A Coupling for Triple Stochastic Integrals

Xìling Zh¯ang *

November 29, 2018

Abstract

This article presents a coupling approach for the approximation of iterated stochastic integrals of length three. The generation of such integrals is the central problem of higher-order pathwise approximations for SDEs, which still lacks a satisfactory answer due to the restriction of dimensionality and computational load. Here we start from the Fourier representation of the triple stochastic integral and investigate the global behaviour of the joint density of the representation. Finally in the main result we give a coupling in the quadratic Vaserstein distance.

1 Introduction

Let $d, q \in \mathbb{Z}^+$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space equipped with a right-continuous filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$. Consider an $\mathbb{R}^d$-valued autonomous stochastic differential equation driven by a $q$-dimensional $\mathcal{F}$-Wiener martingale $W$:

$$x_t = x_0 + \int_0^t b(x_s) \, ds + \int_0^t \sigma(x_s) \, dW_s. \tag{1.1}$$

Assume that the coefficients $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times q}$ are sufficiently smooth. It is well-known that one can derive numerical schemes that converge in the strong $L^p$ sense of order greater than $1/2$ from stochastic Taylor expansions, as is shown in [7]. For example, by applying Itô’s formula to $b$ and $\sigma$, one obtains the Itô-Taylor expansion of length 2: for each

---

*School of Mathematics, The University of Edinburgh. e-mail: xiling.zhang@ed.ac.uk

1Throughout the paper $\mathbb{Z}^+$ denotes the set of positive integers and $\mathbb{N}$ denotes the set of natural numbers $\mathbb{Z}^+ \cup \{0\}$. 
component $i = 1, \cdots, d$ on the interval $[s, t]$,

$$x^i_t = x^i_s + b^i(x_s)(t - s) + \sum_{j=1}^{q} \sigma_{ij}(x_s)(W^j_t - W^j_s)$$

$$+ \int_{s}^{t} \int_{s}^{r} \mathcal{L}b^i(x_u)du + \sum_{j=1}^{q} \int_{s}^{t} \int_{s}^{r} \sum_{k=1}^{d} \sigma_{kj}(x_u)\partial_k b^i(x_u)dW^j_u dr$$

$$+ \sum_{j=1}^{q} \int_{s}^{t} \int_{s}^{r} \mathcal{L}\sigma_{ij}(x_u)dudu + \sum_{j,k=1}^{q} \int_{s}^{t} \int_{s}^{r} \sum_{l=1}^{d} \sigma_{lk}(x_u)\partial_l \sigma_{ij}(x_u)dW^k_u dW^j_r, \quad (1.2)$$

where $\partial_k$ is the partial derivative w.r.t. the $k$-th coordinate. The last term in (1.2) involves an iterated stochastic integral, and it gives rise to Milstein’s method: for each component $i = 1, \cdots, d$,

$$X^i_{k+1} = X^i_k + b^i(X_k)h + \left( \sum_{j=1}^{q} \sigma_{ij}(X_k)\Delta W^j_{k+1} + \sum_{j,l=1}^{q} \varsigma_{ijl}(X_k)A_k(j,l) \right), \quad (1.3)$$

where $h \in (0, 1)$ is the step size, $\Delta W^j_{k+1} = W^j_{tk+1} - W^j_{tk}$, $\varsigma_{ijl}(x) := \sum_{m=1}^{d} \sigma_{mj}(x)\partial_m \sigma_{il}(x)$ and

$$A_k(j,l) := \int_{tk}^{tk+1} (W^j_{\tau} - W^j_{tk})dW^l_{\tau}.$$ 

The scheme (1.3) has strong-$L^2$ convergence rate $O(h)$ according to Kloeden and Platen [7] (Section 10.3), but the problem lies in the generation of the double integral $I_{jl} = \int_{0}^{h} W^j_t dW^l_t$, which is non-trivial for $q \geq 2$.

As mentioned by Wiktorsson [13] and Davie [2] (Section 2), if the diffusion matrix satisfies the commutativity condition $\varsigma_{ijl}(x) = \varsigma_{ijl}(x)$ for all $x \in \mathbb{R}^d$ and all $i = 1, \cdots, d$, $j,l = 1, \cdots, q$, one only needs to generate the Wiener increments $\Delta W^j_{k+1}$ to achieve the order-1 convergence. But this is not always the case: using only the Wiener increments $\Delta W^j_{k+1}$ to implement a numerical method will, in general, result in a convergence rate no more than $O(h^{1/2})$, according to [1].

One attempt to generate the double integral $I_{jl}$ was made by Lyons and Gaines [8], but their method only works for $q = 2$. Recently a strong result for any dimension has been proved by Davie [2] (Theorem 4) under the condition that the diffusion matrix $\sigma$ has rank $d$ everywhere, and it provides a way to approximate the SDE up to an arbitrary order. This is a significant improvement concerning higher-order approximations. The idea is that, rather than generating the double integrals at each step $k$, one approximates the quantity inside the big parentheses in (1.3) as a whole. This is a completely different approach than the usual ones, as Davie’s arguments are based on the coupling method, quantifying the strong-$L^p$ convergence in terms of the Vaserstein metrics.

\[2\text{Also spelt as “Wasserstein”}\]
The coupling method. For probability measures \( P, Q \) on \( \mathbb{R}^d \) and \( p \geq 1 \), the Vaserstein \( p \)-distance is defined by

\[
W_p(P, Q) := \inf_{\pi \in \Pi(P,Q)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p},
\]

where \( \Pi(P, Q) \) is the set of all joint probability measures on \( \mathbb{R}^d \times \mathbb{R}^d \) with marginal laws \( P \) and \( Q \). In general \( P \) and \( Q \) need not be defined on the same probability space, but this definition is enough for the purpose of this article. The notation \( W_p(X, Y) \) will not cause any confusion for random variables \( X \) and \( Y \) having laws \( P \) and \( Q \), respectively. If one can show a bound for the distance between the two laws, we then say there is a coupling between \( X \) and \( Y \) (or \( P \) and \( Q \)).

The significance of using the Vaserstein distances instead of other ones is that, when generating numerical schemes for an SDE, the convergence in the Vaserstein-type distance \( W_{p,\infty}(\text{replacing } |x - y|^p \text{ in the definition above by } \max_k |x_k - y_k|^p) \) is equivalent to the usual strong \( L^p \)-convergence, for the purpose of simulation at least. To see this, suppose we have found a coupling between the grid points of the solution \( x = \{x_{tk}\}_k \) and a numerical scheme \( X = \{X_k\}_k \) with \( W_{p,\infty}(x, X) \leq Ch^\gamma \) for some \( \gamma > 0 \). Then by definition, \( \forall \varepsilon > 0 \) there is a random vector \( Y^\varepsilon \) on the same probability space as the solution \( x \), having the same distribution as \( X \), s.t. \( (\mathbb{E} \max_k |x_{tk} - Y_{tk}|^p)^{1/p} \leq W_{p,\infty}(x, X) + \varepsilon \). Choose \( \varepsilon = h^\gamma \) and in practice one generates \( Y \) instead of \( x \) to approximate \( x \). The reader is referred to Section 12 in [2] for a detailed discussion on the contexts where such a substitution holds or fails.

Although there is no general formulas for the quantity \( W_p(P, Q) \), if \( P \) and \( Q \) have densities \( f \) and \( g \), respectively, then there is the elementary and yet important inequality

\[
W_p(P, Q) \leq C_p \left( \int_{\mathbb{R}^d} |x|^p |f(x) - g(x)|dx \right)^{1/p}, \tag{1.4}
\]

for all \( p \geq 1 \), as a variant of Proposition 7.10 in [12]. This inequality serves as a main tool to give an \( W_2 \)-estimate in [2] and [3], and will be used for all the main result in this article.

The more difficult situation is that \( \sigma \) has rank less than \( d \), which could well happen. In Section 9 in [2] a different approach based on the Fourier expansion introduced in Section 5.8 in [7] is proposed, giving a coupling for the double integral \( I_{ij} \). The motivation of this article is to provide a feasible approximation for SDEs of a higher order. For the equation \( (1.1) \) on the interval \([0, T]\), by applying Itô’s formula again to the term \( \sigma_{kl}(X_u)\partial_k\sigma_{ij}(X_u) \) in \( (1.2) \), one obtains, for each component \( i = 1, \ldots, d \) on the interval \([s, t]\),

\[
X_t^i = X_s^i + b_i(X_s)(t - s) + \sigma_{ij}(X_s)(W_t^j - W_s^j) + \sigma_{kl}(X_s)\partial_k\sigma_{ij}(X_s) \int_s^t \int_s^r dW_u^l dW_r^j
\]

\[
+ \int_s^t \int_s^r \mathbb{L}b_i(X_u)du dr + \int_s^t \int_s^r \sigma_{kl}(X_u)\partial_kb_i(X_u)du dW_r^j
\]

\[
+ \int_s^t \int_s^r \mathbb{L}\sigma_{ij}(X_u)du dW_r^j + \int_s^t \int_s^r \mathbb{L}(\sigma_{kl}(X_v)\partial_k\sigma_{ij}(X_v)) dv dW_r^j
\]

\[
+ \int_s^t \int_s^r \mathbb{L}(\sigma_{kl}(X_v)\partial_k\sigma_{ij}(X_v)) dv dW_r^j
\]

\[
+ \int_s^t \int_s^r \partial_m(\sigma_{kl}(X_v)\partial_k\sigma_{ij}(X_v)) \sigma_{mn}(X_v) dW_u^l dW_r^j dW_r^j,
\]

3
where the summation signs over repeated indices are omitted. From this expression one can obtain a suitable numerical scheme (formula (10.4.6) in [2]) with strong convergence order $O(h^{3/2})$. Just as the Milstein scheme, the crucial ingredient to achieve such a higher-order convergence is the generation of the triple integrals

$$I_{jkl}(s,t) := \int_s^t \int_s^r \int_s^u dW'_j dW'_k dW'_l,$$

for indices $(j,k,l) \in \{1, \cdots, q\}^3$.

Similar to the way the double stochastic integral is treated in [2], one would expect the same method to be extended to treat triple integrals. For the simplicity of formulation, the Stratonovich triple integral $I_{jkl}^S(s,t) := \int_s^t \int_s^r \int_s^u dW_u^j \circ dW_u^k \circ dW_u^l$ will be considered instead of the Itô version, since the Fourier representation of the former has a relatively simpler form. This is due to the fact that the product of two Stratonovich integrals is a shuffle product - see Proposition 2.2 in [5]. In other words, an iterated Stratonovich integral of longer length can be represented by shorter ones in a much simpler way compared its Itô counterpart.

The double integral case. The goal of this paper is to find a random variable $\tilde{I}_{jkl}$ whose law is close to that of $I_{jkl}^S$ in the Vaserstein distance, which in turn gives a feasible $O(h^{3/2})$-approximation for the SDE (1.1). In order to have a better understanding of the method let us briefly review Davie’s Fourier method (Section 9 in [2]). Consider the interval $[0,1]$ for simplicity. According to [7] (Section 5.8), the Brownian bridge process $W_t - t W_1$ has Fourier expansion

$$W^j_t - t W^j_1 = \frac{1}{2\sqrt{2\pi}} x_{j0} + \frac{1}{\sqrt{2\pi}} \sum_{r=1}^{\infty} x_{jr} \cos(2\pi r t) + \frac{1}{\sqrt{2\pi}} \sum_{r=1}^{\infty} y_{jr} \sin(2\pi r t), \quad (1.5)$$

where $x_{jr}$, $y_{jr}$ are $N(0,1)$-random variables mutually independent for different values of $j = 1, \cdots, q$ or $r \in \mathbb{N}$, all independent of $W_1$. Then the double integral $I_{jk}^S = \int_0^1 W^j_u \circ dW^k_u$ has Fourier representation

$$I_{jk}^S = \frac{1}{2} W^j_1 W^k_1 + \frac{1}{2\pi} \left( W^j_1 z_k - W^k_1 z_j \right) + \frac{1}{2\pi} \lambda_{jk}, \quad (1.6)$$

where $\lambda_{jk} = \sum_{r \geq 1} r^{-1} (x_{jr} y_{kr} - y_{jr} x_{kr})$ and $z_j = \sum_{r \geq 1} r^{-1} x_{jr}$. One then needs to approximate each $\lambda_{jk}$ and $z_j$ by their partial sums $\lambda_{jk} = \sum_{r=1}^{p} r^{-1} (x_{jr} y_{kr} - y_{jr} x_{kr})$ and $z_j = \sum_{r=1}^{p} r^{-1} x_{jr}$. Denote $\tilde{\lambda}^{(p)}_{jk} = \lambda_{jk} - \lambda_{jk}^{(p)}$, $\tilde{z}^{(p)}_j = z_j - z_j^{(p)}$ and $U := (\lambda, z)$, $U_p := (\lambda^{(p)}, z^{(p)})$, $\tilde{U}_p := (\tilde{\lambda}^{(p)}, \tilde{z}^{(p)})$.

Davie’s result states that if there is a random variable $\tilde{U}_p$, independent of $U_p$, having the same moments as $U_p$ up to order $m-1$ and satisfying $\mathbb{E} \exp(a\sqrt{|p|} |\tilde{U}_p|) \leq b$ for some positive constants $a, b$ for all $p$, then $W_2(U, U_p + \tilde{U}_p) = O(p^{-m/2})$ for $p$ sufficiently large. The idea is to estimate the densities $g(\zeta)$ of $U$ and $h(\zeta)$ of $U_p + \tilde{U}_p$. If $f_p$ is the density of $U_p$, then
\(g(\zeta) = \mathbb{E}f_p(\zeta - \bar{U}_p)\) and \(h(\zeta) = \mathbb{E}f_p(\zeta - \bar{U}_p)\). By Taylor’s theorem, for all \(\zeta, w \in \mathbb{R}^d\),

\[
f_p(\zeta - w) = \sum_{|\beta| = 0}^{m-1} \frac{(-1)^{|\beta|}}{\beta!} w^\beta \partial^\beta f_p(\zeta)
+ \sum_{|\beta| = m} \frac{|\beta|(-1)^{|\beta|}}{\beta!} \int_0^1 (1 - \theta)^{|\beta| - 1} w^\beta \partial^\beta f_p(\zeta - \theta w) d\theta. \tag{1.7}
\]

Since up to the \((m - 1)\)-th moments of \(\bar{U}_p\) and \(\bar{U}\) match, when taking the difference \(g(\zeta) - h(\zeta)\) the first summation vanishes, and hence \(\forall \zeta \in \mathbb{R}^d,\)

\[
g(\zeta) - h(\zeta) = \sum_{|\beta| = m} \int_0^1 C_{\beta, \theta} \mathbb{E} \left( \bar{U}_p^\beta \partial^\beta f_p(\zeta - \theta \bar{U}_p) - \bar{U}^\beta \partial^\beta f_p(\zeta - \theta \bar{U}_p) \right) d\theta, \tag{1.8}
\]

where \(C_{\beta, \theta} = |\beta|(-1)^{|\beta|}(1 - \theta)^{|\beta| - 1}/\beta!\). If one can give a uniform bound for some higher derivatives of \(f_p\) in terms of \(p\), then using an interpolation argument one can show a reasonable decay for the \(m\)-th derivative of \(f_p\), and finally one finds a coupling between \(U\) and \(U_p + \bar{U}_p\) by the inequality \((1.4)\).

The main advantage of the double integral \(I_{jk}^p\) compared to the triple one is the fact that its Fourier representation only involves \(\lambda\) and \(z\), whose summands are independent. This ensures that \(U\) has a smooth density (as the convolution of the density \(f_p\) of \(U_p\) and the law of \(\bar{U}_p\)), which significantly simplifies the analysis. More importantly, the characteristic function of \(U_p\) can be explicitly calculated - see formula (32) in the proof of Lemma 11 in [2]. This provides some convenience for investigating the global and local behaviour of the density \(f_p\) (Lemma 12, 13 and 14). In particular, Lemma 14 therein gives a lower bound for \(f_p\), which is essential for achieving a coupling for \(U\) of the optimal order \(O(p^{-m/2})\) in the \(W_2\) distance.

Without Lemma 14, one can still achieve a suboptimal \(W_2\)-rate \(O(p^{-m/4})\) by directly showing a decay of the difference \(|g(\zeta) - h(\zeta)|\) - this is the goal of the present paper, but the treatment of the densities is quite different from the double integral case.

The latter is much more straightforward to see. For \(p\) sufficiently large, the vector \(D^{2m}f_p\) of partial derivatives of order \(2m\) is uniformly bounded everywhere due to part (1) of Lemma 11 in [2]. Also by Lemma 12 therein, one has \(f_p(\zeta) \leq e^{-c_4|\zeta|}\) for \(|\zeta|\) sufficiently large. Then one can apply Lemma 9 therein to get a rapid decay for \(D^m f_p(\zeta)\). To see this, consider \(|\zeta| > p\) sufficiently large and the ball \(B(\zeta, 1)\) that is disjoint with \(B(0, p)\). Then \(\sup_{y \in B(\zeta, 1)} f_p(y) \leq e^{-c_6(|\zeta| - 1)}\), and by applying Lemma 9 to the ball \(B(\zeta, 1)\) one sees the following bound for (the Euclidean norm of) the \(m\)-th derivatives:

\[
|D^m f_p(\zeta)| \leq C_{q,m} \max \left\{ \sup_{y \in B(\zeta, 1)} \sqrt{f_p(y)} \sup_{y \in B(\zeta, 1)} \sqrt{|D^{2m} f_p(y)|}, \sup_{y \in B(\zeta, 1)} f_p(y) \right\}. \tag{1.9}
\]

This yields \(|D^m f_p(\zeta)| \leq C_{q,m} e^{-c_6|\zeta|}\). Therefore from (1.8) and part (2) of Lemma 11 in [2]
one has, by the Cauchy-Schwartz inequality, that for all \( \zeta \in \mathbb{R}^{q(q+1)/2} \),

\[
|g(\zeta) - h(\zeta)| \leq C_{d,m} \sum_{|\beta| = m} \left( |\mathbb{E}[\tilde{U}^\beta_p \partial^\beta f_p(\zeta - \tilde{U}_p)]| + |\mathbb{E}[\tilde{U}^\beta_p \partial^\beta f_p(\zeta - \tilde{U}_p)]| \right) \\
\leq C_{d,m} p^{-m/2} \left( \sqrt{\mathbb{E}|D^m f_p(\zeta - \tilde{U}_p)|^2} + \sqrt{\mathbb{E}|D^m f_p(\zeta - \tilde{U}_p)|^2} \right).
\]

Notice that, on the set \( \{ \omega : |\tilde{U}_p| \leq 1 \} \) one has \( \|D^m f_p(\zeta - \tilde{U}_p)\| \leq C_0 e^{-c_0|\zeta|} \) by the rapid decay of \( D^m f_p \); on the complement \( \{ \omega : |\tilde{U}_p| > 1 \} \), part (2) of Lemma 11 and Chebyshev’s inequality imply that \( \mathbb{P}(|\tilde{U}_p| > 1) \leq C_M p^{-M} \) for any \( M > 0 \). The same argument works for the second term above involving \( \bar{U} \), and so by the inequality (1.4) for the quadratic distance,

\[
W_2(U, \tilde{U}_p + \bar{U}) \leq C \left( \int_{\mathbb{R}^{q(q+1)/2}} |\zeta|^2 |g(\zeta) - h(\zeta)| d\zeta \right)^{1/2} \leq C_{q,m} p^{-m/4}.
\]

From this calculation one sees that the key step towards a good coupling result depends on how well the behaviour of \( f_p \) is understood. Davie’s result is a significant improvement to the existing rate of approximation - see the discussion following the proof of Theorem 15 therein. This is due to some careful estimates (Lemma 12, 13 and 14 in [2]) for the density \( f_p \). For the triple integral \( I_{jkl}^0 \), however, showing similar estimates becomes much more complicated as the Fourier coefficients for \( I_{jkl}^0 \) have summands that are not independent of each other - see the definition of the random variable \( \Delta_{jkl} \) below.

**Notation.** Throughout this paper we will denote by \( \phi \) the standard normal density of dimension 1, by \( B(x,r) \) the open ball of radius \( r \) centred at \( x \), and by \( \Lambda^d \) the Lebesgue measure on \( \mathbb{R}^d \). The notation \( C_0^\infty \) stands for the set of functions that are infinitely times continuously differentiable with compact support. Unless specified otherwise, the single bars \( | \cdot | \) stand for the Euclidean norm, modulus of a complex number, or the cardinality of a set, and the double bars \( \| \cdot \| \) stand for the operator norm, which in the context of matrices is equivalent to any other matrix norm. The letter \( C \) will be used for a generic constant that may change value from line to line, with subscripts specifying its dependence on the parameters. The symbol \( \lesssim_\alpha \) means that the inequality \( \leq (\geq) \) holds up to a multiplicative constant \( C_\alpha \), and \( \simeq_\alpha \) is used when both inequalities hold.

## 2 The Fourier Representation

For the simplicity of presentation let us consider the triple integral on the unit interval \([0,1]\). Following Section 5.8 in [7], from the Fourier expansion (1.5) the triple Stratonovich integral

\[
I^0_{jkl} = \int_0^1 \int_0^t dW^j_s \circ dW^k_s \circ dW^l_t,
\]
for each \((j, k, l) \in \{1, \cdots, q\}^3\) has the following representation:

\[
\begin{align*}
I_{jkl}^p &= \frac{1}{6} W_1^j W_1^k W_1^l - \frac{1}{2\sqrt{2\pi}} W_1^j W_1^k \left( z_l - \frac{1}{\pi} u_l \right) - \frac{1}{2\sqrt{2\pi}} W_1^k W_1^l \left( z_j - \frac{1}{\pi} u_j \right) \\
&\quad - \frac{1}{\sqrt{2\pi}^2} W_1^j W_1^k u_k - \frac{1}{2\pi^2} z_j \left( W_1^k z_l - W_1^l z_k \right) + \frac{1}{2\pi} \left( \frac{1}{2} \lambda_{jk} + \frac{1}{\pi} \nu_{kj} \right) \\
&\quad + \frac{1}{2\pi} W_1^j \left( \frac{1}{2} \lambda_{kl} - \frac{1}{\pi} \nu_{kl} \right) + \frac{1}{4\pi^2} \left( W_1^j \mu_{kl} - W_1^k \mu_{jl} \right) - \frac{1}{2\sqrt{2\pi}^2} z_j \lambda_{kl} \\
&\quad + \frac{1}{4\sqrt{2\pi}} \Delta_{jkl},
\end{align*}
\]

where the coefficients \(z, u, \lambda, \mu, \nu\) are defined as

\[
\begin{align*}
z_j &= \sum_{r=1}^{\infty} \frac{1}{r} x_{jr}, \quad u_j = \sum_{r=1}^{\infty} \frac{1}{r^2} y_{jr}, \\
\lambda_{jk} &= \sum_{r=1}^{\infty} \frac{1}{r} \left( x_{jr} y_{kr} - y_{jr} x_{kr} \right), \quad \mu_{jk} = \sum_{r=1}^{\infty} \frac{1}{r^2} \left( x_{jr} x_{kr} + y_{jr} y_{kr} \right), \\
\nu_{jk} &= \sum_{r,s=1 \atop r \neq s} \frac{1}{r^2 - s^2} \left( \frac{r}{s} x_{jr} x_{ks} - y_{jr} y_{ks} \right),
\end{align*}
\]

with \(x_{jr}, y_{jr}\), again, being \(\mathcal{N}(0,1)\)-random variables independent for different indices \(j = 1, \cdots, q\), \(r \in \mathbb{Z}^+\) and all independent of \(W_1^j\), and the last coefficient \(\Delta\) is given by

\[
\Delta_{jkl} = \sum_{r,s=1}^{\infty} \left\{ -\frac{1}{r(r+s)} \left[ (x_{jr} y_{ks} + y_{jr} x_{ks}) x_{l,r+s} + (-x_{jr} x_{ks} + y_{jr} y_{ks}) y_{l,r+s} \right] \\
+ \frac{1}{rs} \left[ (x_{jr} y_{ls} + y_{jr} x_{ls}) x_{k,r+s} + (-x_{jr} x_{ls} + y_{jr} y_{ls}) y_{k,r+s} \right] \\
+ \frac{1}{s(r+s)} \left[ (-x_{kr} y_{ls} + y_{kr} x_{ls}) x_{j,r+s} + (x_{kr} x_{ls} + y_{kr} y_{ls}) y_{j,r+s} \right] \right\}
\]

For a positive integer \(p\), write \(z^{(p)}\) as the \(p\)-th partial sum of \(z\) and \(z^{(p)} = z - z^{(p)}\). Similar notations are applied to \(u, \lambda\) and \(\mu\). Let \(\nu^{(p)}\) be the partial sum of \(\nu\) over \(r, s \leq p\), \(r \neq s\) and \(\nu^{(p)} = \nu - \nu^{(p)}\), whilst \(\Delta^{(p)}\) denotes the partial sum of \(\Delta\) up to \(r+s \leq p\) and \(\Delta^{(p)} = \Delta - \Delta^{(p)}\).

From the definition of \(\nu^{(p)}_{jk}\) one observes that, by splitting each variable \(\mu^{(1,p)}_{jk}\) into \(\mu^{(2,p)}_{jk} := \sum_{r=1}^{p} r^{-2} x_{jr} x_{kr}\) and \(\mu^{(2,p)}_{jk} := \sum_{r=1}^{p} r^{-2} y_{jr} y_{kr}\), one need only generate \(\nu^{(p)}_{jk}\) for \(j < k\), since

\[
\nu^{(p)}_{jk} + \nu^{(p)}_{kj} = z^{(p)}_{j} z^{(p)}_{k} - \mu^{(1,p)}_{jk}.
\]

Equivalent notations for the infinite sums are used by omitting the superscript \((p)\) and the identity still holds. Therefore one need only consider \(\nu_{jk}\) for \(j < k\).

Another observation is that one need not consider all possible choices of the 3-tuple \((j, k, l) \in \{1, \cdots, q\}^3\) for \(\Delta\); it suffices to focus on those terms with \((j, k, l)\) being a Lyndon word - a
word that is strictly less than all of its proper right factors in the lexicographic order. This is due to the fact that all triple Stratonovich integrals $I_{jkl}^p$ can be expressed by the Lyndon words of length at most 3 - see Corollary 3.3 in [5].

For a word $w$ in a totally ordered set $A$, if it is the concatenation of two non-empty words $u, v \in A$, i.e. $w = uv$, then $v$ is called a proper right factor of $w$. For example, $(1, 1, 2)$ and $(1, 3, 2)$ are both Lyndon words but $(1, 2, 1)$ is not. By definition, a triple $(j, k, l)$ is a Lyndon word if and only if $j < k$ and $l$ or $j = k < l$. Denote by $\mathcal{L}_{3,q} \subset \{1, \cdots , q\}^3$ the set of Lyndon words of length 3, then according to [5], $|\mathcal{L}_{3,q}| = (q^3 - q)/3$.

As an analogue of the work by Davie [2] (Section 9), one seeks to approximate the variable $V = (z, u, \lambda, \mu, \nu, \Delta)$ by studying the distribution of the partial sums

$$V_p = (z^{(p)}, u^{(p)}, \lambda^{(p)}, \mu^{(p)}, \nu^{(p)}, \Delta^{(p)}),$$

and that of the remainder $\tilde{V}_p := (z^{(p)}, u^{(p)}, \lambda^{(p)}, \mu^{(p)}, \nu^{(p)}, \Delta^{(p)})$. Note that for an $O(h^{3/2})$-approximation of the SDE (1.1), one also needs to simulate the double integrals (1.6) along with the triple ones. But they are determined by the variables $(z, \lambda)$, which are already included in $V$.

To develop an analogue of Davie’s results in [2], it is necessary to give some suitable moment estimates for the remainder $\tilde{V}_p$. For simplicity denote the dimension of $V$ by

$$d = 2q^2 + 2q + (q^3 - q)/3,$$

and denote by $v_p$ the $\mathbb{R}^{2qp}$-vector consisting of $x_{jr}, y_{ks}$ for $j, k = 1, \cdots , q$ and $r, s = 1, \cdots , p$.

For a unit vector $\omega = (\alpha, \beta^{(1)}, \beta^{(2)}, \gamma, a, b, \rho)$ and $v = (x_{jr}, y_{jr})_{j,r} \in \mathbb{R}^{2qp}$, define the cubic phase function $\Phi_p : \mathbb{R}^{2qp} \times \mathbb{S}^{d-1} \rightarrow \mathbb{R}$ by

$$\Phi_p(v; \omega) = \sum_{j<k} \left( \alpha_{jk}\lambda_{jk}^{(p)} + \gamma_{jk}v_{jk}^{(p)} \right) + \sum_{j<k} \left( \beta_{jk}^{(1)}\mu_{jk}^{(1,p)} + \beta_{jk}^{(2)}\mu_{jk}^{(2,p)} \right)$$

$$+ \sum_{j=1}^q \left( a_{jz_j^{(p)}} + b_{jy_j^{(p)}} \right) + \sum_{(j,k,l) \in \mathcal{L}_{3,q}} \rho_{jkl}\Delta_{jkl}^{(p)}.$$  (2.1)

Then by definition the characteristic function $\psi_p(\xi)$ of $V_p$ is given by

$$\psi_p(\xi) = \int_{\mathbb{R}^{2qp}} \exp\{i\xi|\Phi_p(x, y; \xi/|\xi|)\} \prod_{j=1}^q \prod_{r=1}^p \phi(x_{jr})\phi(y_{jr}) dx dy$$

$$= \int_{\mathbb{R}^{2qp}} \exp\{i\xi|\Phi_p(v; \xi/|\xi|)\} \phi_p(v) dv,$$

where $\phi$ is the density function of $\mathcal{N}(0,1)$. Observe that the matrices $\lambda$ and $\mu$ are skew-symmetric and symmetric, respectively, so it would be convenient to extend the values of the coefficients $\alpha$, $\beta := (\beta^{(1)}, \beta^{(2)})$ to their lower-triangles by setting $\alpha_{jk} = -\alpha_{kj}$, $\beta_{jk}^{(i)} = \beta_{kj}^{(i)}$ for all $i = 1, 2$, $j, k = 1, \cdots , q$. Set $\gamma_{jk} = 0$ for all $j \geq k$ and $\rho_{jkl} = 0$ if $(j, k, l)$ is not a Lyndon word.
Throughout this article we will be frequently dealing with oscillatory integrals of the form $\psi_p(\xi)$, and we will conveniently call the function $\phi_p$ the \textbf{amplitude}. In order to give a good estimate for magnitude of $\psi_p(\xi)$ one resorts to the method of stationary phase, and for that one needs to study the derivatives of the phase function $\Phi_p$.

\textit{N.B.} Throughout this article all derivatives of the phase function $\Phi_p(v; \omega)$ are with respect to the first variable $v$.

To find the gradient $\nabla \Phi_p(v; \omega)$, one can make use the extended definitions of $\alpha, \beta, \gamma$ and write down the partial derivatives. For each $j = 1, \ldots, q$ and $r = 1, \ldots, p$, differentiating w.r.t. $x_{jr}$ and $y_{jr}$ gives

\[
\partial_{x_{jr}} \Phi_p(v; \omega) = \frac{1}{r} \alpha_{jk} x_{kr} + \frac{1}{r^2} (1 + \delta_{kj}) \beta_{jk}^{(1)} x_{kr} + \sum_{s=1}^{p} \frac{1}{r^2 - s^2} \left( \frac{\gamma_{jk} - s}{s} \right) x_{ks} + \frac{1}{r} \delta_j
\]

\[
+ \sum_{s=1}^{r-1} \left[ \left( \frac{-\rho_{jkl} + \rho_{lkj}}{r(r + s)} - \frac{\rho_{klj}}{s(r + s)} \right) y_{ks} x_{l,r+s} + \left( \frac{\rho_{jkl} + \rho_{lkj}}{r(r + s)} + \frac{\rho_{klj}}{s(r + s)} \right) x_{ks} y_{l,r+s} \right]
\]

\[
+ \sum_{s=1}^{r-1} \left[ \left( \frac{-\rho_{jkl}}{r(s)} + \frac{\rho_{klj}}{(r-s)s} \right) x_{k,r-s} y_{ls} + \left( \frac{\rho_{jkl}}{r(s)} + \frac{\rho_{klj}}{(r-s)s} \right) y_{k,r-s} x_{ls} \right]
\]

\[
- \frac{\rho_{lkj}}{(r-s)r} (x_{l,r-s} y_{ks} + y_{l,r-s} x_{ks}), \quad (2.2)
\]

\[
\partial_{y_{jr}} \Phi_p(v; \omega) = -\frac{1}{r} \alpha_{jk} x_{kr} + \frac{1}{r^2} (1 + \delta_{kj}) \beta_{jk}^{(2)} x_{kr} - \sum_{s=1}^{p} \frac{1}{r^2 - s^2} (\gamma_{jk} - \gamma_{kj}) y_{ks} + \frac{1}{r^2} \delta_j
\]

\[
+ \sum_{s=1}^{r-1} \left[ \left( \frac{-\rho_{jkl}}{r(r + s)} - \frac{\rho_{klj}}{s(r + s)} \right) x_{ks} x_{l,r+s} + \left( \frac{-\rho_{jkl} + \rho_{lkj}}{r(r + s)} - \frac{\rho_{klj}}{s(r + s)} \right) y_{ks} y_{l,r+s} \right]
\]

\[
+ \sum_{s=1}^{r-1} \left[ \left( \frac{\rho_{jkl}}{(r-s)r} - \frac{\rho_{klj}}{(r-s)s} \right) x_{k,r-s} x_{ls} + \left( \frac{\rho_{jkl}}{r(s)} + \frac{\rho_{klj}}{(r-s)s} \right) y_{k,r-s} y_{ls} \right]
\]

\[
+ \frac{\rho_{lkj}}{(r-s)r} (x_{l,r-s} y_{ks} - y_{l,r-s} y_{ks}), \quad (2.3)
\]

where $\delta_{jk}$ is the Kronecker delta, the summation signs over the repeated indices $k, l = 1, \ldots, q$ are omitted, and all $x$ and $y$-terms with second subscripts outwith the interval $[1, p]$ are
assumed to vanish. The Hessian matrix of $\Phi_p$ takes the form

$$D^2\Phi_p(v;\omega) = \begin{pmatrix}
H_{xx}(1,1) & \cdots & H_{xx}(1,q) & H_{xy}(1,1) & \cdots & H_{xy}(1,q) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
H_{xx}(q,1) & \cdots & H_{xx}(q,q) & H_{xy}(q,1) & \cdots & H_{xy}(q,q) \\
H_{yx}(1,1) & \cdots & H_{yx}(1,q) & H_{yy}(1,1) & \cdots & H_{yy}(1,q) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
H_{yx}(q,1) & \cdots & H_{yx}(q,q) & H_{yy}(q,1) & \cdots & H_{yy}(q,q)
\end{pmatrix}, \quad (2.4)$$

where for each pair $(j,k) \in \{1, \cdots, q\}^2$ the blocks $H_{xx}(j,k)$, $H_{xy}(j,k)$, $H_{yy}(j,k)$ are $p \times p$ matrices, e.g.,

$$H_{xx}(j,k) = \begin{pmatrix}
\frac{\partial^2 \Phi_p}{\partial x_{j1}^2} & \frac{\partial^2 \Phi_p}{\partial x_{j2}^2} & \cdots & \frac{\partial^2 \Phi_p}{\partial x_{jp}^2} \\
\frac{\partial^2 \Phi_p}{\partial x_{j1} \partial x_{k1}} & \frac{\partial^2 \Phi_p}{\partial x_{j2} \partial x_{k1}} & \cdots & \frac{\partial^2 \Phi_p}{\partial x_{jp} \partial x_{k1}} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{\partial^2 \Phi_p}{\partial x_{j1} \partial x_{kp}} & \frac{\partial^2 \Phi_p}{\partial x_{j2} \partial x_{kp}} & \cdots & \frac{\partial^2 \Phi_p}{\partial x_{jp} \partial x_{kp}}
\end{pmatrix} \Phi_p(v,\omega), \quad (2.5)$$

and the rest are similarly defined. From the gradient of $\Phi_p$ in $v$ one can compute the second derivative $D^2\Phi_p$ by finding the mixed derivatives for each pair $(j,k)$ and $(r,s)$. The $(r,s)$-th entries of the blocks $H_{xx}(j,k)$, $H_{yy}(j,k)$ and $H_{xy}(j,k)$ are given by

$$\frac{\partial^2 \Phi_p}{\partial x_{jr} \partial x_{ks}} = \frac{1}{r^2}(1 + \delta_{jk})(\frac{1}{r} \delta_{rs} + \frac{1}{r^2 - s^2} (\frac{r}{s} \gamma_{jk} - \frac{s}{r} \gamma_{kj})(1 - \delta_{rs})$$

$$+ \left( \frac{\rho_{jkl} + \rho_{jlk}}{r(r + s)} + \frac{\rho_{klj} + \rho_{lkj}}{s(r + s)} - \frac{\rho_{klj} + \rho_{lkj}}{rs} \right) y_{l,r+s}$$

$$+ \left( -\frac{\rho_{jkl} + \rho_{jlk}}{rs} - \frac{\rho_{klj} + \rho_{lkj}}{(s-r)s} + \frac{\rho_{klj} + \rho_{lkj}}{r(s-r)} \right) y_{l,s-r},$$

$$\frac{\partial^2 \Phi_p}{\partial y_{jr} \partial y_{ks}} = \frac{1}{r^2}(1 + \delta_{jk})(\frac{2}{r^2} \delta_{rs} - \frac{1}{r^2 - s^2} (\gamma_{jk} - \gamma_{kj})(1 - \delta_{rs})$$

$$+ \left( -\frac{\rho_{jkl} + \rho_{jlk}}{r(r + s)} - \frac{\rho_{klj} + \rho_{lkj}}{s(r + s)} + \frac{\rho_{klj} + \rho_{lkj}}{rs} \right) y_{l,r+s}$$

$$+ \left( -\frac{\rho_{jkl} + \rho_{jlk}}{rs} + \frac{\rho_{klj} + \rho_{lkj}}{(s-r)s} + \frac{\rho_{klj} + \rho_{lkj}}{(r-s)r} \right) y_{l,s-r},$$

$$\frac{\partial^2 \Phi_p}{\partial x_{jr} \partial y_{ks}} = \frac{1}{r} \alpha_{jk} \delta_{rs} + \left( -\frac{\rho_{jkl} + \rho_{jlk}}{r(r + s)} - \frac{\rho_{klj} + \rho_{lkj}}{s(r + s)} + \frac{\rho_{klj} + \rho_{lkj}}{rs} \right) x_{l,r+s}$$

$$+ \left( \frac{\rho_{jkl} + \rho_{jlk}}{rs} + \frac{\rho_{klj} + \rho_{lkj}}{(s-r)s} - \frac{\rho_{klj} + \rho_{lkj}}{r(s-r)} \right) x_{l,s-r}$$

where, again, the summation sign over the repeated index $l = 1, \cdots, q$ is omitted, and all $x$ and $y$-terms with second subscripts outwith the interval $[1, p]$ are assumed to vanish.
3 The Joint Characteristic Function of the Partial Sums

With the gradient and the Hessian matrix of the phase function $\Phi_p(v; \omega)$ in $v$ given above, one can apply the method of stationary phase to study the asymptotic behaviour of the oscillatory integral $\psi_p(\xi)$. A useful tool for this is provided in [11] (Lemma 0.4.7), and the first estimate given in the following lemma is a more quantitative version of it.

Before stating the lemma let us introduce the norm

$$|\varphi|_{K,\Omega} := \max_{0 \leq n \leq K} \sup_{x \in \Omega} |D^n \varphi(x)|$$

for any smooth function $\varphi$ on a bounded domain $\Omega \subset \mathbb{R}^d$ and any natural number $K$.

**Lemma 1.** \(\Psi, \varphi \in C^\infty(\mathbb{R}^k)\) with \(\text{supp} \varphi = \Omega \) bounded. Then for all \(\delta > 0\) and \(K \in \mathbb{N}\),

$$\left| \int_\Omega e^{i|\Psi(x)|} \varphi(x) \, dx \right| \leq C|\varphi|_{K,\Omega} \left( |\Psi|_{K,\Omega}^{\delta - 2K} + 2|\varphi|_{0,\Omega} \Lambda^k(\Omega \setminus \Omega_{\delta}) \right),$$

where \(\Omega_{\delta} := \{x \in \Omega : |\nabla \Psi(x)| > \delta\}\) and the constant \(C\) depends on \(k, K\) and \(\Lambda^k(\Omega)\).

**Proof.** It suffices to bound the integral on \(\Omega_{\delta}\). For any fixed \(K > 0\) write \(M = |\Psi|_{K,\Omega} \vee 1\) and further divide the set \(\Omega_{\delta}\) into several level sets of the gradient:

$$\Omega_r := \{x \in \Omega_{\delta} : 2^{-r} M < |\nabla \Psi(x)| \leq 2^{-r + 1} M\},$$

for \(r = 1, \ldots, r_0 := \lfloor \log_2(M/\delta) \rfloor\); there are at most \(\lfloor \log_2(M/\delta) \rfloor + 1\) non-empty \(\Omega_r\)’s. On each \(\Omega_r\), which is bounded, choose \(\varepsilon_r = 2^{-r} M/(M + 1)\) and let \(N_r = N_r(d, \varepsilon_r)\) be the maximum number s.t. there are \(x_1, \ldots, x_{N_r} \in \Omega_r\) so that the balls \(B(x_j, \varepsilon_r/2)\) are all disjoint. Then one can choose a partition of unity \((\tilde{\phi})\) for \(\{B(x_j, \varepsilon_r/2)\}_j\) as \(\tilde{\phi}(\xi) := \log_{M/\delta}(|\xi|)\). Then the balls \(B(x_j, \varepsilon_r/2)\) must cover \(\Omega_r\): if there is \(x_{\ast} \in \Omega_r\) s.t. \(|x_{\ast} - x_j| > \varepsilon_r\) for all \(j\), then \(B(x_{\ast}, \varepsilon_r/2)\) is disjoint from all other balls \(B(x_j, \varepsilon_r)\) or those with half radius, which contradicts the maximality of \(N_r\). Note that \(\bigcup_{j=1}^{N_r} B(x_j, \varepsilon_r/2) \subset \Omega_{\varepsilon_r/2}\), the \(\varepsilon_r/2\)-neighbourhood of \(\Omega_r\), and therefore

$$N_r \leq \frac{\Lambda^k(\Omega_{\varepsilon_r/2})}{\Lambda^k(\Omega_{\varepsilon_r/2})} \leq C^2 \varepsilon_r^{-k} \Lambda^k(\Omega^{1/2}) \leq C \varepsilon_r^{-k},$$

where \(C\) is a constant depending on \(k\) and the Lebesgue measure of \(\Omega\). This provides a finite open cover for the entire \(\Omega_{\delta}\), and there exist non-negative functions \(\alpha_{j,r} \in C^\infty_0(B(x_j, \varepsilon_r))\) that give a partition of unity (§1.4 in [6]): \(\forall x \in \Omega_{\delta}\),

$$\sum_{r} \sum_{j=1}^{N_r} \alpha_{j,r}(x) = 1,$$

with derivatives satisfying \(|\alpha_{j,r}|_{K,\Omega, B(x_j, \varepsilon_r)} \leq C_{d,K} \varepsilon_r^{-K}\) for all \(K, j, r\). For each \(j\) and \(r\) let \(\tilde{\Psi}_{j,r}(y) := M^{-1} \varepsilon_r^{-2} (\Psi(\varepsilon_r y + x_j) - \Psi(x_j))\). Then for each \(y \in B(0, 1)\), the point \(\varepsilon_r y + x_j \in B(x_j, \varepsilon_r)\), and by Taylor’s theorem, there is some \(x' \in B(x_j, \varepsilon_r)\) s.t.

$$\left| \nabla \tilde{\Psi}_{j,r}(y) \right| = M^{-1} \varepsilon_r^{-1} |\nabla \Psi(\varepsilon_r y + x_j)| \geq M^{-1} \varepsilon_r^{-1} |\nabla \Psi(x_j)| - \frac{1}{2} M^{-1} |\nabla^2 \Psi(x')|$$

$$\geq \varepsilon_r^{-1} 2^{-r} - \frac{1}{2} > \frac{1}{2},$$

11
Since each \( x_j \in \Omega_r \), one applies Taylor’s theorem again to get, for all \( y \in B(0,1) \) and some \( x'' \in B(x_j, \varepsilon_r) \),
\[
|\tilde{\psi}_{j,r}(y)| \leq M^{-1}\varepsilon_r^{-1}|\nabla \psi(x_j)| + \frac{1}{2} M^{-1} \left\| D^2 \psi(x'') \right\| \leq \varepsilon_r^{-2} - r + 1 + \frac{1}{2} \leq \frac{9}{2},
\]
the same argument gives the same upper bound for \( |\nabla \tilde{\psi}_{j,r}(y)| \). For all \( n \geq 2 \), one also has the Euclidean norm \( |D^n \tilde{\psi}_{j,r}(y)| \leq M^{-1}\varepsilon_r^{n-2}|D^n \psi(x_j)| \leq 1 \). Therefore \( \tilde{\psi}_{j,r} \) is in a (uniformly) bounded subset of \( C^\infty(B(0,1)) \).

Now that each function \( \varphi_{j,r} := \alpha_{j,r} \phi \) is supported on the ball \( B(x_j, \varepsilon_r) \), the function \( \psi_{j,r}(y) := \varphi_{j,r}(\varepsilon_r y + x_j) \) is then supported on \( B(0,1) \), satisfying \( |\psi_{j,r}|_{K,B(0,1)} \leq C_{k,K}|\varphi|_{K,\Omega} \) for all \( K,j,r \). Hence using the same arguments as in the proof of Lemma 0.4.7 in [11] one sees:
\[
\left| \int_{B(x_j, \varepsilon_r)} e^{i\xi(x)} \varphi_{j,r}(x) dx \right| = \varepsilon_r^k \left| \int_{B(0,1)} e^{i\varepsilon_r^2\xi(y)} \varphi_{j,r}(\varepsilon_r y + x_j) dy \right| \leq C_{k,K}|\varphi|_{K,\Omega} M^{-K} \varepsilon_r^{k-2K} |\xi|^{-K}.
\]

Finally, since \( \text{supp} \varphi = \Omega \), by the triangle inequality one deduces that
\[
\left| \int_{\Omega_\delta} e^{i\xi(x)} \varphi(x) dx \right| \leq \sum_{r=1}^{N_r} \left| \int_{B(x_j, \varepsilon_r)} e^{i\xi(x)} \varphi_{j,r}(x) dx \right| + |\varphi|_{0,\Omega} \Lambda^k \left( \bigcup_{r,j} B(x_j, \varepsilon_r) \setminus \Omega_\delta \right) \leq C|\varphi|_{K,\Omega} M^{-K} \sum_r N_r \varepsilon_r^{-2K} |\xi|^{-K} + |\varphi|_{0,\Omega} \Lambda^k (\Omega \setminus \Omega_\delta) \leq C|\varphi|_{K,\Omega} M^{-K} \delta^{-2K} |\xi|^{-K} + |\varphi|_{0,\Omega} \Lambda^k (\Omega \setminus \Omega_\delta),
\]
where \( C \) is a constant depending on \( k, K \) and \( \Lambda^k(\Omega) \).

This lemma is to be applied to \( \Psi(v) = \Phi_p(v; \xi/|\xi|) \) and \( \Omega_\delta = \{ v \in \Omega, \ |\nabla_v \Phi_p(v; \xi/|\xi|) | > \delta \} \) for some bounded domain \( \Omega \subset \mathbb{R}^{2dp} \) and any \( \delta > 0 \); in this case the phase function \( \Psi \) also depends on the parameter \( \xi/|\xi| \). Instead of a unit vector consider \( \omega \in \mathbb{R}^d \) s.t. \( |\omega| \geq c \) for some \( c > 0 \); if the \( v \)-derivatives of \( \Psi(v; \omega) \) have no singularity in \( \omega \), then the result holds with \( |\Psi|_{K,\Omega} \) replaced by \( \sup_{|\omega| \geq c} |\Psi(\omega)|_{K,\Omega} \). If the amplitude \( \varphi \) also depends on \( \omega \), then \( |\varphi|_{K,\Omega} \) should be replaced by \( \sup_{|\omega| \geq c} |\varphi(\omega)|_{K,\Omega} \).

It then remains to estimate the Lebesgue measure of the exceptional set \( \Omega \setminus \Omega_\delta \), which would also depend on \( \omega \) if \( \Psi = \Psi(v; \omega) \). The next three lemmas are devoted to this; the idea is to study the degeneracy of the Hessian matrix \( D^2 \Phi_p(v; \omega) \) described by (2.6), (2.8) and (2.7). We start with the following general fact.

**Lemma 2.** Let \( \Omega \subset \mathbb{R}^k \) be open and bounded, \( f : \Omega \to \mathbb{R}^l \) be a \( C^1 \) function. For each \( x \), let \( \sigma_1(x) \geq \sigma_2(x) \geq \cdots \geq \sigma_{k\wedge(l)}(x) \) be the singular values of its derivative \( Df(x) \). For any
\( n \in [1, k \wedge l] \cap \mathbb{N} \) and \( \eta > 0 \), define \( G_{n, \eta}(f) := \{ x \in \Omega : \sigma_n(x) > \eta \} \). If \( Df \) is Lipschitz continuous with Lipschitz constant \( L \), then \( \forall \delta > 0 \),
\[
\Lambda^d(G_{n, \eta}(f) \cap \{ |f| \leq \delta \}) \leq C L^n \eta^{-2n} \delta^n,
\]
where the constant \( C \) depends on \( k, l \) and \( \Lambda^k(\Omega) \).

**Proof.** For fixed \( n, \eta \) and any \( z \in G_{n, \eta} \), by definition the matrix \( Df(z) \) has rank \( n \). This implies that for each \( z \) there are \( n \)-dimensional subspaces \( E_z \) of \( \mathbb{R}^k \) and \( F_z \) of \( \mathbb{R}^l \) s.t., with \( g_z(\cdot) := \pi_{E_z} \circ f|_{E_z}(\cdot) \) and \( \pi \) being the orthogonal projection, the linear map \( Dg_z(z) \) is invertible. Denote by \( E_z^{\perp} \) the orthogonal complement of \( E_z \) for each \( z \).

By the continuity of \( Df \) the set \( G_{n, \eta}(f) \) is open, and the inverse function theorem implies that \( g_z \) is a diffeomorphism in some neighbourhood \( B^{(n)}(z, \varepsilon) \subset E_z \). Moreover, in the proof of the inverse function theorem (see, e.g., Theorem 9.24 in [10] or Theorem 1.17 in [6]), the ball \( B^{(n)}(z, \varepsilon) \) can be typically constructed with radius \( \varepsilon \leq 1/(2L \| (Dg_z(z))^{-1}\|) \leq \| Df(z) \|/(2L) \).

As \( z \in G_{n, \eta}(f) \), one can choose e.g. \( \varepsilon = \eta/(4L) \wedge 1 \).

Since \( G_{n, \eta}(f) \) is bounded, similar to the proof of Lemma [1] there are finitely many points \( z_1, \ldots, z_{N_\varepsilon} \in G_{n, \eta}(f) \) s.t. \( G_{n, \eta}(f) \subset \bigcup_{j=1}^{N_\varepsilon} B(z_j, \varepsilon) \), with the number of balls satisfying
\[
N_\varepsilon \leq \frac{\Lambda^k \left( C_{n, \eta}^{\varepsilon/2}(f) \right)}{\Lambda^k(B(z_j, \varepsilon/2))} \leq C 2^k \varepsilon^{-k} \Lambda^k \left( \Omega^{1/2} \right) \leq C \varepsilon^{-k},
\]
for some constant \( C \) depending on \( k \) and \( \Lambda^k(\Omega) \). Write \( \Gamma_j = B(z_j, \varepsilon) \cap G_{n, \eta}(f) \) and \( P_{j, \delta} = \pi_{E_{z_j}^{\perp}}(\Gamma_j \cap \{ |f| \leq \delta \}) \) for \( \delta > 0 \). For each \((k - n)\)-dimensional vector \( (x_{n+1}, \ldots, x_k) \in P_{j, \delta} \) let \( S_{j, \delta} = S_{j, \delta}(x_{n+1}, \ldots, x_k) \) be the corresponding ‘slice’ of \( \Gamma_j \cap \{ |f| \leq \delta \} \) parallel to \( E_{z_j} \). Then
\[
\Gamma_j \cap \{ |f| \leq \delta \} = \bigcup_{(x_{n+1}, \ldots, x_k) \in P_{j, \delta}} S_{j, \delta}.
\]

Notice that all the singular values of \( Dg_z \) are greater than \( \eta \) on \( \Gamma_j \). Then by a change of coordinates and variables, one has that
\[
\Lambda^k (G_{n, \eta}(f) \cap \{ |f| \leq \delta \}) \leq \sum_{j=1}^{N_\varepsilon} \int_{\Gamma_j \cap \{ |f| \leq \delta \}} dx_1 \cdots dx_k
\]
\[
= \sum_{j=1}^{N_\varepsilon} \int_{P_{j, \delta}} dx_{n+1} \cdots dx_k \int_{S_{j, \delta}} dx_1 \cdots dx_n
\]
\[
= \sum_{j=1}^{N_\varepsilon} \int_{P_{j, \delta}} dx_{n+1} \cdots dx_k \int_{g_{z_j}(\Gamma_j) \cap \{ |y| \leq \delta \}} |\det(Dg_{z_j}(y))|^{-1} dy_1 \cdots dy_n
\]
\[
\leq C \left( \min_j \inf_{x \in \Gamma_j} |\det Dg_{z_j}(x)| \right)^{-1} \delta^n \sum_{j=1}^{N_\varepsilon} \Lambda^{k-n}(B^{(k-n)}(z_j, \varepsilon))
\]
\[
\leq C \eta^{-n} \delta^n N_\varepsilon \varepsilon^{k-n}.
\]

\(^3\)The superscript \((n)\) signifies that it is a ball in \( \mathbb{R}^n \). Balls without superscripts lie in the whole space \( \mathbb{R}^k \).
where the constant $C$ depends on $k, l$ and $\Lambda^k(\Omega)$. Then the result follows from the bound for $N_\varepsilon$ and the choice of $\varepsilon$. \hfill \Box

Now write $G_{n,\eta} = G_{n,\eta}(\nabla \Phi_p(\cdot; \omega))$ as defined in Lemma 2 with $k = l = 2qp$. One then needs to estimate the measure of the complement $\Omega \setminus G_{n,\eta}$ for suitable values of $\eta$ and $n \leq 2qp$. From the expressions (2.6), (2.7) and (2.8) one sees that the behaviour of the second derivatives depends on the magnitude of the parameter $\rho$. Since the differentiation is done w.r.t. the variable $v$, the measure $\Lambda^{2qp}(\Omega \setminus G_{n,\eta})$ may depend on $\omega$, which for now we do not assume to be a unit vector.

The following result gives an estimate for the case where $\rho$ is not too small.

**Lemma 3.** Let $\Omega \subset \mathbb{R}^{2qp}$ be bounded and $n \leq \sqrt{2p}/4$ be an integer. If $|\rho| > \varepsilon$ for some fixed $\varepsilon \in (0, |\omega|)$, then one has $\Lambda^{2qp}(\Omega \setminus G_{n,\eta}) \leq C\varepsilon^{-2n}\eta^n$, where $C$ is a constant depending on $q, p, n$ and $\text{diam}(\Omega)$.

**Proof.** It suffices to focus on a submatrix of $D^2\Phi_p(v; \omega)$ since $\hat{G}_{n,\eta} \subset G_{n,\eta}$ where $\hat{G}_{n,\eta}$ is similarly defined by the singular values of the submatrix. Since $|\rho| > \varepsilon$, locate the (Lyndon) word $(j, k, l^*)$ that gives the maximum entry $|\rho_{jkl^*}| \geq \varepsilon\sqrt{3/(q^3 - q)}$. Then for the fixed pair $(j, k)$ we will focus on the submatrix $H_{xx}(j, k)$.

For a particular pair $(r, s)$, observe from (2.6) that $\partial^2_{x_j x_{ks}} \Phi_p(x, y; \omega)$ contains all the permutations of the word $(j, k, l)$ for each index $l$. Recall that all non-Lyndon entries of $\rho$ are defined to be 0, and that if $(j, k, l)$ is a Lyndon word, we have either $j < k \land l$ or $j = k < l$. Thus for every Lyndon word $(j, k, l)$, out of the rest five permutations only one of $\rho_{jkl}$ and $\rho_{kjl}$ may not vanish, corresponding to the aforementioned two cases respectively. Moreover, in the case $j < k \land l$ one has that

$$\partial^2_{x_j x_{ks}} \Phi_p(v; \omega) = \frac{1}{r^2} \frac{\delta^{(1)}_{jk}}{r - s} \left( \frac{r}{s} \gamma_{jk} - \frac{s}{r} \gamma_{kj} \right) (1 - \delta_{rs})$$

$$+ \left( \frac{\rho_{jkl}}{r(r + s)} - \frac{\rho_{jlk}}{r} \right) y_{l,r+s} + \left( \frac{\rho_{jkl}}{r[s^2 - r]} - \frac{\rho_{jlk}}{rs} \right) \text{sgn}(s-r) y_{l,|s-r|}.$$

Clearly, when $r \neq s$ the coefficients of $y_{l,r+s}$ and $y_{l,|s-r|}$ cannot vanish simultaneously. This is trivial for the case $j = k < l$, where

$$\partial^2_{x_j x_{js}} \Phi_p(v; \omega) = \frac{2}{r^2} \frac{\delta^{(1)}_{jk}}{r} \delta_{rs} + \frac{\rho_{jkl}}{rs} (y_{l,r+s} + y_{l,|s-r|}).$$

The hidden summation in $l = 1, \cdots, q$ in these derivatives then gives linear combinations of $q$ different components $y_{l,r+s}$ and $y_{l,|s-r|}$ of the vector $y$ (for fixed $r \geq s$).

For integers $n \leq m \leq \sqrt{p/2} - 1$, one can choose $r_1, \cdots, r_m, s_1, \cdots, s_m \leq p$ s.t. $r_a \neq s_b$ and the integers $r_a + s_b, |r_c - s_d|$ are all different from one another for all choices of $a, b, c, d = 1, \cdots, m$. For example, one can choose $r_a = a, s_a = a(2m+1)$. In this case, the only choice of $(a, b, c, d)$ s.t. $r_a + s_b = r_c + s_d$, i.e. $(c-a) + (d-b)(2m+1) = 0$, is that $a = c$ and $b = d$; the same for $r_a - s_b = r_c - s_d$. There is no choice of $(a, b, c, d)$ for the equation $(a+c) + (b+d)(2m+1) = 0$.
to hold so \( r_a + s_b = s_c - r_d \) is never satisfied. Since we also require that all of them are no greater than \( p \), it is necessary that \( \max_{a,b}(r_a + s_b) = 2m(m + 1) \leq p \).

Thus one obtains an \( m \times m \) submatrix \( Q_m(y) = Q_m(y; \rho, \beta^{(1)}, \gamma) \) of \( H_{xx}(j,k) \), of which each entry, indexed \((a,b)\), takes the form \( \kappa_{ab} + w_{ab} \cdot y \) with a scalar \( \kappa_{ab} = \kappa_{ab}(\beta^{(1)}_{jk}, \gamma; r_a, s_b) \) and a vector \( w_{ab} = w_{ab}(\rho_{jk}, r_a, s_b) \in \mathbb{R}^{mp} \) satisfying \( |w_{ab}| \geq c_q \varepsilon \) according to (2.6). By the selection of the indices \( r_1, \ldots, r_m, s_1, \ldots, s_m \) in the preceding paragraph, each entry of the submatrix \( Q_m(y) \), translated by the constant \( \kappa_{ab} \), is a linear combination of different components of \( y \) that are all distinct from those appearing in other entries; in other words, the \( m^2 \) vectors \( \{w_{ab}\}_{a,b} \) are mutually orthogonal. Denote the rows of \( Q_m(y) \) by \( q_1(y), \ldots, q_m(y) \), then each \( q_a(y) = \kappa_a + W_a y \) where \( \kappa_a = (\kappa_{a1}, \ldots, \kappa_{am})^\top \) and \( W_a \) is the \( m \times qp \) matrix consisting of the rows \( w_{a1}, \ldots, w_{am} \).

Now define for \( a = 1, \ldots, n \) the set
\[
F_a := \{ (x, y) \in \Omega : \text{dist}(q_a(y), \text{span}\{q_b(y) : b = 1, \ldots, n, b \neq a\}) > \sqrt{n} \eta \},
\]
then \( Q_m(y) \) has rank at least \( n \) for \( (x, y) \in \bigcap_{a=1}^n F_a \). Every point \((x, y) \in F_a \) satisfy
\[
\inf_{c_1, \ldots, c_n \in \mathbb{R}} \left| \kappa_a - \sum_{b \neq a} c_b \kappa_b + \left( W_a - \sum_{b \neq a} c_b W_b \right) y \right| > \sqrt{n} \eta.
\]
The mutual orthogonality of the vectors \( \{w_{ab}\}_{a,b} \) implies the mutual orthogonality of the \( m \) rows of the matrix \( U_a := W_a - \sum_{b \neq a} c_b W_b \), which therefore has a right inverse on an \( m \)-dimensional subspace \( E_m \) of \( \mathbb{R}^{qp} \). Note also that each \( |w_{ab}| \geq c_q \varepsilon \) implies that \( U_a \) restricted on \( E_m \) has norm at least \( c_q \varepsilon \). Hence by the translation-invariance of the Lebesgue measure and the boundedness of \( \Omega \), that for each \( a \),
\[
\Lambda^{2mp}(\Omega \setminus F_a) \leq C |\det(U_a|E_m)|^{-1} (\sqrt{n} \eta)^{m-n+1}
\leq C \|(U_a|E_m)\|^{-m} (\sqrt{n} \eta)^{m-n+1} \leq C \varepsilon^{-m} (\sqrt{n} \eta)^{m-n+1},
\]
where the constant \( C = C(q, p, m, \text{diam}(\Omega)) \) grows at most exponentially in \( m \).

For each point \((x, y) \in \bigcap_{a=1}^n F_a \) and any unit vector \( e = (e_1, \ldots, e_n) \), consider the linear combination \( e \cdot (q_1(y), \ldots, q_n(y)) \) of the \( n \) rows. Choose \( a \) s.t. \( |e_a| = \max\{|e_1|, \ldots, |e_n|\} \geq 1/\sqrt{n} \), then
\[
|e_1 q_1(y) + \cdots + e_n q_n(y)| = |e_a| \left| q_a(y) + \sum_{b \neq a} e_a^{-1} e_b q_b(y) \right| \geq \eta.
\]
Thus, the \( n \times m \) submatrix \( \widehat{Q}_n(y) := (q_1(y)^\top, \ldots, q_n(y)^\top)^\top \) has a right inverse \( R_n(y) \) on an \( n \)-dimensional subspace \( E_n \) of \( \mathbb{R}^m \), and
\[
\|R_n(y)\| = \sup_{|e|=1} \left| R_n(y) e \right| \leq \left( \inf_{|e|=1} \left| e \widehat{Q}_n(y) \right| \right)^{-1} \leq \eta^{-1}.
\]
It then follows from the singular-value decomposition that the singular values of the matrix \( \widehat{Q}_n(y) \) are all bounded from below by \( \|R_n(y)\|^{-1} \geq \eta \), which in turn gives an estimate for the
measure of the exceptional set:
\[
\Lambda^{2q\rho}(\Omega \setminus G_{n,\eta}) \leq \Lambda^{2q\rho}(\Omega \setminus \bar{G}_{n,\eta}) \leq \Lambda^{2q\rho}\left(\bigcup_{a=1}^{n}(\Omega \setminus F_a)\right) \leq Cn\varepsilon^{-m}(\sqrt{n\eta})^{m-n+1},
\]
and the result follows by taking \( m = 2n - 1 \). \( \Box \)

The result of Lemma 3 is meaningful for small values of \( \varepsilon \) and \( \eta \). It remains to show that the measure \( \Lambda^{2q\rho}(\Omega \setminus G_{n,\eta}) \) is also small when \( \rho \) is small.

**Lemma 4.** Let \( \Omega \subset \mathbb{R}^{2q\rho} \) be bounded and \( n \) be an even integer s.t. \( n + 1 \) is prime. Then, depending on \( q, p, n \) and \( \text{diam}(\Omega) \), one can choose \( \varepsilon, \delta, \eta > 0 \) sufficiently small s.t. for \( |\rho| \leq \varepsilon \), either \( \Omega_{\delta} = \Omega \) or \( G_{n,\eta} = \Omega \).

**Proof.** For \( \varepsilon \in (0, |\omega|/\sqrt{2}) \) define \( \varepsilon' = \sqrt{|\omega|^2 - \varepsilon^2} \in (|\omega|/\sqrt{2}, |\omega|) \), and assume \( \text{diam}(\Omega) = 1 \) w.l.o.g., otherwise replace \( \varepsilon \) with \( \varepsilon/(1 \vee \text{diam}(\Omega)) \). First of all that \( ||\rho|| \leq \varepsilon \) implies that the vector \( (\alpha, \beta, \gamma, a, b) \) has modulus no less than \( \varepsilon' \). The proof is divided into several cases depending on which components of this vector are dominant in modulus or norm.

Let us first consider the case where the coefficients \((a, b)\) are ‘dominant’ in the sense that \( |(a, b)| > \varepsilon' \sqrt{1 - \theta^2} \geq |\omega|/2 \) for some \( \theta \in (0, 1/\sqrt{2}) \) to be chosen later. In this case \( |(\alpha, \beta, \gamma)| \leq \varepsilon' \theta \). From the expression (2.2) one has the following bound:
\[
|\partial_{x_{jr}} \Phi_p(v; \omega)|^2 \geq \frac{1}{r^2}a_j^2 - \frac{2}{r}|a_j||Q_{x_{jr}}(v; \rho)| - 2\left(\frac{1}{r}|a_j| + |Q_{x_{jr}}(v; \rho)|\right)|L_{x_{jr}}(v; \alpha, \beta, \gamma)|,
\]
where \( L_{x_{jr}}(v; \alpha, \beta, \gamma) \) and \( Q_{x_{jr}}(v; \rho) \) denote the linear and quadratic parts for \( v \) in (2.2). Since \( x \) and \( y \) are bounded, one has that
\[
|Q_{x_{jr}}(v; \rho)| \leq C_q||p||\frac{1}{r}\left(\sum_{s=1}^{p-r}1 + \sum_{s=1}^{r-1}1\right) \leq C_q\frac{\varepsilon}{r} \log p,
\]
and that, omitting the summation in \( k \),
\[
|L_{x_{jr}}(v; \alpha, \beta, \gamma)| \leq \frac{1}{r}|\alpha_{jk}||y_{kr}| + \frac{2}{r^2}|\beta^{(1)}_{jk}||x_{kr}| + \frac{1}{r}(|\gamma_{jk}| + |\gamma_{kj}|)\sum_{s \neq r}1 |x_{ks}|
\leq C_q \frac{\log p}{r} \varepsilon' \theta.
\]
Hence one derives that
\[
|\partial_{x_{jr}} \Phi_p(v; \omega)|^2 \geq \frac{1}{r^2}a_j^2 - C_q\frac{\varepsilon \log p}{r^2}|a_j| - C_q\varepsilon' \theta \frac{\log p}{r}\left(\frac{1}{r}|a_j| + \frac{\varepsilon}{r} \log p\right),
\]
and a similar inequality for \( |\partial_{y_{jr}} \Phi_p/v|^2 \) with \( a_j/r \) replaced with \( b_j/r^2 \) as per (2.3). Thus, summing up \( j \) and \( r \) one has that
\[
|\nabla \Phi_p(v; \omega)|^2 \geq C|(a, b)|^2 - C_q|(a, b)|\varepsilon \log p - C_q\varepsilon' \theta (\log p)^2 (|a, b| + \varepsilon \log p)
\geq C|\omega|^2 - C_q|\omega|\varepsilon \log p - C_q|\omega|\theta (\log p)^2 (|\omega| + \varepsilon \log p),
\]
(3.1)
which has a fixed lower bound for \( \varepsilon < |\omega|(\log p)^{-1} \) and \( \theta < (\log p)^{-2} \) sufficiently small. Then for small values of \( \delta < C|\omega| \) we have \( \Omega = \Omega_\delta \).

Now suppose that \( |(a, b)| \leq \varepsilon' \sqrt{1 - \theta^2} \), then \( |(\alpha, \beta, \gamma)| \geq \varepsilon' \theta \). The latter corresponds to the constant terms in the second derivative \( D^2 \Phi_p(v, \omega) \). Write

\[
D^2 \Phi_p(v; \omega) = A_p + L_p(v; \rho)
\]

according to the expressions (2.6), (2.7) and (2.8), where \( A_p = A_p(\alpha, \beta, \gamma) \) and \( L_p(v; \rho) = \{(L_{x_jy_jk_1}, L_{y_jy_1k_1}, L_{x_jy_1k_1})(v; \rho)\} \), where \( j, k, r, s \) are the constant and linear parts in \( v \), respectively. Then for each \((j, k)\) and \((r, s)\),

\[
\sup_{v \in \Omega} |L_{x_jy_jk_1}(v; \rho)| \leq C_q \|\rho\| \left( \frac{1}{rs} + \frac{\delta_{rs}}{r|r - s|} + \frac{\delta_{rs}}{s|s - r|} \right) \leq C_q \varepsilon \rho \land \theta,
\]

and the same bound holds for \( L_{y_jy_1k_1} \) and \( L_{x_jy_1k_1} \). Let \( H_n(v; \omega) = A_n + L_n(v; \rho) \) be an \( n \times n \) submatrix of \( D^2 \Phi_p(v; \omega) \) with ‘constant’ part \( A_n = A_n(\alpha, \beta, \gamma) \) and linear part \( L_n(v; \rho) \). The estimate above then gives a bound for the 2-norm \( \|L_n(v; \rho)\| \leq C_q \varepsilon \) for all \( v \in \Omega \). By definition, \( \det H_n(v; \omega) = \det A_n + P_n(v; \rho) \), where \( P_n(v; \rho) \) is a polynomial of degree \( n \) in \( v \) and \( \rho \) with no constant terms. Therefore for all \( v \in \Omega \), \( \omega \in \mathbb{R}^d \),

\[
|\det H_n(v; \omega)| \geq |\det A_n| - C_q \varepsilon.
\]

If \( A_n \) is invertible with a bound \( \|A_n^{-1}\| \leq \tau^{-1} \) for some \( \tau \in (0, 1) \), then for \( \varepsilon \ll \tau \),

\begin{enumerate}
  \item \( |\det A_n| \geq \|A_n^{-1}\|^{-n} \geq \tau^n \), so \( H_n \) is invertible for all \( v \in \Omega \), and
  \item \( \|H_n^{-1}\| \leq \|A_n^{-1}\|/(1 + \|A_n^{-1}\|\|L_n\|) \leq 2\tau^{-1} \).
\end{enumerate}

The second inequality in (b) comes from the fact that \( \|(I + B)^{-1}\| = \sum_{k \geq 0} \|(-B)^k\| \leq \sum_{k \geq 0} \|B\|^k = \frac{1}{1 - \|B\|} \) for any square matrix \( B \) s.t. \( \|B\| < 1 \). Thus, by the singular-value decomposition (a) and (b) will imply that \( H_n(v; \omega) \) as an \( n \times n \) submatrix of \( D^2 \Phi_p(v; \omega) \) has singular values no less than \( \tau/2 \) for all \( v \in \Omega \), in other words, \( \Omega \ \setminus \ G_{n, \tau/2} = \emptyset \). In particular, for any \( \eta \leq \tau/2 \) we have \( \Omega \ \setminus \ G_{n, \eta} = \emptyset \), too. Henceforth, one looks for an invertible \( n \times n \) submatrix \( A_n \) of \( A_p \) with an appropriate bound for \( \|A_n^{-1}\| \), and the result will follow by choosing sufficiently small values of \( \varepsilon \) and \( \eta \).

Write \( D_n = \text{diag}(1, 1/2, \ldots, 1/n) \), \( n \leq p \) for simplicity. If the component \( \alpha \) is ‘dominant’ amongst \( \alpha, \beta, \gamma \) in the sense that, for example, \( \|\alpha\| > \varepsilon' \theta / \sqrt{3} \geq |\omega| \theta / \sqrt{6} \), choose the largest entry \( |\alpha_{jk}| \geq c_q |\omega| \theta \). Then by (2.8) the constant part of the \( n \)-th principle submatrix of the block \( H_{xy}(j, k) \) is \( A_n = \alpha_{jk} D_n \), and \( \|A_n^{-1}\| \leq |\alpha_{jk}|^{-1} n \). Thus the result holds for \( \tau \lesssim \varepsilon |\omega| \theta / n \).

On the other hand we need to consider the case where \( |(\beta, \gamma)| \geq \varepsilon' \theta / \sqrt{2} \). If the largest entry of \( (\beta, \gamma) \) is located on the diagonal, i.e. \( |(\beta\gamma)| \geq c_q \varepsilon' \theta \) (recall that \( \gamma_{jj} = 0 \) for \( i = 1 \) or \( 2 \) and some \( j \)), then the constant part of the \( n \)-th principle submatrix of the block \( H_{xx}(j, j) \) or the block \( H_{yy}(j, j) \) is \( A_n = 2\beta_{jj} D_n^2 \) by (2.6) and (2.7). Hence we have that \( \|A_n^{-1}\| \leq 2\beta_{jj}^{-1} n^2 \) and we need \( \tau \lesssim \varepsilon |\omega| \theta / n^2 \).
The situation is trickier when the largest entry is found off diagonal, i.e. for some pair \((j, k)\) (assuming \(j < k\) w.l.o.g.). Consider the constant part \(A^{(2)}_n\) of the \(n\)-th principle submatrix of the block \(H_{yy}(j, k)\). By \((2, 7)\) it takes the form

\[
A^{(2)}_n = \begin{pmatrix}
\frac{1}{\sqrt{n}} \gamma_{jk} & \frac{1}{\sqrt{n}} \gamma_{jk} & \cdots & \frac{1}{\sqrt{n}} \gamma_{jk} \\
\frac{1}{\sqrt{n}} \gamma_{jk} & \frac{1}{\sqrt{n}} \gamma_{jk} & \cdots & \frac{1}{\sqrt{n}} \gamma_{jk} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{\sqrt{n}} \gamma_{jk} & \frac{1}{\sqrt{n}} \gamma_{jk} & \cdots & \frac{1}{\sqrt{n}} \gamma_{jk}
\end{pmatrix}
\]

where \(S_n\) is the skew-symmetric matrix with \((r, s)\)-th entry \((s^2 - r^2)^{-1}\), \(r \neq s\) and 0 on the diagonal. If \(|\beta_{jk}^{(2)}| \geq c_q \varepsilon \theta\), then the matrix \(A^{(2)}_n\) has full rank. To see this, notice that the matrix \(S_n := D_n^{-1} S_n D_n^{-1}\) is also skew-symmetric and has purely imaginary eigenvalues only. Then all the eigenvalues of the scaled matrix \(A^{(2)}_n := I + \gamma_{jk} S_n / \beta_{jk}^{(2)}\) have real parts 1, which serves as a lower bound for the operator norm of \(A^{(2)}_n\) as it is in fact a normal matrix, and so \(\|A^{(2)}_n\| < 1\). Therefore \(\|A^{(2)}_n\|^{-1} = \|\beta_{jk}^{(2)} D_n A^{(2)}_n D_n^{-1}\|^{-1} \leq |\beta_{jk}^{(2)}|^{-1} n^2\) and again we need \(\tau \lesssim q |\omega|/n^2\).

The same applies to the case where \(|\beta_{jk}^{(1)}| \geq c_q \varepsilon \theta\): instead of \(A^{(2)}_n\) consider the constant part \(A^{(1)}_n\) of the \(n\)-th principle submatrix of the block \(H_{xx}(j, k)\), which by \((2, 6)\) takes the form

\[
A^{(1)}_n = \beta_{jk}^{(1)} D_n^2 + \gamma_{jk} S_n^1 \text{ where } S_n^1 \text{ is the matrix with } (r, s)\text{-th entry } (r^2 - s^2)^{-1} r/s. \text{ Then it suffices to observe that } S_n^1 = -D_n^{-1} S_n D_n \text{ and } A^{(1)}_n = \beta_{jk}^{(1)} (I - \gamma_{jk} S_n / \beta_{jk}^{(1)}) D_n^2.
\]

Finally, if \(|\gamma_{jk}| \geq c_q \varepsilon \theta\) is the largest entry of \((\beta, \gamma)\), we return to the matrix \(A^{(2)}_n\). Since \(S_n\) is skew-symmetric, \(\det S_n = 0\) for all odd \(n\). If \(n\) is even, by definition the determinant of \(S_n\) is given by the expansion

\[
\det S_n = \sum_{\sigma \in \Pi_n} \text{sgn}(\sigma) \frac{1}{1 - \sigma_1^2} \frac{1}{4 - \sigma_2^2} \cdots \frac{1}{n^2 - \sigma_n^2},
\]

where \(\Pi_n\) is the set of permutations of \((1, \cdots, n)\) with no fixed points. Notice that this summation includes the product of all the entries along the reflected diagonal \(r + s = n + 1\), each of which has denominator divisible by \(n + 1\). Clearly, out of all the permutations this product is the only term in the above expansion whose denominator is divisible by \((n + 1)^n\) if \(n + 1\) is prime. Then it follows from the fundamental theorem of arithmetic that \(P_n := \det S_n \neq 0\). It is rather difficult to compute the the value \(P_n\) explicitly; computer results for large values of \(n\) up to 400 shows that it decays roughly exponentially. Notice that \(A^{(2)}_n = D_n (I + \beta_{jk}^{(2)} S_n^{-1} \gamma_{jk} S_n^{-1}) D_n^{-1} S_n^{-1} \gamma_{jk} S_n\) and that the matrix \(S_n^{-1}\) is still skew-symmetric, the same argument used in the previous cases still applies. Therefore \(\|A^{(2)}_n\|^{-1} \leq |\gamma_{jk}|^{-1} \|S_n^{-1}\| \|n\), and we need \(\tau \lesssim q |\omega| P_n^{1/n} / n\).

Combining all the criteria above, for an even integer \(n\) s.t. \(n + 1\) is prime one can choose \(\tau \leq q |\omega|(p \log p)^{-1} P_n^{1/n} n^{-2}\) s.t. the result holds true for \(\varepsilon \lesssim q |\omega| \log p \wedge \tau^n\) sufficiently small, and any \(\delta \leq q |\omega|/4\) and \(\eta < \tau\) sufficiently small. \(\blacksquare\)
These lemmas altogether give an estimate for oscillatory integrals of the type

\[ T(\xi) = \int_{\Omega} e^{i\xi|\Phi_p(v;\xi/|\xi|)} \varphi(v,\xi)dv, \]

for a bounded domain \( \Omega \) and a smooth amplitude \( \varphi \) supported on \( \Omega \times \mathbb{R}^d \). In order to study the global behaviour of it, in particular, the characteristic function \( \psi_p(\xi) \) of \( V_p \), some cut-off arguments will be needed to derive a similar estimate as in Lemma \[ \Pi \] on the whole space \( \mathbb{R}^{2q} \).

But as the reader will realise later, to find a desired coupling for \( V_p \) it is necessary to estimate oscillatory integrals with amplitudes other than just \( \phi_p \). For a Schwartz function \( \varphi \in \mathcal{S}(\mathbb{R}^q) \) and \( k, l \in \mathbb{N} \), introduce the norm

\[ \| \varphi \|_{k,l} = \max_{|\theta| \leq q, |j| \leq k, \xi \in \mathbb{R}^q} |\xi^\theta \partial^j \varphi(\xi)|, \]

where \( \theta, \tau \in \mathbb{N}^q \) are multi-indices. Then for \( \varphi \in C_0^\infty(\Omega) \) it holds that \( \| \varphi \|_{k,\Omega} \approx_q \| \varphi \|_{0,k} \).

**Theorem 5.** For any \( K > 0 \), let \( p_0 > 8K^2 \) be a fixed even integer s.t. \( \lceil \sqrt{2p_0}/4 \rceil + 1 \) is prime. For any \( p > p_0 \) and a Schwartz function \( \varphi \in \mathcal{S}(\mathbb{R}^{2q}) \), let \( \Phi_p(v,\omega) \) be the phase function for \( V_p \) and for \( \xi \in \mathbb{R}^d \) define

\[ I_p(\xi) := \int_{\mathbb{R}^{2q}} e^{i\xi|\Phi_p(v;\xi/|\xi|)|} \varphi(v)dv. \]

If the amplitude \( \varphi \) can be factorised as the product of two Schwartz functions \( \varphi_0 \in \mathcal{S}(\mathbb{R}^{2q}) \) and \( \varphi_1 \in \mathcal{S}(\mathbb{R}^{2q}) \), \( p' := p - p_0 \), then \( I_p \in C^\infty(\mathbb{R}^d) \) and for any \( k \in \mathbb{N} \) and \( |\xi| \) sufficiently large it holds that

\[ |D^k I_p(\xi)| \leq C_{q,p_0,k,K} \xi^{-K/16} \| \varphi_0 \|_{(2q+1)/2p_0+K+1+3k,k} \int_{\mathbb{R}^{2q}} (1 + |v'|^{2p_0+3k}) \varphi_1(v')dv', \]

uniformly for all \( p > p_0 \).

We highlight the crucial fact here that the constants in the estimates for \( I_p(\xi) \) and its derivatives are all independent of the parameter \( p \).

**Proof.** Let us prove the uniform (in \( p \)) decay of \( |I_p(\xi)| \) first. Instead of \( I_p(\xi) \) let us consider for now the oscillatory integral (with \( \omega := \xi/|\xi| \))

\[ I_{p_0}(\xi) = \int_{\mathbb{R}^{2q}} e^{i\xi|\Phi_{p_0}(v_0;\omega)|} \varphi_0(v_0)dv_0 \]

for a fixed positive integer \( p_0 \). First choose a non-negative, smooth cut-off function \( \zeta_0 \in C_0^\infty(B(0,2)) \) s.t. \( \zeta_0 \equiv 1 \) on \( B(0,1) \) and all its derivatives are bounded on \( B(0,2) \setminus B(0,1) \). Divide the rest of \( \mathbb{R}^{2q} \) by the annuli

\[ A_r := \{ v_0 \in \mathbb{R}^{2q} : 2^{r-1} \leq |v_0| < 2^r \}, \quad r \in \mathbb{N}, \]

and define the fattened annuli \( A'_r := \{ 2^{r-2} \leq |v_0| < 2^{r+1} \} \). Choose another non-negative, smooth cut-off \( \zeta_1 \in C_0^\infty(A'_1) \) taking value 1 on \( A_1 \) and bounded derivatives on \( A'_1 \setminus A_1 \), and...
define \( \zeta_r(v_0) := \zeta_1(2^{-r+1}v_0) \), \( \forall r \geq 2 \). Then for each \( r \geq 1 \), the smooth function \( \zeta_r \) is supported on the fattened annulus \( A'_r \) with value 1 on \( A_r \) and bounded derivatives on \( A'_r \setminus A_r \); the sum \( \sigma(v_0) := \sum_{r=0}^{\infty} \zeta_r(v_0) \) is then supported on the whole of \( \mathbb{R}^{2qp_0} \).

If one further sets \( \tilde{\zeta}_r(v_0) := \zeta_r(v_0)/\sigma(v_0) \), then each \( \tilde{\zeta}_r \) has the same properties as those of \( \zeta_r \), and \( \sum_{r=0}^{\infty} \tilde{\zeta}_r \equiv 1 \) trivially. Then one can write

\[
I_{p_0}(\xi) = \int_{B(0,2)} e^{i\xi|\Phi_{p_0}(v_0,\omega)|} \varphi_0(v_0)\tilde{\zeta}_0(v_0) dv_0 + \sum_{r=1}^{\infty} \int_{A'_r} e^{i\xi|\Phi_{p_0}(v_0,\omega)|} \varphi_0(v_0)\tilde{\zeta}_r(v_0) dv_0
\]

where the first integral can be readily estimated by the lemmas above. Since the function \( \Phi_{p_0}(v_0;\omega) \) is a cubic polynomial and the vector \( \omega = (a,b,\alpha,\beta,\gamma,\rho) \) is normalised, all the derivatives of \( \tilde{\Phi}_{p_0} \) are uniformly bounded on \( B(0,2) \); so do all the derivatives of \( \tilde{\zeta}_0 \) by its construction.

Thus, applying Lemma 1 we have, \( \forall K, \delta_0 > 0, \ \xi \in \mathbb{R}^d \),

\[
|T_0(\xi)| \leq C_{q,p_0,K} |\varphi_0|_{K,B(0,2)} \delta_0^{-2K} |\xi|^{-K} + 2|\varphi_0|_{0,B(0,2)} \Lambda^{2qp_0}(\Gamma_{\delta_0}),
\]

where \( \Gamma_{\delta_0} = \{ v_0 \in B(0,2) : |\nabla \Phi_{p_0}(v_0,\omega)| \leq \delta_0 \} \). The set \( \Gamma_{\delta_0} \) can be further split by the set \( G_{n,\eta_0} = G_{n,\eta_0}(\nabla \Phi_{p_0}) \) as defined in Lemma 2 and its complement for some \( \eta_0 > 0 \) and some integer \( n \). Note that the Lipschitz constant of \( D^2\Phi_{p_0} \) is at most \( |p| \leq 1 \). Then by Lemma 2 3 and 4 one sees that for any \( \delta_0, \eta_0, \varepsilon_0 > 0 \) sufficiently small and any even integer \( n \leq \sqrt{2p_0}/4 \) s.t. \( n + 1 \) is prime, one has that

\[
\Lambda^{2qp_0}(\Gamma_{\delta_0}) \leq \Lambda^{2qp_0}(\Gamma_{\delta_0} \cap G_{n,\eta_0}) + \Lambda^{2qp_0}(\Gamma_{\delta_0} \setminus G_{n,\eta_0})
\]

\[
\lesssim q_{p_0,n} \eta_0^{-2n \delta_0^{-n}} + \varepsilon_0^{-2n \eta_0^{-n}}.
\]

(3.4)

Thus, choosing \( \eta_0 = \delta_0^{1/4} \), \( \delta_0 = |\xi|^{-1/4} \), and \( \varepsilon_0 \lesssim q (np \log p)^{-n} (n^{-n} \wedge p_n) \) one has that

\[
|T_0(\xi)| \leq C_{q,p_0,K} |\varphi_0|_{K,B(0,2)} |\xi|^{-K} + C_{q,p_0,n} |\varphi_0|_{0,B(0,2)} \left( |\xi|^{-\frac{n}{2}} + |\xi|^{-\frac{1}{4}n} \right)
\]

\[
\leq C_{q,p_0,n,K} |\varphi_0|_{0,K} |\xi|^{-\frac{1}{4}K}
\]

for \( |\xi| \) sufficiently large and \( n > K \). Hence we choose \( p_0 > 8K^2 \) s.t. \( \sqrt{2p_0}/4 \) + 1 is prime and set \( n = \left[ \sqrt{2p_0}/4 \right] \).

For each \( r \geq 1 \), let \( \tilde{\omega} = (\rho, 2^{-r} \alpha, 2^{-r} \beta, 2^{-r} \gamma, 2^{-2r} a, 2^{-2r} b) \) and consider the scaled phase function \( \tilde{\Phi}_{p_0}(u_0;\tilde{\omega}) = 2^{-3r} \Phi_{p_0}(2^{-r} u_0;\omega) \) for all \( u_0 \in A'_0, \ \omega \in S^{d-1} \). This is again a cubic polynomial with bounded coefficients and so \( |\Phi_{p_0}(\cdot;\tilde{\omega})|_{K,A'_0} \leq C_{q,p_0} \) for any \( K > 0 \). Scaling each annulus \( A'_r \) down to \( A'_0 \subset B(0,2) \) one has that

\[
T_r(\xi) = \int_{A'_0} e^{i2\pi r |\xi|\Phi_{p_0}(u_0,\tilde{\omega})} \chi_r(u_0) du_0,
\]
where \( \chi_r(u_0) = 2^{2q_0r}\varphi_0(2^{-r}u_0)\zeta_1(2u_0)/\sigma(2^r u_0) \in C_0^\infty(A'_0) \). Applying Lemma \[ again to this new expression of \( T_r \) on \( A'_0 \) one sees that \( \forall \delta_r, K > 0, \)

\[
|T_r(\xi)| \leq C_{q,p_0,K}|\chi_r|_{K,A'_0} \delta_r^{-2K}(2^{3r}|\xi|)^{-K} + 2|\chi_r|_{0,A'_0} 2^{q_0p_0}(\tilde{\Gamma}_{\delta_r}),
\]

where \( \tilde{\Gamma}_{\delta_r} = \{v_0 \in A'_0 : |\nabla \Phi_{p_0}(v_0; \tilde{\omega})| \leq \delta_r \}. \) Splitting \( \tilde{\Gamma}_{\delta_r} \) according to the set \( \tilde{G}_{n,q_r} := G_{n,q_r}(\nabla \Phi_{p_0}(\cdot; \tilde{\omega})) \) one obtains the estimate \( (3.4) \) for the measure of the set \( \tilde{\Gamma}_{\delta_r} \) again from Lemma \[ and \[. \) Note that here instead of unit frequencies \( 1 \)

\[
\Phi_r(\xi), v_r
\]

\[
\chi
\]

\[
\sup_{v_0 \in A'_0} |\partial^\theta \varphi_0(2^r u_0)| \leq 4^{|\theta|} \sup_{v_0 \in A'_0} |2^r u_0|^\theta |\partial^\theta \varphi_0(2^r u_0)| \leq 4^{|\theta|}\|\varphi_0\|_{\theta,|\theta|,|r|}.
\]

Therefore for any \( k, l \geq 0, \) differentiating \( \chi_r \) up to \( k \) times by Leibniz’s rule one sees that \( 2^r|\chi_r|_{k,A'_0} \leq C_{q,p_0,k,l} \|\varphi_0\|_{2q_0+l+k,k,k} \) from the boundedness of the derivatives of \( \zeta_1 \) and \( \sigma. \) This in turn implies that \( |T_r(\xi)| \leq C_{q,p_0,K} 2^{-r} |\xi|^{-K/16} \|\varphi_0\|_{(2q+1/2)p_0+K+1,K}. \) Summing up in \( r \) and one achieves the bound

\[
|I_{p_0}(\xi)| \leq C_{q,p_0,K}|\xi|^{-K/16}\|\varphi_0\|_{(2q+1/2)p_0+K+1,K}.
\]

It remains to bound the original integral \( I_p(\xi) \) in question for all \( p \geq p_0. \)

Write \( v'_p := \{(x_j,y_{ks}) : j,k = 1, \cdots, q; r, s = p_0 + 1, \cdots, p\}, \) then conditional on \( v'_p \) the integral \( I_p(\xi) \) can be written as, by the factorisation assumption,

\[
I_p(\xi) = \int_{\mathbb{R}^{2q_0'}} \varphi_1(v')dv' \int_{\mathbb{R}^{2q_0}} e^{i\xi|\Phi_p(v_0,v';\tilde{\omega})\varphi_0(v_0)}dv_0
\]

\[
= \int_{\mathbb{R}^{2q_0'}} J_p(\xi,v')\varphi_1(v')dv'. \tag{3.5}
\]

If one can show that \( |J_p(\xi,v')| \) has a global decay in \( |\xi| \) uniformly in \( p \) and at most polynomial growth in \( v' \), then such a decay should be passed on to \( |I_p(\xi)| \) by the rapid decay of \( \varphi_1. \) The idea is that for a fixed value of \( v' \) (equivalently, conditional on the random variable \( v'_p \) the oscillatory integral \( J_p(\xi,v') \) has the same global behaviour in \( \xi \) as \( I_{p_0}(\xi) \).

Using the same cut-off arguments, it suffices to focus on the case where the amplitude \( \varphi_0 \) is compactly supported on \( B(0,2) \subset \mathbb{R}^{2q_0}. \) Fixing the value of \( v' \) one easily sees that \( |\Phi_p(\cdot,v')|_{K,B(0,2)} \leq C_{q,p_0,K}(1 + |v'|^3) \) for any \( K > 0, \) and Lemma \[ can be readily applied w.r.t. \( v_0 : \forall K, \delta' > 0, \)

\[
|J_p(\xi,v')| \leq C_{q,p_0,K}|\varphi_0|_{K,B(0,2)}(1 + |v'|^3K)(\delta')^{-2K}|\xi|^{-K} + 2|\varphi_0|_{0,B(0,2)}A^{2q_0p_0}(\Gamma_{\delta'}/r),
\]
Hence Lemma 2 applies directly and gives \( \Lambda \). Proof, a constant vector \( \theta \)...

**Figure 1:** Monomials of \( \Theta(v_0, v') \) in the shaded areas.

where \( \Gamma_{\delta'} = \{ v_0 \in B(0, 2) : |\nabla_1 \Phi_p(v_0, v')| \leq \delta' \} \). To estimate the measure of the set \( \Gamma_{\delta'} \), which may depend on \( v' \), one divides it by the set \( G'_{n, \eta'} := G_{n, \eta'}(\nabla_1 \Phi_p(\cdot, v')) \) just like before. The key is then to recognise the \( v_0 \)-derivatives of \( \Phi_p(v_0, v') \).

Recall the definition \((2.1)\) and write

\[
\Phi_p(v_0, v'; \omega) = \Phi_{p_0}(v_0; \omega) + \sum_{j=1}^{q} \sum_{r=p_0+1}^{p} \frac{a_j}{r} x_{jr} + \sum_{j=1}^{q} \sum_{r=p_0+1}^{p} \frac{b_j}{r^2} y_{jr} + \sum_{j<k} \sum_{r=p_0+1}^{p} \frac{\beta_{jk}}{r^2} (x_{jr} x_{kr} + y_{jr} y_{kr}) + \sum_{j<k} \sum_{r=p_0+1}^{p} \frac{\alpha_{jk}}{r} (x_{jr} y_{kr} - y_{jr} x_{kr}) + \Theta_p(v_0, v'; \gamma, \rho) + \Upsilon_p(v', \gamma, \rho),
\]

\((3.6)\)

where \( \Theta_p(v_0, v'; \gamma, \rho) + \Upsilon_p(v'; \gamma, \rho) = \sum_{j<k} \gamma_{jk} (\nu_{jk}^{(p)} - \nu_{jk}^{(p_0)}) + \sum_{(j,k,l) \in \mathbb{Z}_+} \rho_{jk} (\Delta_{jk}^{(p)} - \Delta_{jk}^{(p_0)}) \) and \( \Upsilon_p(v') \) is the sum of monomials that do not involve \( v_0 \). Then it is clear from the expression above that the function \( \Phi_p(v_0, v') \) has the same derivatives in \( v_0 \) as the function

\[
\Psi_p(v_0, v') := \Phi_{p_0}(v_0) + \Theta_p(v_0, v').
\]

It is rather cumbersome to write down \( \Theta_p(v_0, v') \) explicitly, but it is not hard to see the ranges of the indices \( r, s \) for its monomials - they are the shaded areas in Figure 1a and 1b, corresponding to \( \nu(q) - \nu(p_0) \) and \( \Delta(q) - \Delta(p_0) \), respectively. What is more important is that the polynomial \( \Theta(v_0, v') \) is at most quadratic in \( v_0 \). The monomials corresponding to the grey areas are linear in \( v_0 \) while the ones corresponding to the black area are quadratic in \( v_0 \).

Considering the variable \( v_0 \) only, the Lipschitz constant of the linear function \( D_2^2 \Psi_p(v_0, v') \) is identical to that of \( D_2^2 \Phi_{p_0}(v_0) \) since they only differ by a constant matrix \( D_1^2 \Theta_p(v_0, v') \). Hence Lemma 2 applies directly and gives \( \Lambda_{2np}(\Gamma_{\delta'} \cap G_{n, \eta'}') \leq C_{q,p_0}(\eta')^{-2n}(\delta')^n \) uniformly in \( p \). This uniformity also holds when applying Lemma 3; the difference here is that, in its proof, a constant vector \( \theta_a(v') \) from the matrix \( D_1^2 \Theta_p(v_0, v') \) is added to each row \( q_a(y) \) of the submatrix \( Q_{n}(y) \) therein, and the sets \( F_a \) are replaced by

\[
F_a' = \{ (x, y) : \text{dist}(q_a(y) + \theta_a(v'), \text{span}\{q_b(y) + \theta_b(v') : a \neq b = 1, \cdots, n\}) > \sqrt{n} \eta' \}.
\]
The function follows from the relation (3.5).

By the same scaling argument for a general Schwartz function and so with the same values of and \( \epsilon \), defined as \( \sum \) where \( \rho, \alpha \), both and \( \beta, \gamma \) are similarly defined as \( \alpha' \) and \( \beta', \gamma' \) are similarly defined as \( \alpha', \beta', \gamma' \). Then recalling that \( |\omega| = 1 \) we have that, for \( |\rho| \lesssim \epsilon \),

\[
|\omega|^2 \gtrsim |\rho|^2 + |(\alpha, \beta, \gamma)|^2 + |(\alpha'', \beta'')|^2 - 2|\rho||\omega'|-2(\alpha'', \beta'')|(|\rho||\omega|^2)
\]

If \( |(\alpha'', \beta'')| > |(\alpha, \beta)/\|v\|^2 \) then \( |\omega| \gtrsim 1 - \epsilon(1 + |\omega|^3) \geq 1 \) for \( \epsilon' \ll 1/(1 + |\omega|^3) \). On the other hand if \( |(\alpha'', \beta'')| \leq |(\alpha, \beta)/\|v\|^2 \), then \( \omega' \neq 0 \) and \( |\gamma| \simeq |(\alpha, \beta)/\|v\|^2 \), implying

\[
|\omega|^2 \gtrsim |\rho|^2 + |(\alpha, \beta, \gamma)|^2 - \epsilon(1 + |\omega|^3) \gtrsim 1 - \epsilon(1 + |\omega|^3).
\]

Therefore, recalling the choices for \( \epsilon_0, \delta_0, \eta_0 \) altogether we have that, for \( |\xi| \) sufficiently large,

\[
\Lambda^{2q\rho_0}(\Gamma_{\rho'} \setminus G_{n,\eta'}^\prime) \lesssim_{q, p_0, n} \eta_0^{-2n} \delta_0^{-2n} (1 + |\omega'|^2)^n + \epsilon_0^{-2n} \eta_0 (1 + |\omega'|^4)^n \lesssim_{q, p_0, n} |\xi|^{n/16}(1 + |\omega'|^4)^n,
\]

and so with the same values of \( n > K \) and \( p_0 \) as before we have that, for any \( K > 0 \), \( p \geq p_0 \) and \( \xi \) sufficiently large,

\[
|J_p(\xi, \omega')| \lesssim_{q, p_0, K} \|\varphi_0\|_{0, K} (1 + |\omega'|^{2p_0}) |\xi|^{-K/16}.
\]

By the same scaling argument for a general Schwartz function \( \varphi_0 \) we have the above result with \( \|\varphi_0\|_{0, K} \) replace by the norm \( \|\varphi_0\|_{(2q+1/2)p_0+K+1, K} \), and the desired bound for \( |I_p(\xi)| \) follows from the relation (3.5).

The function \( I_p(\xi) \) is obviously smooth, and since the phase function \( f_p(\varphi) \) is a cubic polynomial, differentiating \( I_p(\xi) \) by \( k \) times only changes the amplitude \( \varphi(v) \) to \( \varphi(v)P(v) \) where \( P \) is a polynomial of degree at most \( 3k \). This is still a Schwartz function, and by further separating the monomials of \( P \) in \( v_0 \) one can rewrite

\[
\varphi(v)P(v) = \sum_M \varphi_0(v_0)M(v_0)\varphi_1(v')P_M(v'), \tag{3.7}
\]

where the summation is over the monomials \( M \) in \( v_0 \) and each \( P_M \) is a polynomial in \( v' \), both of degree at most \( 3k \). Thus the \( k \)-th derivative \( D^kI_p(\xi) \) consists of oscillatory integrals of the
same type as $I_p(\xi)$ itself. The result then follows from the observation that the number of terms in (3.7) depends only on $q, p_0, k$ and that

$$\max_{|\theta| \leq j, |\sigma| \leq l, |r| \leq m} \sup_{v_0 \in \mathbb{R}^{2p_0}} |v_0^\theta \partial^r (v_0^\sigma \varphi_0(v_0))| \leq \|\varphi_0\|_{j, l, m},$$

for any $j, l, m \in \mathbb{N}$ and multi-indices $\theta, \sigma, \tau \in \mathbb{N}^{2q_0}$.

The same cut-off and scaling argument in the first half of the proof above can be used to estimate the Fourier transform-type integral with a sufficiently smooth amplitude $h$:

$$T(z) = \int_{\mathbb{R}^d} e^{\pm iz \cdot u} h(u; z) du = \int_{B(0, 2)} e^{\pm iz \cdot (u - \omega)} h(u; z) du + \sum_{r \geq 1} \int_{A_0'} e^{\pm iz \cdot (u - \omega)} \chi_r(u; z) du,$$

where $\omega = z/|z|$ and $\chi_r(u; z) = 2^{dz} h(2^r u; z) \zeta_1(2u)/\sigma(2u)$. The phase function $\Psi(u; \omega) = u \cdot \omega$ is linear with gradient bounded from below by $\delta = 1$, and

$$|\chi_r(\cdot; z)|_{K, A_0'} \lesssim_{d, K} \sup_{z \in \mathbb{R}^d} \|h(\cdot; z)\|_{d + K, K}.$$

Apply Lemma 1 on the annulus $A_0'$ for $\delta = 1$ and any $K \in \mathbb{N}$, then each integral on $A_0'$ in the summation above is bounded by

$$C_{d, K} \sup_{z \in \mathbb{R}^d} \|h(\cdot; z)\|_{d + K, K} |z|^{-K} 2^{-rK}.$$

It also follows that in the expression for $T(z)$ above the first integral on the ball $B(0, 2)$ is bounded by $C_{d, K} \sup_z \|h(\cdot; z)\|_{0, K} |z|^{-K}$ for all $z$. Hence we assert the following:

**Lemma 6.** For an arbitrary $K \in \mathbb{N}$ and $h : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ sufficiently smooth w.r.t the first variable s.t. $\sup_z \|h(\cdot; z)\|_{d + K, K} < \infty$, it holds that $\forall z \in \mathbb{R}^d$,

$$\left| \int_{\mathbb{R}^d} e^{\pm iz \cdot u} h(u; z) du \right| \leq C_{d, K} \sup_{z \in \mathbb{R}^d} \|h(\cdot; z)\|_{d + K, K} |z|^{-K}.$$

Returning to Theorem 5 as a special case the characteristic function $\psi_p(\xi)$ of $V_p$ has Gaussian amplitude $\varphi(v) = \phi_p(v)$, which can be factorised as the product of $\varphi_0(v_0) = \phi_{p_0}(v_0)$ and $\varphi_1(v') = \phi_p(v)/\phi_{p_0}(v_0) = \prod_{j=1}^q \prod_{r=p_0+1}^p \phi(x_{jr})\phi(y_{jr})$. Then Theorem 5 immediately implies the following:

**Theorem 7.** The density $f_p$ of $V_p$ has continuous and uniformly bounded derivatives up to order $N$ if $p \geq p_0$, where $p_0 > 2048(N + d)^2$ is an even integer s.t. $[\sqrt{2p_0}/4] + 1$ is prime.

This is an analogue of part (1) of Lemma 11 in [2]; it is not clear whether part (2) of that lemma holds in the triple integral case. But at least we can conclude that the density $f_p$ converges uniformly in $p$ and we have the following:
Corollary 8. The random variable \( V \) has a density with continuous and bounded derivatives up to any order.

By the mean value theorem for \( p > p_0 \) all the derivatives of \( f_p \) up to order \( N \) are Lipschitz. It will be shown in the next section that the random variable \( V_p \) has bounded moments for all \( p \). These facts will be useful to deduce the rapid decay of the density \( f_p \) for large \( p \) due to the following observation:

Lemma 9. Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a Lipschitz function with Lipschitz constant \( L \). If the moments
\[
\int_{\mathbb{R}^d} |x|^M |f(x)| \, dx < \infty \text{ for any } M > 0,
\]
then \( |f(x)| \leq C_{d,L,r} |x|^{-r} \) for any \( r > 0 \) and \( |x| \) sufficiently large.

Proof. If \( f \equiv 0 \) outside a large ball then the claim is trivial. Otherwise take an \( x \notin B(0, |f(0)|/L) \) s.t. \( \alpha := |f(x)| > 0 \). By the Lipschitz condition \( |f(x)| < 2L|x| \). Moreover, for any \( y \in B(x,3\alpha/(4L)) \) one has that
\[
|f(y)| \geq |f(x)| - L|x - y| \geq \alpha/4.
\]
Thus, by assumption for any \( M > 0 \), there is a constant \( C_M \) s.t.
\[
C_M \geq \int_{B(x,\frac{2\alpha}{4L})} |y|^M |f(y)| \, dy \geq C_d \left( \frac{\alpha}{L} \right)^d |x|^M \frac{\alpha}{4} \geq C_{d,L,M} |x|^M \alpha^{d+1},
\]
which in turn implies that \( |f(x)| \leq C_{d,L,M} |x|^{-M/(d+1)} \). For any \( r > 0 \) let \( M = r(d+1) \), and the result follows from the arbitrary choice of \( x \). \( \square \)

4 Moment Estimates and Main Result

Recall the notations \( d = 2q^2 + 2q + (q^3 - q)/3 \) and \( v_p = \{(x_{jr}, y_{ks}) : j,k = 1, \cdots, q, r, s = 1, \cdots, p\} \subset \mathbb{R}^{2qp} \). It is not quite clear yet how the method described in the introduction for the double integral can be applied to the triple integral case. In fact, the expression (1.8) no longer holds as \( g \) is no longer the convolution of \( f_p \) and the law of \( V_p \) - the latter is not independent of \( V_p \). Instead, let \( \kappa_y, \chi_y \) be the densities of \( V_{\tilde{p}} \) and \( \tilde{V}_p \) conditional on that \( V_p = y \), respectively. Then one has that
\[
g(z) = \int_{\mathbb{R}^d} f_p(z - w) \kappa_{z-w}(w) \, dw, \quad h(z) = \int_{\mathbb{R}^d} f_p(z - w) \chi_{z-w}(w) \, dw,
\]
and by (1.7), for all \( z \in \mathbb{R}^d \) one arrives at
\[
|g(z) - h(z)| \leq C_{d,m} \sum_{|\beta| = 0}^{m-1} \left| \int_{\mathbb{R}^d} \left( w^\beta \kappa_{z-w}(w) - w^\beta \chi_{z-w}(w) \right) \, dw \right|
\]
\[
+ C_{d,m} \sum_{|\beta| = m} \int_{\mathbb{R}^d} \left| w^\beta \kappa_{z-w}(w) - w^\beta \chi_{z-w}(w) \right| \, dw.
\]
One then sees the complication of estimating the integrands above, compared to the proof of Theorem 15 in [2]: in the double integral case, due to the independence the first integral above
will just be $E\tilde{U}_p^\beta - E\tilde{U}_p^{\beta}$, which vanishes by assumption, and the rest is of order $O(p^{-m/2})$ by Lemma 10 (or Lemma 11 in [2]). However, here $\int_{\mathbb{R}^d} w^\beta \kappa_{z=w} dw$ is not even the conditional moment of $\tilde{V}_p$ due to the appearance of $w$ in the subscript of $\kappa$. One may apply Taylor’s theorem about $z$ in the subscript and impose certain smoothness condition on $\chi_a$, but whether $\kappa a$ is smooth in $a$ is not clear.

Instead of this approach we follow a somewhat more primitive way of deriving the coupling bound via the Fourier inversion formula, for which the following moment estimate is crucial.

**Lemma 10.** For fixed $p_0 \in \mathbb{Z}^+$ and any $p, N \in \mathbb{Z}^+$, $N > p \geq 2p_0$, recall the notation $v'_N = \{(x_{jr}, y_{jr}) : j = 1, \cdots, q; r = p_0 + 1, \cdots, N\}$ and let $\tilde{V}_{p,N} = V_N - V_p = \tilde{V}_p - \tilde{V}_N$. Then for any $2 \leq m \in \mathbb{Z}^+$ and $\alpha \in \mathbb{N}^d$ s.t. $|\alpha| = m$,

(i) the power $\tilde{V}_{p,N}^\alpha$ is a polynomial of degree at most $m$ in $v_{p_0}$ and at most $3m$ in $v'_{N}$, and the number of monomials involving $v_{p_0}$ depends on $p_0, q$ and $\alpha$ only;

(ii) It holds that $E(\tilde{V}_{p,N}^\alpha | v_{p_0}) \leq C_{q,m} (1 + |v_{p_0}|)^m p^{-m/2}$ uniformly in $N$, where $|v_{p_0}| := \sum_{j=1}^q \sum_{r=1}^{p_0} r^{-1}(|x_{jr}| + |y_{jr}|)$.

**Proof.** One can write $\alpha = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$ s.t. $\sum_{i=1}^6 |\beta_i| = m$ and

$$\tilde{V}_{p,N}^\alpha = \left(\tilde{z}^{(p,N)}\right)^{\beta_1} \left(\tilde{u}^{(p,N)}\right)^{\beta_2} \left(\tilde{\lambda}^{(p,N)}\right)^{\beta_3} \left(\tilde{\mu}^{(p,N)}\right)^{\beta_4} \left(\tilde{\Delta}^{(p,N)}\right)^{\beta_5}$$

where the terms on the right-hand side are similarly defined as the components of $\tilde{V}_p$ and each multi-index $\beta$ is of corresponding dimension. It is then easier to work with powers of each component.

Clearly the contribution of $v_{p_0}$ comes from $\tilde{z}^{(p,N)}$ and $\tilde{\Delta}^{(p,N)}$ only. For each admissible $j, k, l$, the truncated sum $\tilde{V}_{jkl}^{p,N}$ of $\tilde{z}_{jkl}^{(p)}$ over $p < r \leq s \leq N$, $r \neq s$ is at most linear in $v_{p_0}$, and the truncated sum $\tilde{\Delta}_{jkl}^{p,N}$ of $\tilde{\Delta}_{jkl}^{(p)}$ over $p < r + \Delta s \leq N$ is also at most linear in $v_{p_0}$ as the assumption $p \geq 2p_0$ implies that at least one of $r$ and $s$ must be greater than $p_0$. One also sees that the number of occurrences of those monomials in $v_{p_0}$ only depends on $p_0$. Thus the claim (i) follows by multiplying out the powers and counting the highest possible degrees.

The estimate in (ii) can be shown by deriving such a bound for the $m$-th moments of each component of $\tilde{V}_{p,N}$. First of all, it is clear that the random variables $\tilde{z}^{(p,N)}$, $\tilde{u}^{(p,N)}$, $\tilde{\lambda}^{(p,N)}$, $\tilde{\mu}^{(p,N)}$ are independent of $v_p$ (and thus $v_{p_0}$), so their conditional moments given $v_{p_0}$ are just their corresponding ordinary moments. For the random variable $\tilde{u}^{(p,N)} = \tilde{u}^{(p)} - \tilde{u}^{(N)}$, each component $\tilde{u}_j^{(p,N)}$ follows $\mathcal{N}(0, \sum_{r=p+1}^N r^{-1})$ and one sees that $E|\tilde{u}^{(p,N)}|^m \leq C_{q,m} p^{-3m/2}$. For $\tilde{\mu}^{(p,N)}$, notice that each component is an infinite sum of independent random variables. Consider $\tilde{\mu}_{jkl}^{1,p,N} := \tilde{\mu}_{jkl}^{(1,p)} - \tilde{\mu}_{jkl}^{(1,N)}$ for instance: by Rosenthal’s inequality (see Theorem 3 in [2] or
Theorem 2.1 in [3], for any $N > p$,

$$\mathbb{E} \left[ \sum_{r=p+1}^{N} \frac{1}{r^2} x_{jr} x_{kr} \right]^m \lesssim_m \sum_{r=p+1}^{N} \frac{1}{r^m} \mathbb{E}[x_{jr}]^m \mathbb{E}[x_{kr}]^m + \left( \sum_{r=p+1}^{N} \frac{1}{r^3} \mathbb{E}[x_{jr}]^2 \mathbb{E}[x_{kr}]^2 \right)^{m/2} \lesssim_m (p+1)^{-2m} + (p+1)^{-3m/2}.$$

Obviously the same bound holds for the $m$-th moment of $\tilde{\nu}_{2p,N}$, too. One also sees a bound $C_{q,m}p^{-m/2}$ for the $m$-th moments of $\tilde{\nu}_{(p,N)}$ and $\tilde{\lambda}_{(p,N)}$ for the same reason, but this is in fact implied by part (2) of Lemma 11 in [2] where a stronger estimate is given.

It is much less straightforward to compute the conditional moments of $\tilde{\nu}_{(p,N)}$ and $\tilde{\lambda}_{(p,N)}$ as they are not sums of independent random variables. For the rest of the proof use the shorthand notation $\mathbb{E}_{p_0} := \mathbb{E}(\cdot | \nu_{p_0})$. For each pair $(j, k)$ write $\tilde{\nu}_{jk} = A_{jk} - B_{jk}$ where $A_{jk}$ is the corresponding sum of $(r^2 - s^2)^{-1} r s^{-1} x_{jr} x_{ks}$ and $B_{jk}$ of $(r^2 - s^2)^{-1} y_{jr} y_{ks}$. One can further write (see Figure 2a below)

$$A_{jk} = \left( \sum_{\alpha \leq p_0} + \sum_{p < r < s} \sum_{p \leq r < s} + \sum_{p < r < s} + \sum_{p > r < s} \right) \frac{1}{r^2 - s^2} \frac{r}{s} x_{jr} x_{ks} =: T_1 - T_2 + T_3 - T_4,$$

and $B_{jk}$ can be similarly split into four smaller sums. Hence it suffices to bound the $m$-th conditional moment of each of those smaller sums. Moreover, it suffices to consider the case where $m$ is even, as the odd moments can be derived from the Cauchy-Schwarz inequality.

Multiplying out the power one obtains that

$$T_1^m = \sum_{p < r \leq N} \left( \sum_{s \leq p_0} \frac{1}{r} \right) \frac{1}{r_1 \cdots r_m} \frac{r_1 \cdots r_m}{s_1 \cdots s_m} x_{jr_1} \cdots x_{jr_m} x_{ks_1} \cdots x_{ks_m}$$

and similar expressions for $T_2^m, T_3^m$ and $T_4^m$, where the summations are in $s_\alpha, r_\alpha$ accordingly for all $\alpha = 1, \cdots, m$. Note that the random variables $x_{jr_1}, \cdots, x_{jr_m}$ are independent of $x_{ks_1}, \cdots, x_{ks_m}$ as $j < k$. For the conditional expectation not to vanish, the indices $r_1, \cdots, r_m$ must match in pairs; meanwhile since $p > 2p_0$, one has that $r_\alpha > 2s_\alpha$ and $r_\alpha/(r_\alpha - s_\alpha) \leq 2$ for each $\alpha$. Thus

$$\mathbb{E}_{p_0} T_1^m \lesssim_m \left( \sum_{p < r \leq N} \frac{1}{r^2} \right)^{m/2} \left( \sum_{s \leq p_0} \frac{1}{s} \right)^m,$$

and by symmetry $\mathbb{E}_{p_0} T_2^m$ has the same bound with $k$ replaced by $j$.

As for $T_3$ the indices $s_1, \cdots, s_m$ must also match in pairs. If $r_\alpha > 2s_\alpha$ then one immediately obtains a bound $C_{m}p^{-m/2}$; if $r_\alpha \leq 2s_\alpha$, then the corresponding expected sum is bounded by (up to the number of matchings)

$$\left( \sum_{r > p} \sum_{s < r} \frac{1}{(r - s)^2 s^2} \right)^{m/2} \lesssim_m \left( \sum_{r > p} \frac{1}{r^2} \sum_{s < r/2} \frac{1}{s} \right)^{m/2} \lesssim_m p^{-m/2}.$$
Thus $\mathbb{E}_p T_3^m \leq C_m p^{m/2}$, and the same holds for $\mathbb{E}_p T_4^m$ by symmetry. So altogether we have that $\mathbb{E}_p A_{jk}^m \leq C_m (1 + (\sum_{r \leq p_0} r^{-1}|x_{jr}|)^m + (\sum_{r \leq p_0} r^{-1}|x_{kr}|)^m) p^{-m/2}$, and it is easy to see that the same bound with $x$ replaced by $y$ holds for $\mathbb{E}_p B_{jk}^m$.

It remains to estimate the conditional moments of $\tilde{\Delta}_{jkl}^{(p,N)}$ for each Lyndon word $(j, k, l)$. Take, for instance, the sum

$$\Sigma_{jkl} := \sum_{s \leq N} 1 \frac{1}{r_s} x_{jr_1} y_{ls_1} \cdots x_{jr_m} y_{ls_m}$$

for a certain Lyndon word $(j, k, l)$. Write $\Sigma_{jkl} = S_1 + S_2 + S_3$ where, as is illustrated by Figure 2b, $S_1$ is the sum over $p - s < r \leq p, s \leq p$, $S_2$ is the sum over $p < s \leq N - r, r \leq p$, and $S_3$ is the sum over $p < r \leq N - s, s < N - p$. Further write $t = r + s$ for simplicity. Then the $m$-th power of $\Sigma_{jkl}$ can be expressed as

$$\Sigma_{jkl}^m = \sum_{s_s \leq N} \frac{1}{s_1 \cdots s_m} y_{ls_1} \cdots y_{ls_m} \left( \sum_{r_s \leq N} \frac{1}{r_1 \cdots r_m} x_{jr_1} \cdots x_{jr_m} x_{kt_1} \cdots x_{kt_m} \right),$$

and $S_1, S_2$ and $S_3$ can also be written in this form. We also write $r_s$ as the $m$-tuple $(r_1, \cdots, r_m)$ and $s, t$ likewise. Denote by $\Pi_m$ the set of all pair-matching patterns for an $m$-tuple.

Notice that in $S_1$ the random variables $x_{jr_1}, \cdots, x_{jr_m}$ are all independent of $x_{kt_1}, \cdots, x_{kt_m}$. For the conditional expectation not to vanish, the indices $t$ must match in pairs. Thus

$$\mathbb{E}_p S_1^m = \sum_{s_s \leq p} \frac{1}{s_1 \cdots s_m} \mathbb{E}_p y_{ls_1} \cdots y_{ls_m} \left( \sum_{t_s \in \Pi_m} \frac{1}{r_1 \cdots r_m} \mathbb{E}_p x_{jr_1} \cdots x_{jr_m} \right),$$

where the last summation is over $p - s < r_\alpha \leq p$ for all $\alpha = 1, \cdots, m$ subject to a fixed pair-matching pattern of $t_s$, and the constant $C_t$ is the product of the corresponding even moments of $x_{kt_\alpha}$. Note that the indices $r_s$ and $s_s$ cannot be both less than or equal to $p_0$.
as \( p \geq 2p_0\), so they must match in pairs, respectively, too. Hence \( E_{p_0}S_1^m \) is a polynomial in \( x_{j_1}, \ldots, x_{j_{p_0}}, y_{l_1}, \ldots, y_{l_{p_0}} \) of degree \( m \). Distributing out the summation above and using the restriction \( p - p_0 \geq p/2 \) one sees that

\[
E_{p_0}S_1^m \lesssim_m \left( \sum_{p - s < r \leq p, s \leq p_0} \frac{1}{s^{r/2} + y_{l_s}^2} \right)^{m/2} + \left( \sum_{p - s < r \leq p_0, p - p_0 < s \leq p} \frac{1}{s^{r/2} + x_{j_r}^2} \right)^{m/2} + \left( \sum_{p_0 < s \leq p} \frac{1}{s^{r/2}} \right)^{m/2}
\]

\[
\lesssim_m \left( \frac{1}{p} |v_{p_0}|^2 s \right)^{m/2} + \left( \frac{1}{p} \sum_{s \leq p} \frac{1}{s(p - s + 1)} \right)^{m/2}.
\]

The last summation is bounded by \( 2 \sum_{s \leq p} s^{-2} \), and therefore \( E_{p_0}S_1^m \leq C_m (1 + |v_{p_0}|^m) p^{-m/2} \).

For \( S_2 \), the random variables \( x_{j_r} \) and \( x_{k_t} \) are still independent as \( r \leq p \). In addition to \( t \), under conditional expectation the indices \( s \) must match in pairs. This means that the indices \( r \) must also match in pairs, and hence

\[
E_{p_0}S_2^m = \left( \sum_{p < s, r \leq N} \frac{C_{s, r}}{s_1^{r_1} \cdots s_m^{r_m/2}} \right) \left( \sum_{r \leq p_0} \frac{C_{t, r}}{r_1^{r_1} \cdots r_m^{r_m/2}} x_{j_{r_1}}^2 \cdots x_{j_{r_m}}^2 + \sum_{p_0 < r \leq N} \frac{C_{t, r}}{r_1^{r_1} \cdots r_m^{r_m/2}} \right),
\]

where the constants reflect the number of pair-matching patterns for \( s \), \( t \) and \( r \), and the even moments of \( y_{l_s}, x_{k_t}, \) and \( x_{j_r} \). This is a polynomial in \( x_{j_1}, \ldots, x_{j_{p_0}} \) of degree \( m \), and it has the same bound as \( E_{p_0}S_1^m \) uniformly in \( N \).

In \( S_3 \) some index \( r_\alpha \) may match some other \( t_\beta \), thus the we need to consider some more specific cases. Recall that there are only two types of Lyndon words for 3-tuple \((j, k, l)\). For the case where \( j < k \wedge l \), we still have the independence between the random variables \( x_{j_{r_1}}, \ldots, x_{j_{r_m}} \) and \( x_{k_{t_1}}, \ldots, x_{k_{t_m}} \). Thus, after taking conditional expectation, only those with pair-matching indices \( r \) and \( t \), respectively, remain non-vanishing. The indices \( s \) must also match in pairs, and therefore similar to \( S_2 \) one has that

\[
E_{p_0}S_3^m = \left( \sum_{p < r, s \leq N} \frac{C_{r, t}}{r_1^{r_1} \cdots r_m^{r_m/2}} \right) \left( \sum_{s \leq p_0} \frac{1}{s_1^{r_1} \cdots s_m^{r_m/2}} y_{l_{s_1}}^2 \cdots y_{l_{s_m}}^2 + \sum_{p_0 < s, r \leq N} \frac{C_{s, r}}{s_1^{r_1} \cdots s_m^{r_m/2}} \right),
\]

which again leads to the same bound.

It is more intricate to deal with the case where \( j = k < l \), as the independence between \( x_{j_{r_1}}, \ldots, x_{j_{r_m}} \) and \( x_{j_{t_1}}, \ldots, x_{j_{t_m}} \) is no longer true. For the conditional expectation not to vanish, the 2m-tuple \( \tau := (r_1, \ldots, r_m, t_1, \ldots, t_m) \) must match in pairs. So it suffices to consider the following terms in \( S_3 \):

\[
\sum_{s \leq N} \frac{1}{s_1 \cdots s_m} y_{l_{s_1}} \cdots y_{l_{s_m}} \left( \sum_{\tau \in \Pi_{2m}} \sum_{r} \frac{1}{r_1 \cdots r_m} x_{j_{r_1}} \cdots x_{j_{r_m}} x_{j_{t_1}} \cdots x_{j_{t_m}} \right),
\]

where the last summation is over \( p < r \leq N \) subject to a pair-matching pattern of \( \tau \). For fixed values of \( s_1, \ldots, s_m \) and a pattern \( \tau \), define \( \alpha \) and \( \beta \) as equivalent on \( \{1, \ldots, m\} \) if \( r_\alpha
Another observation from (ii) is that, if one sets

\[ n_r = \text{min}\{r_\alpha : \alpha \in E\} \]

and \( t_\alpha \) is equal to \( r_\beta \) or \( t_\beta \). Consider the equivalence relation generated by this relation. If \( \alpha \) and \( \beta \) are in an equivalent class \( E \subset \{1, \ldots, m\} \), then the difference \( r_\alpha - r_\beta \) is determined by the fixed choice of \( s_\alpha, s_\beta \) and the matching constraint of \( r_\alpha, t_\alpha, r_\beta, t_\beta \). In effect, for any \( \alpha \in E \) the value of \( r_\alpha \) determines the values of \( r_\beta \) for all \( \beta \in E \). Thus, one can choose \( r_\alpha = \text{min}\{r_\alpha : \alpha \in E\} \) and rewrite the last summation above as

\[
\sum_r \frac{1}{r_1 \cdots r_m} x_{jr_1} \cdots x_{jr_m} x_{kt_1} \cdots x_{kt_m} = \prod_{E} \sum_{p < r_\alpha \in N} \frac{1}{r_\alpha} x_{jr_\alpha} \cdots x_{jr_{|E|}} x_{jt_\alpha} \cdots x_{jt_{|E|}}
\]

where the product is taken over the equivalent class partition of the set \( \{1, \ldots, m\} \). Then the expectation of the terms in the big parentheses in (4.1) is bounded by

\[
C_m \prod_{E} \sum_{r_\alpha > p} r_\alpha^{-|E|} \leq C_m \prod_{E} p^{-|E|} = C_m p^m - m,
\]

where \( n_m \) is the number of equivalent classes, which is at most \( m/2 \), giving an upper bound \( C_m p^m - m/2 \) for the quantity above. Thus \( E_m p_0 S_3^m \) is a polynomial in \( y_{l_1}, \ldots, y_{l_p} \) of degree \( m \), and one has that

\[
E_m p_0 S_3^m \leq m p^{-m/2} \left( \sum_{s \in \mathbb{N}} \frac{1}{s} y_{l_s} \right)^m + p^{-m/2} \left( \sum_{p_0 < s \in \mathbb{N}} \frac{1}{s^2} \right)^{m/2},
\]

which again gives the same bound as previous cases.

It then follows from the triangle inequality that \( E \sum_{jkl}^m \leq C_m p^{-m/2} (1 + |v_p|^m) \). The same arguments apply to all other terms in \( \Delta^{(p)} \) (as for the terms \( y_{jr} y_{ls} y_{kr}, \beta \) the indices \( j, k, l \) are never all the same), and the result is proven.

As a side note, if one removes the restriction \( p \geq 2p_0 \) and takes \( p_0 = p \) instead of a fixed value, it is not hard to see from the proof above that the conditional moment \( E_p \nu_{p,N}^{n} \) is a polynomial in \( v_p \) of degree at most \( 2m \), and so is \( E_p \nu_{p,N}^{n} \) due to the uniformity in \( N \). In addition, one has that \( E_p \nu_{p,N}^{n} \leq q_m (1 + |v_p|^m) p^{-m/2} + |v_p|^{2m} \), with the extra contribution coming from the fact that \( E_p S_1^m = S_1^m \).

Taking expectation again and noticing the pair-matching requirement one immediately sees that \( E \nu_{p,N}^{n} \leq C_q m p^{-m/2} \).

Another observation from (ii) is that, if one sets \( p_0 = 1, \ p = 2 \), then taking expectation again one sees the following from the triangle inequality (obviously \( V_2 \) has bounded moments):

**Corollary 11.** For any integer \( m \geq 2 \) there exists a constant \( C_q,m \) s.t. \( \sup_{p \geq 1} E \nu_{p,N}^{n} \leq C_q,m \).

Then Lemma 10 is validated by Corollary 11 and Theorem 7 giving \( |f_{p}(y)| \leq C_{d,p_0,r} |y|^{-r} \) for any \( r > 0, \ p > p_0 \) and \( y \in \mathbb{R}^d \) sufficiently far away from 0. Moreover, the argument for the interpolation (1.9) also applies here, with \( m \) replace by any \( k \geq 0 \) and the decay rate \( C_{q,m} e^{-\epsilon_k |y|} \) replaced by \( C_{d,p_0,k} |y|^{-r} \) for the derivative \( D^k f_{p} \).
Corollary 12. For any $N \geq 1$ let $p_0$ be defined as in Theorem 7. Then for all $p \geq 2p_0$, all the derivatives of $f_p$ up to order $N$ have rapid decay.

We are now finally ready to proceed to the coupling result, by which we wish to characterise a candidate random variable $\bar{V}_p$ s.t. the distance $W_2(V, \bar{V}_p + \bar{V}_p)$ is small.

Theorem 13. Let $2 \leq m \in \mathbb{Z}^+$ and $p_0 > 2048(m+3d+3)^2$ be an even integer s.t. $[\sqrt{2p_0/4}] + 1$ is prime. For any $p \geq 2p_0$, suppose there exists an $\mathbb{R}^d$-random variable $\bar{V}_p$ having the same conditional moments given $v_p$ as those of $V_p$ up to order $m - 1$ and having density $\chi_y$ conditional on that $V_p = y$. If the function $y \mapsto \chi_y(w)$ is at least $C^{m+2d+3}(\mathbb{R}^d)$ and for $n,k \geq 0$ there is a constant $M(n,k) \geq 0$ s.t.

$$\int_{\mathbb{R}^d} |w|^n |D_y^k \chi_y(w)| dw \leq C_{d,n,k}(1 + |y|^{M(n,k)}) p^{-n/2},$$

then there exists a constant $C$ depending on $d, m, p_0$ and the density $f_p$ of $V_p$ s.t. $\forall p \geq 2p_0$,

$$W_2(V, V_p + \bar{V}_p) \leq C p^{-m/4}.$$

Proof. Let $g$ and $h$ be the densities of $V$ and $\bar{V} := V_p + \bar{V}_p$, respectively. Then by the inversion formula and the tower property, for all $z \in \mathbb{R}^d$,

$$g(z) - h(z) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iz\cdot\xi} \left( E e^{i\xi\cdot V} - E e^{i\xi\cdot \bar{V}} \right) d\xi$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iz\cdot\xi} \left( e^{i\xi\cdot V_p} E_p \left( e^{i\xi\cdot \bar{V}_p} \right) - e^{i\xi\cdot \bar{V}_p} \right) d\xi,$$

where $E_p := E(\cdot|v_p)$. Applying Taylor’s theorem to $\exp(i\xi \cdot \bar{V}_p)$ and $\exp(i\xi \cdot \bar{V})$ inside the conditional expectation up to order $m$, one sees that the first $m - 1$ differences vanish due to the moment matching assumption and hence

$$E_p \left( e^{i\xi\cdot \bar{V}_p} - e^{i\xi\cdot \bar{V}} \right) = i^m \sum_{|\alpha| = m} \frac{m!}{\alpha!} \xi^\alpha \int_0^1 (1 - \theta)^{m-1} E_p \left( \bar{V}_p^\alpha e^{i\theta \xi \bar{V}_p} - \bar{V}_p^\alpha e^{i\theta \xi \bar{V}} \right) d\theta. $$

Thus we have the identity

$$g(z) - h(z) = \frac{1}{(2\pi)^d} \sum_{|\alpha| = m} i^m \frac{m!}{\alpha!} \int_0^1 (1 - \theta)^{m-1} \int_{\mathbb{R}^d} e^{-iz\cdot\xi} \left( \rho(\xi) - \eta(\xi) \right) d\xi d\theta,$$

where

$$\rho(\xi) = \rho(\xi; p, \alpha, \theta) := E \left( e^{i\xi\cdot (V_p + \theta \bar{V}_p)} \bar{V}_p^\alpha \right),$$

and $\eta(\xi)$ is similarly defined by replacing $\bar{V}_p$ with $\bar{V}$. The goal is then to show the rapid decay in $|z|$ of the $d\xi$ integral above with an appropriate rate of decay in $p$. This should follow from the rapid decay in $|\xi|$ of the derivatives of $\rho(\xi)$ and $\eta(\xi)$.

To this end it is easier to work with, instead of $\bar{V}_p$, the ‘truncated remainder’ $\bar{V}_{p,N} := V_N - V_p$ up to some integer $N \gg p$. Replace $\bar{V}_p$ with $\bar{V}_{p,N}$ in $\rho(\xi)$ and denote it by $\rho_N(\xi)$. Recall the
notations \( v'_N = \{(x_{jr}, y_{jr}) : j = 1, \ldots, q, r = p_0 + 1, \ldots, N'\} \) and \( N' = N - p_0 \). As \( p \geq 2p_0 \), according to Lemma 11 (i) the power \( \tilde{V}^\alpha_{p,N} \), as a function of \( v_{p_0} \) and \( v'_N \), takes the form

\[
\tilde{V}^\alpha_{p,N} = \sum_M M(v_{p_0})P_M(v'_N),
\]

where the summation is over the monomials \( M \) of degree at most \( m \) and each \( P_M \) is a polynomial of degree at most \( 3m \). Similar to the decomposition (3.7), the number of summands depends only on \( p_0 \) and \( \alpha \). By the tower property again,

\[
\rho_N(\xi) := \mathbb{E}\left(e^{i\xi(\nu + \theta \tilde{V}^\alpha_{p,N})} \tilde{V}^\alpha_{p,N}\right) = \mathbb{E}\sum_M P_M(v'_N)\mathbb{E}\left(e^{i\xi(\nu + \theta \tilde{V}^\alpha_{p,N})} M(v_{p_0})\right) v'_N
\]

\[
= \sum_M \int_{\mathbb{R}^{2qN'}} P_M(\nu') \frac{\phi_N(\nu_0, \nu')}{\phi_{p_0}(\nu_0)} d\nu' \int_{\mathbb{R}^{2q\nu_0}} e^{i\xi|\Xi_N(v_{p_0}, \nu'; \theta, \omega)} M(v_{p_0})\phi_{p_0}(v_0) dv_0,
\]

where \( \omega = \xi/|\xi| \) and the phase function \( \Xi_N(v_0, \nu'; \theta, \omega) \) is defined to be the right-hand side terms in (3.7) (\( p \) replaced by \( N \), with the first term \( \Phi_p(v_0; \omega) \) fixed but the rest having dilated frequency \( \theta \omega \).

Note that the functions \( \phi_0(v_0) := M(v_0)\phi_{p_0}(v_0) \) and \( \varphi_1(\nu') := P_M(\nu')\phi_N(v_0, \nu')/\phi_{p_0}(v_0) \) are both Schwartz, and for any \( k \in \mathbb{N} \), \( \|\phi_0\|_{(2q+1)/2p_0+m+1+3k,m} \leq C_{d, p_0, k} \) and

\[
\int_{\mathbb{R}^{2qN'}} (1 + |\nu'|^{2p_0+3k})\varphi_1(\nu') d\nu' = \mathbb{E}|(1 + |\nu'|^{2p_0+3k})P_M(\nu'_N)| \leq C_{q, p_0, m, kp^{-m/2}}
\]

uniformly in \( N \) by Lemma 11 (ii) and the Cauchy-Schwartz inequality. Moreover, for fixed \( \theta \) and \( \nu' \) the function \( \Xi_N(v_0, \nu'; \theta, \omega) \) only differs from \( \Phi_{p_0}(v_0; \omega) \) by a quadratic polynomial in \( v_0 \) with no singularity in \( \theta \), which is insignificant according to the last part of the proof of Theorem 5. Thus, by Theorem 5 (for \( K = m + 2d + 3 \)) the integral \( \rho_N(\xi) \) is a smooth function s.t. for all \( k \geq 0, N \gg p, \theta \in (0, 1) \) and \( |\xi| \) sufficiently large,

\[
|D^k \rho_N(\xi)| \leq C_{d, p_0, k, |\xi|}^{-m-2d-3} p^{-m/2}. \tag{4.2}
\]

Thereby for all \( N, \theta \) the function \( G_N(\xi) := \xi^\alpha \rho_N(\xi) \) has uniformly bounded derivatives, and in particular \( \|G_N\|_{2d+3, d+3} \leq \|\rho_N\|_{m+2d+3, d+3} \leq C_{d, p_0, m} p^{-m/2} \). Then applying Lemma 6 with \( k = K = d + 3 \) one deduces that

\[
\int_{\mathbb{R}^d} e^{-i\xi \cdot \xi^\alpha\rho_N(\xi)} d\xi \leq C_{d, p_0, m} |z|^{-d-3} p^{-m/2}. \tag{4.3}
\]

This estimate is uniform in \( N \), and therefore by taking the limit \( N \to \infty \) the same bound holds for the integral \( \int_{\mathbb{R}^d} e^{-i\xi \cdot \xi^\alpha\rho(\xi)} d\xi \).

For the other integral \( \int_{\mathbb{R}^d} e^{-i\xi \cdot \xi^\alpha\eta(\xi)} d\xi \), from the same arguments above it suffices to show that \( \forall \xi \leq d + 3, N \gg p, \theta \in (0, 1) \) and \( |\xi| \) sufficiently large the estimate (4.2) also holds for \( \eta \). By conditioning on the value of \( V_p \) one finds the identity

\[
\eta(\xi) = e^{i\xi \cdot V_p} \mathbb{E}\left(e^{i\theta \xi \cdot V_p \tilde{V}^\alpha_p} |V_p\right)
\]

\[
= \int_{\mathbb{R}^d} e^{i\xi \cdot y} f_p(y) \int_{\mathbb{R}^d} e^{i\theta \xi \cdot w} \chi_y(w) dw dy.
\]

32
Differentiating $\eta(\xi)$ by $k$ times by Leibniz’s rule, one has that $\forall \tau \in \mathbb{N}^d$, $|\tau| = k$,

$$
\partial^\tau \eta(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot y} f_p(y) \sum_{\sigma \leq \tau} \binom{\tau}{\sigma} (iy)^{\tau - \sigma} \partial_\xi^\sigma \left( \int_{\mathbb{R}^d} e^{i\theta \xi \cdot w} \chi_y(w) dw \right) dy
$$

$$
= i^k \sum_{\sigma \leq \tau} \binom{\tau}{\sigma} \theta^{|\sigma|} \int_{\mathbb{R}^d} e^{i\xi \cdot y} y^{\tau - \sigma} f_p(y) \int_{\mathbb{R}^d} e^{i\theta \xi \cdot w} w^{\sigma + \alpha} \chi_y(w) dw dy.
$$

Then by Leibniz’s rule again for any $\beta \in \mathbb{N}^d$, $|\beta| \leq m + 2d + 3$,

$$
\partial^\beta_y H_{\tau,\sigma}(y;\xi,\theta) = \sum_{\nu \leq \beta} \binom{\beta}{\nu} \partial^\nu_y \left( y^{\tau - \sigma} f_p(y) \right) \int_{\mathbb{R}^d} e^{i\theta \xi \cdot w} w^{\beta - \nu} \partial_\chi_y \chi_y(w) dw,
$$

which is bounded in $\xi$ and $\theta$. Thus, by the assumption on the conditional law of $\bar{V}_p$ and Corollary 12 one sees that

$$
\sup_{\xi \in \mathbb{R}^d, \theta \in (0,1)} \|H_{\tau,\sigma}(\cdot;\xi,\theta)\|_m \leq C_{m,d,k,p} \|f_p\|_{k+m+3d+3+M(m+k,m+2d+3,m+2d+3)}
$$

for all $\theta \in (0,1)$ and $\xi \in \mathbb{R}^d$. Thus the sought-after estimate (4.2) for $|D^k \eta(\xi)|$ follows from Lemma 6 for $K = m + 2d + 3$, with the norm of $f_p$ above as a multiplicative factor. Apply Lemma 6 again for $k = K = d + 3$ we obtain (4.3) with $\rho_N$ replaced by $\eta$ and a multiplicative factor $\|f_p\|_{m+4d+6+M(m+d+3,m+2d+3,m+2d+3)}$.

The result then follows from the inequality (1.4) for $p = 2$.

To finish off this section we remark that, as opposed to the rate $O(p^{-m/2})$ obtained in [2] for the double integral, the rate $O(p^{-m/4})$ is probably the best one can expect simply from Theorem 5 and Theorem 7 alone. This is because the particular form of the phase function $\Phi_p$ and its derivatives are not fully exploited. In fact we have only used the fact that the phase function $\Phi_p$ is a cubic polynomial in Lemma 3 and Lemma 4.

Despite this limitation, to my best knowledge what is proved so far is the first attempt to find a coupling for triple stochastic integrals. I believe that Davie’s rate $O(p^{-m/2})$ could be achieved if analogues of Lemma 12, 13 and especially 14 in [2] can be proven, but the question remains open.

Acknowledgement. This work was completed under the patient guidance of my Ph.D. advisor, Prof. Alexander M. Davie, who suggested the problem and gave many crucial advices on the main steps of the arguments, as well as technical details.

References

[1] Clark, J. M. C., and Cameron, R. J. The maximum rate of convergence of discrete approximations for stochastic differential equations. *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius)* (1978), 162–171.
[2] Davie, A. M. Pathwise approximation of stochastic differential equations using coupling. *Preprint* (2014), www.maths.ed.ac.uk/~sandy/coum.pdf.

[3] Davie, A. M. Polynomial perturbations of normal distributions. *Preprint* (2016), www.maths.ed.ac.uk/~sandy/polg.pdf.

[4] de Acosta, A. Inequalities for $B$-valued random vectors with applications to the strong law of large numbers. *The Annals of Probability* 9, 1 (1981), 157–161.

[5] Gaines, J. G. The algebra of iterated stochastic integrals. *Stochastics and Stochastics Reports* 49 (1994), 169–179.

[6] Hörmander, L. *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*, 2nd ed. Springer-Verlag, 1990.

[7] Kloeden, P. E., and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, 1995.

[8] Lyons, T. J., and Gaines, J. G. Random generation of stochastic area integrals. *SIAM Journal on Applied Mathematics* 54, 4 (1994), 1132–1146.

[9] Rosenthal, H. P. On the subspaces of $l^p (p > 2)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics* 8 (1970), 273–303.

[10] Rudin, W. *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill Inc., 1976.

[11] Sogge, C. D. *Fourier Integrals in Classical Analysis*. Cambridge Tracts in Mathematics. Cambridge University Press, 1993.

[12] Villani, C. *Topics in Optimal Transportation*, vol. 58. American Mathematical Society, 2003.

[13] Wiktorsson, M. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Annals of Applied Probability* 11 (2001), 470–487.