omega^*\) of full probability such that they can be written as a finite linear combination of deterministic vectors for every \(\omega \in \Omega^*\). Therefore, \([M]_T v_i\) is null a.s. for every \(i \in \{p + 1, \ldots, d\}\) which implies that the last \(d - p\) rows of \(\mathcal{V}(\omega)\circ[M]_T(\omega)\) are null for every \(\omega \in \Omega^*\). A direct application of Lemma 5.1 yields \(\{(\mathcal{V}(\omega)M)_i; 1 \leq i \leq p\}\) and \(\{(\mathcal{V}(\omega)M)_i; p + 1 \leq i \leq d\}\) are basis for \(\mathcal{W}\) and \(\mathcal{D}\), respectively, for each \(\omega \in \Omega^*\). \(\square\)

**Remark 5.1.** We stress there exists a deterministic linear operator \(\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d\) satisfying conditions in Lemma 5.1. Indeed, from Proposition 3.1, we know that \(\ker \ ([M]_T)\) is a deterministic set. Let \(\{w_{p+1}, \ldots, w_d\}\) be a basis for the deterministic set \(\mathcal{H} \subset \mathbb{R}^d\) such that \(\mathcal{H} = \ker \ ([M]_T)\) a.s and let \(\{w_1, \ldots, w_p\}\) be a basis to its orthogonal complement in \(\mathbb{R}^d\). Then, \(\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d\) can be taken to be \(\mathcal{T}e_j = w_j; 1 \leq j \leq d\) where \(\{e_j\}_{j=1}^d\) is the standard Euclidean basis. This shows the existence of such operator.

**Remark 5.2.** In practice, no distinction can be made between \(\mathcal{V}\) in \((5.3)\) and a deterministic operator \(\mathcal{T}\) satisfying conditions in Lemma 5.1. Indeed, since we are dealing with a single realization \(\omega\) and \(\mathcal{V}(\omega)\) works as a candidate for \(\mathcal{T}\), then in statistical applications, one can use the matrix \(\mathcal{V}\) as a device to split the semimartingale \(M\).
5.2. An illustrative example. Let us now stress that usual techniques based on the computation of quadratic variation or simply looking at a the graphic is not enough to guess correctly the hidden structure encoded by \((W, D)\).

Let \(Z_t = (Z^1_t, Z^2_t, Z^3_t)\) be the solution of the following system of SDEs of the system

\[
\begin{align*}
\mathrm{d}Z^1_t &= -8Z^3_t \, \mathrm{d}t + \mathrm{d}B_t \\
\mathrm{d}Z^2_t &= (Z^1_t - 12Z^3_t) \, \mathrm{d}t \\
\mathrm{d}Z^3_t &= (Z^2_t - 6Z^3_t) \, \mathrm{d}t.
\end{align*}
\] (5.4)

Figure 1 shows a realization of this semimartingale.

Suppose now, we observe the vector \(M = (M^1, M^2, M^3)\), where

\[
M^1 = Z^1, \quad M^2 = 2Z^2 - 1/2Z^1 \quad \text{and} \quad M^3 = Z^3 - 1/3Z^1.
\] (5.5)

It is clear that \(W = \text{span}\{Z^1\}\) and \(D = \text{span}\{Z^2, Z^3\}\). Figure 2 contains a realization of \(M\). We stress that in practice one frequently observes data which is generated by finite linear combinations of semimartingales of type \(Z\) rather than \(Z\) itself. So the natural question is how to estimate \((W, D)\) from a discretely-observed semimartingale of type \(M\).

![Figure 1. Example of a realization of the SDE in (5.4).](image-url)
5.3. **Estimation of the spaces** \((W, D)\). Let us suppose that we are in the same setup of the previous section, but now we have a high-frequency of observations at hand from a truly \(d\)-dimensional process \(M = (M^1, \ldots, M^d)\) satisfying Assumption 2.2. In our asymptotic considerations, we always assume that the number of observations in \([0, T]\) goes to infinity, and also that the maximum distance \(\|\Pi\|\) in time between two consecutive observations goes to zero. In this section, the high-frequency data is assumed to be observed at common regular times for each \(M^i; i = 1, \ldots, d\). We leave the case of non-synchronous data to a future research. Throughout this section, we assume the existence of a consistent estimator \(\hat{[M]}_T\) for \([M]_T\) which satisfies the following assumption:

**Assumption 5.1.** \(\hat{[M]}_T\) is a sequence of non-negative definite and self-adjoint matrices such that \(\hat{[M]}_T \overset{p}{\to} [M]_T\) as \(\|\Pi\| \to 0\).

We refer the reader to the monograph by Ait-Sahalia and Jacod [2], Wang and Zou [13], Zheng and Li [44], Mallliavin and Mancino [33], Tao, Wang and Zhou [45], Bibinger et al [13], Fa, Li and Yu [21], Christensen, Podołskij and Vetter [15] and other references therein for a complete view of the estimation of \([M]_T\). The goal of this section is to describe a generic estimation methodology based on the existence of \(\hat{[M]}_T\) satisfying Assumption 5.1. More importantly, we stress the results of this section do not depend on the estimator of the quadratic variation matrix.

In the sequel, we fix \(\hat{[M]}_T\) satisfying Assumption 5.1 and we choose any consistent estimator \(\hat{b}\) for \(\text{dim} \ [M]_T\). For instance, the number of non-zero eigenvalues of \(\hat{[M]}_T\) is a consistent estimator for \(\text{dim} \ [M]_T\). Let \(\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_{\hat{b}} \geq \ldots \geq 0\).

**Figure 2.** Example of a realization of the SDE in (5.5).
\[ \hat{\lambda}_d \geq 0 \] be its eigenvalues and let us denote by \( \hat{v}_1, \ldots, \hat{v}_p, \ldots, \hat{v}_d \) the corresponding eigenvectors.

**Lemma 5.2.** Let \( M \) be a discretely-observed \( d \)-dimensional process satisfying Assumptions 2.1 and 2.2 and \( \text{span}\{M^1, \ldots, M^d\} = \mathcal{M} \). Let \([M]_T\) be a consistent estimator for \([M]_T\) satisfying Assumption 5.1. Let \( \hat{\lambda}_1, \ldots, \hat{\lambda}_p, \ldots, \hat{\lambda}_d \geq 0 \) be the ordered eigenvalues of \([\hat{M}]_T\). Then

\[
\sup_{\hat{p} + 1 \leq i \leq d} \hat{\lambda}_i \overset{p}{\to} 0
\]
as \( \|\Pi\| \to 0 \).

**Proof.** The eigenvalues are continuous functions of the entries of the matrix and \( \hat{p} \) is an integer-valued consistent estimator. In this case, we shall take \( \hat{p} = p \) for \( \|\Pi\| \) small enough so that the statement of the lemma holds. \( \square \)

Now, we need to define a metric notion on the set of finite-dimensional subspaces embedded on a possibly infinite-dimensional vector space. For this task, we make use of the same metric between subspaces defined by Bathia et al. [7]. Let \( N_1 \) and \( N_2 \) be two finite-dimensional Hilbert subspaces of an inner product vector space \( H \) with dimensions \( m_1 \) and \( m_2 \), respectively. Let \( \{\zeta_{i1}, \ldots, \zeta_{im}\} \) be an orthonormal basis of \( N_i \), \( i = 1, 2 \). Then, we define

\[
(5.6) \quad D(N_1, N_2) := \sqrt{1 - \frac{1}{\max\{m_1, m_2\}} \sum_{k=1}^{m_1} \sum_{j=1}^{m_2} (\langle \zeta_{2j}, \zeta_{1k} \rangle_H)^2}. 
\]

In the sequel, we need to compute distances for finite-dimensional subspaces which are not embedded in a natural common Hilbert space. For this reason, let \( A \) be a finite-dimensional linear space. If \( A_1 \) and \( A_2 \) are finite-dimensional subspaces of \( A \), then we define

\[
(5.7) \quad d(A_1, A_2) := D(\Phi(A_1), \Phi(A_2))
\]
where \( \Phi : A \to \mathbb{R}^m; \quad i = 1, 2 \) is the canonical isomorphism and \( \text{dim} \ A = m \). One can easily check that \( d \) is indeed a metric over the set of all finite-dimensional subspaces of \( A \). The metric \( d \) in (5.7) is very convenient to study consistency of subspace estimators.

Next, we provide a technical lemma which plays a key role in the proof of the main result.

**Lemma 5.3.** Let \( C_n, C : \Omega \to M_{d \times d} \) be a sequence of self-adjoint real \( d \times d \) matrices such that \( C_n \overset{p}{\to} C \) as \( n \to \infty \). Assume that \( q = \dim \text{Ker}(C) \) a.s and let us denote by \( v^n_1, \ldots, v^n_q \) the eigenvectors associated to the \( q \) least eigenvalues of \( C_n \), and \( v_1, \ldots, v_q \) the eigenbasis of \( \text{Ker}(C) \) associated to the eigenvalues \( \alpha_1, \ldots, \alpha_q \), respectively. Then, if we denote by \( K_n = \text{span} \{v^n_1, \ldots, v^n_q\} \) and \( K = \text{Ker}(C) \), we have that

\[
D(K_n, K) \overset{p}{\to} 0
\]
as \( n \to \infty \).
Proof. To shorten notation, in the sequel we denote by \( \langle \cdot , \cdot \rangle = \| \cdot \|_{1/2} \) the inner product over Euclidean spaces. We may assume that \( 0 < q < d \). Let us choose \( \{ v_{q+1}, \ldots, v_d \} \) be a basis of \( K^\perp \). At first, we notice that since \( K_n \) and \( K \) have the same dimension, it is sufficient to prove that \( D(K_n, K^\perp) \overset{p}{\to} 1 \). This is equivalent to prove that

\[
\sum_{j=1}^{q} \sum_{i=1}^{d-q} (\langle v^n_j, v_{q+i} \rangle)^2 \overset{p}{\to} 0 \quad \text{as } n \to \infty.
\]

To do so, let \( q_{i,j} = \langle v^n_j, v_{q+i} \rangle v_{q+i} \), and note that \( \| q_{i,j} \| \leq 1 \) a.s and \( C_{q_{i,j}} = \langle v^n_j, v_{q+i} \rangle \alpha_{q+i} v_{q+i} \). Therefore,

\[
\langle C_{q_{i,j}}, v^n_j \rangle = \alpha_{q+i} (\langle v^n_j, v_{q+i} \rangle)^2 = \sum_{i=1}^{d-q} \sum_{j=1}^{d-q} \langle C_{q_{i,j}}, v^n_j \rangle = \sum_{i=1}^{d-q} \alpha_{q+i} (\langle v^n_j, v_{q+i} \rangle)^2,
\]

and since \( \sum_{i,j} \alpha_{q+i} (\langle v^n_j, v_{q+i} \rangle)^2 \geq \alpha_{q+1} \sum_{i,j} (\langle v^n_j, v_{q+i} \rangle)^2 \) a.s we may conclude that

\[
\sum_{i=1}^{d-q} \sum_{j=1}^{d-q} (\langle v^n_j, v_{q+i} \rangle)^2 \leq \frac{1}{\alpha_{q+1}} \sum_{i=1}^{d-q} \sum_{j=1}^{d-q} \langle q_{i,j}, C_{v^n_j} \rangle \leq \frac{1}{\alpha_{q+1}} \sum_{i=1}^{d-q} \| q_{i,j} \| \cdot \| C_{v^n_j} \| \leq \frac{1}{\alpha_{q+1}} \sum_{i=1}^{d-q} \| C_{v^n_j} \| \text{ a.s } \forall n \geq 1.
\]

We now claim that

\[
(5.8) \quad \sup_{v \in K_n, \|v\|=1} \| Cv \| \to 0.
\]

Let \( \alpha^n_1 \geq \alpha^n_2 \geq \ldots \geq \alpha^n_d \) be the ordered eigenvalues of \( C_n \) related to the \( q \) least eigenvalues. Let \( \gamma_n \) be the number of non-zero eigenvalues of \( C_n \). We have \( P\{ \gamma_n = d - q \} = 1 \) for every \( n \) sufficiently large so that

\[
\sup_{v \in K_n, \|v\|=1} \| C_n v \| \leq \alpha^n_1 \overset{p}{\to} 0
\]

as \( n \to \infty \).

On the other hand, \( C_n \overset{p}{\to} C \) as \( n \to \infty \) and hence

\[
\sup_{v \in \mathbb{R}^p, \|v\|=1} \| C_n v - C v \| \overset{p}{\to} 0
\]
as \( n \to \infty \). Therefore, triangle inequality yields
\[
\sup_{\|v\|=1} \|Cv\| \leq \sup_{v \in K_n} \|C_n v - C v\| + \sup_{v \in K_n} \|C_n v\|
\]
\[
\leq \sup_{\|v\|=1} \|C_n v - C v\| + \sup_{\|v\|=1} \|C_n v\|
\]
\[
\xrightarrow{p} 0
\]
as \( n \to \infty \). This shows (5.8) and we may conclude the proof. \( \square \)

We are now able to state and prove the main result of this section.

**Theorem 5.1.** Let \( M = (M^1, \ldots, M^d) \) be a discretely-observed process satisfying Assumptions 2.1, 2.2. Let \( \hat{M}_T \) be a consistent estimator for \( [M]_T \) satisfying Assumption 5.1 and let \( \hat{p} \) be any consistent estimator for \( \dim [M]_T \). Let \( \hat{V} \) be the orthogonal matrix whose rows are formed by the eigenvectors of \( \hat{M}_T \). If
\[
(\hat{J}_1, \ldots, \hat{J}_d) := \hat{V} M, \text{ then let us define } \hat{W} := \text{span} \{\hat{J}_1, \ldots, \hat{J}_p\} \text{ and } \hat{D} := \text{span} \{\hat{J}_{p+1}, \ldots, \hat{J}_d\}.
\]
Under such conditions, we have
\[
d(\hat{W}, W) \xrightarrow{p} 0 \text{ and } d(\hat{D}, \hat{D}) \xrightarrow{p} 0,
\]
as \( ||\Pi|| \to 0 \). In particular, if \( \hat{M} := \hat{W} \oplus \hat{D} \) then \( d(\hat{M}, M) \xrightarrow{p} 0 \) as \( ||\Pi|| \to 0 \).

**Proof.** Recall the definition of the isomorphism \( \Phi \) used in equation (5.7). From Lemma 5.1 we have \( \Phi(\hat{D}) = \text{Ker}(\hat{M}_T) \), and from the very definition of \( \hat{D} \), \( \Phi(\hat{D}) = \text{Ker}(\hat{M}_T) \). Thus, from Lemma 5.3,
\[
d(\hat{D}, \hat{D}) \xrightarrow{p} 0.
\]
Now, notice that
\[
\mathbb{R}^d = \Phi(\hat{D}) \oplus \Phi(W) = \Phi(\hat{D}) \oplus \Phi(\hat{W}).
\]
Therefore, it follows from the definition of the metric \( d \) that
\[
d(\hat{W}, W) \xrightarrow{p} 0.
\]
\( \square \)

**Corollary 5.1.** Under assumptions in Theorem 5.1, if \( M \) is discretely-observed at \( \{M_t; 1 \leq r \leq n\} \) over \([0, T]\), then for each \( i = 1, \ldots, d \), there exists \( \alpha^i = (\alpha^{i1}, \ldots, \alpha^{id}) \in \mathbb{R}^d \) such that
\[
(5.9) \quad \max_{1 \leq r \leq n} |M^i_t - \sum_{\ell=1}^{\hat{p}} \alpha^{i\ell} \hat{J}^\ell_t - \sum_{k=\hat{p}+1}^d \alpha^{ik} \hat{J}^k_t| \xrightarrow{p} 0,
\]
as \( ||\Pi|| \to 0 \).

**Proof.** Let us equip \( \mathcal{X} \) with the topology of the uniform convergence in probability. Let \( \mathcal{H} \) be the smallest finite-dimensional subspace of \( \mathcal{X} \) which contains \( \{M^1, \ldots, M^d, J^1, \ldots, J^d\} \). Let \( \Phi : \mathcal{H} \to \mathbb{R}^m \) be the canonical isomorphism for some \( m > 0 \). We notice that \( \Phi \) is actually an homeomorphism when \( \mathcal{H} \) is endowed
with the subspace topology. From Theorem 5.1 and the definition of the metric \( d \) in (5.7) and (5.6), we know that

\[
(5.10) \quad d(\mathcal{M}, \hat{\mathcal{M}}) = D(\Phi(\mathcal{M}), \Phi(\hat{\mathcal{M}})) = \sqrt{2d} \sup_{\|v\|_{\mathbb{R}^d} = 1} \|T_{\Phi(\mathcal{M})}v - T_{\Phi(\hat{\mathcal{M}})}v\|_{\mathbb{R}^d} \to 0
\]

as \( \|\Pi\| \to 0 \), where \( T_A \) denotes the projection onto a closed subspace \( A \subset \mathbb{R}^d \). Then from (5.10) and using the fact that \( \Phi \) is an homeomorphism, we get the existence of \( \alpha^i = (\alpha_1^i, \ldots, \alpha_d^i) \in \mathbb{R}^d \) such that

\[
\left| \Phi(M^i) - \sum_{\ell=1}^{\hat{p}} \alpha^{id}_{\ell} \Phi(J^\ell) - \sum_{k=\hat{p}+1}^{d} \alpha^{ik}_{\ell} \Phi(J^k) \right| R^p \to 0
\]

as \( \|\Pi\| \to 0 \) which implies (5.9).

\( \square \)

Under the assumptions of Theorem 5.1, if \( M \) is discretely-observed semimartingale at \( \{M_{tk}; 1 \leq k \leq n\} \) over \([0, T]\), then we shall use Corollary 5.1 to estimate by OLS

\[
(5.11) \quad \hat{\beta}_{tk} := \arg\min_{\beta \in \mathbb{R}_d} \|M_{tk} - \beta(\hat{V}M_{tk})\|_{\mathbb{R}^d}; 1 \leq k \leq n,
\]

the regression coefficients which provide us the precise linear contribution of non-null quadratic variation and pure drift components of \( \mathcal{V} \) and \( \mathcal{D} \), respectively. In this case, the following linear combination

\[
(5.12) \quad \hat{M}^i_k := \sum_{\ell=1}^{\hat{p}} \hat{\beta}^i_{tk} \hat{J}^\ell_{tk} + \sum_{r=\hat{p}+1}^{d} \hat{\beta}^i_{tk} \hat{J}^r_{tk}; i = 1, \ldots, d, k = 1, \ldots, n.
\]

depicts the direct sum \( \mathcal{W} \oplus \mathcal{D} \) over the sample \( \{M_{tk}; 1 \leq k \leq n\} \).

6. Estimation of Finite-Dimensional Invariant Manifolds

In this section, we apply the theory developed in previous sections to present a methodology for the estimation of finite-dimensional invariant manifolds related to space-time data generated by stochastic PDEs of the form

\[
(6.1) \quad dr_t = (A(r_t) + F(r_t))dt + \sum_{j=1}^{m} \sigma_j(r_t)dB^j_t; t \geq 0; r_0 = h \in E,
\]

where \( A \) is an infinitesimal generator of a \( C_0 \)-semigroup on a separable Hilbert space \( E \) which we assume to be a subspace of continuous functions \( g : K \to \mathbb{R} \) where for simplicity of exposition we work with the one-dimensional space \( K = [a, b] \) where \( -\infty < a \leq x \leq b < +\infty \). Throughout this section, we assume that \( (F, \sigma_1, \ldots, \sigma_m) \) satisfy usual conditions to ensure the existence of a unique (at least) mild solution for (6.1) for a given initial condition \( r_0 \in E \). See e.g. [17], Th 7.2. Actually, since we are interested in stochastic PDEs with finite-dimensional
realizations as described in section 4.2, then (6.1) will exhibit strong solutions for a class of initial conditions lying in a finite-dimensional invariant manifold.

6.1. Splitting the invariant manifold. Let us now introduce the basic geometric objects related to the stochastic PDE (6.1) that we are interested in estimating. We refer the reader to Tappe [41] for a very clear treatment of these objects.

**Definition 6.1.** A family $(V_t)_{t \geq 0}$ of affine manifolds in $E$ is called a foliation generated by a finite-dimensional subspace $V \subset E$ if there exists $\phi \in C^1([0, T]; E)$ such that

$$V_t = \phi(t) + V; \ t \geq 0.$$ 

The map $\phi$ is a parametrization of $(V_t)_{t \geq 0}$.

**Remark 6.1.** We notice that the parametrizations of $(V_t)_{t \geq 0}$ are not unique, but for any distinct parametrizations $\phi^1$ and $\phi^2$ we have $\phi^1(t) - \phi^2(t) \in V$ for every $t \in [0, T]$.

In the remainder of this paper, $(V_t)_{t \geq 0}$ denotes a foliation generated by a finite-dimensional subspace.

**Definition 6.2.** The foliation $(V_t)_{t \geq 0}$ of affine manifolds is invariant w.r.t the stochastic PDE (6.1) if for every $t_0 \in \mathbb{R}^+$ and $h \in V_{t_0}$ we have

$$P\{r_t \in V_{t_0 + t}, \ for\ all\ t \geq 0\} = 1$$

for $r_0 = h$.

The above objects lead us to the following definition which is the main object of statistical study in this section.

**Definition 6.3.** We say that the stochastic PDE (6.1) has an affine realization generated by a finite-dimensional subspace $V \subset E$ if for each $h_0 \in \text{dom}(A)$ there exists a foliation $(V_t^{h_0})_{t \geq 0}$ generated by $V$ with $h_0 \in V_{t_0}^{h_0}$ which is invariant w.r.t (6.1). An affine realization with a generator $V$ is called minimal, if for another affine realization generated by some subspace $W$ we have $V \subset W$.

**Remark 6.2.** Suppose that the stochastic PDE (6.1) has an affine realization generated by a subspace $V$. We recall that for each $h_0 \in \text{dom}(A)$ the foliation $(V_t^{h_0})_{t \geq 0}$ generated by $V$ is uniquely defined. See e.g. [Lemma 2.7 [41]].

See Section 4.2 for a brief discussion on affine realizations in the context of Mathematical Finance. Throughout this paper, we assume that the stochastic PDE data generating process satisfies the following assumption.

**Assumption (A1):** The stochastic PDE (6.1) has an affine realization generated by a finite-dimensional subspace.

Let us now introduce the basic operators which will encode the underlying loading factors of the stochastic PDE that we are interested in estimating. We fix once and for all a terminal time $0 < T < \infty$, $r_0 \in \text{dom}(A)$, the minimal subspace generator $V$ of (4.2) spanned by linearly independent vectors $\{w_1, \ldots, w_d\}$ and a parametrization $\phi \in C([0, T]; E)$ so that it has null quadratic variation $[\phi(u)]_T = 0; \ u \in [a, b]$. Under assumption (A1), the stochastic PDE (4.2) has
a strong solution. From the reproducing kernel property of $E \subset C([a, b]; \mathbb{R})$, the evaluation map $\tau_u : f \mapsto f(u)$ is a bounded linear functional and therefore pointwise evaluation of the stochastic PDE is well-defined for every point-space and the following representation holds

\begin{equation}
(6.2) \quad r_t(u) = r_0(u) + \int_0^t (A(r_s)(u) + F(r_s)(u))ds + \sum_{i=1}^m \int_0^t \sigma_i^r(s)(u)dB_s^i,
\end{equation}

where we set $r_t(u) := \tau_u r_t$ for $0 \leq t \leq T$ and $u \in [a, b]$. Let us consider the following kernels

$$
\sigma_t(u, v) := \sum_{j=1}^m \sigma_j^r(r_t)(u)\sigma_j^r(r_t)(v);
$$

$$
Q_T(u, v) := [r(u), r(v)]_T = \int_0^T \sigma_s(u, v)ds, \ u, v \in [a, b], 0 \leq t \leq T.
$$

The above kernels induce random linear operators $Q_T$ and $\sigma_t$ defined almost everywhere by

$$
(Q_T f)(\cdot) := (Q_T(\cdot, \cdot), f)_E; f \in E.
$$

$$
\sigma_t f(\cdot) := (\sigma_t(\cdot, \cdot), f)_E; \ f \in E, 0 \leq t \leq T.
$$

By the very definition, the random linear operator $Q_T$ can be written as

\begin{equation}
(6.3) \quad (Q_T f)(u) = \int_0^T (\sigma_s f)(u)ds; \ f \in E.
\end{equation}

where we denote $Q := \text{Range } Q_T$. In the remainder of this article, we denote by $\mathcal{N}$ the supplementary subspace of $Q$ in the minimal subspace $\mathcal{V}$.

From (A1), we know (see e.g Th. 2.11 and (2.27) in [41]) that there exists a truly $d$-dimensional semimartingale $Z = (Z^1, \ldots, Z^p)$ which realizes the strong solution (6.2) as follows

\begin{equation}
(6.4) \quad r_t(u) = \phi_t(u) + \sum_{i=1}^d Z^i_t w_i(u); \ 0 \leq t \leq T, u \in [a, b].
\end{equation}

**Definition 6.4.** We say that the stochastic PDE in (4.2) admits a finite-dimensional realization (FDR) if for each $h \in \text{dom } (A)$ there exists a truly $d$-dimensional semimartingale $Z \in \mathcal{S}^d$, a parametrization $\phi \in C([0, T]; E)$ and a linearly independent set $\{w_1 \ldots, w_d\} \subset E$ which realize (6.4).

See e.g [10, 41, 22, 23] for more details on this affine construction of the stochastic PDE. Representation (6.4) is not unique but it will be the basis for our splitting scheme as follows. At first, in order to apply the spectral analysis in previous sections, we will assume the following hypothesis on the stochastic PDE (6.1):

**Assumption (A2):** For each initial condition $h \in \text{dom } (A)$, there exists a latent factor representation $Z$ which realizes (6.4) and it satisfies Assumption 2.2.
In the sequel, if \( L \in \mathbb{M}_{d \times d} \) and \( \eta = (\eta_1, \ldots, \eta_d) \) is a list of real-valued functions on \([a, b]\), then \( \eta(x) = (\eta_1(x), \ldots, \eta_d(x)) \in \mathbb{M}_{d \times 1} \) and we set \( L\eta \) meaning the \( \mathbb{R}^d \)-valued function \( x \mapsto L\eta(x) \).

**Remark 6.3.** Let \( r_t(u) = \phi_t(u) + \sum_{i=1}^d Z^i_t w_i(u) \); \( 0 \leq t \leq T, u \in [a, b] \) be a representation of the FDR of \((6.1)\). Let \( A \in \mathbb{M}_{d \times d} \) be an invertible random matrix. Then

\[
\phi_t(x) + \sum_{j=1}^d Y^j_t \varphi_j(x); 0 \leq t \leq T, x \in [a, b]
\]

where \( \varphi = (A^{-1})^T w \) is a random basis for \( V \) and \( Y = AZ \in X^d \).

We can actually write \( Q_T \) in terms of any representation \((6.4)\) as follows

\[
(Q_T f)(u) = \sum_{i,j=1}^d \langle f, w_i \rangle E w_j(u) [Z^i, Z^j]_T; \ f \in E; u \in [a, b].
\]

Of course, a similar representation holds by using \((6.5)\) and the following remark holds.

**Remark 6.4.** From Lemma 2.1, one can easily see that under assumption \((A2)\), any truly \( d \)-dimensional factor process realizing \((6.4)\) (or \((6.7)\)) will satisfy Assumption 2.2.

In the sequel, we need to introduce new notation. For a given \( Z \in X^d \) satisfying Assumptions 2.1 and 2.2, we denote \( \mathcal{M}(Z) := \text{span} \{Z^1, \ldots, Z^d\} \), \( \bar{\mathcal{M}}(Z) := \mathcal{M}(Z)/D(Z) \) where \( D(Z) := \{X \in \mathcal{M}(Z); [X] = 0 \text{ a.s on } [0, T]\} \) and the quotient space is defined by the equivalence relation \((2.7)\) over \([0, T]\). We stress that \( \mathcal{M}(Z), D(Z) \) and \( \bar{\mathcal{M}}(Z) \) are \( \mathcal{M}, D \) and \( \bar{\mathcal{M}} \), respectively, which are defined in \((2.6)\) for the specific choice \( M = Z \).

**Proposition 6.1.** Let \( r \) be the stochastic PDE \((6.1)\) satisfying assumptions \((A1-A2)\) and admitting a FDR generated by the minimal foliation \( \mathcal{V}^h_t = \{\phi_t + V\}; 0 \leq t \leq T \) where \( \dim V = d \) and \( r_0 = h \). Then, we shall represent \((6.1)\) as follows

\[
r_t = \phi_t + \sum_{i=1}^p Y^i_t \varphi_i + \sum_{j=p+1}^d Y^j_t \varphi_j; 0 \leq t \leq T,
\]

where \( Y \) is a truly \( d \)-dimensional semimartingale \( Y \) satisfying \( \mathcal{W}(Y) = \text{span} \{Y^1, \ldots, Y^p\} \), \( D(Y) = \text{span} \{Y^{p+1}, \ldots, Y^d\} \) and \( V = Q \oplus N \), where \( Q = \text{span} \{\varphi_1, \ldots, \varphi_p\} \) and \( N = \text{span} \{\varphi_{p+1}, \ldots, \varphi_d\} \).

**Proof.** By assumption, there exists a truly \( d \)-dimensional semimartingale \( Z = (Z^1, \ldots, Z^d) \) satisfying Assumption 2.2 and a basis \( w = \{w_i\}_{i=1}^d \) for \( V \) such that

\[
r_t = \phi_t + \sum_{i=1}^d Z^i_t w_i; 0 \leq t \leq T.
\]

From \((6.6)\), we have \( Q \subset V \) a.s so that we shall consider the random operator \( Q_T \) restricted to \( V \) as follows \( Q_T : \Omega \times V \rightarrow V \). Moreover, from \((6.6)\) we readily see that
the random matrix of the linear operator $Q_T$ is given by \([Z^i, Z^j]_T; 1 \leq i, j \leq d\) for any pair \((Z, w)\) of latent semimartingale representation $Z$ and a basis $w$ for $V$. By Lemma 2.1 we have $\dim Q = \dim \tilde{M}(Z)$ a.s. Let $Y = \{Y^1, \ldots, Y^d\}$ be a truly $d$-dimensional semimartingale such that $\{Y^1, \ldots, Y^p\}$ is a basis for $\mathcal{W}(Z)$ and $\{Y^{p+1}, \ldots, Y^d\}$ is a basis for $\mathcal{D}(Z)$ where $p = \dim Q$. Then span$\{Y^1, \ldots, Y^d\} = \mathcal{M}(Z)$ and $Y$ satisfies Assumptions 2.1 and 2.2.

Let $I : \mathcal{M}(Z) \to \mathcal{M}(Z)$ be the linear isomorphism given by the change of basis from $Z$ to $Y$. If $[I]_Y^Z = \{a_{ij}; 1 \leq i, j \leq d\}$ is the matrix of $I$, then we shall write

\[
(6.8) \quad r_t = \phi_t + \sum_{i=1}^{d} Y_t^i \varphi_i; 0 \leq t \leq T,
\]

where $\varphi_j := \sum_{i=1}^{d} a_{ij} w_i; 1 \leq j \leq d$. By writing $Q_T$ in terms of the basis $\{\varphi_j\}_{j=1}^{d}$ and using (6.8), we clearly see that $Q = \text{span}\{\varphi_1, \ldots, \varphi_p\}$. By taking $\mathcal{N} = \text{span}\{\varphi_{p+1}, \ldots, \varphi_d\}$, we then conclude (6.7).

The main message of Proposition 6.1 is the following. When the stochastic PDE is projected onto $\mathcal{Q}$ ($\mathcal{N}$), then the associated coefficients are non-null quadratic variation (bounded variation) semimartingales.

**Remark 6.5.** One should notice that functional PCA is only capable to identify the factors which explain the variance but not all the factors explaining the whole dynamics of the process. In fact, let $\{\nu_t; t \geq 0\}$ be an invariant foliation for a stochastic PDE with a parametrization $\phi$, subspace generator $V$ and covariance operator $C_t$. By the very definition if $g \in V^\perp$, then $\langle C_t f, g \rangle_E = 0 \forall f \in E$ so that $\dim \text{Range } C_t \leq \dim V$. In particular, one can easily check that the inequality may be strict.

The main goal of this section is to construct a consistent estimator for $(\mathcal{Q}, \mathcal{N})$. For this, the following results are very useful for the approach taken in this work.

**Lemma 6.1.** Let $r$ be the stochastic PDE (6.7) satisfying assumptions (A1-A2). For a given $h \in \text{dom } (A)$, let $\nu_t^h = \phi_t + V_t; 0 \leq t \leq T$ be the minimal foliation generated by $V$ such that $r_0 = h \in \nu_0^h$. Let

\[
(6.9) \quad r_t = \phi_t + \sum_{i=1}^{d} Z_t^i \eta_i; 0 \leq t \leq T,
\]

be a latent factor semimartingale representation, where $V = \text{span}\{\eta_1, \ldots, \eta_d\}$ and $Z$ satisfies Assumptions 2.1 and 2.2. Let $L \in M_{d \times d}$ be a deterministic orthogonal matrix whose the last $p - d$ rows $\{L_{p+1}, \ldots, L_d\}$ is a basis for $\text{Ker } [Z]_T$ where $\dim [Z]_T = p$. Then

\[
(6.10) \quad \mathcal{N} = \text{span}\ \{(L\eta)_{p+1}, \ldots, (L\eta)_d\}
\]

and

\[
(6.11) \quad \mathcal{Q} = \text{span}\ \{(L\eta)_1, \ldots, (L\eta)_p\}.
\]
Proof. Let us define $X = LZ$. By Lemma 5.1 we know that \{\textit{X}^1, \ldots, \textit{X}^p\} is a basis for $\mathcal{W}(\textit{Z})$ and \{\textit{X}^{p+1}, \ldots, \textit{X}^d\} is a basis for $\mathcal{D}(\textit{Z})$. We also recall that from Corollary 3.1 that $L$ is indeed a deterministic matrix. Moreover,

$$\langle L\eta_i, \textit{X}_t \rangle_{\mathbb{R}^d} = \langle \eta_i, L^{-1}\textit{X}_t \rangle_{\mathbb{R}^d} = \langle \eta_i, \textit{Z}_t \rangle_{\mathbb{R}^d}; 0 \leq t \leq T.$$ 

By assumption, identity (6.9) holds and hence we conclude that $r = \phi + \sum_{j=1}^d \textit{X}^j \xi_j$, where $\xi_i := (L\eta)_i; 1 \leq i \leq d$. Since \{\{\textit{X}^i, \textit{X}^j\}_T; 1 \leq i, j \leq d\} is the matrix of $Q_T$ w.r.t the basis \{\xi_k; 1 \leq k \leq d\}, then we conclude that both (6.11) and (6.10) holds.

In practice, we are not able to observe any semimartingale latent factor $\textit{Z} = (\textit{Z}^1, \ldots, \textit{Z}^d)$ of a stochastic PDE admitting a FDR. But it will be very important for our estimation strategy to identify the pair $(Q, N)$ in terms of the random matrix $\mathcal{Z}_T$, or more precisely, in terms of the quadratic variation of random rotations of $\textit{Z}$. Lemma 6.1 and Proposition 5.1 yield such representations.

Corollary 6.1. Under the same assumptions in Lemma 6.1, let $L : \Omega \to \mathbb{M}_{d \times d}$ be the random matrix whose rows are given by $L_i = v_i; 1 \leq i \leq d$ where \{\textit{v}_1, \ldots, \textit{v}_d\} is an eigenvector set of $[\mathcal{Z}]_T$ associated to the ordered eigenvalues $q_1 \geq q_2 \geq \ldots \geq q_d$ a.s. Then there exists a set $\Omega^*$ of full measure such that for each realization $\omega \in \Omega^*$, we have

$$\mathcal{N} = \text{span} \left\{ (L(\omega)\eta)_{p+1}, \ldots, (L(\omega)\eta)_d \right\}$$

and

$$\mathcal{Q} = \text{span} \left\{ (L(\omega)\eta)_1, \ldots, (L(\omega)\eta)_p \right\}.$$  

In order to estimate $(Q, N)$, one has to work with factor representations which are random rotations of semimartingales. Indeed, from Remark 6.4 the following consequence of Lemma 6.1 also holds.

Corollary 6.2. Let us assume the same assumptions in Lemma 6.1. Let $A \in \mathbb{M}_{d \times d}$ be an invertible random matrix. Let

$$r_t(x) = \phi_t(x) + \sum_{j=1}^d Y_t^j \varphi_j(x); 0 \leq t \leq T; x \in [a, b]$$

where $\varphi = (A^{-1})^T \eta$ is a random basis for $V$ and $Y = AZ \in \mathcal{X}_d$. Let $L : \Omega \to \mathbb{M}_{d \times d}$ be the random matrix whose rows are given by $L_i = v_i; 1 \leq i \leq d$ where \{\textit{v}_1, \ldots, \textit{v}_d\} is an eigenvector set of $[Y]_T$ associated to the ordered eigenvalues $q_1 \geq q_2 \geq \ldots \geq q_d$ a.s. Then there exists a set $\Omega^*$ of full measure such that for each realization $\omega \in \Omega^*$, we have

$$\mathcal{N} = \text{span} \left\{ (L(\omega)\varphi(\omega))_{p+1}, \ldots, (L(\omega)\varphi(\omega))_d \right\}$$

and

$$\mathcal{Q} = \text{span} \left\{ (L(\omega)\varphi(\omega))_1, \ldots, (L(\omega)\varphi(\omega))_p \right\}. $$
6.2. Preliminaries on Factor models. The goal os this section is to describe a methodology estimation for the pair \((\mathcal{Q}, \mathcal{N})\) which generates invariant foliations for stochastic PDEs of the form \(6.1\). The methodology will be inspired by the so-called Factor Analysis developed in the Econometrics literature (see e.g \([40], [5], [6], [27]\)), but with some fundamental differences: Unlike the classical discrete Factor Analysis, we are working with an underlying continuous time process sampled in high-frequency at discrete points in time and space. More importantly, the spaces \((\mathcal{Q}, \mathcal{N})\) cannot be identified by applying standard techniques from Factor Analysis due to the rather distinct behavior between quadratic variation and covariance matrices in the high-frequency setup. Throughout this section, Assumptions \((\text{A1-A2})\) are in force. We also assume the underlying state-space \(E\) is the Sobolev space of absolutely continuous functions \(f : [a, b] \rightarrow \mathbb{R}\) such that

\[
\|f\|_E^2 := |f(a)|^2 + \int_a^b |f'(x)|^2 \mu(dx) < \infty
\]

where \(\mu\) is absolutely continuous w.r.t Lebesgue measure (see e.g \([22]\)). For simplicity of exposition, we work with the closed subspace of \(E\) formed by functions \(f(a) = 0\) and we set \(\frac{df}{dx} = 1\). With a slight abuse of notation we denote it by \(E\).

We are going to fix the minimal invariant foliation \(\mathcal{V}_i = \psi_i + V\) generated by a \(d\)-dimensional subspace \(V\) equipped with a basis \(\{\lambda_1, \ldots, \lambda_d\}\) and a truly \(d\)-dimensional semimartingale \((Z^1, \ldots, Z^d)\) satisfying Assumption \(2.2\) such that

\[
(6.16) \quad r_t = \psi_t + \sum_{j=1}^d Z_t^j \lambda_j; 0 \leq t \leq T.
\]

In this section, we work in a high-frequency setup as follows. To shorten notation, the points of partition in time \((t^n_i)_{i=1}^\bar{n}\) and space \((x^n_j)_{j=1}^\bar{N}\) will be denoted by \(t_i = t^n_i\) and \(x_j = x^n_j\), respectively, and we set \(\rho(n) := \sup_{1 \leq i \leq \bar{n}-1} |t_{i+1} - t_i| = |t_2 - t_1|\) and \(\delta(N) := \sup_{1 \leq j \leq \bar{N}-1} |x_{j+1} - x_j| = |x_2 - x_1|\). That is, the samplings in time and space are equally spaced. For the sake of preciseness, it should be noted we are dealing with a sequence of refining partitions and we always assume that \(\rho(n) \rightarrow 0\), \(\delta(N) \rightarrow 0\), \(\bar{n} \rightarrow \infty\), \(\bar{N} \rightarrow \infty\) as \(n, N \rightarrow \infty\), where both \(n\) and \(N\) goes to infinity.

We assume that the observations are generated by a space-time process

\[
(6.17) \quad X_t(x) := r_t(x) + \varepsilon_t(x); 0 \leq t \leq T, x \in [a, b]
\]

where \(\varepsilon\) represents a space-time error component satisfying some regularity conditions. In this section, we assume that one is able to sample the curves \(x \mapsto X_t(x)\) in high-frequency in time. For instance, term-structure objects like interpolated forward rate curves are examples of this type of data. See e.g Diebold and Li [18] and other references therein.

In particular, under \((\text{A1-A2})\), the \((n \times N)\)-matrix \(X_t(x_j)\) of observations admits an affine noisy representation

\[
(6.18) \quad X_t(x_j) = \phi_{t_i}(x_j) + \sum_{k=1}^d Z_{t_i}^k \lambda_k(x_j) + \varepsilon_{t_i}(x_j)
\]
for $i = 1, \ldots, \bar{n}$ and $j = 1, \ldots, \bar{N}$. Throughout this section, we assume that $\phi$ is known by the observer and with a slight abuse of notation we write $X$ for the difference $X - \phi$. In matrix representation, we shall write

$$X = Z\Lambda^\top + \mathcal{E}, \quad X_i = \Lambda Z_i + \mathcal{E}_i; 1 \leq i \leq n,$$

where $\Lambda := \{\lambda_j(x_i); 1 \leq i \leq \bar{n}, 1 \leq j \leq d\}$, $X := \{X_{t_i}(x_j); 1 \leq i \leq \bar{n}, 1 \leq j \leq \bar{N}\}$, $Z := \{Z_t^j; 1 \leq i \leq \bar{n}, 1 \leq j \leq d\}$ and $\mathcal{E} := \{\varepsilon_{t_i}(x_j); 1 \leq i \leq \bar{n}, 1 \leq j \leq \bar{N}\}$. Let $\mathcal{H}^q$ be the space of $q$-integrable continuous Brownian semimartingales. The following assumptions are inspired by Bai and Ng \[5\] and Bai \[6\] but in the context of a continuous time setup sampled at discrete intervals. For the sake of completeness, we list them here.

**(D1)** $Z^j \in \mathcal{H}^4$ for each $j = 1, \ldots, d$ and

$$\rho(n) \sum_{i=1}^{\bar{n}} Z_t, Z^\top_t \rightarrow \Sigma_Z := (Z^j, Z^j)_{L^2([0,T];\mathbb{R})} \leq i, j \leq d$$

in probability as $n \rightarrow \infty$ and $\Sigma_Z$ is a $d \times d$ positive definite matrix a.s.

**(D2)** $\sup_{j \geq 1} \|\lambda(x_j)\|_{\mathbb{R}^d} < \infty$ and

$$\|\delta(N) \sum_{j=1}^N \lambda^\top(x_j)\lambda(x_j) - \int_a^b \lambda^\top(x)\lambda(x)dx\| \rightarrow 0$$

as $\delta(N) \rightarrow 0$. Moreover, $\Sigma_{\lambda} := \int_a^b \lambda^\top(x)\lambda(x)dx$ is a $d \times d$-positive definite matrix.

**(D3)** The error process $\varepsilon$ satisfies assumptions:

- $\mathbb{E}[\varepsilon_{t_1}(x_j)] = 0$, $\mathbb{E}[\sup_{x_j} |\varepsilon_{t_1}(x_j)|^6 < \infty$

- If $\gamma_N(t_i, t_j) := \mathbb{E}[\varepsilon_{t_i}(x_j)\varepsilon_{t_j}(x_j)]$ then $\sup_t \gamma_N(t_i, t_j) < \infty$ and the sum $\rho(n) \sum_{i,j=1}^N |\gamma_N(t_i, t_j)|$ is bounded in $n, N$.

- $\sup_{N \geq 1} \delta(N) \sup_{t,m=1}^N \sup_t |\mathbb{E}[\varepsilon_{t_1}(x_m)\varepsilon_{t_i}(x_j)| \leq \infty$.

- $\sup_{N \geq 1} \delta(N) \rho(n) \sum_{i,j=1}^N \sum_{t,m=1}^N |\mathbb{E}[\varepsilon_{t_1}(x_k)\varepsilon_{t_1}(x_m)| \leq \infty$.

- $\mathbb{E}[\delta^{1/2}(N) \sum_{i=1}^N |\varepsilon_{t_i}(x_k)\varepsilon_{t_i}(x_k) - \mathbb{E}[\varepsilon_{t_i}(x_k)\varepsilon_{t_i}(x_k)]|^4$.

- The error $\varepsilon$ and the factors $Z$ are mutually independent.

**(D4)**

$$\sup_{n, N} \sup_{t} \mathbb{E}\left\|\sqrt{\rho(n)} \delta(N) \sum_{i=1}^{\bar{n}} Z_t \left[\varepsilon_{t_1}(x_j)\varepsilon_{t_i}(x_j) - \mathbb{E}[\varepsilon_{t_1}(x_j)\varepsilon_{t_i}(x_j)]\right]\right\|_{\mathbb{F}}^2 < \infty$$

$$\sup_{n, N} \mathbb{E}\left\|\sqrt{\rho(n)} \delta(N) \sum_{i=1}^{\bar{n}} Z_t^\top \lambda(x_j)\varepsilon_{t_i}(x_j)\right\|_{\mathbb{F}}^2 < \infty$$

**Remark 6.6.** Assumptions (D1-D2) imply that the sampling scheme in time and space must be in high-frequency so that $\lim_{n \rightarrow \infty} \rho(n) = \lim_{N \rightarrow \infty} \delta(N) = 0$. Moreover, since the target matrices are Gramian matrices then $\Sigma_Z$ and $\Sigma_{\lambda}$ are positive
definite if, and only if, \( \{ Z^1, \ldots, Z^d \} \) and \( \{ \lambda_1, \ldots, \lambda_d \} \) are linearly independent in \( L^2([0,T];\mathbb{R}) \) and \( L^2([a,b];\mathbb{R}) \), respectively. In other words, if \( Z \) does not satisfy (D1) then we shall reduce the effective dimension without losing information so that assumption \( \text{Rank} \Sigma_Z = d \) a.s is not restrictive at all. The fact that \( \Sigma_\lambda \) is a positive definite matrix is equivalent to the fact that \( \{ \lambda_1, \ldots, \lambda_d \} \) is linearly independent on the state space \( E \) equipped with the \( L^2([a,b];\mathbb{R}) \)-inner product. In contrast to the usual factor analysis (see e.g. [23]), we stress that \( \Sigma_Z \) is random. The assumption \( \sup_{j \geq 1} \| \lambda(x_j) \|_{\mathbb{R}^d} < \infty \) holds because the loadings \( \{ \lambda_i; i = 1, \ldots, d \} \) are continuous functions on \([a,b]\).

The first step in the estimation procedure is to estimate the underlying dimension. But this is an almost straightforward application of the (APCA). For the sake of completeness, we give the details here. In general, we are interested in solving the following optimization problem (for large \( n, N \))

\[
\min_{\Lambda^k, \mathcal{Y}(k)} \rho(n)\delta(N) \sum_{i=1}^{\bar{n}} \sum_{j=1}^{\bar{N}} \left( X_{tk}(x_j) - \langle g^k(x_j), Y_{tk}(k) \rangle_{\mathbb{R}^k} \right)^2,
\]

where the minimum is taken over the set of real matrices with columns

\[
\Lambda^k = (g^1, \ldots, g^k) \in \mathbb{M}_{\bar{n} \times k} ; \mathcal{Y}(k) = (Y(1), \ldots, Y(k)) \in \mathbb{M}_{\bar{N} \times k},
\]

subject to either \( \delta(N)\Lambda^k \Lambda_k = I_k \) or \( \rho(n)\mathcal{Y}^\top(k)\mathcal{Y}(k) = I_k \) (Identity matrix in \( \mathbb{M}_{k \times k} \)). Here \( g^i := (g^i(x_1), \ldots, g^i(x_{\bar{n}}))^\top \) and \( Y(i) := (Y_{t_1}(i), \ldots, Y_{t_n}(i))^\top \) for \( 1 \leq i \leq k \). The index \( k \) encodes the allowance of \( k \) factors in the estimation procedure.

**Remark 6.7.** In order to avoid curse of dimensionality issues, we do assume \( k < \min\{\bar{n}, \bar{N}\} \) and \( n, N \rightarrow \infty \) jointly.

The factor estimator is defined as follows. Let \( \hat{Y}(k) \in \mathbb{M}_{\bar{n} \times k} \) be the random matrix defined by \( \hat{Y}_{tk}(k) := \rho(n)^{-1/2}y^i_{tk} ; 1 \leq j \leq k, 1 \leq i \leq \bar{n} \) whose \( j \)-th column

\[
y^j := (y^j_1, \ldots, y^j_{\bar{n}}) \in \mathbb{M}_{\bar{n} \times 1}
\]

is an eigenvector associated to the \( j \)-th largest eigenvalue of \( XX^\top \in \mathbb{M}_{\bar{n} \times \bar{n}} \) subject to \( \rho(n)XX^\top(k)\hat{Y}(k) = I_k \). The loading factor estimator is given by \( \hat{\Lambda}^k := \rho(n)XX^\top\hat{Y}(k) \)

In the sequel, we denote

\[
\tag{6.19}
V(k, \hat{Y}(k)) := \min_{\Lambda^k} \rho(n)\delta(N) \sum_{i=1}^{\bar{n}} \sum_{j=1}^{\bar{N}} \left( X_{tk}(x_j) - \langle g^k(x_j), \hat{Y}_{tk}(k) \rangle_{\mathbb{R}^k} \right)^2.
\]

The estimation procedure for the underlying dimension of \( V \) is due to Bai and Ng [5]. They propose a class of information criteria of the form

\[
\tag{6.20}
PC(k) := V(k, \hat{Y}(k)) + kq(n, N)
\]

for penalty functions \( q(n, N) \).

---

6Since we work with the subspace of functions \( f \in E \) such that \( f(a) = 0 \), then \( \langle \cdot, \cdot \rangle_{L^2} \) is indeed an inner product over \( E \).
6.3. Main Results. Let us now present the main results of this section. The starting point for the estimation of \((\hat{Q}, \hat{N})\) is to take advantage of the identities [6.12] and [6.13] based on a quadratic variation matrix \([Y]_T\) constructed from an asymptotic \(Y \in \mathcal{X}^d\) satisfying Assumptions 2.1 and 2.2. In the sequel, \(\hat{d}\) is a consistent estimator for \(\text{dim } V\) as described in Appendix I. The following list of assumptions will be in force throughout this section.

(Q1) The eigenvalues of \(\Sigma_{\lambda} \Sigma_Z \in \mathbb{M}_{d \times d}\) are distinct almost surely.

(Q2) We assume
\[\rho(n) \sum_{1 \leq t < s \leq \bar{n}} |y^k_{t,i} y^k_{s,i}| \]
is bounded in probability for every \(k \in \{1, \ldots, d\}\).

(Q3)
\[\rho(n) \sum_{1 \leq t < s \leq \bar{n}} y^k_{t,i} y^k_{s,i} \langle \varepsilon_t, \lambda_r \rangle_{\mathbb{R}^N} \delta(N) \langle \varepsilon_s, \lambda_j \rangle_{\mathbb{R}^N} \delta(N) \overset{P}{\to} 0\]
as \(n, N \to \infty\) for each \(k, r, j \in \{1, \ldots, d\}\).

(Q4) There exists a sequence of natural numbers \(\{\gamma(n); n \geq 1\}\) decaying to zero such that
\[\mathbb{E} \sum_{i=1}^{\bar{n}} \|\Delta \varepsilon_{t,i}\|_{\mathbb{R}^N}^2 \delta(N) = O(\gamma(n)).\]

(Q5) \(\sup_{0 \leq t \leq T} \|\varepsilon_t\|_{E}^2\) is bounded in probability and for each \(i \in \{1, \ldots, d\}\),
\[\rho(n) \sum_{1 \leq t < s \leq \bar{n}} |y^k_{t,i} y^k_{s,i}| \|\varepsilon_{t,i}\|_E \|\varepsilon_{s,i}\|_E \overset{P}{\to} 0\]
as \(n \to \infty\).

Remark 6.8. Assumption (Q1) is essential to our estimation procedure because it yields an asymptotic \(Y \in \mathcal{X}^d\) and a random basis for \(V\) which will allow us to construct \((\hat{Q}, \hat{N})\). The technical conditions (Q2, Q3, Q5) are not strong since they impose a very mild growth condition on the eigenvectors of \(X^\top X\). Assumption (Q4) is quite natural for error structures arising in space-time data generated by stochastic PDEs. For example, as far as the consistency problem of the HJM model (see section 4.2), assumption (Q4) means that the initial fitting method used to interpolate points which generates \(X\) cannot introduce an extrinsic volatility for the market. In other words, (Q4) rules out pure martingale error structures.

Since \(G\) is invertible a.s and \(\Sigma_{\lambda}\) is positive definite, then the random matrix matrix \(A = C^{-1} G \Sigma_{\lambda} \in \mathbb{M}_{d \times d}\) given by
\[A_{ij} = \sum_{k=1}^{d} c_{i}^{-1} G_{ik} \int_{a}^{b} \lambda_k(x) \lambda_j(x) dx; \quad 1 \leq i, j \leq d,\]
is invertible a.s. Then we shall apply Remark 6.3 to state that $Y = AZ \in \mathcal{X}^d$ is a truly $d$-dimensional process and it is a factor measurable process realizing (6.16) for the basis (loading factors) $(A^{-1})^T \lambda$. From Remark 6.4 $Y$ satisfies Assumptions 2.1 and 2.2. Let us define $[\hat{Y}]_T := (\hat{m}_{tk})_{1 \leq t, k \leq \hat{d}}$ and $[Y]_T := (m_{tk})_{1 \leq t, k \leq d}$ the matrices given, respectively, by

\[(6.21) \hat{m}_{tk} := \sum_{i=1}^{n-1} (\hat{Y}_{t+1,i}(d) - \hat{Y}_{t,i}(d))(\hat{Y}_{t+1,k}(d) - \hat{Y}_{t,k}(d)),\]

for $1 \leq t, k \leq \hat{d}$ and

\[m_{sv} := [Y^s, Y^v]_T; \quad 1 \leq s, v \leq d.\]

In the sequel, let us denote the following scalar processes

\[(6.22) \gamma_N(t, t_i) := \delta(N)\mathbb{E}(\varepsilon_{t, t_i})_{\mathbb{R}^N} \quad \theta_N(t, t_i) := \delta(N)\langle \varepsilon_{t, t_i} \rangle_{\mathbb{R}^N} - \gamma_N(t, t_i)\]

\[(6.23) \xi_N(t, t_i) := Z_i^T \Lambda^T \varepsilon_{t, t_i} \delta(N); \quad \eta_N(t, t_i) := Z_i^T \Lambda^T \varepsilon_{t, t_i} \delta(N)\]

for $1 \leq i, \ell \leq n$ and $n, N$ positive integers.

**Proposition 6.2.** If assumptions (Q1-Q4) hold then

\[\|\|Y\|_T - [Y]_T\|_F^2 \xrightarrow{p} 0\]

as $n, N \to \infty$.

**Proof.** At first, by taking $n, N$ large enough, we shall assume that $\hat{d} = d$ because $\hat{d}$ is an integer-valued consistent estimator. In the sequel, if $P$ is a real-valued process then we write $\Delta_{Si}P := P_{i+1} - P_i; 1 \leq i \leq n - 1$. By using the definition of $\hat{Y}(d)$, one can actually write (see e.g. proofs of Theorem 1 in [5, 6])

\[\hat{Y}(d) = H_d^T Z_t + L_{nN}^{-1}(\rho(n)) \sum_{i=1}^{\hat{n}} Y_{t,i}(d) \left( \gamma_N(t, t_i) + \theta_N(t, t_i) + \xi_N(t, t_i) + \eta_N(t, t_i) \right)\]

\[= H_d^T Z_t + \hat{R}_t(n, N),\]

where $H_d^T = L_{nN}^{-1}(\rho(n)) \Lambda^T \delta(N)$. To shorten notation, we set $\tilde{W}_t := H_d^T Z_t$ and $\varphi_N(t, t_i) := \gamma_N(t, t_i) + \theta_N(t, t_i) + \xi_N(t, t_i) + \eta_N(t, t_i)$ for $1 \leq i, \ell \leq \hat{n}$. In the sequel, for $r, \ell = 1, \ldots, d$ we denote $O_{\mathbb{P}(r, \ell)}(\xi_n)$ a constant which may differ from line to line and let us denote the $d \times d$-matrix given by $\hat{W} := (\hat{w}_{sq})$ where

\[\hat{w}_{sq} := \sum_{i=1}^{\hat{n}-1} \Delta \hat{W}_t(s) \Delta \hat{W}_t(q)\]

for $s, q = 1, \ldots, d$. We claim that

\[(6.24) \sum_{m=1}^{d} \sum_{i=1}^{\hat{n}-1} \left( \Delta \hat{R}_t^m(n, N) \right)^2 \xrightarrow{p} 0\]
and

\[ (6.25) \quad \text{vec}(\hat{W}) \overset{p}{\to} \text{vec}(\{Y\}_T) \]

as \( n, N \to \infty \). Let \( L_{n,N} = \text{diag}(\gamma_1, \ldots, \gamma_n) \). By the very definition,

\[ (H^T_d)_{ij} = \sqrt{\rho(n)\delta(N)} \sum_{k=1}^{d} \left( \sum_{i=1}^{\bar{n}} \gamma_i^{-1} y_{i}^j Z_{i}^k \right) \left( \sum_{m=1}^{N} \lambda_k(x_m)\lambda_j(x_m) \right); 1 \leq i, j \leq d. \]

By Lemma 7.2, we know that \( L_{n,N} \overset{p}{\to} \text{diag}(c_1, \ldots, c_d) \) as \( n, N \to \infty \), where \((c_1, \ldots, c_d)\) are the eigenvalues of \( \Sigma \Sigma^\dagger \). Then Lemma 7.3 yields

\[ \hat{w}_sq = \sum_{j=1}^{d} \sum_{r=1}^{d} (H^T_d)_{jqr} (H^T_d)_{sr} \sum_{i=1}^{\bar{n}} \Delta Z^i \Delta Z^r \]

\[ \overset{p}{\to} \sum_{j=1}^{d} \sum_{k=1}^{d} \sum_{m=1}^{d} \sum_{l=1}^{d} \left( \lambda_k \lambda_l \right) L^2_{\{a,b\}R} c^{-1}_q G_{jk} (\lambda_m \lambda_l) L^2_{\{a,b\}R} c^{-1}_s G_{lm} [Z^j, Z^r]_T \]

\[ = [Y^s, Y^q]_T; 1 \leq s, q \leq d, \]

as \( n, N \to \infty \). This shows that \((6.25)\) holds. By noting that

\[ \Delta \hat{Y}_{1,\ell}(d) \Delta \hat{Y}_{1,k}(d) = \Delta \hat{W}_{1,k}(\ell) + \Delta \hat{W}_{1,\ell}(k) \Delta \hat{R}_{1,k}(n,N) \]

\[ \overset{(6.26)}{=} \Delta \hat{R}_{1,k}(n,N) \Delta \hat{W}_{1,\ell}(\ell) + \Delta \hat{R}_{1,\ell}(n,N) \Delta \hat{R}_{1,k}(n,N); 1 \leq k, \ell \leq d, \]

we only need to check \((6.24)\) in order to conclude the proof. Let \( \hat{S}_{1,k}(n,N) := L_{n,N} \hat{R}_{1,k}(n,N) \in M_{d \times 1} \). From Lemma 7.2, we know that \( \|L_{n,N}^{-1}\|_F = O_p(1) \), so we only need to check that

\[ \overset{(6.27)}{=} \sum_{m=1}^{\bar{n}} \sum_{i=1}^{\bar{n}-1} \left( \Delta \hat{S}_{1,k}^m(n,N) \right)^2 \overset{p}{\to} 0 \quad \text{as } n, N \to \infty. \]

At first, for each \( k \in \{1, \ldots, d\} \) we shall write

\[ \sum_{i=1}^{\bar{n}-1} \left( \Delta \hat{S}_{1,k}^i(n,N) \right)^2 = \rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^k|^2 (\Delta_i \varphi_N(t\ell, t_i))^2 \]

\[ + 2 \rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq \ell < s \leq \bar{n}} y_{i\ell}^k \Delta_i \varphi_N(t\ell, t_i) y_{is}^k \Delta_i \varphi_N(t_s, t_i) \]

\[ \overset{(6.28)}{=} T_1(n,N) + T_2(n,N) \]

where \( \Delta_i \varphi_N(t\ell, t_i) := \varphi_N(t\ell, t_i) - \varphi_N(t\ell, t_{i-1}); 1 \leq i \leq \bar{n} - 1, 1 \leq \ell \leq \bar{n} \). We divide the argument into two steps.

**Analysis of** \( T_1(n,N) \). It is sufficient to prove that
\[
\rho(n) \sum_{i=1}^{n} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 \left[ (\Delta_i \gamma_N(t_{\ell}, t_i))^2 + (\Delta_i \theta_N(t_{\ell}, t_i))^2 \right] \\
+ (\Delta \xi_N(t_{\ell}, t_i))^2 + (\Delta \eta_N(t_{\ell}, t_i))^2 \right] = O_P(\rho(n))
\]
for each \( k \in \{1, \ldots, d\} \). In fact, a simple application of Cauchy-Schwartz inequality and the fact that \( \sum_{i=1}^{\bar{n}} |y_{i\ell}^{k}|^2 = 1 \) yield the following estimates

\[
(6.29)
\rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 (\Delta_i \gamma_N(t_{\ell}, t_i))^2 \leq \rho(n) \mathbb{E} \sum_{m=1}^{\bar{N}} \sup_s |\varepsilon_{m}(x_m)|^2 \delta(N) \sum_{i=1}^{\bar{n}} \sum_{k=1}^{n} |\Delta \varepsilon_{t_i}(x_k)|^2 \delta(N),
\]

\[
(6.30)
\rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 (\Delta_i \theta_N(t_{\ell}, t_i))^2 \leq 2 \left( \rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 (\Delta_i \gamma_N(t_{\ell}, t_i))^2 \right)^{1/2} \times \left( \rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |\delta(N) \varepsilon_{t_i}^T \Delta \varepsilon_{t_i} |^2 \right)^{1/2} \\
+ \rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 \left((\Delta_i \gamma_N(t_{\ell}, t_i))^2 + (\delta(N) \varepsilon_{t_i}^T \Delta \varepsilon_{t_i})^2 \right),
\]

\[
(6.31)
\rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 (\Delta_i \xi_N(t_{\ell}, t_i))^2 \leq C \rho(n) \sup_{t} \|\varepsilon_t\|_{R_N}^2 \delta(N) \sum_{r=1}^{d} \sum_{i=1}^{\bar{n}} |\Delta Z_{t_i}^r|^2 \|\lambda_r\|_{R_N}^2 \delta(N)
\]

\[
(6.32)
\rho(n) \sum_{i=1}^{\bar{n}} \sum_{\ell=1}^{\bar{n}} |y_{i\ell}^{k}|^2 (\Delta_i \eta_N(t_{\ell}, t_i))^2 \leq C \rho(n) \sum_{i=1}^{\bar{n}} \|\Delta \varepsilon_{t_i}\|_{R_N}^2 \delta(N) \sum_{r=1}^{d} \sup_{t} |Z_{t_i}^r|^2 \|\lambda_r\|_{R_N}^2 \delta(N)
\]

The estimates (6.29), (6.30), (6.31) and (6.32) allow us to conclude that \( T_1(n, N) = O_P(\rho(n)) \).

**Analysis of \( T_2(n, N) \).** The estimates for the crossing terms are more involved. Let us split \( T_2(n, N) \) according to the terms \( \Delta_i \gamma_N(t_{\ell}, t_i), \Delta_i \theta_N(t_{\ell}, t_i), \Delta_i \xi_N(t_{\ell}, t_i) \) and \( \Delta_i \eta_N(t_{\ell}, t_i) \) as follows. To shorten notation, in the sequel we denote \( J(k, n) = \rho(n) \sum_{1 \leq \ell < s \leq \bar{n}} |y_{i\ell}^{k} y_{is}^{k}| \). Cauchy-Schwartz inequalities and routine algebraic manipulations yield the following estimates

\[
\rho(n) \sum_{i=1}^{\bar{n}} \sum_{1 \leq \ell < s \leq \bar{n}} |y_{i\ell}^{k} \Delta_i \gamma_N(t_{\ell}, t_i) y_{is}^{k} \Delta_i \xi_N(t_{s}, t_i)| \leq C J(k, n) \left( \sup_{t} \|\varepsilon_t\|_{R_N}^2 \delta(N) \right)^{1/2} \times \left[ (\mathbb{E} \sup_{t} \|\varepsilon_t\|_{R_N}^2 \delta(N)) \sum_{i=1}^{\bar{n}} \sum_{r=1}^{d} |\Delta Z_{t_i}^r|^2 \|\lambda_r\|_{R_N}^2 \delta(N) \right]^{1/2},
\]
\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} |y_i^k \Delta \xi_N(t_i, t_s)y_i^k \Delta \eta_N(t_s, t_i)| \leq C J(k, n) \sum_{r, q=1}^{P} O_{P(r, q)}(1) \times \left[ \sum_{i=1}^{\bar{n}-1} \| \Delta \varepsilon_{t_i} \|_{\mathbb{R}^N}^2 \delta(N) \right]^{1/2},
\]

\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} |y_i^k \Delta \theta_N(t_i, t_s)y_i^k \Delta \xi_N(t_s, t_i)| \leq C J(k, n) \sum_{r=1}^{P} O_{P(r)}(1) \sup_{t} \| \varepsilon \|_{\mathbb{R}^N} \times \left( \sum_{i=1}^{\bar{n}-1} |\Delta Z_{t_i}^r|^2 \sum_{i=1}^{\bar{n}-1} \| \Delta \varepsilon_{t_i} \|_{\mathbb{R}^N}^2 \delta(N) \right)^{1/2}.
\]

We also shall write
\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} y_i^k \Delta \gamma_N(t_i, t_s)y_i^k \Delta \gamma_N(t_s, t_i) = \sum_{r, j=1}^{P} O_{P(r, j)}(1) \rho(n) \sum_{1 \leq t < s \leq \bar{n}} y_i^k y_i^k \times \langle \varepsilon_{t_i}, \lambda_r \rangle_{\mathbb{R}^N} \delta(N) \langle \varepsilon_{t_s}, \lambda_j \rangle_{\mathbb{R}^N} \delta(N)
\]

and
\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} |y_i^k \Delta \gamma_N(t_i, t_s)y_i^k \Delta \gamma_N(t_s, t_i)| \leq J(k, n) O_{P(1)} \sum_{i=1}^{\bar{n}-1} \| \varepsilon_{t_i} \|_{\mathbb{R}^N}^2 \delta(N),
\]

\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} |y_i^k \Delta \gamma_N(t_i, t_s)y_i^k \Delta \gamma_N(t_s, t_i)| \leq C J(k, n) \left( \sum_{i=1}^{\bar{n}-1} \| \Delta \varepsilon_{t_i} \|_{\mathbb{R}^N}^2 \delta(N) \right)^{1/2},
\]

\[
\rho(n) \sum_{i=1}^{\bar{n}-1} \sum_{1 \leq t < s \leq \bar{n}} |y_i^k \Delta \eta_N(t_i, t_s)y_i^k \Delta \eta_N(t_s, t_i)| \leq C J(k, n) \sum_{q, r=1}^{P} O_{P(q, r)} \left( \sum_{i=1}^{\bar{n}-1} \| \Delta \varepsilon_{t_i} \|_{\mathbb{R}^N}^2 \delta(N) \right).
\]

The remainder terms in \( T_2(n, N) \) are analogous. Summing up the above estimates, we conclude that \( T_2(n, N) \to 0 \) in probability as \( n, N \to \infty \). From identities \( 6.26 \), \( 6.27 \) and \( 6.24 \), we conclude the proof.
\[ r_t = \phi_t + \sum_{k=1}^{d} Y_t^k \xi_k; 0 \leq t \leq T. \]

where \( Y = AZ \) and \( Z \) satisfies (6.16).

**Proposition 6.3.** If Assumptions (D1-D2-D3-D4-Q1-Q5) hold then
\[
\sum_{j=1}^{d} \|\hat{\varphi}_j - \xi_j\|_{E}^2 \xrightarrow{p} 0
\]
as \( n, N \to \infty \).

**Proof.** Let us fix \( i \in \{1, \ldots, d\} \). Since \( \hat{d} \) is an integer-valued consistent estimator for \( d \), then we shall assume that \( \hat{d} = d \). Under (D1-D2-D3-D4-Q1), \( \{\xi_i; 1 \leq i \leq d\} \) is a well-defined random basis for \( V \). By the very definition,

\[
\hat{\varphi}_i(x) = \sqrt{\rho(n)} \sum_{k=1}^{\hat{n}} y_{tk}^i X_{tk}(x) = \sum_{m=1}^{d} \left( \sum_{k=1}^{\hat{n}} \sqrt{\rho(n)} y_{tk}^i Z_{tk}^m \right) \lambda_m(x)
\]

\[+ \sum_{k=1}^{\hat{n}} \sqrt{\rho(n)} y_{tk}^i \varepsilon_{tk}(x) =: R_{1,i}(x) + R_{2,i}(x), x \in [a, b]. \]

Let us recall that for any \( f \in E \), we can compute the Sobolev norm as follows
\[ \|f\|_{E}^2 = \sup_{\Pi} \sum_{s_j \in \Pi} \frac{\|\Delta f(s_j)\|_{E}^2}{\Delta s_j} < \infty \]
where the sup is taken over all partitions \( \Pi \) of \( [a, b] \). See e.g Prop 1.45 in Friz and Victoir [28] for further details. If \( \Pi = \{s_j\}_{j=1}^{M} \) is a partition of \( [a, b] \), then

\[
\sum_{s_j \in \Pi} \frac{\|\Delta R_{2,i}(s_j)\|_{E}^2}{\Delta s_j} = \sum_{s_j \in \Pi} \sum_{k=1}^{\hat{n}} \rho(n) |y_{tk}^i|^2 \frac{\|\Delta \varepsilon_{tk}(s_j)\|_{E}^2}{\Delta s_j}
\]

\[+ 2\rho(n) \sum_{1 \leq k < m \leq \hat{n}} y_{tk}^i y_{tm}^i \sum_{s_j \in \Pi} \frac{\Delta \varepsilon_{tk}(s_j)}{\Delta s_j} \frac{\Delta \varepsilon_{tm}(s_j)}{\Delta s_j}
\]

\[= I_{1,i} + I_{2,i} \]

Since \( \rho(n)Y^T(d)\dot{Y}(d) = I_d \) a.s., then (Q5) yields
\[
|I_{1,i}| \leq \sum_{t=1}^{\hat{n}} \rho(n) |y_{t}^i|^2 \|\varepsilon_{t}||_{E}^2 \leq \sup_{0 \leq t \leq T} \|\varepsilon_{t}||_{E}^2 \rho(n) \xrightarrow{p} 0
\]
as \( n \to \infty \). Cauchy-Schwartz inequality and (Q5) yield
\[
|I_{2,i}| \leq 2\rho(n) \sum_{1 \leq k < s \leq \hat{n}} |y_{tk}^i y_{ts}^i| \|\varepsilon_{t}||_{E} \|\varepsilon_{s}||_{E} \xrightarrow{p} 0
\]
as \( n \to \infty \). From Lemma 7.3 we know that \( (G^T)^{-1} = A \) so that \( (A^{-1})^T = G \).

Since \( \{\lambda_1, \ldots, \lambda_d\} \subset E \), then we obviously have \( \|R_{1,i}(\cdot) - \xi_i(\cdot)\|_{E} \xrightarrow{p} 0 \) as \( n \to \infty \). This concludes the proof. \( \square \)

We will now state and prove an elementary lemma from linear algebra.
Lemma 6.2. Let $v_1, \ldots, v_d$ be a set of $d$ linearly independent vectors in a real Hilbert space $H$. Let, also, $V = \text{span}\{v_1, \ldots, v_d\}$, and $T : V \rightarrow V$ be an orthogonal matrix. If $\tau_1, \ldots, \tau_d$ is the Gram-Schmidt orthonormalization of $v_1, \ldots, v_d$ and $w_1, \ldots, w_d$ is the Gram-Schmidt orthonormalization of $Tv_1, \ldots, Tv_d$, then we have

$$w_i = T\tau_i; \quad i = 1, \ldots, d. \quad \Box$$

Proof. Follows directly from observing that, for each $v \in V$, $\|Tv\| = \|v\|$, and for each $u \in V$, $T(\text{Proj}_v u) = \text{Proj}_{Tv}(Tu)$, where $\text{Proj}_v u = v(u,v)/\|v\|$. \hfill \Box

Now we are able to present the main result of this section. In the sequel, $\hat{p}$ is any consistent estimator for $\text{dim } Q$ based on $X$ (see e.g Appendix II).

Theorem 6.1. Let $r$ be the stochastic PDE satisfying (A1-A2) and let us assume that (D1, D2, D3, D4) and (Q1, Q2, Q3, Q4, Q5) hold. For a given $h \in \text{dom } (A)$, let $V^h_t = \phi_t + V; 0 \leq t \leq T$ be the minimal foliation generated by $V$ such that $\tau_0 = h \in V_0^h$. Let $\hat{L} \in M_{d \times d}$ be the matrix whose rows are given by $\hat{L}_i := \hat{v}_i; 1 \leq i \leq \hat{d}$, where $\{\hat{v}_1, \ldots, \hat{v}_d\}$ is an eigenvector set of $[\hat{Y}]_T$ given by (6.21) associated to the ordered eigenvalues $\hat{q}_1 \geq \hat{q}_2 \geq \ldots \geq \hat{q}_d$. Let

$$\hat{N} := \text{span } \{(\hat{L}\hat{\varphi})_{\hat{p}+1}, \ldots, (\hat{L}\hat{\varphi})_d\} \quad \hat{Q} := \text{span } \{(\hat{L}\hat{\varphi})_1, \ldots, (\hat{L}\hat{\varphi})_{\hat{p}}\}. \quad \Box$$

Then,

$$\max \{d(\hat{N}, N), d(\hat{Q}, Q)\} \overset{P}{\rightarrow} 0 \text{ as } n, N \rightarrow \infty. \quad \Box$$

Proof. Let $\hat{N} = \text{span } \{(\hat{L}\hat{\varphi})_{\hat{p}+1}, \ldots, (\hat{L}\hat{\varphi})_d\}$, and $\hat{Q} = \text{span } \{(\hat{L}\hat{\varphi})_1, \ldots, (\hat{L}\hat{\varphi})_{\hat{p}}\}$.

Following the same lines as in the proof of Theorem 5.1 and noting that

$$\Phi(\hat{N}) = \text{Ker}(\hat{[Y]}_T) \quad \text{and} \quad \Phi(N) = \text{Ker}([Y]_T),$$

we obtain

$$d(\hat{N}, N) \overset{P}{\rightarrow} 0, \quad \text{and} \quad d(\hat{Q}, Q) \overset{P}{\rightarrow} 0. \quad \Box$$

as $\hat{n}, \hat{N} \rightarrow \infty$. By using the triangle inequality, we obtain

$$d(\hat{N}, N) \leq d(\hat{N}, \hat{N}) + d(\hat{N}, N),$$

and from equation (6.33), it is enough to prove that $d(\hat{N}, \hat{N}) \overset{P}{\rightarrow} 0$ as $n, N \rightarrow \infty$.

Let $\hat{\varphi}_1, \ldots, \hat{\varphi}_{\hat{p}}; \hat{\xi}_1, \ldots, \hat{\xi}_d$ be as in Proposition 6.3. Let, also, $\pi$ and $\hat{N}$ be large enough so that $\pi = \hat{p}$ and $\hat{d} = d$.

From Lemma 6.2, if we denote by $\tau_1, \ldots, \tau_d$, the Gram-Schmidt orthonormalization of $\xi_1, \ldots, \xi_d$, and by $\hat{\tau}_1, \ldots, \hat{\tau}_d$, the Gram-Schmidt orthonormalization of $\hat{\varphi}_1, \ldots, \hat{\varphi}_d$, we have that $\hat{\tau}_1, \ldots, \hat{\tau}_d$ is the Gram-Schmidt orthonormalization of $\hat{\xi}_1, \ldots, \hat{\xi}_d$, and $\hat{\tau}_1, \ldots, \hat{\tau}_d$ is the Gram-Schmidt orthonormalization of $\hat{\varphi}_1, \ldots, \hat{\varphi}_d$.

Furthermore, from the orthonormalization procedure, for each $k \leq d$,

$$\text{span } \{(\hat{L}\hat{\varphi}_1), \ldots, (\hat{L}\hat{\varphi}_k)\} = \text{span } \{(\hat{L}\hat{\tau}_1), \ldots, (\hat{L}\hat{\tau}_k)\},$$

and

$$\text{span } \{(\hat{L}\xi_1), \ldots, (\hat{L}\xi_k)\} = \text{span } \{(\hat{L}\tau_1), \ldots, (\hat{L}\tau_k)\}.$$

Thus,

$$\hat{N} = \text{span } \{(\hat{L}\hat{\tau}_1), \ldots, (\hat{L}\hat{\tau}_p)\}, \quad \text{and} \quad N = \text{span } \{(\hat{L}\tau_1), \ldots, (\hat{L}\tau_p)\}. \quad \Box$$
Therefore, since $\Phi$ is an isometry, we have

$$d(\hat{N},\tilde{N}) = D(\hat{N},\tilde{N}) = 1 - \frac{1}{d} \sum_{i,j} (\langle (\hat{L}\hat{\tau})_i, (\hat{L}\tau)_j \rangle)^2 \ a.s$$

Let us work with the quantity inside the brackets, and let us introduce some notation: Denote the matrix of $\hat{L}$ by $\{\hat{a}_{ij}\}$, i.e., for any vector $v \in \mathbb{R}^d$,

$$(\hat{L}v)_i := \sum_j \hat{a}_{ij} v_j.$$ 

Note that since the transformation $\hat{L}$ is orthogonal, we have

$$\sum_k \hat{a}_{ik} \hat{a}_{jk} = \delta_{ij} \ a.s.$$ 

Observe that from Proposition 6.3 we have that $\langle \hat{\tau}_i, \tau_j \rangle \xrightarrow{p} \delta_{ij}$ as $\bar{n}, \bar{N} \to \infty$. Since $\hat{L}$ is orthogonal and the set of orthogonal matrices is compact, the set $\{\hat{a}_{i,j}\}$ is uniformly bounded in $n$ and $N$, so that

$$\left| \sum_{k \neq p} \hat{a}_{ik} \hat{a}_{jp} \langle \hat{\tau}_k, \tau_p \rangle \right| \xrightarrow{P} 0,$$

and

$$\left| \sum_k \hat{a}_{ik} \hat{a}_{jk} (\langle \hat{\tau}_k, \tau_k \rangle - 1) \right| \xrightarrow{P} 0$$

as $n, N \to \infty$.

Therefore,

$$\left| \sum_{k,p} \hat{a}_{ik} \hat{a}_{jp} \langle \hat{\tau}_k, \tau_p \rangle - \delta_{ij} \right| = \left| \sum_{k,p} \hat{a}_{ik} \hat{a}_{jp} \langle \hat{\tau}_k, \tau_p \rangle - \sum_k \hat{a}_{ik} \hat{a}_{jk} \right| \leq \left| \sum_k \hat{a}_{ik} \hat{a}_{jk} (\langle \hat{\tau}_k, \tau_k \rangle - 1) \right| + \left| \sum_{k \neq p} \hat{a}_{ik} \hat{a}_{jp} \langle \hat{\tau}_k, \tau_p \rangle \right| \xrightarrow{P} 0,$$

as $n, N \to \infty$, which thus implies that

$$\frac{1}{m} \sum_{i,j} ((\hat{L}\hat{\tau})_i, (\hat{L}\tau)_j)^2 = \frac{1}{m} \sum_{i,j} \left( \sum_{k,p} \hat{a}_{ik} \hat{a}_{jp} \langle \hat{\tau}_k, \tau_p \rangle \right)^2 \xrightarrow{P} 1,$$

and then,

$$d(\hat{N},\tilde{N}) \xrightarrow{P} 0$$

as $n, N \to \infty$. The proof for $\lim_{n,N \to \infty} d(\hat{Q},\tilde{Q}) = 0$ in probability, follows from the same reasoning as in the proof of Theorem 6.1. This concludes the proof. \qed
7. Appendix I: Estimating the dimension of invariant manifolds

In this section, we present proofs of the lemmas related to the estimation of $\text{dim} \ V$, where $V$ generates a finite-dimensional foliation invariant w.r.t. the $(6.1)$. The proofs of Lemmas $[7.1]$ $[7.2]$ $[7.3]$ and $[7.4]$ are inspired by the arguments given by Bai and Ng $[5]$ and Bai $[6]$. In one hand, in contrast to $[5]$ and $[6]$, our asymptotic matrix $\Sigma Z$ is random and the sampling should be in high-frequency. On the other hand, Assumption D1 allows us to prove similar results without significant extra effort. For sake of completeness, we give the details here.

**Lemma 7.1.** If Assumptions $\text{(D1-D2-D3-D4)}$ hold, then

(a) $\rho(n) \sum_{l=1}^{\bar{n}} \hat{Y}_l(d) = O_p\big(\frac{1}{C_n^2}\big)$

(b) $\rho(n) \sum_{l=1}^{\bar{n}} \hat{Y}_l(d) = O_p\big(\frac{1}{C_n^2}\big)$

(c) $\rho(n) \sum_{l=1}^{\bar{n}} \hat{Y}_l(d) = O_p\big(\frac{1}{C_n^2}\big)$

(d) $\rho(n) \sum_{l=1}^{\bar{n}} \hat{Y}_l(d) = O_p\big(\frac{1}{C_n^2}\big)$

Proof. Let $L_{n,N}$ be the diagonal matrix of the eigenvalues of $\rho(n)\delta(N)XX^\top$ arranged in decreasing order. From $(\text{D1-D2-D3-D4})$, one can easily check that $\|\rho(n)\delta(N)XX^\top\|_F = O_p(1)$ and hence $\|L_{n,N}\|_F = O_p(1)$. In this case, the same argument given in the proof of Lemma A1 in $[6]$ allows us to state that

$$C_n^2\left(\rho(n) \sum_{l=1}^{\bar{n}} \|\hat{Y}_l(d) - H_d Z_l\|_2^2\right) = O_p(1)$$

where $H_d := \delta(N)\Lambda^\top \Lambda^\top \hat{Y}(d)\rho(n)L_{n,N}^{-1} \in \mathbb{M}_{d \times d}$. Assumptions $(\text{D1-D2-D3-D4})$ together with $(7.1)$ allow us to repeat the same argument given in the proof of Lemma A2 in $[6]$ to conclude that the statement hold true. We omit the details. □

The next result was enunciated by Bai and Ng $[5]$ in Lemma A3 (in the context of a discrete-time model and deterministic $\Sigma Z$) without a formal proof. For sake of completeness, we give the details here in our context.

**Lemma 7.2.** Let $L_{n,N}$ be the diagonal matrix of the eigenvalues of $\rho(n)\delta(N)XX^\top$ arranged in decreasing order. If Assumptions $(\text{D1-D2-D3-D4})$ hold then

$$L_{n,N} \overset{p}{\rightarrow} \mathcal{C} := \text{diag}(c_1, \ldots, c_d)$$

as $n, N \rightarrow \infty$, where $(c_1, \ldots, c_d)$ are the eigenvalues (in decreasing order) of $\Sigma \Lambda \Sigma Z$.

Proof. We follow closely the idea contained in the proof of Proposition 1 in $[6]$. By the very definition, $\rho(n)\delta(N)XX^\top \hat{Y}(d) = \hat{Y}(d)L_{n,N}$ a.s and hence

$$\left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \rho(n)Z^\top \left(\rho(n)\delta(N)XX^\top \hat{Y}(d)\right) = \left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \rho(n)Z^\top \hat{Y}(d) = \left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \rho(n)Z^\top \hat{Y}(d)$$

From the identity $\mathcal{X} = Z\Lambda^\top + \mathcal{E}$, we actually have

$$\left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \left(\rho(n)Z^\top \mathcal{Z}\right) = \left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \left(\rho(n)Z^\top \hat{Y}(d)\right) + s_{nN} = \left(\delta(N)\Lambda^\top \Lambda\right)^{1/2} \left(\rho(n)Z^\top \hat{Y}(d)\right) L_{n,N}$$

(7.2)
where
\[
\begin{align*}
    s_{n,N} &:= (\delta(N)\Lambda^\top\Lambda)^{1/2} \left[ \rho(n) (Z^\top Z) \rho(n) \delta(N) \Lambda^\top \hat{Y}(d) + \rho(n) \delta(N) Z^\top \Sigma Z \rho(n) \right] \\
    &\quad + \rho(n) \delta(N) Z^\top \Sigma \Sigma Z \rho(n) = o_p(1) 
\end{align*}
\] (7.3)
due to Lemma 7.1. Let \( U_{n,N} := (\delta(N)\Lambda^\top\Lambda)^{1/2} (\rho(n) Z^\top Z) (\delta(N)\Lambda^\top\Lambda)^{1/2} \) and \( E_{n,N} := (\delta(N)\Lambda^\top\Lambda)^{1/2} \left( \rho(n) Z^\top \hat{Y}(d) \right) \). We shall write (7.2) as follows

\[
[U_{n,N} + s_{n,N} E_{n,N} E_{n,N}^+] E_{n,N} = E_{n,N} L_{n,N}
\]
where \( E_{n,N}^+ \) is the pseudoinverse of \( E_{n,N} \). Then each column of \( E_{n,N}^+ \) is an eigenvector of \( N_{n,N} + s_{n,N} E_{n,N} E_{n,N}^+ \). Since \( E_{n,N} E_{n,N}^+ = O_p(1) \) then (7.3) and Assumptions (D1, D2) yield

\[
\|U_{n,N} + s_{n,N} E_{n,N} E_{n,N}^+ - \Sigma^1/2 \Sigma Z \Sigma^1/2\| \xrightarrow{P} 0
\]
as \( n, N \to \infty \). By the continuity of the eigenvalues, we do have \( \|L_{n,N} - C\|_F \xrightarrow{P} 0 \) as \( n, N \to \infty \). Since \( \Sigma^1/2 \Sigma Z \Sigma^1/2 \) and \( \Sigma \Lambda \Sigma Z \) have the same random eigenvalues, we conclude the proof. \( \Box \)

In the sequel, an eigenvector matrix of \( W \) refers to the matrix whose columns are the eigenvectors of \( W \) with unit length and the \( i \)th column corresponds to the \( i \)th largest eigenvalue.

**Lemma 7.3.** Under Assumptions (D1-D2-D3-D4-Q1), for every \( j = 1, \ldots, d \), there exists a random vector \((G_{1j}, \ldots, G_{dj})\) such that

\[
\left( \sum_{i=1}^\infty \sqrt{\rho(n) g_{i1}^j Z_{i1}} \ldots \sum_{i=1}^\infty \sqrt{\rho(n) g_{id}^j Z_{id}} \right) \xrightarrow{P} (G_{1j}, \ldots, G_{dj})
\]
as \( n, N \to \infty \). Moreover, the matrix \( G := (G_{ij})_{1 \leq i, j \leq d} \) is invertible a.s and it is given by \( G = C^{1/2} \Phi^\top \Sigma^1/2 \) and \( \Phi \) is the eigenvector matrix related to \( C \) subject to \( \Phi^\top \Phi = I_d \) a.s.

**Proof.** By using Lemma 7.2, the proof is identical to Proposition 1 in Bai [6] even in the case when \( \Sigma Z \) is random. We refer the reader to the discussion in page 162 in [6]. \( \Box \)

In the sequel, \( kmax \) is a finite integer such that \( d \leq kmax \).

**Lemma 7.4.** Let us assume that assumptions (D1,D2, D3, D4, Q1) hold and let \( \hat{d} = \arg \min_{1 \leq k \leq kmax} PC(k) \). Then \( \lim_{n,N \to \infty} \mathbb{P}[\hat{d} = d] = 1 \) if (i) \( q(N, n) \to 0 \) and (ii) \( C_{nN} q(N, T) \to \infty \) as \( n, N \to \infty \) where \( C_{nN} = \min \{ \delta(N)^{-1/2}\rho(n)^{-1/2} \} \).

**Proof.** The same arguments given in the proof of Theorem 1 in Bai and Ng [5] apply in our context. In particular, Lemmas 2, 3 and 4 in Bai and Ng [5] can be similarly proved in our context as well by using Assumptions (D1, D2, D3, D4, Q1). In particular, the fact that \( \Sigma Z \) is not deterministic is not essential for the validity of Lemmas 2, 3 and 4 as long as \( \text{rank } \Sigma Z = d \) a.s (Assumption D1). In particular, for \( k < d \) let us define \( \hat{J}_k^\top := \hat{Y}^\top (k) Z \rho(n) \Lambda^\top \Lambda \delta(N) \in \mathbb{M}_{k \times d} \). In our
context, Lemma 3 in [3] can be written as follows: There exists \( \tau_k > 0 \) a.s such that

\[
\liminf_{n,N \to \infty} V(k, ZJ_k) - V(d, Z) = \text{tr}(R_k, \Sigma) =: \tau_k
\]

in probability, where \( R_k := \Sigma_Z - \Sigma_Z \Sigma_k (\Sigma_k^\top \Sigma_Z)^{-1} \Sigma_k^\top \Sigma_Z \) and \( \Sigma_k := \lim_{n,N \to \infty} J_k \) exists due to Lemma 7.3. By construction rank \( \Sigma_k = k < d \) a.s. Assumptions (D1-D2) yield \( \text{tr}(R_k, \Sigma) > 0 \) a.s. By writing

\[
PC(k) - PC(d) = V(k, \hat{Y}(k)) - V(d, \hat{Y}(d)) - (d - k)q(n, N)
\]

and splitting

\[
V(k, \hat{Y}(k)) - V(d, \hat{Y}(d)) = [V(k, \hat{Y}(k)) - V(k, ZJ_k)] + [V(k, ZJ_k) - V(d, ZJ_d)]
\]

\[+ [V(d, ZJ_d) - V(d, \hat{Y}(d))],\]

we shall use the same argument in the proof of Th 1 in [5] to conclude that

\[(7.4) \quad \lim_{n,N \to \infty} \mathbb{P}\{PC(k) < PC(d)\} = 0,\]

for each \( k < d \). If \( k_{max} \geq k \geq d \), then similar to Lemma 4 in [5], Assumptions (D1,D2, D3, D4, Q1) yield \( V(k, \hat{Y}(k)) - V(d, \hat{Y}(d)) = O_p(C_{nN}^{-2}) \). The rest of the proof is identical to the proof of Th 1 in [5], so we omit the details. \( \square \)

8. Appendix II: Estimating \( \dim \mathbb{Q} \)

In this section, we give a concrete alternative for estimating \( p = \dim \mathbb{Q} \) which is an important step in Theorem 6.1. We choose to work with the Fourier-type estimator proposed by Malliavin and Mancino [33] but we stress that other quadratic variation estimators can be certainly used as well. In particular, we stress that the results of Section 6.3 do not depend on the particular choice for the estimator of \( p \). Nevertheless, it is instructive to provide a concrete estimator based on \( X \).

We assume that the continuous-time version of \( X \) given by (6.18) is an \( E \)-valued \( \text{Itô semimartingale field of the form} \)

\[(8.1) \quad X_t(x) = X_0(x) + \int_0^t U_s(x) ds + \sum_{i=1}^q \int_0^t V^i_s(x) dB^i_s, \quad (t, x) \in [0, T] \times [a, b]\]

where \( U(x) \) and \( V^i(x) \) are adapted processes for every \( x \in [a, b] \) and they satisfy the usual integrability hypothesis to ensure that it is indeed a well-defined semimartingale random field.

The strategy is to find the minimum requirements on the residual process \( \varepsilon \) in (6.17) in such way that one can estimate the random operator \( Q_T \) via an observed curve process \( X \) of the form (8.1). If \( \varepsilon \) is negligible in the quadratic variation sense, then the method of the estimation of the kernel \( Q_T(u, v) \) is fully based on any reasonable non-parametric estimator of the integrated volatility.

In the sequel, without any loss of generality we assume that \( [0, T] = [0, 2\pi] \). Let \( \Pi = \{t^n_i; i = 0, \ldots, n \} \) be the instant times of observations and \( \rho(n) = \max_{0 \leq h \leq n-1} |t^n_{h+1} - t^n_h| \). In this section, we assume that \( \rho(n) \to 0 \) as \( n \to \infty \).
so we are able to sample the curves \( x \mapsto X_t(x) \) in high-frequency in time. In the sequel, we make use of the following notation

\[
\varphi_n(t) := \sup\{ t_k^n; t_k^n \leq t \}
\]

For given positive integers \( M \geq 1 \) and \( n \geq 1 \), we define

\[
\hat{Q}_T(u, v) := \frac{1}{2M+1} \sum_{|s| \leq M} \int_0^T e^{i s \varphi_n(t)} dX_t(u) \int_0^T e^{-i s \varphi_n(t)} dX_t(v)
\]

\[
= \sum_{\ell, j=0}^{n-1} d_M(t^n_\ell - t^n_j) \Delta X_{t^\ell_{\ell+1}}(u) \Delta X_{t^j_{j+1}}(v), \quad u, v \in [a, b],
\]

where \( \Delta X_{t^\ell_{\ell+1}}(u) := X_{t^\ell_{\ell+1}}(u) - X_{t^\ell_{\ell}}(u) \) and

\[
d_M(t) := \frac{1}{2M+1} \sum_{|s| \leq M} e^{i s t} = \begin{cases} 1; \ t = sT, \ s \in \mathbb{Z} \\ \frac{1}{2M+1} \sin((M+1/2)t); \ t \neq sT \end{cases}
\]

is the normalized Dirichlet kernel. Here \( M \) encodes the Bohr convolution product and \( n \) the discretization level of the Fourier transform of \( \sigma_t(u, v) \). See Malliavin and Mancino [33] for more details.

To keep notation simple, from now on we set \( t_i := t^n_i, 0 \leq i \leq \bar{n} \). The kernel \( \hat{Q}_T(u, v) \) induces a random linear operator \( \hat{Q}_T \) on the complexification \( E_\mathbb{C} \) as follows

\[
(\hat{Q}_T f)(u) = \langle \hat{Q}_T(u, \cdot), f \rangle_{E_\mathbb{C}}
\]

\[
= \frac{1}{2M+1} \sum_{|s| \leq M} \sum_{\ell=0}^{n-1} \sum_{k=0}^{n-1} \exp(is(t_k - t_\ell)) \Delta X_{t_{\ell+1}}(u) \langle \Delta X_{t_{\ell+1}}, f \rangle_{E_\mathbb{C}},
\]

for \( f \in E_\mathbb{C} \). The reason to consider \( E \) on the field \( \mathbb{C} \) is due to a nice representation as follows. If \( A \) and \( B \) are two linear operators on Hilbert spaces then \( AB \) and \( BA \) share the same nonzero eigenvalues. Furthermore, if \( \gamma \) is an eigenvector of \( BA \), \( A\gamma \) is an eigenvector of \( AB \) with the same eigenvalue. So the strategy is to write \( \hat{Q}_T = AB \) in such way that \( BA : \mathbb{C}^p \to \mathbb{C}^p \) for some \( p \geq 1 \) and therefore one can easily relate the eigenvalues of \( BA \) to \( AB \). In fact, by the very definition of \( \hat{Q}_T \) we have

\[
(8.2) \quad \hat{Q}_T = AB
\]

where \( B : E_\mathbb{C} \to \mathbb{C}^{2M+1} \) is defined by

\[
B(\cdot) := \left( \sum_{\ell=0}^{n-1} \exp(-i(M)t_\ell) \langle \Delta X_{t_{\ell+1}}, \cdot \rangle_{E_\mathbb{C}}, \ldots, \sum_{\ell=0}^{n-1} \exp(-i(M)t_\ell) \langle \Delta X_{t_{\ell+1}}, \cdot \rangle_{E_\mathbb{C}} \right)
\]

and \( A : \mathbb{C}^{2M+1} \to E_\mathbb{C} \) is defined by
(8.3) \((Ax)() := \frac{1}{2M + 1} \sum_{|s| \leq N} \sum_{k=0}^{n-1} x_s \exp(ist_k) \Delta X_{t_{k+1}}() ; x \in \mathbb{C}^{2M+1} \).

By the very definition, \(\bar{Q}T := BA : \mathbb{C}^{2M+1} \rightarrow \mathbb{C}^{2M+1} \) is given componentwise by

\[
\bar{Q}Ty = \frac{1}{2M + 1} \sum_{|s| \leq M} \sum_{k,\ell=1}^{n-1} y_s \exp{(i(st_k - mt_\ell))} \langle \Delta X_{t_{\ell+1}}, \Delta X_{t_{k+1}} \rangle_{\mathbb{E}C},
\]

for \(y \in \mathbb{C}^{2M+1} \), \(m = -M, \ldots, M \). We then arrive at the following elementary result.

**Lemma 8.1.** The random linear operators \(Q_T\) and \(\bar{Q}_T\) share the same nonzero eigenvalues in \(\mathbb{C} \). Let \(\hat{p}\) be the number of nonzero eigenvalues \(\{\hat{\theta}_j = 1, \ldots, \hat{p}\}\) of \(\bar{Q}_T\) and let \(\gamma_j = (\gamma_j(-M), \ldots, \gamma_j(M))\), \(j = 1, \ldots, \hat{p}\) be the corresponding eigenvectors in \(\mathbb{C}^{2M+1} \). Then

\[
\frac{1}{2M + 1} \sum_{|s| \leq M} \gamma_j(s) \left( \sum_{k=0}^{n-1} \exp{(ist_k)} \Delta X_{t_{k+1}} \right), \quad j = 1, \ldots, \hat{p}
\]

are the \(\hat{p}\) eigenfunctions of \(\hat{Q}_T\).

**Proof.** Let \(\hat{\theta}_j \in \mathbb{C}\) be a nonzero eigenvalue of \(\bar{Q}_T\) and let \(\gamma_j \in \mathbb{C}^{2M+1}\) be the corresponding eigenvector. Then

(8.4) \(\bar{Q}_T\gamma_j = \hat{\theta}_j \gamma_j\) and \(\hat{Q}_T(A\gamma_j) = \hat{\theta}_j A\gamma_j \) a.s,

where \(A\) is the operator given by (8.3). By writing (8.4) component by component we have

\[
\frac{1}{2M + 1} \sum_{|s| \leq M} \sum_{k,\ell=0}^{n-1} \gamma_j(s) \exp{(i(st_k - rt_\ell))} \langle \Delta X_{t_{\ell+1}}, \Delta X_{t_{k+1}} \rangle_{\mathbb{E}C} = \hat{\theta}_j \gamma_j(r) \) a.s
\]

for \(r = -M, \ldots, M\). On the other hand,

\[
A\gamma_j = \frac{1}{2M + 1} \sum_{|s| \leq M} \gamma_j(s) \left( \sum_{k=0}^{n-1} \exp{(ist_k)} \Delta X_{t_{k+1}} \right), \quad j = 1, \ldots, \hat{p} \) a.s.
\]

This concludes the proof of the Lemma. \(\square\)

**Remark 8.1.** Let

(8.5) \[
\frac{1}{2M + 1} \sum_{|s| \leq M} \gamma_j(s) \left( \sum_{k=0}^{n-1} \exp{(ist_k)} \Delta X_{t_{k+1}} \right), \quad j = 1, \ldots, \hat{p} \) a.s
\]

be the eigenvectors of \(\bar{Q}_T\) related to its nonzero eigenvalues \(\{\hat{\theta}_j ; j = 1, \ldots, \hat{p}\}\). Since \(\bar{Q}_T\) is a self-adjoint finite-rank operator then the following spectral decomposition holds a.s
\[
\hat{Q}_T f = \sum_{i=1}^{\hat{p}} \hat{\theta}_i \langle f, \hat{\varphi}_i \rangle_{E_C} \hat{\varphi}_i; \quad f \in E_C,
\]

where \( \hat{p} \leq 2M + 1 \) a.s for every \( n, M \) and \( \{ \hat{\varphi}_i; i = 1, \ldots, \hat{p} \} \) is an orthonormal set by applying a Gram-Schmidt algorithm to the functions given by (8.5).

Let us now introduce the basic assumptions on the residual process \( \varepsilon \). Since the estimation is based on a high-frequency sampling we need to impose some structure on the continuous-time dynamics.

**(B1)** The residual process is an Itô semimartingale field of the form (8.1) where the drift component satisfies

\[
\sup_{0 \leq t \leq T} \| h_t \|_E \in L^p
\]

for every \( p > 1 \).

**(B2)** The quadratic variation of \( \varepsilon(u) \) at time \( T \) satisfies

\[
\int_{\mathbb{R}} [\varepsilon(u), \varepsilon(u)]_T \mu(du) = 0 \text{ a.s.}
\]

**(B3)** The following growth assumption holds

\[
0 < \liminf_{n, M \to \infty} M \rho(n) \leq \limsup_{n, M \to \infty} M \rho(n) < \infty.
\]

**(H1)** The vector fields \( F, \sigma^i : E \to E \) are globally Lipschitz for each \( i = 1, \ldots, m \).

**(H2)** Linear growth condition on the vector fields \( F, \sigma^1; 1 \leq i \leq m \) in (6.1): There exists a constant \( C > 0 \) such that

\[
\| F(x) \|_E^2 + \sum_{i=1}^m \| \sigma^i(x) \|_E^2 \leq C^2 (1 + \| x \|_E^2)
\]

for every \( x \in E \).

**Remark 8.2.** As far as the consistency problem of the HJM model (see Section 4.2), Assumption (B2) means that the initial fitting method used to interpolate points which generates \( X \) cannot introduce an extrinsic volatility. The interpolation must be chosen in such way that the resulting observed volatility on the whole curve must be fully dictated by the market and not to the particular choice of fitting. See also Assumption (Q4). The semimartingale decomposition yields the following structure on the residual process

\[
\varepsilon_t = e + \int_0^t h_s ds
\]

for some \( \mathcal{F}_0 \)-measurable random variable \( e = X_0 - r_0 \) and an integrable adapted process \( h \) satisfying (B1). Assumption (B3) is a technical assumption in order to get optimal bounds but it is also linked with different flavors between the exact Fourier estimator and the usual quadratic variation estimator for \( Q_T \). See Malliavin and Mancino [33] and Clement and Gloter [14] for further details.
Under \((H1)\) and \((H2)\), for every initial condition \(\xi \in E\) there exists a unique mild solution \(r^\xi_t\) of (6.1). Moreover, the following integrability property holds

\[
\mathbb{E} \sup_{0 \leq t \leq T} \|r^\xi_t\|^q_E < \infty
\]

for every \(q > 1\) and \(\xi \in E\). See e.g \([17], \text{Th 7.2}\).

The following result is a functional version of (and almost straightforward consequence) of Proposition 1 and Lemma 3 in \([14]\). In the sequel, \(\| \cdot \|_{HS}\) is the Hilbert-Schmidt norm operator over \(E_C\) and to keep notation simple, we write \(\| \cdot \| = \| \cdot \|_{E_C}\).

**Proposition 8.1.** Assume that \((A1, A2, B1, B2, B3, H1, H2)\) hold and in addition

\[
\mathbb{E}^{1/2} \sup_{0 \leq t \leq T} |\partial_v \sigma^j(r_t)(\cdot)|^4 \in L^1(\mu)
\]

for each \(j = 1, \ldots, m\). Then

\[
\mathbb{E}\|Q_T - \hat{Q}_T\|^2_{HS} = O(\rho(n)).
\]

**Proof.** In the sequel, we denote by \(C\) a positive constant which may differ from line to line. We also decompose

\[
X_t(u) = \tilde{r}_t(u) + \tilde{\varepsilon}_t(u),
\]

into

\[
\tilde{r}_t(u) := \sum_{j=1}^m \int_0^t \sigma^j(r_s)(u)dB^j_s; \ 0 \leq t \leq T, \ u \in [a,b],
\]

\[
\tilde{\varepsilon}_t(u) = \int_0^t \xi_s(u)ds
\]

where

\[
\xi_t(u) := Ar_t(u) + F(r_t)(u) + h_t(u); 0 \leq t \leq T, \ u \in [a,b].
\]

For a given \((u, v) \in [a, b] \times [a, b]\), integration by parts and \((B2)\) yield

\[
Q_T(u, v) - Q_T(u, v) = \int_0^T \left( \int_0^t d_{M,n}(\ell, t)dX_\ell(u) \right)dX_t(v)
\]

\[
= \int_0^T \left( \int_0^t d_{M,n}(\ell, t)dX_\ell(v) \right)dX_t(u)
\]

\[
= R_{n,M}(u, v)
\]

where

\[
d_{M,n}(\ell, t) := d_M(\varphi_n(\ell) - \varphi_n(t)),
\]

\[
R_{n,M}(u, v) = J_{n,M,1}(u, v) + J_{n,M,2}(u, v) + \sum_{i=1}^3 I_{n,M,i}(u, v) + \sum_{i=1}^3 \hat{I}_{n,M,i}(u, v),
\]
and

\[ J_{n,M,1}(u, v) := \int_0^T \left( \int_0^t d_{M,n}(\ell, t)d\hat{\xi}(u) \right)d\hat{\xi}(v), \]

\[ J_{n,M,2}(u, v) := \int_0^T \left( \int_0^t d_{M,n}(\ell, t)d\xi(\ell) \right)d\xi(u), \]

\[ I_{n,M,1}(u, v) := \int_0^T \left( \int_0^t d_{M,n}(\ell, t)d\hat{\xi}(u) \right)d\hat{\xi}(v), \]

\[ I_{n,M,2}(u, v) := \int_0^T \left( \int_0^t d_{M,n}(\ell, t)d\hat{\xi}(u) \right)d\hat{\xi}(v), \]

\[ I_{n,M,3}(u, v) := \int_0^T \left( \int_0^t d_{M,n}(\ell, t)d\hat{\xi}(u) \right)d\hat{\xi}(v), \]

where \( \hat{I}_{n,M,i} \) are the symmetric quantities w.r.t \( I_{n,M,i} \). By the very definition

\[ ||Q_T - \hat{Q}_T||_{HS}^2 = ||(Q_T - \hat{Q}_T)(0, \cdot)||^2 + \int_{[a,b]} \left\| \partial_u (Q_T - \hat{Q}_T)(u, \cdot) \right\|^2 \mu(du) \]

\[ = |Q_T(0,0) - \hat{Q}_T(0,0)|^2 + \int_{[a,b]} |\partial_u Q_T(0, v) - \partial_u \hat{Q}_T(0, v)|^2 \mu(du) \]

\[ + \int_{[a,b]} |\partial_u Q_T(u, 0) - \partial_u \hat{Q}_T(u, 0)|^2 \mu(du) \]

\[ + \int_{[a,b]^2} |\partial^2_{uu} Q_T(u, v) - \partial^2_{uu} \hat{Q}_T(u, v)|^2 \mu(du) \mu(dv) \]

\[ =: T_1(n, M) + T_2(n, M) + T_3(n, M) + T_4(n, M). \]

**Step 1:** The term \( T_1 \). By the invariance hypothesis (A1), we know that [see [41]; Corollary 2.13] \( V \subset dom(A) \) so we may consider \( (dom(A), A) \) as a bounded operator restricted to \( V \). Moreover, we shall represent

\[ r_t = p_{V^p}r_t + p_{V^h}r_t = p_{V^p}V_t^h + p_{V^h}r_t, \]

where \( p \) is the usual projection and \( h = r_0 \). From Theorem 2.11 in [41], we also know that \( t \to A(\pi_{V^p} V^h_t) \) is continuous and therefore there exists a constant \( C \) such that \( ||Ar_t|| \leq C + C||r_t|| \) for every \( t \in [0, T] \). Based on these facts, we may use the linear growth conditions (H1-H2) and (B1) to arrive at the following estimate

\[ \sup_{0 \leq \ell \leq T} ||\xi_{\ell}|| \leq C + \sup_{0 \leq \ell \leq T} ||r_{\ell}|| + C \sup_{0 \leq \ell \leq T} ||h_{\ell}||. \]

In this case, one can easily check that assumptions in Proposition 1 of [14] hold trivially and all their estimates as well. In this case, we have

\[ T_1(n, M) \leq C \int_0^T \int_0^t d_{M,n}(\ell, t)d\ell dt. \]

Lemma 3 in [14] yields \( T_1(n, M) = O(\rho(n)) \).

**Step 2:** The term \( T_2 + T_3 \). Let us now treat \( T_2(n, M) + T_3(n, M) \). For a given \( (i, j) \in \{1, \ldots, m\}^2 \), Burkholder-Davis-Gundy inequality and (H1-H2) yield
The drift part is estimated as follows. Cauchy-Schwartz and Burkholder-Davis-Gundy inequalities, the estimate \( SS \), \((H1-H2)\), \(A1, B1\) and Lemma 3 in \cite{14} yield

\[
\mathbb{E} \int_{[a,b]} \left| \int_0^T \int_0^t d_{M,n}(\ell, t) \sigma^j(\eta_t) \left( 0 \right) dB_{i}^{j}(\ell) \partial_{v} \sigma^i(\eta_t) \left( v \right) dB_{i}^{i}(\ell) \right|^2 d\mu(dv) \leq C \int_0^T \int_0^t d_{M,n}^2(\ell, t) d\ell dt
\]

This yields \( \mathbb{E} |\partial_{v} J_{n,M,1}(0, v)|^2 d\mu(dv) = O(\rho(n)) \). The same argument also holds for \( \partial_{v} J_{n,M,2}(0, v) \) and we conclude that

\[
\int_{[a,b]} \mathbb{E} |\partial_{v} J_{n,M,1}(0, v) + \partial_{v} J_{n,M,2}(0, v)|^2 d\mu(dv) = O(\rho(n))
\]

The drift part is estimated as follows. Cauchy-Schwartz and Burkholder-Davis-Gundy inequalities, the estimate \( SS \), \((H1-H2)\), \(A1, B1\) and Lemma 3 in \cite{14} yield

\[
\int_{[a,b]} \mathbb{E} |\partial_{v} I_{n,M,1}(0, v)|^2 d\mu(dv) \leq \mathbb{E} \sum_{i=1}^d \int_0^T \left( \int_0^t d_{M,n}^2(\ell, t) dt \times \int_0^t \|\xi_i\|^2 dt \right) d\ell \times \|\sigma^i(\eta_t)\|^2 dt \\
\leq C \int_0^T \int_0^t d_{M,n}^2(\ell, t) d\ell dt = O(\rho(n)).
\]

The term \( \partial_{v} I_{n,M,2}(0, v) \) is more evolved but we can repeat the same steps as in the proof of Theorem 1 in \cite{14} in page 1114 to represent

\[
|I_{n,M,2}(u, v)|^2 = \int_{[0,T]^2} Y_{n,M}(u, t, t) \xi_t(v) Y_{n,M}(u, t', t') \xi_{t'}(v) dt dt',
\]

where

\[
Y_{n,M}(u, t, s) := \int_0^s d_{M,n}(\ell, t) d\ell(u).
\]

We fix \( \eta > 0 \) and we split \( |I_{n,M,2}(0, v)|^2 = A_{n,M,1}(u, v, \eta) + A_{n,M,2}(u, v, \eta) \) so that

\[
\partial_{v} A_{n,M,1}(0, v, \eta) := \int_0^T \int_{t-\eta}^t Y_{n,M}(0, t, t) \partial_v \xi_t(v) Y_{n,M}(0, t', t') \partial_v \xi_{t'}(v) dt' dt
\]

\[
\partial_{v} A_{n,M,2}(0, v, \eta) := \int_0^T \int_{0-\eta}^{t-\eta} Y_{n,M}(0, t, t) \partial_v \xi_t(v) Y_{n,M}(0, t', t') \partial_v \xi_{t'}(v) dt' dt
\]

By applying the same arguments in the proof of Theorem 1 in \cite{14} with small \( \eta \) together with Cauchy-Schwartz inequalities on \( E \) and assumptions \((H1-H2)\), \(B1, B3\), we also have

\[
\int_{[a,b]} \mathbb{E} |I_{n,M,2}(0, v)|^2 d\mu(dv) = O(\rho(n)).
\]

Moreover, \((B1,B3)\) and Lemma 3 in \cite{14} yield

\[
\int_{[a,b]} \mathbb{E} |I_{n,M,3}(0, v)|^2 d\mu(dv) \leq C \int_0^T \int_0^t d_{M,n}^2(\ell, t) d\ell dt \leq C \rho(n).
\]
By the symmetry of the other terms, these estimates allow us to conclude that $T_2(n, M) + T_3(n, M) = O(\rho(n))$.

**Step 3:** The term $T_4$. For given $(i, j) \in \{1, \ldots, m\}^2$, Burkholder-Davis-Gundy and Cauchy-Schwartz inequalities and (H1-H2) with (8.6) yield

$$
\mathbb{E} \int_{[a,b]^2} |\partial^2_{vu} J_{n,M,1}(u,v)|^2 \mu(du) \mu (dv) = \int_{[a,b]} \mathbb{E} \int_0^T \left( \int_0^t d_{M,n}(\ell,t) \partial_u \sigma^i(r_v)(u) dB_t^i \right)^2 \| \sigma^i(r_v) \|^2 dt \mu (du) 
\leq C \int_0^T \int_0^t d_{M,n}(\ell,t) d\ell dt \int_{[a,b]} \mathbb{E}^{1/2} \sup_{0 \leq \ell \leq T} |\partial_u \sigma^i(r_v)(\cdot)|^4 \mu (du).
$$

Therefore, by the symmetry of the martingale terms and Lemma 3 in [14] we have

$$
\int_{\mathbb{R}_+^4} \mathbb{E} |\partial^2_{vu} J_{n,M,1}(u,v) + \partial^2_{vu} J_{n,M,2}(u,v)|^2 \mu (dv) = O(\rho(n)).
$$

Similarly, Lemma 3 in [14] and (8.8) yield

$$
\int_{\mathbb{R}_+^4} \mathbb{E} |\partial^2_{vu} I_{n,M,1}(u,v)|^2 \mu (du) \mu (dv) \leq \int_0^T \int_0^t d_{M,n}(\ell,t) d\ell dt 
\times \mathbb{E} \sup_{0 \leq \ell \leq T} \| \xi_{\ell} \|^2 
\leq C \int_0^T \int_0^t d_{M,n}(\ell,t) d\ell dt = O(\rho(n)).
$$

Summing up all the inequalities for $T_1$, $T_2$, $T_3$ and $T_4$, we conclude the proof. □

In the sequel, for each $\epsilon > 0$, we define

$$
\hat{p}^\epsilon := \text{number of non-zero eigenvalues of } \tilde{Q}_T \text{ greater than } \epsilon \text{ a.s}
$$

**Corollary 8.1.** Assume that Assumptions in Proposition 8.1 hold and let $Q = \text{Range } Q_T$ with dimension $p$. Let $\epsilon \to 0$ in such a way that $\epsilon^2 p(n)^{-1} \to \infty$ as $n \to \infty$. Then, $P(\hat{p}^\epsilon \neq p) = O(\epsilon^{-2} \rho(n))$.

**Proof.** From Proposition 8.1 we know that $\mathbb{E} \|Q_T - \tilde{Q}_T\|^2_{HS} = O(\rho_n)$. Since we are considering the ordered eigenvalues, $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \cdots \geq 0$, we have that $\{\hat{p}^\epsilon > p\} = \{\hat{\theta}_{p+1} > \epsilon\}$.

A simple calculation on the Hilbert-Schmidt norm (see e.g [20] p. 1091) together with $\hat{\theta}_{p+1} = 0$ yield

$$
\hat{\theta}_{p+1} = |\hat{\theta}_{p+1} - \theta_{p+1}| \leq \| \tilde{Q}_T - Q_T \|_{HS}.
$$

Therefore,

$$
P(\hat{p}^\epsilon > p) \leq \epsilon^{-2} \mathbb{E} \| \tilde{Q}_T - Q_T \|^2_{HS} = O(\epsilon^{-2} \rho(n)).
$$

The other inequality follows analogously. Since $P(\hat{p}^\epsilon \neq p) = P(\hat{p}^\epsilon > p) + P(\hat{p}^\epsilon < p)$, the result follows. □
REFERENCES

[1] Ait-Sahalia, Y. and Xiu, D. Increased correlation among asset classes: Are volatility or jumps to Blame, or both? Chicago Booth Paper No. 14-11. Preprint.
[2] Ait-Sahalia, Y. and Jacod, J. High-Frequency Financial Econometrics. Princeton University Press.
[3] Alekseevsky, D., Kriegl, A., Losik, M. and Michor, P.W. (1998). Choosing roots of polynomials smoothly. Israel J. Math, 105, 2013-233.
[4] Angelina, F. and Herzel, S. (2002). Consistent Initial Curves for Interest Rate Models. The Journal of Derivatives 9, 4, 8-17.
[5] Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. Econometrica, 70, 191-221.
[6] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. Econometrica, 71, 1, 135-171.
[7] Bathia, N. Yao, Q. and Ziegelmann, F. (2010). Identifying Finite Dimensionality of Curve Time Series. Ann. Stat, 38, 3352-3386.
[8] Baudoin, F. and Teichmann, J. (2005). Hypoellipticity in infinite dimensions and an application in interest rate theory. Ann. Appl. Probab, 15, no. 3, 1765-1777.
[9] Bjork, T. and Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. Math. Finance, 9 323-348.
[10] Bjork, T. C. Landm. (2002). On the construction of finite dimensional realizations for nonlinear forward rate models. Finance Stochast. 6, 303-331.
[11] Burgers, C.J.C. (2009). Dimension reduction:A guided tour. Foundations and Trends in Machine Learning, 2, 4, 275-365.
[12] Bibinger, M., Hautsch, N., Malec, P. and Rei, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. Annals of Stat, 42, 4, 1312-1346.
[13] Clement, E. and Gloter, A. (2011). Limit theorems in the Fourier transform method for the estimation of multivariate volatility. Stoch. Proc. Appl. 121, 1097-1124.
[14] Christensen, K., Podolskij, M. and Vetter, M. (2013): On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. Journal of Multivariate Analysis, 120, 59-84
[15] Diebold, F. X and Li, C. (2006). Forecasting the term structure of government bond yields. J. of Econometrics, 130, 337-364.
[16] Dunford, N. and Schwartz, J. T. Linear Operators. Vol II. Willey Classics Library.
[17] Forni, M., Hallin, M., Lippi, M; and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and Estimation. Review of Economics and Statistics, 82, 540-554.
[29] Harms, P., Stefanovits, D., Teichmann, J., and Wuthrich, M. Consistent Recalibration of Yield Curve Models. arXiv:1502.02926.

[30] Heston, S.L (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of financial studies, 6(2): 327-343.

[31] Jacod, J. and Podolskij, M. A test for the rank of the volatility process: the random perturbation approach. to appear in Annals of Statistics.

[32] Lee, J. and Verleysen, M. Nonlinear Dimensionality Reduction. Springer, 2007.

[33] Malliavin, P. and Mancino, M.E. (2009). A Fourier transform method for nonparametric estimation of multivariate volatility. Ann. Statist. 37, No. 4, 1983-2010.

[34] Malliavin, P; Mancino, M. E.; Recchioni, M. C. (2007). A non-parametric calibration of the HJM geometry: an application of Itô calculus to financial statistics. Jpn. J. Math. 2, no. 1, 55-77.

[35] Ohashi, A. (2009). Fractional term structure models: no-arbitrage and consistency. Ann. Appl. Probab. 19, no. 4, 1553-1580.

[36] Richter, A. and Teichmann, J. Discrete Time Term Structure Theory and Consistent Recalibration Models. Preprint arXiv:/1409.1830.

[37] Rutter, J. W. Geometry of Curves, CRC Press, 2009.

[38] Shang, H. L. (2014). A survey of functional principal component analysis. AStA Advances in Statistical Analysis, 98, 2, 121-142

[39] Slinko, I. (2010). On finite-dimensional realizations of two-country interest rate models. Mathematical Finance, 20, 1, 117-143.

[40] Stock, J. H. nad Watson, M. W. (2002). Forecasting Using Principal Components from a Large Number of Predictors. Journal of the American Statistical Association, 97, 460, 1167-1179.

[41] Tappe, S. (2010). An alternative approach on the existence of affine realizations for HJM term-structure models. Proceedings of The Royal Society of London, Series A, Mathematical, Physical and Engineering Sciences. 466(2122), 3033-3060.

[42] Teichmann, J. and Wthrich, M. Consistent Long-Term Yield Curve Prediction, arXiv/1203.2017, preprint.

[43] Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. Ann. Statist. 38, 2, 943-978.

[44] Zeng, X. and Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. Ann. Statist. 39, 6, 3121-3151.

[45] Tao, M., Wang, Y.and Zhou, H.H. (2013). Optimal sparse volatility matrix estimation for high-dimensional It processes with measurement errors. Ann. Statist, 41, 4, 1816-1864.

Departamento de Matemática, Universidade Federal da Paraíba, 13560-970, João Pessoa - Paraíba, Brazil

E-mail address: alberto.ohashi@pq.cnpq.br; ohashi@mat.ufpb.br

Departamento de Matemática, Universidade Federal da Paraíba, 13560-970, João Pessoa - Paraíba, Brazil

E-mail address: alexandre@mat.ufpb.br