Fourier Methods for Estimating Integrated Volatility and Occupation Time Functionals

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

M.Sc. Randolf Altmeyer

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Markus Reiß (Humboldt-Universität zu Berlin)
2. Prof. Dr. Mark Podolskij (Aarhus University, Denmark)
3. Prof. Dr. Nicolas Perkowski (Humboldt-Universität zu Berlin)

Tag der Verteidigung: 24.11.2017

Abstract

We study two estimation problems for continuous-time stochastic processes: the estimation of *integrated volatility* and the estimation of *occupation time functionals*.

In the first problem high-frequency observations of a continuous Itô semimartingale are considered, which are perturbed by additive microstructure noise. Our main contribution is the analysis of the prominent spectral estimation approach for very general random and time-varying volatility in any dimension. We prove stable central limit theorems, which show that the estimators attain the optimal rates of convergence with quasi-efficient asymptotic variances. Adaptive estimators together with feasible limit theorems are provided.

In the second problem we estimate occupation time functionals $\int_0^t f(X_r)dr$ for a function f and a d-dimensional càdlàg process X with respect to discrete observations by a Riemann-sum estimator. Based on novel semimartingale approximations in the Fourier domain, central limit theorems are proved for L^2-Sobolev functions f with fractional smoothness and continuous Itô semimartingales X. General $L^2(\mathbb{P})$-upper bounds on the error are given under weak assumptions. These bounds cover all previously obtained results in the literature and apply also to non-Markovian processes. Particularly simple and revealing results are obtained for stationary Markov processes. Several detailed examples are discussed. As an application the approximation of local times for fractional Brownian motion is studied. The optimality of the $L^2(\mathbb{P})$-upper bounds is shown by proving the corresponding lower bounds in case of Brownian motion.

The same methods as for studying occupation time functionals are used in the third part to obtain generalized Itô formulas for continuous Itô semimartingales and L^2-Sobolev functions. For this the existence of certain quadratic covariations is proved. As opposed to the usual assumption in the literature, however, X is not required to be reversible. The Itô formulas hold in any dimension and also for rough drift and volatility coefficients.

Zusammenfassung

Wir untersuchen zwei Schätzprobleme für zeitstetige stochastische Prozesse: Das Schätzen der *integrierten Volatilität* und das Schätzen von *occupation-time-Funktionalen.*

Das erste Problem beschäftigt sich mit hochfrequenten Beobachtungen von stetigen Itô-Semimartingalen, wobei die Beobachtungen durch additives Mikrostrukturrauschen gestört sind. Unser Hauptbeitrag in diesem Fall ist die Analyse des vielbeachteten Spektralansatzes für sehr allgemeine, zufällige und zeitveränderlichere Volatilitäten in beliebiger Dimension. Wir beweisen stabile Grenzwertsätze. Mit diesen kann nachgewiesen werden, dass die optimale Konvergenzrate erreicht wird mit quasi-effizienter asymptotischer Varianz. Darüberhinaus zeigen wir Grenzwertsätze für Daten-adaptive Schätzer.

Im zweiten Problem werden occupation-time-Funktionale $\int_0^t f(X_r)dr$ geschätzt für eine Funktion f und diskrete Beobachtungen eines d-dimensionalen Prozesses X mit càdlàg-Pfaden. Dafür verwenden wir einen Riemann-Schätzer. Mit Hilfe von innovativen Semimartingalapproximationen im Fourierbereich werden zentrale Grenzwertsätze bewiesen für L^2-Sobolev-Funktionen f mit fraktioneller Glattheit und stetige Itô-Semimartingale X. Außerdem werden allgemeine obere Schranken für den $L^2(\mathbb{P})$-Fehler gezeigt unter schwachen Voraussetzungen. Diese oberen Schranken umfassen alle bisher erzielten Resultate in der Literatur und gelten auch für nicht-Markovsche Prozesse. Besonders einfache und aufschlussreiche Resultate werden für stationäre Markovprozesse erzielt. Dies wird an mehreren detaillierten Beispielen verdeutlicht. Als Anwendung betrachten wir die Approximation von Lokalzeiten für fraktionelle Brownsche Bewegungen. Die Optimalität der oberen Schranken für den $L^2(\mathbb{P})$-Fehler wird bewiesen durch entsprechende untere Schranken, wenn X eine Brownsche Bewegung ist.

Die Methoden, die wir für das zweite Problem entwickelt haben, werden im dritten Teil der Arbeit verwendet, um verallgemeinerte Itô-Formeln für stetige Itô-Semimartingale und L^2-Sobolev-Funktionen zu zeigen. Dafür beweisen wir die Existenz von bestimmten quadratischen Variationen. Im Gegensatz zur üblichen Annahme in der Literatur muss X jedoch nicht reversibel sein. Die Itô-Formeln gelten in jeder Dimension und auch für weniger glatte Drift- und Volatilitätskoeffizienten.

Contents

Chapter 1.

Introduction

The theory of stochastic processes is an important subject in modern mathematics. It is applied in all sciences, in engineering and in finance to model complex phenomena that appear to be random and vary in time. Continuous-time models are often mathematically more tractable than discrete-time models, but statistical inference is usually limited to discrete observations. If the time grid is fine enough, it can be expected that the statistical properties of the discretized model are close to the properties of the continuous one. The quality of the approximation, however, depends crucially on the underlying model, possibly unknown parameters and observation errors. It is therefore important for theoretical and practical reasons to study the properties of these approximations.

We focus in this thesis on two estimation problems for continuous-time stochastic processes: The estimation of integrated volatility for continuous Itô semimartingales and the estimation of occupation time functionals for general processes with càdlàg paths. The fundamental novelty is to use Fourier analysis, together with pathwise and distributional approximations to derive new and surprisingly strong results. The methods are developed in order to be applicable to rather general processes and independent of the dimension. In the third part of the thesis we will use these methods to generalize one of the most important tools in stochastic analysis, namely Itô's formula.

The problems and methods discussed in this thesis lie in the intersection between statistics and stochastic analysis. For basic definitions and a general overview the reader may consult the monographs of Revuz and Yor (1999), Jacod and Shiryaev (2013) and Jacod and Protter (2011). The next three sections provide a general overview of the problems that we study. We will give detailed account of previous work. This is followed by an outline of the main results.

Estimating integrated volatility

For the first estimation problem let $X = (X_t)_{0\leq t\leq T}$ for $T > 0$ be a d-dimensional continuous Itô semimartingale of the form

$$X_t = X_0 + \int_0^t b_r dr + \int_0^t \sigma_r dW_r, \quad 0 \leq t \leq T, \tag{1.0.1}$$

with drift $b = (b_t)_{0\leq t\leq T}$, volatility $\sigma = (\sigma_t)_{0\leq t\leq T}$ and a Brownian motion $W = (W_t)_{0\leq t\leq T}$. Log-price models based on Itô semimartingales are widely used in econometrics and finance. A key problem is to estimate the *integrated volatility*

$$IV_t = \int_0^t \sigma_r \sigma_r^\top dr, \quad 0 \leq t \leq T,$$

from discrete observations of X at $t_k = k\Delta_n$, where $\Delta_n = T/n$ and $k = 0, \ldots, n$. Since

$$IV_t = \langle X, X \rangle_t = \lim_{n\to\infty} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} (X_{t_k} - X_{t_{k-1}})(X_{t_k} - X_{t_{k-1}})^\top$$

is the quadratic variation of X, where the limit is taken in probability, a natural estimator for this is the realized variance $\widehat{RV_t} = \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} (X_{t_k} - X_{t_{k-1}})(X_{t_k} - X_{t_{k-1}})^\top$. It is variance efficient with optimal rate of convergence $\Delta_n^{1/2}$, even for more general jump processes X (cf. Jacod and Protter (1998), Jacod and Todorov (2014)). When being applied to financial data obtained at high frequency, however, it turns out that the realized variance is unreliable (cf. Zhang et al. (2005)). The well-established concept in the financial econometrics literature to explain this phenomenon is that effects like bid-ask spreads, transaction costs and round-off errors caused by the discreteness of prices perturb the true price process when observed at high frequency (Ait-Sahalia and Jacod (2014, Section 2.3)). This means that the semimartingale model alone has severe limitations in describing stylized facts of high-frequency data. Many authors therefore study instead the observation model $Y_k = X_{t_k} + \varepsilon_k$ with additive iid *microstructure noise*. In this case the realized variance is provably explosive (Zhang et al. (2005)), which makes it necessary to develop new estimation methods for IV_t.

The one-dimensional parametric experiment with constant volatility σ, without drift and with Gaussian iid noise has been well understood through a LAN (local asymptotic normality) result by Gloter and Jacod (2001). They showed that the optimal rate of convergence drops to $\Delta_n^{1/4}$ and the efficient variance is $8\eta\sigma^3$, where η^2 is the variance of the noise, while the efficient variance is $2\sigma^4$ in the model without noise. The nonparametric estimation problem for time-varying volatility has been studied in recent years by many different authors and several rate optimal estimators have been developed, all satisfying central limit theorems. Prominent approaches are Zhang (2006), Barndorff-Nielsen et al. (2008), Jacod et al. (2009) and Xiu (2010) for $d = 1$ and Ait-Sahalia et al. (2010), Barndorff-Nielsen et al. (2011), Bibinger (2011) and Christensen et al. (2013) for the multi-dimensional setup. A major focus has been to attain a minimum asymptotic variance, which at the slow optimal convergence rate could result in substantial finite sample precision gains.

A fundamentally different approach was introduced by Reiß (2011). He proved for $d = 1$, deterministic, but time-varying volatility, without drift and with Gaussian iid noise that the noisy observation model is asymptotically equivalent in the Le Cam sense to a Gaussian white noise experiment. This made it possible to construct a rate optimal estimator for IV_t based on a spectral decomposition of the covariance operator in the equivalent white noise experiment and using locally a method of moments. In contrast to all previous approaches, the *spectral estimator* achieves the Cramér-Rao efficiency lower bound $8\eta \int_0^t \sigma_r^3 dr$ for the asymptotic variance. This approach was extended by Bibinger and Reiß (2014) and Bibinger et al. (2014) to a multi-dimensional non-synchronous framework.

In the general setting, where σ is an adapted stochastic process, the asymptotic variance of the estimators above is random, as well. The notion of nonparametric efficiency, however, has been restricted so far to deterministic situations. A first step to generalize this was taken by Clément et al. (2013) who proved a convolution theorem for random functionals when the LAMN property (local asymptotic mixed normality) is satisfied. This indeed shows efficiency for the realized variance estimator in the observation model without noise. A similar result has not been achieved yet in the model with noise, but it is generally believed that the nonparametric efficiency bound from the submodel with deterministic volatility $8\eta\int_0^t \sigma_r^3 dr$ also holds here. For instance, Jacod and Mykland (2015) have proposed an adaptive version of their pre-average estimator which achieves an asymptotic variance of about $1.07\cdot 8\eta\int_0^t \sigma_r^3 dr$. The first main result of this thesis will be to show that the spectral estimator for general dimension d indeed attains the conjectured optimal asymptotic variance, when σ is a rather general càdlàg process. This allows for using more complex, non-Markovian models which are of central interest in finance and econometrics.

We want to mention a few other results related to this estimation problem. Recent advances have been made in analyzing properties of the volatility process itself, and not only in its integrated form. The focus has been here on describing smoothness properties of σ as observed in financial data (see for example Gatheral et al. (2018), Bibinger et al. (2017); see also Renault et al. (2017)). In the model without noise there are also several works on estimating different functionals of σ, not only the integrated squared process. Interesting results in this context have been obtained by Jacod and Rosenbaum (2013) and Li et al. (2013). Without noise it was also shown that random endogenous observation times can influence the results (Li et al. (2014)). An observation model with one-sided errors has been studied by Bibinger et al. (2016), exhibiting different optimal rates of convergence.

Estimating occupation time functionals

For the second estimation problem consider a general d-dimensional stochastic process $X=(X_t)_{0\leq t\leq T}$ with càdlàg paths. Similar to the last section, an important problem in many fields is to estimate integral-type functionals of the form

$$\Gamma_t(f)=\int_0^t f(X_r)\,dr,\quad 0\leq t\leq T,$$

for a function f from discrete observations of X at $t_k=k\Delta_n$, where $\Delta_n=T/n$ and $k=0,\dots,n$. $\Gamma_t(f)$ is called the *occupation time functional* of X.

If A is a Borel set, then $\Gamma_T(\mathbf{1}_A)$ is known as the *occupation time* which measures the time X spends in A. General functions f appear for example in mathematical finance to model path dependent derivatives (Hugonnier (1999), Chesney et al. (1997)) or in evolutionary dynamics (Pollett (2003)). Occupation time functionals are also an important tool from a statistical point of view to estimate functionals with respect to the invariant measure μ of an ergodic process X, because $T^{-1}\Gamma_T(f)\to\int f d\mu$ as $T\to\infty$ by the ergodic theorem under appropriate regularity assumptions (Dalalyan (2005), Mattingly et al. (2010)). Moreover, the smoothness properties of $x\mapsto\int_0^T f(x+X_r)dr$ play an important role in solving ordinary differential equations, for example in combination with the phenomenon of regularization by noise (Catellier and Gubinelli (2016)).

The natural estimator for discrete observations is the Riemann-sum estimator

$$\widehat{\Gamma}_{n,t}\left(f\right)=\Delta_{n}\sum_{k=1}^{\lfloor t/\Delta_{n}\rfloor}f(X_{t_{k-1}}).$$

It has been applied in the statistics literature, for instance, in order to estimate the occupation time (Chorowski (2018), Gobet and Matulewicz (2016)) or functionals of the local time of a diffusion (Florens-Zmirou (1993), Jacod (1998)). For general f see also Dion and Genon-Catalot (2016). The obtained error bounds for $\Gamma_t(f)-\widehat{\Gamma}_{n,t}(f)$ are often suboptimal and very specific to the problem at hand. The Riemann-estimator is moreover commonly used for simulating from the law of $\Gamma_t(f)$. For this the X_{t_k} usually have to be approximated by some $X^n_{t_k}$, obtained for example by an Euler-scheme (Mattingly et al. (2010)). However, the increasing availability of exact simulation methods alleviates this problem to some extent (Beskos and Roberts (2005)). Jacod et al. (2003) considered the Riemann-sum estimator for $f(x)=x$ in order to find the rate of convergence of the integrated error $\int_0^t(X_r-X_{\lfloor r/\Delta_n\rfloor\Delta_n})dr$ for semimartingales with jump discontinuities, because in this case the error $X_t-X_{\lfloor t/\Delta_n\rfloor\Delta_n}$ does not converge to zero in the Skorokhod sense.

The theoretical properties of $\widehat{\Gamma}_{n,t}(f)$ have been studied systematically only in few works and only for rather specific processes X and functions f. Consistency as $\Delta_n\to 0$ follows from Riemann approximation already under weak assumptions. A central limit theorem for Itô semimartingales and $f\in C^2(\mathbb{R}^d)$ was proven in the monograph of Jacod and Protter (2011, Chapter 6) with rate of convergence Δ_n. This is much faster than the $\Delta_n^{1/2}$-rate when approximating $f(X_t)$ by $f(X_{\lfloor t/\Delta_n\rfloor\Delta_n})$ for continuous X. Interestingly, the weak limit depends only on ∇f and therefore it seems that the CLT might also hold for $C^1(\mathbb{R}^d)$-functions. The proof, however, works only for $f\in C^2(\mathbb{R}^d)$, using Itô's formula.

For less smooth functions no CLT has been obtained so far. Instead, several authors considered $L^2(\mathbb{P})$-bounds for the estimation error $\Gamma_t(f)-\widehat{\Gamma}_{n,t}(f)$. For α-Hölder functions f and $0\leq\alpha\leq 1$ the rate of convergence $\Delta_n^{(1+\alpha)/2}$, up to log factors, has been obtained by Malliavin calculus for one dimensional diffusions (Kohatsu-Higa et al. (2014)) and by assuming heat kernel bounds on the transition densities for Markov processes in $\mathbb{R}^d$ (Ganychenko (2015); Ganychenko and Kulik (2014)). For applications the most important case is when f is the indicator function of a Borel set. The only available result in the literature applies to one-dimensional Brownian motion and indicator functions $f=\mathbf{1}_{[a,b)}$, $a<b$. Ngo and Ogawa (2011) found the surprising rate $\Delta_n^{3/4}$, which corresponds to the Hölder-rate for $\alpha=1/2$. It is not clear if a similar result holds in higher dimensions or for different processes. Ngo and Ogawa (2011) also showed that the $\Delta_n^{3/4}$-rate for indicators is optimal in the $L^2(\mathbb{P})$-sense. This is the only proof of optimality for estimating occupation time functionals so far. Note that all studied processes until now are Markov processes.

In principle, the Riemann-sum estimator is not the only possible estimator for approximating $\Gamma_t(f)$. A possible alternative is the trapezoid rule

$$\widehat{\Theta}_{n,t}\left(f\right)=\Delta_{n}\sum_{k=1}^{\lfloor t/\Delta_{n}\rfloor}\frac{f\left(X_{t_{k-1}}\right)+f\left(X_{t_{k}}\right)}{2},$$

which is also the optimal estimator from a filtering point of view if X is a Brownian motion and $f(x) = x$, i.e. $\mathbb{E}[\Gamma_t(f)|X_0, \ldots, X_T] = \widehat{\Theta}_{t,n}(f)$ (cf. Diaconis (1988)). We will show that both estimators have the same properties asymptotically, independent of the smoothness of f, which is different from deterministic numerical integration. Interesting connections to probabilistic quadrature rules can also be found in Briol et al. (2015). Estimation of occupation time functionals, where the process is not observed directly, has been studied for example by Li et al. (2013) when X is the volatility of an Itô semimartingale.

Generalized Itô formulas

For the third problem studied in this thesis let X be again a d-dimensional continuous Itô semimartingale as in (1.0.1). One of the most important tools in stochastic analysis is Itô's formula which says in its classical form that

$$f(X_t) - f(X_0) = \int_0^t \langle \nabla f(X_r), dX_r \rangle + \frac{1}{2} \sum_{k,m=1}^{d} \int_0^t \partial^2_{km} f(X_r)\, d\left\langle X^{(k)}, X^{(m)} \right\rangle_r$$

for $f \in C^2(\mathbb{R}^d)$ and $0 \leq t \leq T$. Various generalizations of this formula have been obtained in the past in order to relax the regularity assumptions on f. The main distinction between the different formulas is the Itô-correction term. It can be expressed by generalized integrals with respect to local times (cf. Bouleau and Yor (1981)) or as $\frac{1}{2}\sum_{m=1}^d [\partial_m f(X), X^{(m)}]_t$, where $[\partial_m f(X), X^{(m)}]_t$ is the *quadratic covariation* of $\partial_m f(X)$ and $X^{(m)}$ (cf. Föllmer et al. (1995)). It is defined by

$$\left[\partial_m f(X), X^{(m)}\right]_t = \lim_{n\to\infty} \sum_{t_k \in \pi_n, t_k \leq t} \left(\partial_m f(X_{t_k}) - \partial_m f\left(X_{t_{k-1}}\right)\right)\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right), \tag{1.0.2}$$

if this limit exists in probability for any sequence of partitions $(\pi_n)_{n\geq 1}$ of $[0,T]$ such that the mesh size $|\pi_n| = \max_k |t_k - t_{k-1}|$ tends to zero as $n \to \infty$, where the points in π_n are $0 = t_0 < t_1 < \cdots < t_n = T$. For one-dimensional semimartingales both approaches give essentially the same results (Bardina and Rovira (2007)). In higher dimensions the local time method does not apply anymore (with the notable exception of Eisenbaum (2006) for Lévy processes with independent components).

A major problem in applying the second method is to prove the existence of $[\partial_m f(X), X^{(m)}]_t$. A sufficient condition is reversibility of X, i.e. the time reversed process $t \mapsto X_{T-t}$ is again a diffusion or just a semimartingale. In this case the partial sums in (1.0.2) can be decomposed into the sum of $\sum_{t_k\in\pi_n, t_k\leq t} \partial_m f(X_{t_{k-1}})(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}})$ and $\sum_{t_k\in\pi_n, t_k\leq t} \partial_m f(X_{t_k})(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}})$, which separately converge to the forward and backward stochastic integrals $\int_0^t \partial_m f(X_r) dX_r$ and $\int_0^t \partial_m f(X_r) d^* X_r$. Reversibility holds, for example, for non-degenerate diffusions with Lipschitz coefficients (Millet et al. (1989)). Under this assumption, several Itô formulas were obtained: by Föllmer and Protter (2000) for d-dimensional Brownian motion and L^2-Sobolev functions $f \in H^1(\mathbb{R}^d)$, by Moret and Nualart (2001) for d-dimensional non-degenerate diffusions and functions in

L^p-Sobolev spaces with $p > d$ and regularity one and by Errami et al. (2002) for $C^1(\mathbb{R}^d)$-functions and reversible semimartingales. All formulas (including the ones based on local times, but except for Föllmer and Protter (2000)) apply only to continuous f. Note also that there are simple counterexamples for reversibility as soon as we give up on the assumption that X is a diffusion (Walsh (1982)).

In view of (1.0.2) the existence of the quadratic covariation $[\partial_m f(X), X^{(m)}]_t$ is a discretization problem. Instead of identifying it as the limit of certain stochastic integrals, we consider the approximation *directly* without the assumption of reversibility. Note that $(f(X_t))_{0\leq t\leq T}$ is in general *not* a semimartingale anymore, if $f \notin C^2(\mathbb{R}^d)$ (see Theorem 71 of Protter (2013) for a counterexample). This means that $[\partial_m f(X), X^{(m)}]_t$ may have paths of unbounded variation. It is, however, a process of zero quadratic variation for $d = 1$ and Lipschitz continuous f (cf. Lowther (2010), see also Walsh (2013)). It is also interesting to note that Itô formulas can even be proven for Hölder continuous functions, however only if X is a pure-jump process (cf. Jacod et al. (2003)).

Outline of main results

We now give an overview of the main results for the three problems above.

Chapter 2 investigates the spectral approaches for estimating integrated volatility of Reiß (2011), Bibinger and Reiß (2014) and Bibinger et al. (2014) in a very general setup. We study the noisy high-frequency non-synchronous observation model with general error distribution and where X satisfies (1.0.1) for a general drift and random time varying volatility. The one-dimensional case is considered first and in more detail in order to give a clearer picture of the main ideas. This is followed by the bivariate and the general multi-dimensional setting, including non-synchronous observations. We prove functional stable limit theorems at the optimal convergence rate $\Delta_n^{1/4}$ and with asymptotic variances coinciding with the lower bounds in the nonparametric subexperiments with deterministic volatility. As there is no suitable variance efficiency concept in this case yet, we call the spectral estimators *quasi-efficient*. They are the only estimators so far in this general setting to achieve the minimal asymptotic variance.

The asymptotic analysis is based on the theory of Jacod (1997), applied in a similar context also in Fukasawa (2010) and Hayashi and Yoshida (2011), and incorporates Fourier analysis and matrix algebra. The fundamental idea of the spectral estimator is to smooth the noisy observations in the Fourier domain. The smoothed observations are then combined by a local method of moments with optimal weights which depend on the local covolatility matrix. The optimal weights thus require knowledge of the unknown volatility process. We therefore prove adaptive versions of the spectral estimators where in a first step the local covolatility matrices are pre-estimated from the same data. This two stage method yields feasible limit theorems that are fully data driven.

Chapter 3 studies the estimation of occupation time functionals from several different points of views. Related to the classical work of Geman and Horowitz (1980) on occupation densities, we use Fourier methods for the estimation error $\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)$. The key idea is that the error is equal to

$$(2\pi)^{-d}\int \mathcal{F}f\,(u)\Bigg(\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\Big(e^{-i\langle u,X_r\rangle}-e^{-i\big\langle u,X_{t_{k-1}}\big\rangle}\Big)dr\Bigg)du$$

by inverse Fourier transform under suitable regularity assumptions. Together with a pathwise analysis of the exponentials $e^{-i\langle u,X_r\rangle}$ and with functions f having sufficiently regular Fourier transforms this is just the right idea to control the estimation error. The pathwise analysis is inspired by the one-step Euler approximations of Fournier and Printems (2008) and is also related to Catellier and Gubinelli (2016). These ideas allow us in Section 3.1 to extend the central limit theorem of Jacod and Protter (2011) to L^2-Sobolev functions $f \in H^1(\mathbb{R}^d)$ and non-degenerate continuous Itô semimartingales with the same rate of convergence Δ_n. The proof is based on tight bounds for the Itô-correction term in the classical Itô formula. Note that a function $f \in H^1(\mathbb{R}^d)$ is not necessarily continuous for $d > 1$.

For less smooth functions it is in general not possible to prove central limit theorems, because the bias may become degenerate asymptotically. Instead, Section 3.2 provides non-asymptotic upper bounds for the $L^2(\mathbb{P})$-error $\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)$ and general d-dimensional càdlàg processes X under weak assumptions. Only the smoothness of the bivariate distributions of (X_h, X_r) in $0 \leq h < r \leq T$ is required, i.e. either the joint densities or the characteristic functions are differentiable in h and r. This allows us to prove the rate $\Delta_n^{(1+s)/2}$ for a large class of d-dimensional processes and L^2-Sobolev functions with fractional smoothness $0 \leq s \leq 1$. In particular, this covers the previous results for Hölder and indicator functions. We therefore obtain a unifying mathematical explanation for the different rates. Several examples demonstrate the applicability of the upper bounds, for example to Markov processes, but also to fractional Brownian motion. As an interesting application we prove rates of convergence for approximating the local times of fractional Brownian motion. Note that the $L^2(\mathbb{P})$-bounds also yield improved bounds for the so-called *weak approximations* $\mathbb{E}[\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)]$, which are of key importance in Monte-Carlo simulations (cf. Gobet and Labart (2008), see also Kohatsu-Higa et al. (2014)).

Section 3.3 studies the special case of stationary Markov processes. In this case the $L^2(\mathbb{P})$-error of $\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)$ can be calculated explicitly with respect to the associated semigroup. This yields upper bounds in terms of fractional powers of the infinitesimal generator of the process applied to f. While the assumption of stationarity is a limitation, these bounds are more precise than the ones above (when both methods apply), because they also hold for functions $f \in L^2(\mu)$ which are square integrable with respect to the invariant measure μ. Moreover, this method easily extends to infinite dimensions.

Rate optimality is addressed in Section 3.4. We prove the corresponding lower bounds for the $L^2(\mathbb{P})$-error in case of L^2-Sobolev functions and d-dimensional Brownian motion. In this case we can even conclude the efficiency of the Riemann-sum estimator in terms of its asymptotic variance.

We want to emphasize that the $L^2(\mathbb{P})$-bounds are not only optimal and explicit with respect to their dependence on Δ_n, but also with respect to T. More precisely, the typical upper bound for L^2-Sobolev functions is of order $T^{1/2}\Delta_n^{(1+s)/2}$. This allows for approximating functionals $\int f d\mu$ in an ergodic setting with respect to the invariant measure μ at the optimal rate $T^{-1/2}$ by the estimator $T^{-1}\widehat{\Gamma}_{n,T}(f)$, independent of Δ_n being fixed or $\Delta_n \to 0$. We therefore believe that our results may be instrumental in bridging the gap between results in statistics obtained for high-frequency and low-frequency observations. In fact, the results of Section 3.3 have been crucial for approximating $\int_0^t \mathbf{1}_{[a,b)}(X_r)dr$, $a < b$, with respect to a one-dimensional stationary diffusion X in an effort to find a universal estimator for the volatility process which is minimax optimal at high and low frequency (cf. Chorowski (2018)). Moreover, it is well-known that, under suitable regularity assumptions, $T^{-1}\Gamma_T(f)$ converges to $\int f d\mu$ at the rate $T^{-1/2}$. This is the same rate as for $T^{-1}\widehat{\Gamma}_{n,T}(f)$. This suggests that our results can also be applied to transfer results obtained in statistics for continuous observations to discrete observations by approximating the corresponding integral functionals.

Chapter 4 is devoted to showing the existence of the quadratic covariations $[\partial_m f(X), X^{(m)}]_t$ for d-dimensional non-degenerate continuous Itô semimartingales. We apply the methods developed in Chapter 3, working under the same assumptions as for the central limit theorems in Section 3.1. This yields surprisingly strong results. Most importantly, reversibility of X is not necessary. This further immediately leads to generalized Itô formulas. The main result is that, up to some minor conditions, Itô's formula holds for L^2-Sobolev functions $f \in H^s(\mathbb{R}^d)$, $s > 1$, if σ is Lipschitz continuous and uniformly elliptic. This result achieves two things. First, it generalizes the Itô formulas mentioned above for continuous diffusions and in any dimension $d \geq 1$. In particular, we have a precise relation between the regularity of f, σ and b, such that Itô's formula also holds for rough coefficients, if f is slightly more regular. Second, the formula also holds for continuous Itô semimartingales and therefore allows for very complex processes that are important in practice.

Note that we have to study all possible partitions $(\pi_n)_{n\geq 1}$ with mesh size $|\pi_n| \to 0$ for the existence of $[\partial_m f(X), X^{(m)}]_t$. This suggests that also the results in Chapter 3 should hold for more general sampling schemes.

Acknowledgements

This thesis would not have been possible without many wonderful people around me during the past few years. First of all, I want to express my deepest gratitude to Markus Reiß who gave me this incredible opportunity of being a member in his working group and in the Humboldt family. He was always supportive and guided me with clear ideas and his infinite enthusiasm for mathematics.

I was lucky to find a welcoming and creative atmosphere in the Stochastics group at Humboldt-University. Special thanks go to Moritz Jirak, David Prömel, Jörg Martin, Sebastian Holtz, Ester Mariucci, Katerina Papagiannouli and, in particularly, my office mates Mathias Trabs and Martin Wahl, who helped me enormously by listening to my simple questions.

I am very thankful to have been a member of the research training group "Stochastic Analysis with Applications in Biology, Finance and Physics". I especially acknowledge the support for my stay at Yale University. I want to thank Harrison Zhou at the department of statistics at Yale for welcoming me into his group and connecting me to his student Derek Feng, which lead to many fruitful and challenging discussions.

Many thanks to my collaborators Markus Bibinger and Jakub Chorowski for productive discussions and for always supporting my crazy ideas. I especially have to thank Jakub Chorowski for sharing his research problem with me, which opened up many new possibilities for me.

I would not be here at this point without the support from my family and especially from my girlfriend Poy. Without her I would not be in Berlin and I probably would have never dared to study mathematics in the first place. I could not have finished this thesis without her love and constant encouragement.

Music has always been an important part of my life. It was also essential for finishing this thesis. I recommend the following playlist of songs, which inspired me the most in the past few years, during reading (not necessarily in this order):

1. *Day 3*, Marilyn Manson
2. *Angeles*, Elliot Smith
3. *Pictures of me*, Elliot Smith
4. *City of stars*, La La Land Soundtrack
5. *Disparity by design*, Rise Against
6. *The rains of Castamere*, Game of Thrones Soundtrack
7. *Dead man's eye*, Apocalyptica
8. *Sacred world*, Blind Guardian
9. *Seabeast*, Mastodon
10. *Let it go*, Frozen Soundtrack
11. *Dance of the knights*, Prokofiev (from Romeo and Juliet)
12. *Karma Chameleon*, Culture Club

If you listen closely, you might find parts of these songs hidden in the thesis ...

Nomenclature

The notation follows the usual conventions. Throughout the thesis C or C_p denote non-negative absolute constants which may change from line to line. We give now a list of general mathematical symbols that will be used.

$\xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{d}$	weak convergence
$\xrightarrow{st}$	stable convergence
$\xrightarrow{ucp}$	uniform convergence in probability
$a_n \lesssim b_n$, $a_n = O(b_n)$	$a \leq Cb$
$a \asymp b$	$a \lesssim b$ and $b \lesssim a$
$a_n = o(b_n)$	$a_n/b_n \to 0$
$X_n = O_{\mathbb{P}}(a_n)$	$(X_n/a_n)_{n\geq 1}$ is tight
$X_n = o_{\mathbb{P}}(a_n)$	$X_n/a_n \to 0$ in probability
$X_n = O_{ucp}(a_n)$	$(X_n/a_n)_{n\geq 1}$ is tight with respect to the ucp topology
$X_n = o_{ucp}(a_n)$	$X_n/a_n \to 0$ with respect to the ucp topology
$C_c^\infty(\mathbb{R}^d)$	smooth functions with compact support
$L^p(\mathbb{R}^d)$, $L^p(\mu)$	p-integrable functions, with respect to measure μ
$\mathcal{S}(\mathbb{R}^d)$	Schwartz functions
$C^s(\mathbb{R}^d)$	$\lfloor s \rfloor$-times differentiable functions f such that all partial derivatives of order $\lfloor s \rfloor$ are $s - \lfloor s \rfloor$-Hölder continuous
$\mathcal{D}([0,T], \mathbb{R}^d)$	Skorokhod space
$\lVert\cdot\rVert$, $\lVert\cdot\rVert_\infty$	Euclidean and sup norms on $\mathbb{R}^d$ or $\mathbb{R}^{d\times d}$
$\lVert\cdot\rVert_{L^p}$, $\lVert\cdot\rVert_{L^p(\mathbb{P})}$	L^p norms on $L^p(\mathbb{R}^d)$ and $L^p(\mathbb{P})$
$\lVert\cdot\rVert_{\infty,\mu}$	sup norm with respect to measure μ
$\lVert\cdot\rVert_\alpha$	Hölder seminorm
$\lVert\cdot\rVert_\mu$	L^2-norm with respect to measure μ
δ_{lm}	Kronecker's delta, i.e. $\delta_{ll} = 1$ and $\delta_{lm} = 0$ for $l \neq m$
I_d	$d \times d$-dimensional identity matrix
$[f,g]_n$, $\langle f,g \rangle_n$	empirical scalar products
$A^\top$	transpose of matrix A
$\mathrm{dom}(L)$	domain of operator L
$\mathrm{vec}(A)$	vectorization of A
$A \otimes B$	Kronecker product of A and B
$\mathcal{F}f$	Fourier transform of f
Id	identity function
$\mathcal{B}(\mathcal{M})$	Borel σ-algebra with respect to metric space M
$(P_t)_{t\geq 0}$	Markov semigroup
$\sigma(L)$	spectrum of operator L
$\sigma(X_{t_k} : 0 \leq k \leq n)$	sigma algebra generated by random variables X_{t_k}

δ_x	dirac measure or dirac delta function in $x \in \mathbb{R}^d$
Σ_t	short for $\sigma_t \sigma_t^\top$
$\Gamma_t(f)$	occupation time functional
$\widehat{\Gamma}_{n,t}(f)$	Riemann-sum estimator
$\widehat{\Theta}_{n,t}(f)$	trapezoid rule estimator
$\widehat{IV_{n,t}}$, $\widehat{ICV_{n,t}^{p,q}}$, $LMM_{n,t}$	adaptive spectral estimators

Chapter 2.

Central limit theorems for spectral estimators of integrated volatility

This chapter is adapted from Altmeyer and Bibinger (2015).

We present in this chapter several central limit theorems for the spectral estimators of Reiß (2011), Bibinger and Reiß (2014) and Bibinger et al. (2014) in a general setting. The first two sections describe the assumptions for the underlying process and the observations, as well as the main results. The estimators will only be defined in the third section, which also discusses in detail the spectral approach. In Section four the asymptotic theory is presented with a focus on the one-dimensional case. For the multi-dimensional estimators only the main steps are proved, including the effects of non-synchronicity. Section five presents a small simulation study in order to investigate the finite sample performance of the one-dimensional estimator. The proofs can be found in Section six.

2.1. Statistical model

We first introduce the statistical model and provide all assumptions. For $T > 0$ let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ be a filtered probability space which satisfies the usual conditions, i.e. the filtration is right continuous and complete. On this space consider a d-dimensional continuous Itô semimartingale X such that

$$X_t = X_0 + \int_0^t b_r dr + \int_0^t \sigma_r dW_r, \quad 0 \le t \le T, \tag{2.1.1}$$

where X_0 is $\mathcal{F}_0$-measurable, $(W_t)_{0\le t\le T}$ is a standard d-dimensional Brownian motion, the drift $b = (b_t)_{0\le t\le T}$ is a locally bounded $\mathbb{R}^d$-valued process and the volatility $\sigma = (\sigma_t)_{0\le t\le T}$ is a càdlàg $\mathbb{R}^{d\times d}$-valued process, all adapted to $(\mathcal{F}_t)_{0\le t\le T}$.

We pursue the asymptotic analysis with respect to the following structural assumptions on b and σ.

Assumption 2.1.1 (SM-α-β)**.** *Let $0 \le \alpha, \beta \le 1$. There exists a constant C and a sequence of stopping times $(\tau_K)_{K\ge1}$ with $\tau_K \to \infty$ as $K \to \infty$ such that*

$$\mathbb{E}\left[\sup_{0\le r\le t} \|\sigma_{(s+r)\wedge\tau_K} - \sigma_{s\wedge\tau_K}\|^2\right] \le Ct^{2\alpha}, \quad \mathbb{E}\left[\sup_{0\le r\le t} \|b_{(s+r)\wedge\tau_K} - b_{s\wedge\tau_K}\|^2\right] \le Ct^{2\beta}$$

for all $0 \le s, t \le T$ with $s + t \le T$. Moreover, $(\sigma_t, \sigma_t^\top)_{0\le t\le T}$ is elliptic in the sense that for $\mathbb{P}$-almost all $\omega \in \Omega$ there exist constants $C(\omega)$ with $\inf_{0\le t\le T}(\sigma_t\sigma_t^\top)(\omega) \ge C(\omega)$.

The smoothness assumptions on σ and b are rather general and appear frequently in the literature (see e.g. Jacod and Mykland (2015), Jacod and Protter (2011, Section 2.1.5)). They exclude fixed times of discontinuities, but allow for non-predictable jumps. The assumptions are satisfied, if σ and b are themselves Itô semimartingales (with $\alpha = 1/2$ or $\beta = 1/2$) or if their paths are Hölder continuous with regularity α or β. In particular, they hold with $\alpha = \beta = 1/2$ if X is a diffusion process such that $\sigma_t = \tilde{\sigma}(X_t)$, $b_t = \tilde{b}(X_t)$ with Lipschitz continuous functions $\tilde{\sigma}$, $\tilde{b}$. The semimartingale model in (2.1.1) and Assumption (SM-α-β) will appear again in Chapters 3 and 4. In this chapter we will work under (SM-α-β) for $\alpha \geq 1/2$ and any $\beta > 0$.

In the following consider without loss of generality $T = 1$. We work within the model where we have indirect observations of X diluted by noise. For this let $\varepsilon = (\varepsilon_t)_{0\leq t\leq 1}$ be a d-dimensional iid white noise process with independent components. We assume that ε is independent of $\mathcal{F}$. Set $\mathcal{G}_t = \mathcal{F}_t \otimes \sigma(\varepsilon_s : s \leq t)$ for $0 \leq t \leq 1$ and let $(\Omega^0, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P}^0)$ be a filtered probability space which accommodates the signal and the noise processes and extends the space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, \mathbb{P})$. For simplicity, denote it by $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P})$.

With respect to the observation scheme of the perturbed process we consider two different assumptions.

Assumption 2.1.2 (N-1)**.** *Let $d = 1$. The white noise process satisfies $\mathbb{E}[\varepsilon_t^8] < \infty$ and $\mathbb{E}[\varepsilon_t^2] = \eta^2 > 0$ for all $0 \leq t \leq 1$. On the extension $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P})$ we observe $Y \in \mathbb{R}^{n+1}$ with $Y_k = X_{t_k} + \varepsilon_{t_k}$ at regular times $t_k = k/n$, $k = 0, \ldots, n$.*

For general dimensions $d \geq 1$ we consider a very general framework with noise and in which observations arrive at non-synchronous sampling times. Denote the *integrated covolatility matrix* by $\int_0^t \Sigma_s\, ds$, $\Sigma_s = \sigma_s\sigma_s^\top$. It coincides with the quadratic variation $\langle X, X\rangle_t$ of X.

Assumption 2.1.3 (N-d)**.** *Let $d \geq 1$. The white noise process satisfies $\mathbb{E}[(\varepsilon_t^{(p)})^8] < \infty$ and $Var(\varepsilon_t^{(p)}) = \eta_p^2 > 0$ for all $0 \leq t \leq 1$ and $p = 1, \ldots, d$. On the extension $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P})$ we observe*

$$Y_k^{(p)} = X_{t_k^{(p)}}^{(p)} + \varepsilon_k^{(p)}, \quad k = 0, \ldots, n_p, \quad p = 1, \ldots, d,$$

at non-synchronous observation times $0 \leq t_k^{(p)} \leq 1$ which are described by quantile transformations $t_k^{(p)} = F_p^{-1}(k/n_p)$ with differentiable, possibly random (but independent of X and ε) distribution functions F_p with $F_p(0) = 0$, $F_p(1) = 1$, $F_p' \in C^\alpha([0,1])$ with values in $[0,1]$ for some $1/2 < \alpha \leq 1$ such that F_p' is strictly positive. Moreover, $n/n_p \to \nu_p$ as $n \to \infty$ with constants $0 < \nu_p < \infty$ for all $p = 1, \ldots, d$.

For $d = 1$, $\nu_1 = 1$ and $F_1 = Id$ this assumption reduces to (N-1). Working under (N-1) in $d = 1$ allows for a simpler proof which may help to understand the method better. For $d \geq 1$ Assumption (N-d) includes deterministic and random observation times which are independent of Y. While this is still an idealization of realistic market microstructure dynamics, our observation model constitutes the established setup in the related literature and captures the main ingredients of realistic log-price models.

2.2. Main results

In this section we present the three major results of this chapter in Theorems 2.2.1, 2.2.2 and 2.2.3 and discuss the consequences. Theorems 2.2.1 and 2.2.2 establish functional stable central limit theorems in a general semimartingale setting for the spectral estimators of Reiß (2011) and Bibinger and Reiß (2014). Theorem 2.2.3 gives a multivariate limit theorem for the localized method of moments approach of Bibinger et al. (2014). These methods are briefly explained in Section 2.3. All estimators rely on optimal weight functions, which have to be estimated from the data, as well. The following theorems apply to the corresponding *adaptive* estimators, which are all defined in Section 2.3:

- the adaptive spectral estimator $\widehat{IV}_{n,t}$ of integrated volatility,
- the adaptive spectral estimator $\widehat{ICV}_{n,t}^{(p,q)}$ of integrated covolatility and
- the local method of moments estimator $LMM_{n,t}$ of the integrated covolatility matrix.

These estimators are defined in (2.3.8), in (2.3.16) and (2.3.19), respectively. They attain asymptotic efficiency lower variance bounds in the simplified models without drift, with independent volatility processes and Gaussian noise. The limit theorems are all based on the concept of *stable convergence* (see Section A.1 for details).

Theorem 2.2.1. *Let $d = 1$. Assume (SM-α-β) for some $\alpha \geq 1/2$, $\beta > 0$ and (N-1). Then we have the stable convergence*

$$n^{1/4}\left(\widehat{IV}_{n,t} - \int_0^t \sigma_s^2 ds\right) \xrightarrow{st} \int_0^t \sqrt{8\eta\,|\sigma_s^3|}d\widetilde{W}_s \tag{2.2.1}$$

as processes on $\mathcal{D}([0,1],\mathbb{R})$, where $\widetilde{W}$ is a Brownian motion defined on an independent extension of the original probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P})$. Moreover, the variance estimator $\widehat{\mathcal{V}}_{n,t}^{\mathcal{IV}}$ defined in (2.3.9) provides for fixed $0 \leq t \leq 1$ the feasible central limit theorem:

$$\left(\widehat{\mathcal{V}}_{n,t}^{\mathcal{IV}}\right)^{-1/2}\left(\widehat{IV}_{n,t} - \int_0^t \sigma_s^2 ds\right) \xrightarrow{d} N(0,1). \tag{2.2.2}$$

The convergence rate in (2.2.1) and (2.2.2) is *optimal* already in the parametric subexperiment (see Gloter and Jacod (2001)). $\widehat{IV}_{n,1}$ is asymptotically mixed normally distributed with random asymptotic variance $\int_0^1 8\eta|\sigma_s^3|ds$. This asymptotic variance coincides with the lower bound derived by Reiß (2011) in the subexperiment with time-varying but deterministic volatility, without drift and Gaussian error distribution. The spectral estimator of squared integrated volatility is hence *asymptotically efficient* in this setting. For the general semimartingale experiment the concept of asymptotic efficiency is not developed yet. It is conjectured that the lower bound has analogous structure (cf. Remark 3.1 of Jacod and Mykland (2015)). Theorem 2.2.1 establishes that the asymptotic variance of the estimator has the same form in the very general framework, what we call *quasi-efficient*, and stable convergence holds true. The *feasible limit theorem* (2.2.2) allows to provide confidence bands and is of pivotal importance for applications.

For the multi-dimensional setting define a diagonal matrix of asymptotic noise levels $\mathcal{H}(t) = \operatorname{diag}(\eta_p \nu_p^{1/2} F_p'(t)^{-1/2})_p \in \mathbb{R}^{d\times d}$, $0 \le t \le 1$.

Theorem 2.2.2. *Let $d \ge 1$. Assume (SM-α-β) for some $\alpha \ge 1/2$, $\beta > 0$ and (N-d). Then we have the stable convergence*

$$n^{1/4}\left(\widehat{ICV}_{n,t}^{(p,q)} - \int_0^t \Sigma_s^{(pq)} ds\right) \xrightarrow{st} \int_0^t v_s^{(p,q)} d\widetilde{W}_s \tag{2.2.3}$$

for $n/n_p \to \nu_p$ and $n/n_q \to \nu_q$ with $0 < \nu_p < \infty, 0 < \nu_q < \infty$, as processes on $\mathcal{D}([0,1],\mathbb{R})$, where $\widetilde{W}$ is a Brownian motion defined on an independent extension of the original probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\le t\le 1}, \mathbb{P})$. The asymptotic variance process satisfies

$$\begin{aligned}\left(v_s^{(p,q)}\right)^2 = 2\left(\left(\mathcal{H}(s)^{(p)}\right)^2\left(\mathcal{H}(s)^{(q)}\right)^2 (A_s^2 - B_s)B_s\right)^{1/2} \\ \times\left(\sqrt{A_s + \sqrt{A_s^2 - B_s}} - \operatorname{sgn}(A_s^2 - B_s)\sqrt{A_s - \sqrt{A_s^2 - B_s}}\right),\end{aligned} \tag{2.2.4}$$

where sgn denotes the sign function taking values in $\{-1,1\}$ such that $(v_s^{(p,q)})^2$ is always non-negative, and where

$$A_s = \Sigma_s^{(pp)}\left(\frac{\mathcal{H}(s)^{(q)}}{\mathcal{H}(s)^{(p)}}\right)^2 + \Sigma_s^{(qq)}\left(\frac{\mathcal{H}(s)^{(p)}}{\mathcal{H}(s)^{(q)}}\right)^2, \quad B_s = 4\left(\Sigma_s^{(pp)}\Sigma_s^{(qq)} + \left(\Sigma_s^{(pq)}\right)^2\right).$$

Moreover, the variance estimator $\widehat{\mathcal{V}}_{n,t}^{\mathcal{ICV}(p,q)}$ defined in (2.3.17) provides for fixed $0 \le t \le 1$ the feasible central limit theorem:

$$\left(\widehat{\mathcal{V}}_{n,t}^{\mathcal{ICV}(p,q)}\right)^{-1/2}\left(\widehat{ICV}_{n,t}^{(p,q)} - \int_0^t \Sigma_s^{(pq)} ds\right) \xrightarrow{d} N(0,1). \tag{2.2.5}$$

The bivariate extension of the spectral method outperforms by its local adaptivity and Fourier domain smoothing previous approaches for integrated covolatility estimation in most cases, see Bibinger and Reiß (2014) for a detailed discussion and survey on the different methods. It attains the multi-dimensional variance lower bound in the submodel for estimating the integrated covolatility $\int_0^1 \Sigma_s^{(pq)}\, ds$ only in case of zero correlation. On the other hand, the estimator already achieves a high efficiency and since it does not involve Fisher information weight matrices, it is less computationally costly than the efficient local method of moments approach. The general form of the asymptotic variance given in (2.2.4) is complicated. When $\Sigma_t^{(12)} = 0$ and for equal volatilities $\Sigma_t^{(11)} = \Sigma_t^{(22)} = \sigma_t$, it simplifies to $\int_0^t 4\eta|\sigma_s^3|\, ds$ which is efficient for this setup. The rescaled version in (2.2.5) allows for confidence bounds and is of high practical value.

For the general multi-dimensional setting we first have to introduce some additional notation. For matrices $A, B \in \mathbb{R}^{d\times d}$ the vec-operator $\operatorname{vec}(A) \in \mathbb{R}^{d^2}$ and the Kronecker product $A \otimes B \in \mathbb{R}^{d^2\times d^2}$ are defined by

$$\begin{aligned}&\operatorname{vec}(A) = \left(A_{11}, A_{21}, \ldots, A_{d1}, A_{12}, A_{22}, \ldots, A_{d2}, \ldots, A_{d(d-1)}, A_{dd}\right)^\top \in \mathbb{R}^{d^2},\\ &(A\otimes B)_{d(p-1)+q, d(p'-1)+q'} = A_{pp'}B_{qq'}, \quad p, q, p', q' = 1, \ldots, d.\end{aligned}$$

Note the crucial relation between the Kronecker product and the vec-operator $\operatorname{vec}(ABC) = (C^\top \otimes A)\operatorname{vec}(B)$. In the limit theorems, we need to account for effects by the non-commutativity of matrix multiplication. It will be useful to standardize the limit theorems such that the matrix

$$\mathcal{Z} = \operatorname{cov}(\operatorname{vec}(ZZ^\top)), \quad \text{for } Z \sim N(0, I_d) \ \text{ standard Gaussian,} \tag{2.2.6}$$

appears as covariance matrix of the standardized form instead of the identity matrix. This matrix is the sum of the d^2-dimensional identity matrix I_{d^2} and the so-called *commutation matrix* $C_{d,d}$ that maps a $(d \times d)$ matrix to the vectorization of its transpose, i.e. $C_{d,d}\operatorname{vec}(A) = \operatorname{vec}(A^\top)$. The *symmetrizer matrix* $\mathcal{Z}/2$ is idempotent (Abadir and Magnus (2009, Chapter 11)). For background information on matrix algebra, especially tensor calculus using the Kronecker product and vec-operator we refer to Abadir and Magnus (2009). With this preparation we can state the third main result in this chapter.

Theorem 2.2.3. *Let $d \geq 1$. Assume (SM-α-β) for some $\alpha \geq 1/2$, $\beta > 0$ and (N-d). Then we have the stable convergence*

$$n^{1/4}\Big(LMM_{n,t} - vec\Big(\int_0^t \Sigma_s\, ds\Big)\Big) \xrightarrow{st} \int_0^t \big(\Sigma_s^{1/2} \otimes \big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\big)\mathcal{Z} d\widetilde{W}_s$$
$$+ \int_0^t \big(\big(\Sigma_s^{\mathcal{H}}\big)^{1/4} \otimes \Sigma_s^{1/2}\big)\mathcal{Z} d\widetilde{W}_s^{\perp} \tag{2.2.7}$$

for $n/n_p \to \nu_p$ as $n \to \infty$ for $p = 1, \ldots, d$ as processes on $\mathcal{D}([0,1], \mathbb{R}^d)$, where $\widetilde{W}$ and $\widetilde{W}^{\perp}$ are independent d-dimensional Brownian motions defined on an independent extension of the original probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{0\leq t\leq 1}, \mathbb{P})$. Moreover, $\big(\Sigma^{\mathcal{H}}\big)^{1/4}$ is the square root of $\big(\Sigma^{\mathcal{H}}\big)^{1/2} := \mathcal{H}(\mathcal{H}^{-1}\Sigma\mathcal{H}^{-1})^{1/2}\mathcal{H}$ and $\mathcal{Z}$ is the matrix defined in (2.2.6). For $t = 1$ the pointwise marginal central limit theorem becomes

$$n^{1/4}\left(LMM_{n,1} - vec\Big(\int_0^1 \Sigma_s ds\Big)\right) \xrightarrow{st} MN\left(0, I^{-1}\mathcal{Z}\right), \tag{2.2.8}$$

where MN means mixed normal distribution with asymptotic covariance matrix

$$I^{-1} = 2\int_0^1 \Big(\Sigma_s \otimes \big(\Sigma_s^{\mathcal{H}}\big)^{1/2} + \big(\Sigma_s^{\mathcal{H}}\big)^{1/2} \otimes \Sigma_s\Big)\, ds. \tag{2.2.9}$$

Moreover, the covariance matrix estimator $I_{n,t}^{-1}$ defined in (2.3.20) provides for fixed $0 \leq t \leq 1$ the feasible central limit theorem:

$$I_{n,t}^{1/2}\left(LMM_{n,t} - vec\Big(\int_0^t \Sigma_s ds\Big)\right) \xrightarrow{d} N\left(0, \mathcal{Z}\right). \tag{2.2.10}$$

The local method of moments attains the lower asymptotic variance bound derived in Bibinger et al. (2014) for a nonparametric experiment with deterministic covolatility matrix, without drift and with Gaussian error distribution. Thus, the local method of moments is *asymptotically efficient* in this subexperiment.

The asymptotic variance of estimating integrated volatility decreases as we can benefit from observing correlated components. In the multi-dimensional observation model the minimum asymptotic variance can become much smaller than the bound in (2.2.1) for $d = 1$. In an idealized parametric model with $\sigma > 0$, the variance can be reduced up to $(8/\sqrt{d})\eta\sigma^3$ in comparison to the one-dimensional lower bound $8\eta\sigma^3$. See Bibinger et al. (2014) for a deeper discussion of the lower bound. In view of the complex geometry of the general multi-dimensional parameter space, expression (2.2.9) provides a neat description of the asymptotic variance bound.

2.3. Spectral estimation

In this section we review the building blocks of the spectral estimation method. In the following, denote by $\Delta_i^n Y^{(l)} = Y_i^{(l)} - Y_{i-1}^{(l)}$, $i = 1, \ldots, n_l$, $l = 1, \ldots, d$ the increments of $Y^{(l)}$ and analogously for X and other processes. For simplicity we assume first that $d = 1$, in the setting of Assumption (N-1). The time interval $[0, 1]$ is partitioned into equidistant bins $[(k-1)h_n, kh_n)$, $k = 1, \ldots, h_n^{-1} \in \mathbb{N}$, $h_n \to 0$ as $n \to \infty$ such that nh_n is the average number of observations per bin. Under Assumption (SM-α-β) for $\alpha \geq 1/2$, $\beta > 0$ we can always assume by an approximation argument that $\sigma_t^2 = \sigma_{(k-1)h_n}^2$ is locally constant on each bin for $t \in [(k-1)h_n, kh_n)$. On every block the integrated volatility $\int_{(k-1)h_n}^{kh_n} \sigma_s^2 ds = h_n \sigma_{(k-1)h_n}^2$ can then be estimated by $h_n \hat{\sigma}_{(k-1)h_n}^2$, solving locally parametric estimation problems. For this purpose Reiß (2011) uses the *spectral statistics*

$$S_{jk} = \sum_{i=1}^{n} \Delta_i^n Y \Phi_{jk}\left(\frac{i}{n}\right), \ j = 1, \ldots, \lfloor nh_n \rfloor - 1, \ k = 1, \ldots, h_n^{-1}, \tag{2.3.1}$$

which are discrete analogues of expressions obtained from diagonalizing the covariance operator of observations in an equivalent white noise experiment. Here,

$$\Phi_j(t) = \sqrt{\frac{2}{h_n}} \sin\left(j\pi h_n^{-1} t\right) \mathbf{1}_{[0,h_n]}(t), \quad j \geq 1, \, 0 \leq t \leq 1, \tag{2.3.2}$$

$$\varphi_j(t) = 2n\sqrt{\frac{2}{h_n}} \sin\left(\frac{j\pi}{2nh_n}\right) \cos\left(j\pi h_n^{-1} t\right) \mathbf{1}_{[0,h_n]}(t), \tag{2.3.3}$$

and $\Phi_{jk}(t) = \Phi_j(t-(k-1)h_n)$, $\varphi_{jk}(t) = \varphi_j(t-(k-1)h_n)$ are systems of trigonometric function orthogonal with respect to empirical scalar products $\langle \cdot, \cdot \rangle_n$ and $[\cdot, \cdot]_n$, respectively (see Section 2.6.1 for details). Efficient estimators $\hat{\sigma}_{(k-1)h_n}^2$ for $\sigma_{(k-1)h_n}^2$ are then constructed by weighted linear combinations of bias corrected squared spectral statistics over different frequencies.

The spectral statistics are the principal elements of the considered estimation techniques. They are related to the pre-averages of Jacod et al. (2009). An important difference is that we keep the bins fixed which simplifies the construction of the spectral approach. Bin-wise the spectral estimation benefits from an advanced smoothing method in the frequency domain by using the weight function of a discrete sine transformation. The spectral statistics hence *de-correlate* the observations and form their bin-wise principal components. The methodology can be viewed also as localizing the estimator of

Curci and Corsi (2012) on bins. Reiß (2011) showed that this leads to a semiparametrically efficient estimation approach of integrated volatility in a nonparametric setup with deterministic volatility, without drift and with Gaussian noise.

The bin-width is chosen as $h_n \asymp n^{-1/2}\log n$ in order to attain the optimal convergence rates for the results in Section 2.2. This becomes clear in the proofs in Section 2.6. The log-factor plays a role in the convergence of the sum of variances over different frequencies. The leading asymptotic order $n^{-1/2}$ for the bin-width is analogous to the pre-average and kernel bandwidths, cf. Jacod et al. (2009) and Barndorff-Nielsen et al. (2008), and balances the discretization error which increases with increasing h_n and the error due to noise which decreases as h_n increases. Let us point out that the basis functions (2.3.2) and (2.3.3) are scaled versions of the respective basis functions in Bibinger and Reiß (2014) and Bibinger et al. (2014) for a more convenient exposition.

2.3.1. The spectral estimator of integrated volatility

As above let $d = 1$ and assume (N-1). In order to estimate $\sigma^2_{(k-1)h_n}$ on block k, the spectral statistics S_{jk} in (2.3.1) have to be squared. This creates a bias. After correcting for this bias, we form weighted linear combinations:

$$\widehat{\sigma}^2_{(k-1)h_n} = \sum_{j=1}^{\lfloor nh_n\rfloor -1} w_{jk}\Big(S_{jk}^2 - [\varphi_{jk},\varphi_{jk}]_n\,\frac{\eta^2}{n}\Big). \tag{2.3.4}$$

The weights w_{jk} are defined below. This means that we consider only asymptotically infinitely many frequencies. As the proofs reveal it is even sufficient to consider only the first $J_n \asymp \log n$ frequencies, because higher frequencies are asymptotically negligible. The bias correction incorporates the noise level η which is unknown in general. It can be consistently estimated from the data with rate of convergence $n^{-1/2}$, for instance by $\hat{\eta}^2 = (2n)^{-1}\sum_{i=1}^n(\Delta_i^n Y)^2$, see Zhang et al. (2005) for an asymptotic analysis of this estimator. The principle of bias-correcting the squared spectral statistics relates to the early estimator of Zhou (1998) for volatility estimation under microstructure noise. Set $I_{jk} = \big(\mathrm{Var}\big(S_{jk}^2|\mathcal{G}_{(k-1)h_n}\big)\big)^{-1}$ to be the inverse conditional variance of the S_{jk}^2 and let $I_k = \sum_{j=1}^{\lfloor nh_n\rfloor -1} I_{jk}$. The variance of $\widehat{\sigma}^2_{(k-1)h_n}$ becomes minimal and equal to $\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} I_k^{-1}$ with oracle weights

$$w_{jk} = I_k^{-1} I_{jk} = \frac{\Big(\sigma^2_{(k-1)h_n} + \frac{\eta^2}{n}\,[\varphi_{jk},\varphi_{jk}]_n\Big)^{-2}}{\sum_{m=1}^{\lfloor nh_n\rfloor -1}\Big(\sigma^2_{(k-1)h_n} + \frac{\eta^2}{n}\,[\varphi_{mk},\varphi_{mk}]_n\Big)^{-2}} \tag{2.3.5}$$

for $k = 1,\ldots,h_n^{-1}$ and $j = 1,\ldots,\lfloor nh_n\rfloor - 1$, when the noise is Gaussian. For a general noise distribution the first-order variance of $\widehat{\sigma}^2_{(k-1)h_n}$ is not affected. The estimator of $\int_0^t \sigma_s^2\,ds$ is constructed as Riemann-sum

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n\widehat{\sigma}^2_{(k-1)h_n} = \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n \sum_{j=1}^{\lfloor nh_n\rfloor -1} w_{jk}\Big(S_{jk}^2 - [\varphi_{jk},\varphi_{jk}]_n\,\frac{\eta^2}{n}\Big)\,, \tag{2.3.6}$$

such that the estimator at $t = 1$ becomes simply the average of local estimates in the case of equispaced bins.

It is essential to develop an adaptive version of the estimator, for which we replace the optimal oracle weights by data-driven weights. Additionally to the estimated noise variance, a bin-wise consistent estimator of the $\sigma^2_{(k-1)h_n}$ with some convergence rate suffices. Local pre-estimates can be constructed by using the same ansatz as in (2.3.4), but involving only $J_n \ll \lfloor nh_n \rfloor - 1$ frequencies with constant weights $w_{jk} = J_n^{-1}$ and then averaging over $(2K_n + 1) \asymp n^{1/4}$ bins in a neighborhood of $(k-1)h_n$:

$$\widehat{\sigma}^{2,\text{pilot}}_{(k-1)h_n} = (2K_n+1)^{-1} \sum_{m=(k-1-K_n)\vee 1}^{(k-1+K_n)\wedge h_n^{-1}} J_n^{-1} \sum_{j=1}^{J_n} \Big(S^2_{jm} - [\varphi_{jm}, \varphi_{jm}]_n \frac{\widehat{\eta}^2}{n} \Big) . \tag{2.3.7}$$

This estimator attains $n^{-1/8}$ as rate of convergence under Assumption (SM-α-β) for $\alpha \geq 1/2$, $\beta > 0$. The estimated weights are then given by $\widehat{w}_{jk} = \widehat{I}_k^{-1} \widehat{I}_{jk}$ where $\widehat{I}_k, \widehat{I}_{jk}$ are obtained as above but plugging in the pre-estimates of $\sigma^2_{(k-1)h_n}$ and of η^2. The fully adaptive spectral estimator of integrated volatility and the estimator for its variance are then given by the following two-stage approach:

$$\widehat{IV}_{n,t} = \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n \sum_{j=1}^{\lfloor nh_n \rfloor -1} \widehat{w}_{jk} \Big(S^2_{jk} - [\varphi_{jk}, \varphi_{jk}]_n \frac{\widehat{\eta}^2}{n} \Big), \tag{2.3.8}$$

$$\widehat{\mathcal{V}}^{\mathcal{IV}}_{n,t} = \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n^2 \, \hat{I}_k^{-1}. \tag{2.3.9}$$

2.3.2. The spectral covolatility estimator

Let now $d \geq 1$ and assume (N-d). The spectral covolatility estimator from Bibinger and Reiß (2014) is the obvious extension of the one-dimensional estimator using cross products of the corresponding spectral statistics

$$S^{(p)}_{jk} = \sum_{i=1}^{n_p} \Delta_i^n Y^{(p)} \Phi_{jk} \left(\frac{t_i^{(p)} + t_{i-1}^{(p)}}{2} \right) \quad , j \geq 1, p = 1, \ldots, d, k = 1, \ldots, h_n^{-1}. \tag{2.3.10}$$

The basis functions Φ_{jk} are defined as in (2.3.2). Instead of (2.3.4) we then have the local estimator for $p \neq q$

$$\widehat{\Sigma}^{(p,q)}_{(k-1)h_n} = \sum_{j=1}^{\lfloor nh_n \rfloor -1} w^{p,q}_{jk} \Big(S^{(p)}_{jk} S^{(q)}_{jk} \Big). \tag{2.3.11}$$

Differently from (2.3.8) there is no bias correction for the cross product $S^{(p)}_{jk} S^{(q)}_{jk}$, because the noise is component-wise independent. For the optimal weights observe that by a locally constant approximation of the observation frequencies that we get a bin-wise locally constant approximation of $\mathcal{H}$:

$$\mathrm{H}^n_k = \operatorname{diag}\left(n_p^{-1} \eta_p^2 \nu_p F_p'((k-1)\, h_n)^{-1} \right)_p = \operatorname{diag}\left(H_p^{(k-1)h_n} \right)_p . \tag{2.3.12}$$

Moreover, $[\varphi_{jk},\varphi_{jk}]=\int_0^1\varphi_{jk}^2(t)\,dt=h_n^{-2}\pi^2j^2$, where we use for non-synchronous observations instead of (2.3.3) the simpler expression $\varphi_{jk}=\Phi'_{jk}$ (*). The optimal weights are then given by $w_{jk}^{p,q}=(I_k^{(p,q)})^{-1}I_{jk}^{(p,q)}$ with

$$I_{jk}^{(p,q)}=\Big(\Sigma^{(pp)}_{(k-1)h_n}\Sigma^{(qq)}_{(k-1)h_n}+(\Sigma^{(pq)}_{(k-1)h_n})^2+H_p^{(k-1)h_n}H_q^{(k-1)h_n}\,[\varphi_{jk},\varphi_{jk}]^2 \\ +\big(\Sigma^{(pp)}_{(k-1)h_n}H_q^{(k-1)h_n}+\Sigma^{(qq)}_{(k-1)h_n}H_p^{(k-1)h_n}\big)\,[\varphi_{jk},\varphi_{jk}]\Big)^{-1}. \tag{2.3.13}$$

They depend on the volatilities, the covolatility and the noise levels with respect to p,q. The local noise levels $H_p^{(k-1)h_n}$ can be estimated by

$$\widehat{H}_p^{(k-1)h_n}=\frac{\sum_{i=1}^{n_p}\big(\Delta_iY^{(p)}\big)^2}{2h_n}\sum_{(k-1)h_n\le t_v^{(p)}\le kh_n}\big(t_v^{(p)}-t_{v-1}^{(p)}\big)^2\,, \tag{2.3.14}$$

(cf. (2.6.5)). Averaging empirical covariances $S_{jk}S_{jk}^\top$ over different spectral frequencies $j=1,\ldots,J_n$ and over a set of $(2K_n+1)$ adjacent bins yields a consistent estimator of the instantaneous covolatility matrix:

$$\widehat{\Sigma}_s^{pilot}=(2K_n+1)^{-1}\sum_{k=\lfloor sh_n^{-1}\rfloor-K_n}^{\lfloor sh_n^{-1}\rfloor+K_n}J_n^{-1}\sum_{j=1}^{J_n}\Big(S_{jk}S_{jk}^\top-[\varphi_{jk},\varphi_{jk}]\,\widehat{\mathrm{H}}_k^n\Big), \tag{2.3.15}$$

where $\widehat{\mathrm{H}}_k^n=\mathrm{diag}(\widehat{H}_p^{kh_n})$. End effects for $s<K_nh_n$ and $s>1-K_nh_n$ are not discussed here. Adaptive pre-estimated weights $\hat{w}_{jk}^{p,q}$ can then obtained again by plug-in from (2.3.13). The bivariate spectral covolatility estimator with adaptive weights is then given by

$$\widehat{ICV}_{n,t}^{(p,q)}=\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}h_n\sum_{j=1}^{\lfloor nh_n\rfloor-1}\widehat{w}_{jk}^{p,q}\Big(S_{jk}^{(p)}S_{jk}^{(q)}\Big). \tag{2.3.16}$$

The estimator of the variance is

$$\widehat{\mathcal{V}}_{n,t}^{\mathcal{ICV}(p,q)}=\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}h_n^2\,\big((\widehat{I}_k)^{(p,q)}\big)^{-1}\,. \tag{2.3.17}$$

A more general version of the spectral covolatility estimator for a model including cross-correlation of the noise (in a synchronous framework) can be found in Bibinger and Reiß (2014). For a simpler exposition and since this notion of cross-correlation is not adequate for the more important non-synchronous case, we restrict ourselves here to noise according to Assumption (N-d).

* This meets the original idea by Reiß (2011) for continuous-time observations to use orthogonal systems of functions and their derivatives. While in the case of regular observations on the grid $i/n, i=0,\ldots,n$, we can slightly benefit from discrete Fourier analysis and the exact form of the φ_{jk}, for non-synchronous observations we rely on continuous-time analogues as approximations which coincide discrete expressions at first order.

2.3.3. Local method of moments

Let $S_{jk} = (S_{jk}^{(p)})_p$ be the vector of the spectral statistics in (2.3.10) for $k = 1, \ldots, h_n^{-1}$ and $j \geq 1$. The fundamental novelty of the local method of moments approach is to involve multivariate Fisher informations as optimal weight matrices which are $(d^2 \times d^2)$ matrices of the following form:

$$W_{jk} = I_k^{-1} I_{jk} = \Big(\sum_{u=1}^{\lfloor nh_n \rfloor - 1} \big(\Sigma_{(k-1)h_n} + [\varphi_{uk}, \varphi_{uk}] \mathrm{H}_k^n\big)^{-\otimes 2}\Big)^{-1} \cdot \big(\Sigma_{(k-1)h_n} + [\varphi_{jk}, \varphi_{jk}] \mathrm{H}_k^n\big)^{-\otimes 2}, \tag{2.3.18}$$

where $A^{\otimes 2} = A \otimes A$ and $A^{-\otimes 2} = (A^{\otimes 2})^{-1} = (A^{-1})^{\otimes 2}$. The main difference to the estimators in (2.3.8) and (2.3.16) is that for estimating one specific (co-)volatility of one (two) components, the estimator in (2.3.18) does not only rely on observations of the one (two) considered component(s) but benefits from information contained in all other correlated components. In general, this facilitates a much smaller variance in the multivariate model.

We immediately define the adaptive estimator. With the pilot estimates (2.3.15) and estimators for the noise level (2.3.14) at hand, we derive pre-estimated weight matrices $\widehat{W}_{jk}$ similar as above. The final estimator of $\mathrm{vec}(\int_0^t \Sigma_s \, ds)$ is

$$LMM_{n,t} = \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n \sum_{j=1}^{\lfloor nh_n \rfloor - 1} \widehat{W}_{jk} \mathrm{vec}\Big(S_{jk} S_{jk}^{\top} - [\varphi_{jk}, \varphi_{jk}] \, \widehat{\mathrm{H}}_k^n\Big) \,, \tag{2.3.19}$$

and the estimator of its covariance matrix is

$$\widehat{I}_{n,t}^{-1} = \sum_{k=0}^{\lfloor th_n^{-1} \rfloor} h_n^2 \Big(\sum_{j=1}^{\lfloor nh_n \rfloor - 1} \widehat{I}_{jk} \Big)^{-1} . \tag{2.3.20}$$

Our method is different from the approach of **?**, even though they have similar names. A common feature is the two-stage adaptivity where pre-estimated spot volatilities are plugged in for the final estimator.

2.4. Asymptotic theory

We begin with the one-dimensional experiment. X is decomposed by a locally constant approximation of σ and the approximation error:

$$X_t = X_0 + \widetilde{X}_t + (X_t - X_0 - \widetilde{X}_t) \,, \tag{2.4.1}$$

where $\widetilde{X}_t = \int_0^t \sigma_{\lfloor sh_n^{-1} \rfloor h_n} dW_s$ is a simplified process without drift and with locally constant volatility. We distinguish between $\widehat{IV}_{n,t}^{or}(Y)$, the oracle version of the spectral volatility estimator (2.3.8) from noisy observations, and $\widehat{IV}_{n,t}^{or}(\widetilde{X} + \varepsilon)$ for the oracle estimator in a simplified experiment in which $\widetilde{X}$ instead of X is observed with noise.

In order to establish a functional limit theorem, decompose the estimation error of the oracle version of (2.3.8) in the following way:

$$\widehat{IV}^{or}_{n,t}(Y) - \int_0^t \sigma_s^2\, ds = \widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon) - \int_0^t \sigma^2_{\lfloor s h_n^{-1}\rfloor h_n}\, ds \tag{2.4.2}$$

$$+ \widehat{IV}^{or}_{n,t}(Y) - \widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon) - \int_0^t \left(\sigma_s^2 - \sigma^2_{\lfloor s h_n^{-1}\rfloor h_n}\right) ds\,. \tag{2.4.3}$$

We first prove the result of Theorem 2.2.1 for the right-hand side of (2.4.2). In the second step the approximation error in (2.4.3) is shown to be asymptotically negligible. Finally, we establish that the same functional stable CLT carries over to the adaptive estimators by proving that the error of the plug-in estimation of optimal weights is asymptotically negligible.

Proposition 2.4.1. *We have under the assumptions of Theorem 2.2.1 that*

$$n^{1/4}\left(\widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon) - h_n \sum_{k=1}^{\lfloor t h_n^{-1}\rfloor} \sigma^2_{(k-1)h_n}\right) \xrightarrow{st} \int_0^t \sqrt{8\eta\,|\sigma_s^3|}\, d\widetilde{W}_s, \tag{2.4.4}$$

as $n\to\infty$ on $\mathcal{D}([0,1],\mathbb{R})$, where $\widetilde{W}$ is a Brownian motion defined on an independent extension of the original probability space $(\Omega,\mathcal{G},(\mathcal{G}_t)_{0\le t\le 1},\mathbb{P})$.

Proposition 2.4.2. *We have under the assumptions of Theorem 2.2.1 that*

$$n^{1/4}\left(\widehat{IV}^{or}_{n,t}(Y) - \widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon) - \int_0^t \left(\sigma_s^2 - \sigma^2_{\lfloor s h_n^{-1}\rfloor h_n}\right) ds\right) \xrightarrow{ucp} 0,\ \text{as } n\to\infty\,. \tag{2.4.5}$$

Theorem 2.2.1 is then an immediate consequence of the following proposition:

Proposition 2.4.3. *We have under the assumptions of Theorem 2.2.1 that*

$$n^{1/4}\left|\widehat{IV}_{n,t} - \widehat{IV}^{or}_{n,t}(Y)\right| \xrightarrow{ucp} 0, \qquad \text{as } n\to\infty. \tag{2.4.6}$$

Finally, by consistency of the variance estimators and Slutsky's Lemma the feasible limit theorems for the adaptive estimators are valid. The proof of Proposition 2.4.1 is based on Theorem A.1.2. For this rewrite the rescaled estimation error as a sum of increments:

$$n^{1/4}\left(\widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon) - h_n \sum_{k=1}^{\lfloor t h_n^{-1}\rfloor} \sigma^2_{(k-1)h_n}\right) = \sum_{k=1}^{\lfloor t h_n^{-1}\rfloor} \zeta_k^n, \tag{2.4.7}$$

$$\zeta_k^n = n^{1/4} h_n \sum_{j=1}^{\lfloor n h_n\rfloor -1} w_{jk}\left(\tilde{S}_{jk}^2 - \mathbb{E}\left[\tilde{S}_{jk}^2\,\middle|\,\mathcal{G}_{(k-1)h_n}\right]\right),\ k=1,\ldots,h_n^{-1}, \tag{2.4.8}$$

where the $\tilde{S}_{jk}$ correspond to the spectral statistics in (2.3.1) with respect to $\widetilde{X}+\varepsilon$ instead of $X+\varepsilon$. In order to apply Theorem A.1.2 we will verify the following five conditions:

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[\zeta_k^n\middle|\mathcal{G}_{(k-1)h_n}\right]\stackrel{ucp}{\longrightarrow}0, \tag{J1}$$

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[(\zeta_k^n)^2\middle|\mathcal{G}_{(k-1)h_n}\right]\stackrel{\mathbb{P}}{\to}\int_0^t v_s^2ds, \tag{J2}$$

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[(\zeta_k^n)^4\middle|\mathcal{G}_{(k-1)h_n}\right]\stackrel{\mathbb{P}}{\to}0, \tag{J3}$$

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[\zeta_k^n(W_{kh_n}-W_{(k-1)h_n})\middle|\mathcal{G}_{(k-1)h_n}\right]\stackrel{\mathbb{P}}{\to}0, \tag{J4}$$

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[\zeta_k^n(N_{kh_n}-N_{(k-1)h_n})\middle|\mathcal{G}_{(k-1)h_n}\right]\stackrel{\mathbb{P}}{\to}0, \tag{J5}$$

with $v_s=\sqrt{8\eta|\sigma_s|^3}$ and where N in (J5) is any bounded martingale orthogonal to W.

Next, we consider the covolatility estimator (2.3.16) and the local method of moments approach (2.3.19). A non-degenerate asymptotic variance is obtained when $n/n_p\to\nu_p$ with $0<\nu_p<\infty$ as $n\to\infty$ for all $p=1,\dots,d$. In the idealized martingale framework Bibinger et al. (2014) have found that non-synchronicity effects are asymptotically negligible in terms of the information content of the underlying experiments by a (strong) asymptotic equivalence in the sense of Le Cam of the discrete non-synchronous and a continuous-time observation model. This yields a fundamental difference to the no-noise case where the asymptotic variance of the prominent Hayashi-Yoshida estimator in the functional CLT hinges on interpolation effects, see Hayashi and Yoshida (2011). In the presence of the dominant noise part, however, at the slower optimal convergence rate, the influence of sampling schemes boils down to local observation densities. These time-varying local observation densities are shifted to locally time-varying noise levels (indeed locally increased noise is equivalent to locally less frequent observations). Here, we shall explicitly prove that the error incurred by passing from a non-synchronous to a synchronous reference scheme is asymptotically negligible.

Lemma 2.4.4. *Let $\bar{t}_i^{(l)}=\big(t_i^{(l)}+t_{i-1}^{(l)}\big)/2$ for $l=1,\dots,d$. Under Assumptions (SM-α-β) for $\alpha\ge 1/2$ and $\beta>0$ and (N-d) we can work with synchronous sampling when considering the signal part X, i.e. for $l,m=1,\dots,d$ uniformly in t and with $w_{jk}^{l,m}$ as in*

Section 2.3.2 or defined as entries of (2.3.18):

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n \sum_{j\geq 1} w_{jk}^{l,m} \sum_{v=1}^{n_l} \Big(X_{t_v^{(l)}}^{(l)} - X_{t_{v-1}^{(l)}}^{(l)}\Big)\Phi_{jk}(\bar{t}_v^{(l)})$$
$$\cdot \sum_{i=1}^{n_m} \Big(X_{t_i^{(m)}}^{(m)} - X_{t_{i-1}^{(m)}}^{(m)}\Big)\Phi_{jk}(\bar{t}_i^{(m)}) + o_{\mathbb{P}}(n^{-1/4})$$
$$= \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n \sum_{j\geq 1} w_{jk}^{l,m} \sum_{v=1}^{n_l} \Big(X_{t_v^{(l)}}^{(l)} - X_{t_{v-1}^{(l)}}^{(l)}\Big)\Phi_{jk}(\bar{t}_v^{(l)}) \sum_{i=1}^{n_l} \Big(X_{t_i^{(l)}}^{(m)} - X_{t_{i-1}^{(l)}}^{(m)}\Big)\Phi_{jk}(\bar{t}_i^{(l)}).$$

Note that $(F_l^{-1})', (F_m^{-1})'$ affect the asymptotics of our estimators as can be seen in (2.2.4). They are treated as part of the summands due to noise.

Under a synchronous reference observation scheme the strategy of the asymptotic analysis is similar to the one-dimensional setup. Analogous decompositions of the leading terms from the simplified model without drift and with a locally constant covolatility matrix and remainders are considered for the multivariate method of moments estimator (2.3.19) and the spectral covolatility estimator (2.3.16). In order to prove Theorem 2.2.2 for instance, we apply Theorem A.1.2 to the sum of increments

$$\zeta_k^n = n^{\frac{1}{4}} h_n \Big(\sum_{j\geq 1} w_{jk}^{p,q} \big(\tilde{S}_{jk}^{(p)} \tilde{S}_{jk}^{(q)}\big) - \mathbb{E}\left[\tilde{S}_{jk}^{(p)} \tilde{S}_{jk}^{(q)} \Big| \mathcal{G}_{(k-1)h_n}\right] \Big) \tag{2.4.9}$$

for $k = 1, \ldots, h_n^{-1}$ with $\tilde{S}_{jk}^{(p)}$ as defined in (2.3.10) with respect to $\widetilde{X} + \varepsilon$ instead of $X + \varepsilon$. By including the case $p = q$ with a bias correction the one-dimensional result is generalized to non-equidistant sampling.

The asymptotic negligibility of the plug-in estimation in Proposition 2.4.3 is proven in Section 2.6 exploiting a uniform bound on the derivative of the weights as function of σ_t. In fact, it turns out that the weights are robust enough under misspecification of the pre-estimated local volatility to render the difference between oracle and adaptive estimators asymptotically negligible. This carries over to the multivariate methods.

2.5. Simulations

We study now the finite sample performance of the one-dimensional spectral estimator in (2.3.8) in a random volatility scenario. We sample regular observations $Y_1, \ldots, Y_n$ as in Assumption (N-1) with $\varepsilon_i \overset{iid}{\sim} N(0, \eta^2)$ and the simulated diffusion

$$X_t = bt + \int_0^t \sigma_s dW_s.$$

In a first baseline scenario configuration let $\sigma_s = 1$ be constant. In a second more realistic scenario we consider

$$\sigma_t^2 = \Big(\int_0^t \tilde{\sigma} \cdot \lambda \, dW_s + \int_0^t \sqrt{1-\lambda^2} \cdot \tilde{\sigma} \, dW_s^{\perp} \Big) \cdot f(t)\,, \tag{2.5.1}$$

n	σ	h_n^{-1}	η	λ	$\mathrm{RE}(\widehat{IV}^{or}_{n,1})$	$\mathrm{RE}(\widehat{IV}_{n,1})$
30000	1	25	0.01	–	1.01	1.43
5000	1	25	0.01	–	1.02	1.47
30000	Eq. (2.5.1)	25	0.01	0.5	1.09	1.75
30000	Eq. (2.5.1)	25	0.01	0.2	1.06	1.77
30000	Eq. (2.5.1)	25	0.01	0.8	1.09	1.75
30000	Eq. (2.5.1)	25	0.001	0.5	1.62	1.88
30000	Eq. (2.5.1)	25	0.1	0.5	1.20	1.69
30000	Eq. (2.5.1)	50	0.01	0.5	1.09	1.84
30000	Eq. (2.5.1)	10	0.01	0.5	1.16	1.86
5000	Eq. (2.5.1)	25	0.01	0.5	1.13	1.92
5000	Eq. (2.5.1)	50	0.01	0.5	1.08	1.75
5000	Eq. (2.5.1)	10	0.01	0.5	1.09	1.87

Table 2.1.: Relative Efficiencies (RE) of oracle and adaptive spectral integrated volatility estimator in finite-sample Monte Carlo study.

with $W^{\perp}$ being a standard Brownian motion independent of W and f being a deterministic seasonality function

$$f(t) = 0.1(1 - t^{\frac{1}{3}} + 0.5 \cdot t^2)$$

such that $\sigma_0^2 = 0.1$. Set $b = 0.1$ and $\tilde{\sigma} = 0.01$.
The superposition of a continuous semimartingale as random component with a time-varying seasonality modeling the volatility's typical U-shape mimics very general realistic volatility characteristics. We implement the oracle version of the estimator (2.3.8) and the adaptive two-stage procedure with pre-estimated optimal weights. Table 2.1 presents Monte Carlo results for different scenario configurations. In particular, we consider different tuning parameters (bin-widths) and possible dependence of the finite-sample behavior on the leverage magnitude and the magnitude of the noise variance. We compute the estimators' root mean squared errors (RMSE) at $t = 1$, for each configuration based on 1000 Monte Carlo iterations, and fix in each configuration one realization of a volatility path to compare the RMSEs to the theoretical asymptotic counterparts in the realized relative efficiency (RE):

$$\mathrm{RE}(\widehat{IV}_{n,1}) = \frac{\sqrt{\Big((\mathrm{mean}(\widehat{IV}_{n,1}) - \int_0^1 \sigma_s^2\,ds)^2 + \mathrm{Var}(\widehat{IV}_{n,1})\Big) \cdot \sqrt{n}}}{\sqrt{8\eta \int_0^1 \sigma_s^3\,ds}}\,. \tag{2.5.2}$$

The standard sample size is $n = 30000$, a realistic number of observations in usual high-frequency applications as number of ticks over one trading day for liquid assets at NASDAQ. We also focus on smaller samples, $n = 5000$. Throughout all simulations we fix a maximum spectral cut-off $J_p = 100$ in the pre-estimation step and $J = 150$ for the final estimator, which is large enough to render the approximation error by neglecting higher frequencies negligible.

The Monte Carlo study confirms that the estimator performs well in practice and the Monte Carlo variances come very close to the predicted lower bound, even in the complex "wiggly" volatility setting. The fully adaptive approach performs worse than the oracle estimator which is in light of previous results on related estimation approaches not surprising, see e.g. Bibinger and Reiß (2014) for a study including an adaptive multi-scale estimator (global smoothing parameter, but chosen data-driven). Still the adaptive estimator's performance is remarkably well in almost all configurations. For very small noise level, the relative efficiency is not as close to 1 any more. Apart from this case, the RE comes very close to 1 for the oracle estimator, not depending on the magnitude of leverage, also for small samples, and being very robust with respect to different bin-widths. A simulation study of the multivariate method of moments estimator in a random volatility setup can be found in Bibinger et al. (2014).

2.6. Proofs

2.6.1. Preliminaries

The section prepares the actual proofs by introducing some additional notation and properties of empirical scalar products, as well as some necessary reductions for the process X.

Empirical scalar products

Definition 2.6.1. Let $f, g : [0,1] \to \mathbb{R}$ be functions and let $z = (z_i)_{1\le i\le n} \in \mathbb{R}^n$. We call

$$\langle f, g\rangle_n = \frac{1}{n}\sum_{i=1}^{n} f\left(\frac{i}{n}\right) g\left(\frac{i}{n}\right) ,$$
$$\langle z, g\rangle_n = \frac{1}{n}\sum_{i=1}^{n} z_i\, g\left(\frac{i}{n}\right) ,$$

empirical scalar product of f, g and of z, g, respectively. We further define the "shifted" empirical scalar products

$$[f, g]_n = \frac{1}{n}\sum_{i=1}^{n} f\left(\frac{i-\frac{1}{2}}{n}\right) g\left(\frac{i-\frac{1}{2}}{n}\right) ,$$
$$[z, g]_n = \frac{1}{n}\sum_{i=1}^{n} z_i\, g\left(\frac{i-\frac{1}{2}}{n}\right) .$$

Recall the notation $\Delta^n Y = \left(\Delta_i^n Y\right)_{1\le i\le n} \in \mathbb{R}^n$, the vector of increments and analogously $\Delta^n X$ and let $\varepsilon = (\varepsilon_i)_{0\le i\le (n-1)}$.

Lemma 2.6.2. *It holds that*

$$\langle \Phi_{jk}, \Phi_{mk} \rangle_n = \delta_{jm}\,, \tag{2.6.1}$$

$$[\varphi_{jk}, \varphi_{mk}]_n = \delta_{jm} 4n^2 \sin^2\left(\frac{j\pi}{2\lfloor nh_n \rfloor}\right) \tag{2.6.2}$$

$$\left[\varphi_{jk}^2, \varphi_{mk}^2\right]_n = (2+\delta_{jm})\, n^2 \sin\left(\frac{j\pi}{\lfloor nh_n \rfloor}\right) \sin\left(\frac{m\pi}{\lfloor nh_n \rfloor}\right). \tag{2.6.3}$$

Furthermore, we have the summation by parts decomposition of spectral statistics:

$$\langle n\Delta^n Y, \Phi_{jk} \rangle_n = \langle n\Delta^n X, \Phi_{jk} \rangle_n - [\varepsilon, \varphi_{jk}]_n\,. \tag{2.6.4}$$

Proof. The proofs of the orthogonality relations (2.6.1) and (2.6.2) are similar and we restrict ourselves to proving (2.6.2). In the following we use the shortcut $N = \lfloor nh_n \rfloor$ and without loss of generality we consider the first bin $k = 1$. We make use of the trigonometric addition formulas which yield for $N \geq j \geq r \geq 1$:

$$\begin{aligned} &\cos(j\pi N^{-1}(l+\tfrac{1}{2}))\cos(r\pi N^{-1}(l+\tfrac{1}{2})) \\ &\quad = \cos((j+r)\pi N^{-1}(l+\tfrac{1}{2})) + \cos((j-r)\pi N^{-1}(l+\tfrac{1}{2})). \end{aligned}$$

We show that $\sum_{i=0}^{N-1}\cos(m\pi N^{-1}(i+\frac{1}{2})) = 0$ for $m \in \mathbb{N}$. First, consider m odd:

$$\begin{aligned} &\sum_{i=0}^{N-1}\cos(m\pi N^{-1}(i+\tfrac{1}{2})) \\ &= \sum_{i=0}^{\lfloor (N-2)/2 \rfloor}\cos(m\pi N^{-1}(i+\tfrac{1}{2})) + \sum_{i=\lceil N/2 \rceil}^{N-1}\cos(m\pi N^{-1}(i+\tfrac{1}{2})) \\ &= \sum_{i=0}^{\lfloor (N-2)/2 \rfloor}\Big(\cos(m\pi N^{-1}(i+\tfrac{1}{2})) + \cos(m\pi N^{-1}(N-(i+\tfrac{1}{2})))\Big) \\ &= 0, \end{aligned}$$

since $\cos(x+\pi m) = -\cos(x)$ for m odd. Note that for $i = (N-1)/2 \in \mathbb{N}$, we leave out one addend which equals $\cos(m\pi/2) = 0$, and also that for m even by $\cos(x) = \cos(x+m\pi)$ the two sums are equal. Since $\cos(0) = 1$, this also implies the empirical norm for $j = r$.

For $m \in \mathbb{N}$ with m even, we differentiate the cases $N = 4k, k \in \mathbb{N}$; $N = 4k+2, k \in \mathbb{N}$ and $N = 2k+1, k \in \mathbb{N}$. If $N = 4k+2$, we decompose the sum as follows:

$$\begin{aligned} &\sum_{i=0}^{N-1}\cos(m\pi N^{-1}(i+\tfrac{1}{2})) \\ &= \sum_{i=0}^{2k}\cos(m\pi(4k+2)^{-1}(i+\tfrac{1}{2})) + \sum_{i=2k+1}^{4k+1}\cos(m\pi(4k+2)^{-1}(i+\tfrac{1}{2})). \end{aligned}$$

The addends of the left-hand sum are symmetric around the point $m\pi/4$ at $i = k$ and of the right-hand sum around $3m\pi/4$ at $i = 3k+1$. Thereby, both sums equal zero by

symmetry. More precisely, for m being not a multiple of 4 the sums directly yield zero. If m is a multiple of 4, we can split the sum into two or more sums which then equal zero again.
This observation for the first sum readily implies $\sum_{i=0}^{N-1}\cos(m\pi N^{-1}(i+\frac{1}{2}))=0$ for $N=2k+1$, since in this case

$$\sum_{i=0}^{2k}\cos(m\pi N^{-1}(i+\tfrac{1}{2}))=\sum_{i=1}^{2k}\cos(2m\pi(4k+2)^{-1}(i+\tfrac{1}{2}))=0\,.$$

For $N=4k$, we may as well exploit symmetry relations of the cosine. Decompose the sum

$$\sum_{i=0}^{N-1}\cos(m\pi N^{-1}(i+\tfrac{1}{2}))=\sum_{i=0}^{2k-1}\cos(m\pi(4k)^{-1}(i+\tfrac{1}{2}))+\sum_{i=2k}^{4k-1}\cos(m\pi(4k)^{-1}(i+\tfrac{1}{2}))\,.$$

Symmetry around $m\pi/4$ and $3m\pi/4$ is similar as above, but these points lie off the discrete grid this time. Yet, analogous reasoning as above yields that both sums equal zero again, what completes the proof of (2.6.2). Likewise and using

$$\begin{aligned}&\cos^2(x)\cos^2(y)\\&=\frac{1}{4}\left(\cos(2x)+1\right)\left(\cos(2y)+1\right)\\&=\frac{1}{4}\left(\frac{1}{2}\cos\left(2\left(x+y\right)\right)+\frac{1}{2}\cos\left(2\left(x-y\right)\right)+\cos(2x)+\cos(2y)+1\right),\end{aligned}$$

we deduce relation (2.6.3). Finally, we show (2.6.4). Applying summation by parts to $\langle n\Delta^n\varepsilon,\Phi_{jk}\rangle_n$ and using $\Phi_{jk}(1)=\Phi_{jk}(0)=0$, yields

$$\langle n\Delta^n\varepsilon,\Phi_{jk}\rangle_n \;=\; \sum_{l=1}^{n}\Delta_l^n\varepsilon\Phi_{jk}\left(\frac{l}{n}\right)=-\sum_{l=1}^{n}\varepsilon_{\frac{l-1}{n}}\left(\Phi_{jk}\left(\frac{l}{n}\right)-\Phi_{jk}\left(\frac{l-1}{n}\right)\right)\,.$$

The equality $\sin(x+h)-\sin(x)=2\sin\left(\frac{h}{2}\right)\cos\left(x+\frac{h}{2}\right)$ for $x,h\in\mathbb{R}$ gives

$$\Phi_{jk}\left(\frac{l}{n}\right)-\Phi_{jk}\left(\frac{l-1}{n}\right)=\frac{1}{n}\varphi_{jk}\left(\frac{l-\frac{1}{2}}{n}\right)$$

which yields the claim. □

Localization

By the localization procedure in Section A.2 it is with Assumption (SM-α-β) sufficient to prove Theorems 2.2.1, 2.2.2 and 2.2.3 under the following assumption.

Assumption 2.6.3 (H-α-β)**.** *Let $0\le\alpha,\beta\le 1$. There exists a constant K such that almost surely*

$$\sup_{0\le t\le T}\left(\|X_t\|+\|b_t\|+\|\sigma_t\|+\|(\sigma_t\sigma_t^\top)^{-1}\|\right)\le K$$

and such that for all $0\le s,t\le T$ with $s+t\le T$

$$\mathbb{E}\left[\sup_{0\le r\le t}\|\sigma_{s+r}-\sigma_s\|^2\right]\le Ct^{2\alpha},\quad \mathbb{E}\left[\sup_{0\le r\le t}\|b_{s+r}-b_s\|^2\right]\le Ct^{2\beta}.$$

We will always assume (H-α-β) for $\alpha\ge 1/2$ and $\beta>0$ in the following proofs.

Local quadratic variations of time

Observe the following asymptotic relation:

$$\sum_{(k-1)h_n \le t_i^{(l)} \le kh_n} \big(t_i^{(l)} - t_{i-1}^{(l)}\big)^2 \asymp \sum_{(k-1)h_n \le t_i^{(l)} \le kh_n} H_l^{(k-1)h_n} \eta_l^{-2} n_l^{-1} \big(t_i^{(l)} - t_{i-1}^{(l)}\big) \tag{2.6.5}$$

$$= H_l^{(k-1)h_n} \eta_l^{-2} n_l^{-1} h_n \tag{2.6.6}$$

with $H_l^{(k-1)h_n}$ defined in (2.3.12). Observe that the latter incorporates the noise variance η_l^2 and the local observation frequency which is linked to the observation time increments above. The left-hand side is a localized measure of variation in observation times similar to the quadratic variation of time by Zhang et al. (2005). It appears in the variance of the estimator and is used to estimate $(F_l^{-1})'((k-1)h_n)$. Under $F_l' \in C^\alpha([0,1])$ with $\alpha > 1/2$ the approximation error by $H_l^{(k-1)h_n}$ is $o_{\mathbb{P}}(n^{-1/4})$. The asymptotic identity applies to deterministic observation times in a deterministic manner and to random exogenous sampling in terms of convergence in probability.

Order of optimal weights

Recall the definition of the optimal weights (2.3.5). An upper bound for these weights is

$$w_{jk} \lesssim I_{jk} = \frac{1}{2}\Big(\sigma^2_{(k-1)h_n} + \frac{\eta^2}{n}\,[\varphi_{jk}, \varphi_{jk}]_n\Big)^{-2} \lesssim \Big(1 + \frac{j^2}{nh_n^2}\Big)^{-2}$$

$$\lesssim \begin{cases} 1 & \text{for } j \le \sqrt{n}h_n \\ j^{-4}n^2h_n^4 & \text{for } j > \sqrt{n}h_n \end{cases} . \tag{2.6.7}$$

This also also implies

$$\sum_{j=1}^{\lfloor nh_n \rfloor -1} w_{jk}\Big(\sigma^2_{(k-1)h_n} + \frac{\eta^2}{n}\,[\varphi_{jk}, \varphi_{jk}]_n\Big)$$

$$\lesssim \sum_{j=1}^{\lfloor \sqrt{n}h_n \rfloor} \Big(1 + \frac{j^2}{h_n^2 n}\Big) + \sum_{j=\lceil \sqrt{n}h_n \rceil}^{\lfloor nh_n \rfloor -1} \Big(1 + \frac{j^2}{nh_n^2}\Big) j^{-4} n^2 h_n^4 \tag{2.6.8}$$

$$\lesssim \sqrt{n}h_n + nh_n^2 . \tag{2.6.9}$$

2.6.2. Proof of Proposition 2.4.1

Recall the definition of the spectral statistics (2.3.1) and denote for $j = 1, \ldots, \lfloor nh_n \rfloor - 1, k = 1, \ldots, h_n^{-1}$:

$$\tilde{S}_{jk} = \Big\langle n(\Delta^n \widetilde{X} + \Delta^n \varepsilon), \Phi_{jk} \Big\rangle_n = \Big\langle n\Delta^n \widetilde{X}, \Phi_{jk} \Big\rangle_n - [\varepsilon, \varphi_{jk}]_n \ ,$$

where $\widetilde{X}$ is the signal process in the locally parametric experiment. It holds that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{S}_{jk}^2\middle|\mathcal{G}_{(k-1)h_n}\right]\\
&=\mathbb{E}\left[\left(\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n-[\varepsilon,\varphi_{jk}]_n\right)^2\middle|\mathcal{G}_{(k-1)h_n}\right] &(2.6.10)\\
&=\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2-2\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n[\varepsilon,\varphi_{jk}]_n+[\varepsilon,\varphi_{jk}]_n^2\middle|\mathcal{G}_{(k-1)h_n}\right]\\
&=\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2\middle|\mathcal{G}_{(k-1)h_n}\right]+\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^2\middle|\mathcal{G}_{(k-1)h_n}\right]\\
&=\sigma^2_{(k-1)h_n}+\frac{\eta^2}{n}[\varphi_{jk},\varphi_{jk}]_n\,. &(2.6.11)
\end{aligned}$$

We have defined ζ_k^n above such that $n^{1/4}\Big(\widetilde{IV}_{n,t}-h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sigma^2_{(k-1)h_n}\Big)$ is equal to

$$n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\Big(\tilde{S}_{jk}^2-\frac{\eta^2}{n}[\varphi_{jk},\varphi_{jk}]_n-\sigma^2_{(k-1)h_n}\Big)=\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\zeta_k^n$$

when we shortly express $\widetilde{IV}_{n,t}=\widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon)$. We have to verify (J1)-(J5). (J1) is trivial as the ζ_k^n are centered conditional on $\mathcal{G}_{(k-1)h_n}$. The proof of (J2) is done in two steps. In paragraph 2.6.2 we calculate explicitly the variance which is the left-hand side of (J2). For this we consider at first general weights $w_{jk}\geq 0$, $\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}=1$ which satisfy $w_{jk}\in\mathcal{G}_{(k-1)h_n}$ for all $k=1,\ldots,h_n^{-1}$, $j=1,\ldots,\lfloor nh_n\rfloor-1$. After that we find optimal weights minimizing the variance. In paragraph 2.6.2 we let $n\to\infty$ and calculate the resulting limiting asymptotic variance. The proofs of (J3), (J4) and (J5) follow in paragraph 2.6.2.

Computation of the variance

$\mathbb{E}[(\zeta_k^n)^2|\mathcal{G}_{(k-1)h_n}]$ is equal to

$$\begin{aligned}
&n^{\frac12}h_n^2\sum_{j,m=1}^{\lfloor nh_n\rfloor-1}w_{jk}w_{mk}\,\mathbb{E}\Big[\left(\tilde{S}_{jk}^2-\mathbb{E}\left[\tilde{S}_{jk}^2\middle|\mathcal{G}_{(k-1)h_n}\right]\right)\\
&\qquad\cdot\left(\tilde{S}_{mk}^2-\mathbb{E}\left[\tilde{S}_{mk}^2\middle|\mathcal{G}_{(k-1)h_n}\right]\right)\Big|\mathcal{G}_{(k-1)h_n}\Big]\\
=\;&n^{\frac12}h_n^2\sum_{j,m=1}^{\lfloor nh_n\rfloor-1}w_{jk}w_{mk}\big(T^n_{j,m,k}(1)+T^n_{j,m,k}(2)+T^n_{j,m,k}(3)\big),
\end{aligned}$$

where the three summands $T^n_{j,m,k}(1)$, $T^n_{j,m,k}(2)$, $T^n_{j,m,k}(3)$ are defined as

$$\begin{aligned}
&\mathbb{E}\left[\left(\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2-\sigma^2_{(k-1)h_n}\right)\left(\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n^2-\sigma^2_{(k-1)h_n}\right)\middle|\mathcal{G}_{(k-1)h_n}\right],\\
&\mathbb{E}\left[4\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n[\varepsilon,\varphi_{jk}]_n\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n[\varepsilon,\varphi_{mk}]_n\middle|\mathcal{G}_{(k-1)h_n}\right],\\
&\mathbb{E}\left[\left([\varepsilon,\varphi_{jk}]_n^2-\frac{\eta^2}{n}[\varphi_{jk},\varphi_{jk}]_n\right)\left([\varepsilon,\varphi_{mk}]_n^2-\frac{\eta^2}{n}[\varphi_{mk},\varphi_{mk}]_n\right)\middle|\mathcal{G}_{(k-1)h_n}\right],
\end{aligned}$$

respectively, for frequencies j,m. The iid structure of the noise and of Brownian increments yields

$$\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n\,[\varepsilon,\varphi_{mk}]_n\right]=\frac{\eta^2}{n}\,[\varphi_{jk},\varphi_{mk}]_n\,,$$
$$\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n\middle|\,\mathcal{G}_{(k-1)h_n}\right]=\delta_{jm}\sigma^2_{(k-1)h_n},$$

which implies with independence of the noise and X that

$$T^n_{j,m,k}\,(2)=4\frac{\eta^2}{n}\delta_{jm}\,[\varphi_{jk},\varphi_{mk}]_n\,\sigma^2_{(k-1)h_n}\,.$$

We further obtain by another polynomial expansion

$$\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^2\,[\varepsilon,\varphi_{mk}]_n^2\right]=n^{-4}\sum_{l,l',p,p'=1}^{n}\Big(\mathbb{E}\left[\varepsilon_l\varepsilon_{l'}\varepsilon_p\varepsilon_{p'}\right]\varphi_{jk}\Big(\frac{l-\frac12}{n}\Big)\varphi_{jk}\Big(\frac{l'-\frac12}{n}\Big)\cdot\varphi_{mk}\Big(\frac{p-\frac12}{n}\Big)\varphi_{mk}\Big(\frac{p'-\frac12}{n}\Big)\Big).$$

Only the cases $l=l'\neq p=p'$, $l=p\neq l'=p'$, $l=p'\neq l'=p$ or $l=l'=p=p'$ produce non-zero results in the expectation. Hence, denoting by $\eta'=\mathbb{E}[\varepsilon_t^4]$ the fourth moment of the observation errors, the last line is equal to

$$\frac{1}{n^4}\sum_{l,l',p,p'}\big(\eta^4\left(\delta_{ll'}\delta_{pp'}+\delta_{lp}\delta_{l'p'}+\delta_{lp'}\delta_{l'p}\right)+\eta'\delta_{lp}\delta_{l'p'}\delta_{ll'}-3\eta^4\delta_{lp}\delta_{l'p'}\delta_{ll'}\big)$$
$$\cdot\Big(\varphi_{jk}\Big(\frac{l-\frac12}{n}\Big)\varphi_{jk}\Big(\frac{l'-\frac12}{n}\Big)\varphi_{mk}\Big(\frac{p-\frac12}{n}\Big)\varphi_{mk}\Big(\frac{p'-\frac12}{n}\Big)\Big)=\frac{\eta^4}{n^2}\big(\,[\varphi_{jk},\varphi_{jk}]_n\cdot[\varphi_{mk},\varphi_{mk}]_n+2\,[\varphi_{jk},\varphi_{mk}]_n^2\,\big)+\frac{\eta'-3\eta^4}{n^4}\sum_{l=1}^{n}\Big(\varphi^2_{jk}\Big(\frac{l-\frac12}{n}\Big)\varphi^2_{mk}\Big(\frac{l-\frac12}{n}\Big)\Big).$$

Arguing similarly and using that $\mathbb{E}[(\Delta_l^nW)^4]=3\,\mathbb{E}[(\Delta_l^nW)^2]$ for $l\in\mathbb{N}$, we obtain

$$\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n^2\middle|\,\mathcal{G}_{(k-1)h_n}\right]$$
$$=\sigma^4_{(k-1)h_n}\left(\langle\Phi_{jk},\Phi_{jk}\rangle_n\langle\Phi_{mk},\Phi_{mk}\rangle_n+2\langle\Phi_{jk},\Phi_{mk}\rangle_n^2\right)=\sigma^4_{(k-1)h_n}\left(1+2\delta_{jm}\right)\,.$$

From the identities so far we obtain

$$\begin{aligned}T^n_{j,m,k}\,(1)&=\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n^2\middle|\,\mathcal{G}_{(k-1)h_n}\right]\\&\quad-\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^2\middle|\,\mathcal{G}_{(k-1)h_n}\right]\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{mk}\right\rangle_n^2\middle|\,\mathcal{G}_{(k-1)h_n}\right]\\&=\sigma^4_{(k-1)h_n}\left(1+2\delta_{mj}\right)-\sigma^4_{(k-1)h_n}=2\delta_{jm}\sigma^4_{(k-1)h_n},\\T^n_{j,m,k}\,(3)&=\mathbb{E}\left[\left([\varepsilon,\varphi_{jk}]_n^2-\frac{\eta^2}{n}\,[\varphi_{jk},\varphi_{jk}]_n\right)\left([\varepsilon,\varphi_{mk}]_n^2-\frac{\eta^2}{n}\,[\varphi_{mk},\varphi_{mk}]_n\right)\right]\\&=\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^2\,[\varepsilon,\varphi_{mk}]_n^2\right]-\frac{\eta^4}{n^2}\,[\varphi_{jk},\varphi_{jk}]_n\,[\varphi_{mk},\varphi_{mk}]_n\\&=\frac{2\eta^4}{n^2}\,[\varphi_{jk},\varphi_{mk}]_n^2+\frac{\eta'-3\eta^4}{n^3}\,\left[\varphi^2_{jk},\varphi^2_{mk}\right]_n\,.\end{aligned}$$

In all, the conditional variance is given by

$$\mathbb{E}\left[(\zeta_k^n)^2\middle|\mathcal{G}_{(k-1)h_n}\right]=\sqrt{n}h_n^2\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}^2\,2\left(\sigma^2_{(k-1)h_n}+\frac{\eta^2}{n}\left[\varphi_{jk},\varphi_{jk}\right]_n\right)^2+R_n$$

with remainder R_n. R_n vanishes for Gaussian noise. In this case, analogous to Bibinger and Reiß (2014), we find that the optimal weights minimizing the variance, under the constraint $\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}=1$, which ensures unbiasedness of the estimator, are given by (2.3.5). The optimization can be done with Lagrange multipliers. R_n is then a remainder in case that $\eta'\neq 3\eta^4$. With the weights (2.3.5) and using (2.6.3) and (2.6.7) we can bound R_n by:

$$R_n\lesssim\frac{\sqrt{n}h_n^2}{n^3}\left(\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}n\left|\sin\left(\frac{j\pi}{nh_n}\right)\right|\right)^2\leq\frac{h_n^2}{n}\,.$$

Therefore, the variance of the estimator $\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}[(\zeta_k^n)^2|\mathcal{G}_{(k-1)h_n}]$ is equal to

$$\sqrt{n}h_n^2\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1}(I_k^{-2}I_{jk}^2)I_{jk}^{-1}+o_{\mathbb{P}}(1)=\sqrt{n}h_n^2\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}I_k^{-1}+o_{\mathbb{P}}(1).$$

The asymptotic variance of the estimator

The key to the asymptotic variance is to recognize

$$(\sqrt{n}h_n)^{-1}I_k=\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor nh_n\rfloor-1}\frac{1}{2}\left(\sigma^2_{(k-1)h_n}+\frac{\eta^2}{n}\left[\varphi_{jk},\varphi_{jk}\right]_n\right)^{-2}$$

as a Riemann-sum, ending up with the "double-Riemann-sum" $\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}h_n((\sqrt{n}h_n)^{-1}I_k)^{-1}$. The scaling factor $(\sqrt{n}h_n)^{-1}$ is the right choice for the first Riemann-sum which becomes clear after two Taylor expansions. First, expanding the sine for each frequency j we find $0\leq\xi_j\leq j\pi/(2nh_n)$ with

$$I_{jk}=\frac{1}{2}\left(\sigma^2_{(k-1)h_n}+4\eta^2n\left(\frac{j\pi}{2nh_n}-\frac{\xi_j^3}{6}\right)^2\right)^{-2}.$$

Second, we expand $x\mapsto\frac{1}{2}\left(\sigma^2_{(k-1)h_n}+4\eta^2nx^2\right)^{-2}$ which yields $\frac{j\pi}{2nh_n}-\frac{\xi_j^3}{6}\leq\xi_j'\leq\frac{j\pi}{2nh_n}$ such that

$$I_{jk}=\tilde{I}_{jk}+R_{jk}\quad\text{with}\quad R_{jk}=\frac{4\eta^2n\xi_j'}{(\sigma^2_{(k-1)h_n}+4\eta^2n\xi_j'^2)^3}\frac{\xi_j^3}{6}\tag{2.6.12}$$

where we define $\tilde{I}_{jk} = \frac{1}{2}(\sigma^2_{(k-1)h_n} + \eta^2(\frac{j\pi}{2\sqrt{n}h_n})^2)^{-2}$. Now it becomes clear that $\sqrt{n}h_n$ is indeed the right factor because

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{nh_n-1}\tilde{I}_{jk} - \int_0^{\sqrt{n}-\frac{1}{\sqrt{n}h_n}}\frac{1}{2}\left(\sigma^2_{(k-1)h_n}+\eta^2\pi^2x^2\right)^{-2}dx\right| \\
&= \Big|\sum_{j=1}^{nh_n-1}\int_{\frac{j-1}{\sqrt{n}h_n}}^{\frac{j}{\sqrt{n}h_n}}\Big(\frac{1}{2}\big(\sigma^2_{(k-1)h_n}+\eta^2\pi^2j^2h_n^{-2}n^{-1}\big)^{-2} \\
&\quad -\frac{1}{2}\big(\sigma^2_{(k-1)h_n}+\eta^2\pi^2x^2\big)^{-2}\Big)dx\Big| \\
&\lesssim \sum_{j=1}^{nh_n-1}\int_{\frac{j-1}{\sqrt{n}h_n}}^{\frac{j}{\sqrt{n}h_n}}\left|x-\frac{j}{\sqrt{n}h_n}\right|dx \max_{\frac{j-1}{\sqrt{n}h_n}\le y\le\frac{j}{\sqrt{n}h_n}}\big(y\big(\sigma^2_{(k-1)h_n}+\eta^2\pi^2y^2\big)^{-3}\big) \\
&\le \left(\frac{1}{\sqrt{n}h_n}\right)^2\left(\sum_{j=1}^{nh_n-1}\Big(\max_{\frac{j-1}{\sqrt{n}h_n}\le y\le\frac{j}{\sqrt{n}h_n}}\big(y\big(\sigma^2_{(k-1)h_n}+\eta^2\pi^2y^2\big)^{-3}\big)\Big)\right) \\
&= \left(\frac{1}{\sqrt{n}h_n}\right)^2\left(\sum_{j=1}^{\lfloor\sqrt{n}h_n\rfloor}\frac{j}{\sqrt{n}h_n}+\sum_{j=\lceil\sqrt{n}h_n\rceil}^{nh_n-1}\left(\frac{\sqrt{n}h_n}{j-1}\right)^5\right) \\
&\lesssim \left(\frac{1}{\sqrt{n}h_n}\right)^2\left(\sqrt{n}h_n+\sum_{j=1}^{nh_n-1-\lceil\sqrt{n}h_n\rceil}\left(\frac{\sqrt{n}h_n}{j+\lceil\sqrt{n}h_n\rceil}\right)^5\right)\lesssim\frac{1}{\sqrt{n}h_n}\,.
\end{aligned}$$

We choose h_n such that $\sqrt{n}h_n \to \infty$. Though we consider all possible spectral frequencies $j = 1, \dots, \lfloor nh_n \rfloor - 1$, we shall see in the following that the I_{jk} for $j \ge \lceil n^\beta h_n \rceil$ become asymptotically negligible for a suitable $0 \le \beta < 1$. By virtue of monotonicity of the sine on $\left[0, \frac{\pi}{2}\right]$ and $\sin(x) \ge\ x/2$ for $0 \le x \le 1$, it follows that

$$\begin{aligned}
\frac{1}{\sqrt{n}h_n}\sum_{j=\lceil n^\beta h_n\rceil}^{\lfloor nh_n\rfloor-1}I_{jk} &\lesssim \frac{1}{\sqrt{n}h_n}\sum_{j=\lceil n^\beta h_n\rceil}^{\lfloor nh_n\rfloor-1}\left(n\sin^2\left(\frac{n^\beta h_n\pi}{2nh_n}\right)\right)^{-2} \\
&\le \frac{1}{\sqrt{n}h_n}nh_n\left(n\sin^2\left(\frac{n^\beta h_n\pi}{2nh_n}\right)\right)^{-2} \\
&\le \sqrt{n}\left(n\left(\frac{n^{\beta-1}\pi}{4}\right)^2\right)^{-2}\lesssim n^{\frac{1}{2}-4\beta+2} = n^{\frac{5}{2}-4\beta}\,.
\end{aligned}$$

We deduce that $\frac{1}{\sqrt{n}h_n}\sum_{j=\lceil n^\beta h_n\rceil}^{\lfloor nh_n\rfloor-1} I_{jk} = o_{\mathbb{P}}(1)$, for every $5/8<\beta<1$. Moreover, we obtain for the first $\lfloor n^\beta h_n\rfloor$ summands of the remainder term

$$\begin{aligned}\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor n^\beta h_n\rfloor} R_{jk} &= \frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor n^\beta h_n\rfloor}\frac{4\eta^2 n\xi_j'}{\left(\sigma^2_{(k-1)h_n}+4\eta^2 n\xi_j'^2\right)^3}\frac{\xi_j^3}{6}\\ &\lesssim \frac{n}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor n^\beta h_n\rfloor}(\xi_j^3\xi_j') \le \frac{n}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor n^\beta h_n\rfloor}\left(\frac{j\pi}{nh_n}\right)^4\\ &\lesssim \frac{1}{\sqrt{n}h_n}n^\beta h_n n^{4(\beta-1)+1} = n^{5\beta-\frac{7}{2}}.\end{aligned}$$

Hence $\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor n^\beta h_n\rfloor} R_{jk} = o_{\mathbb{P}}(1)$ for every $\beta<7/10$. As the tails are asymptotic negligible we thus have $\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor nh_n\rfloor-1} R_{jk} = o_{\mathbb{P}}(1)$ and, in particular,

$$\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor nh_n\rfloor-1} I_{jk} = \int_0^{\sqrt{n}-\frac{1}{\sqrt{n}h_n}}\frac{1}{2}\left(\sigma^2_{(k-1)h_n}+\eta^2\pi^2x^2\right)^{-2}dx + o_{\mathbb{P}}(1).$$

Computing the integral expression yields

$$\int_0^y \frac{1}{2}\left(\sigma^2_{(k-1)h_n}+\eta^2\pi^2x^2\right)^{-2}dx =$$
$$\frac{y}{4\left|\sigma_{(k-1)h_n}\right|^4\left(1+\left(\frac{\eta\pi}{\left|\sigma_{(k-1)h_n}\right|}y\right)^2\right)} + \frac{1}{4\eta\pi\left|\sigma_{(k-1)h_n}\right|^3}\arctan\left(\frac{\eta\pi}{\left|\sigma_{(k-1)h_n}\right|}y\right).$$

As $c<|\sigma_s|<C$ uniformly for all $0\le s\le 1$ with constants c, C and because $\arctan(x)\to\pi/2$ as $x\to\infty$, as well as $\sqrt{n}-\frac{1}{\sqrt{n}h_n}\to\infty$ as $n\to\infty$, we have

$$\begin{aligned}\frac{1}{\sqrt{n}h_n}\sum_{j=1}^{\lfloor nh_n\rfloor-1} I_{jk} = &\frac{\sqrt{n}-\frac{1}{\sqrt{n}h_n}}{4\left|\sigma_{(k-1)h_n}\right|^4\left(1+\left(\frac{\eta\pi}{\left|\sigma_{(k-1)h_n}\right|}\left(\sqrt{n}-\frac{1}{\sqrt{n}h_n}\right)\right)^2\right)}\\ &+\frac{1}{4\eta\pi\left|\sigma_{(k-1)h_n}\right|^3}\arctan\left(\frac{\eta\pi}{\left|\sigma_{(k-1)h_n}\right|}\left(\sqrt{n}-\frac{1}{\sqrt{n}h_n}\right)\right)+o_{\mathbb{P}}(1)\\ =&\frac{1}{8\eta\left|\sigma_{(k-1)h_n}\right|^3}+o_{\mathbb{P}}(1).\end{aligned}$$

The final step in the proof is another Taylor approximation:

$$\begin{aligned}
&\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} \mathbb{E}\left[(\zeta_k^n)^2 \middle| \mathcal{G}_{(k-1)h_n}\right] \\
&= \sqrt{n}h_n^2 \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} I_k^{-1} + o_{\mathbb{P}}(1) = h_n \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} \left(\frac{1}{\sqrt{n}h_n} \sum_{j=1}^{\lfloor nh_n\rfloor -1} I_{jk} \right)^{-1} + o_{\mathbb{P}}(1) \qquad (2.6.13)\\
&= h_n \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} \left(\frac{1}{8\eta \left|\sigma_{(k-1)h_n}\right|^3} + o_{\mathbb{P}}(1)\right)^{-1} + o_{\mathbb{P}}(1) \\
&= \left(h_n \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} 8\eta \left|\sigma_{(k-1)h_n}\right|^3 \right) + o_{\mathbb{P}}(1) . \qquad (2.6.14)
\end{aligned}$$

The last equality is true by Taylor and because σ is uniformly bounded. Because σ is continuous we obtain the claim by Riemann approximation, i.e.

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} \mathbb{E}\left[(\zeta_k^n)^2 \middle| \mathcal{G}_{(k-1)h_n}\right] \to 8\eta \int_0^t |\sigma_s|^3 \, ds$$

in probability as $n \to \infty$ establishing (J2) with the asymptotic expression of Theorem 2.2.1.

Lyapunov's criterion and stability of convergence

So far, we have proved (J1) and (J2). Next, we shall prove that the Lyapunov condition (J3) is satisfied. For the sum of fourth moments, we obtain by Jensen's inequality and $w_{jk} \in \mathcal{G}_{(k-1)h_n}$ for all $k = 1, \dots, h_n^{-1}$ and $j = 1, \dots, \lfloor nh_n \rfloor - 1$:

$$\begin{aligned}
&\mathbb{E}\left[(\zeta_k^n)^4 \middle| \mathcal{G}_{(k-1)h_n}\right] \\
&= (nh_n)^4 \left(\mathbb{E}\left[\left(\sum_{j=1}^{\lfloor nh_n\rfloor -1} w_{jk} \left(\tilde{S}_{jk}^2 - \mathbb{E}\left[\tilde{S}_{jk}^2 \middle| \mathcal{G}_{(k-1)h_n}\right]\right)\right)^4 \middle| \mathcal{G}_{(k-1)h_n}\right]\right) \\
&\le nh_n^4 \left(\sum_{j=1}^{\lfloor nh_n\rfloor -1} w_{jk} \left(\mathbb{E}\left[\left(\tilde{S}_{jk}^2 - \mathbb{E}\left[\tilde{S}_{jk}^2 \middle| \mathcal{G}_{(k-1)h_n}\right]\right)^4 \middle| \mathcal{G}_{(k-1)h_n}\right]\right)^{\frac{1}{4}} \right)^4 \\
&\lesssim nh_n^4 \left(\sum_{j=1}^{\lfloor nh_n\rfloor -1} w_{jk} \left(\mathbb{E}\left[\tilde{S}_{jk}^8 \middle| \mathcal{G}_{(k-1)h_n}\right]\right)^{\frac{1}{4}} \right)^4 .
\end{aligned}$$

If we can show

$$\mathbb{E}\left[\left\langle n\Delta^n \widetilde{X}, \Phi_{jk} \right\rangle_n^8 \middle| \mathcal{G}_{(k-1)h_n}\right] \lesssim \sigma_{(k-1)h_n}^8, \qquad (2.6.15)$$

$$\mathbb{E}\left[[\varepsilon, \varphi_{jk}]_n^8 \right] \lesssim \left(\eta^2\right)^4 \frac{[\varphi_{jk}, \varphi_{jk}]_n^4}{n^4}, \qquad (2.6.16)$$

then we are able to conclude that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{S}_{jk}^{8}\middle|\mathcal{G}_{(k-1)h_n}\right]\\
&\lesssim \mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n^8\middle|\mathcal{G}_{(k-1)h_n}\right]+\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^8\right]\\
&\lesssim \sigma_{(k-1)h_n}^8+\left(\eta^2\right)^4\frac{[\varphi_{jk},\varphi_{jk}]_n^4}{n^4}\lesssim\left(\sigma_{(k-1)h_n}^2+\eta^2\frac{[\varphi_{jk},\varphi_{jk}]_n}{n}\right)^4. \qquad (2.6.17)
\end{aligned}$$

Hence, we obtain from (2.6.9)

$$\begin{aligned}
&\sum_{k=1}^{h_n^{-1}}\mathbb{E}\left[(\zeta_k^n)^4\middle|\mathcal{G}_{(k-1)h_n}\right]\lesssim\sum_{k=1}^{h_n^{-1}}nh_n^4\left(\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left(\sigma_{(k-1)h_n}^2+\frac{\eta^2}{n}[\varphi_{jk},\varphi_{jk}]_n\right)\right)^4\\
&\lesssim n^2h_n^6=o(1)
\end{aligned}$$

which proves (J3). We are therefore left with proving (2.6.15) and (2.6.16). The first inequality holds because $\langle n\Delta^n\widetilde{X},\Phi_{jk}\rangle_n$ is $N(0,\sigma^2_{(k-1)h_n})$-distributed conditional on $\mathcal{G}_{(k-1)h_n}$. In order to see why the second inequality is satisfied, let $g_l=\varepsilon_{((k-1)\lceil nh_n\rceil+l)/n}$ $\varphi_{jk}\big(\frac{(k-1)\lceil nh_n\rceil+l-\frac{1}{2}}{n}\big)$ for $l=1,\ldots,\lfloor nh_n\rfloor$. The g_l are independent and centered such that for any $1\le l_1,\ldots,l_8\le\lfloor nh_n\rfloor$ with $\mathbb{E}[g_{l_1}\cdots g_{l_8}]\neq0$ each g_l appears at least twice and there are at most four distinct g_l. If there are exactly four distinct g_l, e.g. $l_1=l_2,l_3=l_4,l_5=l_6,l_7=l_8$, we arrive at the bound

$$\sum_{l_1,l_3,l_5,l_7}\mathbb{E}\left[g_{l_1}\cdots g_{l_8}\right]\le\left(\eta^2\right)^4n^4\left[\varphi_{jk},\varphi_{jk}\right]_n^4.$$

The leading term does not include eighth moments of the noise, but the fourth power of the second moment which we denote $(\eta^2)^4$ to prevent any confusion. If there are less than four distinct g_l, we obtain from (2.6.2) and (2.6.3) with the assumption $\mathbb{E}[\varepsilon_t^8]<\infty$, that the respective sums are asymptotically of smaller order. The terms with eighth moments are thus negligible. This implies (2.6.16):

$$\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^8\right]=n^{-8}\sum_{1\le l_1,\ldots,l_8\le nh_n}\mathbb{E}\left[g_{l_1}\cdots g_{l_8}\right]\lesssim\left(\eta^2\right)^4\frac{[\varphi_{jk},\varphi_{jk}]_n^4}{n^4}.$$

It remains to verify (J4) and (J5). The proof follows a similar strategy as the proofs of Proposition 5.10, step 4, of Jacod et al. (2010) and Lemma 5.7 of Jacod et al. (2009). It is sufficient to show with $\delta_k^n(M)=M_{kh_n}-M_{(k-1)h_n}$ that

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[\zeta_k^n\delta_k^n(M)\,\middle|\mathcal{G}_{(k-1)h_n}\right]\xrightarrow{\mathbb{P}}0 \qquad (2.6.18)$$

for any $M\in\mathcal{N}$, the set of square-integrable $(\mathcal{G}_t)_{0\le t\le1}$-martingales. Note that (2.6.18) is closed under L^2-convergence with respect to the terminal variables $M_1\in L^2(\mathcal{G})$ for $M\in\mathcal{N}$ what follows by Cauchy-Schwarz inequality. Define subsets $\mathcal{N}^0,\mathcal{N}^1$, $\mathcal{N}^2$ of $\mathcal{N}$, where $\mathcal{N}^0$ is the space of all square-integrable martingales adapted to $\mathcal{W} = \sigma(W_s : s \le 1)$, i.e. every such martingale has the form $C + \int_0^t h_s \, dW_s$ for some constant C and a predictable square-integrable process $h \in \mathcal{W}$. $\mathcal{N}^1$ is the set of all square-integrable $(\mathcal{F}_t)$-martingales which are orthogonal to W, and $\mathcal{N}^2$ is the space of all square-integrable martingales adapted to the filtration $\mathcal{E}_t = \sigma(\varepsilon_s : s \le t)$, generated by the noise process. Then the set of square-integrable martingales of the form $M \cdot N$, for $M \in \mathcal{N}^0 \cup \mathcal{N}^1$, $N \in \mathcal{N}^2$, is total in $\mathcal{N}$ (by independence any process of the form $M \cdot N$ is again a martingale) and it is enough to show (2.6.18) for such processes. Using the decomposition

$$\delta_k^n(MN) = \delta_k^n(M)\,\delta_k^n(N) + N_{(k-1)h_n}\delta_k^n(M) + M_{(k-1)h_n}\delta_k^n(N) \tag{2.6.19}$$

we have by independence of W and noise for any $k = 1, \dots, h_n^{-1}$:

$$\begin{aligned}
&\mathbb{E}\left[\zeta_k^n \delta_k^n(MN) \middle| \mathcal{G}_{(k-1)h_n}\right] \\
&= n^{1/4} h_n \sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk} \Bigg(\mathbb{E}\left[\tilde{S}_{jk}^2 \delta_k^n(MN) \middle| \mathcal{G}_{(k-1)h_n}\right] \\
&\qquad - \mathbb{E}\left[\tilde{S}_{jk}^2 \middle| \mathcal{G}_{(k-1)h_n}\right] \mathbb{E}\left[\delta_k^n(MN) \middle| \mathcal{G}_{(k-1)h_n}\right] \Bigg) \\
&= n^{1/4} h_n \sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk} \Bigg(N_{(k-1)h_n} \mathbb{E}\left[\left\langle n\Delta^n \widetilde{X}, \Phi_{jk} \right\rangle_n^2 \delta_k^n(M) \middle| \mathcal{F}_{(k-1)h_n}\right] \\
&\qquad - 2\mathbb{E}\left[\left\langle n\Delta^n \widetilde{X}, \Phi_{jk} \right\rangle_n \delta_k^n(M) \middle| \mathcal{F}_{(k-1)h_n}\right] \mathbb{E}\left[[\varepsilon, \varphi_{jk}]_n \delta_k^n(N) \middle| \mathcal{E}_{(k-1)h_n}\right] \\
&\qquad + M_{(k-1)h_n} \mathbb{E}\left[[\varepsilon, \varphi_{jk}]_n^2 \delta_k^n(N) \middle| \mathcal{E}_{(k-1)h_n}\right] \Bigg).
\end{aligned}$$

Let first $M \in \mathcal{N}^0$. As $\mathcal{N}^0$ is closed and because the case M constant is trivial, we can assume that $M = \int_0^{\cdot} \gamma_s \, dW_s$ for γ bounded, adapted to $\mathcal{W}$ and piecewise constant on intervals $(T_q, T_{q+1}]$ for some $0 = T_0 < T_1 < \dots, T_m = 1$, $m \ge 1$, such that

$$\begin{aligned}
&\mathbb{E}\left[\left\langle n\Delta^n \widetilde{X}, \Phi_{jk} \right\rangle_n^2 \delta_k^n(M) \middle| \mathcal{F}_{(k-1)h_n}\right] \\
&= \sum_{l,p=1}^{n} \sigma_{(k-1)h_n}^2 \sum_{q=1}^{m} \mathbb{E}\left[\Delta_l^n W \Delta_p^n W \gamma_{t_q} \left(W_{T_{q+1} \wedge kh_n} - W_{T_q \vee (k-1)h_n}\right) \middle| \mathcal{F}_{(k-1)h_n}\right] \\
&\qquad \cdot \Phi_{jk}\Big(\frac{l}{n}\Big) \Phi_{jk}\Big(\frac{p}{n}\Big).
\end{aligned} \tag{2.6.20}$$

For n large enough there is at most one T_q per block. If there is no T_q on the k-th block, the conditional expectation above vanishes by independence of the Brownian increments for any l, p. On the other hand, there are only m blocks containing some T_q and for every such block the left-hand side of (2.6.20) is bounded, what can be seen e.g. by (2.6.15) and because M is square-integrable. Hence,

$$n^{1/4} h_n \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk} N_{(k-1)h_n} \mathbb{E}\left[\left\langle n\Delta^n \widetilde{X}, \Phi_{jk} \right\rangle_n^2 \delta_k^n(M) \middle| \mathcal{F}_{(k-1)h_n}\right] = o_{\mathbb{P}}(1). \tag{2.6.21}$$

Next, let $M \in \mathcal{N}^1$, i.e. M is orthogonal to W. The left-hand side in (2.6.20) is now equal to

$$\sum_{l,p,q=1}^{n} \sigma^2_{(k-1)h_n} \mathbb{E}\left[\Delta_l^n W \Delta_p^n W \Delta_q^n M \middle| \mathcal{F}_{(k-1)h_n}\right] \Phi_{jk}\Big(\frac{l}{n}\Big)\Phi_{jk}\Big(\frac{p}{n}\Big).$$

The conditional expectation vanishes, except for $l=p=q$, $p<l=q$ or $l<p=q$. For $l=p=q$ we obtain by Itô's formula

$$\mathbb{E}\left[(\Delta_l^n W)^2\,\Delta_l^n M\middle|\,\mathcal{F}_{(k-1)h_n}\right] = \mathbb{E}\left[\left((\Delta_l^n W)^2 - \left(\frac{1}{n}\right)\right)\Delta_l^n M\middle|\,\mathcal{F}_{(k-1)h_n}\right]$$
$$+\,\mathbb{E}\left[\left(\frac{1}{n}\right)\Delta_l^n M\middle|\,\mathcal{F}_{(k-1)h_n}\right] = \mathbb{E}\left[\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}} W_s\,dW_s\right)\Delta_l^n M\middle|\,\mathcal{F}_{(k-1)h_n}\right].$$

However, $((\int_0^t W_s\,dW_s)\cdot M_t)_{0\le t\le 1}$ is an $(\mathcal{F}_t)$-martingale by orthogonality such that the last expression vanishes. The cases $p<l=q$ and $l<p=q$ follow similarly. Hence, (2.6.21) is still satisfied; the left-hand side is actually zero.

With respect to N, as $\mathcal{N}^2$ is closed, we can assume without loss of generality that $N_1 = f(\varepsilon_{T_1},\ldots,\varepsilon_{T_{m'}})$ for some measurable function f and some $0\le T_1<\cdots<T_{m'}\le 1$, $m'\ge 1$. Similar as before, for n large there is at most one $T_{q'}$ per block. On any block not containing such a $T_{q'}$ it holds that $\delta_k^n(N)=0$. Bounding the terms for the m' other blocks yields for $M\in\mathcal{N}^1\cup\mathcal{N}^2$

$$n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1} w_{jk}M_{(k-1)h_n}\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n^2\,\delta_k^n(N)\middle|\,\mathcal{E}_{(k-1)h_n}\right] = o_{\mathbb{P}}(1)\,. \tag{2.6.22}$$

From the previous discussion we further see that for all but at most $m+m'$ blocks:

$$\sum_{j=1}^{nh_n-1} w_{jk}\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n \delta_k^n(M)\middle|\,\mathcal{F}_{(k-1)h_n}\right]\mathbb{E}\left[[\varepsilon,\varphi_{jk}]_n\,\delta_k^n(N)\middle|\,\mathcal{E}_{(k-1)h_n}\right]=0,$$

whereas bounds on the remaining $m+m'$ blocks guarantee that the cross terms tend to zero in probability. We conclude that (2.6.18) holds. This completes the proof of Proposition 2.4.1.

2.6.3. Proof of Proposition 2.4.2

We first give a general outline of the proof, deferring some technical details to the end of this section. By Taylor we have for all $k=1,\ldots,h_n^{-1}$ and $j=1,\ldots,\lfloor nh_n\rfloor-1$, the existence of random variables ξ_{jk} such that $S_{jk}^2-\tilde{S}_{jk}^2 = 2\tilde{S}_{jk}(S_{jk}-\tilde{S}_{jk})+2(\xi_{jk}-$

$\tilde{S}_{jk})(S_{jk}-\tilde{S}_{jk})$ and $|\xi_{jk}-\tilde{S}_{jk}|\le|S_{jk}-\tilde{S}_{jk}|$. This yields

$$\begin{aligned}
&n^{1/4}\left(\widehat{IV}^{or}_{n,t}(Y)-\widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon)\right)\\
&=n^{1/4}\Big(h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left(S_{jk}^2-\tilde{S}_{jk}^2\right)\Big)\\
&=\Big(n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left(2\tilde{S}_{jk}\left(S_{jk}-\tilde{S}_{jk}\right)\right)\Big)\\
&+\Big(n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left(\xi_{jk}-\tilde{S}_{jk}\right)\left(S_{jk}-\tilde{S}_{jk}\right)\Big).
\end{aligned}$$

For the second sum above, which we denote by Z_t^n, we obtain by the Markov inequality and Step 1 below for any $\varepsilon>0$

$$\begin{aligned}
\mathbb{P}\left(\sup_{0\le t\le 1}|Z_t^n|>\varepsilon\right) &\le \mathbb{P}\Bigg(\Big(n^{1/4}h_n\sum_{k=1}^{h_n^{-1}}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left|S_{jk}-\tilde{S}_{jk}\right|^2\Big)>\varepsilon\Bigg)\\
&\le \varepsilon^{-1}n^{1/4}h_n\sum_{k=1}^{h_n^{-1}}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\mathbb{E}\left[\left(S_{jk}-\tilde{S}_{jk}\right)^2\right]\\
&\lesssim \varepsilon^{-1}n^{1/4}h_n\to 0.
\end{aligned}$$

Let $T_k^n=\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left(2\tilde{S}_{jk}\left(S_{jk}-\tilde{S}_{jk}\right)\right)$ and write the first sum above as $M_t^n+R_t^n$ with

$$\begin{aligned}
M_t^n &= n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\left(T_k^n-\mathbb{E}\left[T_k^n|\,\mathcal{G}_{(k-1)h_n}\right]\right),\\
R_t^n &= n^{1/4}h_n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[T_k^n|\,\mathcal{G}_{(k-1)h_n}\right].
\end{aligned}$$

In Step 2 we show that

$$\sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\left[\left(n^{\frac{1}{4}}h_nT_k^n\right)^2\right]\to 0,\quad n\to\infty.$$

A well known result thereby yields $M_t^n\xrightarrow{ucp}0$. Finally, observe that

$$\begin{aligned}
&\mathbb{E}\left[\left(2\tilde{S}_{jk}\left(S_{jk}-\tilde{S}_{jk}\right)\right)\Big|\,\mathcal{G}_{(k-1)h_n}\right]\\
&=\mathbb{E}\left[2\left(\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n-[\varepsilon,\varphi_{jk}]_n\right)\left\langle n\Delta^n\left(X-\widetilde{X}\right),\Phi_{jk}\right\rangle_n\Big|\,\mathcal{G}_{(k-1)h_n}\right]\\
&=\mathbb{E}\left[2\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\left\langle n\Delta^n\left(X-\widetilde{X}\right),\Phi_{jk}\right\rangle_n\Big|\,\mathcal{G}_{(k-1)h_n}\right],
\end{aligned}$$

i.e. the noise terms vanish, thereby simplifying the following calculations. Write $\mathbb{E}[(2\tilde{S}_{jk}(S_{jk}-\tilde{S}_{jk}))|\mathcal{G}_{(k-1)h_n}]$ as the sum $D^n_{jk}+V^n_{jk}$, where D^n_{jk} and V^n_{jk} are defined as

$$\mathbb{E}\left[2\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\left(\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}b_s\,ds\right)\Phi_{jk}\left(\frac{l}{n}\right)\right)\Bigg|\mathcal{G}_{(k-1)h_n}\right],$$
$$\mathbb{E}\left[2\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\left(\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s\right)\Phi_{jk}\left(\frac{l}{n}\right)\right)\Bigg|\mathcal{G}_{(k-1)h_n}\right].$$

In Step 3 we show that $\left|D^n_{jk}+V^n_{jk}\right|\lesssim h_n^{\widetilde{\beta}}$ for some $\widetilde{\beta}>1/2$. This yields immediately

$$\sup_{0\le t\le 1}|R^n_t| \;\le\; n^{1/4}h_n\sum_{k=1}^{h_n^{-1}}\sum_{j=1}^{\lfloor nh_n\rfloor-1}w_{jk}\left|D^n_{jk}+V^n_{jk}\right|\lesssim n^{1/4}h_n^{\widetilde{\beta}}=o\left(1\right),$$

implying *ucp*-convergence. We therefore conclude that

$$n^{1/4}\left(\widehat{IV}^{or}_{n,t}(Y)-\widehat{IV}^{or}_{n,t}(\widetilde{X}+\varepsilon)\right)\xrightarrow{ucp}0,\;\; n\to\infty.$$

The second claim

$$n^{1/4}\int_0^t\left(\sigma_s^2-\sigma^2_{\lfloor sh_n^{-1}\rfloor h_n}\right)ds\xrightarrow{ucp}0,\;\; n\to\infty,$$

follows from (H-α-β), because $\alpha\ge 1/2$. This proves Proposition 2.4.2. We end this section with detailed proofs of Steps 1 – 3.

Step 1. We show that $\mathbb{E}[(S_{jk}-\tilde{S}_{jk})^4]\lesssim h_n^2$. Using the decomposition

$$S_{jk}-\tilde{S}_{jk}=\left\langle n\Delta^n(X-\widetilde{X}),\Phi_{jk}\right\rangle_n \tag{2.6.23}$$
$$=\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}b_s\,ds\right)\Phi_{jk}\left(\frac{l}{n}\right)+\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s\right)\Phi_{jk}\left(\frac{l}{n}\right)$$

into drift and volatility terms we obtain

$$\mathbb{E}\left[\left(S_{jk}-\tilde{S}_{jk}\right)^4\right]\lesssim\;\mathbb{E}\left[\left(\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}b_s\,ds\right)\Phi_{jk}\left(\frac{l}{n}\right)\right)^4\right]$$
$$+\,\mathbb{E}\left[\left(\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s\right)\Phi_{jk}\left(\frac{l}{n}\right)\right)^4\right].$$

The first summand is bounded by h_n^2. For the second let $\kappa_l=\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s$, such that

$$\mathbb{E}\left[\left(\sum_{l=1}^n\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s\right)\Phi_{jk}\left(\frac{l}{n}\right)\right)^4\right]$$
$$=\sum_{l,l',p,p'}\mathbb{E}\left[\kappa_l\kappa_{l'}\kappa_p\kappa_{p'}\right]\Phi_{jk}\left(\frac{l}{n}\right)\Phi_{jk}\left(\frac{l'}{n}\right)\Phi_{jk}\left(\frac{p}{n}\right)\Phi_{jk}\left(\frac{p'}{n}\right).$$

Properties of the conditional expectation show that the only choices for l, l', p, p' with non-vanishing results are $l, l' < p = p'$, $l < l' = p = p'$ and $l = l' = p = p'$. In all three cases we can conclude by Proposition A.3.2 that

$$\left| \mathbb{E}\left[\kappa_l \kappa_{l'} \kappa_p \kappa_{p'}\right] \Phi_{jk}\left(\frac{l}{n}\right) \Phi_{jk}\left(\frac{l'}{n}\right) \Phi_{jk}\left(\frac{p}{n}\right) \Phi_{jk}\left(\frac{p'}{n}\right) \right| \lesssim n^{-4} h_n^{-2}.$$

Observe that in any of the three mentioned cases we find at least two identical integers l, l', p or p'. In all, there are $\lfloor nh_n \rfloor \cdot \binom{\lfloor nh_n \rfloor - 1}{2} \cdot 4!$ possibilities to choose such indices. Hence, we obtain

$$\mathbb{E}\left[\left(\sum_{l=1}^{n}\left(\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_s - \sigma_{(k-1)h_n}\right) dW_s\right) \Phi_{jk}\left(\frac{l}{n}\right)\right)^4\right] \lesssim (nh_n)^3 n^{-4} h_n^{-2} = n^{-1} h_n,$$

which is up to a constant bounded by h_n^2, and therefore the claim holds.

Step 2. We show that $\sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \mathbb{E}[(n^{\frac{1}{4}} h_n T_k^n)^2] \to 0$ as $n \to \infty$. The Minkowski and Cauchy-Schwarz inequality yield

$$\begin{aligned}
&\left\| n^{\frac{1}{4}} h_n \sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk}\left(2\tilde{S}_{jk}\left(S_{jk} - \tilde{S}_{jk}\right)\right) \right\|_{L^2(\mathbb{P})}^2 \\
&\leq n^{\frac{1}{2}} h_n^2 \left(\sum_{j=1}^{\lfloor nh_n \rfloor - 1} \left\| w_{jk}\left(2\tilde{S}_{jk}\left(S_{jk} - \tilde{S}_{jk}\right)\right) \right\|_{L^2(\mathbb{P})} \right)^2 \\
&\leq n^{\frac{1}{2}} h_n^2 \left(\sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk} \left(\mathbb{E}\left[\tilde{S}_{jk}^4\right]\right)^{\frac{1}{4}} \left(\mathbb{E}\left[\left(S_{jk} - \tilde{S}_{jk}\right)^4\right]\right)^{\frac{1}{4}} \right)^2 .
\end{aligned}$$

By Step 1 we already know that $\mathbb{E}\big[\big(S_{jk} - \tilde{S}_{jk}\big)^4\big] \lesssim h_n^2$. Because σ is bounded, we obtain by (2.6.17) the bound

$$\begin{aligned}
\mathbb{E}\left[\tilde{S}_{jk}^4\right] &\leq \mathbb{E}^{\frac{1}{2}}\left[\mathbb{E}\left[\tilde{S}_{jk}^8 \middle| \mathcal{G}_{(k-1)h_n}\right]\right] \lesssim \mathbb{E}^{\frac{1}{2}}\left[\left(\sigma_{(k-1)h_n}^2 + \frac{\eta^2}{n}\left[\varphi_{jk}, \varphi_{jk}\right]_n\right)^4\right] \\
&\lesssim \left(1 + \frac{\eta^2}{n}\left[\varphi_{jk}, \varphi_{jk}\right]_n\right)^2 \leq \left(1 + \frac{\eta^2}{n}\left[\varphi_{jk}, \varphi_{jk}\right]_n\right)^4 .
\end{aligned}$$

Together with (2.6.9) it follows that

$$\begin{aligned}
\sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \mathbb{E}\left[\left(n^{\frac{1}{4}} h_n T_k^n\right)^2\right] &\lesssim n^{\frac{1}{2}} h_n^3 \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \left(\sum_{j=1}^{\lfloor nh_n \rfloor - 1} w_{jk}\left(1 + \frac{\eta^2}{n}\left[\varphi_{jk}, \varphi_{jk}\right]_n\right)\right)^2 \\
&\lesssim n^{\frac{1}{2}} h_n^2 \cdot n^2 h_n^4 = o(1) .
\end{aligned}$$

Step 3. We show that $\left| D_{jk}^n + V_{jk}^n \right| \lesssim h_n^\beta$ for some $\beta > 1/2$. Expanding the sums in

V_{jk}^n and Itô isometry yield for V_{jk}^n

$$\sum_{l,m=1}^{n}\left(\mathbb{E}\left[\Delta_l^n\widetilde{X}\left(\int_{\frac{m-1}{n}}^{\frac{m}{n}}\left(\sigma_s-\sigma_{(k-1)h_n}\right)dW_s\right)\middle|\mathcal{G}_{(k-1)h_n}\right]\Phi_{jk}\left(\frac{l}{n}\right)\Phi_{jk}\left(\frac{m}{n}\right)\right)$$
$$=\sum_{l=1}^{n}\mathbb{E}\left[\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(\sigma_{(k-1)h_n}\left(\sigma_s-\sigma_{(k-1)h_n}\right)\right)ds\right]\Phi_{jk}^2\left(\frac{l}{n}\right).$$

From (H-α-β) it follows for $s\in[(k-1)h_n,kh_n]$ that

$$\left|\mathbb{E}\left[\left(\sigma_{(k-1)h_n}\left(\sigma_s-\sigma_{(k-1)h_n}\right)\right)\right]\right|=\left|\mathbb{E}\left[\sigma_{(k-1)h_n}\mathbb{E}\left[\sigma_s-\sigma_{(k-1)h_n}\middle|\mathcal{G}_{(k-1)h_n}\right]\right]\right|,$$

which is up to a constant bounded by h_n^{α}, and hence by Fubini $|V_{jk}|\lesssim h_n^{\alpha}$, as well. With respect to D_{jk}^n, we need an additional approximation. By the boundedness of $\mathbb{E}[|\langle n\Delta^n\widetilde{X},\Phi_{jk}\rangle_n|]$ from (2.6.15):

$$\left|\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\int_{\frac{l-1}{n}}^{\frac{l}{n}}b_s\,ds\middle|\mathcal{G}_{(k-1)h_n}\right]\right|$$
$$\lesssim\left|\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\int_{\frac{l-1}{n}}^{\frac{l}{n}}\left(b_s-b_{(k-1)h_n}\right)ds\middle|\mathcal{G}_{(k-1)h_n}\right]\right|$$
$$+\left|\frac{b_{(k-1)h_n}}{n}\mathbb{E}\left[\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\middle|\mathcal{G}_{(k-1)h_n}\right]\right|\lesssim h_n^{\nu\wedge\frac{1}{2}}n^{-1}.$$

Using this bound we find that

$$\left|D_{jk}^n\right|\;\le\;\sum_{l=1}^{n}\left|\mathbb{E}\left[2\left\langle n\Delta^n\widetilde{X},\Phi_{jk}\right\rangle_n\int_{\frac{l-1}{n}}^{\frac{l}{n}}b_s\,ds\middle|\mathcal{G}_{(k-1)h_n}\right]\right|\left|\Phi_{jk}\left(\frac{l}{n}\right)\right|\lesssim h_n^{\nu\wedge\frac{1}{2}+\frac{1}{2}}.$$

We obtain the claim with $\widetilde{\beta}=\min\left\{\nu\wedge\frac{1}{2}+\frac{1}{2},\alpha\right\}$. This is the only time we need the smoothness of the drift in Assumption (H-α-β) with $\beta>0$. This is necessary because of the log-term in the definition of h_n.

2.6.4. Proofs of Theorem 2.2.2 and Theorem 2.2.3 for oracle estimation

We decompose X similarly as in the proof of Theorem 2.2.1:

$$X_t=X_0+\bar{B}_t+\tilde{B}_t+\bar{C}_t+\tilde{C}_t\,,\qquad(2.6.24)$$

where we denote

$$\bar{B}_t=\int_0^t b_{\lfloor sh_n^{-1}\rfloor h_n}\,ds\;,\;\tilde{B}_t=\int_0^t\left(b_s-b_{\lfloor sh_n^{-1}\rfloor h_n}\right)ds\,,\qquad(2.6.25)$$

$$\bar{C}_t=\int_0^t\sigma_{\lfloor sh_n^{-1}\rfloor h_n}\,dW_s\;,\;\tilde{C}_t=\int_0^t\left(\sigma_s-\sigma_{\lfloor sh_n^{-1}\rfloor h_n}\right)dW_s\,.\qquad(2.6.26)$$

In order to establish a functional CLT, we decompose the estimation errors of (2.3.19) (and likewise (2.3.16)) in the following way:

$$LMM_{n,t}^{or}(Y) - \text{vec}\Big(\int_0^t \Sigma_s\,ds\Big) = LMM_{n,t}^{or}(\bar{C}+\varepsilon) - \text{vec}\left(\int_0^t \Sigma_{\lfloor sh_n^{-1}\rfloor h_n}\,ds\right) \quad (2.6.27)$$

$$+ LMM_{n,t}^{or}(Y) - LMM_{n,t}^{or}(\bar{C}+\varepsilon) - \text{vec}\left(\int_0^t \big(\Sigma_s - \Sigma_{\lfloor sh_n^{-1}\rfloor h_n}\big)\,ds\right). \quad (2.6.28)$$

One crucial step to cope with multi-dimensional non-synchronous data is Lemma 2.4.4 which is proved next. Below, we give a concise proof of the functional CLTs for the estimators (2.3.19) and (2.3.16), where after restricting to a synchronous reference scheme many steps follow as direct extensions of the one-dimensional case. The stable CLTs for the *leading terms*, namely the right-hand side of (2.6.27) and the analogue for estimator (2.3.16), are established in paragraph 2.6.4. The *remainder terms* (2.6.28) and their analogues are handled in paragraph 2.6.4.

Proof of Lemma 2.4.4

Consider for $l,m = 1,\ldots,d$, observation times $t_i^{(l)} = F_l^{-1}(i/n_l)$ and $t_i^{(m)} = F_m^{-1}(i/n_m)$. Define a next-tick interpolation function by

$$t_+^{(l)}(s) = \min\left(t_v^{(l)}, v = 0,\ldots,n_l | t_v^{(l)} \ge s\right), l = 1,\ldots,d,$$

and analogously a previous-tick interpolation function by

$$t_-^{(l)}(s) = \max\left(t_v^{(l)}, v = 0,\ldots,n_l | t_v^{(l)} \le s\right), l = 1,\ldots,d.$$

We decompose increments of $X^{(l)}$ between adjacent observation times $t_{v-1}^{(l)}, t_v^{(l)}, v = 1,\ldots,n_l$, in the sum of increments of $X^{(l)}$ over all time intervals $[t_{i-1}^{(m)}, t_i^{(m)}]$ contained in $[t_{v-1}^{(l)}, t_v^{(l)}]$ and the remaining time intervals at the left $\big[t_{v-1}^{(l)}, t_+^{(m)}\big(t_{v-1}^{(l)}\big)\big]$ and the right border $\big[t_-^{(m)}\big(t_v^{(l)}\big), t_v^{(l)}\big]$:

$$X_{t_v^{(l)}}^{(l)} - X_{t_{v-1}^{(l)}}^{(l)} = \Big(X_{t_v^{(l)}}^{(l)} - X_{t_-^{(m)}(t_v^{(l)})}^{(l)}\Big) + \sum_{\Delta_i t^{(m)} \subset \Delta_v t^{(l)}} \Big(X_{t_i^{(m)}}^{(l)} - X_{t_{i-1}^{(m)}}^{(l)}\Big)$$
$$+ \Big(X_{t_+^{(m)}(t_{v-1}^{(l)})}^{(l)} - X_{t_{v-1}^{(l)}}^{(l)}\Big).$$

If there is only one observation of $X^{(m)}$ in $[t_{v-1}^{(l)}, t_v^{(l)}]$, set $\sum_{\Delta_i t^{(m)} \subset \Delta_v t^{(l)}} \big(X_{t_i^{(m)}}^{(l)} - X_{t_{i-1}^{(m)}}^{(l)}\big) = 0$. If there is no observation of $X^{(m)}$ in $[t_{v-1}^{(l)}, t_v^{(l)}]$ we take the union of a set of intervals $\bigcup_{v\in V}[t_{v-1}^{(l)}, t_v^{(l)}]$ which contains at least one observation time of $X^{(m)}$. We use an expansion of $\Phi_{jk}(t) - \Phi_{jk}(s)$. By virtue of $\sin(t) - \sin(s) = 2\cos((t+s)/2)\sin((t-s)/2)$ and the sine expansion, we obtain for $s,t \in [kh_n,(k+1)h_n)$:

$$\Phi_{jk}(t) - \Phi_{jk}(s) \asymp \sqrt{2}h_n^{-3/2} j\pi \cos\left(j\pi h_n^{-1}\left(\tfrac{t+s}{2} - kh_n\right)\right)(t-s)\,. \quad (2.6.29)$$

In particular, for $t-s=O(n^{-1})$ we have that $\Phi_{jk}(t)-\Phi_{jk}(s)=O\left(\varphi_{jk}(\frac{t+s}{2})n^{-1}\right)$. With $u_v^{(m)}=(1/2)(t_+^{(m)}(t_v^{(l)})-t_-^{(m)}(t_v^{(l)}))$ and $\tilde u_v^{(m)}=(1/2)(t_+^{(m)}(t_{v-1}^{(l)})-t_-^{(m)}(t_{v-1}^{(l)}))$, we infer

$$\begin{aligned}
&\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n\sum_{j\ge1} w_{jk}^{l,m}\sum_{i=1}^{n_l}\Big(X^{(l)}_{t_i^{(l)}}-X^{(l)}_{t_{i-1}^{(l)}}\Big)X^{(l)}\Phi_{jk}(\bar t_i^{(l)})\sum_{v=1}^{n_m}\Big(X^{(m)}_{t_v^{(m)}}-X^{(m)}_{t_{v-1}^{(m)}}\Big)\Phi_{jk}(\bar t_v^{(m)})\\
&=\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n\sum_{j\ge1} w_{jk}^{l,m}\sum_{i=1}^{n_l}\Big(X^{(l)}_{t_i^{(l)}}-X^{(l)}_{t_{i-1}^{(l)}}\Big)\Phi_{jk}(\bar t_i^{(l)})\sum_{v=1}^{n_l}\Big(X^{(m)}_{t_v^{(l)}}-X^{(m)}_{t_{v-1}^{(l)}}\Big)\Phi_{jk}(\bar t_v^{(l)})\\
&+\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n\sum_{j\ge1} w_{jk}^{l,m}\sum_{v=1}^{n_l}\Big(X^{(l)}_{t_v^{(l)}}-X^{(l)}_{t_{v-1}^{(l)}}\Big)\Phi_{jk}(\bar t_v^{(l)})\\
&\times\Big(\sum_{\Delta_i t^{(m)}\subset\Delta_v t^{(l)}}\Big(X^{(m)}_{t_i^{(m)}}-X^{(m)}_{t_{i-1}^{(m)}}\Big)(\Phi_{jk}(\bar t_i^{(m)})-\Phi_{jk}(\bar t_v^{(l)}))+\\
&\Big(X^{(m)}_{t_+^{(m)}(t_{v-1}^{(l)})}-X^{(m)}_{t_{v-1}^{(l)}}\Big)\big(\Phi_{jk}(\tilde u_v^{(m)})-\Phi_{jk}(\bar t_v^{(l)})\big)+\\
&\Big(X^{(m)}_{t_v^{(l)}}-X^{(m)}_{t_-^{(m)}(t_v^{(l)})}\Big)\big(\Phi_{jk}(u_v^{(m)})-\Phi_{jk}(\bar t_v^{(l)})\big)\Big).
\end{aligned}$$

Since the observation times are independent of X according to Assumption (N-d), we can employ basic estimates from Proposition A.3.2 to the above increments of X. Applying the bound (2.6.29), we find that the order of the last summand is $\sum_k h_n\sum_j w_{jk}^{l,m}j/(nh_n)$ and since for all weights the bound (2.6.7) holds we conclude that the approximation error is uniformly of order $O_{\mathbb{P}}(h_n)=o_{\mathbb{P}}(n^{-1/4})$.

Leading terms

This paragraph develops the asymptotics for the right-hand side of (2.6.27) and the sum of the increments in (2.4.9). We focus on the oracle versions of (2.3.19) and (2.3.16) with their deterministic optimal weights. The proof follows the same methodology as the proof of Proposition 2.4.1 after restricting to a synchronous reference observation scheme. We concisely go through the details for cross terms and the proof for the bivariate spectral covolatility estimator.

We apply again Theorem A.1.2. For the spectral estimator in (2.3.16) consider

$$\zeta_k^n=n^{1/4}\,h_n\Big(\sum_{j\ge1}w_{jk}^{p,q}\zeta_{jk}^{(pq)}-\Sigma^{(pq)}_{(k-1)h_n}\Big)\,, \tag{2.6.30}$$

with the random variables

$$\begin{aligned}
\zeta_{jk}^{(pq)}=\Bigg(&\Big(\sum_{i=1}^{n_p}\Delta_i^n\bar C^{(p)}\Phi_{jk}\big(\bar t_i^{(p)}\big)-\sum_{i=1}^{n_p-1}\varepsilon^{(p)}_{t_i^{(p)}}\varphi_{jk}\big(t_i^{(p)}\big)\frac{t_{i+1}^{(p)}-t_{i-1}^{(p)}}{2}\Big)\\
\times&\Big(\sum_{v=1}^{n_q}\Delta_v^n\bar C^{(q)}\Phi_{jk}\big(\bar t_v^{(q)}\big)-\sum_{v=1}^{n_q-1}\varepsilon^{(q)}_{t_v^{(q)}}\varphi_{jk}\big(t_v^{(q)}\big)\frac{t_{v+1}^{(q)}-t_{v-1}^{(q)}}{2}\Big)\Bigg).
\end{aligned}$$

The agreement with (2.4.9) follows from a generalization of the summation by parts identity (2.6.4):

$$S_{jk}^{(p)} \asymp_p -\sum_{v=1}^{n_p-1} Y_v^{(p)}\Big(\Phi_{jk}\big(\bar t_{v+1}^{(p)}\big)-\Phi_{jk}\big(\bar t_v^{(p)}\big)\Big)$$
$$\asymp_p -\sum_{v=1}^{n_p-1} Y_v^{(p)}\varphi_{jk}(t_v^{(p)})\frac{t_{v+1}^{(p)}-t_{v-1}^{(p)}}{2}\,.$$

The first relation is an equality under (N-d) when $t_0^{(p)}=0$ and $t_{n_p}^{(p)}=1$. $t_0^{(p)}\neq 0$ or $t_{n_p}^{(p)}\neq 1$ are possible for more general observation schemes, but the distances from the edges are asymptotically small, the remainder due to end-effects is asymptotically negligible. Also, the second remainder by application of mean value theorem and passing to arguments $t_v^{(p)}$ is asymptotically negligible. This remainder can be treated as the approximation error between discrete and continuous-time norm of the (φ_{jk}) in the following.

By Lemma 2.4.4 we may without loss of generality work under synchronous observations $t_i, i=0,\dots,n$, when considering the signal part X. Set $\bar t_i=(t_{i+1}-t_i)/2$. We shall write in the sequel terms of the signal part as coming from observations on a synchronous grid (t_i), while keeping to the actual grids for the noise terms. For the expectation we have

$$\mathbb{E}\left[\zeta_{jk}^{(pq)}\right]=\sum_{i=1}^n \Phi_{jk}^2(\bar t_i)\mathbb{E}\big[\Delta_i^n\bar C^{(p)}\Delta_i^n\bar C^{(q)}\big]$$
$$+\sum_{i,v=1}^{(n_p\vee n_q)-1}\mathbb{E}\left[\varepsilon^{(p)}_{t_i^{(p)}}\varepsilon^{(q)}_{t_i^{(q)}}\right]\varphi_{jk}\big(t_i^{(p)}\big)\Big(\tfrac{t_{i+1}^{(p)}-t_{i-1}^{(p)}}{2}\Big)\varphi_{jk}\big(t_v^{(q)}\big)\Big(\tfrac{t_{v+1}^{(q)}-t_{v-1}^{(q)}}{2}\Big)$$
$$=\sum_{i=1}^n \Phi_{jk}^2(\bar t_i)(t_{i+1}-t_i)\Sigma_{(k-1)h_n}^{(pq)}=\Sigma_{(k-1)h_n}^{(pq)}+R_{n,k}$$

by Itô isometry. Observe that

$$\sum_{i=1}^{n_l-1}\varphi_{jk}^2\big(t_i^{(l)}\big)\Big(\tfrac{t_{i+1}^{(l)}-t_{i-1}^{(l)}}{2}\Big)^2\asymp\sum_{i=1}^{n_l-1}\varphi_{jk}^2\big(t_i^{(l)}\big)\tfrac{t_{i+1}^{(l)}-t_{i-1}^{(l)}}{2}\,\frac{H_l^{(k-1)h_n}}{\eta_l^2 n_l}$$
$$\asymp\Big(\int_0^1\varphi_{jk}^2(t)\,dt\Big)\frac{H_l^{(k-1)h_n}}{\eta_l^2 n_l}. \tag{2.6.31}$$

The left approximation uses $(t_{i+1}^{(l)}-t_i^{(l)})/2=(H_l^{kh_n}+O(h_n^\alpha))/(\eta_l^2 n_l)$ as in (2.6.5) with $\alpha>1/2$ by (N-d). Writing the integral on the right-hand side as sum over the subintervals and using mean value theorem, the differences when passing to the arguments $(t_i^{(l)})_i$ induce approximation errors of order $jh_n^{-1}n^{-1}$. Thus, the total approximation errors in (2.6.31) are of order $(h_n^\alpha+j(nh_n)^{-1})j^2(nh_n^2)^{-1}$.

The remainders $R_{n,k}$ due to the approximation in (2.6.31) satisfy with (2.6.7) uniformly

$$R_{n,k} \lesssim \sum_{j=1}^{\lfloor \sqrt{n}h_n \rfloor} j^2 n^{-1} h_n^{-2}\big(h_n^{\alpha} + j n^{-1} h_n^{-1}\big) + \sum_{\lceil \sqrt{n}h_n \rceil}^{\lfloor nh_n \rfloor -1} \big(j^{-1} h_n + j^{-2} h_n^2 n h_n^{\alpha}\big),$$

which is of order $o\big(n^{-1/4}\big)$. Since $\sum_{j\ge 1} w_{jk}^{p,q} = 1$, asymptotic unbiasedness is ensured:

$$\sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \mathbb{E}\left[\zeta_k^n \middle| \mathcal{G}_{(k-1)h_n}\right] = \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} n^{1/4} h_n \left(\sum_{j\ge 1} w_{jk}^{p,q} \mathbb{E}[\zeta_{jk}^{(pq)}] - \Sigma_{(k-1)h_n}^{(pq)} \right) \xrightarrow{ucp} 0\,.$$

We now determine the asymptotic variance expression in (2.2.4):

$$\begin{aligned}
\operatorname{Var}\big(\zeta_{jk}^{(pq)}\big) &= \Big(\sum_{i=1}^{n} \Phi_{jk}^2(\bar{t}_i)(t_{i+1}-t_i)\Big)^2 \Big(\big(\Sigma_{(k-1)h_n}^{(pq)}\big)^2 + \Sigma_{(k-1)h_n}^{(pp)} \Sigma_{(k-1)h_n}^{(qq)}\Big) \\
&\quad + \eta_p^2 \eta_q^2 \Big(\sum_{i=1}^{n_p-1} \varphi_{jk}^2\big(t_i^{(p)}\big) \Big(\tfrac{t_{i+1}^{(p)} - t_{i-1}^{(p)}}{2}\Big)^2 \Big) \Big(\sum_{i=1}^{n_q-1} \varphi_{jk}^2\big(t_i^{(q)}\big) \Big(\tfrac{t_{i+1}^{(q)} - t_{i-1}^{(q)}}{2}\Big)^2 \Big) \\
&\quad + \left(\sum_{i=1}^{n} \Phi_{jk}^2(\bar{t}_i)(t_{i+1}-t_i) \Big(\eta_p^2 \Sigma_{(k-1)h_n}^{(qq)} \sum_{i=1}^{n_p-1} \varphi_{jk}^2\big(t_i^{(p)}\big) \Big(\tfrac{t_{i+1}^{(p)} - t_{i-1}^{(p)}}{2}\Big)^2 \right. \\
&\quad \left. + \eta_m^2 \Sigma_{(k-1)h_n}^{(pp)} \sum_{i=1}^{n_q-1} \varphi_{jk}^2\big(t_i^{(q)}\big) \Big(\tfrac{t_{i+1}^{(q)} - t_{i-1}^{(q)}}{2}\Big)^2 \Big) \right) \\
&\asymp \big(\Sigma_{(k-1)h_n}^{(pq)}\big)^2 + \Sigma_{(k-1)h_n}^{(pp)} \Sigma_{(k-1)h_n}^{(qq)} + \pi^2 j^2 h_n^{-2} \big(H_p^{(k-1)h_n} n_p^{-1} \Sigma_{(k-1)h_n}^{(qq)} \\
&\quad + H_q^{(k-1)h_n} n_q^{-1} \Sigma_{(k-1)h_n}^{(pp)} \big) + \pi^4 j^4 h_n^{-4} n_p^{-1} n_q^{-1} H_p^{(k-1)h_n} H_q^{(k-1)h_n},
\end{aligned}$$

where the remainder is negligible by the same bounds as for the bias above. The sum of conditional variances with $w_{jk}^{p,q} = I_k^{-1} I_{jk}$, $I_k = \sum_{j\ge 1} I_{jk}$, thus yields

$$\begin{aligned}
\sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \mathbb{E}\left[(\zeta_k^n)^2 \,\middle|\, \mathcal{G}_{(k-1)h_n} \right] + o_{\mathbb{P}}(1) &= \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n^2 n^{1/2} \sum_{j\ge 1} \big(w_{jk}^{(pq)}\big)^2 \operatorname{Var}\big(\zeta_{jk}^{(pq)}\big) \\
&= \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n^2 n^{1/2} \sum_{j\ge 1} I_{jk} I_k^{-2} = \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n^2 n^{1/2} I_k^{-1}.
\end{aligned}$$

As $h_n\sqrt{n} \to \infty$, we obtain an asymptotic expression as the solution of an integral

$$\begin{aligned}
&\sum_{k=1}^{\lfloor th_n^{-1} \rfloor} \mathbb{E}\left[(\zeta_k^n)^2 \,\middle|\, \mathcal{G}_{(k-1)h_n} \right] \\
&= \sum_{k=1}^{\lfloor th_n^{-1} \rfloor} h_n(\sqrt{n} h_n) I_k^{-1} \to \int_0^t \left(\int_0^\infty (f(\Sigma, \mathcal{H}(t), \nu_p, \nu_q; z))^{-1} dz \right)^{-1} ds
\end{aligned}$$

with a continuous limit function f which is the same as in Bibinger and Reiß (2014). Computing the solution of the integral using the explicit form of I_k and f yields the variance $\int_0^t \big(v_s^{(p,q)}\big)^2\,ds$ with

$$\big(v_s^{(p,q)}\big)^2 = 2\Big(\mathcal{H}^2\,(s)^{(p)}\,\mathcal{H}^2\,(s)^{(q)}\,(A_s^2 - B_s)B_s\Big)^{\frac{1}{2}} \\ \times \Big(\sqrt{A_s + \sqrt{A_s^2 - B_s}} - \operatorname{sgn}(A_s^2 - B_s)\sqrt{A_s - \sqrt{A_s^2 - B_s}}\Big)\,,$$

and the terms

$$A_s = \Sigma_s^{(pp)}\frac{\mathcal{H}^2\,(s)^{(q)}}{\mathcal{H}^2\,(s)^{(p)}} + \Sigma_s^{(qq)}\frac{\mathcal{H}^2\,(s)^{(p)}}{\mathcal{H}^2\,(s)^{(q)}}, \quad B_s = 4\Big(\Sigma_s^{(pp)}\Sigma_s^{(qq)} + \big(\Sigma_s^{(pq)}\big)^2\Big)\,.$$

The detailed computation is carried out in Bibinger and Reiß (2014). sgn denotes the sign taking values in $\{-1,+1\}$ and ensuring that the value of $\big(v_s^{(p,q)}\big)^2$ is always a positive real number. Contrarily to the one-dimensional case, in the cross term there is no effect of non-Gaussian noise on the variance because fourth noise moments do not occur and because of component-wise independence.

The Lyapunov criterion follows from

$$\mathbb{E}\Big[\big(\zeta_{jk}^{(pq)}\big)^4|\mathcal{G}_{(k-1)h_n}\Big] \asymp 3\sum_{j\ge1}\big(w_{jk}^{p,q}\big)^4 I_{jk}^{-2} \asymp 3\,I_k^{-4}\sum_{j\ge1} I_{jk}^2 = O_{\mathbb{P}}(1) \\ \Rightarrow \sum_{k=1}^{\lfloor th_n^{-1}\rfloor}\mathbb{E}\Big[\Big(\zeta_k^n\Big)^4|\mathcal{G}_{(k-1)h_n}\Big] = O_{\mathbb{P}}\Big(n\sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n^4\Big) = o_{\mathbb{P}}\Big(n^{-1/4}\Big)\,.$$

By Cauchy-Schwarz and Burkholder-Davis-Gundy inequalities, we deduce

$$\mathbb{E}\left[h_n\sum_{j\ge1} w_{jk}^{p,q}\sum_{i=1}^n \Delta_i^n\bar{C}^{(p)}\Delta_i^n\bar{C}^{(q)}\Phi_{jk}^2(\bar{t}_i)\sum_{i=1}^n\Delta_i^n W^{(p)}\right] \\ = h_n\sum_{j\ge1} w_{jk}^{p,q}\sum_{i=1}^n \mathbb{E}\Big[\Delta_i^n\bar{C}^{(p)}\Delta_i^n\bar{C}^{(q)}\Delta_i^n W^{(p)}\Big]\,\Phi_{jk}^2(\bar{t}_i) \\ \le h_n\sum_{j\ge1} w_{jk}^{p,q}\sum_{i=1}^n (t_i - t_{i-1})^{3/2}\Phi_{jk}^2(\bar{t}_i) = o_{\mathbb{P}}\big(n^{-1/4}\big).$$

By the analogous estimate with $\Delta_i^n W^{(q)}$ the stability conditions are valid. This proves stable convergence of the leading term to the limit given in Theorem 2.2.2.

The heart of the proof of Theorem 2.2.3 is the asymptotic theory for the leading term (2.6.27), namely the analysis of the asymptotic variance-covariance structure. This is carried out in detail in Bibinger et al. (2014) for the idealized locally parametric experiment using bin-wise orthogonal transformation to a diagonal covariance structure. The only difference between our main term and the setup considered in Bibinger et al. (2014) is the Gaussianity of the noise component. Yet, in the deduction of the variance this only affects the terms with fourth noise moments where $\mathbb{E}[\varepsilon_i^4] \neq 3\mathbb{E}[\varepsilon_i^2]$ in general.

Above, we explicitly proved that the resulting remainder converges to zero for the one-dimensional estimator and this directly extends to the diagonal elements here. An intuitive heuristic reason why this holds is that the smoothed statistics are asymptotically still close to a normal distribution, though the normality which could have been used in Bibinger et al. (2014) does not hold here for fixed n in general. Based on the expressions of variances for cross products and squared spectral statistics above, coinciding their counterparts in the normal noise model when separating the remainder induced for the squares, we can pursue the asymptotics along the same lines as the proof of Corollary 4.3 in Bibinger et al. (2014).

At this stage, we restrict to shed light on the connection between the expressions in (2.2.7) and the asymptotic covariance matrix. Observe that $(A\otimes B)^\top = A^\top \otimes B^\top$ for matrices A, B, $\mathcal{Z}\mathcal{Z} = 2\mathcal{Z}$ and that $(A\otimes B)(C\otimes D) = (AC\otimes BD)$ for matrices A, B, C, D, such that

$$\begin{aligned}
&\Big(\Sigma_s^{1/2}\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\Big)\mathcal{Z}\Big(\Big(\Sigma_s^{1/2}\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\Big)\mathcal{Z}\Big)^\top\\
&\quad=\Big(\Sigma_s^{1/2}\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\Big)2\mathcal{Z}\Big(\Sigma_s^{1/2}\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\Big)^\top\\
&\quad=2\Big(\Sigma_s\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/2}\Big)\mathcal{Z}\,,
\end{aligned}$$

since $\mathcal{Z}$ commutes with $\Big(\Sigma_s^{1/2}\otimes\big(\Sigma_s^{\mathcal{H}}\big)^{1/4}\Big)$. Therefore, the expression in (2.2.7) is natural for the matrix square root of the asymptotic covariance, where we use two independent terms because of non-commutativity of matrix multiplication. Conditions (J1) and (J3) and the stability conditions (J4) and (J5) can be analogously shown by element-wise adopting the results for squared and cross products of spectral statistics from above. Since any component of the estimator is a weighted sum of the entries of $S_{jk}S_{jk}^\top$, bias-corrected on the diagonal, the convergences to zero in probability follow likewise.

Remainder terms

After applying the triangular inequality to (2.6.28), it suffices to prove that

$$n^{1/4}\|LMM_{n,t}^{or}(Y)-LMM_{n,t}^{or}(\bar{C}+\varepsilon)\|\stackrel{ucp}{\longrightarrow}0\,,\tag{2.6.32}$$

$$n^{1/4}\left\|\int_0^t \operatorname{vec}\big(\Sigma_s-\Sigma_{\lfloor sh_n^{-1}\rfloor h_n}\big)\,ds\right\|\stackrel{ucp}{\longrightarrow}0\,.\tag{2.6.33}$$

For $A, B\in\mathbb{R}^d$, we use in the following several times the elementary bound:

$$\left\|AA^\top-BB^\top\right\|=\left\|B(A^\top-B^\top)+(A-B)A^\top\right\|\le\big(\|A\|+\|B\|\big)\|A-B\|\,.\tag{2.6.34}$$

Define analogously as above $\tilde S_{jk}=\big(\sum_{i=1}^{n_p}\Delta_i^n\bar C^{(p)}\Phi_{jk}\big(\bar t_i^{(p)}\big)\big)_{1\le p\le d}$, the spectral statistics in the locally constant volatility experiment. Then we can bound uniformly for all t:

$$\begin{aligned}
&\|LMM_{n,t}^{or}(Y)-LMM_{n,t}^{or}(\bar C+\varepsilon)\|\\
&\le\sum_{k=1}^{h_n^{-1}}h_n\left\|\sum_{j=1}^{\lfloor nh_n\rfloor-1}W_{jk}\mathrm{vec}\big(S_{jk}S_{jk}^\top-\tilde S_{jk}\tilde S_{jk}^\top\big)\right\|\\
&\le\sum_{k=1}^{h_n^{-1}}h_n\sum_{j=1}^{\lfloor nh_n\rfloor-1}\|W_{jk}\|\big(\|S_{jk}\|+\|\tilde S_{jk}\|\big)\|S_{jk}-\tilde S_{jk}\|\\
&\lesssim\sum_{k=1}^{h_n^{-1}}h_n\sum_{j=1}^{\lfloor nh_n\rfloor-1}\Big(1+\frac{j^2}{nh_n^2}\Big)^{-2}\|S_{jk}-\tilde S_{jk}\|=O_{\mathbb P}\big(h_n\big)=o_{\mathbb P}(n^{-1/4}),
\end{aligned}$$

what yields (2.6.32). We have used Lemma C.1 from Bibinger et al. (2014) for the magnitude of $\|W_{jk}\|$, the bound (2.6.34) and a bound for the sum over j, for which holds

$$\frac{1}{\sqrt n h_n}\sum_{j=1}^{\lfloor nh_n\rfloor-1}\Big(1+\frac{j^2}{nh_n^2}\Big)^{-2}\to\frac{\pi}{2}$$

by an analogous integral approximation as used in the limiting variance before. Drift terms and cross terms including the drift are asymptotically negligible and are handled similarly as before. Directly neglecting drift terms, we deduce $\|S_{jk}-\tilde S_{jk}\|=O_{\mathbb P}(h_n)$ uniformly from $(S_{jk}-\tilde S_{jk})^{(p)}\asymp_p\sum_{i=1}^{n_p}\Delta_i^n\tilde C^{(p)}\Phi_{jk}(\bar t_i)$ by Proposition A.3.2. (2.6.33) is equivalent to

$$n^{1/4}\left\|\sum_{k=1}^{h_n^{-1}}\int_{(k-1)h_n}^{kh_n}\big(\Sigma_s-\Sigma_{(k-1)h_n}\big)\,ds\right\|\xrightarrow{ucp}0. \tag{2.6.35}$$

Using the decomposition

$$\begin{aligned}
\Sigma_s-\Sigma_{(k-1)h_n}&=\sigma_s\sigma_s^\top-\sigma_{(k-1)h_n}\sigma_{(k-1)h_n}^\top\\
&=(\sigma_s-\sigma_{(k-1)h_n})\sigma_{(k-1)h_n}^\top+\sigma_{(k-1)h_n}(\sigma_s^\top-\sigma_{(k-1)h_n}^\top)\\
&\quad+(\sigma_s-\sigma_{(k-1)h_n})(\sigma_s^\top-\sigma_{(k-1)h_n}^\top)
\end{aligned}$$

for $s\in[(k-1)h_n,kh_n]$, it is easy to find that it suffices to bound terms $\|\sigma_s-\sigma_{(k-1)h_n}\|$. Then, Assumption (H-$\alpha$-$\beta$) guarantees (2.6.35) and (2.6.33) in the same way as for the one-dimensional model.

For the spectral covolatility estimator (2.3.16) we may conduct an analysis of the remainder similarly as in the proof of Proposition 2.4.2. One can as well employ integration by parts of Itô integrals after supposing again a synchronous observation design $t_i, i=0,\dots,n$, possible according to Lemma 2.4.4:

$$\begin{aligned}
&\Delta_i^n\tilde C^{(p)}\Delta_i^n\tilde C^{(q)}-\int_{t_{i-1}}^{t_i}\big(\Sigma_s^{(pq)}-\Sigma^{(pq)}_{\lfloor sh_n^{-1}\rfloor h_n}\big)ds\\
&\quad=\int_{t_{i-1}}^{t_i}\big(\tilde C_s^{(p)}-\tilde C_{t_{i-1}}^{(p)}\big)d\tilde C_s^{(q)}+\int_{t_{i-1}}^{t_i}\big(\tilde C_s^{(q)}-\tilde C_{t_{i-1}}^{(q)}\big)d\tilde C_s^{(p)}\,.
\end{aligned} \tag{2.6.36}$$

with $\tilde{C}$ approximation errors as in (2.6.26). Consider the random variables

$$\tilde{\zeta}_{jk}^{(pq)} = \sum_{i=1}^{n} \Delta_i \tilde{C}^{(p)} \Phi_{jk}(\bar{t}_i) \sum_{v=1}^{n} \Delta_v \tilde{C}^{(q)} \Phi_{jk}(\bar{t}_v)\,,$$
$$\tilde{\zeta}_k^n = h_n \sum_{j\ge 1} w_{jk}^{p,q} \tilde{\zeta}_{jk}^{(pq)} - \int_{kh_n}^{(k+1)h_n} \big(\Sigma_s^{(pq)} - \Sigma_{\lfloor sh_n^{-1}\rfloor h_n}^{(pq)}\big) ds\,.$$

Inserting (2.6.36) for $\Delta_i^n \tilde{C}^{(p)} \Delta_i^n \tilde{C}^{(q)}$, using $\langle \int Z\,dX, \int Z\,dX] \rangle = \int Z^2\,d\langle X,X\rangle$ for Itô integrals and applying Burkholder-Davis-Gundy inequalities and using Proposition A.3.2 for $\mathbb{E}\big[\big(\Delta_i^n \tilde{C}^{(p)}\big)^2\big]$, $\mathbb{E}\big[\big(\Delta_i^n \tilde{C}^{(q)}\big)^2\big]$, it follows that $\mathbb{E}\big[\big(\tilde{\zeta}_k^n\big)^2\big] = O(n^{-1})$. Bounds for cross terms with $\tilde{C}$ and $\bar{C}$ readily follow by standard estimates and we conclude our claim.

2.6.5. Proofs for adaptive estimation

We carry out the proof of Proposition 2.4.3 in the case $d=1$ explicitly. We need to show that

$$n^{\frac{1}{4}} \left| \widehat{IV}_{n,t} - \widehat{IV}_{n,t}^{or}(Y) \right| \xrightarrow{ucp} 0 \quad \text{as } n \to \infty\,. \tag{2.6.37}$$

Let us first act as if the noise level η was known and concentrate on the harder problem of analyzing the plug-in estimation of the instantaneous squared volatility process σ_t^2 in the weights. We have to bound

$$\widehat{IV}_{n,t} - \widehat{IV}_{n,t}^{or}(Y) = \sum_{k=1}^{\lfloor th_n^{-1}\rfloor} h_n \sum_{j=1}^{\lfloor nh_n\rfloor -1} \big(\widehat{w}_{jk} - w_{jk}\big)\Big(S_{jk}^2 - [\varphi_{jk},\varphi_{jk}]_n \frac{\eta^2}{n}\Big)\,,$$

uniformly with w_{jk} being the optimal oracle weights (2.3.5) and $\widehat{w}_{jk}$ their adaptive estimates. We introduce a coarse grid of blocks of lengths r_n such that $r_n h_n^{-1} \to \infty$ as $n \to \infty$. We analyze the above difference in this double asymptotic framework, where the plug-in estimators are evaluated on the coarse grid first. Denoting the adaptive and oracle estimators with weights evaluated on the coarse grid by $\widehat{IV}_{n,t}^{c}$ and $\widehat{IV}_{n,t}^{or,c}(Y)$, respectively, $\widehat{IV}_{n,t}^{c} - \widehat{IV}_{n,t}^{or,c}(Y)$ is equal to

$$\sum_{m=1}^{\lfloor tr_n^{-1}\rfloor} h_n \sum_{k=(m-1)r_nh_n^{-1}+1}^{mr_nh_n^{-1}} \sum_{j=1}^{\lfloor nh_n\rfloor -1} \big(w_j(\widehat{\sigma}^2_{(m-1)r_n}) - w_j(\sigma^2_{(m-1)r_n})\big)\, Z_{jk} \tag{2.6.38}$$

with $Z_{jk} = S_{jk}^2 - [\varphi_{jk},\varphi_{jk}]_n \eta^2/n - \sigma^2_{(k-1)h_n}$ where the weights are functions (independent of the block k, as $[\varphi_{jk},\varphi_{jk}]_n$ does not depend on k)

$$w_j(x) = \frac{\big(x + \frac{\eta^2}{n}[\varphi_{jk},\varphi_{jk}]_n\big)^{-2}}{\sum_{l=1}^{\lfloor nh_n\rfloor -1} \big(x + \frac{\eta^2}{n}[\varphi_{lk},\varphi_{lk}]_n\big)^{-2}}\,,$$

which are well-defined for $x \in \mathbb{R}_+$ and satisfy $\sum w_j(x) = 1$. As σ is uniformly bounded from below and from above, there exists a constant $C_1 > 0$ such that uniformly

$$w_j\left(\sigma_t^2\right) \lesssim w_j\left(C_1\right) \tag{2.6.39}$$

for all $j \geq 1$. By the proof of Proposition 2.4.2 we may directly consider $\widetilde{Z}_{jk} = \tilde{S}_{jk}^2 - [\varphi_{jk}, \varphi_{jk}]_n \eta^2/n - \sigma^2_{(k-1)h_n}$ where the $\tilde{S}_{jk}$ are the statistics under locally parametric volatility and without drift. Moreover, by subtracting $\sigma^2_{(k-1)h_n}$ in the definition of $\tilde{Z}_{jk}$ equation (2.6.11) shows that the $\tilde{Z}_{jk}$ are uncorrelated for different k. Hence, $\mathrm{Var}\big(\sum_k \tilde{Z}_{jk}\big) = \sum_k \mathrm{Var}(\tilde{Z}_{jk})$ and thus

$$\mathrm{Var}\Big(\sum_k Z_{jk}\Big) = \sum_k \mathrm{Var}(Z_{jk}) + o(1). \tag{2.6.40}$$

We prove (2.6.37) in two steps. We show first that (2.6.38) is $o_{\mathbb{P}}(n^{-1/4})$ uniformly and then that the difference between estimating on the coarse and finer grid is $o_{\mathbb{P}}(n^{-1/4})$, as well. The crucial property to ensure tightness of the adaptive approach is a uniform bound on the first derivatives of the weight functions: $w_j(x)$ is continuously differentiable with derivatives satisfying:

$$\big|w_j'(x)\big| \lesssim w_j(x)\log^2(n). \tag{2.6.41}$$

To see why this holds set $c_j = \frac{\eta^2}{n}[\varphi_{jk}, \varphi_{jk}]_n$ and observe that $|w_j'(x)|$ is equal to

$$\left|\frac{-2(x+c_j)^{-3}\sum_{m=1}^{\lfloor nh\rfloor-1}(x+c_m)^{-2} - (x+c_j)^{-2}\sum_{m=1}^{\lfloor nh\rfloor-1}\Big((-2)(x+c_m)^{-3}\Big)}{\Big(\sum_{m=1}^{\lfloor nh\rfloor-1}(x+c_m)^{-2}\Big)^2}\right|$$
$$\leq 2w_j(x)\,\frac{\sum_{m=1}^{\lfloor nh\rfloor-1}(x+c_m)^{-2}\left|(x+c_j)^{-1}-(x+c_m)^{-1}\right|}{\sum_{m=1}^{\lfloor nh\rfloor-1}(x+c_m)^{-2}} \lesssim w_j(x)\log^2(n)$$

for n sufficiently large. The last inequality follows from

$$\left|(x+c_j)^{-1}-(x+c_m)^{-1}\right| \leq \frac{1}{c_j}+\frac{1}{c_m} \lesssim \frac{1}{c_1} = O\left(\log^2(n)\right).$$

The plug-in estimator (2.3.7) satisfies $\|\widehat{\sigma}^2-\sigma^2\|_{L^1} = \mathcal{O}_{\mathbb{P}}\big(\delta_n\big)$ for the L^1-norm $\|\cdot\|_{L^1}$ with a sequence $\delta_n \to 0$ as $n \to \infty$, $\delta_n \lesssim n^{-1/8}$ for optimal window length under (H-α-β). Hence, by (2.6.39) $w_j(\widehat{\sigma}^2_{(m-1)r_n} - w_j(\sigma^2_{(m-1)r_n}) = \mathcal{O}_{\mathbb{P}}(w_j(C_1)\delta_n\log^2(n))$. This, (2.6.40), the Cauchy-Schwarz inequality and (2.6.7) show that $\mathbb{E}[|\widehat{IV}^{c}_{n,t} - \widehat{IV}^{or,c}_{n,t}(Y)|]$ is up to a constant bounded by

$$\mathbb{E}\left[\sum_{m=1}^{\lfloor tr_n^{-1}\rfloor} h_n \sum_{j=1}^{\lfloor nh_n\rfloor-1}\left|w_j(\hat{\sigma}^2_{(m-1)r_n})-w_j(\sigma^2_{(m-1)r_n})\right|\left|\sum_{k=(m-1)r_nh_n^{-1}+1}^{mr_nh_n^{-1}} Z_{jk}\right|\right]$$
$$\lesssim \delta_n\left(\log^2(n)\right)\sum_{m=1}^{\lfloor tr_n^{-1}\rfloor} h_n \sum_{j=1}^{\lfloor nh_n\rfloor-1} w_j(C_1)\Big(\mathrm{Var}\Big(\sum_{k=(m-1)r_nh_n^{-1}+1}^{mr_nh_n^{-1}} Z_{jk}\Big)\Big)^{1/2}$$
$$\lesssim \sqrt{\frac{h_n}{r_n}}\delta_n\log^2 n + o(1). \tag{2.6.42}$$

The required order $o_{\mathbb{P}}(n^{-1/4})$ for (2.6.38) is thus achieved if $r_n \to 0$ not too fast, i.e. $r_n^{-1} \lesssim n^{1/4}(\log n)^{-5}$. Consider the remainder by the difference of coarse and fine grid. Since for $\widehat{IV}_{n,t}^{or,c}(Y)$ and $\widehat{IV}_{n,t}^{or}(Y)$ the statistics for each block k are uncorrelated, it is enough to bound the variance of the difference by

$$\sum_{m=1}^{r_n^{-1}} \sum_{k=(m-1)r_n h_n^{-1}+1}^{m r_n h_n^{-1}} h_n^2 \left(\sum_{j=1}^{\lfloor n h_n \rfloor -1} \left(\mathbb{E}\left[\left(w_j(\sigma^2_{(k-1)h_n}) - w_j(\sigma^2_{(m-1)r_n}) \right)^2 \tilde{Z}_{jk}^2 \right] \right)^{\frac{1}{2}} \right)^2 ,$$

which is of order $O(h_n r_n \log^4(n))$ using (2.6.41) and (2.6.42). This shows that $|\widehat{IV}_{n,t}^{or,c}(Y) - \widehat{IV}_{n,t}^{or}(Y)| = o_{\mathbb{P}}(n^{-1/4})$ uniformly. Exploiting the same ingredients as above we obtain likewise that $|\widehat{IV}_{n,t}^{c} - \widehat{IV}_{n,t}| = o_{\mathbb{P}}(n^{-1/4})$ uniformly. In order to analyze the estimation error induced by pre-estimation of η^2, we can consider the weights as functions of η^2 and compute their derivatives. As η^2 does not depend on time and we have $|\widehat{\eta}^2 - \eta^2| = O_{\mathbb{P}}(n^{-1/2})$, a simpler computation yields that the pre-estimation of η^2 is of smaller order as the error by plug-in estimation of local volatilities. Thus, using triangle inequality we conclude (2.6.37).

The proofs that Theorem 2.2.2 and Theorem 2.2.3 extend from the oracle to the adaptive versions of the estimators (2.3.16) and (2.3.19) can be conducted in an analogous way. For covariation matrix estimation, the key ingredient is the uniform bound on the norm of the matrix derivative of the weight matrix function $W_j(\Sigma)$ w.r.t. Σ, which is a matrix with d^6 entries and requires a notion of matrix derivatives, see Lemma C.2 in Bibinger et al. (2014). The proof is then almost along the same lines as the proof of Theorem 4.4 in Bibinger et al. (2014), with the only difference in the construction being that the Z_{jk} are not independent, but still have negligible correlations. The adaptivity in the proof of Theorem 4.4 of Bibinger et al. (2014) is proved under more delicate asymptotics of asymptotically separating sample sizes. For this reason, but at the same time not having the remainders, the restrictions on r_n are different there.

Chapter 3.

Estimating occupation time functionals

This chapter is adapted from Altmeyer (2017) (Sections 3.1, 3.2 and 3.4) and Altmeyer and Chorowski (2017) (Section 3.3).

In this chapter we estimate occupation time functionals with respect to discrete observations by a Riemann-sum estimator. In the first section central limit theorems are proved for L^2-Sobolev functions f and continuous Itô semimartingales X. We then provide general $L^2(\mathbb{P})$-upper bounds on the error in the second section and discuss several examples in detail. Section three provides a different method for obtaining $L^2(\mathbb{P})$-upper bounds for stationary Markov processes. The fourth section studies the optimality of the $L^2(\mathbb{P})$-upper bounds in case of Brownian motion. Proofs can be found in section five.

If not stated otherwise, we assume that X is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{0\leq t\leq T}, \mathbb{P})$.

3.1. Central limit theorems

Let X be a continuous d-dimensional Itô semimartingale as in (2.1.1). Recall from the introduction the definition of the occupation time functional

$$\Gamma_t(f) = \int_0^t f(X_r)\,dr$$

and the corresponding Riemann-sum estimator

$$\widehat{\Gamma}_{n,t}(f) = \Delta_n \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} f\left(X_{t_{k-1}}\right),$$

where $t_k = \Delta_n k$ for $\Delta_n = T/n$ and $k = 0, \ldots, n$. We will derive in this section central limit theorems for the error $\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)$ as $\Delta_n \to 0$ with $0 \leq t \leq T$ and T fixed.

3.1.1. CLT for C^2-functions

We first review the basic situation when $f \in C^2(\mathbb{R}^d)$. The following is a special case of Theorem 6.1.2 of Jacod and Protter (2011) for continuous X.

Theorem 3.1.1. *Let $f \in C^2(\mathbb{R}^d)$. Then we have the stable convergence*

$$\Delta_n^{-1}\left(\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f)\right) \xrightarrow{st} \frac{f(X_t) - f(X_0)}{2} + \frac{1}{\sqrt{12}} \int_0^t \left\langle \nabla f(X_r), \sigma_r d\widetilde{W}_r \right\rangle \qquad (3.1.1)$$

as processes on $\mathcal{D}([0,T],\mathbb{R}^d)$*, where* $\widetilde{W}$ *is a* d*-dimensional Brownian motion defined on an independent extension of* $(\Omega,\mathcal{F},(\mathcal{F}_t)_{0\leq t\leq T},\mathbb{P})$*.*

In order to explain the main ideas of the proof consider the decomposition $\Gamma_t(f)-\widehat{\Gamma}_{n,t}(f)=M_{n,t}(f)+D_{n,t}(f)$, where

$$M_{n,t}(f)=\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(f(X_r)-\mathbb{E}\left[f(X_r)\middle|\mathcal{F}_{t_{k-1}}\right]\right)dr, \tag{3.1.2}$$

$$D_{n,t}(f)=\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\mathbb{E}\left[f(X_r)-f\left(X_{t_{k-1}}\right)\middle|\mathcal{F}_{t_{k-1}}\right]dr. \tag{3.1.3}$$

This decomposition is similar to the one in Section 2.4. By the martingale structure of $M_{n,t}(f)$ and using Itô's formula it is easy to check from Theorem A.1.2 that the central limit theorem

$$\Delta_n^{-1}M_{n,t}(f)\overset{st}{\to}\frac{1}{2}\int_0^t\langle\nabla f(X_r),\sigma_r dW_r\rangle+\frac{1}{\sqrt{12}}\int_0^t\left\langle\nabla f(X_r),\sigma_r d\widetilde{W}_r\right\rangle \tag{3.1.4}$$

holds for $n\to\infty$ as processes on $\mathcal{D}([0,T],\mathbb{R}^d)$, where $\widetilde{W}$ is a d-dimensional Brownian motion defined on an independent extension of $(\Omega,\mathcal{F},(\mathcal{F}_t)_{0\leq t\leq T},\mathbb{P})$. In fact, for this $f\in C^1(\mathbb{R}^d)$ is sufficient (for a proof see Proposition 3.5.4). With respect to $D_{n,t}(f)$ Itô's formula shows that $(\Delta_n^{-1}D_{n,t}(f))_{0\leq t\leq T}$ converges uniformly on $[0,T]$ in probability to

$$\left(\frac{f(X_t)-f(X_0)}{2}-\frac{1}{2}\int_0^t\langle\nabla f(X_r),\sigma_r dW_r\rangle\right)_{0\leq t\leq T}. \tag{3.1.5}$$

In particular, $\Delta_n^{-1}D_{n,t}(f)$ is not negligible asymptotically (this is different compared to (J1) in Section 2.4). Summing up $\Delta_n^{-1}M_{n,t}(f)$ and $\Delta_n^{-1}D_{n,t}(f)$ as well as the corresponding limits yields the theorem. It is interesting to note that the CLT implies the stable convergence of $\Delta_n^{-1}(\Gamma_t(f)-\widehat{\Theta}_{n,t}(f))$ to $1/\sqrt{12}\int_0^t\langle\nabla f(X_r),\sigma_r d\widetilde{W}_r\rangle$, where

$$\widehat{\Theta}_{n,t}(f)=\Delta_n\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\frac{f\left(X_{t_{k-1}}\right)+f(X_{t_k})}{2}$$

is the trapezoid rule estimator. Therefore $\widehat{\Theta}_{n,t}(f)$ is actually the more natural estimator for $\Gamma_t(f)$. It is remarkable that the trapezoid rule and the Riemann-sum approximation have the same rate of convergence. This is not true generally for deterministic integrands. We will see in Section 3.4 that both estimators are rate optimal and that the asymptotic variance is optimal.

Remark 3.1.2. From a statistical point of view the stable central limit theorem can be exploited to obtain a *feasible* central limit theorem, like in the theorems of Section 2.2. More precisely, the estimator $\widehat{AVAR_T}(f)=1/12\sum_{k=1}^n\langle\nabla f(X_{t_{k-1}}),X_{t_k}-X_{t_{k-1}}\rangle^2$ converges in probability to $1/12\int_0^T\|\sigma_r^\top\nabla f(X_r)\|^2dr$, which is equal to $\operatorname{Var}(1/\sqrt{12}\int_0^T\langle\nabla f(X_r),\sigma_r d\widetilde{W}_r\rangle)$. The stable convergence and the continuous mapping theorem therefore yield $\Delta_n^{-1}(\widehat{AVAR_T}(f))^{-1/2}(\Gamma_T(f)-\widehat{\Theta}_{n,T}(f))\overset{d}{\to}N(0,1)$. This can be used to derive asymptotic confidence intervals for $\widehat{\Theta}_{n,T}(f)$.

3.1.2. CLT for Fourier-Lebesgue functions

Interestingly, the weak limit in (3.1.1) is also well-defined for less smooth functions. As the argument above relies on Itô's formula, it breaks already for $f \in C^1(\mathbb{R}^d)$. In order to study the limit of $\Delta_n^{-1} D_{n,t}(f)$ for more general f note that we can write

$$f\left(X_r\right) - f\left(X_{t_{k-1}}\right) = (2\pi)^{-d} \int \mathcal{F}f\left(u\right)\left(e^{-i\langle u, X_r\rangle} - e^{-i\langle u, X_{t_{k-1}}\rangle}\right) du, \qquad (3.1.6)$$

if f is sufficiently regular, where $\mathcal{F}f(u) = \int f(x)e^{i\langle u,x\rangle}dx$ is the Fourier transform of f. In principle, we can now study $e^{-i\langle u,X_r\rangle} - e^{-i\langle u,X_{t_{k-1}}\rangle}$ instead of $f(X_r) - f(X_{t_{k-1}})$. The error can be calculated exactly, if the characteristic functions of the marginals X_r are known. For the general Itô semimartingale X in (2.1.1), however, this is a difficult issue. The key idea is to replace the marginals X_r by the close approximations $X_{r-\varepsilon} + b_{r-\varepsilon}(r - \varepsilon) + \sigma_{r-\varepsilon}(W_r - W_{r-\varepsilon})$ for some $\varepsilon = \varepsilon(u, n)$ whose distributions are Gaussian conditional on $\mathcal{F}_{r-\varepsilon}$. This idea is inspired by the one-step Euler approximation of Fournier and Printems (2008). For this σ needs to be non-degenerate and the approximation error has to be sufficiently small. We therefore work as in the last chapter under Assumption (SM-α-β). This time, however, $0 \le \alpha \le 1$ is arbitrary.

The right hand side in (3.1.6) shows that it is natural to assume that the Fourier transform of f is integrable, which leads to the the Fourier-Lebesgue spaces. They are introduced in Section A.4. If $f \in FL^s_{loc}(\mathbb{R}^d)$ for $s \ge 1$, then $f \in C^1(\mathbb{R}^d)$ such that (3.1.4) remains true. Moreover, for sufficiently smooth σ also the limit for $\Delta_n^{-1} D_{t,n}(f)$ in (3.1.5) remains valid. This yields the wanted CLT. For a concise statement we use the trapezoid rule estimator from the last section.

Theorem 3.1.3. *Assume (SM-α-β) for $0 \le \alpha, \beta \le 1$. Let $s > 2 - 2\alpha$, $s \ge 1$, $s + \beta > 1$. Then we have for $f \in FL^s_{loc}(\mathbb{R}^d)$ the stable convergence*

$$\Delta_n^{-1}\left(\Gamma_t\left(f\right) - \widehat{\Theta}_{n,t}\left(f\right)\right) \xrightarrow{st} \frac{1}{\sqrt{12}} \int_0^t \left\langle \nabla f\left(X_r\right), \sigma_r d\widetilde{W}_r\right\rangle$$

as processes on $\mathcal{D}([0,T], \mathbb{R}^d)$, where $\widetilde{W}$ is a d-dimensional Brownian motion defined on an independent extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$. The feasible central limit theorem of Remark 3.1.2 remains valid.

This result is remarkable since it is only based on regularity assumptions for f and σ. In particular, if σ is smoother, then the conditions on f can be relaxed. For $\alpha > 1/2$, $f \in FL^1_{loc}(\mathbb{R}^d)$ is allowed. For $\alpha \le 1/2$ there is a trade-off between the regularities of f and σ. The theorem also extends to L^2-Sobolev functions for sufficiently large regularity, because $H^s_{loc}(\mathbb{R}^d) \subset FL^{s'}_{loc}(\mathbb{R}^d)$, if $s > s' + d/2$ (cf. Proposition A.4.2).

Remark 3.1.4. As the proof of Theorem 3.1.3 reveals, it is not possible to argue as in Section 3.1.1 using a generalized Itô formula for $f \in C^1(\mathbb{R}^d)$.

3.1.3. CLT for L^2-Sobolev functions

The proof of Theorem 3.1.3 does not apply to all $C^1(\mathbb{R}^d)$-functions. The weak limit, however, is also well-defined for $f \in H^1_{loc}(\mathbb{R}^d)$. A minor issue in this case is that the random variables $f(X_r)$ depend on the version of f that we choose in its equivalence class in $L^2_{loc}(\mathbb{R}^d)$. This problem disappears if f is continuous or if X_r has a density. Note that $H^1(\mathbb{R}^d) \subset C(\mathbb{R}^d)$ only for $d = 1$. Interestingly, it can be shown by the methods of Debussche and Romito (2014, Section 5), which are also inspired by Fournier and Printems (2008), under Assumption (SM-α-β) that the marginals X_r have Lebesgue densities p_r for $r > 0$ (*).

In order to extend the central limit theorem to $f \in H^1_{loc}(\mathbb{R}^d)$, it turns out that we need to make the following stronger assumption.

Assumption (X0). *X_0 is independent of $(X_t - X_0)_{0\leq t\leq T}$ and Lebesgue density μ. Either, $\mathcal{F}\mu \in L^1(\mathbb{R}^d)$ or $\mathcal{F}\mu$ is non-negative and μ is bounded.*

This assumption can be understood in two ways. First, the independence and the boundedness of μ imply that the marginals X_r have uniformly bounded Lebesgue densities (this follows without assuming the existence of the densities as motivated above). Second, f itself becomes more regular, as by independence $\mathbb{E}[\Gamma_t(f)|(X_r - X_0)_{0\leq r\leq t}] = \int_0^t (f * \tilde{\mu})(X_r - X_0)dr$ with $\tilde{\mu}(x) = \mu(-x)$. Unfortunately, this property can not be used directly in the proof.

We can show under this assumption that (3.1.4) remains true for $f \in H^1_{loc}(\mathbb{R}^d)$. Moreover, for $f \in H^s_{loc}(\mathbb{R}^d)$ and sufficiently large $s \geq 1$ we can prove that $\Delta_n^{-1} D_{n,T}(f)$ converges to (3.1.5) in probability. This convergence is not uniform in $0 \leq t \leq T$ anymore. Therefore the weak convergence is not functional and holds only at the fixed time T.

Theorem 3.1.5. *Assume (SM-α-β) for $0 \leq \alpha, \beta \leq 1$ and (X0). Let $s > 2 - 2\alpha$, $s \geq 1$, $s + \beta > 1$. Then we have for $f \in H^s_{loc}(\mathbb{R}^d)$ the stable convergence*

$$\Delta_n^{-1}\left(\Gamma_T(f) - \widehat{\Theta}_{n,T}(f)\right) \xrightarrow{st} \frac{1}{\sqrt{12}}\int_0^T \left\langle \nabla f(X_r), \sigma_r d\widetilde{W}_r\right\rangle,$$

where $\widetilde{W}$ is a d-dimensional Brownian motion defined on an independent extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$. The feasible central limit theorem of Remark 3.1.2 remains valid.

Because of independence, Assumption (X0) can be relaxed by randomizing the initial condition and then using a coupling argument. This yields the following corollary.

Corollary 3.1.6. *Assume (SM-α-β) for $0 \leq \alpha, \beta \leq 1$. Let $s > 2 - 2\alpha$, $s \geq 1$, $s + \beta > 1$. For any function $f \in H^s_{loc}(\mathbb{R}^d)$ there exists a set $E \subset \mathbb{R}^d$ such that $\mathbb{R}^d \backslash E$ has Lebesgue measure 0 and such that the stable convergence in Theorem 3.1.5 holds for all $X_0 = x_0 \in E$.*

This result generalizes Theorem 3.1.1 considerably. The set E depends in general on the function f, i.e. it can change if we consider a different function $\tilde{f}$ with $f = \tilde{f}$ almost everywhere.

Remark 3.1.7. In some cases it is possible to derive similar CLTs for $f \in H^s_{loc}(\mathbb{R}^d)$ with $0 \leq s < 1$. For example, we have $f = \mathbf{1}_{[a,\infty)} \in H^{1/2-}_{loc}(\mathbb{R})$ and the proof of Theorem

*This will be studied more explicitly in the forthcoming paper of Romito (2017).

3.1.5 implies a CLT for $\Delta_n^{-3/4}(\Gamma_T(f_\varepsilon)-\widehat{\Gamma}_{n,T}(f_\varepsilon))$, where $f_\varepsilon = f * \varphi_\varepsilon$ with $\varphi \in C_c^\infty(\mathbb{R}^d)$, $\varphi_\varepsilon = \varepsilon^{-1}\varphi(\varepsilon^{-1}(\cdot))$ and $\varepsilon = \Delta_n^{1/2}$. The limiting distribution is similar to Corollary 3.4 of Ngo and Ogawa (2011), which applies to Brownian motion only, and involves local times of X. The rate $\Delta_n^{3/4}$ will be explained in the next section. This proof does not yield a CLT for $\Delta_n^{-3/4}(\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f))$, as the error $\Gamma_T(f-f_\varepsilon)-\widehat{\Gamma}_{n,T}(f-f_\varepsilon)$ is only of order $O_{\mathbb{P}}(\Delta_n^{3/4})$.

3.2. Upper bounds for less smooth functions

The aim of this section is to derive finite sample upper bounds on $\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}$ with explicit dependence on Δ_n, T and f. The function f is possibly much rougher than in the last section. It is therefore not possible to use arguments based on Taylor's theorem such as Itô's formula. Except for special cases, it is impossible to prove central limit theorems for $\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)$ in this case (cf. Remark 3.1.7). Instead of using martingale arguments, the results here are based on direct calculations with respect to the distribution of X. The following is inspired by the proof of Ganychenko (2015, Theorem 1).

We always assume that $X=(X_t)_{0\le t\le T}$ is a càdlàg process with respect to $(\Omega,\mathcal{F},(\mathcal{F})_{0\le t\le T},\mathbb{P})$, not necessarily a semimartingale or a Markov process. Then

$$\begin{aligned}&\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2\\&=\sum_{k,j=1}^n\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\mathbb{E}\Big[\left(f(X_r)-f\left(X_{t_{k-1}}\right)\right)\left(f(X_h)-f\left(X_{t_{j-1}}\right)\right)\Big]dhdr.\end{aligned}$$

Assume that the bivariate distributions of (X_a,X_b), $a<b$, have Lebesgue densities $p_{a,b}$. Under suitable regularity assumptions the expectation in the last display can be written as

$$\begin{aligned}&\int_{t_{k-1}}^r\left(\int f(x)f(y)\left(\partial_b p_{h,b}(x,y)-\partial_b p_{t_{j-1},b}(x,y)\right)d(x,y)\right)db\\&=\int_{t_{k-1}}^r\int_{t_{j-1}}^h\left(\int f(x)f(y)\,\partial_{ab}^2 p_{a,b}(x,y)\,d(x,y)\right)dadb.\end{aligned}\tag{3.2.1}$$

From this we can obtain general upper bounds on $\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2$. Their structure reflects that the distributions of (X_a,X_b) degenerate for $a=b$, therefore requiring a different argument.

Proposition 3.2.1. *Assume that the joint densities $p_{a,b}$ of (X_a,X_b) exist for all $0<a<b\le T$.*

(i) Assume that $b\mapsto p_{a,b}(x,y)$ is differentiable for all $x,y\in\mathbb{R}^d$, $0<a<b<T$ with $\partial_b p_{a,b}\in L^\infty_{loc}(\mathbb{R}^{2d})$. Then there exists a constant C such that for all bounded f with

compact support

$$\begin{aligned}
&\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P})} \\
&\leq C\Delta_n \int (f(y) - f(x))^2 \Bigg(\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} p_{t_{k-1},r}(x,y)\, dr \\
&\quad + \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \Big(|\partial_r p_{h,r}(x,y)| + |\partial_r p_{t_{j-1},r}(x,y)| \Big) dhdr \Bigg) d(x,y).
\end{aligned}$$

(ii) *In addition, assume that $a \mapsto \partial_b p_{a,b}(x,y)$ is differentiable for all $x, y \in \mathbb{R}^d$, $0 < a < b < T$ with $\partial^2_{ab} p_{a,b} \in L^\infty_{loc}(\mathbb{R}^{2d})$. Then we also have*

$$\begin{aligned}
&\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P})} \\
&\leq C\Delta_n^2 \int (f(y) - f(x))^2 \Bigg(\Delta_n^{-1} \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} p_{t_{k-1},r}(x,y)\, dr \\
&\quad + \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left| \partial^2_{hr} p_{h,r}(x,y) \right| dhdr \Bigg) d(x,y).
\end{aligned}$$

Concrete upper bounds can be obtained from this by combining the smoothness of f with bounds on $\partial_b p_{a,b}$ and $\partial^2_{ab} p_{a,b}$. Another way for getting upper bounds comes from formally applying the Plancherel theorem to (3.2.1). Denote by $\varphi_{a,b} = \mathcal{F} p_{a,b}$ the characteristic function of (X_a, X_b). Under sufficient regularity conditions (3.2.1) is equal to

$$(2\pi)^{-2d} \int_{t_{k-1}}^{r} \int_{t_{j-1}}^{h} \left(\int \mathcal{F}f(u)\, \mathcal{F}f(v)\, \overline{\partial^2_{ab} \varphi_{a,b}(u,v)} d(u,v) \right) dadb.$$

This yields the following version of the last proposition.

Proposition 3.2.2. *Let $\varphi_{a,b}$ be the characteristic functions of (X_a, X_b) for $0 \leq a, b \leq T$ with $\varphi_{a,a}(u,v) = \varphi_a(u+v)$ for $u, v \in \mathbb{R}^d$.*

(i) *Assume that $b \mapsto \varphi_{a,b}(u,v)$ is differentiable for $u, v \in \mathbb{R}^d$, $0 < a < b < T$ with $\partial_b \varphi_{a,b} \in L^\infty_{loc}(\mathbb{R}^{2d})$. Then there exists a constant C such that for all $f \in L^1(\mathbb{R}^d)$ with $\mathcal{F}f \in C_c^\infty(\mathbb{R}^d)$*

$$\begin{aligned}
&\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P})} \\
&\leq C\Delta_n \int |\mathcal{F}f(u)|\, |\mathcal{F}f(v)| \Bigg(\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} g_{t_{k-1},r}(u,v) dr \\
&\quad + \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left(|\partial_r \varphi_{h,r}(u,v)| + \left|\partial_r \varphi_{t_{j-1},r}(u,v)\right| \right) dhdr \Bigg) d(u,v),
\end{aligned}$$

with $g_{t_{k-1},r}(u,v) = |\varphi_{r,r}(u,v)| + |\varphi_{t_{k-1},r}(u,v)| + |\varphi_{t_{k-1},t_{k-1}}(u,v)|$.

(ii) In addition, assume that $a \mapsto \partial_b \varphi_{a,b}(u,v)$ *is differentiable for all* $u, v \in \mathbb{R}^d$, $0 < a < b < T$ *with* $\partial^2_{ab}\varphi_{a,b} \in L^\infty_{loc}(\mathbb{R}^{2d})$. *Then we also have*

$$\begin{aligned}&\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P})}\\ &\leq C\Delta_n^2 \int |\mathcal{F}f(u)|\,|\mathcal{F}f(v)| \Bigg(\Delta_n^{-1}\sum_{k=1}^n \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t_k} (|\partial_r \varphi_{h,r}(u,v)|\\ &+ \left|\partial_r \varphi_{t_{k-1},r}(u,v)\right|) dhdr + \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j} \left|\partial^2_{hr}\varphi_{h,r}(u,v)\right| dhdr\Bigg) d(u,v).\end{aligned}$$

The second proposition is useful if the characteristic functions $\varphi_{a,b}$ are explicitly known, while the densities $p_{a,b}$ are not. This is true for many Lévy or affine processes. Moreover, it can be easier to find upper bounds on characteristic functions than for the respective densities. Note that the second proposition does not require the joint densities $p_{a,b}$ to exist. This is relevant, for instance, when studying jump processes without marginal densities (cf. Example 3.2.12). In some cases both propositions apply and the results can differ as we will see in the next section.

We will now study several concrete examples of processes X and function spaces for f and derive explicit upper bounds.

3.2.1. Markov processes

Let X be a continuous-time Markov process on $\mathbb{R}^d$ with transition densities $\xi_{h,r}$, $0 \leq h < r \leq T$, such that $\mathbb{E}[g(X_r)|X_h = x] = \int g(y)\xi_{h,r}(x,y)dy$ for $x \in \mathbb{R}^d$ and all continuous, bounded functions g. Denote by $\mathbb{P}_{x_0}$ the law of X conditional on $X_0 = x_0$. The joint density of (X_h, X_r), conditional on $X_0 = x_0$, is $p_{h,r}(x,y;x_0) = \xi_{0,h}(x_0,x)\xi_{h,r}(x,y)$. The necessary differentiability conditions on $p_{h,r}$ from Proposition 3.2.1 translate to assumptions on $\xi_{h,r}$. The following heat kernel bounds are similar to the ones in Ganychenko (2015).

Assumption 3.2.3. *The transition densities* $\xi_{h,r}$ *for* $0 \leq h < r < T$ *satisfy one of the following conditions:*

(*A*) *The function* $r \mapsto \xi_{h,r}(x,y)$ *is continuously differentiable for all* $x, y \in \mathbb{R}^d$ *and there exist probability densities* q_r *on* $\mathbb{R}^d$ *satisfying for some constant* $C > 0$

$$\sup_{x,y\in\mathbb{R}^d} \frac{|\xi_{h,r}(x,y)|}{q_{r-h}(y-x)} \leq C, \quad \sup_{x,y\in\mathbb{R}^d} \frac{|\partial_r \xi_{h,r}(x,y)|}{q_{h-r}(y-x)} \leq \frac{C}{h-r}. \tag{3.2.2}$$

(*B*-γ) *Let* $0 < \gamma \leq 2$. *In addition to (A), the function* $h \mapsto \partial_r \xi_{h,r}(x,y)$ *is continuously differentiable for all* $x, y \in \mathbb{R}^d$ *and the* q_h *satisfy*

$$\sup_{x,y\in\mathbb{R}^d} \frac{\left|\partial^2_{hr}\xi_{h,r}(x,y)\right|}{q_{r-h}(y-x)} \leq \frac{C}{(r-h)^2}. \tag{3.2.3}$$

Moreover, if $\gamma < 2$, *then* $\sup_{x\in\mathbb{R}^d}\left(\|x\|^{2s+d} q_h(x)\right) \lesssim h^{2s/\gamma}$ *for* $0 < s \leq \gamma/2$, *while, if* $\gamma = 2$, *then* $\int \|x\|^{2s} q_h(x)dx \lesssim h^s$ *for* $0 < s \leq 1$.

These conditions are satisfied in case of elliptic diffusions with Hölder continuous coefficients with $q_h(x) = c_1 h^{-d/2} e^{-c_2\|xh^{-1/2}\|^2}$ and $\gamma = 2$ for some constants c_1, c_2. They are also satisfied for many Lévy driven diffusions with $q_h(x) = c_1 h^{-d/\gamma}(1 + \|xh^{-1/\gamma}\|^{\gamma+d})^{-1}$ and $0 < \gamma < 2$ (Ganychenko et al. (2015)). Different upper bounds in (3.2.2), (3.2.3) are possible yielding different results below.

Based on Proposition 3.2.1 we recover the main results of Ganychenko (2015) and Ganychenko et al. (2015). For $0 \leq s \leq 1$ denote by $\|f\|_{C^s}$ the Hölder seminorm $\sup_{x\neq y} \frac{\|f(x)-f(y)\|}{\|x-y\|^s}$.

Theorem 3.2.4. *Let $n \geq 2$ and $x_0 \in \mathbb{R}^d$. Let X be a Markov process with transition densities $\xi_{a,b}$.*

(i) Assume (A). There exists a constant C such that for every bounded f

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P}_{x_0})} \leq C\|f\|_\infty T^{1/2} \Delta_n^{1/2} (\log n)^{1/2}.$$

(ii) Assume (B-γ) for $0 < \gamma \leq 2$. There exists a constant C such that for $f \in C^s(\mathbb{R}^d)$ with $0 \leq s \leq \gamma/2$

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P}_{x_0})} \leq C\|f\|_{C^s} T^{1/2} \begin{cases} \Delta_n^{\frac{1+2s/\gamma}{2}}, & 2s/\gamma < 1, \\ \Delta_n (\log n)^{1/2}, & 2s/\gamma = 1. \end{cases}$$

Up to log factors the rate of convergence (for fixed T) is $\Delta_n^{(1+2s/\gamma)/2}$ for $f \in C^s(\mathbb{R}^d)$, interpolating between the worst-case rates $\Delta_n^{1/2}$ and the "best" rate Δ_n. Interestingly, smaller γ means faster convergence for the same smoothness s.

Remark 3.2.5. The $T^{1/2}$-term in the upper bound is optimal and appears in almost every other example below (observe however Theorem 3.2.13). If X is ergodic with invariant measure μ, then this can be used to estimate functionals $\int f d\mu$ with respect to μ by the estimator $T^{-1}\widehat{\Gamma}_{n,T}(f)$ with optimal rate $T^{-1/2}$, independent of any condition on the discretization order Δ_n, i.e. there is essentially no difference between the high and the low frequency setting (cf. Theorem 3.3.4).

Theorem 3.2.4 yields for the bounded function $f = \mathbf{1}_{[a,b]}$, $a < b$, only the rate $\Delta_n^{1/2}(\log n)^{1/2}$. This cannot explain the $\Delta_n^{3/4}$-rate obtained for Brownian motion in Ngo and Ogawa (2011). In order to find a unifying view consider now $f \in H^s(\mathbb{R}^d)$, $0 \leq s \leq 1$ (cf. Section A.4).

Theorem 3.2.6. *Let X be a Markov process with transition densities $\xi_{a,b}$ and bounded initial density μ.*

(i) Assume (A). There exists a constant C such that for $f \in L^2(\mathbb{R}^d)$

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \leq C\|\mu\|_\infty^{1/2} \|f\|_{L^2} T^{1/2} \Delta_n^{1/2} (\log n)^{1/2}.$$

(ii) *Assume (B-γ) for $0<\gamma\leq 2$. There exists a constant C such that for $f\in H^s(\mathbb{R}^d)$ with $0\leq s\leq \gamma/2$*

$$\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}\leq C\|\mu\|_\infty^{1/2}\|f\|_{H^s}T^{1/2}\begin{cases}\Delta_n^{\frac{1+2s/\gamma}{2}}, \gamma<2, & 2s/\gamma<1,\\ \Delta_n^{\frac{1+s}{2}}(\log n)^{1/2}, \gamma=2, & 2s/\gamma<1,\\ \Delta_n(\log n)^{1/2}, & 2s/\gamma=1.\end{cases}$$

While the regularity of f is now measured in the L^2-Sobolev sense, we still obtain the interpolating rate $\Delta_n^{(1+2s/\gamma)/2}$ up to log factors. Since $C^s(K)\subset H^{s-}(\mathbb{R}^d)$ for compacts $K\subset\mathbb{R}^d$ and because $f=\mathbf{1}_{[a,b]}\in H^{1/2-}(\mathbb{R})$, this theorem also yields the rates $\Delta_n^{(1+2s/\gamma)/2-}$ for s-Hölder functions on compacts and $\Delta_n^{3/4-}$ (up to log factors) for indicators. By an explicit interpolation as in Theorems 3.3.9 and 3.3.10 this can be improved to $\Delta_n^{(1+2s/\gamma)/2}$ and $\Delta_n^{3/4}$, respectively. By considering L^2-Sobolev spaces we therefore unify the different rates obtained for Markov processes. The log factors in Theorem 3.2.6 can be removed in many cases (cf. Section 3.2.2).

Remark 3.2.7. *(i)* The role of μ in the proof of Theorem 3.2.6 is to ensure that the marginals have uniformly bounded densities p_h, i.e. $\sup_{0\leq h\leq T}\|p_h\|_\infty\leq\|\mu\|_\infty$. This is necessary, because the bounds in Assumption 3.2.3 degenerate at 0. Otherwise it is not even clear that $\|\Gamma_T(f)\|_{L^2(\mathbb{P})}<\infty$ for $f\in L^2(\mathbb{R}^d)$. If $\sup_{x\in\mathbb{R}^d}\int_0^T\xi_{0,r}(x)dr<\infty$, then the initial distribution can be arbitrary. This holds, for instance, when $d=1$ and $q_h(x)=c_1h^{-1/2}e^{-c_2\|xh^{-1/2}\|^2}$.

(ii) A different possibility for removing the initial condition is to wait until $T_0>0$ such that X_{T_0} has dispersed enough to have a bounded Lebesgue density. The proof of Theorem 3.2.6 can then be applied to $\|\int_{T_0}^T f(X_r)dr-\widehat{\Gamma}_{n,T_0,T}(f)\|_{L^2}$, where $\widehat{\Gamma}_{n,T_0,T}(f)$ is a Riemann-sum estimator taking only observations in $[T_0,T]$ into account.

(iii) A similar argument as in the proof of Corollary 3.1.6 shows $\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)=O_{\mathbb{P}_{x_0}}(a_n)$ for almost all initial conditions $X_0=x_0\in\mathbb{R}^d$, where a_n corresponds to the rates in Theorem 3.2.6 up to an additional log factor.

3.2.2. Additive processes

Let $Y=(Y_t)_{0\leq t\leq T}$ be an additive process on $\mathbb{R}^d$ with $Y_0=0$ and local characteristics (σ_t^2,F_t,b_t), where $\sigma^2=(\sigma_t^2)_{0\leq t\leq T}$ is a continuous $\mathbb{R}^{d\times d}$-valued function such that σ_t^2 is symmetric for all t, $b=(b_t)_{0\leq t\leq T}$ is a locally integrable $\mathbb{R}^d$-valued function and $(F_t)_{0\leq t\leq T}$ is a family of positive measures on $\mathbb{R}^d$ with $F_t(\{0\})=0$ and $\sup_{0\leq t\leq T}\{\int(\|x\|^2\wedge 1)dF_t(x)\}<\infty$ (cf. Tankov (2003, Section 14.1)). The increments Y_r-Y_h, $0\leq h<r\leq T$, are independent and have infinitely divisible distributions. In particular, the corresponding characteristic functions are $e^{\psi_{h,r}(u)}$, $u\in\mathbb{R}^d$, by the Lévy-Khintchine representation (Tankov (2003, Theorem 14.1)), where the characteristic exponents $\psi_{h,r}(u)$ are given by

$$i\int_h^r\langle u,b_t\rangle\,dt-\frac{1}{2}\int_h^r\|\sigma_t^\top u\|^2dt+\int_h^r\int\left(e^{i\langle u,x\rangle}-1-i\langle u,x\rangle\mathbf{1}_{\{\|x\|\leq 1\}}\right)dF_t(x)\,dt.$$

Applying Proposition 3.2.2 yields the following result. The independence in (X0) is always satisfied, because Y has independent increments.

Theorem 3.2.8. *Let $T \geq 1$. Consider the process $X_t = X_0 + Y_t$, where $Y = (Y_t)_{0\leq t\leq T}$ is an additive process with local characteristics (σ_t^2, F_t, b_t) as above and such that X_0 satisfies (X0).*

(i) Let $0 < \gamma \leq 2$ and assume that $|\partial_r \psi_{h,r}(v)| \leq c(1+\|v\|)^{\gamma+\beta_r}$ and $|e^{\psi_{h,r}(v)}| \leq ce^{-c\|v\|^{\gamma}(r-h)}$ for a constant c and all $0 \leq h < r \leq T$, $v \in \mathbb{R}^d$ and some $0 \leq \beta_r \leq \beta^* \leq \gamma/2$ with $0 < \gamma + \beta_r \leq 2$. Then there exists a constant C_μ such that for $f \in H^s(\mathbb{R}^d)$ with $\beta^*/2 \leq s \leq \gamma/2 + \beta^*$

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \leq C_\mu \|f\|_{H^s} T \Delta_n^{\frac{1}{2}+\frac{s-\beta^*/2}{\gamma-\beta^*}}.$$

If $\mathcal{F}\mu \in L^1(\mathbb{R}^d)$, then $C_\mu = C\|\mathcal{F}\mu\|_{L^1}^{1/2}$ and otherwise $C_\mu = C\|\mu\|_\infty^{1/2}$. If even $|\partial_r \psi_{h,r}(v)| \leq c\|v\|^{\gamma+\beta_r}$, then the same upper bound holds with $T^{1/2}$ instead of T.

(ii) If $|\partial_r \psi_{h,r}(v)| \leq c$, then we have for $f \in L^2(\mathbb{R}^d)$

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \leq C_\mu \|f\|_{L^2} T \Delta_n.$$

The same upper bound holds with $T^{1/2}$ instead of T, if $c_1 \leq \rho(v) \leq \partial_r \psi_{h,r}(v) \leq c_2 \rho(v) \leq 0$ for a bounded function $v \mapsto \rho(v)$ and constants $c_1 \leq c_2$.

By the comments before Remark 3.2.7 we can obtain from this upper bounds for Hölder and indicator functions. The condition $|\partial_r \psi_{h,r}(v)| \leq c(1+\|v\|)^{\gamma+\beta_r}$ gives an additional degree of freedom in order to account for time-inhomogeneity (cf. Example 3.2.11). Note that there are no log terms as compared to Theorem 3.2.6. The smaller $\gamma/2 + \beta^*$, the less smoothness is necessary for f to achieve a Δ_n rate.

Remark 3.2.9. In some situations it is sufficient to consider directly $X_t = Y_t$. This is true, for instance, if $d = 1$ and $\gamma > 1$ (cf. Remark 3.5.13). For different d or γ it follows in (i) that Y_{T_0} for any $T_0 > 0$ has a density p_{T_0} with $\mathcal{F}p_{T_0} \in L^1(\mathbb{R}^d)$. Similarly to Remark 3.2.7(ii) the proof of Theorem 3.2.8 can then be applied to $\|\int_{T_0}^T f(X_r)dr - \widehat{\Gamma}_{n,T_0,T}(f)\|_{L^2}$. For $O_{\mathbb{P}}$ bounds and almost all initial values $X_0 = x_0 \in \mathbb{R}^d$ refer to Remark 3.2.7(iii).

We study now a few examples.

Example 3.2.10 (*Non-vanishing Gaussian part*)**.** Assume that Y has local characteristics $(\sigma_t^2, F_t, 0)$ with $\sup_{0\leq r\leq T}\|(\sigma_r\sigma_r^\top)^{-1}\| < \infty$. Then $\gamma = 2$, $\beta^* = 0$ and $|\partial_r \psi_{h,r}(v)| \lesssim \|v\|^2$ (cf. Sato (1999, Equation (8.9))). Part (i) of Theorem 3.2.8 therefore yields up to a constant the upper bound $\|f\|_{H^s} T^{1/2} \Delta_n^{(1+s)/2}$ for $f \in H^s(\mathbb{R}^d)$ with $0 \leq s \leq 1$, thus improving on Theorem 3.2.6 by removing the log-factor.

Example 3.2.11 (γ*-stable processes*)**.** Let $\psi_{h,r}(v) = -c\int_h^r \|v\|^{\gamma+\beta_t} dt$ with c, γ, β_r as in Theorem 3.2.8. A process with these characteristic exponents exists if β is continuous. X is a generalized symmetric γ-stable process with stability index $\gamma + \beta_r$ changing in time. For $d = 1$ it is a multistable Lévy motion (cf. Example 4.1 in Falconer and Liu (2012)). If $\beta^* = 0$, then X is just a symmetric γ-stable process and Theorem 3.2.8 yields the upper bound $\|f\|_{H^s} T^{1/2} \Delta_n^{1/2+s/\gamma}$ for $f \in H^s(\mathbb{R}^d)$ and $0 \leq s \leq \gamma/2$. Again, this removes the log-factor of Theorem 3.2.6.

Example 3.2.12 (*Compound Poisson process*)**.** Let X be a compound Poisson process. Then $\psi_{h,r}(v) = (r-h)\int(e^{i\langle v,x\rangle}-1)dF(x)$ for all $0 \leq h < r \leq T$ and a finite measure F. Observe that the marginals X_r do not have Lebesgue densities unless X_0 does. Since $\rho(v) := \partial_r \psi_{h,r}(v) = \int(e^{i\langle v,x\rangle}-1)dF(x)$ is bounded in v, part (ii) of the theorem yields the upper bound $\|f\|_{L^2} T\Delta_n$ for *all* $f \in L^2(\mathbb{R}^d)$. The improved bound applies, if F is symmetric (cf. Section 3.3.1).

3.2.3. Fractional Brownian motion

Let $B = (B_t)_{0\leq t\leq T}$ be a fractional Brownian motion in $\mathbb{R}^d$ with Hurst index $0 < H < 1$. This means that the d component processes $(B_t^{(m)})_{0\leq t\leq T}$ for $m = 1,\dots,d$ are independent and centered Gaussian processes with covariance function $c(h,r) := \mathbb{E}[B_h^{(m)}B_r^{(m)}] = \frac{1}{2}(r^{2H}+h^{2H}-(r-h)^{2H})$, $0 \leq h \leq r \leq T$. If $H = 1/2$, then B is just a Brownian motion. For $H \neq 1/2$ it is an important example of a non-Markovian process which is also not a semimartingale.

Theorem 3.2.13. *Let $T \geq 1$, $n \geq 2$. Consider the process $X_t = X_0 + B_t$, where $(B_t)_{0\leq t\leq T}$ is a fractional Brownian motion with Hurst index $0 < H < 1$ and where X_0 satisfies (X0). Then there exists a constant C_μ as in Theorem 3.2.8 such that for any $f \in H^s(\mathbb{R}^d)$ and $0 \leq s \leq 1$*

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \leq C_\mu \|f\|_{H^s} \begin{cases} T^H \Delta_n^{\frac{1+s}{2}}, & H \geq 1/2, \\ T^{1/2}\Delta_n^{\frac{1+2sH}{2}}, & H < 1/2. \end{cases}$$

Again, from this we can obtain upper bounds for Hölder and indicator functions (cf. comments before Remark 3.2.7). It is interesting that the rate remains unchanged but the dependency on T differs for $H > 1/2$, while this effect is reversed for $H < 1/2$. The dependency on H is optimal. Indeed, if f is the identity, then for some constant C

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \geq C \begin{cases} T^H \Delta_n, & H > 1/2, \\ T^{1/2}\Delta_n^{\frac{1+2H}{2}}, & H < 1/2. \end{cases} \tag{3.2.4}$$

Remark 3.2.9 applies here as well in order to relax the assumption on X_0. In particular, we can directly consider $X_t = B_t$ if $d = 1$. Comparing the theorem (at least for $H < 1/2$) to Example 3.2.11 suggests that there is a more general result for self-similar processes with self-similarity index α and upper bound $\|f\|_{H^s} T^{1/2}\Delta_n^{1/2+\alpha s}$.

The key idea in the proof is that fractional Brownian motion is locally nondeterministic. There are many more processes (and random fields) with this property. In principle, the proof of the theorem will apply in these cases as well, as long as the time derivatives of $\Phi_{h,r}(u,v)$ can be controlled. This holds, for instance, for multifractional Brownian motion with time varying Hurst index $H(t)$ (cf. Boufoussi et al. (2007)) and stochastic differential equations driven by fractional Brownian motion (cf. Lou and Ouyang (2017))

We will now apply Theorem 3.2.13 to approximate local times from discrete data. Let $d = 1$ and let $(L_T^a)_{a\in\mathbb{R}}$ be the family of *local times* of B until T which satisfies the occupation time formula $\int_0^T g(B_r)dr = \int_{\mathbb{R}} g(x)L_T^x dx$ for every continuous and bounded function g (cf. Nualart (1995, Chapter 5)). We can write $L_T^a = \delta_a(L_T)$ for $a \in \mathbb{R}$, where δ_a is the Dirac delta function. Note that $\delta_a \in H^{-1/2-}(\mathbb{R})$ has negative regularity. Theorem 3.2.13 therefore suggests the rate $T^{1/2}\Delta_n^{1/4}$ (for $H = 1/2$). This turns out to be almost correct.

Theorem 3.2.14. *Let $T \geq 1$, $n \geq 2$, $d = 1$. Let $X_t = B_t$, where $(B_t)_{0\leq t\leq T}$ is a fractional Brownian motion with Hurst index $0 < H < 1$. Consider $f_{a,\varepsilon}(x) = (2\varepsilon)^{-1}\mathbf{1}_{(a-\varepsilon,a+\varepsilon)}(x)$ for $x, a \in \mathbb{R}$ and $\varepsilon = \Delta_n^{\alpha_H}$ with $\alpha_H = \frac{3}{2}\cdot\frac{H}{1+H} - \rho$ when $H \geq 1/2$ and $\alpha_H = H - \rho$ when $H < 1/2$ for any small $\rho > 0$. Then we have for some constant C, independent of a, that*

$$\|L_T^a - \widehat{\Gamma}_{n,T}(f_{a,\varepsilon})\|_{L^2(\mathbb{P})} \leq C \begin{cases} T^H \Delta_n^{\frac{3}{4}\cdot\frac{1-H}{1+H}-\rho}, & H \geq 1/2, \\ T^{1/2}\Delta_n^{\frac{1-H}{2}-\rho}, & H < 1/2. \end{cases}$$

This considerably generalizes Theorem 2.6 of Kohatsu-Higa et al. (2014), which applies only to Brownian motion. For H close to 1 the rate becomes arbitrarily slow, because the paths of B are almost differentiable and the occupation measure becomes more and more singular with respect to the Lebesgue measure.

3.3. Upper bounds for stationary Markov processes

While we obtained in the last section $L^2(\mathbb{P})$-upper bounds on $\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)$ for general càdlàg processes, this section considers the case of a stationary Markov process X. On the one hand, this leads to very elegant proofs, based on calculus for Markov semigroup and its infinitesimal generator with respect to X. On the other hand, the upper bounds apply to more general functions f and are even better in some cases compared to the last section by removing log terms. Moreover, a stationary Markov process is not necessarily $\mathbb{R}^d$-valued or has to satisfy any of the assumptions of the various examples studied in the last section. We will recover some of the previous results, but based on the methods in this section.

In the following, X is a continuous-time Markov process on $[0, T]$ with values in a Polish space $\mathcal{S}$. For any measure μ on $\mathcal{S}$ denote by $L^2(\mu) := L^2(\mathcal{S}, \mu)$ the space of square integrable functions $f : \mathcal{S} \to \mathbb{R}$ with respect to μ and with norm $\|f\|_\mu = (\int f^2 d\mu)^{1/2}$. $\|\cdot\|_{\infty,\mu}$ denotes the sup-norm in $L^\infty(\mu)$. For the basic concepts of semigroup theory and functional calculus refer to Section A.5.

Our main assumptions are the following:

Assumption 3.3.1. *X is a stationary time-homogeneous Markov process with stationary measure μ. The associated semigroup $(P_r)_{r\geq 0}$ is Feller and its infinitesimal generator L with respect to $L^2(\mu)$ is a normal operator.*

These assumptions are satisfied for many important processes. The leading example is the standard Ornstein-Uhlenbeck process in $\mathbb{R}^d$. Note that for $f \in L^2(\mu)$ both $\Gamma_T(f)$ and $\widehat{\Gamma}_{n,T}(f)$ are μ-almost surely well-defined random variables in $L^2(\mathbb{P})$ (cf. Section 3.1.3). Consider the operators $|L|^{s/2}$, $s \geq 0$, which are defined via the functional calculus of L. They have domains $\mathrm{dom}(|L|^{s/2}) \subset L^2(\mu)$ and thus contain all $f \in L^2(\mu)$ with $\||L|^{s/2}f\|_\mu < \infty$. If X is an Ornstein-Uhlenbeck process, then the related spaces

$\mathrm{dom}((I-L)^{s/2}) \subset \mathrm{dom}(|L|^{s/2})$ are known as Bessel potential spaces and play an important role in Malliavin calculus (Watanabe (1984)). We are now ready to state the general upper bound.

Theorem 3.3.2. *Let X be a Markov process satisfying Assumption 3.3.1 with $X_0 \overset{d}{\sim} \mu$. There exists a universal constant C such that for all $f \in \mathrm{dom}(|L|^{s/2})$, $0 \leq s \leq 1$,*

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \leq C\||L|^{s/2}f\|_\mu T^{1/2}\Delta_n^{\frac{1+s}{2}}.$$

The proof of this theorem is remarkably short. For $s=0$ it follows that $\mathrm{dom}(|L|^0) = L^2(\mu)$ and the rate is $T^{1/2}\Delta_n^{1/2}$. Since $\mathrm{dom}(|L|^{s/2}) \subset \mathrm{dom}(|L|^{1/2})$ for $s \geq 1$, the rate is never better than $T^{1/2}\Delta_n$. For $0 < s < 1$ the bound interpolates between the two extreme cases. Note that there are no log terms as in Theorems 3.2.4 or 3.2.6.

A deeper understanding of the spaces $\mathrm{dom}(|L|^{s/2})$ requires more explicit knowledge about the generator. For example, if L is self-adjoint, then $|L| = -L$ and thus $\||L|^{1/2}f\|_\mu^2 = \|(-L)^{1/2}f\|_\mu^2 = \langle -Lf, f\rangle_\mu$. This is the Dirichlet form associated with L and μ. It is typically easier to analyze than studying $\mathrm{dom}(|L|^{1/2})$ directly in terms of the functional calculus. Examples are diffusions on $\mathbb{R}^d$ such that for sufficiently smooth functions f the Dirichlet form is bounded by $\int \|\nabla f(x)\|^2 d\mu(x)$. This immediately leads to upper bounds for Hölder and indicator functions. Up to some additional conditions, we will show that α-Hölder functions lie in $\mathrm{dom}(|L|^{\alpha/2})$ and indicator functions of certain cylinder sets of $\mathbb{R}^d$ lie in $\mathrm{dom}(|L|^{1/4})$, implying the rates $\Delta_n^{(1+\alpha)/2}$ and $\Delta_n^{3/4}$, respectively. This explains the different rates obtained in the literature similar to the arguments before Remark 3.2.7 (see also Remark 3.3.11). An important difference in case of $\mathcal{S} \subset \mathbb{R}^d$ is that the considered functions have to lie only in $L^2(\mu)$ and not in $L^2(\mathbb{R}^d)$. For instance, bounded functions are in $L^2(\mu)$, but not in $L^2(\mathbb{R}^d)$.

The assumption of starting in the stationary distribution can be relaxed to some extent. Indeed, if the initial distribution is absolutely continuous with respect to μ, then the result of Theorem 3.3.2 remains valid. More generally, if the distribution of X_{T_0}, $T_0 \geq 0$, is absolutely continuous with respect to μ, then the result still applies, if instead of $\Gamma_T(f)$ the alternative occupation time functional $\Gamma_{T_n,T}(f) = \int_{T_n}^T f(X_r)dr$ is estimated by $\widehat{\Gamma}_{n,T_n,T}(f) = \Delta_n \sum_{k=T_n\Delta_n^{-1}+1}^n f(X_{t_{k-1}})$, where $T_n = \lceil T_0/\Delta_n^{-1}\rceil\Delta_n$. Clearly, $\Gamma_T(f) = \Gamma_{0,T}(f)$ and $\widehat{\Gamma}_{n,T}(f) = \widehat{\Gamma}_{n,0,T}(f)$. This yields the following corollary.

Corollary 3.3.3. *Let X be a Markov process satisfying Assumption 3.3.1. Assume that $X_{T_0} \overset{d}{\sim} \eta$, $T_0 \geq 0$, for a probability measure η such that $\eta \ll \mu$ with density $d\eta/d\mu$. There exists a universal constant C such that for all $f \in \mathrm{dom}(|L|^{s/2})$, $0 \leq s \leq 1$,*

$$\left\|\Gamma_{T_n,T}(f) - \widehat{\Gamma}_{n,T_n,T}(f)\right\|_{L^2(\mathbb{P})} \leq C\left\|\frac{d\eta}{d\mu}\right\|_{\infty,\mu}^{1/2}\||L|^{s/2}f\|_\mu T^{1/2}\Delta_n^{\frac{1+s}{2}}.$$

As an example consider the Ornstein-Uhlenbeck process which is a Gaussian process and therefore every X_{T_0} for $T_0 > 0$ is normally distributed such that the distribution of X_{T_0} is absolutely continuous with respect to the stationary measure of X (cf. Example 3.3.5).

Consider now the situation of Remark 3.2.5. With respect to ergodicity, it is well-known that $T^{-1}\Gamma_T(f)$ is $T^{1/2}$-consistent for $\int f d\mu$, i.e. $T^{-1}\Gamma_T(f) - \int f d\mu = O_{\mathbb{P}}(T^{-1/2})$, when L is self-adjoint and $f \in \mathrm{dom}((-L)^{-1/2})$ (see e.g. Kipnis and Varadhan (1986, Theorem 1.8)). By Theorem 3.3.2 this can be extended to the estimator $T^{-1}\widehat{\Gamma}_{n,T}(f)$ and more general $L^2(\mu)$-functions (cf. Remark 3.2.5).

Theorem 3.3.4. *Let X be a Markov process satisfying Assumption 3.3.1 with $X_0 \overset{d}{\sim} \mu$. There exists a universal constant C such that for all $f \in L^2(\mu)$ with $f_0 \in dom(|L|^{-1/2})$, $f_0 = f - \int f d\mu$,*

$$\left\| T^{-1}\widehat{\Gamma}_{n,T}(f) - \int_{\mathcal{S}} f(x)\, d\mu(x) \right\|_{L^2(\mathbb{P})} \leq CT^{-1/2}\left(\|f\|_\mu \Delta_n^{1/2} + \left\| |L|^{-1/2} f_0 \right\|_\mu \right).$$

Using Corollary 3.3.3 the assumption of starting in the stationary distribution can be relaxed. As an example for $\mathrm{dom}(|L|^{-1/2})$ being non-trivial, assume that 0 is a simple eigenvalue of L and that L has a spectral gap, i.e. $s_0 > 0$, where $s_0 = \sup\{r > 0 : B(0,r) \cap \sigma(L) = \{0\}\}$ and $B(0,r) = \{z \in \mathbb{C} : |z| \leq r\}$. In that case, X is ergodic and it can be shown that $f_0 \in \mathrm{dom}(|L|^{-1/2})$ is satisfied whenever f is non-constant (Bakry et al. (2013, Section 4.2.1)). Furthermore, the upper bound of the theorem simplifies, since

$$\left\| |L|^{-1/2} f_0 \right\|_\mu \leq s_0^{-1/2} \|f_0\|_\mu \leq s_0^{-1/2} \|f\|_\mu.$$

A concrete example of a process with spectral gap is the Ornstein-Uhlenbeck process (Bakry et al. (2013, Chapter 4)). Theorem 3.3.4 shows that in order to achieve the rate $T^{1/2}$ as $n, T \to \infty$ there is essentially no gain in the high-frequency case, i.e. $\Delta_n \to 0$, compared to the low-frequency case with Δ_n fixed. The error bound improves on the commonly used condition in the literature that $T\Delta_n \lesssim 1$ to achieve $T^{1/2}$-consistency (see e.g. Dion and Genon-Catalot (2016, Section 5)). It is interesting to note that Theorem 3.3.4 depends on negative powers of $|L|$, while Theorem 3.3.2 depends on positive powers of $|L|$.

Next, we apply the general bound from Theorem 3.3.2 to several important examples. We first study Markov jump processes, i.e. continuous time Markov processes with countable state spaces. Then a special class of diffusion processes is considered for which the spaces $\mathrm{dom}(|L|^{s/2})$ can be described via the Dirichlet form $\langle -Lf, f \rangle_\mu$. After that, we show for the example of Brownian motion how the assumption of stationarity can be removed. Finally, we show that the method also applies to infinite dimensional diffusions.

3.3.1. Markov-jump processes

Consider a continuous-time Markov process $(X_r)_{r\geq 0}$ on a countable state space $\mathcal{S}$. Such a process can always be realized as $X_r = Y_{N_r}$ for a Markov chain $(Y_s)_{s\in\mathcal{S}}$ starting in some initial distribution μ with transition probabilities $(P_{xy})_{x,y\in S}$ and an independent Poisson process $(N_r)_{r\geq 0}$ with intensity $0 < \lambda < \infty$ (Ethier and Kurtz (1986, Section 4.2)). Observing a path of X at the discrete times $0, \Delta_n, 2\Delta_n, ..., (n-1)\Delta_n$, the jump times can be identified with Δ_n precision. Hence, if the function f is bounded, then every jump contributes at most $2\|f\|_\infty$ to the estimation error $|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)|$. This yields the bound

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \leq 2\|f\|_\infty \mathbb{E}\left[N_T^2\right]^{1/2} \Delta_n = 2\|f\|_\infty \left(\lambda T + (\lambda T)^2\right)^{1/2} \Delta_n.$$

This gives the optimal rate Δ_n but requires the function f to be bounded. Moreover, the error grows linearly in T as opposed to $T^{1/2}$ in Theorem 3.3.2. This can be improved, if X is stationary with stationary measure μ and reversible, i.e. $P^\top = P$, where $P^\top$ is the transpose of P. Then the infinitesimal generator $L = \lambda(P - I)$ is a bounded, non-negative self-adjoint operator. Therefore, $\|(-L)^{1/2} f\|_\mu \leq \|(-L)^{1/2}\|\|f\|_\mu \leq \lambda^{1/2}\|f\|_\mu$ with operator norm $\|(-L)^{1/2}\|$. It follows that $\mathrm{dom}((-L)^{1/2}) = L^2(\mu)$ and Theorem 3.3.2 implies

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \leq C\lambda^{1/2}\|f\|_\mu T^{1/2}\Delta_n.$$

(cf. Example 3.2.12). Note that the results of Section 3.2.1 do not apply here, because the state space is countable and therefore heat kernel bounds are not available.

3.3.2. Diffusions with generator in divergence form

In the following we write $\|A\|_{\infty,\mu} = \sup_{j,k}\|A_{jk}\|_{\infty,\mu}$ and $\|g\|_\mu = \int \|g(x)\|^2 d\mu(x)$ for measurable functions $x \mapsto A(x) \in \mathbb{R}^{d\times d}$ and $x \mapsto g(x) \in \mathbb{R}^d$.

Let L be an elliptic operator in divergence form (c.f. Bass (2006, Chapter VII)) and let $(X_r)_{r\geq 0}$ be the associated diffusion process (in the sense of Bass (2006, Section I.2)) with or without reflection arising as the solution of some stochastic differential equation. Assume that the process is stationary and takes its values in some closed subset $U \subset \mathcal{S} := \mathbb{R}^d$. Then the stationary measure μ has support in U. In case $U \subsetneq \mathbb{R}^d$ we think of μ as a measure on $\mathbb{R}^d$ and embed the domain of the infinitesimal generator $\mathrm{dom}(L) \subset L^2(U,\mu)$ canonically into $L^2(\mathbb{R}^d,\mu)$ by letting $Lf := L\tilde{f}$, whenever $f|_U = \tilde{f}$ for $f \in L^2(\mathbb{R}^d,\mu)$, $\tilde{f} \in L^2(U,\mu)$. Finally, assume that L satisfies

$$\langle -Lf, g\rangle_\mu = \int_{\mathbb{R}^d} \langle A(x)\nabla f(x), \nabla g(x)\rangle_{\mathbb{R}^d}\, d\mu(x), \quad f,g \in \mathrm{dom}(L) \cap C^2(\mathbb{R}^d), \qquad (3.3.1)$$

for a measurable function $x \mapsto A(x) \in \mathbb{R}^{d\times d}$ such that $A(x)$ is symmetric, positive definite for all $x \in \mathbb{R}^d$ and such that $\|A\|_{\infty,\mu}$ is finite. Observe that the right hand side of the last line is also well-defined for $L^2(\mathbb{R}^d,\mu)$-integrable functions $f, g \in C^1(\mathbb{R}^d)$. An operator L satisfying (3.3.1) is self-adjoint on $\mathrm{dom}(L) \cap C^2(\mathbb{R}^d)$. Observe for $f \in \mathrm{dom}(L) \cap C^2(\mathbb{R}^d) \subset \mathrm{dom}(L) \subset \mathrm{dom}(|L|^{s/2})$ and $0 \leq s \leq 1$ that

$$\begin{aligned}\||L|^{s/2} f\|_\mu^2 &= \|(-L)^{s/2} f\|_\mu^2 \leq \|(I-L)^{s/2} f\|_\mu^2 \leq \|(I-L)^{1/2} f\|_\mu^2 \\ &= \langle f - Lf, f\rangle_\mu \leq \|f\|_\mu^2 + \|A\|_{\infty,\mu}\|\nabla f\|_\mu^2 \\ &\leq \max(1, \|A\|_{\infty,\mu})\, \|f\|_{H^1(\mu)}^2, \end{aligned} \qquad (3.3.2)$$

where

$$\|f\|_{H^1(\mu)} = \|f\|_\mu + \|\nabla f\|_\mu$$

is the μ-weighted Sobolev norm. Combining this with Theorem 3.3.2 yields

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \leq \begin{cases} C \max\left(1, \|A\|_{\infty,\mu}^{1/2}\right) \|f\|_{H^1(\mu)} T^{1/2} \Delta_n, & f \in \mathrm{dom}(L) \cap C^2(\mathbb{R}^d), \\ C\|f\|_\mu T^{1/2} \Delta_n^{1/2}, & f \in L^2(\mu). \end{cases} \tag{3.3.3}$$

By interpolating between the two cases $f \in L^2(\mu)$ and $f \in \mathrm{dom}(L) \cap C^2(\mathbb{R}^d)$ we will study Hölder and indicator functions. Compared to the last section this interpolation will be done explicitly. The unified view using L^2-Sobolev spaces follows afterwards. Before doing this let us discuss some important examples where (3.3.1) holds.

Example 3.3.5 (Ornstein-Uhlenbeck process)**.** Assume that $(X_r)_{r\geq 0}$ satisfies the stochastic differential equation

$$dX_r = -X_r dr + \sqrt{2} dW_r$$

in $\mathbb{R}^d$ where $(W_r)_{r\geq 0}$ is a d-dimensional Brownian motion. If $X_0 \overset{d}{\sim} \mu$, where μ has Lebesgue density $d\mu(x)/d\lambda = (2\pi)^{-d/2} \exp(-|x|^2/2)$, then X is stationary with stationary measure μ. The infinitesimal generator L satisfies

$$Lf(x) = -\langle x, \nabla f(x)\rangle_{\mathbb{R}^d} + \Delta f(x), \quad x \in \mathbb{R}^d, \tag{3.3.4}$$

with $f \in \mathrm{dom}(L) = H^2(\mu)$, the μ-weighted Sobolev space of twice weakly differentiable functions with all partial derivatives up to order two belonging to $L^2(\mu)$. Using integration by parts it follows that

$$\langle -Lf, g\rangle_\mu = \int_{\mathbb{R}^d} \langle \nabla f(x), \nabla g(x)\rangle_{\mathbb{R}^d} \, d\mu(x), \quad f, g \in C^2(\mathbb{R}^d) \tag{3.3.5}$$

(cf. Pavliotis (2014, Section 4.4)). Hence, L is a self-adjoint operator of the form (3.3.1) with $A_{jk} = \mathbf{1}(j = k)$ for all $1 \leq j, k \leq d$. This example can be generalized considerably (see Chojnowska-Michalik and Goldys (2002) and Subsection 3.3.3 below).

Example 3.3.6 (Scalar diffusion with possibly attracting boundaries)**.** Fix some boundaries $-\infty \leq \beta < \rho \leq \infty$. Assume that $(X_r)_{r\geq 0}$ is a stationary diffusion process on $[\beta, \rho]$ solving the one-dimensional stochastic differential equation

$$dX_r = b(X_r)dr + \sigma(X_r)dW_r, \tag{3.3.6}$$

for a continuous drift $b : [\beta, \rho] \to \mathbb{R}$, strictly positive continuous volatility $\sigma : [\beta, \rho] \to (0, \infty)$ and a one-dimensional Brownian motion $(W_r)_{r\geq 0}$. Sufficient conditions for the existence have been provided by Hansen et al. (1998, Section 3.1). In particular, stationarity is guaranteed if the *speed density*

$$m(x) = \frac{1}{\sigma^2(x)} \exp\Big(\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)} dy\Big), \quad \beta \leq x_0, x \leq \rho,$$

is integrable on $[\beta, \rho]$. Then the stationary measure has the density

$$\frac{d\mu(x)}{d\lambda} = C_0 m(x) \mathbf{1}(\beta < x < \rho),$$

where C_0 is a normalizing constant. The infinitesimal generator L satisfies

$$\begin{aligned} Lf\,(x) &= b(x)f'(x) + \frac{\sigma^2(x)}{2} f''(x) \\ &= \frac{1}{2}\left(\frac{d\mu(x)}{d\lambda}\right)^{-1}\left(f'(x)\sigma^2(x)\frac{d\mu(x)}{d\lambda}\right)', \quad \beta < x < \rho, \end{aligned}$$

with $f \in \operatorname{dom}(L)$, where

$$\begin{aligned} \operatorname{dom}(L) = \Big\{ & f \in L^2([\beta,\rho],\mu) : f \text{ and } f' \text{ are absolutely continuous with} \\ & \lim_{x \searrow \lambda} f'(x)m\,(x)\,\sigma^2\,(x) = \lim_{x \nearrow \rho} f'(x)m\,(x)\,\sigma^2\,(x) = 0 \text{ and} \\ & Lf \in L^2([\beta,\rho],\mu)\Big\}. \end{aligned}$$

For details see Section 3.3 of Hansen et al. (1998). Embedding the domain into $L^2(\mu)$ as mentioned before and integrating by parts it follows that

$$\langle -Lf, g\rangle_\mu = \int_{\mathbb{R}} f'(x)g'(x)\sigma^2(x)\,d\mu(x), \quad f,g \in \operatorname{dom}(L) \cap C^2(\mathbb{R}), \tag{3.3.7}$$

which is of the form (3.3.1) with $A = \sigma^2$. For $b(x) = -x$ and $\sigma(x) = \sqrt{2}$, X is just the one-dimensional Ornstein-Uhlenbeck process.

Example 3.3.7 (Reflected diffusion)**.** Fix some boundaries $-\infty < \beta < \rho < \infty$. Assume that X is a one-dimensional reflected diffusion on $[\beta,\rho]$. By this we mean that X satisfies the Skorokhod type stochastic differential equation

$$dX_r = b(X_r)dr + \sigma(X_r)dW_r + dK_r, \tag{3.3.8}$$

for a bounded measurable drift $b : [\beta,\rho] \to \mathbb{R}$, strictly positive continuous volatility $\sigma : [\beta,\rho] \to (0,\infty)$, $(W_r)_{r\geq 0}$ is a Brownian motion and $(K_r)_{r\geq 0}$ is an adapted continuous process with finite variation starting from 0 and such that for every $r \geq 0$

$$\int_0^r \mathbf{1}_{(\beta,\rho)}(X_s)dK_s = 0.$$

The stationary measure and the generator L are as in the last example. Since $[\beta,\rho]$ is compact, the domain simplifies to

$$\begin{aligned} \operatorname{dom}(L) = \Big\{ & f \in L^2([\beta,\rho],\mu) : f \text{ and } f' \text{ are absolutely continuous with} \\ & f'(\lambda) = f'(\rho) = 0 \text{ and } Lf \in L^2([\beta,\rho],\mu)\Big\}. \end{aligned}$$

Therefore (3.3.1) holds here, as well. For more details see Chorowski (2018, Section 1.1).

Hölder functions

Consider an α-Hölder continuous function $f : \mathbb{R}^d \to \mathbb{R}$, $0 \leq \alpha \leq 1$, with finite Hölder-norm

$$\|f\|_\alpha = \sup_{x \neq y \in \mathbb{R}^d} \frac{|f(x) - f(y)|}{|x-y|^\alpha},$$

and such that $f \in L^2(\mu)$. Let $(\varphi_\varepsilon)_{\varepsilon \geq 0}$ be a non-negative smooth kernel, i.e. $\varphi_\varepsilon(x) = \varepsilon^{-1}\varphi(\varepsilon^{-1}x)$, $0 \leq \varphi \in C_c^\infty(\mathbb{R}^d)$, $\operatorname{supp}(\varphi) \subset [-1,1]^d$, $\int_{\mathbb{R}^d} \varphi(x)\,dx = 1$. Then the convolution $f_\varepsilon = f * \varphi_\varepsilon$ lies in $C^\infty(\mathbb{R}^d)$ and has bounded derivatives. Hence $f_\varepsilon \in L^2(\mu) \cap C^2(\mathbb{R}^d)$. It is not clear that $f_\varepsilon \in \operatorname{dom}(L)$ due to possible boundary conditions as in the examples above. In order to extend (3.3.3) assume the following:

Assumption 3.3.8. *$\operatorname{dom}(L) \cap C^2(\mathbb{R}^d)$ is dense in $L^2(\mu) \cap C^1(\mathbb{R}^d)$ with respect to $\|\cdot\|_{H^1(\mu)}$.*

This assumption is relatively weak and is satisfied in all the examples above. In particular, if there are no boundary conditions for $f \in \operatorname{dom}(L)$, then $L^2(\mu) \cap C^2(\mathbb{R}^d) = \operatorname{dom}(L) \cap C^2(\mathbb{R}^d)$, as is the case for the Ornstein-Uhlenbeck process. By approximation (3.3.3) can thus be extended to

$$\begin{aligned} &\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \\ &\leq \begin{cases} C \max\left(1, \|A\|_{\infty,\mu}^{1/2}\right) \|f\|_{H^1(\mu)} T^{1/2} \Delta_n, & f \in L^2(\mu) \cap C^1(\mathbb{R}^d), \\ C\|f\|_\mu T^{1/2} \Delta_n^{1/2}, & f \in L^2(\mu). \end{cases} \end{aligned} \tag{3.3.9}$$

Note that $f_\varepsilon \in L^2(\mu) \cap C^1(\mathbb{R}^d)$. Using $\int \varphi(x)\,dx = 1$ and $\int \nabla\varphi(x)dx = 0$, it follows that

$$\begin{aligned} \|f - f_\varepsilon\|_\mu^2 &= \int \left| \int (f(x) - f(x+\varepsilon y))\,\varphi(y)\,dy \right|^2 d\mu(x) \lesssim \|f\|_\alpha^2 \varepsilon^{2\alpha}, \\ \|\nabla f_\varepsilon\|_\mu^2 &= \int \| \frac{f(x)}{\varepsilon} \int \nabla\varphi(y)\,dy - \nabla f_\varepsilon(x)\|^2 d\mu(x) \\ &= \frac{1}{\varepsilon^2} \int \| \int (f(x) - f(x+\varepsilon y))\,\nabla\varphi(y)\,dy\|^2 d\mu(x) \\ &\lesssim \|f\|_\alpha^2 \varepsilon^{2\alpha-2}. \end{aligned}$$

From $\|f_\varepsilon\|_{H^1(\mu)} \lesssim \|f_\varepsilon\|_\mu + \|\nabla f_\varepsilon\|_\mu \leq \|f - f_\varepsilon\|_\mu + \|f\|_\mu + \|\nabla f_\varepsilon\|_\mu$ this yields with (3.3.9) that

$$\begin{aligned} &\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \\ &\quad \leq \left\|\Gamma_T(f - f_\varepsilon) - \widehat{\Gamma}_{n,T}(f - f_\varepsilon)\right\|_{L^2(\mathbb{P})} + \left\|\Gamma_T(f_\varepsilon) - \widehat{\Gamma}_{n,T}(f_\varepsilon)\right\|_{L^2(\mathbb{P})} \\ &\quad \lesssim \|f\|_\alpha T^{1/2} \Delta_n^{1/2} \varepsilon^\alpha + \|f\|_\alpha T^{1/2} \Delta_n \varepsilon^{\alpha-1} + \|f\|_\mu T^{1/2} \Delta_n. \end{aligned}$$

Choosing $\varepsilon = \Delta_n^{1/2}$ implies the bound $\|f\|_\alpha T^{1/2} \Delta_n^{(1+\alpha)/2} + \|f\|_\mu T^{1/2} \Delta_n$. Up to the second term, which is of smaller order as long as $\alpha < 1$, these are the rates obtained in Section 3.2.1. This can be improved, if L satisfies a Poincaré type inequality, i.e. if there exists a constant $c < \infty$ such that for all $f \in \operatorname{dom}(L)$

$$\|f_0\|_\mu \leq c\|\nabla f\|_\mu, \tag{3.3.10}$$

where $f_0 = f - \int f d\mu$. Let $f_{0,\varepsilon} = f_0 * \varphi_\varepsilon$. Then $\|f_\varepsilon\|_\alpha = \|f_{0,\varepsilon}\|_\alpha$ and it follows that $\|f_{0,\varepsilon}\|_{H^1(\mu)} \lesssim \|\nabla f_\varepsilon\|_\mu \lesssim \|f\|_\alpha \varepsilon^{\alpha-1}$. With $\varepsilon = \Delta_n^{1/2}$ this implies

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} = \left\|\Gamma_T(f_0) - \widehat{\Gamma}_{n,T}(f_0)\right\|_{L^2(\mathbb{P})} \lesssim \|f\|_\alpha T^{1/2} \Delta_n^{\frac{1+\alpha}{2}}. \tag{3.3.11}$$

Poincaré inequalities hold for many stationary measures μ, for example for the Ornstein-Uhlenbeck process and in Example 3.3.6 when $m(x)$ is uniformly bounded from above and below. For other examples see Bakry et al. (2013, Chapter 4) and Chen (2006). Observe that for $\alpha = 1$ the upper bound is $\|f\|_1 T^{1/2}\Delta_n$, removing an additional $\sqrt{\log n}$ term present in Theorem 3.2.4. In summary, we have shown the following:

Theorem 3.3.9. *Let X be a stationary diffusion with values in $\mathbb{R}^d$ and stationary measure μ, whose generator L satisfies (3.3.1) and Assumption 3.3.8. There exists a constant $C < \infty$ such that for all α-Hölder continuous functions f, $0 \leq \alpha \leq 1$,*

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \leq C\left(\|f\|_\alpha T^{1/2}\Delta_n^{\frac{1+\alpha}{2}} + \|f\|_\mu T^{1/2}\Delta_n\right).$$

If L satisfies a Poincaré type inequality as in (3.3.10) for some $c < \infty$, then the upper bound is $C\|f\|_\alpha T^{1/2}\Delta_n^{\frac{1+\alpha}{2}}$.

Indicator functions

Let $d = 1$ and consider $f = \mathbf{1}_{[K,\infty)}$, $K \in \mathbb{R}$, such that $f \in L^2(\mu)$. Let $(\varphi_\varepsilon)_{\varepsilon>0}$ be a non-negative smooth kernel as in the previous example. Then $f_\varepsilon = f * \varphi_\varepsilon$ is bounded by 1 and lies in $L^2(\mu) \cap C^2(\mathbb{R})$. $f - f_\varepsilon$ has support in $[K-\varepsilon, K+\varepsilon]$ such that

$$\begin{aligned}
\|f - f_\varepsilon\|_\mu^2 &\leq \int_{K-\varepsilon}^{K+\varepsilon} d\mu, \\
\|f_\varepsilon'\|_\mu^2 &= \int \left|\frac{f(x)}{\varepsilon}\int \varphi'(y)\,dy - f_\varepsilon'(x)\right|^2 d\mu(x) \\
&= \frac{1}{\varepsilon^2}\int_{K-\varepsilon}^{K+\varepsilon}\left|\int (f(x) - f(x+\varepsilon y))\,\varphi'(y)\,dy\right|^2 d\mu(x) \\
&\lesssim \frac{1}{\varepsilon^2}\int_{K-\varepsilon}^{K+\varepsilon} d\mu.
\end{aligned}$$

As before, $\|f_\varepsilon\|_\mu \leq \|f - f_\varepsilon\|_\mu + \|f\|_\mu$. If μ is absolutely continuous with respect to the Lebesgue measure λ with bounded density $d\mu/d\lambda$, then $\varepsilon^{-1}\int_{K-\varepsilon}^{K+\varepsilon} d\mu$ is bounded and in that case it follows from (3.3.3), uniformly in K, that

$$\left\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\right\|_{L^2(\mathbb{P})} \lesssim T^{1/2}(\Delta_n\varepsilon)^{1/2} + T^{1/2}\Delta_n\varepsilon^{-1/2} + T^{1/2}\Delta_n. \tag{3.3.12}$$

The last term is of lower order compared to the first two. Hence, choosing $\varepsilon = \Delta_n^{1/2}$ yields the rate $T^{1/2}\Delta_n^{3/4}$ obtained by Ngo and Ogawa (2011) and Kohatsu-Higa et al. (2014) for one-dimensional diffusions. However, now the rate is uniform in K with explicit dependence on T. These arguments can easily be extended to general dimension d implying the following theorem.

Theorem 3.3.10. *Let X be a stationary diffusion with values in $\mathbb{R}^d$ and stationary measure μ, whose generator L satisfies (3.3.1) and Assumption 3.3.8. Assume that μ has bounded Lebesgue density. If f is an indicator function in $\mathbb{R}^d$ of $[K_1, L_1) \times \cdots \times [K_d, L_d)$, $-\infty < K_j < L_j \leq \infty$, $1 \leq j \leq d$, then*

$$\left\| \Gamma_T(f) - \widehat{\Gamma}_{n,T}(f) \right\|_{L^2(\mathbb{P})} \lesssim T^{1/2} \Delta_n^{3/4},$$

uniformly in K_j, L_j.

The same rate clearly holds up to constants for finite linear combinations of such indicators.

Sobolev functions

The closure of $L^2(\mu) \cap C^1(\mathbb{R}^d)$ with respect to $\|\cdot\|_{H^1(\mu)}$ yields the space $H^1(\mu)$, a μ-weighted Sobolev space. This is not a Banach space in general (Kufner (1985)). This issue can be avoided if μ has a bounded Lebesgue density $d\mu/d\lambda$. Then $L^2(\mathbb{R}^d) \subset L^2(\mathbb{R}^d, \mu)$ and

$$\|f\|_{H^1(\mu)} \leq \|\frac{d\mu}{d\lambda}\|_\infty \|f\|_{H^1}, \quad f \in L^2(\mathbb{R}^d) \cap C^1(\mathbb{R}^d).$$

Taking the closure of $L^2(\mu) \cap C^2(\mathbb{R}^d)$ with respect to $\|\cdot\|_{H^1}$ yields $H^1(\mathbb{R}^d)$. This implies, instead of (3.3.9), that

$$\begin{aligned} &\left\| \Gamma_T(f) - \widehat{\Gamma}_{n,T}(f) \right\|_{L^2(\mathbb{P})} \\ &\leq \begin{cases} C \max\left(1, \|A\|_{\infty,\mu}^{1/2}\right) \|\frac{d\mu}{d\lambda}\|_\infty^{1/2} \|f\|_{H^1} T^{1/2} \Delta_n, & f \in H^1(\mathbb{R}^d), \\ C \|\frac{d\mu}{d\lambda}\|_\infty^{1/2} \|f\|_\lambda T^{1/2} \Delta_n^{1/2}, & f \in L^2(\mathbb{R}^d). \end{cases} \end{aligned} \tag{3.3.13}$$

From this we can obtain a special case of Theorem 3.2.6 by an interpolation argument as in the proof of that theorem.

Assume now that $H^1(\mu)$ is actually a Banach space. This is true, for instance, in the Examples 3.3.5 and 3.3.6, when $m(x)$ is uniformly bounded from above and below. In that case interpolation (in the sense of Adams and Fournier (2003, Chapter 7)) between $H^1(\mu)$ and $L^2(\mu)$ is possible and yields a similar bound as in Theorem 3.2.6, but with $\|\cdot\|_{H^s}$ replaced by an appropriate interpolation norm. The results in Theorems 3.3.9 and 3.3.10 are explicit cases of this. Up to boundary conditions this implies that α-Hölder functions lie in $\mathrm{dom}((-L)^{\alpha/2})$, $0 \leq \alpha \leq 1$, and indicator functions $f = \mathbf{1}_{[K,L)}$ lie in $\mathrm{dom}((-L)^{1/4})$, $-\infty < K < L \leq \infty$.

Depending on the boundary conditions for $f \in \mathrm{dom}(L)$ and assuming that μ has bounded Lebesgue density, it can be shown in many examples that $H^1(\mathbb{R}^d)$ embeds continuously into $\mathrm{dom}((I-L)^{1/2}) \subset \mathrm{dom}((-L)^{1/2})$. This holds, for instance, for the Ornstein-Uhlenbeck process and for the reflected diffusions in Example 3.3.6. Since $L^2(\mathbb{R}^d) \subset \mathrm{dom}((-L)^0) = L^2(\mathbb{R}^d, \mu)$, interpolation implies that $H^s(\mathbb{R}^d)$ embeds continuously into $\mathrm{dom}((I-L)^{s/2}) \subset \mathrm{dom}((-L)^{s/2})$. In particular, the indicator functions $f = \mathbf{1}_{[K,L)}$ lie in $\mathrm{dom}((-L)^{1/4-\varepsilon})$ for any small $\varepsilon > 0$.

Remark 3.3.11. Arguing like in the proof of Corollary 3.3.3 the strict stationarity assumption can be relaxed here and in Theorems 3.3.9 and 3.3.10.

3.3.3. Infinite dimensional diffusions

Since the general state space $\mathcal{S}$ of X is only assumed to be Polish, it is also possible to study infinite dimensional diffusions. The results of Section 3.2.1 do not apply then, because, in general, heat kernel bounds are not available in this setting. Example 3.3.5 can be generalized considerably. If X satisfies the stochastic differential equation

$$dX_r = AX_r dr + Q^{1/2} dW_r,$$

where A and Q are operators on a separable Hilbert space $\mathcal{H}$, with Q being bounded self-adjoint, then X is a Gaussian Markov process and the generator L satisfies a similar formula as in (3.3.4) with ∇ and Δ replaced by the corresponding Fréchet derivatives D and D^2. Under certain conditions on A and Q the generator is reversible and X has a stationary measure μ. The domain is again a μ-weighted Sobolev space and the associated Dirichlet form is

$$\langle -Lf, g\rangle_\mu = \frac{1}{2}\int_{\mathcal{H}} \left\langle Q^{1/2}Df\left(x\right), Q^{1/2}Dg\left(x\right)\right\rangle_{\mathcal{H}} d\mu\left(x\right).$$

The results of Section 3.3.2 therefore remain formally the same. For details see Chojnowska-Michalik and Goldys (2002). For a different kind of example consider an infinite dimensional system of the form

$$dX_r^{(i)} = \left(pV'\left(X_r^{(i+1)} - X_r^{(i)}\right) - qV'\left(X_r^{(i)} - X_r^{(i-1)}\right)\right)dr + dW_r^{(i)},$$

where $(r,i) \in [0,\infty) \times \mathbb{Z}$, $p,q \geq 0$ with $p = (1+\sqrt{\varepsilon})/2$, $q = (1-\sqrt{\varepsilon})/2$, $\varepsilon > 0$, where $\{(W_r^{(i)})_{r\geq 0} : i \in \mathbb{Z}\}$ is an independent family of Brownian motions and where V is some potential function (Diehl et al. (2017)). $X = (X_r^{(i)})_{r\geq 0, i\in\mathbb{Z}}$ is stationary and the infinitesimal generator $L^{(\varepsilon)} = L_S + \sqrt{\varepsilon}L_A$ can be studied via its symmetric and antisymmetric parts L_S and L_A. The Dirichlet form for the symmetric part is given in Lemma 2.1 of Diehl et al. (2017). If $\varepsilon = 0$, then the generator is symmetric and a similar analysis as in Section 3.3.2 can be applied.

3.4. Lower bounds

We will now address the important question if the upper bounds for $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}$ derived in the last two sections are optimal. Optimality here means that the upper bounds cannot be improved uniformly for all f belonging to a given class of functions. For this it is sufficient to find a candidate f where the error $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}$ matches the upper bound up to an absolute constant. The only explicit lower bound in the literature has been established by Ngo and Ogawa (2011) for Brownian motion in $d = 1$ and indicator functions $f = \mathbf{1}_{[a,b]}$, matching the upper bound $\Delta_n^{3/4}$.

Apart from optimality with respect to the Riemann-sum estimator, it is interesting from a statistical point of view to ask for optimality across *all* possible estimators. Note that $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}$ is bounded from below by

$$\inf_g \|\Gamma_T(f) - g\|_{L^2(\mathbb{P})} = \|\Gamma_T(f) - \mathbb{E}\left[\Gamma_T(f)\middle|\, \mathcal{G}_n\right]\|_{L^2(\mathbb{P})}, \tag{3.4.1}$$

where $\mathcal{G}_n = \sigma(X_{t_k} : 0 \leq k \leq n)$ and where the infimum is taken over all $\mathcal{G}_n$-measurable random variables. If f is the identity, then it is well-known that $\mathbb{E}[\Gamma_T(f)|\mathcal{G}_n] = \widehat{\Theta}_{n,T}(f)$, where $\widehat{\Theta}_{n,T}(f)$ is the trapezoid rule estimator from Section 3.1.1 (see e.g. (Diaconis, 1988)). If $f \in H^1(\mathbb{R}^d)$, then this still holds approximately. The methods from Section 3.1 allow for identifying the limit of the right hand side in (3.4.1) as $n \to \infty$, yielding an explicit lower bound valid for all $f \in H^1(\mathbb{R}^d)$. For the L^2-Sobolev spaces $H^s(\mathbb{R}^d)$ with $0 < s < 1$ such a universal result is not possible. Instead, we derive a lower bound for an explicit candidate matching the upper bound established in Example 3.2.10.

Theorem 3.4.1. *Let $T \geq 1$ and let $X_t = X_0 + W_t$, where $(W_t)_{0 \leq t \leq T}$ is a d-dimensional Brownian motion and where X_0 satisfies (X0).*

(i) We have for any $f \in H^1(\mathbb{R}^d)$ the asymptotic lower bound

$$\begin{aligned}
&\liminf_{n\to\infty} \left(\Delta_n^{-1} \|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})} \right) \\
&\geq \liminf_{n\to\infty} \left(\Delta_n^{-1} \inf_g \|\Gamma_T(f) - g\|_{L^2(\mathbb{P})} \right) \\
&= \mathbb{E}\left[\frac{1}{12} \int_0^T \|\nabla f(X_r)\|^2 dr \right]^{1/2},
\end{aligned}$$

where the infimum is taken over all $\mathcal{G}_n$-measurable random variables.

(ii) Let $f_\alpha \in L^2(\mathbb{R}^d)$, $0 < \alpha < 1$, be the $L^2(\mathbb{R}^d)$ function with Fourier transform $\mathcal{F} f_\alpha(u) = (1 + \|u\|)^{-\alpha - d/2}$, $u \in \mathbb{R}^d$. Then $f_\alpha \in H^s(\mathbb{R}^d)$ for all $0 \leq s < \alpha$, but $f_\alpha \notin H^\alpha(\mathbb{R}^d)$. Moreover, f_α satisfies for all $0 \leq s < \alpha$ the asymptotic lower bound

$$\begin{aligned}
&\liminf_{n\to\infty} \left(\Delta_n^{-\frac{1+s}{2}} \|\Gamma_T(f_\alpha) - \widehat{\Gamma}_{n,T}(f_\alpha)\|_{L^2(\mathbb{P})} \right) \\
&\geq \liminf_{n\to\infty} \left(\Delta_n^{-\frac{1+s}{2}} \inf_g \|\Gamma_T(f_\alpha) - g\|_{L^2(\mathbb{P})} \right) > 0.
\end{aligned}$$

For $d = 1$ the lower bounds also hold for $X_t = W_t$ (cf. Remark 3.2.9). Interestingly, the asymptotic lower bound in (i) corresponds exactly to the asymptotic variance obtained for the CLTs in Section 3.1. This proves the asymptotic efficiency of $\widehat{\Gamma}_{n,T}(f)$ and $\widehat{\Theta}_{n,T}(f)$ for $f \in H^1(\mathbb{R}^d)$. Note that Brownian motion is a major example for the upper bounds derived in the last section.

The key step in the proof is to calculate the conditional expectation $\mathbb{E}[\Gamma_T(f)|\mathcal{G}_n]$, which reduces to Brownian bridges interpolating between the observations. The same calculations hold when X is a Lévy process with finite first moments (cf. Jacod and Protter (1988, Theorem 2.6)) and similarly when X belongs to a certain class of Markov processes (cf. Chaumont and Uribe Bravo (2011)).

3.5. Proofs

3.5.1. Proofs of Section 1

In the following, T is fixed and $\Delta_n \to 0$ as $n \to \infty$. Consider first the following preliminary observations.

Preliminaries

By the localization procedure in Section A.2 and Assumption (SM-α) it is sufficient to prove Theorems 3.1.3 and 3.1.5 under Assumption (H-α-β) from the last chapter.

In this case it is enough to prove Theorems 3.1.3 and 3.1.5 for f with compact support. Indeed, if $f \in FL^s_{loc}(\mathbb{R}^d)$ or $f \in H^s_{loc}(\mathbb{R}^d)$ is replaced by $\tilde{f} = f\varphi$ for a smooth cutoff function φ with compact support in a ball $B_{K+\varepsilon} = \{x \in \mathbb{R}^d : \|x\| \leq K + \varepsilon\}$ of radius $K + \varepsilon$, $\varepsilon > 0$, and $\varphi = 1$ on B_K, then $\tilde{f} = f$ on B_K and $\tilde{f} \in FL^s(\mathbb{R}^d)$ or $\tilde{f} \in H^s(\mathbb{R}^d)$.

Moreover, in order to replace b and σ by piecewise constant approximations let $\lfloor t \rfloor_{\Delta_n} = \lfloor t/\Delta_n \rfloor \Delta_n$, $t \geq 0$, and define the process $X(\Delta_n) = (X_t(\Delta_n))_{0\leq t\leq T}$, where $X_t(\Delta_n) = X_{\lfloor t \rfloor_{\Delta_n}} + b_{\lfloor t \rfloor_{\Delta_n}}(t - \lfloor t \rfloor_{\Delta_n}) + \sigma_{\lfloor t \rfloor_{\Delta_n}}(W_t - W_{\lfloor t \rfloor_{\Delta_n}})$.

The main estimates distinguishing the proofs of Theorems 3.1.3 and 3.1.5 are collected in the following two lemmas. Recall that $FL^1(\mathbb{R}^d) \subset C^1(\mathbb{R}^d)$.

Lemma 3.5.1. *Assume (H-α-β) for $0 \leq \alpha, \beta \leq 1$. Let $f \in C^1(\mathbb{R}^d)$ have compact support. Then it follows for $k = 1, \ldots, n$ and $t_{k-1} \leq r \leq t_k$, uniformly in r and k:*

(i) $\mathbb{E}[\|\nabla f(X_r)\|^2] = O(\|\nabla f\|_\infty^2)$,

(ii) $\mathbb{E}[\langle \nabla f(X_{t_{k-1}}), X_r - X_r(\Delta_n)\rangle^2] = o(\Delta_n \|\nabla f\|_\infty^2)$,

(iii) $\mathbb{E}[(f(X_r) - f(X_{t_{k-1}}) - \langle \nabla f(X_{t_{k-1}}), X_r - X_{t_{k-1}}\rangle)^2] = o(\Delta_n \|\nabla f\|_\infty^2)$,

(iv) $\mathbb{E}[\|\nabla f(X_r) - \nabla f(X_{t_{k-1}})\|^2] = o(\|\nabla f\|_\infty^2)$,

(v) $\mathbb{E}[\sup_t |\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \int_{t_{k-1}}^{t_k} (t_k - r - \Delta_n/2)\mathbb{E}[\langle \nabla f(X_r), b_r\rangle | \mathcal{F}_{t_{k-1}}]dr|] = o(\Delta_n \|\nabla f\|_\infty)$.

Proof. We only prove (v). The other statements follow from the boundedness of ∇f and Proposition A.3.2. (v) follows immediately for bounded and continuous b, because $\langle \nabla f(X_r), b_r\rangle$ can be approximated uniformly at the left end $\langle \nabla f(X_{t_{k-1}}), b_{t_{k-1}}\rangle$. For bounded b let g_ε be continuous and adapted processes such that $\sup_{0\leq t\leq T}\|g_{\varepsilon,t}\| < \infty$ uniformly in ε and $\mathbb{E}[\int_0^T \|b_h - g_{\varepsilon,h}\|dh] \to 0$ as $\varepsilon \to 0$. Then (v) holds for g_ε and by approximation also for b. ☐

Lemma 3.5.2. *Assume (H-α-β) for $0 \leq \alpha, \beta \leq 1$ and (X0). Let $f \in H^1(\mathbb{R}^d)$. Then it follows for $k = 1, \ldots, n$ and $t_{k-1} \leq r \leq t_k$, uniformly in r and k:*

(i) $\mathbb{E}[\|\nabla f(X_r)\|^2] = O(\|f\|_{H^1}^2)$,

(ii) $\mathbb{E}[\langle \nabla f(X_h), X_r - X_r(\Delta_n)\rangle^2] = o(\Delta_n \|f\|_{H^1}^2)$,

(iii) $\mathbb{E}[(f(X_r) - f(X_h) - \langle \nabla f(X_h), X_r - X_h\rangle)^2] = o(\Delta_n)$,

(iv) $\mathbb{E}[\|\nabla f(X_r) - \nabla f(X_{t_{k-1}})\|^2] = o(1)$,

(v) $\mathbb{E}[|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} (t_k - r - \Delta_n/2)\mathbb{E}[\langle \nabla f(X_r), b_r\rangle | \mathcal{F}_{t_{k-1}}]dr|] = o(\Delta_n \|f\|_{H^1})$.

Proof. The marginals X_r have uniformly bounded Lebesgue densities p_r by (X0). Hence (i) follows from

$$\mathbb{E}\left[\|\nabla f(X_r)\|^2\right] = \sum_{m=1}^d \int (\partial_m f(x))^2 p_r(x)\, dx \lesssim \|f\|_{H^1}^2. \tag{3.5.1}$$

With respect to (ii) consider first $f \in \mathcal{S}(\mathbb{R}^d)$. By inverse Fourier transform and $\mathcal{F}(\nabla f)(u) = iu\mathcal{F}f(u)$, $u \in \mathbb{R}^d$, it follows that $\langle \nabla f(X_{t_{k-1}}), X_r - X_r(\Delta_n)\rangle^2$ is equal to

$$\begin{aligned}
&(2\pi)^{-2d}\left(\int \mathcal{F}f(u) i \left\langle u, X_r - X_r\left(\Delta_n\right)\right\rangle e^{-i\left\langle u, X_{t_{k-1}}-X_0\right\rangle} e^{-i\langle u, X_0\rangle} du\right)^2 \\
&= -(2\pi)^{-2d} \int \mathcal{F}f(u)\mathcal{F}f\left(v\right) \left\langle u, X_r - X_r\left(\Delta_n\right)\right\rangle \\
&\quad \cdot \left\langle v, X_r - X_r\left(\Delta_n\right)\right\rangle e^{-i\left\langle u+v, X_{t_{k-1}}-X_0\right\rangle} e^{-i\langle u+v, X_0\rangle} d\left(u, v\right).
\end{aligned}$$

As X_0 and $(X_t - X_0)_{0\leq t\leq T}$ are independent, $\mathbb{E}[\langle \nabla f(X_{t_{k-1}}), X_r - X_r(\Delta_n)\rangle^2]$ is up to a constant bounded by

$$\left(\int \left|\mathcal{F}f\left(u\right)\right| \left|\mathcal{F}f\left(v\right)\right| \|u\|\|v\| \left|\mathcal{F}\mu\left(u+v\right)\right| d\left(u, v\right)\right) \mathbb{E}\left[\|X_r - X_r(\Delta_n)\|^2\right],$$

which is of order $o(\Delta_n\|f\|_{H^1}^2)$ by Lemma 3.5.3 and Proposition A.3.2. This yields (ii) for $f \in \mathcal{S}(\mathbb{R}^d)$. For $f \in H^1(\mathbb{R}^d)$ consider a sequence $(f_m)_{m\geq 1} \subset \mathcal{S}(\mathbb{R}^d)$ converging to f with respect to $\|\cdot\|_{H^1}$. Then $\|X_r - X_r(\Delta_n)\| \leq \|X_r\| + \|X_r(\Delta_n\| \lesssim 1 + \|W_r - W_{t_{k-1}}\|$. Independence yields

$$\begin{aligned}
&\left|\|\langle \nabla f(X_{t_{k-1}}), X_r - X_r(\Delta_n)\rangle\|_{L^2(\mathbb{P})} - \|\langle \nabla f_m(X_{t_{k-1}}), X_r - X_r(\Delta_n)\rangle\|_{L^2(\mathbb{P})}\right| \\
&\lesssim \mathbb{E}[\|\nabla\left(f - f_m\right)\left(X_{t_{k-1}}\right)\|^2]^{1/2}\mathbb{E}[\left(1 + \|W_r - W_{t_{k-1}}\|\right)^2]^{1/2} \lesssim \|f - f_m\|_{H^1} \to 0,
\end{aligned}$$

as $m \to \infty$. Hence (ii) also holds for $f \in H^1(\mathbb{R}^d)$. With respect to (iii) consider again first $f \in \mathcal{S}(\mathbb{R}^d)$. Arguing by inverse Fourier transform, the left hand side is because of Taylor's theorem bounded by

$$\begin{aligned}
&\int_0^1 \mathbb{E}\left[\left\langle \nabla f\left(X_{t_{k-1}} + t\left(X_r - X_{t_{k-1}}\right)\right) - \nabla f\left(X_{t_{k-1}}\right), X_r - X_{t_{k-1}}\right\rangle^2\right] dt \\
&\lesssim \int \left|\mathcal{F}f\left(u\right)\right| \left|\mathcal{F}f\left(v\right)\right| \|u\|\|v\| \mathbb{E}\left[g_n\left(u\right) g_n\left(v\right)\right] \left|\mathcal{F}\mu\left(u+v\right)\right| d\left(u, v\right) \\
&\quad \cdot \mathbb{E}\left[\|X_r - X_{t_{k-1}}\|^4\right]^{1/2},
\end{aligned}$$

where $g_n(u) = \sup_{r,h:|r-h|\leq\Delta_n} \int_0^1 |1 - e^{-it\langle u, X_r - X_h\rangle}|^2 dt$ and where we applied the Cauchy-Schwarz inequality twice. Lemma 3.5.3 together with $\mathbb{E}[\|X_r - X_{t_{k-1}}\|^4]^{1/2} = O(\Delta_n)$ shows that the left hand side in (iii) is for $f \in \mathcal{S}(\mathbb{R}^d)$ up to a constant bounded by

$$\Delta_n \int \left|\mathcal{F}f\left(u\right)\right|^2 \|u\|^2 \mathbb{E}\left[g_n^2\left(u\right)\right]^{1/2} du.$$

A similar approximation argument as for (ii) yields the same bound for $f \in H^1(\mathbb{R}^d)$. $g_n(u)$ is bounded in n and u and converges $\mathbb{P}$-almost surely to 0 as $n \to \infty$ for any $u \in \mathbb{R}^d$. By dominated convergence the last display is thus of order $o(\Delta_n)$. This yields (iii). (iv) is proved similarly. For (v) and bounded and continuous b the claim follows from

$$\begin{aligned}
&\langle \nabla f(X_r), b_r\rangle - \langle \nabla f(X_{t_{k-1}}), b_{t_{k-1}}\rangle \\
&= \langle \nabla f(X_r), b_r - b_{t_{k-1}}\rangle + \langle \nabla f(X_r) - \nabla f(X_{t_{k-1}}), b_{t_{k-1}}\rangle,
\end{aligned}$$

part (iv) and (3.5.1). For bounded b argue as in part (v) of the last lemma. □

Lemma 3.5.3. *Let $\xi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and let μ be a probability density on $\mathbb{R}^d$.*

(i) If $\mathcal{F}\mu \in L^1(\mathbb{R}^d)$, then

$$\int |\xi(u)\xi(v)| \, |\mathcal{F}\mu(u+v)| \, d(u,v) \lesssim \|\mathcal{F}\mu\|_{L^1} \|\xi\|_{L^2}^2.$$

(ii) If $\mathcal{F}\mu$ is non-negative and μ is bounded, then the upper bound is instead $\|\mu\|_\infty \|\xi\|_{L^2}^2$.

Proof. By a density argument we can assume that $\xi, \mu \in \mathcal{S}(\mathbb{R}^d)$ and that $\mathcal{F}\mu$ is non-negative in (ii). Let $g, h \in L^2(\mathbb{R}^d)$ with $\mathcal{F}g(u) = |\xi(u)|$, $\mathcal{F}h(u) = |\mathcal{F}\mu(u)|$ such that the $d(u,v)$ integral is equal to

$$\begin{aligned}\int \mathcal{F}g(u)\,\mathcal{F}g(v)\,\mathcal{F}h(u+v)\,d(u,v) &= \int \mathcal{F}g(u-v)\,\mathcal{F}g(v)\,\mathcal{F}h(u)\,d(u,v) \\ = \int (\mathcal{F}g * \mathcal{F}g)(u)\,\mathcal{F}h(u)\,du &= \int \mathcal{F}g^2(u)\,\mathcal{F}h(u)\,du = C\int g^2(u)\,h(-u)\,du,\end{aligned} \tag{3.5.2}$$

where we used the Plancherel Theorem in the last line. If $\mathcal{F}\mu \in L^1(\mathbb{R}^d)$, then the last line is bounded by

$$C\|g\|_{L^2}^2 \|h\|_\infty \lesssim \|\xi\|_{L^2}^2 \sup_{u \in \mathbb{R}^d} \left| \int \mathcal{F}h(x)\, e^{i\langle u,x\rangle} dx \right| \lesssim \|\mathcal{F}\mu\|_{L^1} \|\xi\|_{L^2}^2.$$

If, on the other hand, $\mathcal{F}\mu$ is non-negative, then $h(u) = \mathcal{F}\mathcal{F}h(-u) = \mu(-u)$ and therefore (3.5.2) is bounded by

$$C\|g\|_{L^2}^2 \|h\|_\infty \lesssim \|\mu\|_\infty \|\xi\|_{L^2}^2.$$

This shows (i) and (ii). □

Proof of Theorem 3.1.3

It is enough to show the CLT in (3.1.1) for $f \in FL^s_{loc}(\mathbb{R}^d)$, which immediately yields the claim in terms of $\Gamma_t(f) - \widehat{\Theta}_{n,t}(f)$. Recall the decomposition $\Gamma_t(f) - \widehat{\Gamma}_{n,t}(f) = M_{n,t}(f) + D_{n,t}(f)$ with $M_{n,t}(f)$ and $D_{n,t}(f)$ as in (3.1.2) and (3.1.3). By the localization argument in the preliminaries above the proof follows from the following two propositions.

Proposition 3.5.4. *Assume (H-α-β) for $0 \le \alpha, \beta \le 1$. Let $f \in C^1(\mathbb{R}^d)$ have compact support. Then we have the stable convergence*

$$\Delta_n^{-1} M_{n,t}(f) \xrightarrow{st} \frac{1}{2}\int_0^t \langle \nabla f(X_r), \sigma_r dW_r\rangle + \frac{1}{\sqrt{12}}\int_0^t \left\langle \nabla f(X_r), \sigma_r d\widetilde{W}_r \right\rangle$$

as processes on $\mathcal{D}([0,T], \mathbb{R}^d)$, where $\widetilde{W}$ is a d-dimensional Brownian motion defined on an independent extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$.

Proposition 3.5.5. *Assume (H-α-β) for $0 \le \alpha, \beta \le 1$. Let $s > 2 - 2\alpha$, $s \ge 1$, $s + \beta > 1$. Then we have for $f \in FL^s(\mathbb{R}^d)$ with compact support that*

$$\Delta_n^{-1} D_{n,t}(f) \xrightarrow{ucp} \frac{1}{2}\left(f(X_t) - f(X_0)\right) - \frac{1}{2}\int_0^t \langle \nabla f(X_r), \sigma_r dW_r\rangle. \tag{3.5.3}$$

We note in the proofs precisely where Lemma 3.5.1 is used, which allows for deducing Theorem 3.1.5 with small modifications.

Proof of Proposition 3.5.4. We write $M_{n,t}(f) = \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} Z_k$ and $\tilde{M}_{n,t}(f) = \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \tilde{Z}_k$ for random variables

$$Z_k = \int_{t_{k-1}}^{t_k} \left(f\left(X_r\right) - \mathbb{E}\left[f(X_r) \middle| \mathcal{F}_{t_{k-1}} \right] \right) dr, \tag{3.5.4}$$

$$\tilde{Z}_k = \int_{t_{k-1}}^{t_k} \left\langle \nabla f(X_{t_{k-1}}), X_r\left(\Delta_n\right) - \mathbb{E}\left[X_r\left(\Delta_n\right) \middle| \mathcal{F}_{t_{k-1}} \right] \right\rangle dr. \tag{3.5.5}$$

$\tilde{Z}_k$ "linearizes" Z_k with respect to f. The proof is based on the following statements for $0 \leq t \leq T$:

$$\Delta_n^{-1} \sup_{0 \leq t \leq T} \left| M_{n,t}\left(f\right) - \tilde{M}_{n,t}\left(f\right) \right| \xrightarrow{\mathbb{P}} 0, \tag{3.5.6}$$

$$\Delta_n^{-2} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\Big[\tilde{Z}_k^2 \Big| \mathcal{F}_{t_{k-1}} \Big] \xrightarrow{\mathbb{P}} \frac{1}{3} \int_0^t \|\sigma_r^\top \nabla f(X_r)\|^2 dr, \tag{3.5.7}$$

$$\Delta_n^{-2} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\Big[\tilde{Z}_k^2 \mathbf{1}_{\{|\tilde{Z}_k| > \varepsilon\}} \Big| \mathcal{F}_{t_{k-1}} \Big] \xrightarrow{\mathbb{P}} 0, \quad \text{for all } \varepsilon > 0, \tag{3.5.8}$$

$$\Delta_n^{-1} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[\tilde{Z}_k \left(W_{t_k} - W_{t_{k-1}} \right)^\top \middle| \mathcal{F}_{t_{k-1}} \right] \xrightarrow{\mathbb{P}} \frac{1}{2} \int_0^t \nabla f(X_r)^\top \sigma_r dr, \tag{3.5.9}$$

$$\Delta_n^{-1} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[\tilde{Z}_k \left(N_{t_k} - N_{t_{k-1}} \right) \middle| \mathcal{F}_{t_{k-1}} \right] \xrightarrow{\mathbb{P}} 0, \tag{3.5.10}$$

where (3.5.10) holds for all bounded ($\mathbb{R}$-valued) martingales N which are orthogonal to all components of W. (3.5.6) yields $M_{n,t}(f) = \tilde{M}_{n,t}(f) + o_{ucp}(\Delta_n)$. The claim follows thus from the remaining statements (3.5.7) through (3.5.10) and Theorem A.1.2.

We prove now the five statements above. $M_{n,t}(f) - \tilde{M}_{n,t}(f)$ is a discrete martingale such that by the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\Big[\sup_{0 \leq t \leq T} \Big(M_{n,t}\left(f\right) - \tilde{M}_{n,t}\left(f\right) \Big)^2 \Big] \leq \sum_{k=1}^{n} \mathbb{E}\left[\left(Z_k - \tilde{Z}_k \right)^2 \right].$$

Decompose any such $Z_k - \tilde{Z}_k$ into

$$\int_{t_{k-1}}^{t_k} \left(A_{k,r} - \mathbb{E}\left[A_{k,r} \middle| \mathcal{F}_{t_{k-1}} \right] \right) dr \tag{3.5.11}$$

$$+ \int_{t_{k-1}}^{t_k} \left\langle \nabla f\left(X_{t_{k-1}} \right), X_r - X_r(\Delta_n) - \mathbb{E}\left[X_r - X_r(\Delta_n) \middle| \mathcal{F}_{t_{k-1}} \right] \right\rangle dr, \tag{3.5.12}$$

where $A_{k,r} = f(X_r) - f(X_{t_{k-1}}) - \langle \nabla f(X_{t_{k-1}}), X_r - X_{t_{k-1}} \rangle$. The second moment of (3.5.12) is bounded by $2\Delta_n \int_{t_{k-1}}^{t_k} \mathbb{E}[\langle \nabla f(X_{t_{k-1}}), X_r - X_r(\Delta_n) \rangle^2] dr = o(\Delta_n^3)$ using Lemma

3.5.1(ii). The same order follows for the second moment of (3.5.11) from Lemma 3.5.1(iii). This yields (3.5.6). In order to prove the remaining statements observe first by the (stochastic) Fubini theorem that $\tilde{Z}_k$ is equal to

$$\langle \nabla f(X_{t_{k-1}}), \int_{t_{k-1}}^{t_k} (t_k - r)(b_r - \mathbb{E}[b_r|\mathcal{F}_{t_{k-1}}])\rangle dr$$
$$+ \langle \nabla f(X_{t_{k-1}}), \sigma_{t_{k-1}} \int_{t_{k-1}}^{t_k} (t_k - r) dW_r \rangle.$$

The first term is of smaller order than the second one. By Itô isometry, because σ is càdlàg and from Lemma 3.5.1(i),(iv) it therefore follows that the left hand side in (3.5.7) is equal to

$$\frac{\Delta_n}{3} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \|\sigma_{t_{k-1}}^\top \nabla f\left(X_{t_{k-1}}\right)\|^2 + o_{\mathbb{P}}(1) = \frac{1}{3}\int_0^t \|\sigma_r^\top \nabla f\left(X_r\right)\|^2 dr + o_{\mathbb{P}}(1).$$

With respect to (3.5.8) note that $|\tilde{Z}_k| > \varepsilon$ implies $\|\int_{t_{k-1}}^{t_k}(t_k - r)dW_r\| > \varepsilon'$ for some $\varepsilon' > 0$ and sufficiently large n because of the Cauchy-Schwarz inequality. Consequently, it follows from Lemma 3.5.1(i) and independence that

$$\mathbb{E}\left[\tilde{Z}_k^2 \mathbf{1}_{\{|\tilde{Z}_k|>\varepsilon\}}\right] \lesssim \mathbb{E}\left[\|\nabla f(X_{t_{k-1}})\|^2\right]\left(\Delta_n^4 + \mathbb{E}\left[\|\int_{t_{k-1}}^{t_k}(t_k - r)\, dW_r\|^4\right]\right),$$

which is of order $O(\Delta_n^4)$, thus implying (3.5.8). The left hand side of (3.5.9), on the other hand, is equal to $R_n + \frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \nabla f(X_{t_{k-1}})^\top \sigma_{t_{k-1}}$ with $\mathbb{E}[\|R_n\|] = o(1)$ by Itô's isometry (applied coordinatewise). (3.5.9) follows then from σ being càdlàg and 3.5.1(iv). The same argument shows that the left hand side in *(3.5.10)* is of order $o_{\mathbb{P}}(1)$. □

Proof of Proposition 3.5.5. Lemma 3.5.6(i) below shows

$$D_{n,t}(f) = \frac{\Delta_n}{2} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[f(X_{t_k}) - f(X_{t_{k-1}}) \middle| \mathcal{F}_{t_{k-1}}\right] + o_{ucp}(\Delta_n). \tag{3.5.13}$$

In order to find the limit of this sum, write it as

$$\frac{\Delta_n}{2} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[A_k \middle| \mathcal{F}_{t_{k-1}}\right] + \frac{\Delta_n}{2} \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[B_k \middle| \mathcal{F}_{t_{k-1}}\right], \tag{3.5.14}$$

where $A_k = f(X_{t_k}) - f(X_{t_{k-1}}) - B_k$ and $B_k = \langle \nabla f(X_{t_{k-1}}), X_{t_k} - X_{t_{k-1}}\rangle$. Note that by the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{k=1}^{\lfloor t/\Delta_n \rfloor}\left(\mathbb{E}\left[A_k \middle| \mathcal{F}_{t_{k-1}}\right] - A_k\right)\right|^2\right] \lesssim \sum_{k=1}^{n} \mathbb{E}\left[A_k^2\right],$$

which is of order $o(\Delta_n)$ by Lemma 3.5.1(iii). Therefore, (3.5.14) is up to an error of order $o_{ucp}(\Delta_n)$ equal to

$$\frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\left(f\left(X_{t_k}\right)-f\left(X_{t_{k-1}}\right)\right)+\frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\left(\mathbb{E}\left[B_k\middle|\mathcal{F}_{t_{k-1}}\right]-B_k\right),$$

The first sum is just $\frac{\Delta_n}{2}(f(X_{\lfloor t/\Delta_n\rfloor})-f(X_0))=\frac{\Delta_n}{2}(f(X_t)-f(X_0))+o_{ucp}(\Delta_n)$, while the second one is equal to

$$\begin{aligned}&\frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left\langle\nabla f\left(X_{t_{k-1}}\right),\left(\mathbb{E}\left[b_r\middle|\mathcal{F}_{t_{k-1}}\right]-b_r\right)\right\rangle dr\\&-\frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left\langle\nabla f\left(X_{t_{k-1}}\right),\sigma_r dW_r\right\rangle.\end{aligned}$$

This is equal to $-\frac{\Delta_n}{2}\int_0^{\lfloor t/\Delta_n\rfloor}\langle\nabla f(X_r),\sigma_r dW_r\rangle+o_{ucp}(\Delta_n)$ and the claim follows. In the second line use Lemma 3.5.1(iv) and for the first line note that it is a discrete martingale of order $o_{ucp}(\Delta_n)$ by the Burkholder-Davis-Gundy inequality and Lemma 3.5.1(i). □

We now state and prove the lemmas used above.

Lemma 3.5.6. *Assume (H-α-β) for $0\le\alpha,\beta\le 1$. Let $s>2-2\alpha$, $s\ge 1$, $s+\beta>1$. Then we have for $f\in FL^s(\mathbb{R}^d)$ with compact support that*

$$D_{n,t}(f)-\frac{\Delta_n}{2}\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}[f(X_{t_k})-f(X_{t_{k-1}})|\mathcal{F}_{t_{k-1}}]=o_{ucp}(\Delta_n).$$

Proof. Consider first $f\in\mathcal{S}(\mathbb{R}^d)$. By applying Itô's formula and the Fubini theorem the left hand side in the statement is equal to $D_{n,t}(1,f)+D_{n,t}(2,f)$, where $D_{n,t}(1,f)$ and $D_{n,t}(2,f)$ are defined by

$$\begin{aligned}&\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(t_k-r-\frac{\Delta_n}{2}\right)\mathbb{E}\left[\left\langle\nabla f(X_r),b_r\right\rangle\middle|\mathcal{F}_{t_{k-1}}\right]dr,\\&\sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(t_k-r-\frac{\Delta_n}{2}\right)\mathbb{E}\left[\frac{1}{2}\sum_{l,m=1}^{d}\partial_{lm}^2 f\left(X_r\right)\left(\sigma_r\sigma_r^\top\right)^{(l,m)}\middle|\mathcal{F}_{t_{k-1}}\right]dr.\end{aligned}$$

We will show that

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\le t\le T}\left|D_{n,t}(1,f)+D_{n,t}(2,f)\right|\right]\\&\lesssim o\left(\Delta_n\|f\|_{FL^s}\right)+\Delta_n\int\left|\mathcal{F}f\left(u\right)\right|\left(1+\|u\|\right)^s g_n\left(u\right)du,\end{aligned}\tag{3.5.15}$$

with g_n as in Lemma 3.5.7 below. Choose now any sequence $(f_m)\subset\mathcal{S}(\mathbb{R}^d)$ converging to $f\in FL^s(\mathbb{R}^d)$ with respect to $\|\cdot\|_{FL^s}$. This means, in particular, that f_m converges to f uniformly. Therefore (3.5.15) also holds for f. The properties of g_n and dominated convergence therefore imply the claim.

In order to show (3.5.15) note first that $\mathbb{E}[\sup_{0\leq t\leq T}|D_{n,t}(1,f)|] = o(\Delta_n \|f\|_{FL^s})$ follows already from Lemma 3.5.1(v). With respect to $D_{n,t}(2,f)$ write $\Sigma_t = \sigma_t\sigma_t^\top$ and fix $l,m=1,\ldots,d$. For $f\in\mathcal{S}(\mathbb{R}^d)$ it is always justified to exchange integrals in the following calculations. We can write $\partial^2_{lm}f(X_r) = -(2\pi)^{-d}\int \mathcal{F}f(u)u_l u_m e^{-i\langle u,X_r\rangle}du$ such that

$$D_{n,t}\left(2,f\right) = -(2\pi)^{-d}\int \mathcal{F}f\left(u\right)u_l u_m Q_{n,t}\left(u\right)du,$$

where

$$Q_{n,t}(u) = \sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(t_k - r - \frac{\Delta_n}{2}\right)\mathbb{E}\left[e^{-i\langle u,X_r\rangle}\Sigma_r{}^{(l,m)}\middle|\mathcal{F}_{t_{k-1}}\right]dr.$$

Consequently, because of

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int \mathcal{F}f\left(u\right)u_l u_m Q_{n,t}\left(u\right)du\right|\right] \leq \int \left|\mathcal{F}f\left(u\right)\right|\|u\|^2\mathbb{E}\left[\sup_{0\leq t\leq T}\left|Q_{n,t}\left(u\right)\right|\right]du,$$

the remaining part of (3.5.15) follows from Lemma 3.5.7. □

The following lemma is stronger than necessary here. This will become useful for Theorem 3.1.5.

Lemma 3.5.7. *Assume (H-α-β) for $0\leq\alpha,\beta\leq 1$. Let $s>2-2\alpha$, $s\geq 1$, $s+\beta>1$. Then we have uniformly in $u\in\mathbb{R}^d$ that*

$$\|\sup_{0\leq t\leq T}Q_{n,t}\left(u\right)\|_{L^2(\mathbb{P})} \leq C\Delta_n\left(1+\|u\|\right)^{s-2}g_n\left(u\right),$$

where $\sup_{n\geq1}\sup_{u\in\mathbb{R}^d}|g_n(u)|<\infty$ and $g_n(u)\to 0$ for all $u\in\mathbb{R}^d$ as $n\to\infty$.

Proof. The proof is divided into five steps.

Step 1. Let $0<\varepsilon\leq 1$ and define $t_\varepsilon = \max(\lfloor t/\varepsilon\rfloor\varepsilon - 2\varepsilon, 0)$ for $0\leq t\leq T$. t_ε projects t to the grid $\{0,\varepsilon,2\varepsilon,\ldots,\lceil T/\varepsilon\rceil\varepsilon\}$ such that $t-t_\varepsilon\leq 3\varepsilon$ and $t-t_\varepsilon\geq\min(2\varepsilon,t)$. Later, we will choose ε depending on n and u, i.e. $\varepsilon=\varepsilon(u,n)$. Define the process $\tilde{X}_t\left(\varepsilon\right) = X_{t_\varepsilon}+b_{t_\varepsilon}(t-t_\varepsilon)+\sigma_{t_\varepsilon}(W_t-W_{t_\varepsilon})$. Assumption (H-$\alpha$-$\beta$) implies $\mathbb{E}[(\Sigma_t{}^{(l,m)}-\Sigma_{t_\varepsilon}^{(l,m)})^2]\lesssim\varepsilon^{2\alpha}$ and Proposition A.3.2 yields $\mathbb{E}[\|X_t-\widetilde{X}_t(\varepsilon)\|^2]\lesssim(\varepsilon^{2(\beta+1)}+\varepsilon^{2(\alpha+1/2)})$. Define

$$Q_{n,t}\left(\varepsilon,u\right) = \sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(t_k - r - \frac{\Delta_n}{2}\right)\mathbb{E}\left[e^{-i\left\langle u,\widetilde{X}_r(\varepsilon)\right\rangle}\Sigma_{r_\varepsilon}^{(l,m)}\middle|\mathcal{F}_{t_{k-1}}\right]dr.$$

The Lipschitz-continuity of $x\mapsto e^{ix}$ therefore yields

$$\begin{aligned}
&\|\sup_{0\leq t\leq T}\left(Q_{n,t}\left(u\right)-Q_{n,t}\left(\varepsilon,u\right)\right)\|_{L^2(\mathbb{P})}\\
&\lesssim \Delta_n\left(\int_0^T\mathbb{E}\left[\left|e^{-i\langle u,X_r\rangle}\Sigma_r^{(l,m)} - e^{-i\left\langle u,\widetilde{X}_r(\varepsilon)\right\rangle}\Sigma_{r_\varepsilon}^{(l,m)}\right|^2\right]dr\right)^{1/2}\\
&\lesssim \Delta_n\left(1+\|u\|\right)^{s-2}g_{n,1}(u),
\end{aligned}$$

with $g_{n,1}(u) = (1+\|u\|)^{2-s}(\varepsilon^{\alpha} + \|u\|\varepsilon^{1+\beta} + \|u\|\varepsilon^{1/2+\alpha})$. We study now $Q_{n,t}(\varepsilon, u)$.

Step 2. With respect to the new grid $\{0, \varepsilon, 2\varepsilon, \ldots, \lceil T/\varepsilon\rceil\varepsilon\}$ and $0 \leq t \leq T$ let

$$I_j(t) = \{k = 1, \ldots, \lfloor t/\Delta_n \rfloor : (j-1)\varepsilon < t_k \leq j\varepsilon\}, \quad 1 \leq j \leq \lceil T/\varepsilon\rceil,$$

be the set of blocks $k \leq \lfloor t/\Delta_n \rfloor$ with right endpoints $t_k \leq t$ inside the interval $(j-1)\varepsilon, j\varepsilon]$. Then $Q_{n,t}(\varepsilon, u) = \sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}(u) + \sum_{j=1}^{\lceil T/\varepsilon\rceil} \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-1)\varepsilon}]$ for $R_{j,t}(u) = A_{j,t}(u) - \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-1)\varepsilon}]$ and where

$$A_{j,t}(u) = \sum_{k\in I_j(t)} \int_{t_{k-1}}^{t_k} \left(t_k - r - \frac{\Delta_n}{2}\right) \xi_{r,k} dr, \quad \xi_{r,k} = \mathbb{E}\left[e^{-i\langle u, \widetilde{X}_r(\varepsilon)\rangle} \Sigma_{r_\varepsilon}^{(l,m)} \middle| \mathcal{F}_{t_{k-1}}\right],$$

such that $A_{j,t}(u)$ is $\mathcal{F}_{j\varepsilon}$-measurable for fixed u. We want to show that $\sup_{0\leq t\leq T} |\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}(u)|$ is negligible. Note first that $I_j(t) = \emptyset$ for $t \leq (j-1)\varepsilon$ and $I_j(t) = I_j(T)$ for $t > j\varepsilon$. Therefore, $A_{j,t}(u) = 0$ for $t \leq (j-1)\varepsilon$ and $A_{j,t}(u) = A_{j,T}(u)$ for $t > j\varepsilon$. Denote by j^* the unique $j \in \{1, \ldots, \lceil T/\varepsilon\rceil\}$ with $(j-1)\varepsilon < t \leq j\varepsilon$. Then $\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}(u) = B_{j^*-1}(u) + R_{j^*,t}(u)$, where $B_m(u) = \sum_{j=1}^{m} R_{j,T}(u)$ defines a complex-valued martingale $(B_m(u))_{m=0,\ldots,\lceil T/\varepsilon\rceil}$ with respect to the filtration $(\mathcal{F}_{m\varepsilon})_{m=0,\ldots,\lceil T/\varepsilon\rceil}$. The Burkholder-Davis-Gundy inequality then yields

$$\mathbb{E}\Bigg[\sup_{0\leq t\leq T}\Bigg|\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}(u)\Bigg|^2\Bigg] \lesssim \mathbb{E}\Bigg[\sup_{m\in\{0,\ldots,\lceil T/\varepsilon\rceil\}} |B_m(u)|^2 + \sup_{0\leq t\leq T} |R_{j^*,t}(u)|^2\Bigg]$$
$$\lesssim \mathbb{E}\left[\sum_{j=1}^{\lceil T/\varepsilon\rceil} |A_{j,T}(u)|^2\right] + \mathbb{E}\left[\sup_{0\leq t\leq T} |A_{j^*,t}(u)|^2\right].$$

If $\varepsilon < \Delta_n$, then each $I_j(t)$ contains at most one block k and for $t_{k-1} \leq r \leq t_k \leq j\varepsilon$ we have necessarily $t_{k-1} \leq (j-1)\varepsilon = r_\varepsilon$. Hence, $|\xi_{r,k}| \lesssim |\mathbb{E}[e^{-i\langle u, \sigma_{r_\varepsilon}(W_r - W_{r_\varepsilon})\rangle}|\mathcal{F}_{r_\varepsilon}]| \leq e^{-\frac{\|u\|^2}{2K}\varepsilon}$ by Assumption (H-α-β) and thus $|A_{j,t}(u)| \lesssim \Delta_n^2 e^{-\frac{\|u\|^2}{2K}\varepsilon}$. Moreover, there are clearly at most Δ_n^{-1} many non-empty $I_j(t)$. Consequently in this case the last display is up to a constant bounded by $\Delta_n^3 e^{-\frac{\|u\|^2}{K}\varepsilon}$.

Assume in the following that $\varepsilon \geq \Delta_n$. Then $I_j(t)$ contains at most $C\varepsilon\Delta_n^{-1}$ many blocks k and therefore

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |A_{j^*,t}(u)|^2\right] \lesssim \Delta_n^2\varepsilon^2. \tag{3.5.16}$$

Moreover,

$$\mathbb{E}\Bigg[\sum_{j=1}^{\lceil T/\varepsilon\rceil} |A_{j,T}(u)|^2\Bigg]^2$$
$$\lesssim \Delta_n^4 \sum_{j_1,j_2=1}^{\lceil T/\varepsilon\rceil} \sum_{k_1,k_1'\in I_{j_1}(T)} \sum_{k_2,k_2'\in I_{j_2}(T)} \int_{t_{k_1-1}}^{t_{k_1}} \int_{t_{k_1'-1}}^{t_{k_1'}}$$
$$\int_{t_{k_2-1}}^{t_{k_2}} \int_{t_{k_2'-1}}^{t_{k_2'-1}} \left|\mathbb{E}\left[\xi_{r_1k_1}\overline{\xi_{r_1'k_1'}} \cdots \xi_{r_2k_2}\overline{\xi_{r_2'k_2'}}\right]\right| d\left(r_1, r_1', r_2, r_2'\right).$$

Fix j and k_1, k_1', k_2, k_2', $r_1, r_1', \dots, r_2, r_2'$. Let r and h be the largest and second largest indices in the set $\{r_l, r_{l'} : 1 \le l \le 2\}$ with corresponding blocks $k, \tilde{k}$ such that $t_{k-1} \le r \le t_k$, $t_{\tilde{k}-1} \le h \le t_{\tilde{k}}$. Without loss of generality assume $h \le r$. If $t_{k-1} \le r_\varepsilon < t_k$, then

$$\left|\mathbb{E}\left[\xi_{r_1k_1}\overline{\xi_{r_1'k_1'}}\xi_{r_2k_2}\overline{\xi_{r_2'k_2'}}\right]\right| \lesssim \mathbb{E}\left[|\xi_{r,k}|\right] \lesssim e^{-\frac{\|u\|^2}{2K}\varepsilon}.$$

If, on the other hand, $h \le r_\varepsilon < t_{k-1} \le r < t_k$, then

$$\left|\mathbb{E}\left[\xi_{r_1k_1}\overline{\xi_{r_1'k_1'}}\xi_{r_2k_2}\overline{\xi_{r_2'k_2'}}\right]\right| \lesssim \mathbb{E}\left[\left|\mathbb{E}\left[\xi_{r,k}\right|\mathcal{F}_{r_\varepsilon}\right]\right|\right] \lesssim e^{-\frac{\|u\|^2}{2K}\varepsilon}.$$

In the two cases $r_\varepsilon < t_{k-1} \le h \le r < t_k$ and $r_\varepsilon < h < t_{k-1} \le r < t_k$ conditioning on $\mathcal{F}_h$ instead gives

$$\left|\mathbb{E}\left[\xi_{r_1k_1}\overline{\xi_{r_1'k_1'}}\xi_{r_2k_2}\overline{\xi_{r_2'k_2'}}\right]\right| \lesssim \mathbb{E}\left[\left|\mathbb{E}\left[e^{-i\langle u, \sigma_{r_\varepsilon}(W_r - W_h)\rangle}\right|\mathcal{F}_h\right]\right|\right] \lesssim e^{-\frac{\|u\|^2}{2K}|r-h|}.$$

As $\varepsilon \ge \Delta_n$, it follows that $\sum_{k\in I_j(T)}\int_{t_{k-1}}^{t_k} 1dr \le \varepsilon$. In all, we conclude that $\mathbb{E}[(\sum_{j=1}^{\lceil T/\varepsilon\rceil}|A_{j,T}(u)|^2)^{p/2}]^2$ is up to a constant bounded by

$$\Delta_n^4\left(\varepsilon^2 e^{-\frac{\|u\|^2}{2K}\varepsilon} + \varepsilon\sum_{j=1}^{\lceil T/\varepsilon\rceil}\sum_{k,\tilde{k}\in I_j(T)}\int_{t_{\tilde{k}-1}}^{t_{\tilde{k}}}\int_{t_{k-1}}^{t_k} e^{-\frac{\|u\|^2}{2K}|r-h|}drdh\right).$$

By symmetry in r, h we find for $u \neq 0$ that

$$\begin{aligned}
&\sum_{j=1}^{\lceil T/\varepsilon\rceil}\sum_{k,\tilde{k}\in I_j(T)}\int_{t_{\tilde{k}-1}}^{t_{\tilde{k}}}\int_{t_{k-1}}^{t_k} e^{-\frac{\|u\|^2}{2K}|r-h|}drdh \\
&\le 2\sum_{j=1}^{\lceil T/\varepsilon\rceil}\sum_{\tilde{k}\in I_j(T)}\int_{t_{\tilde{k}-1}}^{t_{\tilde{k}}}\int_{h}^{j\varepsilon} e^{-\frac{\|u\|^2}{2K}(r-h)}drdh \\
&\lesssim \sum_{j=1}^{\lceil T/\varepsilon\rceil}\sum_{\tilde{k}\in I_j(T)}\int_{t_{\tilde{k}-1}}^{t_{\tilde{k}}} 1dh\|u\|^{-2}\left(1 - e^{-\frac{\|u\|^2}{2}(\varepsilon+\Delta_n)}\right) \\
&\lesssim \|u\|^{-2}\left(1 - e^{-\frac{\|u\|^2}{2}(\varepsilon+\Delta_n)}\right),
\end{aligned}$$

because $1 - e^{-\frac{\|u\|^2}{2}(j\varepsilon - h)} \le 1 - e^{-\frac{\|u\|^2}{2}(\varepsilon+\Delta_n)}$ for $t_{\tilde{k}-1} \le h \le j\varepsilon$ and $\tilde{k} \in I_j(T)$. Combining the estimates for $\varepsilon < \Delta_n$ and $\varepsilon \ge \Delta_n$ in all we have shown in this step that

$$\begin{aligned}
&\|\sup_{0\le t\le T} Q_{n,t}(\varepsilon, u)\|_{L^2(\mathbb{P})} \lesssim \Delta_n\left(1 + \|u\|\right)^{s-2} g_{n,2}(u) \\
&\quad + \|\sup_{0\le t\le T}\sum_{j=1}^{\lceil T/\varepsilon\rceil}\mathbb{E}\left[A_{j,t}(u)\right|\mathcal{F}_{(j-1)\varepsilon}\right]\|_{L^2(\mathbb{P})}
\end{aligned}$$

with

$$g_{n,2}(u) = (1+\|u\|)^{2-s}(\Delta_n^{1/2}e^{-\frac{\|u\|^2}{2K}\varepsilon} + \varepsilon^{1/2-1/4}\|u\|^{-1/2}\left(1-e^{-\frac{\|u\|^2}{2}(\varepsilon+\Delta_n)}\right)^{1/4} + \varepsilon).$$

Step 3. We need to use two martingale decompositions. Write

$$\sum_{j=1}^{\lceil T/\varepsilon\rceil} \mathbb{E}\left[A_{j,t}(u) \middle| \mathcal{F}_{(j-1)\varepsilon}\right] = \sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}^{(1)}(u) + \sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}^{(2)}(u) + \sum_{j=1}^{\lceil T/\varepsilon\rceil} \mathbb{E}\left[A_{j,t}(u) \middle| \mathcal{F}_{(j-3)\varepsilon}\right],$$

where $R_{j,t}^{(1)}(u) = \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-1)\varepsilon}] - \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-2)\varepsilon}]$, $R_{j,t}^{(2)} = \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-2)\varepsilon}] - \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-3)\varepsilon}]$. The arguments in step *2* can be applied to $\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}^{(1)}(u)$ and $\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}^{(2)}(u)$ instead of $\sum_{j=1}^{\lceil T/\varepsilon\rceil} R_{j,t}(u)$. Moreover, for $r \leq 3\varepsilon$ observe that $r_\varepsilon = 0$. Hence $\mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-3)\varepsilon}]$ is for $j \in \{1,2,3\}$ up to a constant bounded by

$$\sum_{k\in I_j(t)} \int_{t_{k-1}}^{t_k} \left(t_k - r - \frac{\Delta_n}{2}\right) e^{-\frac{\|\sigma_0^\top u\|^2}{2}r} dr \lesssim \Delta_n \int_0^\varepsilon e^{-\frac{\|u\|^2}{2K}r} dr \leq \Delta_n\varepsilon.$$

We conclude that

$$\|\sup_{0\leq t\leq T} \sum_{j=1}^{\lceil T/\varepsilon\rceil} \mathbb{E}\left[A_{j,t}(u) \middle| \mathcal{F}_{(j-1)\varepsilon}\right]\|_{L^2(\mathbb{P})} \lesssim \Delta_n (1+\|u\|)^{s-2} g_{n,2}(u) + \|\sup_{0\leq t\leq T} \sum_{j=4}^{\lceil T/\varepsilon\rceil} \mathbb{E}\left[A_{j,t}(u) \middle| \mathcal{F}_{(j-3)\varepsilon}\right]\|_{L^2(\mathbb{P})}.$$

Step 4. For $t_{k-1} \leq r \leq t_k$ and $k \in I_j(t)$, $j \geq 4$, note that $r_\varepsilon = (j-3)\varepsilon$. Hence $\mathbb{E}[\xi_{r,k}|\mathcal{F}_{(j-3)\varepsilon}] = Y_k V_{r,k}$, where

$$V_{r,k} = e^{-i\langle u, b_{(j-3)\varepsilon}(r-t_{k-1})\rangle - \frac{\|\sigma_{(j-3)\varepsilon}^\top u\|^2}{2}(r-t_{k-1})},$$
$$Y_k = e^{-i\langle u, X_{(j-3)\varepsilon} + b_{(j-3)\varepsilon}(t_{k-1}-(j-3)\varepsilon)\rangle - \frac{\|\sigma_{(j-3)\varepsilon}^\top u\|^2}{2}(t_{k-1}-(j-3)\varepsilon)} \Sigma_{(j-3)\varepsilon}^{(l,m)}.$$

Since also $t_{k-1} - (j-3)\varepsilon > \varepsilon$, it follows that $|Y_k| \lesssim e^{-\frac{\|u\|^2}{2K}\varepsilon}$. Moreover, $\int_{t_{k-1}}^{t_k}(t_k - r - \frac{\Delta_n}{2})Y_k V_{t_k,k} dr = 0$. We therefore conclude that $\sum_{j=4}^{\lceil T/\varepsilon\rceil} \mathbb{E}[A_{j,t}(u)|\mathcal{F}_{(j-3)\varepsilon}]$ is bounded by

$$\Delta_n \left(\sum_{j=4}^{\lceil T/\varepsilon\rceil} \sum_{k\in I_j(t)} \int_{t_{k-1}}^{t_k} |Y_k|\,|V_{r,k} - V_{t_k,k}|\, dr\right) \lesssim \Delta_n^2 (1+\|u\|)^2 e^{-\frac{\|u\|^2}{2K}\varepsilon}.$$

Consequently, it follows with $g_{n,3}(u) = \Delta_n(1+\|u\|)^{4-s}e^{-\frac{\|u\|^2}{2K}\varepsilon}$ that

$$\| \sup_{0\le t\le T} \sum_{j=4}^{\lceil T/\varepsilon\rceil} \mathbb{E}\left[A_{j,t}(u) \middle| \mathcal{F}_{(j-3)\varepsilon}\right] \|_{L^2(\mathbb{P})} \lesssim \Delta_n (1+\|u\|)^{s-2} g_{n,3}(u).$$

Step 5. The four previous steps combined show that $\|\sup_{0\le t\le T} Q_{n,t}(u)\|_{L^2(\mathbb{P})}$ is up to a constant bounded by $\Delta_n(1+\|u\|)^{s-2}g_n(u)$ with $g_n(u) = g_{n,1}(u)+g_{n,2}(u)+g_{n,3}(u)$. Set $\varepsilon = \varepsilon(u,n) := \min\left(\nu_n\|u\|^{-2}, 1\right)$ for $\nu_n = 2K\log(1+\|u\|^3\Delta_n^{1/2})$. Hence $0<\varepsilon\le 1$ and $\varepsilon \to 0$ for fixed u. Choose $C\ge 1$ large enough to ensure that $\varepsilon(u,n) < 1$ for $\|u\| > C$ and $n = 1$ (and thus for all n). For $\|u\|\le C$ this means $\varepsilon \le \nu_n\|u\|^{-2} \lesssim \Delta_n^{1/2}$ and $\sup_{u:\|u\|\le C} g_n(u) = o(1)$. For $\|u\| > C$, on the other hand, it follows that

$$\begin{aligned} g_{n,1}(u) &\lesssim \|u\|^{2-s}\left(\|u\|^{-1-2\beta}\nu_n^{1+\beta} + \|u\|^{-2\alpha}\nu_n^{1/2+\alpha}\right), \\ g_{n,2}(u) &\lesssim \|u\|^{2-s}\Bigg(\Delta_n^{1/2}\left(1+\|u\|^3\Delta_n^{1/2}\right)^{-1} \\ &\quad + \|u\|^{-1}\nu_n^{1/2-1/(2p)}\left(1-e^{-\frac{\|u\|^2}{2}(\varepsilon+\Delta_n)}\right) + \|u\|^{-2}\nu_n\Bigg), \\ g_{n,3}(u) &\lesssim \|u\|^{4-s}\Delta_n\left(1+\|u\|^3\Delta_n^{1/2}\right)^{-1}. \end{aligned}$$

The assumptions that $2-s-2\alpha<0$, $s\ge 1$, $s+\beta>1$ and the fact that ν_n grows in u only logarithmically imply that $\sup_{\|u\|>C} g_n(u)$ is bounded in n. Consequently, $\sup_{n\ge 1}\sup_{u\in\mathbb{R}^d} g_n(u)$ is bounded. Moreover, for fixed u with $\|u\| > C$ it follows that $g_n(u)\to 0$ as $n\to\infty$. This proves the claim. □

Proof of Theorem 3.1.5

Similar to Theorem 3.1.3 and given the preliminaries it is sufficient to prove the following two propositions for $f\in H^s(\mathbb{R}^d)$.

Proposition 3.5.8. *Assume (H-α-β) for $0\le\alpha,\beta\le 1$ and (X0). Then we have for $f\in H^1(\mathbb{R}^d)$ the stable convergence*

$$\Delta_n^{-1}M_{t,n}(f) \xrightarrow{st} \frac{1}{2}\int_0^t \langle \nabla f(X_r), \sigma_r dW_r\rangle + \frac{1}{\sqrt{12}}\int_0^t \left\langle \nabla f(X_r), \sigma_r d\widetilde{W}_r\right\rangle \qquad (3.5.17)$$

as processes on $\mathcal{D}([0,T],\mathbb{R}^d)$, where $\widetilde{W}$ is a d-dimensional Brownian motion defined on an independent extension of $(\Omega,\mathcal{F},(\mathcal{F}_t)_{0\le t\le T},\mathbb{P})$.

Proposition 3.5.9. *Assume (H-α-β) for $0\le\alpha,\beta\le 1$ and (X0). Let $s>2-2\alpha$, $s\ge 1$, $s+\beta>1$. Then we have for $f\in H^s(\mathbb{R}^d)$ that*

$$\Delta_n^{-1}D_{T,n}(f) \xrightarrow{\mathbb{P}} \frac{1}{2}\left(f(X_T)-f(X_0)\right) - \frac{1}{2}\int_0^T \langle \nabla f(X_r), \sigma_r dW_r\rangle. \qquad (3.5.18)$$

Note that the convergence in the second proposition is not functional as compared to Proposition 3.5.5. Since the weak limit in (3.5.17) is a continuous process, convergence with respect to the Skorokhod topology and thus the stable convergence also hold at $t = T$ (Billingsley (2013)). This yields the CLT in (3.1.1) for $f \in H^s(\mathbb{R}^d)$ and at the fixed time T.

Proof of Proposition 3.5.8. The proof of Proposition 3.5.4 can be repeated in exactly the same way after replacing all references to Lemma 3.5.1 by the corresponding statements in Lemma 3.5.2. We only have to argue differently for (3.5.8), because $\nabla f(X_r)$ may not be bounded.

As $\int_{t_{k-1}}^{t_k}(t_k - r)dW_r$ is independent of $\mathcal{F}_{t_{k-1}}$, it follows from the Cauchy-Schwarz inequality that $\mathbb{E}[\tilde{Z}_k^2 \mathbf{1}_{\{|\tilde{Z}_k|>\varepsilon\}}|\mathcal{F}_{t_{k-1}}]$ is up to a constant bounded by $\|\nabla f(X_{(k-1)\Delta_n})\|^2 \mathbb{E}[(\Delta_n^4 + \Delta_n^3 Y_k^2)\mathbf{1}_{\{\|\nabla f(X_{(k-1)\Delta_n})\|\Delta_n^{3/2}(1+|Y_k|)>\varepsilon'\}}|\mathcal{F}_{t_{k-1}}]$ for $\varepsilon' > 0$ and with $Y_k \sim N(0,1)$ independent of $\mathcal{F}_{t_{k-1}}$. Since the marginals have uniformly bounded Lebesgue densities (uniform in time), it follows that the first moment of the left hand side in (3.5.8) is up to a constant bounded by

$$\int \|\nabla f(x)\|^2 \mathbb{E}\left[\left(\Delta_n + Y_1^2\right)\mathbf{1}_{\left\{\|\nabla f(x)\|\Delta_n^{3/2}(1+|Y_1|)>\varepsilon'\right\}}\right] dx.$$

This converges to 0 by dominated convergence, implying (3.5.8). □

Proof of Proposition 3.5.9. The proof follows the one of Proposition 3.5.5. We only have to replace all references to 3.5.1 by the corresponding statements in Lemma 3.5.2 and use Lemma 3.5.10 instead of Lemma 3.5.6, while also replacing all o_{ucp} expressions by the respective $o_{\mathbb{P}}$ terms. □

Lemma 3.5.10. *Assume (H-α-β) for $0 \leq \alpha, \beta \leq 1$ and (X0). Let $s > 2 - 2\alpha$, $s \geq 1$, $s + \beta > 1$. Then we have for $f \in H^s(\mathbb{R}^d)$ with compact support, $s \geq 1$ and $s > 2 - 2\alpha$, that*

$$D_{n,T}(f) - \frac{\Delta_n}{2}\sum_{k=1}^{n}\mathbb{E}[f(X_{t_k}) - f(X_{t_{k-1}})|\mathcal{F}_{t_{k-1}}] = o_{\mathbb{P}}(\Delta_n).$$

Proof. Using the notation from Lemma 3.5.6 we only have to show for $f \in \mathcal{S}(\mathbb{R}^d)$ that

$$\begin{aligned}&\mathbb{E}\left[|D_{n,T}(1,f) + D_{n,T}(2,f)|\right]\\&\lesssim o\left(\Delta_n\|f\|_{H^s}\right) + \Delta_n\left(\int |\mathcal{F}f(u)|^2\left(1+\|u\|\right)^{2s} g_n^2(u)\,du\right)^{1/2},\end{aligned} \tag{3.5.19}$$

with g_n as in Lemma 3.5.7. This can be extended to $f \in H^s(\mathbb{R}^d)$ by an approximation argument as in Lemma 3.5.6.

$\mathbb{E}[|D_{n,T}(1,f)|] = o(\Delta_n\|f\|_{H^s})$ follows from Lemma 3.5.2(*v*). With respect to $D_{n,T}(2,f)$ we write

$$D_{n,T}(2,f) = -(2\pi)^{-d}\int \mathcal{F}f(u)\,u_l u_m e^{-i\langle u,X_0\rangle}\tilde{Q}_{n,T}(u)\,du$$

with

$$\tilde{Q}_{n,T}(u) = \sum_{k=1}^{\lfloor t/\Delta_n\rfloor}\int_{t_{k-1}}^{t_k}\left(t_k - r - \frac{\Delta_n}{2}\right)\mathbb{E}\left[e^{-i\langle u,X_r-X_0\rangle}\Sigma_r^{(l,m)}\middle|\mathcal{F}_{t_{k-1}}\right]dr.$$

This corresponds to $Q_{n,T}(u)$ in Lemma 3.5.7 with $X_r - X_0$ instead of X_r. Consequently, the independence from (X0) shows that $\mathbb{E}[|D_{n,T}(2,f)|^2]$ is equal to

$$(2\pi)^{-2d}\int \mathcal{F}f\left(u\right)\mathcal{F}f\left(v\right)\mathcal{F}\mu\left(u+v\right)u_l u_m v_l v_m \mathbb{E}\left[\tilde{Q}_{n,T}\left(u\right)\tilde{Q}_{n,T}\left(v\right)\right]d\left(u,v\right)$$
$$\lesssim \int \left|\mathcal{F}f(u)\right|^2 \|u\|^4 \mathbb{E}\left[\left|\tilde{Q}_{n,T}\left(u\right)\right|^2\right]du,$$

by Lemma 3.5.3. The remaining part of (3.5.19) follows therefore from Lemma 3.5.7. □

Proof of Corollary 3.1.6

Proof. Without loss of generality we can assume in the following that $\mathcal{F}$ and the corresponding extensions are separable. In fact, it is enough to prove stable convergence for separable $\mathcal{F}$, essentially because the σ-fields generated by X, b and σ are separable (see Jacod and Shiryaev (2013, Theorem IX 7.3) for details). Assume first that $X_0 = 0$. On a suitable extension as in Theorem 3.1.5, denoted by $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{0\le t\le T}, \mathbb{P}')$, let $F_n(X, x_0)$ be defined as the random variables

$$\Delta_n^{-1}\left(\int_0^T f\left(X_r + x_0\right)dr - \Delta_n \sum_{k=1}^n \frac{1}{2}\left(f\left(X_{t_{k-1}} + x_0\right) + f\left(X_{t_k} + x_0\right)\right)\right)$$

and let $F(X, \sigma, \widetilde{W}, x_0) = \sqrt{1/12}\int_0^T \langle \nabla f(X_r + x_0), \sigma_r d\widetilde{W}_r\rangle$, where F_n and F are measurable functions and $x_0 \in \mathbb{R}^d$. The stable convergence in the claim is equivalent to $\mathbb{E}[Ug(F_n(X, x_0))] \to \mathbb{E}[Ug(F(X, \sigma, \widetilde{W}, x_0))]$ for any continuous bounded function $g: \mathbb{R} \to \mathbb{R}$ and any bounded $\mathcal{F}$-measurable random variable U (cf. Section A.1). We have to show that this holds for almost all $x_0 \in \mathbb{R}^d$.

Let $(\Omega'', \mathcal{F}'', (\mathcal{F}''_t)_{0\le t\le T}, \mathbb{P}'')$ be another extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ such that there is a random variable $Y \overset{d}{\sim} N(0, I_d)$, with the d-dimensional identity matrix I_d, which is independent of $\mathcal{F}$ and such that Y is $\mathcal{F}''_0$-measurable. On this space the process $(X_t + Y)_{0\le t\le T}$ satisfies Assumption (X0). Without loss of generality $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{0\le t\le T}, \mathbb{P}')$ also extends this space. Theorem 3.1.5 yields $\mathbb{E}[Ug(F_n(X, Y))] \to \mathbb{E}[Ug(F(X, \sigma, \widetilde{W}, Y))]$ for all continuous and bounded g and all $\mathcal{F}''$-measurable random variables U. By independence of Y and $\mathcal{F}$ this holds in particular for all $\mathcal{F}$-measurable U independent of Y.

By a coupling argument (cf. Kallenberg (2002, Corollary 6.12)) there are (again on another extension of the probability space) $\widetilde{X}$, $\tilde{Y}$, $\tilde{\sigma}$, B, $\tilde{U}$ with $(X, \sigma, \widetilde{W}, Y, U) \overset{d}{\sim} (\widetilde{X}, \tilde{\sigma}, B, \tilde{Y}, \tilde{U})$ such that $\tilde{Y}$ is independent of $(\widetilde{X}, \tilde{\sigma}, B, \tilde{U})$ and $(F_n(\widetilde{X}, \tilde{Y}), \tilde{U}) \to (F(\widetilde{X}, \tilde{\sigma}, B, \tilde{Y}), \tilde{U})$ almost surely. By conditioning on $\tilde{Y} = x_0$ and using independence this implies that $\mathbb{E}[Ug(F_n(X, x_0))] \to \mathbb{E}[Ug(F(X, \sigma, \widetilde{W}, x_0))]$ for $\mathbb{P}^{\tilde{Y}}$-almost all x_0 (by dominated convergence for conditional expectations, cf. Kallenberg (2002, Theorem 6.1)). Since $\tilde{Y} \overset{d}{\sim} Y \overset{d}{\sim} N(0,1)$, this holds for almost all x_0. In particular, this holds for all $g \in C_c(\mathbb{R}^d)$, i.e. all continuous functions with compact support. Since this space is separable and because $\mathcal{F}$ is separable, this implies the claim (cf. Theorem Kallenberg (2002, 5.19)). □

3.5.2. Proofs of Section 2

Observe first the following lemma, which will be used frequently.

Lemma 3.5.11. *Let $\alpha, \beta \in \mathbb{R}$. It follows that*

$$\sum_{k-1>j\geq 2}^{n} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} (b-a)^{-\alpha} a^{-\beta} dadb \lesssim \begin{cases} \log^2 n, & \alpha = 1, \beta = 1, \\ T^{2-\alpha-\beta} \log^{\beta} n, & \alpha < 1, \beta \geq 1, \\ T^{2-\alpha-\beta}, & \alpha < 1, \beta < 1, \\ \Delta_n^{2-\alpha-\beta}, & \alpha > 1, \beta > 1, \\ T^{1-\beta} \log n, & \alpha = 1, \beta < 1. \end{cases}$$

Proof. The expression in the statement is equal to $\Delta_n^{2-\alpha-\beta} \sum_{k-1>j\geq 2}^{n} (k-1-j)^{-\alpha} \int_{j-1}^{j} a^{-\beta} da$, which is bounded by $\Delta_n^{2-\alpha-\beta} (\sum_{k=1}^{n} k^{-\alpha})(\int_1^n a^{-\beta} da)$. If $\alpha = 1$, then the sum is of order $\log n$, while it is of order $n^{1-\alpha}$ when $\alpha < 1$ and just finite when $\alpha > 1$. The same statements hold for the integral, depending on β. Considering all possible combinations yields the claim. □

Proof of Proposition 3.2.1

Proof. Write $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2 = A_1 + A_2 + A_3$, where $A_1 = \sum_{|k-j|\leq 1} M_{k,j}$, $A_2 = 2\sum_{k-1>j\geq 2} M_{k,j}$ and $A_3 = 2\sum_{k>2} M_{k,1}$ and where

$$M_{k,j} = \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \mathbb{E}\left[\left(f\left(X_r\right) - f\left(X_{t_{k-1}}\right)\right)\left(f\left(X_h\right) - f\left(X_{t_{j-1}}\right)\right)\right] dhdr.$$

Applying the Cauchy-Schwarz inequality several times yields $A_1 + A_2 + A_3 \lesssim S_1 + S_2$, where $S_1 = \Delta_n \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \mathbb{E}[(f(X_r) - f(X_{t_{k-1}}))^2] dr$ and $S_2 = \sum_{k-1>j\geq 2} |M_{k,j}|$. It follows that

$$S_1 = \Delta_n \int \left(f\left(y\right) - f\left(x\right)\right)^2 \left(\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} p_{t_{k-1},r}\left(x,y\right) dr\right) d\left(x,y\right).$$

The following idea generalizes Equation (8) of Ganychenko (2015) to arbitrary processes. For (i) consider $t_{j-1} < h < t_j < t_{k-1} < r < t_k$ and let $g_{h,t_{j-1},b}(x,y) = p_{h,b}\left(x,y\right) - p_{t_{j-1},b}\left(x,y\right)$. The Fubini theorem implies for bounded f with compact support that $M_{k,j}$ is equal to

$$\int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \int_{t_{k-1}}^{r} \int f\left(x\right) f\left(y\right) \partial_b g_{h,t_{j-1},b}\left(x,y\right) d\left(x,y\right) dbdhdr.$$

By interchanging integration and differentiation the inner integral is equal to $\partial_b(\int f(x)f(y) g_{h,t_{j-1},b}(x,y) d(x,y))$. Observe that $\int g_{h,t_{j-1},b}(x,y) dy$ is independent of b. Consequently, $\partial_b(\int f^2(x) g_{h,t_{j-1},b}(x,y) d(x,y)) = 0$. This holds similarly if $f^2(x)$ is replaced by $f^2(y)$, because $\int g_{h,t_{j-1},b}(x,y) dx = 0$. It follows that $M_{k,j}$ is equal to

$$-\frac{1}{2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \int_{t_{k-1}}^{r} \int \left(f\left(y\right) - f\left(x\right)\right)^2 \partial_b g_{h,t_{j-1},b}\left(x,y\right) d\left(x,y\right) dbdhdr$$

and S_2 is up to a constant bounded by

$$\Delta_n \int \left(f\left(y\right)-f\left(x\right)\right)^2 \left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\partial_r g_{h,t_{j-1},r}\left(x,y\right)\right| dhdr\right) d\left(x,y\right).$$

Together with the bound for S_1 this yields (i). For (ii) it follows similarly that $M_{k,j}$ is equal to

$$-\frac{1}{2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\int_{t_{k-1}}^{r}\int_{t_{j-1}}^{h}\left(\int\left(f\left(y\right)-f\left(x\right)\right)^2\partial_{ab}^2 p_{a,b}\left(x,y\right)d\left(x,y\right)\right) dadbdhdr.$$

(ii) follows from the bound on S_1 and because S_2 is up to a constant bounded by

$$\Delta_n^2 \int \left(f\left(y\right)-f\left(x\right)\right)^2 \left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\partial_{hr}^2 p_{h,r}\left(x,y\right)\right| dhdr\right) d(x,y).$$

□

Proof of Proposition 3.2.2

Proof. As in the proof of Proposition 3.2.1 it is sufficient to bound S_1+S_2. For $f\in\mathcal{S}(\mathbb{R}^d)$ we can write $f(X_r)=(2\pi)^{-d}\int\mathcal{F}f(u)e^{-i\langle u,X_r\rangle}du$ for all $0<r<T$. It follows that $\mathbb{E}[(f(X_r)-f(X_{t_{k-1}})(f(X_h)-f(X_{t_{j-1}}))]$ is equal to

$$\begin{aligned}(2\pi)^{-2d}\int \mathcal{F}f\left(u\right)\mathcal{F}f\left(v\right)\mathbb{E}\Big[&\left(e^{-i\langle v,X_r\rangle}-e^{-i\langle v,X_{t_{k-1}}\rangle}\right)\\ &\cdot\left(e^{-i\langle u,X_h\rangle}-e^{-i\langle u,X_{t_{j-1}}\rangle}\right)\Big]d\left(u,v\right).\end{aligned}$$

With $\varphi_{h,h}(u,v)=\mathbb{E}[e^{i\langle u+v,X_h\rangle}]$ the expectation is for *all* h,r,t_{k-1},t_{j-1} equal to

$$\overline{\varphi_{h,r}\left(u,v\right)}-\overline{\varphi_{t_{j-1},r}\left(u,v\right)}-\overline{\varphi_{h,t_{k-1}}\left(u,v\right)}+\overline{\varphi_{t_{j-1},t_{k-1}}\left(u,v\right)}. \tag{3.5.20}$$

For (i) this implies by symmetry in u,v that S_1 is up to a constant bounded by

$$\Delta_n\int\left|\mathcal{F}f\left(u\right)\right|\left|\mathcal{F}f\left(v\right)\right|\left(\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k} g_{t_{k-1},r}\left(u,v\right)dr\right)d\left(u,v\right) \tag{3.5.21}$$

with $g_{t_{k-1},r}(u,v)$ as in the statement. Let $\tilde{g}_{h,t_{j-1},b}(u,v)=\partial_b\varphi_{h,b}(u,v)-\partial_b\varphi_{t_{j-1},b}(u,v)$. Then (3.5.20) is for $t_{j-1}<h<t_j<t_{k-1}<r<t_k$ equal to $\int_{t_{k-1}}^{r}\tilde{g}_{h,t_{j-1},b}(u,v)db$. Therefore S_2 is up to a constant bounded by

$$\Delta_n\int\left|\mathcal{F}f\left(u\right)\right|\left|\mathcal{F}f\left(v\right)\right|\left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\tilde{g}_{h,t_{j-1},r}\left(u,v\right)\right| dhdr\right)d\left(u,v\right).$$

This yields (i). With respect to (ii) note that the last argument also applies to $r = h$, $k = j$ such that (3.5.21) is bounded by

$$\Delta_n \int |\mathcal{F}f(u)| |\mathcal{F}f(v)| \left(\sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t_k} \left| \tilde{g}_{h,t_{k-1},r}(u,v) \right| dhdr \right) d(u,v),$$

giving a bound on S_1. For S_2 note that (3.5.20) is equal to $\int_{t_{k-1}}^{r} \int_{t_{j-1}}^{h} \partial_{ab}^2 \varphi_{a,b}(u,v) dadb$. This yields (ii), because S_2 is up to a constant bounded by

$$\Delta_n^2 \int |\mathcal{F}f(u)| |\mathcal{F}f(v)| \left(\sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left| \partial_{hr}^2 \varphi_{h,r}(x,y) \right| dhdr \right) d(u,v).$$

□

Proof of Theorem 3.2.4

Proof. If f is bounded, then $f_m(x) = f(x)\mathbf{1}_{\{\|x\|\leq m\}}$ defines a sequence of bounded functions with compact support converging to f pointwise with $\|f_m\|_\infty \leq \|f\|_\infty$ for all m. If f is Hölder-continuous, then we can similarly find a sequence $(f_m)_{m\geq 1} \subset C_c^\infty(\mathbb{R}^d)$ converging to f pointwise with $\|f_m\|_{C^s} \lesssim \|f\|_{C^s}$. In both cases it follows $\mathbb{P}_{x_0}$ almost surely that $\Gamma_T(f_m) - \widehat{\Gamma}_{n,T}(f_m) \to \Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)$ as $m \to \infty$ by dominated convergence. The lemma of Fatou implies

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P}_{x_0})} \leq \liminf_{m\to\infty} \|\Gamma_T(f_m) - \widehat{\Gamma}_{n,T}(f_m)\|^2_{L^2(\mathbb{P}_{x_0})}.$$

It is therefore sufficient to prove the theorem for bounded f with compact support.

Conditional on x_0 the random variables (X_h, X_r), $h \neq r$, have the joint densities $p_{h,r}(x,y;x_0) = \xi_{0,r}(x_0,x)\xi_{h,r}(x,y)$, $x, y \in \mathbb{R}^d$. Moreover, the heat kernel bounds in Assumption 3.2.3 imply

$$\begin{aligned} |p_{h,r}(x,y;x_0)| &\leq q_{r-h}(y-x)\, q_h(x-x_0), \\ |\partial_r p_{h,r}(x,y;x_0)| &\leq \frac{1}{r-h} q_{r-h}(y-x)\, q_h(x-x_0), \\ \left|\partial_{hr}^2 p_{h,r}(x,y;x_0)\right| &\leq \left(\frac{1}{(r-h)^2} + \frac{1}{(r-h)h} \right) q_h(x-x_0)\, q_{r-h}(y-x). \end{aligned}$$

Then $\int (\sum_{k=1}^n \int_{t_{k-1}}^{t_k} p_{t_{k-1},r}(x,y;x_0)dr) d(x,y) = T$ and Lemma 3.5.11 yields

$$\begin{aligned} &\int \left(\sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \Big(|\partial_r p_{h,r}(x,y;x_0)| + |\partial_r p_{t_{j-1},r}(x,y;x_0)| \Big) dhdr \right) d(x,y) \\ &\lesssim \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} (r-h)^{-1} dhdr \lesssim T \log n. \end{aligned}$$

Applying Proposition 3.2.1(i) to $p_{h,r}(\cdot;x_0)$ yields the claim in (i) for bounded f. For (ii), on the other hand, the moment conditions on q_a imply that $\int \|y-x\|^{2s} q_a(x -$

$x_0)q_{b-a}(y-x)d(x,y) \lesssim (b-a)^{2s/\gamma}$ for $0 < s \leq \gamma/2$. Consequently, Lemma 3.5.11 yields for $\Delta_n^{-1}\int\|y-x\|^{2s}(\sum_{k=1}^n\int_{t_{k-1}}^{t_k}p_{t_{k-1},r}(x,y;x_0)dr)d(x,y)$ up to a constant the upper bound $\Delta_n^{-1}\sum_{k=1}^n\int_{t_{k-1}}^{t_k}(r-t_{k-1})^{2s/\gamma}dr$ and also

$$\begin{aligned}&\int\|y-x\|^{2s}\Big(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\partial_{hr}^2 p_{h,r}(x,y;x_0)\right|dhdr\Big)d(x,y)\\ &\lesssim \sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\Big((r-h)^{2s/\gamma-2}+(r-h)^{2s/\gamma-1}h^{-1}\Big)dhdr.\end{aligned}$$

For $2s/\gamma < 1$ Lemma 3.5.11 implies for the sum of these two upper bounds the order $O(T\Delta_n^{2s/\gamma-1}+T^{2s/\gamma}\log n)$, while it is $O(T\log n)$ for $2s/\gamma = 1$. In the first case note that

$$T^{2s/\gamma}\log n = T\Delta_n^{2s/\gamma-1}\Big(T^{2s/\gamma-1}\Delta^{1-2s/\gamma}\Big)\log n \leq T\Delta_n^{2s/\gamma-1}\frac{\log n}{n^{1-2s/\gamma}},$$

which is of order $O(T\Delta_n^{1+2s/\gamma})$, i.e. there is no $\log n$-term. This implies (ii) for $f \in C^s(\mathbb{R}^d)$. □

Proof of Theorem 3.2.6

Proof. Note that $L^2(\mathbb{R}^d) = H^0(\mathbb{R}^d)$. For $f \in H^s(\mathbb{R}^d)$, $0 \leq s \leq 1$, let $(f_m)_{m\geq 1} \subset C_c^\infty(\mathbb{R}^d)$ be a sequence of functions converging to f with respect to $\|\cdot\|_{H^s}$ with $\|f_m\|_{H^s} \leq \|f\|_{H^s}$. Then $\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2$ is bounded by

$$2\|\Gamma_T(f-f_m)-\widehat{\Gamma}_{n,T}(f-f_m)\|_{L^2(\mathbb{P})}^2 + 2\|\Gamma_T(f_m)-\widehat{\Gamma}_{n,T}(f_m)\|_{L^2(\mathbb{P})}^2.$$

Then $\|\Gamma_T(f-f_m)\|_{L^2(\mathbb{P})} \lesssim \int_0^T\mathbb{E}[\int(f(x)-f_m(x))^2p_r(x)dx]^{1/2}dr$, where the marginal densities p_r satisfy $\sup_{0\leq r\leq T}|p_r(x)| = \sup_{0\leq r\leq T}|\int\xi_{0,r}(x_0,x)\mu(x_0)dx_0| \leq \|\mu\|_\infty$ by Assumption A. It follows that $\|\Gamma_T(f-f_m)\|_{L^2(\mathbb{P})}$ is up to a constant bounded by $\|f-f_m\|_{L^2}$, which converges to 0 as $m\to\infty$. A similar argument shows $\|\widehat{\Gamma}_{n,T}(f-f_m)\|_{L^2(\mathbb{P})} \to 0$ as $m \to \infty$. It is therefore sufficient to prove the theorem for $f \in C_c^\infty(\mathbb{R}^d)$.

The random variables (X_h, X_r), $h \neq r$, have the joint densities $p_{h,r}(x,y) = p_r(x)\xi_{h,r}(x,y)$, $x,y\in\mathbb{R}^d$ and the heat kernel bounds in Assumption 3.2.3 imply

$$\begin{aligned}|p_{h,r}(x,y)| &\leq \|\mu\|_\infty q_{r-h}(y-x),\\ |\partial_r p_{h,r}(x,y)| &\leq \|\mu\|_\infty\frac{1}{r-h}q_{r-h}(y-x),\\ \left|\partial_{hr}^2 p_{h,r}(x,y)\right| &\leq \|\mu\|_\infty\left(\frac{1}{(r-h)^2}+\frac{1}{(r-h)h}\right)q_{r-h}(y-x).\end{aligned}$$

Then $\int f^2(x)(\sum_{k=1}^n\int_{t_{k-1}}^{t_k}p_{t_{k-1},r}(x,y)dr)d(x,y) \lesssim \|\mu\|_\infty\|f\|_{L^2}^2T$ and it follows by Lemma 3.5.11 that

$$\begin{aligned}&\int f^2(x)\left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\Big(|\partial_r p_{h,r}(x,y)|+|\partial_r p_{t_{j-1},r}(x,y)|\Big)dhdr\right)d(x,y)\\ &\lesssim \|\mu\|_\infty\|f\|_{L^2}^2\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}(r-h)^{-1}dhdr \lesssim \|\mu\|_\infty\|f\|_{L^2}^2T\log n.\end{aligned}$$

By symmetry the same holds with $f^2(y)$ instead of $f^2(x)$. Applying Proposition 3.2.1(i) along with the trivial bound $(f(x)-f(y))^2 \leq 2f(x)^2+2f(y)^2$ therefore yields (i). For (ii) we distinguish the cases $\gamma<2$ and $\gamma=2$. Let first $0<s\leq\gamma/2<1$. In this case, the L^2-Sobolev norm defined via the Fourier transform is equivalent to the Slobodeckij-norm

$$\|f\|_{\tilde{H}^s} = \left(\|f\|_{L^2}^2 + \int \frac{(f(x)-f(y))^2}{\|x-y\|^{2s+d}} d(x,y)\right)^{1/2}, \tag{3.5.22}$$

(cf. Di et al. (2012) for more details). Similar to the proof of Theorem 3.2.6 the moment conditions on q_a imply for $0<s\leq\gamma/2$ that

$$\begin{aligned}
&\Delta_n^{-1}\int (f(y)-f(x))^2\left(\sum_{k=1}^n\int_{t_{k-1}}^{t_k} p_{t_{k-1},r}(x,y)dr\right)d(x,y)\\
&\quad\leq \|f\|_{H^s}^2\Delta_n^{-1}\sup_{x,y\in\mathbb{R}^d}\left(\sum_{k=1}^n\int_{t_{k-1}}^{t_k}\|y-x\|^{2s+d}p_{t_{k-1},r}(x,y)dr\right)\\
&\quad\lesssim \|f\|_{H^s}^2\Delta_n^{-1}\left(\sum_{k=1}^n\int_{t_{k-1}}^{t_k}(r-t_{k-1})^{2s/\gamma}\,dr\right),\\
&\int (f(y)-f(x))^2\left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\partial_{hr}^2 p_{h,r}(x,y)\right|dhdr\right)d(x,y)\\
&\quad\leq \|f\|_{H^s}^2\sup_{x,y\in\mathbb{R}^d}\left(\|y-x\|^{2s+d}\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}\left|\partial_{hr}^2 p_{h,r}(x,y)\right|dhdr\right)\\
&\quad\lesssim \|f\|_{H^s}^2\left(\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}(r-h)^{2s/\gamma-1}h^{-1}dhdr\right).
\end{aligned}$$

We surprisingly recover the same upper bounds as in the proof of Theorem 3.2.6. This yields the claim in (ii) for $0<s\leq\gamma/2<1$. Consider now $\gamma=2$ and $0<s\leq 1$. Unfortunately, the Slobodeckij-norm is not equivalent to the $\|\cdot\|_{H^s}$-norm when $s=1$. We already know from (i) that the operator $\Gamma_T-\widehat{\Gamma}_{n,T}$ is a continuous linear operator from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{P})$. It is therefore sufficient to show that it is also a continuous linear operator from $H^1(\mathbb{R}^d)$ to $L^2(\mathbb{P})$. Indeed, as the Sobolev spaces $H^s(\mathbb{R}^d)$ for $0\leq s\leq 1$ form interpolation spaces, the general claim is obtained by interpolating the operator norms of $\Gamma_T-\widehat{\Gamma}_{n,T}$ for $s=0$ and $s=1$ (cf. Adams and Fournier (2003, Theorem 7.23)). For $s=1$ we have $f(y)-f(x)=\int_0^1\langle\nabla f(x+t(y-x)),y-x\rangle\,dt$. It follows for any $0<h<r<T$ that

$$\begin{aligned}
&\int (f(y)-f(x))^2 q_{r-h}(y-x)\,d(x,y)\\
&\leq \int_0^1\left(\int\|\nabla f(x+t(y-x))\|^2\|y-x\|^2 q_{r-h}(y-x)\,d(x,y)\,dt\right)\\
&= \int\|\nabla f(x+tz)\|^2\|z\|^2 q_{r-h}(z)\,d(x,z)\lesssim\|f\|_{H^1}^2(r-h),
\end{aligned}$$

using $\int \|x\|^2 q_a(x)dx \lesssim a$. Proposition 3.2.1$(ii)$ therefore implies

$$\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|^2_{L^2(\mathbb{P})} \lesssim \|\mu\|_\infty \|f\|^2_{H^1} \Bigg(\Delta_n \sum_{k=1}^n \int_{t_{k-1}}^{t_k} (r - t_{k-1})\, dr + \Delta_n^2 \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left((r-h)^{-2} + h^{-1} \right) dh dr \Bigg).$$

Using the bounds from above yields the claim in (ii) for $s = 1$. □

Proof of Theorem 3.2.8

Y is independent of $\mathcal{F}_0$ and thus of X_0. Therefore the characteristic function of (X_h, X_r) at $(u,v) \in \mathbb{R}^{2d}$ for $0 \leq h < r \leq T$ is $\varphi_{h,r}(u,v) = \tilde{\varphi}_{h,r}(u,v)\mathcal{F}\mu(u+v)$, where $\tilde{\varphi}_{h,r}(u,v) = e^{\psi_{h,r}(v)+\psi_{0,h}(u+v)}$ is the characteristic function of (Y_h, Y_r). $\psi_{h,r}(u)$ is for almost all r differentiable with

$$\partial_r \psi_{h,r}(u) = i\langle u, b_r\rangle - \frac{1}{2}\|\sigma_r^\top u\|^2 + \int \left(e^{i\langle u,x\rangle} - 1 - i\langle u,x\rangle \mathbf{1}_{\{\|x\|\leq 1\}} \right) dF_r(x),$$

and also $\partial^2_{hr}\psi_{h,r}(u) = 0$. Hence

$$\partial_r \varphi_{h,r}(u,v) = \partial_r \psi_{h,r}(v)\, \tilde{\varphi}_{h,r}(u,v)\mathcal{F}\mu(u+v), \tag{3.5.23}$$
$$\partial^2_{hr}\varphi_{h,r}(u,v) = (\partial_h \psi_{h,r}(v) + \partial_h \psi_{0,h}(u+v))\, \partial_r \psi_{h,r}(v)\, \tilde{\varphi}_{h,r}(u,v)\mathcal{F}\mu(u+v).$$

$\varphi_{h,r}$ as well as the derivatives $\partial_r\varphi_{h,r}$ and $\partial^2_{hr}\varphi_{h,r}$ satisfy the assumptions of Proposition 3.2.2(i) and (ii). Consider first the following lemma.

Lemma 3.5.12. *Fix $u, v \in \mathbb{R}^d$ such that $v \neq 0$ and $\|u+v\| \neq 0$ and let*

$$U_n = \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left(|\tilde{\varphi}_{h,r}(u,v)| + |\tilde{\varphi}_{t_{j-1},r}(u,v)| \right) dh dr,$$
$$V_n = \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t_k} \left(|\tilde{\varphi}_{h,r}(u,v)| + |\tilde{\varphi}_{t_{k-1},r}(u,v)| \right) dh dr.$$

Then we have the following under the assumptions of Theorem 3.2.8(i):

(i) $(1+\|v\|)^{\gamma+\beta^*} U_n \lesssim T^2 (1+\|v\|)^{\beta^*/2}(1+\|u\|)^{\beta^*/2}$.

(ii) $(1+\|v\|)^{\gamma+\beta^*} V_n \lesssim T\Delta_n (1+\|v\|)^{\gamma/2+\beta^*} (1+\|u\|)^{\gamma/2+\beta^*}$.

(iii) $((1+\|v\|)^{2\gamma+2\beta^*} + (1+\|v\|)^{\gamma+\beta^*}(1+\|u+v\|)^{\gamma+\beta^*})U_n \lesssim T^2(1+\|v\|)^{\gamma/2+\beta^*}(1+\|u\|)^{\gamma/2+\beta^*}$.

Proof. Observe first the following estimates:

$$\sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} |\tilde{\varphi}_{h,r}(u,v)|\, dh dr$$
$$\lesssim \int_0^T \int_0^T e^{-c\|v\|^\gamma |r-h| - c\|u+v\|^\gamma (h\wedge r)} dh dr$$
$$\lesssim \|v\|^{-\gamma} \int_0^T \left(1 - e^{-c\|v\|^\gamma (T-h)} \right) e^{-c\|u+v\|^\gamma h} dh \lesssim \begin{cases} \|v\|^{-\gamma} T, \\ \|v\|^{-\gamma}\|u+v\|^{-\gamma}. \end{cases} \tag{3.5.24}$$

The same holds when $\tilde{\varphi}_{h,r}(u,v)$ is replaced by $\tilde{\varphi}_{t_{j-1},r}(u,v)$. Let first $\|v\|, \|u+v\| \geq 1$. Then $(1+\|v\|)^{\gamma+\beta^*} \leq \|v\|^{\gamma+\beta^*/2}\|u\|^{\beta^*/2} + \|v\|^{\gamma+\beta^*/2}\|u+v\|^{\beta^*/2}$ and the last display, together with $T \geq 1$ and $\beta^*/2 \leq \gamma$, yields (i). The same is true if $\|u+v\| \leq 1$, as $\|u+v\|^{\beta^*/2} \leq 1$. If $\|v\| < 1$, then (i) holds trivially, because $|\tilde{\varphi}_{h,r}(u,v)| \leq 1$. Observe next that V_n is bounded by

$$2\Delta_n^2 \sum_{k=1}^{n} e^{-c\|u+v\|^\gamma t_{k-1}} \lesssim \Delta_n \int_0^T e^{-c\|u+v\|^\gamma h} dh \lesssim \begin{cases} T\Delta_n, \\ \Delta_n \|u+v\|^{-\gamma}. \end{cases} \tag{3.5.25}$$

Let first $\|u+v\| \geq 1$. Then $(1+\|v\|)^{\gamma+\beta^*} \leq \|v\|^{\gamma/2+\beta^*/2}\|u\|^{\gamma/2+\beta^*/2} + \|v\|^{\gamma/2+\beta^*/2}\|u+v\|^{\gamma/2+\beta^*/2}$. The last display, together with $T \geq 1$ and $\beta^*/2 \leq \gamma$, yields (ii). Again, this remains true if $\|u+v\| < 1$. With respect to (iii) let $\|v\|, \|u+v\| \geq 1$. Then it follows from $\|u+v\|^{\beta^*} \lesssim \|u\|^{\beta^*} + \|v\|^{\beta^*}$ that

$$\begin{aligned} &\|v\|^{2\gamma+2\beta^*} + \|v\|^{\gamma+\beta^*}\|u+v\|^{\gamma+\beta^*} \\ &\leq \|v\|^{3/2\gamma+\beta^*}\|u\|^{\gamma/2+\beta^*} + \|v\|^{3/2\gamma+\beta^*}\|u+v\|^{\gamma/2+\beta^*} + \|v\|^{\gamma+2\beta^*}\|u+v\|^{\gamma} \\ &\quad + \|u+v\|^{\gamma}\|u\|^{\beta^*}\|v\|^{\gamma+\beta^*}. \end{aligned}$$

(3.5.24) together with $\beta^*/2 \leq \gamma$ implies (iii). The same holds when $\|u+v\| < 1$ as before. For $\|v\| < 1$ the trivial bound from above applies. □

Proof of Theorem 3.2.8. Since Y and X_0 are independent, the marginals X_r have uniformly bounded densities $p_r(x) \leq \|\mu\|_\infty$, $x \in \mathbb{R}^d$, even if the distributions of Y_r have no densities. By the argument at the beginning of the proof of Theorem 3.2.6 it is therefore enough to show the claim for $f \in \mathcal{S}(\mathbb{R}^d)$.

Consider first the claim in (i). We only have to show it for $s = \beta^*/2$ and $s = \gamma/2+\beta^*$. As in the proof of Theorem 3.2.6, the general claim for $\beta^*/2 \leq s \leq \gamma/2+\beta^*$ follows by interpolation. Let $u, v \in \mathbb{R}^d$. Then for any $0 \leq h, r \leq T$ it holds $|g_{h,r}(u,v)| \lesssim |\mathcal{F}\mu(u+v)|$ with g from Proposition 3.2.2(i). Moreover, by assumption $|\partial_r \psi_{h,r}(v)| \leq c(1+\|v\|)^{\gamma+\beta^*}$. Lemma 3.5.12$(i)$ and Proposition 3.2.2(i) therefore imply that $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2$ is up to a constant bounded by

$$T^2\Delta_n \int |\mathcal{F}f(u)|\,|\mathcal{F}f\,(v)|\,(1+\|u\|)^{\beta^*/2}(1+\|v\|)^{\beta^*/2}\,|\mathcal{F}\mu(u+v)|\,d\,(u,v)\,. \tag{3.5.26}$$

Lemma 3.5.3 shows for this the upper bound $T^2\Delta_n\|f\|_{H^s}^2$, implying the claim for $s = \beta^*/2$. With respect to $s = \gamma/2+\beta^*$ it follows similarly by Lemma 3.5.12(ii) and (iii), Proposition 3.2.2(ii) and Lemma 3.5.3 that $\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2$ is up to a constant bounded by $T^2\Delta_n\|f\|_{H^s}^2$. This is the claimed bound for $s = \gamma/2+\beta^*$. To see that the improved bound holds note that $|\partial_r \psi_{h,r}(v)| \leq c\|v\|^{\gamma+\beta_r}$ simplifies the calculations in Lemma 3.5.12, since there is no need to distinguish the cases $\|v\| \geq 1$ or $\|v\| < 1$.

At last, consider (ii). From $|\partial_r \psi_{h,r}(v)| \lesssim 1$ it follows immediately that $\varphi_{h,r}(u,v)$ and the time derivatives $\partial_r \varphi_{h,r}(u,v)$, $\partial_{hr}^2 \varphi_{h,r}(u,v)$ are bounded by $T^2|\mathcal{F}\mu(u+v)|$. As $T \geq 1$, Proposition 3.2.2(ii) and Lemma 3.5.3 imply the claim. If $c_1\rho(v) \leq \partial_r \psi_{h,r}(v) \leq c_2\rho(v) \leq 0$ for all $0 \leq h, r \leq T$, then $|\tilde{\varphi}_{h,r}(u,v)| \leq e^{-cc_2\rho(v)|r-h|-cc_2\rho(u+v)(r\wedge h)}$ and

$\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t_k}|\partial_r\tilde{\varphi}_{h,r}(u,v)|dhdr$ is up to a constant bounded by

$$\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\int_{h}^{t_k}\left(-\rho\left(v\right)\right)e^{-cc_2\rho(v)(r-h)}drdh\lesssim\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\left(e^{-cc_2\rho(v)(t_k-h)}-1\right)dh,$$

and similarly for $\partial_r\tilde{\varphi}_{t_{k-1},r}(u,v)$, while $\sum_{k-1>j\geq 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}|\partial^2_{hr}\tilde{\varphi}_{h,r}(u,v)|dhdr$ is up to a constant bounded by $\int_0^T\int_h^T(-\rho(v))e^{-cc_2\rho(v)(r-h)}drdh$. The first expression is of order $O(T\Delta_n)$ and the second one of order $O(T)$. Again, the claim follows from Proposition 3.2.2(ii) and Lemma 3.5.3. □

Remark 3.5.13. If $d=1$ and $\gamma>1$, $\beta^*=0$, then the proof applies to $X_t=Y_t$. Indeed, replace T by $\int_0^T e^{-c\|u+v\|^2h}dh$ in (3.5.24) and (3.5.25). Together with a slightly different argument for $\|v\|<1$ this yields e.g. instead of (3.5.26) the bound

$$\begin{aligned}&T\Delta_n\int_0^T\int|\mathcal{F}f(u)|\,|\mathcal{F}f\left(v\right)|\,e^{-c\|u+v\|^2h}d\left(u,v\right)dh\\&\leq T\Delta_n\int_0^T\int|\mathcal{F}f(u)|^2\,e^{-c\|u+v\|^2h}d\left(u,v\right)dh\lesssim\|f\|^2_{H^s}T^2\Delta_n.\end{aligned}$$

This works, because $u\mapsto e^{-c\|u\|^\gamma h}$ is integrable and because $\int_0^T h^{-1/\gamma}dh$ is finite.

Proof of Theorem 3.2.13

The characteristic function of (X_h,X_r) at $(u,v)\in\mathbb{R}^{2d}$ for $0\leq h<r\leq T$ is $\varphi_{h,r}(u,v)=\tilde{\varphi}_{h,r}(u,v)\mathcal{F}\mu(u+v)$, where $\tilde{\varphi}_{h,r}(u,v)$ is the characteristic function of (B_h,B_r). As B is a Gaussian process, it follows that $\tilde{\varphi}_{h,r}(u,v)$ is equal to $e^{-\frac{1}{2}\Phi_{h,r}(u,v)}$ with $\Phi_{h,r}(u,v)=\|u\|^2h^{2H}+\|v\|^2r^{2H}+2\left\langle u,v\right\rangle c\left(h,r\right)$. Since fractional Brownian motion is *locally nondeterministic* (cf. Xiao (2006)), there exist constants $\delta,c_2>0$ such that for $b-a\leq\delta$, $0\leq a<b$,

$$\mathrm{Var}\left(\left\langle v,B_b-B_a\right\rangle+\left\langle u,B_a\right\rangle\right)\geq c_2\left(\|v\|^2\sigma^2(a,b)+\|u\|^2\sigma^2\left(0,a\right)\right)$$

with $\sigma^2(a,b)=\mathbb{E}[(B_b^{(1)}-B_a^{(1)})^2]=(b-a)^{2H}$. By self-similarity of B this holds for arbitrary $0\leq a<b$. Therefore

$$\begin{aligned}\Phi_{h,r}\left(u,v\right)&=\mathrm{Var}\left(\left\langle v,B_r\right\rangle+\left\langle u,B_h\right\rangle\right)=\mathrm{Var}\left(\left\langle v,B_r-B_h\right\rangle+\left\langle u+v,B_h\right\rangle\right)\\&\geq c_2\left(\|v\|^2\left(r-h\right)^{2H}+\|u+v\|^2h^{2H}\right)\end{aligned}$$

and $\tilde{\varphi}_{h,r}(u,v)\leq e^{-c\|v\|^2(r-h)^{2H}-c\|u+v\|^2h^{2H}}$. Moreover,

$$\begin{aligned}\partial_r\varphi_{h,r}(u,v)&=-\frac{1}{2}\partial_r\Phi_{h,r}(u,v)\varphi_{h,r}(u,v),\\\partial^2_{hr}\varphi_{h,r}\left(u,v\right)&=\left(-\frac{1}{2}\partial^2_{hr}\Phi_{h,r}\left(u,v\right)+\frac{1}{4}\partial_r\Phi_{h,r}\left(u,v\right)\partial_h\Phi_{h,r}\left(u,v\right)\right)\varphi_{h,r}\left(u,v\right),\\\partial_r\Phi_{h,r}\left(u,v\right)&=2H(\|v\|^2+\left\langle u,v\right\rangle)r^{2H-1}-2H\left\langle u,v\right\rangle\left(r-h\right)^{2H-1},\\\partial_h\Phi_{h,r}\left(u,v\right)&=2H(\|u\|^2+\left\langle u,v\right\rangle)h^{2H-1}+2H\left\langle u,v\right\rangle\left(r-h\right)^{2H-1},\\\partial^2_{hr}\Phi_{h,r}\left(u,v\right)&=2H\left(2H-1\right)\left\langle u,v\right\rangle\left(r-h\right)^{2H-2}.\end{aligned}$$

We first prove a lemma. Denote for any function $(r,h) \mapsto g(r,h)$ and fixed $u, v \in \mathbb{R}^d$ by $U_n(g)$ the sum $\sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} g(r,h)\tilde{\varphi}_{h,r}(u,v)dhdr$.

Lemma 3.5.14. *Let $T \geq 1$ and assume (X0). Fix $u, v \in \mathbb{R}^d \backslash \{0\}$ and let $0 < H < 1$, $H \neq 1/2$. Consider for $0 < h < r < T$ the functions $g_1(r,h) = (r-h)^{2H-1}$, $g_2(r,h) = h^{2H-1}$, $g_3(r,h) = (r-h)^{4H-2}$, $g_4(r,h) = (r-h)^{2H-2}$, $g_5(r,h) = (r-h)^{2H-1}h^{2H-1}$, $g_6(r,h) = r^{2H-1}h^{2H-1}$ and $g_7(r,h) = (r-h)^{2H-1}r^{2H-1}$. Then we have the following estimates with absolute constants:*

(i) $(\|v\|^2 + \|v\|\|u+v\|)(U_n(g_1) + U_n(g_2)) \lesssim T$,

(ii) if $H > 1/2$, then $(\|v\|^2 + \|v\|\|u+v\|)(U_n(g_3) + U_n(g_4)) \lesssim T^{2H}$ and if $H < 1/2$, then the same expression is up to a constant bounded by $T\Delta_n^{2H-1}$,

(iii) if $H > 1/2$, then $(\|v\| + \|u+v\|)^2(U_n(g_5) + U_n(g_6) + U_n(g_7)) \lesssim T^{2H}$ and if $H < 1/2$, then the same expression is up to a constant bounded by $T\Delta_n^{2H-1}$,

(iv) $(1 + \|v\|)\sum_{k=1}^n \int_{t_{k-1}}^{t_k} \int_h^{t_k} (r^{2H-1} + (r-h)^{2H-1})\tilde{\varphi}_{h,r}(u,v)drdh \lesssim T^{2H}\Delta_n$.

Proof. We need to bound the integrals in $U_n(g_i)$ in several different ways. Observe for $0 \leq a < b \leq T$ and $q = 2H-1, 4H-2, 1$ the following estimates for $R_{a,b,v}^{(q)} := \int_a^b r^q e^{-\frac{1}{2}\|v\|^2 r^{2H}} dr$:

$$R_{a,b,v}^{(2H-1)} \lesssim \|v\|^{-2}\left(e^{-\frac{1}{2}\|v\|^2 a^{2H}} - e^{-\frac{1}{2}\|v\|^2 b^{2H}}\right) \lesssim \begin{cases} \|v\|^{-2}, \\ \|v\|^{-1}(b^{2H} - a^{2H})^{1/2}, \\ b^{2H} - a^{2H}, \end{cases} \tag{3.5.27}$$

$$R_{a,b,v}^{(4H-2)} \lesssim \begin{cases} \|v\|^{-2} \int_a^b r^{2H-2} dr \\ \|v\|^{-1} \int_a^b r^{3H-2} dr \end{cases} \lesssim \begin{cases} \|v\|^{-2}\left(b^{2H-1} - a^{2H-1}\right), \\ \|v\|^{-1}(b^{3H-1} - a^{3H-1}), \end{cases} \tag{3.5.28}$$

$$R_{a,b,v}^{(1)} \lesssim \begin{cases} \|v\|^{-1} \int_a^b r^{-H} dr \\ b - a \end{cases} \lesssim \begin{cases} \|v\|^{-1}(b^{1-H} - a^{1-H}), \\ b-a, \end{cases} \tag{3.5.29}$$

where we used that $\sup_{v\in\mathbb{R}^d} \|v\|^p r^{pH} e^{-\frac{1}{2}\|v\|^2 r^{2H}} = \sup_{x\geq 0} x e^{-\frac{1}{2}x^2} < \infty$ for any $p \geq 0$. It follows from (3.5.27) and (3.5.29) that $U_n(g_1)$ is bounded by

$$\int_0^T \int_h^T (r-h)^{2H-1} e^{-c_2\left(\|v\|^2(r-h)^{2H} + \|u+v\|^2 h^{2H}\right)} drdh \leq T \begin{cases} \|v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}. \end{cases}$$

The estimate for g_2 follows in the same way. For g_3 and $H > 1/2$ it follows similarly from (3.5.28), (3.5.29), $T \geq 1$ and Lemma 3.5.11 that

$$\begin{aligned} U_n(g_3) &\lesssim \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left\{ \begin{array}{l} \|v\|^{-2}(r-h)^{2H-2} \\ \|v\|^{-1}\|u+v\|^{-1}(r-h)^{3H-2} h^{-H} \end{array} \right\} dhdr \\ &\lesssim T^{2H} \begin{cases} \|v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}, \end{cases} \end{aligned}$$

while for $H < 1/2$

$$U_n\left(g_3\right) \lesssim T\Delta_n^{2H-1} \begin{cases} \|v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}. \end{cases}$$

The estimates for g_4 follow similarly (they are even easier). With respect to g_5 the integrals decompose and (3.5.27) and (3.5.29) yield for $U_n(g_5)$ the bound

$$R_{0,T,v}^{(2H-1)} R_{0,T,u+v}^{(2H-1)} \lesssim T^{2H} \begin{cases} \|v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}, \\ \|u+v\|^{-2}. \end{cases} \tag{3.5.30}$$

The argument for $U_n(g_7)$ is similar. For $U_n(g_6)$, on the other hand, the same equations imply for $H > 1/2$ the upper bound

$$\int_0^T \int_h^T R_{h,T,v}^{(2H-1)} h^{2H-1} e^{-c_2\|u+v\|^2 h^{2H}} dh \lesssim T^{2H} \begin{cases} \|u+v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}, \end{cases}$$

and for $H < 1/2$ by $r^{2H-1}h^{2H-1} \leq h^{4H-2}$ and Lemma 3.5.11

$$\begin{aligned} &\sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} \left\{ \begin{array}{c} \|u+v\|^{-2} h^{2H-2} \\ \|u+v\|^{-1}\|v\|^{-1}\left(r-h\right)^{-H} h^{3H-2} \\ \|v\|^{-2} r^{-2H} h^{4H-2} \end{array} \right\} dhdrdhdr \\ &\lesssim T\Delta_n^{2H-1} \begin{cases} \|u+v\|^{-2}, \\ \|v\|^{-1}\|u+v\|^{-1}, \\ \|v\|^{-2}, \end{cases} \end{aligned}$$

because $T \geq 1$ and because $1 \lesssim \log n \lesssim \Delta_n^{2H-1}$. Observe that we did not prove any bound on $\|v\|^2 U_n(g_6)$ for $H > 1/2$. For this, we need a different upper bound on $\tilde{\varphi}_{h,r}(u,v)$. If $\|u+v\| \geq \|v\|$, then $\tilde{\varphi}_{h,r}(u,v) \leq e^{-c_2\|v\|^2(r-h)^{2H} - c_2\|u+v\|^2 h^{2H}}$ is clearly bounded by $e^{-c_2\|v\|^2 h^{2H}}$. As $r^{2H-1}h^{2H-1} \lesssim (r-h)^{2H-1}h^{2H-1} + h^{4H-2}$ for $H > 1/2$, it thus follows from (3.5.30) and Lemma 3.5.11 that

$$U_n\left(g_6\right) \lesssim U_n\left(g_5\right) + \|v\|^{-2} \sum_{k-1>j\geq 2} \int_{t_{k-1}}^{t_k} \int_{t_{j-1}}^{t_j} h^{4H-2} dhdr \lesssim T^{2H}\|v\|^{-2}.$$

If $\|u+v\| < \|v\|$, however, then $\tilde{\varphi}_{h,r}(u,v) \leq e^{-c_2(\|v\|^2 r^{2H} + \|u\|^2 h^{2H})}$. To see why this holds note that in this case necessarily $\langle u, v\rangle \geq 0$ by elementary geometric considerations. But then $\Phi_{h,r}\left(u,v\right) \geq \|u\|^2 h^{2H} + \|v\|^2 r^{2H}$, since also $c(h,r) = \mathbb{E}[(Y_r - Y_h)Y_h] + h^{2H} \geq 0$ (recall that increments of fractional Brownian motion are positively correlated when $H > 1/2$). From the new bound and (3.5.27) follows immediately that

$$U_n\left(g_6\right) \lesssim \int_0^T \int_h^T R_{h,T,v}^{(2H-1)} h^{2H-1} e^{-c_2\|u\|^2 h^{2H}} dh \lesssim T^{2H}\|v\|^{-2}.$$

Finally, with respect to (iv), (3.5.27) yields

$$\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\int_{h}^{t_k}(r^{2H-1}+(r-h)^{2H-1})\tilde{\varphi}_{h,r}(u,v)drdh \lesssim T^{2H}\Delta_n.$$

Arguing as for $U_n(g_6)$ with the different upper bounds for $\tilde{\varphi}_{h,r}(u,v)$, it follows that the left hand side is bounded by $\|v\|^{-1}T^{2H}\Delta_n$. This yields (iv). □

Proof of Theorem 3.2.13. As in the proof of Theorem 3.2.8 it is sufficient to prove the claim for $f\in\mathcal{S}(\mathbb{R}^d)$ and $s\in\{0,1\}$. The conclusion follows by interpolation. We consider only $H\neq 1/2$, since the case $H=1/2$ corresponds to Brownian motion and is already covered by Example 3.2.10.

Let $0\le h<r\le T$ and $u,v\in\mathbb{R}^d$. From $\|u\|\le\|v\|+\|u+v\|$ it follows that $|\partial_r\Phi_{h,r}(u,v)|\lesssim(\|v\|^2+\|v\|\|u+v\|)((r-h)^{2H-1}+r^{2H-1})$. Lemma 3.5.14$(i)$ therefore implies that $\sum_{k-1>j\ge 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}(|\partial_r\varphi_{h,r}(u,v)|+|\partial_r\varphi_{t_{j-1},r}(u,v)|)dhdr$ is of order $O(T)$. Moreover, $|g_{t_{k-1},r}(u,v)|\lesssim|\mathcal{F}\mu(u+v)|$ for all $1\le k\le n$ and $t_{k-1}\le r<t_k$ with g from Proposition 3.2.2(i). Applying Proposition 3.2.2(i) and Lemma 3.5.3 shows that $\|\Gamma_T(f)-\widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2$ is up to a constant bounded by $C_\mu T\Delta_n\|f\|_{L^2}^2$. With $1\le T\le T^{2H}$ for $H>1/2$ this yields the claimed bound for $s=0$. With respect to $s=1$ note first that

$$\begin{aligned}
|\partial_r\Phi_{h,r}(u,v)| &\lesssim (1+\|u\|)(1+\|v\|)\Big((\|v\|+1)\,r^{2H-1}+(r-h)^{2H-1}\Big),\\
|\partial_r\Phi_{h,r}(u,v)\,\partial_h\Phi_{h,r}(u,v)| &\lesssim (1+\|u\|)(1+\|v\|)(\|v\|+\|u+v\|)^2\\
&\quad\cdot\left(r^{2H-1}h^{2H-1}+(r-h)^{2H-1}h^{2H-1}+(r-h)^{2H-1}r^{2H-1}\right)\\
&\quad+(1+\|u\|)(1+\|v\|)\left(\|v\|^2+\|v\|\|u+v\|\right)(r-h)^{4H-2},\\
\left|\partial_{hr}^2\Phi_{h,r}(u,v)\right| &\lesssim (1+\|u\|)(1+\|v\|)(r-h)^{2H-2}.
\end{aligned}$$

Lemma 3.5.14(ii), (iii) and (iv) imply

$$\begin{aligned}
&\Delta_n^{-1}\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t_k}(|\partial_r\varphi_{h,r}(u,v)|+|\partial_r\varphi_{t_{k-1},r}(u,v)|)dhdr\\
&\quad\lesssim(1+\|u\|)(1+\|v\|)(\|v\|+1)\,T^{2H}\,|\mathcal{F}\mu(u+v)|,\\
&\sum_{k-1>j\ge 2}\int_{t_{k-1}}^{t_k}\int_{t_{j-1}}^{t_j}|\partial_{hr}^2\varphi_{h,r}(u,v)|dhdr\\
&\quad\lesssim(1+\|u\|)(1+\|v\|)\,|\mathcal{F}\mu(u+v)|\begin{cases}T^{2H}, & H>1/2,\\ T\Delta_n^{2H-1}, & H<1/2.\end{cases}
\end{aligned}$$

This yields the claim for $s=1$ by applying Proposition 3.2.2(ii) and Lemma 3.5.3 as above. □

Proof of Theorem 3.2.14

Proof. We have $f_{a,\varepsilon}\in H^{1/2-\rho}(\mathbb{R})$ for any small $\rho>0$ with $\|f_{a,\varepsilon}\|_{H^{1/2-\rho}}\lesssim\varepsilon^{-1+\rho}$. By the triangle inequality and Theorem 3.2.13 (Assumption (X0) can be removed for $d=1$, cf. Remark 3.2.9) $\|L_T^a - \widehat{\Gamma}_{n,T}(f_{a,\varepsilon})\|_{L^2(\mathbb{P})}$ is bounded by

$$\begin{aligned}
&\|L_T^a - \Gamma_T(f_{a,\varepsilon})\|_{L^2(\mathbb{P})} + \|\Gamma_T(f_{a,\varepsilon}) - \widehat{\Gamma}_{n,T}(f_{a,\varepsilon})\|_{L^2(\mathbb{P})} \\
&\lesssim \|L_T^a - \Gamma_T(f_{a,\varepsilon})\|_{L^2(\mathbb{P})} + \varepsilon^{-1+\rho} \begin{cases} T^H \Delta_n^{\frac{3}{4}-\frac{\rho}{2}}, & H \geq 1/2, \\ T^{1/2} \Delta_n^{\frac{1+H}{2}-\rho H}, & H < 1/2. \end{cases}
\end{aligned}$$

By the occupation time formula (cf. Geman and Horowitz (1980)) and $\int f_{a,\varepsilon}(x)dx = 1$ it follows that $\|L_T^a - \Gamma_T(f_{a,\varepsilon})\|_{L^2(\mathbb{P})}^2$ is equal to $\mathbb{E}[(L_T^a - \int f_{a,\varepsilon}(x) L_T^x dx)^2] = \mathbb{E}[(\frac{1}{2}\int_{-1}^1 (L_T^a - L_T^{\varepsilon x + a})dx)^2]$. Equation Pitt (1978, (4.1)) implies (together with the proof of Pitt (1978, Theorem 4)) that $\mathbb{E}[(L_T^a - L_T^b)^2] \lesssim (a-b)^{2\xi}$ for all $0 < \xi < \frac{1}{2H}(1-H)$. Consequently, $\|L_T^a - \Gamma_T(f_{a,\varepsilon})\|_{L^2(\mathbb{P})} \lesssim \varepsilon^{\frac{1}{2H}(1-H)-\rho}$. Optimizing in ε yields the claim. □

3.5.3. Proofs of Section 3

Proof of Theorem 3.3.2

Proof. Assume first that $f \in L^2(\mu)$. Expanding the squared error yields

$$\begin{aligned}
\|\Gamma_T(f) - \widehat{\Gamma}_{n,T}(f)\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\Big[\Big|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} \left(f(X_r) - f\left(X_{t_{k-1}}\right)\right) dr\Big|^2\Big] \\
&= \sum_{k,l=1}^n \int_{t_{k-1}}^{t_k} \int_{t_{l-1}}^{t_l} \mathbb{E}\big[\left(f(X_r) - f\left(X_{t_{k-1}}\right)\right)\left(f(X_h) - f\left(X_{t_{l-1}}\right)\right)\big] dr dh.
\end{aligned}$$

We bound the diagonal ($l = k$) and off-diagonal terms ($l \neq k$) separately. Consider first the diagonal case and $t_{k-1} \leq r \leq h \leq t_k$. By the Markov property and stationarity of X the expectation above can be calculated explicitly. Indeed,

$$\begin{aligned}
&\mathbb{E}\left[\left(f(X_r) - f\left(X_{t_{k-1}}\right)\right)\left(f(X_h) - f\left(X_{t_{k-1}}\right)\right)\right] \\
&\quad = \langle P_{h-r} f, f\rangle_\mu - \left\langle P_{r-t_{k-1}} f, f\right\rangle_\mu - \left\langle P_{h-t_{k-1}} f, f\right\rangle_\mu + \langle f, f\rangle_\mu \\
&\quad = \left\langle (P_{h-r} - I) f + \left(I - P_{r-t_{k-1}}\right) f + \left(I - P_{h-t_{k-1}}\right) f, f\right\rangle_\mu.
\end{aligned}$$

Consequently, by symmetry in r, h,

$$\begin{aligned}
&\sum_{k=1}^n \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t_k} \mathbb{E}\big[\left(f(X_r) - f\left(X_{t_{k-1}}\right)\right)\left(f(X_h) - f\left(X_{t_{k-1}}\right)\right)\big] dr dh \\
&= 2\sum_{k=1}^n \Big\langle \Big(\int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^h (P_{h-r} - I)\, dr dh + \Delta_n \int_{t_{k-1}}^{t_k} \left(I - P_{h-t_{k-1}}\right) dh \Big) f, f \Big\rangle_\mu \\
&= 2n \Big\langle \Big(\int_0^{\Delta_n} \int_0^h (P_{h-r} - I)\, dr dh + \Delta_n \int_0^{\Delta_n} (I - P_h)\, dh \Big) f, f \Big\rangle_\mu.
\end{aligned}$$

Since the generator L is normal, by the functional calculus of L (see Section A.5) this can be written as

$$\langle \Psi(L) f, f\rangle_\mu = \int_{\sigma(L)} \Psi(\lambda)\, d\langle E_\lambda f, f\rangle_\mu$$

with

$$\Psi(\lambda) = 2n\left(\int_0^{\Delta_n}\int_0^h \left(e^{\lambda(h-r)}-1\right) drdh + \Delta_n \int_0^{\Delta_n}\left(1-e^{\lambda h}\right) dh\right), \quad \lambda \in \mathbb{C}.$$

Since L is the generator of a Feller semigroup, it follows that $\sigma(L) \subset \{\lambda \in \mathbb{C} : Re(\lambda) \leq 0\}$. Fix $0 \leq s \leq 1$ such that $|1-e^z| \leq 2|z|^s$ for $z \in \{\lambda \in \mathbb{C} : Re(\lambda) \leq 0\}$. Then $|\Psi(\lambda)| \leq 8n\Delta_n^{2+s}|\lambda|^s$, $\lambda \in \sigma(L)$. Hence the diagonal terms are bounded by

$$\begin{aligned}\int_{\sigma(L)} |\Psi(\lambda)|\, d\langle E_\lambda f, f\rangle_\mu &\leq 8T\Delta_n^{1+s}\int_{\sigma(L)} |\lambda|^s\, d\langle E_\lambda f, f\rangle_\mu \\ &\leq 8\|\,|L|^{\,s/2} f\|_\mu^2 T\Delta_n^{1+s}, \end{aligned} \tag{3.5.31}$$

which is true as long as $f \in \mathcal{D}(|L|^{s/2})$. For the off-diagonal terms with $l \neq k$ consider $t_{l-1} \leq r \leq t_{k-1} \leq h$. Then, similar as before

$$\begin{aligned}&\mathbb{E}\left[\left(f\left(X_h\right) - f\left(X_{t_{k-1}}\right)\right)\left(f\left(X_r\right) - f\left(X_{t_{l-1}}\right)\right)\right] \\ &\quad = \langle P_{h-r}f, f\rangle_\mu - \left\langle P_{h-t_{l-1}}f, f\right\rangle_\mu - \left\langle P_{t_{k-1}-r}f, f\right\rangle_\mu + \left\langle P_{t_{k-1}-t_{l-1}}f, f\right\rangle_\mu \\ &\quad = \left\langle P_{t_{k-1}-r}(P_{h-t_{k-1}} - I)(I - P_{r-t_{l-1}})f, f\right\rangle_\mu . \end{aligned} \tag{3.5.32}$$

The off-diagonal terms are therefore equal to

$$\begin{aligned}&2\sum_{k>l=1}^n \int_{t_{k-1}}^{t_k}\int_{t_{l-1}}^{t_l} \mathbb{E}\Big[\left(f\left(X_r\right) - f\left(X_{t_{k-1}}\right)\right)\left(f\left(X_h\right) - f\left(X_{t_{l-1}}\right)\right)\Big]drdh \\ &\quad = 2\sum_{k>l=1}^n \left\langle\left(\int_{t_{k-1}}^{t_k}\int_{t_{l-1}}^{t_l} P_{t_{k-1}-r}\left(P_{h-t_{k-1}} - I\right)\left(I - P_{r-t_{l-1}}\right) drdh\right) f, f\right\rangle_\mu \\ &\quad = \left\langle 2\left(\int_0^{\Delta_n}\int_0^{\Delta_n}\left(\sum_{k>l=1}^n P_{t_{k-1}-t_{l-1}-r}\right)\left(P_h - I\right)\left(I - P_r\right) drdh\right) f, f\right\rangle_\mu \\ &\quad = \int_{\sigma(L)} \tilde{\Psi}\left(\lambda\right) d\left\langle E_\lambda f, f\right\rangle_\mu , \end{aligned}$$

where

$$\tilde{\Psi}(\lambda) = 2\int_0^{\Delta_n}\int_0^{\Delta_n}\left(\sum_{k>l=1}^n e^{\lambda(t_{k-1}-t_{l-1}-r))}\right)\left(e^{\lambda h} - 1\right)\left(1 - e^{\lambda r}\right) drdh, \quad \lambda \in \mathbb{C}.$$

We will show that there exists a universal constant $\tilde{C} < \infty$ such that

$$\left|\tilde{\Psi}\left(\lambda\right)\right| \leq \tilde{C}T|\lambda|^s\Delta_n^{1+s}, \quad \lambda \in \sigma(L). \tag{3.5.33}$$

As in (3.5.31), this implies that the off-diagonal terms are bounded by

$$\tilde{C}T\Delta_n^{1+s}\int_{\sigma(L)} |\lambda|^s\, d\langle E_\lambda f, f\rangle_\mu = \tilde{C}\|\,|L|^{\,s/2} f\|_\mu^2 T\Delta_n^{1+s}$$

for $f \in \mathcal{D}(|L|^{s/2})$. Combining this with (3.5.31) yields the claim. In order to show (3.5.33) observe that $\tilde{\Psi}(\lambda) = 0$ for $\lambda = 0$. It is therefore sufficient to consider $\lambda \neq 0$. In order to bound $\tilde{\Psi}$ in that case note that

$$\sum_{k>l=1}^{n} e^{\lambda(k-l-1)\Delta_n} = \sum_{l=1}^{n} \frac{1-e^{\lambda(n-l)\Delta_n}}{1-e^{\lambda\Delta_n}} = \frac{n}{1-e^{\lambda\Delta_n}} - \frac{1-e^{\lambda n\Delta_n}}{(1-e^{\lambda\Delta_n})^2}.$$

Hence, again using $|1-e^z| \leq 2|z|^s$,

$$\begin{aligned}\left|\Delta_n^2\left(1-e^{\lambda\Delta_n}\right)^2 \sum_{k>l=1}^{n} e^{\lambda(k-l-1)\Delta_n)}\right| &\leq 2n\Delta_n^2\left|\lambda\Delta_n\right|^s + 2\Delta_n^2\left|\lambda n\Delta_n\right|^s \\ &\leq 4T\Delta_n^{1+s}\left|\lambda\right|^s.\end{aligned}$$

Therefore, (3.5.33) follows if

$$\left|\Delta_n^{-2}\left(1-e^{\lambda\Delta_n}\right)^{-2}\int_0^{\Delta_n}\int_0^{\Delta_n} e^{\lambda(\Delta_n-r)}\left(e^{\lambda h}-1\right)\left(1-e^{\lambda r}\right)drdh\right| \tag{3.5.34}$$

is bounded by a universal constant. To show this, let $z = \lambda\Delta_n$ and note that

$$\Delta_n^{-1}\left|\frac{\int_0^{\Delta_n}\left(1-e^{\lambda h}\right)dh}{1-e^{\lambda\Delta_n}}\right| = \Delta_n^{-1}\left|\frac{\Delta_n}{1-e^{\lambda\Delta_n}} - \lambda^{-1}\right| = \left|\frac{1}{z} - \frac{1}{e^z-1}\right|, \tag{3.5.35}$$

$$\Delta_n^{-1}\left|\int_0^{\Delta_n}\frac{\left(e^{\lambda(\Delta_n-r)}-e^{\lambda\Delta_n}\right)}{1-e^{\lambda\Delta_n}}dr\right| = \left|-\frac{1}{z} - \frac{e^z}{1-e^z}\right| = \left|\frac{1}{z} - \frac{1}{e^z-1}+1\right|. \tag{3.5.36}$$

(3.5.35) converges to 1 and (3.5.36) converges to 0 as $|z| \to \infty$. If $|z| \to 0$ and $z \in \{\lambda \in \mathbb{C} : \mathrm{Re}(\lambda) \leq 0\}$, then (3.5.35) converges to 1/2 and (3.5.36) converges to 3/2 . This implies a universal constant bounding (3.5.34), thereby proving (3.5.33). □

Remark 3.5.15. *(i)* If the generator is self-adjoint, then the operators P_u, $u \geq 0$, are self-adjoint as well. P_u is positive, $P_u - I$ is negative semidefinite and $I - P_u$ is positive semidefinite. Therefore $P_{t_{k-1}-r}(P_{h-t_{k-1}} - I)(I - P_{r-t_{l-1}})$ is negative semidefinite and (3.5.32) is non-positive. In this case, the off-diagonal terms do not contribute to the estimation error.

(ii) The restriction $0 \leq s \leq 1$ appears in (3.5.31) and (3.5.33) due to the Lipschitz bound $|1-e^z| \leq 2|z|^s$.

Proof of Corollary 3.3.3

Proof. Denote by $\mathbb{E}_\nu$ for a measure ν the expectation with respect to X starting with ν as initial distribution. Let $g(x) = \mathbb{E}[|\Gamma_{T_n,T}(f) - \widehat{\Gamma}_{n,T_n,T}(f)|^2|X_{T_n} = x]$. By conditioning on X_{T_0} the tower and Markov properties yield that

$$\begin{aligned}\left\|\Gamma_{T_n,T}(f) - \widehat{\Gamma}_{n,T_n,T}(f)\right\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\left[\mathbb{E}\left[g\left(X_{T_n}\right)\middle| X_{T_0}\right]\right] \\ = \mathbb{E}\left[P_{T_n-T_0}g\left(X_{T_0}\right)\right] &= \int_{\mathcal{S}} P_{T_n-T_0}g\left(x\right)d\eta(x) \\ \leq \left\|\frac{d\eta}{d\mu}\right\|_{\infty,\mu} &\int_{\mathcal{S}} P_{T_n-T_0}g\left(x\right)d\mu(x).\end{aligned}$$

Since μ is an invariant (stationary) measure of the semigroup, this is equal to (cf. Bakry et al. (2013, Section 1.2.1))

$$\left\|\frac{d\eta}{d\mu}\right\|_{\infty,\mu}\int_{\mathcal{S}} g(x)\,d\mu(x) = \left\|\frac{d\eta}{d\mu}\right\|_{\infty,\mu}\mathbb{E}_\mu\left[\left|\Gamma_{T-T_n}(f) - \widehat{\Gamma}_{n,T-T_n}(f)\right|^2\right],$$

because

$$\begin{aligned} g(x) &= \mathbb{E}\left[\left|\int_0^{T-T_n} f(X_r)\,dr - \Delta_n\sum_{k=1}^{n-T_n} f(X_{t_k})\right|^2 \middle| X_0 = x\right] \\ &= \mathbb{E}\left[\left|\Gamma_{T-T_n}(f) - \widehat{\Gamma}_{n,T-T_n}(f)\right|^2 \middle| X_0 = x\right]. \end{aligned}$$

The conclusion follows by a simple modification of Theorem 3.3.2, because the error is now considered on $[0, T - T_n]$ instead of $[0, T]$. ☐

Proof of Theorem 3.3.4

Proof. By the triangle inequality it follows from $f \in L^2(\mu) = \mathrm{dom}(|L|^0)$ and Theorem 3.3.2 that

$$\begin{aligned} &\left\|T^{-1}\widehat{\Gamma}_{n,T}(f) - \int_{\mathcal{S}} f(x)\,d\mu(x)\right\|_{L^2(\mathbb{P})} \\ &\quad\le T^{-1}\left\|\widehat{\Gamma}_{n,T}(f) - \Gamma_T(f)\right\|_{L^2(\mathbb{P})} + \left\|T^{-1}\Gamma_T(f) - \int_{\mathcal{S}} f(x)\,d\mu(x)\right\|_{L^2(\mathbb{P})} \\ &\quad\le C\|f\|_{L^2(\mu)}T^{-1/2}\Delta_n^{1/2} + \left\|T^{-1}\Gamma_T(f) - \int_{\mathcal{S}} f(x)\,d\mu(x)\right\|_{L^2(\mathbb{P})} \end{aligned}$$

for a universal constant C. The claimed bound for the second term is well-known, but we give the proof here to complement the proof of Theorem 3.3.2. Consider f such that $f_0 = f - \int f d\mu \in \mathrm{dom}(|L|^{-1/2})$. By linearity of the occupation time functional it follows that

$$T^{-1}\Gamma_T(f) - \int_{\mathcal{S}} f d\mu = T^{-1}\Gamma_T(f_0).$$

Fubini's theorem yields

$$\begin{aligned} \mathbb{E}\left[\left|T^{-1}\Gamma_T(f_0)\right|^2\right] &= T^{-2}\int_0^T\int_0^T \mathbb{E}\left[f_0(X_r)\,f_0(X_h)\right]drdh \\ &= 2T^{-2}\int_0^T\int_0^h \langle P_{h-r}f_0, f_0\rangle_\mu\,drdh \\ &= \int_{\sigma(L)} \Psi(\lambda)\,d\langle E_\lambda f_0, f_0\rangle_\mu, \end{aligned}$$

where

$$\Psi(\lambda) = 2T^{-2}\int_0^T\int_0^h e^{\lambda(h-r)}drdh = 2\frac{e^{\lambda T} - 1 - \lambda T}{\lambda^2T^2} = 2\frac{(\lambda T)^{-1}(e^{\lambda T} - 1) - 1}{\lambda T},$$

and where $\Psi(0) = 1$ by continuous extension. Since $z \to z^{-1}(e^z - 1) - 1$ is bounded on the left half-plane $\{z \in \mathbb{C} : \mathrm{Re}(z) \leq 0\}$, there exists a constant $\tilde{C} < \infty$ such that

$$|\Psi(\lambda)| \leq \tilde{C}T^{-1}|\lambda|^{-1}, \quad \lambda \in \sigma(L).$$

Consequently,

$$\mathbb{E}\left[\left|T^{-1}\Gamma_T(f_0)\right|^2\right] \leq CT^{-1}\int_{\sigma(L)} |\lambda|^{-1} d\langle E_\lambda f_0, f_0\rangle_\mu = CT^{-1}\| \, |L|^{-1/2} f_0\|_\mu^2. \qquad \square$$

3.5.4. Proof of Theorem 3.4.1

Consider first the following two lemmas.

Lemma 3.5.16. *Assume (X0). For $f \in H^1(\mathbb{R}^d)$ we have*

$$\|\Gamma_T(f) - \mathbb{E}[\Gamma_T(f)|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2 = \Delta_n^2\mathbb{E}\left[\frac{1}{12}\int_0^T \|\nabla f(X_r)\|^2 dr\right] + o\left(\Delta_n^2\|f\|_{H^1}^2\right).$$

In particular, $\Delta_n^{-2}\|\Gamma_T(f) - \mathbb{E}[\Gamma_T(f)|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2$ converges to $\mathbb{E}[\frac{1}{12}\int_0^T\|\nabla f(X_r)\|^2dr]$ as $n \to \infty$.

Proof. By independence of X_0 and $(X_r - X_0)_{0\leq t\leq T}$ the σ-algebra $\mathcal{G}_n$ is also generated by X_0 and the increments $X_{t_k} - X_{t_{k-1}}$, $1 \leq k \leq n$. The independence of increments and the Markov property then imply for $t_{k-1} \leq r \leq t_k$ that $\mathbb{E}[f(X_r)|\mathcal{G}_n] = \mathbb{E}[f(X_r)|X_{t_{k-1}}, X_{t_k}]$. The same argument shows that the random variables $Y_k = \int_{t_{k-1}}^{t_k}(f(X_r) - \mathbb{E}[f(X_r)|\mathcal{G}_n])dr$ are uncorrelated. Therefore

$$\|\Gamma_T(f) - \mathbb{E}[\Gamma_T(f)|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2 = \sum_{k=1}^n \mathbb{E}\left[Y_k^2\right] = \sum_{k=1}^n \mathbb{E}\left[\mathrm{Var}_k\left(\int_{t_{k-1}}^{t_k} f(X_r)\,dr\right)\right],$$

where $\mathrm{Var}_k(Z)$ is the conditional variance of a random variable Z with respect to the σ-algebra generated by $X_{t_{k-1}}$ and X_{t_k}. In order to linearize f, note that the random variable $\mathrm{Var}_k(\int_{t_{k-1}}^{t_k} f(X_r)dr) = \mathrm{Var}_k(\int_{t_{k-1}}^{t_k}(f(X_r) - f(X_{t_{k-1}}))dr)$ can be written as

$$\begin{aligned}&\mathrm{Var}_k\left(\int_{t_{k-1}}^{t_k} \langle\nabla f(X_{t_{k-1}}), X_r - X_{t_{k-1}}\rangle\, dr\right) + \kappa_n \\ &\quad + \mathrm{Var}_k\left(\int_{t_{k-1}}^{t_k}\left(f(X_r) - f(X_{t_{k-1}}) - \langle\nabla f(X_{t_{k-1}}), X_r - X_{t_{k-1}}\rangle\right)dr\right),\end{aligned}$$

where κ_n is the corresponding crossterm of the decomposition. From Lemma 3.5.2(ii) and (iii) it follows that the first and the last term are of order $o(\Delta_n^3\|f\|_{H^1}^2)$ and $O(\Delta_n^3 a_n(f)) = O(\Delta_n^3\|f\|_{H^1}^2)$, respectively, and thus by the Cauchy-Schwarz inequality $\kappa_n = o(\Delta_n^3\|f\|_{H^1}^2)$. Hence, $\|\Gamma_T(f) - \mathbb{E}[\Gamma_T(f)\,|\mathcal{G}_n]_{L^2(\mathbb{P})}^2$ is equal to

$$\sum_{k=1}^n \mathbb{E}\left[\mathrm{Var}_k\left(\int_{t_{k-1}}^{t_k} \langle\nabla f(X_{t_{k-1}}), X_r\rangle\, dr\right)\right] + o\left(\Delta_n^2\|f\|_{H^1}^2\right).$$

Conditional on $X_{t_{k-1}}$, X_{t_k}, the process $(X_r)_{t_{k-1}\leq r\leq t_k}$ is a Brownian bridge starting from $X_{t_{k-1}}$ and ending at X_{t_k}. In particular, $\mathbb{E}[X_r|X_{t_{k-1}},X_{t_k}]=X_{t_{k-1}}+\frac{r-t_{k-1}}{\Delta_n}(X_{t_k}-X_{t_{k-1}})$ (see e.g. Karatzas and Shreve (1991, 6.10)). The stochastic Fubini theorem and Itô isometry thus imply that the last display is equal to

$$\begin{aligned}
&\sum_{k=1}^{n}\mathbb{E}\left[\left\langle \nabla f\left(X_{t_{k-1}}\right), \int_{t_{k-1}}^{t_k}\left(t_k-r-\frac{\Delta_n}{2}\right)dX_r\right\rangle^2\right]+o\left(\Delta_n^2\|f\|_{H^1}^2\right)\\
&=\frac{\Delta_n^3}{12}\sum_{k=1}^{n}\mathbb{E}\left[\|\nabla f\left(X_{t_{k-1}}\right)\|^2\right]+o\left(\Delta_n^2\|f\|_{H^1}^2\right)\\
&=\frac{\Delta_n^2}{12}\int_0^T\|\nabla f(X_r)\|^2dr+o\left(\Delta_n^2\|f\|_{H^1}^2\right),
\end{aligned}$$

where the last line follows from Lemma 3.5.2(*iv*). □

Lemma 3.5.17. *Assume (X0). Fix* $0\leq s<\alpha$ *and let* $\varphi(x)=(2\pi)^{-d/2}e^{-\|x\|^2/2}$ *for* $x\in\mathbb{R}^d$. *Consider the approximations* $f_{\alpha,\varepsilon}=f_\alpha*\varphi_\varepsilon$, *where* $\varphi_\varepsilon=\varepsilon^{-d}\varphi(\varepsilon^{-1}(\cdot))$ *and* $\varepsilon=\Delta_n^{\frac{1}{2}\cdot\frac{1-s}{1-\alpha}}$. *Then the following statements hold as* $n\to\infty$*:*

(i) $\|\Gamma_T(f_\alpha-f_{\alpha,\varepsilon})-\mathbb{E}\left[\Gamma_T\left(f_\alpha-f_{\alpha,\varepsilon}\right)|\,\mathcal{G}_n\right]\|_{L^2(\mathbb{P})}^2=o(\Delta_n^{1+s})$,

(ii) $\|\Gamma_T(f_{\alpha,\varepsilon})-\mathbb{E}\left[\Gamma_T\left(f_{\alpha,\varepsilon}\right)|\,\mathcal{G}_n\right]\|_{L^2(\mathbb{P})}^2=O(\Delta_n^2\|f_{\alpha,\varepsilon}\|_{H^1}^2)=O(\Delta_n^{1+s})$,

(iii) $\liminf_{n\to\infty}(\varepsilon^{2-2\alpha}\mathbb{E}[\frac{1}{12}\int_0^T\|\nabla f_{\alpha,\varepsilon}(X_r)\|^2dr])>0$.

Proof. Applying (3.4.1) from right to left and Theorem 3.2.13 for the function $f=f_\alpha-f_{\alpha,\varepsilon}\in L^2(\mathbb{R}^d)$ shows that the left hand side of the equation in (*i*) is up to a constant bounded by $\Delta_n\|f_\alpha-f_{\alpha,\varepsilon}\|_{L^2}^2$. The Plancherel theorem and $\mathcal{F}\varphi_\varepsilon(u)=\mathcal{F}\varphi(\varepsilon u)$ yield that this is equal to

$$(2\pi)^{-d}\,\Delta_n\|\mathcal{F}f_\alpha\left(1-\mathcal{F}\varphi_\varepsilon\right)\|_{L^2}^2\lesssim\Delta_n\varepsilon^{2\alpha}\int\|u\|^{-2\alpha-d}\left(1-e^{-\frac{\|u\|^2}{2}}\right)du.$$

The du-integral is finite and therefore the last line is of order $O(\Delta_n\varepsilon^{2\alpha})=o(\Delta_n^{1+s})$, because $\alpha>s$, implying (*i*). Similarly, applying (3.4.1) from right to left and Theorem 3.2.13 for the function $f=f_{\alpha,\varepsilon}\in H^1(\mathbb{R}^d)$, the left hand side of the equation in (*ii*) is up to a constant bounded by $\Delta_n^2\|f_{\alpha,\varepsilon}\|_{H^1}^2$. As above this can be bounded from the Plancherel theorem by

$$\begin{aligned}
&(2\pi)^{-d}\,\Delta_n^2\int|\mathcal{F}f_\alpha\left(u\right)|^2\left|\mathcal{F}\varphi\left(\varepsilon u\right)\right|^2\left(1+\|u\|\right)^2du\\
&\lesssim\Delta_n^2\varepsilon^{2\alpha-2}\int\left(\varepsilon+\|u\|\right)^{2-2\alpha-d}e^{-\frac{\|u\|^2}{2}}du\\
&\lesssim\Delta_n^2\varepsilon^{2\alpha-2}\int_0^\infty\left(\varepsilon+\|r\|\right)^{1-2\alpha}e^{-\frac{r^2}{2}}dr.
\end{aligned}$$

As $\alpha<1$, the dr-integral is finite for $\varepsilon=0$ and thus the last line is of order $O(\Delta_n^2\varepsilon^{2\alpha-2})=O(\Delta_n^{1+s})$. This is the claimed order in (*ii*). Finally, with respect to

(iii), denote by p_r the marginal density of X_r. Then we have by the Plancherel theorem, applied componentwise, for any $T_0 > 0$ that $\mathbb{E}[\frac{1}{12}\int_0^T \|\nabla f_{\alpha,\varepsilon}(X_r)\|^2 dr]$ is bounded from below up to a constant by

$$\int_{T_0}^{T}\left(\int \|\nabla f_{\alpha,\varepsilon}(x)p_r^{1/2}(x)\|^2 dx\right) dr$$
$$= (2\pi)^{-2d}\int_{T_0}^{T}\left(\int \|\int \mathcal{F}f_\alpha(u-y)\,\mathcal{F}\varphi(\varepsilon(u-y))(u-y)\,h_r(y)\,dy\|^2 du\right) dr,$$

where $h_r(y) = 2^{d/2}(2\pi)^{d/4}r^{d/4}e^{-\|y\|^2 r}$ is the Fourier transform of $p_r^{1/2}$. The substitution $\varepsilon u \mapsto u$ then yields that the du-integral above is equal to

$$\varepsilon^{2\alpha-2}\int \|\int \nu_\varepsilon(u-\varepsilon y)\,h_r(y)\,dy\|^2 du = \varepsilon^{2\alpha-2}\int \|(\nu_\varepsilon * h_{r,\varepsilon})(u)\|^2 du,$$

for $h_{r,\varepsilon}(u) = \varepsilon^{-d}h_r(\varepsilon^{-1}u)$ and $\nu_\varepsilon(u) = u(\varepsilon + \|u\|)^{-\alpha-d/2}e^{-\|u\|^2/2}$. Interestingly, $\nu_\varepsilon \in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ for all $\varepsilon \geq 0$ as $\alpha < 1$. As also $h_{r,\varepsilon} \in L^1(\mathbb{R}^d)$, Young's inequality, also applied componentwise, implies that

$$\int \|((\nu_\varepsilon - \nu_0) * h_{r,\varepsilon})(u)\|^2 du \leq \|h_{r,\varepsilon}\|_{L^1}^2\left(\int \|(\nu_\varepsilon - \nu_0)(u)\|^2 du\right).$$

Since $\|h_{r,\varepsilon}\|_{L^1}^2 \lesssim r^{-d/2}$, $\|\nu_\varepsilon(u)\| \leq \|\nu_0(u)\|$ and $\nu_\varepsilon(u) \to \nu_0(u)$ for any $u \in \mathbb{R}^d$ we therefore conclude by dominated convergence that the last line is of order $o(r^{-d/2})$. Moreover, it follows again by the Plancherel theorem with $\mathcal{F}h_{r,\varepsilon}(x) = (2\pi)^d p_r^{1/2}(\varepsilon x)$ that $\int \|(\nu_0 * h_{r,\varepsilon})(u)\|^2 du = (2\pi)^d\int \|\mathcal{F}\nu_0(x)\|^2 p_r(\varepsilon x)\,dx$. Letting $\varepsilon \to 0$ yields the convergence to $(2\pi)^{d/2} r^{-d/2}\int \|\mathcal{F}\nu_0(x)\|^2 dx$. By Pythagoras we thus find for any $r > T_0 > 0$ that also $\int \|(\nu_\varepsilon * h_{r,\varepsilon})(u)\|^2 du \to cr^{-d/2}$ for some constant $0 < c < \infty$. Consequently,

$$\liminf_{n\to\infty}\left(\varepsilon^{2-2\alpha}\mathbb{E}\left[\frac{1}{12}\int_0^T \|\nabla f_{\alpha,\varepsilon}(X_r)\|^2 dr\right]\right) \gtrsim \int_{T_0}^{T} r^{-d/2} dr,$$

which is bounded from below as $T_0 > 0$. □

Now we prove the theorem.

Proof of Theorem 3.4.1. The first inequality in (i) is clear. The limit in the last equality follows from Lemma 3.5.16. With respect to (ii) observe that

$$\|\Gamma_T(f_\alpha) - \mathbb{E}[\Gamma_T(f_\alpha)|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2 = \|\Gamma_T(f_\alpha - f_{\alpha,\varepsilon}) - \mathbb{E}[\Gamma_T(f_\alpha - f_{\alpha,\varepsilon})|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2$$
$$+ \kappa_n + \|\Gamma_T(f_{\alpha,\varepsilon}) - \mathbb{E}[\Gamma_T(f_{\alpha,\varepsilon})|\,\mathcal{G}_n]\|_{L^2(\mathbb{P})}^2,$$

where κ_n is the crossterm of the expansion. From Lemma 3.5.17 it follows that the first term is of order $o(\Delta_n^{1+s})$, while the third one is of order $O(\Delta_n^{1+s})$. Therefore, the crossterm is via the Cauchy-Schwarz inequality itself of order $o(\Delta_n^{1+s})$. Hence, Lemma

3.5.16 implies that $\liminf_{n\to\infty} \Delta_n^{-(1+s)} \|\Gamma_T(f_\alpha) - \mathbb{E}[\Gamma_T(f_\alpha)|\mathcal{G}_n]\|^2_{L^2(\mathbb{P})}$ is equal to

$$\begin{aligned}
&\liminf_{n\to\infty} \Delta_n^{-(1+s)} \|\Gamma_T\left(f_{\alpha,\varepsilon}\right) - \mathbb{E}\left[\Gamma_T\left(f_{\alpha,\varepsilon}\right)\middle|\, \mathcal{G}_n\right]\|^2_{L^2(\mathbb{P})} \\
&= \liminf_{n\to\infty} \left(\Delta_n^{1-s} \mathbb{E}\left[\frac{1}{12} \int_0^T \|\nabla f_{\alpha,\varepsilon}(X_r)\|^2 dr \right] \right) \\
&\quad + \liminf_{n\to\infty} \left(o\left(\Delta_n^{-(1+s)} \Delta_n^2 \|f_{\alpha,\varepsilon}\|^2_{H^1} \right) \right).
\end{aligned}$$

From part (ii) of Lemma 3.5.17 it follows that the last term is 0, while part (iii) implies the wanted lower bound for the first term, as $\Delta_n^{1-s}\varepsilon^{2\alpha-2} = 1$. □

Chapter 4.

Generalized Itô formulas

In this chapter we study the existence of the quadratic covariation process and the related problem of finding generalized Itô formulas. The methods we use are directly related to the ones of Section 3.1. The results are formulated in terms of Fourier-Lebesgue spaces which are introduced in Section A.4.

4.1. Existence of quadratic covariations and Itô formulas

We first define the quadratic covariation and state some of its properties. More details can be found in Russo and Vallois (1996, 1995). Consider two $\mathbb{R}^d$-valued stochastic processes $Z = (Z_t)_{0\le t\le T}$ and $Y = (Y_t)_{0\le t\le T}$. The *quadratic covariation* $[Z, Y]_t$ of Z and Y at $0 \le t \le T$ is defined as

$$[Z,Y]_t = \lim_{n\to\infty} \sum_{t_k\in\pi_n, t_k\le t} \left(Z_{t_k} - Z_{t_{k-1}}\right)\left(Y_{t_k} - Y_{t_{k-1}}\right)^\top, \tag{4.1.1}$$

if this limit exists in probability for any sequence of partitions $(\pi_n)_{n\ge1}$ of $[0, T]$, where the points in π_n are $0 = t_0 < t_1 < \cdots < t_n = T$, and such that the mesh size $|\pi_n| = \max_k |t_k - t_{k-1}|$ tends to zero as $n \to \infty$. $[Z, Y]_t$ is bilinear and independent of $(\pi_n)_{n\ge1}$. Moreover, if Z, Y are continuous semimartingales, then $[Z, Y]_t$ always exists.

From now on let X be a continuous Itô semimartingale as in (2.1.1). It follows that $[f(X), X^{(m)}]_t$ exists for $f \in C^1(\mathbb{R}^d)$ and $1 \le m \le d$ and satisfies $[f(X), X^{(m)}]_t = \sum_{k=1}^d \int_0^t \partial_k f(X_r)(\sigma_r\sigma_r^\top)^{(k,m)} dr$ (cf. Russo and Vallois (1996, Proposition 1.1)). Motivated by the results in Section 3.1.2 we will study this expression by Fourier inversion (cf. (3.1.6)). The results follow the same sequence as in Section 3.1, i.e. we consider first $s \ge 0$ and $f \in FL^s_{loc}(\mathbb{R}^d)$, then $f \in H^s_{loc}(\mathbb{R}^d)$ under a stronger assumption on X_0.

Proposition 4.1.1. *Assume (SM-α-β) for $0 \le \alpha, \beta \le 1$. Let $s > 1 - 2\alpha$, $s \ge 0$, $s + \beta > 0$. Then we have for any $f \in FL^s_{loc}(\mathbb{R}^d)$ and $1 \le m \le d$*

$$\sum_{t_k\in\pi_n, t_k\le t} \left(f\left(X_{t_k}\right) - f\left(X_{t_{k-1}}\right)\right)\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right) \xrightarrow{ucp} [f(X), X^{(m)}]_t \tag{4.1.2}$$

for any sequence of partitions $(\pi_n)_{n\ge1}$ of $[0, T]$ with $|\pi_n| \to 0$, i.e. the sum converges uniformly in probability to $[f(X), X^{(m)}]_t$.

This proposition is remarkable, similar to Theorem 3.1.3, because it is only based on regularity assumptions for σ and b and gives a precise condition on the regularities of σ, b and f. If X is reversible, i.e. if the time reversed process $t \mapsto X_{T-t}$ is again a semimartingale, then this is a special case of Theorem 3.8 of Errami et al. (2002), because $FL^s_{loc}(\mathbb{R}^d) \subset C(\mathbb{R}^d)$. For Proposition 4.1.1, however, reversibility is not necessary.

In order to go beyond $f \in FL^s_{loc}(\mathbb{R}^d)$, we need to work again under Assumption (X0). In this case the sum in (4.1.2) does not converge uniformly in time anymore, in general (cf. Föllmer and Protter (2000)).

Proposition 4.1.2. *Assume (SM-α-β) for $0 \leq \alpha, \beta \leq 1$ and (X0). Let $s > 1 - 2\alpha$, $s \geq 0$, $s + \beta > 0$. Then the quadratic covariation $[f(X), X^{(m)}]_t$ exists for $f \in H^s_{loc}(\mathbb{R}^d)$ and $1 \leq m \leq d$, $0 \leq t \leq T$.*

The condition on X_0 can be relaxed, just as in Corollary 3.1.6.

Corollary 4.1.3. *Assume (SM-α-β) for $0 \leq \alpha, \beta \leq 1$. Let $s > 1-2\alpha$, $s \geq 0$, $s+\beta > 0$. For any function $f \in H^s_{loc}(\mathbb{R}^d)$ there exists a set $E \subset \mathbb{R}^d$ such that $\mathbb{R}^d \backslash E$ has Lebesgue measure 0 and such that the quadratic covariation $[f(X), X^{(m)}]_t$ exists for all $X_0 = x_0 \in E$, $1 \leq m \leq d$, $0 \leq t \leq T$.*

This is, to the best of our knowledge, the most general condition obtained so far for the existence of $[f(X), X^{(m)}]_t$ for a continuous Itô semimartingale X. The set E depends in general on the function f, i.e. it can change if we consider a different function $\tilde{f}$ with $f = \tilde{f}$ almost everywhere. The same restriction appears in Föllmer and Protter (2000) with respect to $f \in L^2_{loc}(\mathbb{R}^d) = H^0_{loc}(\mathbb{R}^d)$ and with X being a Brownian motion.

Remark 4.1.4.

(i) The proof of Proposition 4.1.2 shows that for bounded coefficients b and σ the quadratic covariation $[f(X), X^{(m)}]_t$ even exists as limit in $L^p(\mathbb{P})$ and that $\|[f(X), X^{(m)}]_t\|_{L^2(\mathbb{P})} \lesssim \|f\|_{H^s}$.

(ii) The partition sequences $(\pi_n)_{n\geq 1}$ for which $[f(X), X^{(m)}]_t$ exists are arbitrary, as long as $|\pi_n| \to 0$. In particular, we do not require further conditions such as (3.1) in Moret and Nualart (2001) or (2.2) in Föllmer and Protter (2000).

(iii) A slightly more precise argument shows that $[f(X), X^{(m)}]_t$ exists for all bounded measurable functions f, if X is an additive process (cf. 3.2.2). This is possible, because in this case the characteristic functions in the proof of Lemma 4.2.1 can be calculated explicitly without approximation and thus the expression $(1 + \|u\|)^{-s}\|Q_n(u)\|_{L^2(\mathbb{P})}$ in that lemma is not only bounded, but even integrable with respect to u.

Generalized Itô formulas hold under any of the conditions stated above.

Theorem 4.1.5. *Assume (SM-α-β) for $0 \leq \alpha, \beta \leq 1$. Let $s > 2 - 2\alpha$, $s \geq 1$, $s + \beta > 1$ and consider $0 \leq t \leq T$. Then the Itô formula*

$$f(X_t) = f(X_0) + \int_0^t \langle \nabla f(X_r), dX_r \rangle + \frac{1}{2} \sum_{m=1}^{d} \left[\partial_m f(X), X^{(m)}\right]_t$$

holds under any of the following conditions:

(i) $f \in FL^s_{loc}(\mathbb{R}^d)$,

(ii) $f \in H^s_{loc}(\mathbb{R}^d)$ *and Assumption (X0) is satisfied,*

(iii) f *is a function in* $H^s_{loc}(\mathbb{R}^d)$ *and* $X_0 = x_0 \in E \subset \mathbb{R}^d$*, where* E *is the set of Corollary 4.1.3,*

The conditions in the theorem are precisely the ones for the central limit theorems in Section 3.1. The Itô formula of Russo and Vallois (1996) applies also to $C^1(\mathbb{R}^d)$-functions, assuming reversibility of X.

4.2. Proofs

We first provide some preliminary steps. Define for a partition π_n of $[0,T]$ and a function f the operator

$$\begin{aligned} S_{t,\pi_n}(f) &= S_{t,\pi_n}(f,X) = S^{(m)}_{t,\pi_n}(f,X) \\ &= \sum_{t_k\in\pi_n, t_k\leq t} \left(f\left(X_{t_k}\right) - f\left(X_{t_{k-1}}\right)\right)\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right). \end{aligned}$$

As for the proofs of the central limit theorems in Section 3.1 consider the decomposition $S_{t,\pi_n}(f) = M_{n,t} + D_{n,t}$, where $M_{n,t} = \sum_{t_k\in\pi_n, t_k\leq\pi_n}(A_k - \mathbb{E}[A_k|\mathcal{F}_{t_{k-1}}])$ and $D_{n,t} = \sum_{t_k\in\pi_n, t_k\leq\pi_n}\mathbb{E}[A_k|\mathcal{F}_{t_{k-1}}]$ with $A_k = (f(X_{t_k}) - f(X_{t_{k-1}}))(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}})$. The martingale $M_{n,t}$ can be handled easily, as we will see later. In order to study $D_{n,t}$ we apply the idea from the introduction of this thesis. For $f \in \mathcal{S}(\mathbb{R}^d)$ it follows by inverse Fourier transform that $D_{n,t}$ is equal to $(2\pi)^{-d}\int \mathcal{F}f(u)Q_{n,t}(u)du$, where

$$Q_{n,t}(u) = \sum_{t_k\in\pi_n, t_k\leq t} \mathbb{E}\left[\left(e^{-i\langle u,X_{t_k}\rangle} - e^{-i\langle u,X_{t_{k-1}}\rangle}\right)\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right)\middle|\mathcal{F}_{t_{k-1}}\right].$$

We will first prove upper bounds on $\|Q_{n,t}(u)\|_{L^2(\mathbb{P})}$ uniformly in $u \in \mathbb{R}^d$. For this it is helpful to work under Assumption (H-α-β) first.

Lemma 4.2.1. *Assume (H-α-β) for* $0 \leq \alpha, \beta \leq 1$*. Let* $s > 1 - 2\alpha$*,* $s \geq 0$*,* $s + \beta > 0$*. Let* $(\pi_n)_{n\geq 1}$ *be a sequence of partitions with* $|\pi_n| \to 0$ *and consider* $1 \leq m \leq d$*,* $0 \leq t \leq T$*. Then* $\sup_{u\in\mathbb{R}^d}((1+\|u\|)^{-s}\|Q_{n,t}(u)\|_{L^2(\mathbb{P})}) < \infty$*.*

Proof. In order to simplify the notation we consider in the following only $t = T$ and write $S_{\pi_n}(f) = S_{T,\pi_n}(f)$, $Q_n(u) = Q_{n,T}(u)$ for fixed $1 \leq m \leq d$. The proof is similar to Lemma 3.5.7. For fixed $u \in \mathbb{R}^d$ consider the following four steps.

Step 1. Let $\varepsilon > 0$, which will be chosen later depending on u. Let $t_\varepsilon = \max(\lfloor t/\varepsilon\rfloor\varepsilon - \varepsilon, 0)$, $t \geq 0$, and consider for $k = 1,\dots,n$ the partitions $\{t_{k-1} = a_{k,0} < a_{k,1} < \cdots < a_{k,N_k} = t_k\}$ of $[t_{k-1}, t_k]$ such that either $N_k = 1$ and $(t_k)_\varepsilon < t_{k-1}$, or $N_k > 1$ and $a_{k,N_k-1} = (t_k)_\varepsilon > t_{k-1}$, while $\{a_{k,1},\dots,a_{k,N_k-2}\}$ are all multiples of ε in $(t_{k-1},(t_k)_\varepsilon)$. Let $Z_t = \int_0^t \sigma_r dW_r$, $0 \leq t \leq T$. Then with $U_k = Z_{t_k} - Z_{t_{k-1}}$

$$\begin{aligned} &\sum_{k=1}^n \mathbb{E}\left[\left(e^{-i\langle u,X_{t_k}\rangle} - e^{-i\langle u,X_{t_{k-1}}\rangle}\right)U^{(m)}_k\middle|\mathcal{F}_{t_{k-1}}\right] \\ &= \sum_{k=1}^n\sum_{j=1}^{N_k} \mathbb{E}\left[\left(e^{-i\langle u,X_{a_{k,j}}\rangle} - e^{-i\langle u,X_{a_{k,j-1}}\rangle}\right)U^{(m)}_k\middle|\mathcal{F}_{t_{k-1}}\right]. \end{aligned}$$

Let $X_t(\varepsilon,h)$, $0\le h\le t$, be as in (A.3.1) with $(t)_{\varepsilon,h}=(t)_\varepsilon\vee h$, $0\le h\le t$, and let $Z_t(\varepsilon,h)$ be the corresponding approximation of Z_t. Define furthermore $Q_{n,t}(\varepsilon,u)$ as

$$Q_{n,t}(\varepsilon,u)=\sum_{k=1}^{n}\sum_{j=1}^{N_k}\mathbb{E}\Bigg[\left(e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}-e^{-i\langle u,X_{a_{k,j-1}}\rangle}\right)\cdot\left(Z^{(m)}_{a_{k,j}}(\varepsilon,t_{k-1})-Z^{(m)}_{t_{k-1}}\right)\Bigg|\mathcal{F}_{t_{k-1}}\Bigg]$$

and let $\tilde{Q}_{n,t}(\varepsilon,u)$ be defined similarly but with $X_{a_{k,j}}$ instead of $X_{a_{k,j}}(\varepsilon,t_{k-1})$. It then follows from the Lipschitz continuity of $x\mapsto e^{ix}$ and the Cauchy-Schwarz inequality that $|Q_{n,t}(u)-\tilde{Q}_{n,t}(\varepsilon,u)|$ is upper bounded bounded by

$$\int_0^T\|b_r\|dr+\|u\|\left(\sum_{k=1}^{n}\sum_{j=1}^{N_k}\mathbb{E}\left[\|X_{a_{k,j}}-X_{a_{k,j-1}}\|^2\Big|\mathcal{F}_{t_{k-1}}\right]\right)^{1/2}\cdot\left(\sum_{k=1}^{n}\sum_{j=1}^{N_k}\mathbb{E}\left[\|Z^{(m)}_{a_{k,j}}-Z^{(m)}_{a_{k,j}}(\varepsilon,t_{k-1})\|^2\Big|\mathcal{F}_{t_{k-1}}\right]\right)^{1/2}. \tag{4.2.1}$$

Proposition A.3.2 shows $\|Q_{n,t}(u)-\tilde{Q}_{n,t}(\varepsilon,u)\|_{L^2(\mathbb{P})}\lesssim 1+\|u\|(\varepsilon^{\beta+1/2}+\varepsilon^{\alpha})$. The same bound applies to $\|Q_{n,t}(\varepsilon,u)-\tilde{Q}_{n,t}(\varepsilon,u)\|_{L^2(\mathbb{P})}$ and thus also to $\|Q_{n,t}(u)-Q_{n,t}(\varepsilon,u)\|_{L^2(\mathbb{P})}$.

Step 2. The reduction to $Q_{n,t}(\varepsilon,u)$ allows us to calculate the conditional expectations with respect to $\mathcal{F}_{t_{k-1}}$ up to $(t_k)_{\varepsilon,t_{k-1}}=(t_k)_\varepsilon\vee t_{k-1}$. In order to rewrite $Q_{n,t}(\varepsilon,u)$ let for $1\le j\le N_k$

$$U_{k,j}=Z_{a_{k,j}}(\varepsilon,t_{k-1})-Z_{a_{k,j-1}}=\sigma_{a_{k,j-1}}\left(W_{a_{k,j}}-W_{a_{k,j-1}}\right)$$

and let $U_{k,0}=V_{k,0}=0$. Then $Q_{n,t}(\varepsilon,u)$ is equal to $\sum_{k=1}^n\sum_{j=1}^{N_k}(R_{k,j,1}(u)+R_{k,j,2}(u))$, where

$$\begin{aligned}R_{k,j,1}(u)&=\mathbb{E}\left[\left(e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}-e^{-i\langle u,X_{a_{k,j-1}}\rangle}\right)U^{(m)}_{k,j}\Big|\mathcal{F}_{t_{k-1}}\right]\\&=\mathbb{E}\left[e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}U^{(m)}_{k,j}\Big|\mathcal{F}_{t_{k-1}}\right]\\R_{k,j,2}(u)&=\mathbb{E}\left[\left(e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}-e^{-i\langle u,X_{a_{k,j-1}}\rangle}\right)\left(Z^{(m)}_{a_{k,j-1}}-Z^{(m)}_{t_{k-1}}\right)\Big|\mathcal{F}_{t_{k-1}}\right]\\&=\mathbb{E}\Bigg[\left(e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}-e^{-i\langle u,X_{a_{k,j-1}}\rangle}e^{-\frac12\|\sigma^\top_{a_{k,j-1}}u\|^2\left(a_{k,j}-a_{k,j-1}\right)}\right)\\&\qquad\cdot\left(Z^{(m)}_{a_{k,j-1}}-Z^{(m)}_{t_{k-1}}\right)\Bigg|\mathcal{F}_{t_{k-1}}\Bigg].\end{aligned}$$

Then $X_{a_{k,j}}(\varepsilon,t_{k-1})=X_{a_{k,j-1}}+\tilde{b}_{a_{k,j-1}}(a_{k,j}-a_{k,j-1})+U_{k,j}$ and thus

$$\begin{aligned}&\mathbb{E}\left[e^{-i\langle u,X_{a_{k,j}}(\varepsilon,t_{k-1})\rangle}\Big|\mathcal{F}_{a_{k,j-1}}\right]\\&\quad=e^{-i\langle u,X_{a_{k,j-1}}+\tilde{b}_{a_{k,j-1}}(a_{k,j}-a_{k,j-1})\rangle}e^{-\frac12\|\sigma^\top_{a_{k,j-1}}u\|^2(a_{k,j}-a_{k,j-1})}.\end{aligned}$$

We get therefore according to Assumption (H-α-β) that

$$\begin{aligned}|R_{k,j,2}(u)| &\lesssim \|u\| \left(a_{k,j}-a_{k,j-1}\right) e^{-\frac{\|u\|^2}{2C}\left(a_{k,j}-a_{k,j-1}\right)} \\ &\quad\cdot \mathbb{E}\left[\left\|Z_{a_{k,j-1}}-Z_{t_{k-1}}\right\|\middle|\mathcal{F}_{t_{k-1}}\right] \\ &\lesssim \|u\| \left(a_{k,j}-a_{k,j-1}\right) e^{-\frac{\|u\|^2}{2C}\left(a_{k,j}-a_{k,j-1}\right)},\end{aligned}$$

trivially upper bounding the conditional expectation. On the other hand, since $U_{k,j} \overset{d}{\sim} (\sigma_{a_{k,j-1}}\sigma_{a_{k,j-1}}^{\top})^{1/2}(a_{k,j}-a_{k,j-1})^{1/2}Z$ conditional on $\mathcal{F}_{a_{k,j-1}}$ with $Z \overset{d}{\sim} N(0,I)$ independent of $\mathcal{F}_{a_{k,j-1}}$, $R_{k,j,1}(u)$ is equal to

$$e^{-i\langle u, X_{a_{k,j-1}}+\tilde{b}_{a_{k,j-1}}(a_{k,j}-a_{k,j-1})\rangle}\mathbb{E}\left[e^{-i\langle u,U_{k,j}\rangle}U_{k,j}^{(m)}\middle|\mathcal{F}_{a_{k,j-1}}\right]. \tag{4.2.2}$$

Using integration by parts shows that this equals $\|u\|(a_{k,j}-a_{k,j-1})e^{-\frac{\|u\|^2}{2C}\left(a_{k,j}-a_{k,j-1}\right)}$. In all, for $2 \leq j \leq N_k$ with $a_{k,j}-a_{k,j-1}=\varepsilon$,

$$|R_{k,j,1}(u)+R_{k,j,2}(u)| \lesssim \|u\| \left(a_{k,j}-a_{k,j-1}\right) e^{-\frac{\|u\|^2}{2C}\varepsilon},$$

such that with $\sum_{j=2}^{N_k}(a_{k,j}-a_{k,j-1}) \leq t_k - t_{k-1}$ and $R_{k,1,2}(u)=0$

$$|Q_{n,t}(\varepsilon,u)| \lesssim \|u\|e^{-\frac{\|u\|^2}{2C}\varepsilon} + \left|\sum_{k=1}^{n} R_{k,1,1}(u)\right|.$$

Step 3. With respect to $R_{k,1,1}(u)$ the last step shows only

$$|R_{k,1,1}(u)| \lesssim \|u\| \left(a_{k,1}-t_{k-1}\right) e^{-\frac{\|u\|^2}{2C}\left(a_{k,1}-t_{k-1}\right)}, \tag{4.2.3}$$

and $a_{k,1}-t_{k-1}$ is smaller than ε. In order to obtain the improved factor $e^{-\frac{\|u\|^2}{2C}\varepsilon}$ instead, we use an additional approximation argument. Let $I_j = \{k=1,\ldots,n : (j-1)\varepsilon \leq t_{k-1} < j\varepsilon\}$ for $1 \leq j \leq \lceil T/\varepsilon \rceil$ be the set of blocks k with left endpoints t_{k-1} inside the intervals $[(j-1)\varepsilon, j\varepsilon)$. Let $A_j(u) = \sum_{k\in I_j} R_{k,1,1}(u)$. Then $\sum_{k=1}^{n} R_{k,1,1}(u)$ equals

$$\sum_{j=1}^{\lceil T/\varepsilon \rceil} \left(A_j(u) - \mathbb{E}\left[A_j(u)\middle|\mathcal{F}_{(j-1)\varepsilon}\right]\right) + \sum_{j=1}^{\lceil T/\varepsilon \rceil} \mathbb{E}\left[A_j(u)\middle|\mathcal{F}_{(j-1)\varepsilon}\right].$$

$A_j(u)$ is $\mathcal{F}_{j\varepsilon}$-measurable and vanishes if I_j is empty. It follows by the Burkholder-Davis-Gundy inequality that $\mathbb{E}[|\sum_{k=1}^{n} R_{k,1,1}(u)|^2]$ is up to a constant upper bounded by

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=1}^{\lceil T/\varepsilon \rceil} |A_j(u)|^2\right] + \mathbb{E}\left[\sum_{j=1}^{\lceil T/\varepsilon \rceil} \left|\mathbb{E}\left[A_j(u)\middle|\mathcal{F}_{(j-1)\varepsilon}\right]\right|^2\right] \\ &+ \mathbb{E}\left[\left|\sum_{j=1}^{\lceil T/\varepsilon \rceil} \mathbb{E}\left[A_j(u)\middle|\mathcal{F}_{(j-1)\varepsilon}\right]\right|^2\right].\end{aligned}$$

Since $\|u\|(a_{k,1}-t_{k-1})^{1/2}e^{-\frac{\|u\|^2}{2C}(a_{k,1}-t_{k-1})}\lesssim 1$, (4.2.3) shows for any $j\geq 1$ and with $a_{k,1}-t_{k-1}\leq t_k-t_{k-1}$ that $|A_j(u)|^2\lesssim t_k-t_{k-1}$. Consequently,

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}R_{k,1,1}(u)\right|^2\right]^{1/2}\lesssim 1+\mathbb{E}\left[\left|\sum_{j=1}^{\lceil T/\varepsilon\rceil}\mathbb{E}\left[A_j(u)|\,\mathcal{F}_{(j-1)\varepsilon}\right]\right|^2\right]^{1/2}$$
$$\lesssim 1+\mathbb{E}\left[\left|\sum_{j=2}^{\lceil T/\varepsilon\rceil}\mathbb{E}\left[A_j(u)|\,\mathcal{F}_{(j-2)\varepsilon}\right]\right|^{1/2}\right]^{1/2}, \tag{4.2.4}$$

where we repeat the arguments above for the second inequality. In order to estimate the conditional expectations let $k\in I_j$, $2\leq j\leq\lceil T/\varepsilon\rceil$, such that $(t_{k-1})_\varepsilon=(j-2)\varepsilon$. We will approximate

$$\mathbb{E}[R_{k,1,1}(u)|\mathcal{F}_{(j-2)\varepsilon}]=\mathbb{E}\left[\left(e^{-i\langle u,X_{a_{k,1}}(\varepsilon,t_{k-1})\rangle}-e^{-i\langle u,X_{t_{k-1}}\rangle}\right)U_{k,j}^{(m)}\middle|\mathcal{F}_{(j-2)\varepsilon}\right]$$

by

$$\tilde{R}_k(u)=\mathbb{E}\left[\left(e^{-i\langle u,\tilde{X}_{a_{k,1}}(\varepsilon,a_{k,1}-t_{k-1})\rangle}-e^{-i\langle u,\tilde{X}_{t_{k-1}}(\varepsilon,0)\rangle}\right)\tilde{U}_k^{(m)}\middle|\mathcal{F}_{(j-2)\varepsilon}\right],$$

where $\tilde{X}_t(\varepsilon,h)$, $0\leq h\leq t$, is defined in (A.3.2), and where $\tilde{U}_k=\sigma_{(j-2)\varepsilon}\left(W_{a_{k,1}}-W_{t_{k-1}}\right)$. Define also $\tilde{R}_{k,1}(u)$ as $R_{k,1,1}(u)$, but with $\tilde{U}_k$ instead of $U_{k,1}$. Then the Cauchy-Schwarz inequality implies (cf. (4.2.1)) that $|\mathbb{E}[R_{k,1,1}(u)-\tilde{R}_{k,1}(u)|\mathcal{F}_{(j-2)\varepsilon}]|$ is up to a constant bounded by

$$\|u\|\mathbb{E}\left[\|X_{a_{k,1}}(\varepsilon,t_{k-1})-X_{t_{k-1}}\|^2\middle|\,\mathcal{F}_{(j-2)\varepsilon}\right]^{1/2}\mathbb{E}\left[\|U_{k,1}-\tilde{U}_k\|^2\middle|\,\mathcal{F}_{(j-2)\varepsilon}\right]^{1/2},$$

which is of order $O(\|u\|(a_{k,1}-t_{k-1})(\varepsilon^{\beta+1/2}+\varepsilon^\alpha))$ by Proposition A.3.2 (with $X_{t_{k-1}}=X_{(a_{k,1})_\varepsilon,t_{k-1}}$). On the other hand,

$$e^{-i\left\langle u,X_{a_{k,1}}(\varepsilon,t_{k-1})\right\rangle}-e^{-i\left\langle u,X_{t_{k-1}}\right\rangle}-\left(e^{-i\left\langle u,\tilde{X}_{a_{k,1}}\left(\varepsilon,a_{k,1}-t_{k-1}\right)\right\rangle}-e^{-i\left\langle u,\tilde{X}_{t_{k-1}}(\varepsilon,0)\right\rangle}\right)$$
$$=\left(e^{-i\left\langle u,X_{t_{k-1}}\right\rangle}-e^{-i\left\langle u,\tilde{X}_{t_{k-1}}(\varepsilon,0)\right\rangle}\right)\left(e^{-i\left\langle u,X_{a_{k,1}}(\varepsilon,t_{k-1})-X_{t_{k-1}}\right\rangle}-1\right)$$
$$+e^{-i\left\langle u,\tilde{X}_{t_{k-1}}(\varepsilon,0)\right\rangle}\left(e^{-i\left\langle u,X_{a_{k,1}}(\varepsilon,t_{k-1})-X_{t_{k-1}}\right\rangle}-e^{-i\left\langle u,\tilde{X}_{a_{k,1}}\left(\varepsilon,a_{k,1}-t_{k-1}\right)-\tilde{X}_{t_{k-1}}(\varepsilon,0)\right\rangle}\right).$$

Proposition A.3.2 then yields for $|\mathbb{E}[\tilde{R}_{k,1}(u)-\tilde{R}_k(u)|\mathcal{F}_{(j-2)\varepsilon}]|$ up to a constant the upper bound

$$\left(\|u\|^2\left(\varepsilon^{\beta+1}+\varepsilon^{\alpha+1/2}\right)+\|u\|\left(\varepsilon^{\beta+1/2}+\varepsilon^\alpha\right)\right)(a_{k,1}-t_{k-1}).$$

In all, with $\sum_{k\in I_j}(a_{k,1}-t_{k-1})\leq 2\varepsilon$, we get

$$\mathbb{E}\left[\left|\sum_{j=2}^{\lceil T/\varepsilon\rceil}\left(\mathbb{E}\left[A_j(u)|\,\mathcal{F}_{(j-2)\varepsilon}\right]-\sum_{k\in I_j}\tilde{R}_k(u)\right)\right|^2\right]^{1/2}$$
$$\lesssim\|u\|^2\left(\varepsilon^{\beta+1}+\varepsilon^{\alpha+1/2}\right)+\|u\|\left(\varepsilon^{\beta+1/2}+\varepsilon^\alpha\right).$$

Together with (4.2.4) this means

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} R_{k,1,1}(u)\right|^2\right]^{1/2} \lesssim 1 + \|u\|^2\left(\varepsilon^{\beta+1}+\varepsilon^{\alpha+1/2}\right) + \|u\|\left(\varepsilon^{\beta+1/2}+\varepsilon^{\alpha}\right) + \mathbb{E}\left[\left|\sum_{j=2}^{\lceil T/\varepsilon\rceil}\sum_{k\in I_j}\tilde{R}_k(u)\right|^2\right]^{1/2}.$$

Step 4. Since $\tilde{U}_k$ are centered conditional on $\mathcal{F}_{t_{k-1}}$, $\tilde{R}_k(u)$ is equal to

$$\begin{aligned}&\mathbb{E}\left[e^{-i\langle u,\tilde{X}_{a_{k,1}}(\varepsilon,a_{k,1}-t_{k-1})\rangle}\tilde{U}_k^{(m)}\middle|\mathcal{F}_{(j-2)\varepsilon}\right]\\ &= e^{-i\left\langle u,X_{(j-2)\varepsilon}+\tilde{b}_{(j-2)\varepsilon}(a_{k,1}-(j-2)\varepsilon)\right\rangle}\mathbb{E}\left[e^{-i\left\langle u,\tilde{U}_k\right\rangle}\tilde{U}_k^{(m)}\middle|\mathcal{F}_{(j-2)\varepsilon}\right]\\ &\quad\cdot\mathbb{E}\left[e^{-i\langle u,\sigma_{(j-2)\varepsilon}\left(W_{t_{k-1}}-W_{(j-2)\varepsilon}\right)}\middle|\mathcal{F}_{(j-2)\varepsilon}\right].\end{aligned}$$

The last line is of order $e^{-\frac{\|u\|^2}{2C}(t_{k-1}-(j-2)\varepsilon)}\leq e^{-\frac{\|u\|^2}{2C}\varepsilon}$ (with $(j-2)\varepsilon=(t_{k-1})_\varepsilon$) and the first one is of order $\|u\|(a_{k,1}-t_{k-1})e^{-\frac{\|u\|^2}{2C}(a_{k,1}-t_{k-1})}$ by arguing as after (4.2.2). We conclude that

$$\mathbb{E}\left[\left|\sum_{j=2}^{\lceil T/\varepsilon\rceil}\sum_{k\in I_j}\tilde{R}_k(u)\right|^2\right]^{1/2}\lesssim\|u\|e^{-\frac{\|u\|^2}{2C}\varepsilon}.$$

Step 5. In all we have shown that

$$\begin{aligned}\|Q_n(u)\|_{L^2(\mathbb{P})}&\leq\|Q_n(u)-Q_n(\varepsilon,u)\|_{L^2(\mathbb{P})}+\|Q_n(\varepsilon,u)\|_{L^2(\mathbb{P})}\\ &\lesssim 1+\|u\|\left(\varepsilon^{(\beta+1/2)}+\varepsilon^{\alpha}\right)+\|u\|^2\left(\varepsilon^{\beta+1}+\varepsilon^{\alpha+1/2}\right)+\|u\|e^{-\frac{\|u\|^2}{2C}\varepsilon}.\end{aligned}$$

Choose $\varepsilon=\varepsilon(u)=2C\|u\|^{-2}\log(1+\|u\|^3)$ which is bounded, continuous and tends to zero as $\|u\|\to0$ or $\|u\|\to\infty$. The last line therefore implies $(1+\|u\|)^{-s}\|Q_n(u)\|_{L^2(\mathbb{P})}\lesssim1$ for $\|u\|\leq1/2$. For $\|u\|>1/2$, on the other hand, note that $e^{-\frac{\|u\|^2}{2C}\varepsilon}=(1+\|u\|^3)^{-1}\leq\|u\|^{-3}$ such that for an absolute constant $c>0$

$$(1+\|u\|)^{-s}\|Q_n(u)\|_{L^2(\mathbb{P})}\lesssim1+\left(\|u\|^{-s-2\beta}+\|u\|^{1-s-2\alpha}\right)\log\left(c\|u\|^3\right)^2.$$

This is uniformly bounded in u and n using the conditions on s, α, β and that $\|u\|^{-r}\log(c\|u\|^q)$ is bounded for any $r,q>0$. This yields the claim of the lemma. □

4.2.1. Proof of Proposition 4.1.1

With the help of the last lemma we will now first derive upper bounds on $S_{t,\pi_n}(f)$.

Lemma 4.2.2. *Assume (H-α-β) for* $0 \le \alpha, \beta \le 1$. *Let* $s > 1 - 2\alpha$, $s \ge 0$, $s + \beta > 0$. *Let* $(\pi_n)_{n \ge 1}$ *be a sequence of partitions with* $|\pi_n| \to 0$ *and let* $1 \le m \le d$. *Then we have for any* $f \in FL^s(\mathbb{R}^d)$ *that* $\|\sup_{0 \le t \le T} S_{t,\pi_n}(f)\|_{L^1(\mathbb{P})} \lesssim \|f\|_{FL^s}$.

Proof. Recall the decomposition $S_{t,\pi_n}(f) = M_{n,t}(f) + D_{n,t}(f)$ from the beginning of this section. By a density argument it is sufficient to show $\|\sup_{0 \le t \le T} S_{t,\pi_n}\|_{L^1(\mathbb{P})} \lesssim \|f\|_{FL^s}$ for $f \in \mathcal{S}(\mathbb{R}^d)$, which means we only have to show $\|\sup_{0 \le t \le T} M_{n,t}\|_{L^1(\mathbb{P})} \lesssim \|f\|_{FL^s}$, $\|\sup_{0 \le t \le T} D_{n,t}\|_{L^1(\mathbb{P})} \lesssim \|f\|_{FL^s}$. Since $M_{n,t}$ is a discrete martingale, it follows from the Burkholder-Davis-Gundy inequality that

$$\begin{aligned}\|\sup_{0 \le t \le T} M_{n,t}\|^2_{L^1(\mathbb{P})} \le \|\sup_{0 \le t \le T} M_{n,t}\|_{L^2(\mathbb{P})} &= \sum_{k=1}^n \|A_k - \mathbb{E}\left[A_k \middle| \mathcal{F}_{t_{k-1}}\right]\|^2_{L^2(\mathbb{P})} \\ &\lesssim \sum_{k=1}^n \|A_k\|^2_{L^2(\mathbb{P})} \lesssim \|f\|^2_\infty \sum_{k=1}^n \mathbb{E}[\|X_{t_k} - X_{t_{k-1}}\|^2] \lesssim \|f\|^2_\infty.\end{aligned}$$

This yields $\|\sup_{0 \le t \le T} M_{n,t}\|^2_{L^1(\mathbb{P})} \lesssim \|f\|^2_{FL^s}$ by the inequality $|f(x)| = (2\pi)^{-d} |\int \mathcal{F}f(u) e^{-i\langle u,x\rangle} du| \lesssim \int |\mathcal{F}f(u)| du$ for $x \in \mathbb{R}^d$. Since the estimates in the proof of Lemma 4.2.1 are essentially pathwise, we can further show that

$$\sup_{u \in \mathbb{R}^d} ((1 + \|u\|)^{-s} \sup_{0 \le t \le T} \|Q_{n,t}(u)\|_{L^1(\mathbb{P})}) < \infty.$$

Therefore, we have as at the beginning of this section that

$$\|\sup_{0 \le t \le T} D_{n,t}\|_{L^1(\mathbb{P})} \lesssim \sup_{u \in \mathbb{R}^d} ((1 + \|u\|)^{-s} \sup_{0 \le t \le T} \|Q_{n,t}(u)\|_{L^1(\mathbb{P})}) \|f\|_{FL^s} \lesssim \|f\|_{FL^s}.$$

□

Proof of Proposition 4.1.1. Let $(\pi_n)_{n \ge 1}$ be a sequence of partitions with $|\pi_n| \to 0$. Let $1 \le m \le d$ be fixed. Consider the stopping times $(\rho_K)_{K \ge 1}$ and corresponding processes $(X_t(K))_{0 \le t \le T}$ as defined in Section A.2. We already know that $S_{t,\pi_n}(f, X(K)) - [f(X(K)), X(K)^{(m)}]_t \to 0$ in $L^1(\mathbb{P})$ for all $0 \le t \le T$ and $f \in \mathcal{S}(\mathbb{R}^d) \subset C^1(\mathbb{R}^d)$ (cf. Russo and Vallois (1996, Proposition 1.1)). A simple calculation shows that even $S_{t,\pi_n}(f, X(K)) - [f(X(K)), X(K)^{(m)}]_t \xrightarrow{ucp} 0$. Since $\mathcal{S}(\mathbb{R}^d)$ is dense in $FL^s(\mathbb{R}^d)$ with respect to $\|\cdot\|_{FL^s}$, we can extend $f \mapsto ([f(X(K)), X(K)^{(m)}]_t)_{0 \le t \le T}$ uniquely to an operator on $FL^s(\mathbb{R}^d)$. Lemma 4.2.2 yields that the operators $f \mapsto (S_{t,\pi_n}(f, X(K)))_{0 \le t \le T}$ are equicontinuous on $FL^s(\mathbb{R}^d)$ and therefore $S_{t,\pi_n}(f, X(K)) - [f(X(K)), X(K)^{(m)}]_t \xrightarrow{ucp} 0$ holds also for $f \in FL^s(\mathbb{R}^d)$.

Consider now $f \in FL^s_{loc}(\mathbb{R}^d)$ and let $f^{(K)} = f\varphi$ for a smooth cutoff function φ as in Section 3.5.1 such that $f^{(K)} \in FL^s(\mathbb{R}^d)$ has compact support and $f^{(K)} = f$ on $\{x \in \mathbb{R}^d : \|x\| \le K\}$. It follows that $S_{t,\pi_n}(f, X(K)) = S_{t,\pi_n}(f^{(K)}, X(K)) \xrightarrow{ucp} [f(X(K)), X(K)^{(m)}]_t$. Since $S_{t,\pi_n}(f^{(K+1)}, X(K+1)^{(m)}) = S_{t,\pi_n}(f^{(K)}, X(K)^{(m)})$ almost surely on $\{\rho_K > t\}$, we also have $[f(X(K+1)), X(K+1)^{(m)}]_t = [f(X(K)), X(K)^{(m)}]_t$ almost surely on $\{\rho_K > t\}$ and therefore it is justified to set $[f(X), X^{(m)}]_t := [f(X(K)), X(K)^{(m)}]_t$ on $\{\rho_K > t\}$. In particular, as $S_{t,\pi_n}(f, X) = S_{t,\pi_n}(f^{(K)}, X(K))$ on $\{\rho_K > t\}$, we conclude that $S_{t,\pi_n}(f, X) \xrightarrow{ucp} [f(X), X^{(m)}]_t$. □

4.2.2. Proof of Proposition 4.1.2

Observe first the following corresponding statement to Lemma 4.2.2.

Lemma 4.2.3. *Assume (H-α-β) for $0 \leq \alpha, \beta \leq 1$ and (X0). Let $(\pi_n)_{n\geq 1}$ be a sequence of partitions with $|\pi_n| \to 0$. Let $s > 1 - 2\alpha$, $s \geq 0$, $s + \beta > 0$. Then we have for $f \in H^s(\mathbb{R}^d)$ that $\|S_{t,\pi_n}(f)\|_{L^2(\mathbb{P})} \leq C_2 \|f\|_{H^s}$, $0 \leq t \leq T$, for an absolute constant C_2.*

Proof. As in Lemma 4.2.2 we only have to show that $\|M_{n,t}\|_{L^2(\mathbb{P})} \lesssim \|f\|_{H^s}$, $\|D_{n,t}\|_{L^2(\mathbb{P})} \lesssim \|f\|_{H^s}$. By a density argument it is sufficient to consider $f \in \mathcal{S}(\mathbb{R}^d)$. By the Burkholder-Davis-Gundy inequality it follows that

$$\|M_{n,t}\|^p_{L^2(\mathbb{P})} \lesssim \mathbb{E}\left[\sum_{k=1}^n \left|A_k - \mathbb{E}\left[A_k\right|\mathcal{F}_{t_{k-1}}\right]\right|^2\right] \lesssim \mathbb{E}\left[\sum_{k=1}^n A_k^2\right].$$

Fourier inversion and (X0) yield for $\sum_{k=1}^n A_k^2$ up to a constant the bound

$$\left(\int |\mathcal{F}f(u)|\,|\mathcal{F}f(v)|\,|\mathcal{F}\mu(u+v)|\,d(u,v)\right)\sum_{k=1}^n\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right)^2,$$

which is bounded by $C\|f\|^2_{H^s}\sum_{k=1}^n\left(X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}}\right)^2$ from Lemma 3.5.3. It holds that $\mathbb{E}[\sum_{k=1}^n (X^{(m)}_{t_k} - X^{(m)}_{t_{k-1}})^2] \lesssim 1$. Consequently, $\|M_{n,t}\|_{L^2(\mathbb{P})} \lesssim \|f\|_{H^s}$. With respect to $D_{n,t}$ we find from Fourier inversion and independence via (X0) that $\|D_{n,t}\|^2_{L^2(\mathbb{P})}$ is up to a constant bounded by

$$\int |\mathcal{F}f(u_1)\,\mathcal{F}f(u_2)|\,\mathcal{F}\mu(u_1+u_2)\,|\mathbb{E}\left[Q_{t,n}(u_1)\,Q_{t,n}(u_2)\right]|\,d(u_1,u_2),$$

with $Q_{t,n}(u)$ as in Lemma 4.2.1. Then Lemmas 4.2.1 and 3.5.3(*ii*) immediately yield $\|D_{n,t}\|^2_{L^2(\mathbb{P})} \lesssim \|f\|^2_{H^s}$. □

Proof of Corollary 4.1.3. We repeat the argument from Proposition 4.1.1. Let $(\pi_n)_{n\geq 1}$ be a sequence of partitions with $|\pi_n| \to 0$ and let $1 \leq m \leq d$ be fixed. As in Proposition 4.1.1 we know that $S_{t,\pi_n}(f, X(K)) \to [f(X(K)), X(K)^{(m)}]_t$ in $L^1(\mathbb{P})$ for all $0 \leq t \leq T$ and $f \in \mathcal{S}(\mathbb{R}^d) \subset C^1(\mathbb{R}^d)$. This time, since $\mathcal{S}(\mathbb{R}^d)$ is dense in $H^s(\mathbb{R}^d)$ with respect to $\|\cdot\|_{H^s}$, we can extend $f \mapsto [f(X(K)), X(K)^{(m)}]_t$ uniquely to an operator on $H^s(\mathbb{R}^d)$. Lemma 4.2.3 yields that the operators $f \mapsto S_{t,\pi_n}(f, X(K))$ are equicontinuous from $H^s(\mathbb{R}^d)$ to $L^2(\mathbb{P})$ and therefore $S_{t,\pi_n}(f, X(K)) \xrightarrow{\mathbb{P}} [f(X(K)), X(K)^{(m)}]_t$ holds also for $f \in H^s(\mathbb{R}^d)$. The same localization argument as in Proposition 4.1.1 allows us to conclude that $S_{\pi_n}(f, X) \xrightarrow{\mathbb{P}} [f(X), X^{(m)}]_t$ for $f \in H^s_{loc}(\mathbb{R}^d)$. □

4.2.3. Proof of Corollary 4.1.3

Proof. Let $(\pi_n)_{n\geq 1}$ be a sequence of partitions with $|\pi_n| \to 0$ and fix $1 \leq m \leq d$. We argue as in Corollary 3.1.6. For simplified notation assume $X_0 = 0$ such that the process $X + x_0$ has initial value $x_0 \in \mathbb{R}^d$. Let $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{0\leq t\leq T}, \mathbb{P}')$ be an extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ such that there is a random variable $Y \overset{d}{\sim} N(0, I_d)$ which is independent of $\mathcal{F}$ and such that Y is $\mathcal{F}'_0$-measurable, where I_d is the d-dimensional identity matrix. On this space the process $X+Y$ satisfies Assumption (X0). Proposition 4.1.2 implies therefore $S_{t,\pi_n}(f, X+Y) \xrightarrow{\mathbb{P}'} [f(X+Y), X]_t$. In particular, $x_0 \mapsto [f(X+x_0), X]_t$ is measurable. It follows as in Corollary 3.1.6 by a coupling argument that $S_{t,\pi_n}(f, X+x_0) \xrightarrow{\mathbb{P}} [f(X+x_0), X]_t$ for almost all x_0 (argue by almost sure convergence on the extended space that $Z_n = |S_{t,\pi_n}(f, X+Y) - [f(X+Y), X]_t|$ converges almost surely to zero and apply the bounded test functions $\mathbf{1}_{(\varepsilon,\infty)}(x)$, $\varepsilon \in \mathbb{Q}$, to the Z_n). □

4.2.4. Proof of Theorem 4.1.5

Proof. In order to prove (i) consider the stopping times $(\rho_K)_{K\geq 1}$, processes $(X_t(K))_{0\leq t\leq T}$ and localized functions $f^{(K)} \in FL^s(\mathbb{R}^d)$ for $f \in FL^s_{loc}(\mathbb{R}^d)$ from the proof of Proposition 4.1.1. Let $(f_n^{(K)}) \subset C_c^\infty(\mathbb{R}^d)$ be a sequence of functions converging to $f^{(K)}$ as $n \to \infty$ with respect to $\|\cdot\|_{FL^s}$ such that $\|f_n^{(K)}\|_{FL^s} \leq \|f^{(K)}\|_{FL^s}$. The classical Itô formula together with Proposition 1.1. of Russo and Vallois (1996) yields $F(f_n^{(K)}, X(K)) = G(f_n^{(K)}, X(K))$, where

$$F\left(f_n^{(K)}, X\left(K\right)\right) := f_n^{(K)}\left(X_t(K)\right) - f_n^{(K)}\left(X_0(K)\right) - \int_0^t \langle \nabla f_n^{(K)}(X_r(K)), dX_r(K)\rangle$$

and $G(f_n^{(K)}, X(K)) := \frac{1}{2}\sum_{m=1}^d [\partial_m f_n^{(K)}(X(K)), X(K)^{(m)}]_t$. $f_n^{(K)} \to f^{(K)}$ with respect to $\|\cdot\|_{FL^s}$ implies also $f_n^{(K)} \to f^{(K)}$ uniformly and thus it follows by dominated convergence for stochastic integrals (Revuz and Yor (1999, Theorem 2.12)) that $F(f_n^{(K)}, X(K)) \to F(f^{(K)}, X(K))$ in probability. Moreover, the proof of Lemma 4.2.2 shows that $\|G(f_n^{(K)}, X(K))\|_{L^1(\mathbb{P})} \lesssim \|f_n^{(K)}\|_{FL^s}$ and therefore $G(f_n^{(K)}, X(K)) \to G(f^{(K)}, X(K))$ in probability. From $F(f, X) = F(f^{(K)}, X(K))$ and $G(f, X) = G(f^{(K)}, X(K))$ on $\{\rho_K > t\}$ according to the proof of Proposition 4.1.1, the claim of (i) follows.

With respect to (ii) the argument is similar. This time $f^{(K)} \in H^s(\mathbb{R}^d)$ for $f \in H^s_{loc}(\mathbb{R}^d)$ and the sequence $(f_n^{(K)}) \subset C_c^\infty(\mathbb{R}^d)$ converges to $f^{(K)}$ with respect to $\|\cdot\|_{H^s}$ with $\|f_n^{(K)}\|_{H^s} \leq \|f^{(K)}\|_{H^s}$. Then we can argue as above, noting that $\|G(f_n^{(K)}, X(K))\|_{L^2(\mathbb{P})} \lesssim \|f_n^{(K)}\|_{H^s}$ by Lemma 4.2.3 and using that $f_n^{(K)}(X_r(K)) \to f^{(K)}(X_r(K))$, $\partial_m f_n^{(K)}(X_r(K)) \to \partial_m f^{(K)}(X_r(K))$ in $L^2(\mathbb{R}^d)$ for all $0 \leq r \leq T$, $1 \leq m \leq d$ (argue as in (3.5.1)).

For (iii) assume that $X_0 = 0$ and that Y is as in the proof of Corollary 4.1.3. Then $F(f, X+Y) = G(f, X+Y)$ $\mathbb{P}'$-almost surely by (ii). In particular, $F(f, X+x_0) = G(f, X+x_0)$ $\mathbb{P}$-almost surely for all x_0 in some set $\tilde{E} \subset \mathbb{R}^d$, where $\mathbb{R}^d \backslash \tilde{E}$ has Lebesgue measure zero. By taking the intersection with the set E from Proposition 4.1.1(iii), we can assume without loss of generality that $G(f, X+x_0) = \frac{1}{2}\sum_{m=1}^d [\partial_m f(X+x_0), X^{(m)}]_t$ exists for all $x_0 \in E$. This yields (iii). □

Appendix A.

Technical tools

This appendix provides a number of important tools for the results in this thesis. We first define stable convergence for sequences of random variables and describe some of its properties. Then the localization procedure for continuous Itô semimartingales is presented, which is frequently applied for proving the central limit theorems and the generalized Itô formulas. The third section defines and discusses the Fourier-Lebesgue spaces. We conclude with a short review of Markov semigroup theory and the functional calculus for normal operators.

A.1. Stable convergence

The central limit theorems in Chapters 2 and 3 are based on the concept of *stable convergence*, introduced by Rényi (1963).

Definition A.1.1. Let $(Y_n)_{n\geq 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a Polish space $(E, \mathcal{E})$. We say that Y_n converges *stably* to Y, written $Y_n \xrightarrow{st} Y$, if Y is defined on an extension $(\Omega', \mathcal{F}', \mathbb{P}')$ of the original probability space and if $(Y_n, U) \xrightarrow{d} (Y, U)$ for all $\mathcal{F}$-measurable random variables U.

Equivalently, $Y_n \xrightarrow{st} Y$ if and only if $\mathbb{E}[Uf(Y_n)] \to \mathbb{E}[Uf(Y)]$ for all bounded measurable functions f and all bounded $\mathcal{F}$-measurable random variables U (Jacod and Shiryaev (2013, Definition VIII 5.28)). In that sense stable convergence can be deemed *weak L^1-convergence.* Stable convergence is stronger than convergence in distribution and allows to standardize estimators when the parameters of interest are random.

Let now Y_n and Y be stochastic processes with values in the Skorokhod space $\mathcal{D}([0,T], \mathbb{R}^d)$, $T > 0$, such that $Y_n \xrightarrow{st} Y$. It follows that $(Y_n)_t \xrightarrow{st} Y_t$ for all $0 \leq t \leq T$ if Y is continuous and that $Y_n + Z_n \xrightarrow{st} Y + Z$, if Z_n and Z are processes in $\mathcal{D}([0,T], \mathbb{R}^d)$ with $\sup_{0\leq t\leq T} \|(Z_n)_t - Z_t\| \xrightarrow{\mathbb{P}} 0$ (Billingsley (2013, Chapter 3)). The last convergence $\sup_{0\leq t\leq T} \|(Z_n)_t - Z_t\| \xrightarrow{\mathbb{P}} 0$ is also called *uniform convergence in probability*, which is denoted by $Z_n \xrightarrow{ucp} Z$. For more details on stable convergence we refer to Jacod and Shiryaev (2013). Examples can be found in Podolskij and Vetter (2010).

Proving stable convergence of stochastic processes is generally difficult. An important tool is the following theorem, which appeared in its probably earliest form already in Genon-Catalot and Jacod (1993). We use only a special case adapted to our needs.

Theorem A.1.2 (Jacod and Shiryaev (2013, Theorem 7.28))**.** *Consider on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ a sequence of processes Y_n defined by $(Y_n)_t =$*

$\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \xi_{n,k}$, $\Delta_n = T/n$, *such that each* $\xi_{n,k}$ *is a real-valued* $\mathcal{F}_{k\Delta_n}$*-measurable and square integrable random variable. Let* $W = (W_t)_{0\le t\le T}$ *be a continuous d-dimensional Brownian motion with respect to* $(\mathcal{F}_t)_{0\le t\le T}$*. Assume that there are continuous and adapted* $\mathbb{R}^d$*-valued processes* $(u_t)_{0\le t\le T}$, $(w_t)_{0\le t\le T}$*, zero at* $t = 0$*, such that the following conditions are satisfied:*

$$\sup_{0\le t\le T} \| \sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[\xi_{n,k}\middle| \mathcal{F}_{(k-1)\Delta_n}\right] - B_t\| \xrightarrow{\mathbb{P}} 0, \tag{A.1.1}$$

$$\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \left(\mathbb{E}\left[\xi_{n,k}^2\middle| \mathcal{F}_{(k-1)\Delta_n}\right] - \mathbb{E}\left[\xi_{n,k}\middle| \mathcal{F}_{(k-1)\Delta_n}\right]^2\right) \xrightarrow{\mathbb{P}} \int_0^t \left(\|u_r\|^2 + \|w_r\|^2\right) dr, \tag{A.1.2}$$

$$\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\Big[\, \xi_{n,k}^2 \mathbf{1}_{\{|\xi_{n,k}|>\varepsilon\}} \Big| \mathcal{F}_{(k-1)\Delta_n}\Big] \xrightarrow{\mathbb{P}} 0, \quad \textit{for all } \varepsilon > 0, \tag{A.1.3}$$

$$\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[\xi_{n,k} \left(W_{k\Delta_n} - W_{(k-1)\Delta_n}\right)^\top \Big| \mathcal{F}_{(k-1)\Delta_n}\right] \xrightarrow{\mathbb{P}} \int_0^t u_r^\top dr, \tag{A.1.4}$$

$$\sum_{k=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\left[\xi_{n,k} \left(N_{k\Delta_n} - N_{(k-1)\Delta_n}\right)\middle| \mathcal{F}_{(k-1)\Delta_n}\right] \xrightarrow{\mathbb{P}} 0, \tag{A.1.5}$$

where (A.1.5) holds for all bounded real valued martingales N which are orthogonal to all components of W. Then we have the stable convergence

$$(Y_n)_t \xrightarrow{st} B_t + \int_0^t \langle u_r, dW_r\rangle + \int_0^t \left\langle w_r, d\widetilde{W}_r\right\rangle$$

as processes on $\mathcal{D}([0,T],\mathbb{R}^d)$*, where* $\widetilde{W}$ *is a d-dimensional Brownian motion defined on an independent extension of* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$*.*

A.2. The localization procedure

Let X be a continuous Itô semimartingale as in (2.1.1). In the following we assume that $((\sigma_t\sigma_t^\top)^{-1})_{0\le t\le T}$ is almost surely bounded (cf. Assumption (SM-α-β)). We associate with X a sequence of càdlàg processes $(F_n(X))_{n\ge 1}$ on the same probability space. When proving limiting statements for $F_n(X)$, it is often convenient to *localize* X and the coefficient processes b and σ.

To see how this works note that b being locally bounded implies that there are stopping times τ_K with $\tau_K \to \infty$ as $K \to \infty$ such that $\|b_t\| \le K$ for all $0 \le t \le \tau_K \wedge T$. Since X and σ are càdlàg and because $(\sigma_t\sigma_t^\top)_{0\le t\le T}$ is almost surely invertible, we can further define stopping times $\eta_K = \inf\{0 < t \le T : \|X_t\| + \|\sigma_t\| + \|(\sigma_t\sigma_t^\top)^{-1}\| \ge K\}$, $\inf\emptyset = \infty$, with $\eta_K \to \infty$ as $K \to \infty$. Let $\rho_K = \eta_K \wedge \tau_K$ and set $b_t(K) = b_{t\wedge\rho_K}$, $\sigma_t(K) = \sigma_{t\wedge\rho_K}\mathbf{1}_{\{\|\sigma_{t\wedge\rho_K}\|\le K\}}$. We define the process $X(K) = (X_t(K))_{0\le t\le T}$ by $X_t(K) = 0$ if $\rho_K = 0$ and otherwise by $X_t(K) = X_0 + \int_0^t b_r(K)dr + \int_0^t \sigma_r(K)dW_r$. Then $X_t = X_t(K)$ almost surely for all $0 \le t \le \rho_K$ and

$$\sup_{0\le t\le T} \Big(\|X_t(K)\| + \|b_t(K)\| + \|\sigma_t(K)\| + \|(\sigma_t(K)\sigma_t^\top(K))^{-1}\|\Big) \le K.$$

We can now exploit the fact that convergence in probability and stable convergence are *stable under localization.* For this assume that there exist also càdlàg processes $F_n(X(K))$ associated with $X(K)$.

Proposition A.2.1. *Let $F(X)$ and $F(X(K))$, $K \in \mathbb{N}$, be càdlàg processes, possibly defined on an extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$. Assume that $F_n(X)_t = F_n(X(K))_t$ and $F(X)_t = F(X(K))_t$ almost surely for all $0 \leq t \leq \rho_K$ and all n.*

(i) If $F_n(X(K)) \xrightarrow{\mathbb{P}} F(X(K))$ as $n \to \infty$ for all K, then also $F_n(X) \xrightarrow{\mathbb{P}} F(X)$.

(ii) If $F_n(X(K)) \xrightarrow{st} F(X(K))$ as $n \to \infty$ for all K, then also $F_n(X) \xrightarrow{st} F(X)$.

A proof can be found in Jacod and Protter (2011, Lemma 4.4.9).

A.3. Some inequalities for Itô semimartingales

The following holds in continuous and in discrete time, as well as conditional on $\mathcal{F}_0$.

Theorem A.3.1 (Burkholder-Davis-Gundy inequalities, Revuz and Yor (1999, Theorem 4.1))**.** *For every $0 < p < \infty$ there exist two absolute constants c_p and C_p such that for all continuous local martingales $M = (M_t)_{t\geq 0}$, $M_0 = 0$, and all $t \geq 0$*

$$c_p\mathbb{E}\left[\langle M, M\rangle_t^{p/2}\right] \leq \mathbb{E}\left[\left(\sup_{0\leq s\leq t} |M_s|\right)^p\right] \leq C_p\mathbb{E}\left[\langle M, M\rangle_t^{p/2}\right].$$

Let X be a continuous Itô semimartingale as in (2.1.1). For $\varepsilon > 0$ and $t \geq 0$ let $t_\varepsilon = \max(\lfloor t/\varepsilon\rfloor\varepsilon - \varepsilon, 0)$ and $t_{\varepsilon,h} = h \vee t_\varepsilon$ such that $\varepsilon \leq t - t_\varepsilon \leq 2\varepsilon$ and $(t-h) \wedge \varepsilon \leq t - t_{\varepsilon,h} \leq (t-h) \wedge 2\varepsilon$. Define for $0 \leq h \leq t$ the approximated processes

$$X_t(\varepsilon, h) = X_{t_{\varepsilon,h}} + \tilde{b}_{t_{\varepsilon,h}}(t - t_{\varepsilon,h}) + \sigma_{t_{\varepsilon,h}}(W_t - W_{t_{\varepsilon,h}}), \tag{A.3.1}$$

$$\tilde{X}_t(\varepsilon, h) = X_{(t-h)_\varepsilon} + \tilde{b}_{(t-h)_\varepsilon}(t - (t-h)_\varepsilon) + \sigma_{(t-h)_\varepsilon}(W_t - W_{(t-h)_\varepsilon}). \tag{A.3.2}$$

We write $\tilde{X}_t(\varepsilon) = \tilde{X}_t(\varepsilon, 0)$. Then the following estimates hold by the Burkholder-Davis-Gundy inequality, applied componentwise.

Proposition A.3.2. *Assume (H-α-β) for $0 \leq \alpha, \beta \leq 1$. Then the following holds for some absolute constant C and all $0 \leq h \leq t \leq T$, $t + h \leq T$:*

(i) $\mathbb{E}\left[\|Z_{t+h} - Z_t\|^2\right] \leq Ch$ *for* $Z = X, X(\varepsilon, h), \widetilde{X}(\varepsilon, h)$,

(ii) $\mathbb{E}[\|X_t - X_t(\varepsilon, h)\|^2|\mathcal{F}_{t_{\varepsilon,h}}] \leq C\left(((t-h)\wedge\varepsilon)^{2\beta+2} + ((t-h)\wedge\varepsilon)^{2\alpha+1}\right)$,

(iii) $\mathbb{E}[\|X_{t+h} - X_t - (\widetilde{X}_{t+h}(\varepsilon) - \widetilde{X}_t(\varepsilon))\|^2] \leq Ch\left(\varepsilon^{2\beta+1} + \varepsilon^{2\alpha}\right)$,

(iv) $\mathbb{E}[\|X_{t+h}(\varepsilon, h) - X_t - (\tilde{X}_{t+h}(\varepsilon, h) - \tilde{X}_t(\varepsilon))\|^2|\mathcal{F}_{t_\varepsilon}] \leq Ch\left((\varepsilon+h)^{2\beta+1} + (\varepsilon+h)^{2\alpha}\right)$,

(v) $\mathbb{E}[\|X_t(\varepsilon, h) - X_{t_{\varepsilon,h}}\|^2|\mathcal{F}_{t_{\varepsilon,h}}] \leq C\left(((t-h)\wedge\varepsilon)^2 + ((t-h)\wedge\varepsilon)\right)$.

A.4. Fourier-Lebesgue spaces

We use extensively the following function spaces, which appear in the form below for example in Catellier and Gubinelli (2016).

Definition A.4.1. Let $s \in \mathbb{R}$, $p \geq 1$ and denote by $FL^{s,p}(\mathbb{R}^d) := \{f \in L^p(\mathbb{R}^d) : \|f\|_{FL^{s,p}} < \infty\}$ the *Fourier-Lebesgue* spaces of order (s,p) with norm $\|f\|_{FL^{s,p}} = (\int |\mathcal{F}f(u)|^p(1+\|u\|)^{sp}du)^{1/p}$. Denote by $FL^{s,p}_{loc}(\mathbb{R}^d)$ the localized Fourier-Lebesgue spaces which contain all functions f such that $f\varphi \in FL^{s,p}(\mathbb{R}^d)$ for all $\varphi \in C_c^\infty(\mathbb{R}^d)$.

This definition assumes implicitly for $f \in FL^{s,p}(\mathbb{R}^d)$ that the Fourier transform $\mathcal{F}f$ exists as a function in $L^p(\mathbb{R}^d)$. For $p = 1$ we just write $FL^s(\mathbb{R}^d)$ (or $FL^s_{loc}(\mathbb{R}^d)$) and $\|f\|_{FL^s}$. For $p = 2$ the spaces $H^s(\mathbb{R}^d) := FL^{s,2}(\mathbb{R}^d)$ (or $H^s_{loc}(\mathbb{R}^d) := FL^{s,2}_{loc}(\mathbb{R}^d)$) are the fractional L^2-Sobolev spaces of order s with norm $\|\cdot\|_{H^s} := \|\cdot\|_{FL^{s,2}}$. In particular, a function $f \in H^s(\mathbb{R}^d)$ is $\lfloor s \rfloor$-times weakly differentiable. The Fourier-Lebesgue spaces are also related to the Bessel potential spaces for $1 < p < \infty$ (cf. (Triebel, 2010)). Observe the following embeddings which follow from properties of the Fourier transform.

Proposition A.4.2. *Let $s \geq 0$. Then it holds:*

(i) $FL^s_{loc}(\mathbb{R}^d) \subset C^s(\mathbb{R}^d)$,

(ii) $C^s_c(\mathbb{R}^d) \subset H^{s-\varepsilon}_{loc}(\mathbb{R}^d)$, $\varepsilon > 0$,

(iii) $H^s_{loc}(\mathbb{R}^d) \subset FL^{s'}_{loc}(\mathbb{R}^d)$, $s > s' + d/2$.

Note that we can gain in regularity for some functions by considering larger p. For example, the Fourier transforms of the indicator functions $\mathbf{1}_{[a,b]}$, $a < b$, decay as $|u|^{-1}$ for $|u| \to \infty$ and thus $\mathbf{1}_{[a,b]} \in FL^{0-}(\mathbb{R})$, but also $\mathbf{1}_{[a,b]} \in H^{1/2-}(\mathbb{R})$. Similarly, $x \mapsto e^{-|x|}$ lies in $FL^{1-}(\mathbb{R})$ and in $H^{3/2-}(\mathbb{R})$. For another example of negative regularity see Theorem 3.2.14. More details on these spaces can be found in Adams and Fournier (2003), Di et al. (2012), Triebel (2010).

A.5. Semigroup theory and the functional calculus

This section recalls the basic objects needed in Section 3.3. For more details on semigroup theory and the functional calculus see Bakry et al. (2013, Chapters 1.4.1 and A.4), Rudin (2006, Chapter 13) or Engel and Nagel (1999).

Let X be a Markov process with values in a Polish space $\mathcal{S}$ and let μ be any probability measure $\mathcal{S}$. On the induced Hilbert space $(L^2(\mu), \|\cdot\|_\mu)$ denote by $(P_r)_{r\geq 0}$ the Markov semigroup associated with X which satisfies $P_r f(x) = \mathbb{E}[f(X_r)|X_0 = x]$ for $f \in L^2(\mu)$, $x \in \mathcal{S}$, and $P_{r+s} = P_r P_s$, $r, s \geq 0$. The *infinitesimal generator* of the semigroup is defined as

$$Lf = \lim_{r\to 0} \frac{P_r f - f}{r}, \quad f \in \operatorname{dom}(L),$$

where the limit is taken with respect to $\|\cdot\|_\mu$ and where the *domain* $\operatorname{dom}(L) \subset L^2(\mu)$ is the set of all functions for which this limit exists. If $(P_r)_{r\geq 0}$ is strongly continuous, i.e. $P_r f \to f$ in $L^2(\mu)$ as $r \to 0$ for all $f \in L^2(\mu)$, then the semigroup is called *Feller*.

This is true for most Markov processes in practice, including Lévy processes and many diffusions. In the Feller case, L is a densely defined closed linear and usually unbounded operator on its domain with spectrum $\sigma(L) \subset \{\lambda \in \mathbb{C} : \mathrm{Re}(\lambda) \leq 0\}$.

In order to define fractional powers of the generator assume that L and thus the operators P_r are *normal*, i.e. $LL^* = L^*L$, where L^* is the Hilbert space adjoint of L with respect to $L^2(\mu)$. In that case the spectral theorem (Rudin (2006, Theorem 13.33)) guarantees the existence of a resolution of the identity or spectral measure $(E_A)_{A\in\mathcal{B}(\mathbb{C})}$ on $L^2(\mu)$. This means that $(E_A)_{A\in\mathcal{B}(\mathbb{C})}$ is a family of orthogonal projections $E_A : L^2(\mu) \to L^2(\mu)$ for Borel sets $A \subset \mathbb{C}$ such that for every $f, g \in L^2(\mu)$ the map $A \mapsto \langle E_A f, g\rangle_\mu$ is a complex measure supported on $\sigma(L)$. Moreover, $A \mapsto \langle E_A f, f\rangle_\mu$ is a positive measure with total variation $\langle E_{\mathbb{C}} f, f\rangle_\mu = \|f\|_\mu^2$. By the spectral theorem we can associate to any measurable function $\Psi : \mathbb{C} \mapsto \mathbb{C}$ a densely defined closed operator $\Psi(L)$ by the relation

$$\langle \Psi(L) f, g\rangle_\mu = \int_{\sigma(L)} \Psi(\lambda)\, d\langle E_\lambda f, g\rangle_\mu, \quad f, g \in L^2(\mu),$$

with domain $\mathrm{dom}(\Psi(L)) = \{f \in L^2(\mu) : \int_{\sigma(L)} |\Psi(\lambda)|^2 d\langle E_\lambda f, f\rangle_\mu < \infty\}$. It satisfies $\|\Psi(L)f\|_\mu^2 = \int_{\sigma(L)} |\Psi(\lambda)|^2 d\langle E_\lambda f, f\rangle_\mu$. In particular, we can define the fractional operators $|L|^{s/2}$ on $\mathrm{dom}(|L|^{s/2})$ for $0 \leq s \leq 1$. At last, by the spectral theorem for normal semigroups (Rudin (2006, Theorem 13.38)), the semigroup can be realized in its exponential form, i.e. $P_r = \Psi(L)$ with $\Psi(x) = e^{rx}$, $r \geq 0$.

Bibliography

Abadir, K. M. and Magnus, J. R. (2009). Matrix algebra, Vol. 1, Econometric Exercises.

Adams, R. and Fournier, J. (2003). *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier Science.

Ait-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-Frequency Estimates With Noisy and Asynchronous Financial Data. *Journal of the American Statistical Association*, 105(492):1504–1516.

Ait-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*, volume 1 of *High-Frequency Financial Econometrics*. Princeton University Press.

Altmeyer, R. (2017). Estimating occupation time functionals. *arXiv preprint arXiv:1706.03418*.

Altmeyer, R. and Bibinger, M. (2015). Functional stable limit theorems for quasi-efficient spectral covolatility estimators. *Stochastic Processes and their Applications*, 125(12):4556–4600.

Altmeyer, R. and Chorowski, J. (2017). Estimation error for occupation time functionals of stationary Markov processes. *Stochastic Processes and their Applications*, 128(6):1830–1848.

Bakry, D., Gentil, I., and Ledoux, M. (2013). *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der mathematischen Wissenschaften. Springer International Publishing.

Bardina, X. and Rovira, C. (2007). On Itô's formula for elliptic diffusion processes. *Bernoulli*, 13(3):820–830.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

Bass, R. F. (2006). *Diffusions and Elliptic Operators*. Probability and Its Applications. Springer New York.

Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *The Annals of Applied Probability*, 15(4):2422–2444.

Bibinger, M. (2011). Efficient Covariance Estimation for Asynchronous Noisy High-Frequency Data. *Scandinavian Journal of Statistics*, 38(1):23–45.

Bibinger, M., Hautsch, N., Malec, P., and Reiß, M. (2014). Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency. *The Annals of Statistics*, 42(4):1312–1346.

Bibinger, M., Jirak, M., and Reiß, M. (2016). Volatility estimation under one-sided errors with applications to limit order books. *The Annals of Applied Probability*, 26(5):2754–2790.

Bibinger, M., Jirak, M., and Vetter, M. (2017). Nonparametric change-point analysis of volatility. *The Annals of Statistics*, 45(4):1542–1578.

Bibinger, M. and Reiß, M. (2014). Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics*, 41(1):23–50.

Billingsley, P. (2013). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley.

Boufoussi, B., Dozzi, M., and Guerbaz, R. (2007). Sample path properties of the local time of multifractional Brownian motion. *Bernoulli*, 13(3):849–867.

Bouleau, N. and Yor, M. (1981). Sur la variation quadratique des temps locaux de certaines semimartingales. *CR Acad, Sci. Paris, Ser. I Math*, 292:491–494.

Briol, F.-X., Oates, C. J., Girolami, M., Osborne, M. A., and Sejdinovic, D. (2015). Probabilistic Integration: A Role for Statisticians in Numerical Analysis? *arXiv preprint arXiv:1512.00933*.

Catellier, R. and Gubinelli, M. (2016). Averaging along irregular curves and regularisation of ODEs. *Stochastic Processes and their Applications*, 126(8):2323–2366.

Chaumont, L. i. c. and Uribe Bravo, G. (2011). Markovian bridges: Weak continuity and pathwise constructions. *The Annals of Probability*, 39(2):609–647.

Chen, M. F. (2006). *Eigenvalues, Inequalities, and Ergodic Theory*. Probability and Its Applications. Springer London.

Chesney, M., Jeanblanc-Picqué, M., and Yor, M. (1997). Brownian Excursions and Parisian Barrier Options. *Advances in Applied Probability*, 29(29).

Chojnowska-Michalik, A. and Goldys, B. (2002). Symmetric Ornstein-Uhlenbeck semigroups and their generators. *Probability Theory and Related Fields*, 124(4):459–486.

Chorowski, J. (2018). Nonparametric volatility estimation in scalar diffusions: Optimality across observation frequencies. *Bernoulli*, 24(4A):2934–2990.

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.

Clément, E., Delattre, S., and Gloter, A. (2013). An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications*, 123(7):2500–2521.

Curci, G. and Corsi, F. (2012). Discrete Sine Transform for Multi-Scales Realized Volatility Measures. *Quantitative Finance*, 12:263–279.

Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *The Annals of Statistics*, 33(6):2507–2528.

Debussche, A. and Romito, M. (2014). Existence of densities for the 3D Navier–Stokes equations driven by Gaussian noise. *Probability Theory and Related Fields*, 158(3):575–596.

Di, N., Palatucci, G., and Valdinoci, E. (2012). Hitchhikers guide to the fractional Sobolev spaces. *Bulletin des Sciences Mathématiques*, 136:521–573.

Diaconis, P. (1988). Bayesian numerical analysis. *Statistical decision theory and related topics IV*, 1:163–175.

Diehl, J., Gubinelli, M., and Perkowski, N. (2017). The Kardar–Parisi–Zhang Equation as Scaling Limit of Weakly Asymmetric Interacting Brownian Motions. *Communications in Mathematical Physics*, pages 1–41.

Dion, C. and Genon-Catalot, V. (2016). Bidimensional random effect estimation in mixed stochastic differential model. *Statistical Inference for Stochastic Processes*, 19(2):131–158.

Eisenbaum, N. (2006). Local time-space stochastic calculus for Lévy processes. *Stochastic Processes and their Applications*, 116(5):757–778.

Engel, K.-J. and Nagel, R. (1999). *One-Parameter Semigroups for Linear Evolution Equations*. Graduate Texts in Mathematics. Springer New York.

Errami, M., Russo, F., and Vallois, P. (2002). Itô's formula for C1-functions of a càdlàg process and related calculus. *Probability Theory and Related Fields*, 122(2):191–221.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes : characterization and convergence*. Wiley.

Falconer, K. and Liu, L. (2012). Multistable Processes and Localizability. *Stochastic Models*, 28(3):503–526.

Florens-Zmirou, D. (1993). On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability*, 30(4):pp. 790–804.

Föllmer, H. and Protter, P. (2000). On Itô's formula for multidimensional Brownian motion. *Probability Theory and Related Fields*, 116(1):1–20.

Föllmer, H., Protter, P., and Shiryayev, A. N. (1995). Quadratic covariation and an extension of Itô's formula. *Bernoulli*, 1(1-2):149–169.

Fournier, N. and Printems, J. (2008). Absolute continuity for some one-dimensional processes. *Bernoulli*, 16(2):343–360.

Fukasawa, M. (2010). Realized volatility with stochastic sampling. *Stochastic Processeses and their Applications*, 120:209–233.

Ganychenko, I. (2015). Fast L 2 -approximation of integral-type functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 2:165–171.

Ganychenko, I., Knopova, V., and Kulik, A. (2015). Accuracy of discrete approximation for integral functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 2(4):401–420.

Ganychenko, I. and Kulik, A. (2014). Rates of approximation of non-smooth integral type functionals of Markov processes. *Modern Stochastics: Theory and Applications*, 1(2):117–126.

Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

Geman, D. and Horowitz, J. (1980). Occupation Densities. *The Annals of Probability*, 8(1):1–67.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l'IHP Probabilités et statistiques*, pages 119–151. Gauthier-Villars.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors 1 and 2. *ESAIM: Probability and Statistics*, 5:225–242.

Gobet, E. and Labart, C. (2008). Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electronic Communications in Probability*, 13:352–363.

Gobet, E. and Matulewicz, G. (2016). Parameter estimation of Ornstein–Uhlenbeck process generating a stochastic graph. *Statistical Inference for Stochastic Processes*, pages 1–25.

Hansen, L. P., Alexandre Scheinkman, J., and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, 86(1):1–32.

Hayashi, T. and Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Processes and their Applications*, 121(10):2416–2454.

Hugonnier, J.-N. (1999). The Feynman–Kac formula and pricing occupation time derivatives. *International Journal of Theoretical and Applied Finance*, 2(02):153–178.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. *Séminaire de Probabilitiés*, pages 232–246.

Jacod, J. (1998). Rates of convergence to the local time of a diffusion. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 34(4):505–544.

Jacod, J., Jakubowski, A., and Mémin, J. (2003). On asymptotic errors in discretization of processes. *The Annals of Probability*, 31(2):592–608.

Jacod, J., Li, Y., Mykland, P., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, 119(7):2249–2276.

Jacod, J. and Mykland, P. (2015). Microstructure noise in the continuous case: Approximate efficiency of the adaptive pre-averaging method. *Stochastic Processes and their Applications*.

Jacod, J., Podolskij, M., and Vetter, M. (2010). Limit Theorems for moving averages of discretized processes plus noise. *The Annals of Statistics*, 38:1478–1545.

Jacod, J. and Protter, P. (1988). Time Reversal on Levy Processes. *The Annals of Probability*, 16(2):620–641.

Jacod, J. and Protter, P. (1998). Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations. *The Annals of Probability*, 26:267–307.

Jacod, J. and Protter, P. (2011). *Discretization of Processes*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg.

Jacod, J. and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics*, 41(3):1462–1484.

Jacod, J. and Shiryaev, A. (2013). *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.

Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *The Annals of Statistics*, 42(3):1029–1069.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Probability and Its Applications. Springer New York.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. Springer.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19.

Kohatsu-Higa, A., Makhlouf, R., and Ngo, H.-L. (2014). Approximations of non-smooth integral type functionals of one dimensional diffusion processes. *Stochastic Processes and their Applications*, 124(5):1881–1909.

Kufner, A. (1985). *Weighted Sobolev Spaces*. Teubner-Texte zur Mathematik. B.G. Teubner.

Li, J., Todorov, V., and Tauchen, G. (2013). Volatility occupation times. *The Annals of Statistics*, 41(4):1865–1891.

Li, Y., Mykland, P. A., Renault, E., Zhang, L., and Zheng, X. (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory*, 30(03):580–605.

Lou, S. and Ouyang, C. (2017). Local times of stochastic differential equations driven by fractional Brownian motions. *Stochastic Processes and their Applications*.

Lowther, G. (2010). Nondifferentiable functions of one-dimensional semimartingales. *The Annals of Probability*, 38(1):76–101.

Mattingly, J. C., Stuart, A. M., and Tretyakov, M. V. (2010). Convergence of Numerical Time-Averaging and Stationary Measures via Poisson Equations. *SIAM Journal on Numerical Analysis*, 48(2):552–577.

Millet, A., Nualart, D., and Sanz, M. (1989). Integration by Parts and Time Reversal for Diffusion Processes. *The Annals of Probability*, 17(1):208–238.

Moret, S. and Nualart, D. (2001). Generalization of Itô's formula for smooth nondegenerate martingales. *Stochastic Processes and their Applications*, 91(1):115–149.

Ngo, H.-L. and Ogawa, S. (2011). On the discrete approximation of occupation time of diffusion processes. *Electronic Journal of Statistics*, 5:1374–1393.

Nualart, D. (1995). *The Malliavin Calculus and Related Topics*. Probability and its applications : a series of the applied probability trust. Springer-Verlag.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Texts in Applied Mathematics. Springer New York.

Pitt, L. (1978). Local times for Gaussian vector fields. *Indiana University Mathematics Journal*, 27(2):309–330.

Podolskij, M. and Vetter, M. (2010). Understanding limit theorems for semimartingales: a short survey. *Statistica Neerlandica*, 64(3):329–351.

Pollett, P. (2003). Integrals for continuous-time Markov chains. *Mathematical Biosciences*, 182(2):213–225.

Protter, P. (2013). *Stochastic Integration and Differential Equations*. A New Approach. Springer Science & Business Media, Berlin, Heidelberg.

Reiß, M. (2011). Asymptotic Equivalence for Inference on the volatility from noisy observations. *The Annals of Statistics*, 39(2):772–802.

Renault, E., Sarisoy, C., and Werker, B. J. M. (2017). Efficient estimation of integrated volatility and related processes. *Econometric Theory*, 33(2):439–478.

Rényi, A. (1963). On Stable Sequences of Events. *Sankhya: The Indian Journal of Statistics, Series A*, 25(3):293–302.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*. Springer-Verlag Berlin Heidelberg.

Romito, M. (2017). A simple method for the existence of a density for stochastic evolutions with rough coefficients. *arXiv preprint arXiv:1707.05042*.

Rudin, W. (2006). *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill.

Russo, F. and Vallois, P. (1995). The generalized covariation process and Ito formula. *Stochastic Processes and their Applications*, 59(1):81–104.

Russo, F. and Vallois, P. (1996). Ito formula for C1-functions of semimartingales. *Probability Theory and Related Fields*, 104(1):27–41.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Tankov, P. (2003). *Financial Modelling with Jump Processes*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press.

Triebel, H. (2010). *Theory of Function Spaces*. Modern Birkhäuser Classics. Springer Basel.

Walsh, A. (2013). Stochastic integration with respect to additive functionals of zero quadratic variation. *Bernoulli*, 19(5B):2414–2436.

Walsh, J. B. (1982). A non-reversible semi-martingale. *Séminaire de Probabilités de Strasbourg*, 16:212–212.

Watanabe, S. (1984). *Lectures on Stochastic Differential Equations and Malliavin Calculus*. Lectures on mathematics and physics / Tata institute of fundamental research: Mathematics. Springer.

Xiao, Y. (2006). Properties of local-nondeterminism of Gaussian and stable random fields and their applications. *Annales de la faculté des sciences de Toulouse Mathématiques*, 15(1):157–193.

Xiu, D. (2010). Quasi-Maximum Likelihood Estimation of Volatility With High Frequency Data. *Journal of Econometrics*, 159:235–250.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019.

Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2005). A Tale of Two Time Scales. *Journal of the American Statistical Association*, 100(472):1394–1411.

Zhou, B. (1998). F-consistency, De-volatization and Normalization of High Frequency Financial Data. In Dunis, C. and Zhou, B., editors, *Nonlinear Modelling of High Frequency Financial Time Series*, pages 109–123. John Wiley Sons Ltd.

List of Tables

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, 18.07.2017