Scientific
Research
Publishing

Gradient Density Estimation in Arbitrary Finite Dimensions Using the Method of Stationary Phase

Karthik S. Gurumoorthy[1], Anand Rangarajan[2*], John Corring[3]

[1]International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru, Karnataka, India
[2]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, Florida, USA
[3]Microsoft Research, Seattle, Washington, USA
Email: karthik.gurumoorthy@icts.res.in, *anand@cise.ufl.edu, jocorrin@microsoft.com

Abstract

We prove that the density function of the gradient of a sufficiently smooth function $S : \Omega \subset \mathbb{R}^d \to \mathbb{R}$, obtained via a random variable transformation of a uniformly distributed random variable, is increasingly closely approximated by the normalized power spectrum of $\phi = \exp\left(\dfrac{iS}{\tau}\right)$ as the free parameter $\tau \to 0$. The frequencies act as gradient histogram bins. The result is shown using the stationary phase approximation and standard integration techniques and requires proper ordering of limits. We highlight a relationship with the well-known characteristic function approach to density estimation, and detail why our result is distinct from this method. Our framework for computing the joint density of gradients is extremely fast and straightforward to implement requiring a single Fourier transform operation without explicitly computing the gradients.

Keywords

Stationary Phase Approximation, Density Estimation, Fourier Transform, Wave Functions, Characteristic Functions

1. Introduction

Density estimation methods provide a faithful estimate of a non-observable probability density function based on a given collection of observed data [1] [2] [3] [4]. The observed data are treated as random samples obtained from a large population which is assumed to be distributed according to the underlying den-

sity function. The aim of our current work is to show that the joint density function of the gradient of a sufficiently smooth function S (density function of ∇S) can be obtained from the normalized power spectrum of $\phi = \exp\left(\dfrac{iS}{\tau}\right)$ as the free parameter τ tends to zero. The proof of this relationship relies on the higher order stationary phase approximation [5]-[10]. The joint density function of the gradient vector field is usually obtained via a random variable transformation, of a uniformly distributed random variable X over the compact domain $\Omega \subset \mathbb{R}^d$, using ∇S as the transformation function. In other words, if we define a random variable $Y = \nabla S(X)$ where the random variable X has a *uniform distribution* on the domain Ω ($X \sim UNI(\Omega)$), the density function of Y represents the joint density function of the gradient of S.

In computer vision parlance—a popular application area for density estimation—these gradient density functions are popularly known as the histogram of oriented gradients (HOG) and are primarily employed for human and object detection [11] [12]. The approaches developed in [13] [14] demonstrate an application of HOG—in combination with support vector machines [15]—for detecting pedestrians from infrared images. In a recent article [16], an adaption of the HOG descriptor called the Gradient Field HOG (GF-HOG) is used for sketch-based image retrieval. In these systems, the image intensity is treated as a function $S(X)$ over a $2D$ domain, and the distribution of intensity gradients or edge directions is used as the feature descriptor to characterize the object appearance or shape within an image. In Section 5 we provide experimental evidence to showcase the efficacy of our method in computing the density of these oriented gradients (HOG). The present work has also been influenced by recent work on quantum supremacy [17] [18] [19]. Here, the aim is to draw samples from the density function of random variables corresponding to the measurement bases of a high-dimensional quantum mechanical wave function. This work may initially seem far removed from our efforts. However, as we will show, the core of our density estimation approach is based on evaluating interval measures of the squared magnitude of a wave function in the frequency domain. For this reason, our approach is deemed a wave function approach to density estimation and henceforth we refer to it as such.

In our earlier effort [20], we primarily focused on exploiting the stationary phase approximation to obtain gradient densities of Euclidean distance functions (R) in two dimensions. As the gradient norm of R is identically equal to 1 almost everywhere, the density of the gradient is one-dimensional and defined over the space of orientations. The key point to be noted here is that the dimensionality of the gradient density (one) is one less than the dimensionality of the space (two) and the constancy of the gradient magnitude of R causes its Hessian to vanish almost everywhere. In Lemma 2.3 below, we see that the Hessian is deeply connected to the density function of the gradient. The degeneracy of the Hessian precluded us from directly employing the stationary phase method and hence techniques like symmetry-breaking had to be used to circumvent the va-

nishing Hessian problem. The reader may refer to [20] for a more detailed explanation. In contrast to our previous work, we regard our current effort as a generalization of the gradient density estimation result, now established for *arbitrary* smooth functions in *arbitrary* finite dimensions.

1.1. Main Contribution

We introduce a new approach for computing the density of Y, where we express the given function S as the phase of a wave function ϕ, specifically

$\phi(x) = \exp\left(\dfrac{iS(x)}{\tau}\right)$ for small values of τ, and then consider the normalized

power spectrum—squared magnitude of the Fourier transform—of ϕ [21]. We show that the computation of the joint density function of $Y = \nabla S$ may be approximated by the power spectrum of ϕ, with the approximation becoming increasingly tight point-wise as $\tau \to 0$. Using the stationary phase approximation, a well known technique in asymptotic analysis [9], we show that in the limiting case as $\tau \to 0$, the power spectrum of ϕ converges to the density of Y, and hence can serve as its density estimator at small, non-zero values of τ. In other words, if $P(u)$ denotes the density of Y, and if $P_\tau(u)$ corresponds to the power spectrum of ϕ at a given value of τ, Theorem 3.2 constitutes the following relation,

$$\lim_{\tau \to 0} \int_{\mathcal{N}_\eta(u_0)} P_\tau(u)\,\mathrm{d}u = \int_{\mathcal{N}_\eta(u_0)} P(u)\,\mathrm{d}u$$

where $\overline{\mathcal{N}_\eta(u_0)}$ is a small neighborhood around u_0. We would like to emphasize that our work is fundamentally different from estimating the gradient of a density function [22] and should not be semantically confused with it.

1.2. Significance of Our Result

As mentioned before, the main objective of our current work is to generalize our effort in [20] and demonstrate the fact that the wave function method for obtaining densities can be extended to arbitrary functions in arbitrary finite dimensions. However, one might broach a legitimate question, namely *"What is the primary advantage of this approach over other simpler, effective and traditional techniques like histograms which can compute the HOG say by mildly smoothing the image, computing the gradient via (for example) finite differences and then binning the resulting gradients?"*. The benefits are three fold:

- One of the foremost advantages of our wave function approach is that it recovers the joint gradient density function of S *without* explicitly computing its gradient. Since the stationary points capture gradient information and map them into the corresponding frequency bins, we can directly work with S without the need to compute its derivatives.
- The significance of our work is highlighted when we deal with the more practical finite sample-set setting wherein the gradient density is estimated from a *finite, discrete* set of samples of S rather than assuming the availability

of the complete description of S on Ω. Given the N samples of S on Ω, it is customary to know the *approximation error* of a proposed density estimation method as $N \to \infty$. In [23] we prove that in one dimension, the point-wise approximation error between our wave function method and the true density is bounded above by $O(1/N)$ when $\tau \propto 1/N$. For histograms and kernel density estimators [1] [2], the approximation errors are established for the integrated mean squared error (IMSE) defined as the expected value (with respect to samples of size N) of the square of the ℓ_2 error between the true and the computed probability densities and are shown to be $O\left(N^{-\frac{2}{3}}\right)$ [24] [25] and $O\left(N^{-\frac{4}{5}}\right)$ [26] respectively. Having laid the foundation in this work, we plan to invest our future efforts in pursuit of similar upper bounds in arbitrary finite dimensions.

- Furthermore, obtaining the gradient density using our framework in the finite N sample setting is simple, efficient, and computable in $O(N \log N)$ time as elucidated in the last paragraph of Section 4.

1.3. Motivation from Quantum Mechanics

Our wave function method is motivated by the classical-quantum relation, wherein classical physics is expressed as a limiting case of quantum mechanics [27] [28]. When S is treated as the Hamilton-Jacobi scalar field, the gradients of S correspond to the classical momentum of a particle [29]. In the parlance of quantum mechanics, the squared magnitude of the wave function expressed either in its position or momentum basis corresponds to its position or momentum density respectively. Since these representations (either in the position or momentum basis) are simply (suitably scaled) Fourier transforms of each other, the squared magnitude of the Fourier transform of the wave function expressed in its position basis is its quantum momentum density. However, the time independent Schrödinger wave function $\phi(\boldsymbol{x})$ (expressed in its position basis) can be approximated by $\exp\left(\frac{iS(\boldsymbol{x})}{\tau}\right)$ as $\tau \to 0$ [28]. Here τ (treated as a free parameter in our work) represents Planck's constant. Hence the squared magnitude of the Fourier transform of $\exp\left(\frac{iS(\boldsymbol{x})}{\tau}\right)$ corresponds to the quantum momentum density of S. The principal results proved in the article state that the classical momentum density (denoted by P) can be expressed as a limiting case (as $\tau \to 0$) of its corresponding quantum momentum density (denoted by P_τ), in agreement with the correspondence principle.

2. Existence of Joint Densities of Smooth Function Gradients

We begin with a compact measurable subset Ω of $\mathbb{R}^d$ on which we consider a smooth function $S : \Omega \to \mathbb{R}$. We assume that the boundary of Ω is smooth and the function S is well-behaved on the boundary as elucidated in Appendix B.

Let $\mathcal{H}_x$ denote the *Hessian* of S at a location $x \in \Omega$ and let $\det(\mathcal{H}_x)$ denote its determinant. The *signature* of the Hessian of S at x, defined as the difference between the number of positive and negative eigenvalues of $\mathcal{H}_x$, is represented by σ_x. In order to exactly determine the set of locations where the joint density of the gradient of S exists, consider the following three sets:

$$\mathcal{A}_u = \{x : \nabla S(x) = u\}, \tag{2.1}$$

$$\mathcal{B} = \{x : \det(\mathcal{H}_x) = 0\}, \tag{2.2}$$

and

$$\mathcal{C} = \{\nabla S(x) : x \in \mathcal{B} \cup \partial \Omega\}. \tag{2.3}$$

Let $N(u) = |\mathcal{A}_u|$. We employ a number of useful lemma, stated here and proved in Appendix A.

Lemma 2.1. [*Finiteness Lemma*] $\mathcal{A}_u$ *is finite for every* $u \notin \mathcal{C}$.

As we see from Lemma 2.1 above, for a given $u \notin \mathcal{C}$, there is only a *finite* collection of $x \in \Omega$ that maps to u under the function ∇S. The inverse map $\nabla S^{(-1)}(u)$ which identifies the set of $x \in \Omega$ that maps to u under ∇S is ill-defined as a function as it is a one to many mapping. The objective of the following lemma (Lemma 2.2) is to define appropriate neighborhoods such that the inverse function $\nabla S^{(-1)}$, required in the proof of our main Theorem 3.2, when restricted to those neighborhoods is well-defined.

Lemma 2.2. [*Neighborhood Lemma*] *For every* $u_0 \notin \mathcal{C}$, *there exists a closed neighborhood* $\overline{\mathcal{N}_\eta(u_0)}$ *around* u_0 *such that* $\overline{\mathcal{N}_\eta(u_0)} \cap \mathcal{C}$ *is empty. Furthermore, if* $|\mathcal{A}_{u_0}| > 0$, $\overline{\mathcal{N}_\eta(u_0)}$ *can be chosen such that we can find a closed neighborhood* $\overline{\mathcal{N}_\eta(x)}$ *around each* $x \in \mathcal{A}_{u_0}$ *satisfying the following conditions:*

1) $\nabla S\left(\overline{\mathcal{N}_\eta(x)}\right) = \overline{\mathcal{N}_\eta(u_0)}$.
2) $\det(\mathcal{H}_y) \neq 0, \forall y \in \overline{\mathcal{N}_\eta(x)}$.
3) *The inverse function* $\nabla S_x^{(-1)}(u) : \overline{\mathcal{N}_\eta(u_0)} \to \overline{\mathcal{N}_\eta(x)}$ *is well-defined.*
4) *For* $y, z \in \overline{\mathcal{N}_\eta(x)}, \sigma_y = \sigma_z$.

Lemma 2.3 [*Density Lemma*] *Given* $X \sim UNI(\Omega)$, *the probability density of* $Y = \nabla S(X)$ *on* $\mathbb{R}^d - \mathcal{C}$ *is given by*

$$P(u) = \frac{1}{\mu(\Omega)} \sum_{k=1}^{N(u)} \frac{1}{\left|\det(\mathcal{H}_{x_k})\right|} \tag{2.4}$$

where $x_k \in \mathcal{A}_u, \forall k \in \{1, 2, \cdots, N(u)\}$ *and* μ *is the Lebesgue measure.*

From Lemma 2.3, it is clear that the existence of the density function P at a location $u \in \mathbb{R}^d$ necessitates a non-vanishing Hessian matrix ($\det(\mathcal{H}) \neq 0$) $\forall x \in \mathcal{A}_u$. Since we are interested in the case where the density exists almost everywhere on $\mathbb{R}^d$, we impose the constraint that the set $\mathcal{B}$ in (2.2), comprising all points where the Hessian vanishes, has zero Lebesgue measure. It follows that $\mu(\mathcal{C}) = 0$. Furthermore, the requirement regarding the smoothness of S

$(S \in C^\infty(\Omega))$ can be relaxed to functions S in $C^{\frac{d}{2}+1}(\Omega)$ where d is the dimensionality of Ω as we will see in Section 3.2.2.

3. Equivalence of the Densities of Gradients and the Power Spectrum

Define the function $F_\tau : \mathbb{R}^d \to \mathbb{C}$ as

$$F_\tau(\boldsymbol{u}) = \frac{1}{(2\pi\tau)^{\frac{d}{2}} \mu(\Omega)^{\frac{1}{2}}} \int_\Omega \exp\left(\frac{i}{\tau}\left[S(\boldsymbol{x}) - \boldsymbol{u} \cdot \boldsymbol{x}\right]\right) d\boldsymbol{x} \tag{3.1}$$

or $\tau > 0$. F_τ is very similar to the Fourier transform of the function $\exp\left(\dfrac{iS(\boldsymbol{x})}{\tau}\right)$. The normalizing factor in F_τ comes from the following lemma (Lemma 3.1) whose proof is given in Appendix A.

Lemma 3.1. [*Integral Lemma*] $F_\tau \in L^2(\mathbb{R}^d)$ *and* $\|F_\tau\|_2 = 1$.

The power spectrum defined as

$$P_\tau(\boldsymbol{u}) \equiv F_\tau(\boldsymbol{u})\overline{F_\tau(\boldsymbol{u})} \tag{3.2}$$

equals the squared magnitude of the Fourier transform. Note that $P_\tau \geq 0$. From Lemma (3.1), we see that $\int P_\tau(\boldsymbol{u}) d\boldsymbol{u} = 1$. Our fundamental contribution lies in interpreting $P_\tau(\boldsymbol{u})$ as a density function and showing its equivalence to the density function $P(\boldsymbol{u})$ defined in (2.4). Formally stated:

Theorem 3.2. *For* $\boldsymbol{u}_0 \notin \mathcal{C}$,

$$\lim_{\alpha \to 0} \frac{1}{\mu(\mathcal{N}_\alpha(\boldsymbol{u}_0))} \lim_{\tau \to 0} \int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)} P_\tau(\boldsymbol{u}) d\boldsymbol{u} = P(\boldsymbol{u}_0)$$

where $\mathcal{N}_\alpha(\boldsymbol{u}_0)$ *is a ball around* $\boldsymbol{u}_0$ *of radius* α.

Before embarking on the proof, we would like to emphasize that the ordering of the limits and the integral as given in the theorem statement is crucial and cannot be arbitrarily interchanged. To press this point home, we show below that after solving for P_τ, the $\lim_{\tau \to 0} P_\tau$ does not exist. Hence, the order of the integral followed by the limit $\tau \to 0$ cannot be interchanged. Furthermore, when we swap the limits of α and τ, we get

$$\lim_{\tau \to 0} \lim_{\alpha \to 0} \frac{1}{\mu(\mathcal{N}_\alpha(\boldsymbol{u}_0))} \int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)} P_\tau(\boldsymbol{u}) d\boldsymbol{u} = \lim_{\tau \to 0} P_\tau(\boldsymbol{u}_0)$$

which also does not exist. Hence, the theorem statement is valid *only* for the specified sequence of limits and the integral.

3.1. Brief Exposition of the Result

To understand the result in simpler terms, let us reconsider the definition of the scaled Fourier transform given in (3.1). The first exponential $\exp\left(\dfrac{iS(\boldsymbol{x})}{\tau}\right)$ is a varying complex "sinusoid", whereas the second exponential $\exp\left(-\dfrac{i\boldsymbol{u} \cdot \boldsymbol{x}}{\tau}\right)$ is a

fixed complex sinusoid at frequency $\dfrac{\boldsymbol{u}}{\tau}$. When we multiply these two complex exponentials, at low values of τ, the two sinusoids are usually not "in sync" and tend to cancel each other out. However, around the locations where $\nabla S(\boldsymbol{x}) = \boldsymbol{u}$, the two sinusoids are in perfect sync (as the combined exponent is *stationary*) with the approximate duration of this resonance depending on $\det(\mathcal{H}_x)$. The value of the integral in (3.1) can be increasingly closely approximated via the stationary phase approximation [9] as

$$F_\tau(\boldsymbol{u}) \approx \frac{1}{\mu(\Omega)^{\frac{1}{2}}} \sum_{k=1}^{N(u)} \frac{1}{\sqrt{\left| \det\left(\mathcal{H}_{x_k}\right) \right|}} \exp\left(\frac{i}{\tau} \left[S(\boldsymbol{x}_k) - \boldsymbol{u} \cdot \boldsymbol{x}_k \right] + i\sigma_{x_k} \frac{\pi}{4} \right).$$

The approximation is increasingly tight as $\tau \to 0$. The power spectrum (P_τ) gives us the required result $\dfrac{1}{\mu(\Omega)} \sum_{k=1}^{N(u)} \dfrac{1}{\left| \det\left(\mathcal{H}_{x_k}\right) \right|}$ except for the *cross* phase factors $S(\boldsymbol{x}_k) - S(\boldsymbol{x}_l) - \boldsymbol{u} \cdot (\boldsymbol{x}_k - \boldsymbol{x}_l)$ obtained as a byproduct of two or more remote locations $\boldsymbol{x}_k$ and $\boldsymbol{x}_l$ indexing into the same frequency bin $\boldsymbol{u}$, *i.e.*, $\boldsymbol{x}_k \neq \boldsymbol{x}_l$, but $\nabla S(\boldsymbol{x}_k) = \nabla S(\boldsymbol{x}_l) = \boldsymbol{u}$. The cross phase factors when evaluated are equivalent to $\cos\left(\dfrac{1}{\tau}\right)$, the limit of which does not exist as $\tau \to 0$. However, integrating the power spectrum over a small neighborhood $\mathcal{N}_\alpha(\boldsymbol{u})$ around $\boldsymbol{u}$ removes these cross phase factors as τ tends to zero and we obtain the desired result.

3.2. Formal Proof of Theorem 3.2

We wish to compute the integral

$$F_\tau(\boldsymbol{u}) = \frac{1}{(2\pi\tau)^{\frac{d}{2}} \mu(\Omega)^{\frac{1}{2}}} \int_\Omega \exp\left\{ \frac{i}{\tau} \left(S(\boldsymbol{x}) - \boldsymbol{u} \cdot \boldsymbol{x} \right) \right\} \mathrm{d}\boldsymbol{x} \tag{3.3}$$

at small values of τ and exhibit the connection between the power spectrum $P_\tau(\boldsymbol{u})$ and the density function $P(\boldsymbol{u})$. To this end define $\Psi(\boldsymbol{x}; \boldsymbol{u}) \equiv S(\boldsymbol{x}) - \boldsymbol{u} \cdot \boldsymbol{x}$. The proof follows by considering two cases: the first case in which there are no stationary points and therefore the density should go to zero, and the second case in which stationary points exist and the contribution from the oscillatory integral is shown to increasingly closely approximate the density function of the gradient as $\tau \to 0$.

case (i): We first consider the case where $N(\boldsymbol{u}) = 0$, *i.e.*, $\boldsymbol{u} \notin \nabla S(\Omega)$. In other words there are no stationary points [9] for this value of $\boldsymbol{u}$. The proof that this case yields the anticipated contribution of zero follows clearly from a straightforward technique commonly used in stationary phase expansions. We assume that the function S is sufficiently well-behaved on the boundary such that the total contribution due to the stationary points of the second kind [9] approaches zero as $\tau \to 0$. (We concentrate here on the crux of our work and reserve the discussion concerning the behavior of S on the boundary and the re-

lationship to stationary points of the second kind to Appendix 8.) Under mild conditions (outlined in Appendix B), the contributions from the stationary points of the third kind can also be ignored as they approach zero as τ tends to zero [9]. Higher order terms follow suit.

Lemma 3.3. *Fix* $\boldsymbol{u}_0 \notin \mathcal{C}$. *If* $\mathcal{A}_{\boldsymbol{u}_0} = \varnothing$ *then* $F_\tau(\boldsymbol{u}_0) = O\left(\sqrt{\tau}\right)$ *as* $\tau \to 0$.

Proof. To improve readability, we prove Lemma 3.3 first in the one dimensional setting and separately offer the proof for multiple dimensions.

3.2.1. Proof of Lemma 3.3 in One Dimension

Let s denote the derivative (1D gradient) of S. The bounded closed interval Ω is represented by $\Omega = [b_1, b_2]$, with the length $L = \mu(\Omega) = b_2 - b_1$. As $u_0 \notin \mathcal{C}$, there is no $x \in \Omega$ for which $s(x) = u_0$. Recalling the definition of Ψ, namely $\Psi(x; u) \equiv S(x) - ux$, we see that $\Psi'(x) \neq 0$ and is of constant sign in $[b_1, b_2]$. It follows that $\Psi(x)$ is strictly monotonic. Defining $v = \Psi(x)$, we have from (3.1)

$$F_\tau(u_0) = \frac{1}{\sqrt{2\pi\tau L}} \int_{\Psi(b_1)}^{\Psi(b_2)} \exp\left(\frac{iv}{\tau}\right) t(v) \mathrm{d}v.$$

Here $t(v) = \dfrac{1}{\Psi'\left(\Psi^{-1}(v)\right)}$. The inverse function is guaranteed to exist due to the monotonicity of Ψ. Integrating by parts we get

$$\begin{aligned} F_\tau(u_0)\sqrt{2\pi\tau L} = & \frac{\tau}{i}\left[\exp\left(\frac{i\Psi(b_2)}{\tau}\right)t\left(\Psi(b_2)\right) - \exp\left(\frac{i\Psi(b_1)}{\tau}\right)t\left(\Psi(b_1)\right)\right] \\ & - \frac{\tau}{i}\int_{\Psi(b_1)}^{\Psi(b_2)} \exp\left(\frac{iv}{\tau}\right)t'(v)\mathrm{d}v. \end{aligned} \tag{3.4}$$

It follows that

$$\left|F_\tau(u_0)\right| \le \frac{\sqrt{\tau}}{\sqrt{2\pi L}}\left(\frac{1}{\left|s(b_2) - u_0\right|} + \frac{1}{\left|s(b_1) - u_0\right|} + \int_{\Psi(b_1)}^{\Psi(b_2)}\left|t'(v)\right|\mathrm{d}v\right).$$

3.2.2. Proof of Lemma 3.3 in Finite Dimensions

As $\nabla\Psi(\boldsymbol{x}; \boldsymbol{u}_0) \neq \boldsymbol{0}, \forall \boldsymbol{x}$, the vector field

$$\boldsymbol{v}_1(\boldsymbol{x}) = \frac{\nabla\Psi}{\left\|\nabla\Psi\right\|^2}$$

is well-defined. Choose $m > \dfrac{d}{2}$ (with this choice explained below) and for $j \in \{1, 2, \cdots, m\}$, recursively define the function $g_j(\boldsymbol{x})$ and the vector field $\boldsymbol{v}_{j+1}(\boldsymbol{x})$ as follows:

$$g_1(\boldsymbol{x}) = 1,$$

$$g_{j+1}(\boldsymbol{x}) = \nabla \cdot \boldsymbol{v}_j(\boldsymbol{x})$$

and

$$\boldsymbol{v}_{j+1}(\boldsymbol{x}) = \frac{\nabla\Psi}{\left\|\nabla\Psi\right\|^2} g_{j+1}(\boldsymbol{x}).$$

Using the equality

$$
\exp\left(\frac{i}{\tau}\Psi(x;u_0)\right)g_j(x)
$$
$$
= i\tau\left[\nabla\cdot v_j(x)\right]\exp\left(\frac{i}{\tau}\Psi(x;u_0)\right) - i\tau\nabla\cdot\left[v_j(x)\exp\left(\frac{i}{\tau}\Psi(x;u_0)\right)\right]
$$

(3.5)

where $\nabla\cdot$ is the divergence operator, and applying the divergence theorem m times, the Fourier transform in (3.3) can be rewritten as

$$
F_\tau(u_0) = \frac{1}{(2\pi\tau)^{\frac{d}{2}}\mu(\Omega)^{\frac{1}{2}}}(i\tau)^m\int_\Omega g_{m+1}(x)\exp\left(\frac{i}{\tau}\Psi(x)\right)dx
$$
$$
- \frac{1}{(2\pi\tau)^{\frac{d}{2}}\mu(\Omega)^{\frac{1}{2}}}\sum_{j=1}^m(i\tau)^j\int_{\partial\Omega}\left[v_j(x(y))\cdot n\right]\exp\left(\frac{i}{\tau}\Psi(x(y))\right)dy
$$

(3.6)

which is similar to (3.4).

We would like to add a note on the differentiability of S which we briefly mentioned after Lemma 2.3. The divergence theorem is applied $m > \dfrac{d}{2}$ times to obtain sufficiently higher order powers of τ in the numerator so as to exceed the $\tau^{\frac{d}{2}}$ term in the denominator of the first line of (3.6). This necessitates that S is at least $\dfrac{d}{2}+1$ times differentiable. The smoothness constraint on S can thus be relaxed and replaced by $S\in C^{\frac{d}{2}+1}(\Omega)$.

The additional complication of the d-dimensional proof lies in resolving the geometry of the terms in the second line of (3.6). Here n is the unit outward normal to the positively oriented boundary $\partial\Omega$ parameterized by y. As $m > \dfrac{d}{2}$, the term on the right side of the first line in (3.6) is $o\left(\sqrt{\tau}\right)$ and hence can be overlooked. To evaluate the remaining integrals within the summation in (3.6), we should take note that the stationary points of the second kind for $\Psi(x)$ on Ω correspond to the first kind of stationary points for $\Psi(x(y))$ on the boundary $\partial\Omega$. We show in case (ii) that the contribution of a stationary point of the first kind in a $d-1$ dimensional space is $O\left(\tau^{\frac{d-1}{2}}\right)$. As the dimension of $\partial\Omega$ is $d-1$, we can conclude that the total contribution from the stationary points of the second kind is $O\left(\tau^{\frac{d-1}{2}}\right)$. After multiplying and dividing this contribution by the corresponding τ^j and $\tau^{\frac{d}{2}}$ factors respectively, it is easy to see that the contribution of the jth integral (out of the n integrals in the summation) in (3.6) is $O\left(\tau^{j-\frac{1}{2}}\right)$, and hence the total contribution of the m integrals is of $O\left(\sqrt{\tau}\right)$. Here, we have safely ignored the stationary points of the third kind as their contributions are minuscule compared to the other two kinds

as shown in [9]. Combining all the terms in (3.6) we get the desired result, namely $F_\tau(u_0) = O(\sqrt{\tau})$. For a detailed exposition of the proof, we encourage the reader to refer to Chapter 9 in [9].

We then get $P_\tau(u_0) = O(\tau)$. Since $\nabla S(\Omega)$ is a compact set in $\mathbb{R}^d$ and $u_0 \notin \nabla S(\Omega)$, we can choose a neighborhood $\mathcal{N}_\eta(u_0)$ around u_0 such that for $u \in \mathcal{N}_\eta(u_0)$, no stationary points exist and hence

$$\lim_{\tau \to 0} \int_{\mathcal{N}_\eta(u_0)} P_\tau(u) \, du = 0.$$

Since the cardinality $N(u)$ of the set $\mathcal{A}_u$ defined in (2.1) is zero for $u \in \mathcal{N}_\eta(u_0)$, the true density $P(u)$ of the random variable transformation $Y = \nabla S(X)$ given in (2.4) also vanishes for $u \in \mathcal{N}_\eta(u_0)$.

case (ii): For $u_0 \notin \mathcal{C}$, let $N(u_0) > 0$. In this case, the number of stationary points in the *interior* of Ω is non-zero and *finite* as a consequence of Lemma 2.1. We can then rewrite

$$F_\tau(u_0) = G + \frac{1}{(2\pi\tau)^{\frac{d}{2}} \mu(\Omega)^{\frac{1}{2}}} \sum_{k=1}^{N(u_0)} \int_{\mathcal{N}_\eta(x_k)} \exp\left(\frac{i}{\tau}\Psi(x; u_0)\right) dx, \qquad (3.7)$$

where

$$G \equiv \int_K \exp\left(\frac{i}{\tau}\Psi(x; u_0)\right) dx \qquad (3.8)$$

and the domain $K \equiv \Omega \setminus \bigcup_{k=1}^{N(u_0)} \overline{\mathcal{N}_\eta(x_k)}$. The closed regions $\left\{\overline{\mathcal{N}_\eta(x_k)}\right\}_{k=1}^{N(u_0)}$ are obtained from Lemma 2.2.

Firstly, note that the the set K contains no stationary points by construction. Secondly, the boundaries of K can be classified into two categories: those that overlap with the sets $\overline{\mathcal{N}_\eta(x_k)}$ and those that coincide with $\Gamma = \partial\Omega$. Propitiously, the orientation of the overlapping boundaries between the sets K and each $\overline{\mathcal{N}_\eta(x_k)}$ are in opposite directions as these sets are located at different sides when viewed from the boundary. Hence, we can *exclude* the contributions from the overlapping boundaries between K and $\overline{\mathcal{N}_\eta(x_k)}$ while evaluating $F_\tau(u_0)$ in (3.7) as they *cancel* each other out.

To compute G we leverage case (i), which also includes the contribution from the boundary Γ, and get

$$G = \epsilon_1(u_0, \tau) = O(\sqrt{\tau}). \qquad (3.9)$$

To evaluate the remaining integrals over $\overline{\mathcal{N}_\eta(x_k)}$, we take into account the contribution from the stationary point at x_k and obtain

$$\int_{\overline{\mathcal{N}_\eta(x_k)}} \exp\left(\frac{i}{\tau}\Psi(x; u_0)\right) dx$$

$$= \frac{a(2\pi\tau)^{\frac{d}{2}}}{\sqrt{\left|\det\left(\mathcal{H}_{x_k}\right)\right|}} \exp\left(\frac{i}{\tau}\Psi(x_k; u_0) + i\sigma_{x_k}\frac{\pi}{4}\right) + \epsilon_2(u_0, \tau), \qquad (3.10)$$

where

$$\epsilon_2\left(\boldsymbol{u}_0,\tau\right)=O\left(\tau^{\frac{d+1}{2}}\right)\le\tau^{\frac{d+1}{2}}\gamma_1\left(\boldsymbol{u}_0\right) \tag{3.11}$$

for a continuous bounded function $\gamma_1\left(\boldsymbol{u}\right)$ [9]. The variable a in (3.10) takes the value 1 if $\boldsymbol{x}_k$ lies in the interior of Ω, otherwise equals $\frac{1}{2}$ if $\boldsymbol{x}_k\in\partial\Omega$. Since $\boldsymbol{u}\notin\mathcal{C}$, stationary points do not occur on the boundary and hence $a=1$ for our case. Recall that $\sigma_{\boldsymbol{x}_k}$ is the signature of the Hessian at $\boldsymbol{x}_k$. Note that the main term has the factor $\tau^{\frac{d}{2}}$ in the numerator, when we perform stationary phase in d dimensions, as mentioned under the finite dimensional proof of Lemma 3.3.

Coupling (3.7), (3.8), and (3.10) yields

$$F_\tau\left(\boldsymbol{u}_0\right)=\frac{1}{\mu\left(\Omega\right)^{\frac{1}{2}}}\sum_{k=1}^{N\left(\boldsymbol{u}_0\right)}\exp\left(\frac{i}{\tau}\Psi\left(\boldsymbol{x}_k;\boldsymbol{u}_0\right)+i\sigma_{\boldsymbol{x}_k}\frac{\pi}{4}\right)\frac{1}{\sqrt{\left|\det\left(\mathcal{H}_{\boldsymbol{x}_k}\right)\right|}}+\epsilon_3\left(\boldsymbol{u}_0,\tau\right) \tag{3.12}$$

where

$$\epsilon_3\left(\boldsymbol{u}_0,\tau\right)=\epsilon_1\left(\boldsymbol{u}_0,\tau\right)+\frac{\epsilon_2\left(\boldsymbol{u}_0,\tau\right)}{\left(2\pi\tau\right)^{\frac{d}{2}}\mu\left(\Omega\right)^{\frac{1}{2}}}.$$

As $\epsilon_1\left(\boldsymbol{u}_0,\tau\right)=O\left(\sqrt{\tau}\right)$ and $\epsilon_2\left(\boldsymbol{u}_0,\tau\right)=O\left(\tau^{\frac{d+1}{2}}\right)$ from (3.9) and (3.11) respectively, we have $\epsilon_3\left(\boldsymbol{u}_0,\tau\right)=O\left(\sqrt{\tau}\right)$. Based on the definition of the power spectrum P_τ in (3.2), we get

$$\begin{aligned}P_\tau\left(\boldsymbol{u}_0\right)=&\frac{1}{\mu\left(\Omega\right)}\sum_{k=1}^{N\left(\boldsymbol{u}_0\right)}\frac{1}{\left|\det\left(\mathcal{H}_{\boldsymbol{x}_k}\right)\right|}\\&+\frac{1}{\mu\left(\Omega\right)}\sum_{k=1}^{N\left(\boldsymbol{u}_0\right)}\sum_{\substack{l=1\\l\ne k}}^{N\left(\boldsymbol{u}_0\right)}\frac{\exp\left\{\frac{i}{\tau}\left[\Psi\left(\boldsymbol{x}_k;\boldsymbol{u}_0\right)-\Psi\left(\boldsymbol{x}_l;\boldsymbol{u}_0\right)\right]+i\left(\sigma_{\boldsymbol{x}_k}-\sigma_{\boldsymbol{x}_l}\right)\frac{\pi}{4}\right\}}{\sqrt{\left|\det\left(\mathcal{H}_{\boldsymbol{x}_k}\right)\right|}\sqrt{\left|\det\left(\mathcal{H}_{\boldsymbol{x}_l}\right)\right|}}\\&+\epsilon_4\left(\boldsymbol{u}_0,\tau\right)\end{aligned} \tag{3.13}$$

where $\epsilon_4\left(\boldsymbol{u}_0\right)$ includes both the squared magnitude of $\epsilon_3\left(\boldsymbol{u}_0,\tau\right)$ and the cross terms involving the first term in (3.12) and $\epsilon_3\left(\boldsymbol{u}_0,\tau\right)$. Notice that the main term in (3.12) can be bounded *independently* of τ as

$$\left|\exp\left(\frac{i}{\tau}\Psi\left(\boldsymbol{x}_k;\boldsymbol{u}_0\right)+i\sigma_{\boldsymbol{x}_k}\frac{\pi}{4}\right)\right|=1,\forall\tau\ne0$$

and $\det\left(\mathcal{H}_{\boldsymbol{x}_k}\right)\ne0,\forall k$. Since $\epsilon_3\left(\boldsymbol{u}_0,\tau\right)=O\left(\sqrt{\tau}\right)$, we get $\epsilon_4\left(\boldsymbol{u}_0,\tau\right)=O\left(\sqrt{\tau}\right)$. Furthermore, as $\epsilon_4\left(\boldsymbol{u}_0,\tau\right)$ can also be uniformly bounded by a function of $\boldsymbol{u}$ for small values of τ, we have

$$\lim_{\tau\to0}\int_{\mathcal{N}_\eta\left(\boldsymbol{u}_0\right)}\epsilon_4\left(\boldsymbol{u}_0,\tau\right)\mathrm{d}\boldsymbol{u}=0. \tag{3.14}$$

Observe that the term on the right side of the first line in (3.13) matches the

anticipated expression for the density function $P(u_0)$ given in (2.4). The cross phase factors in the second line of (3.13 arise due to multiple remote locations x_k and x_l indexing into u. The cross phase factors when evaluated can be shown to be proportional to $\cos\left(\dfrac{1}{\tau}\right)$. Since $\lim_{\tau\to 0}\cos\left(\dfrac{1}{\tau}\right)$ is not defined, $\lim_{\tau\to 0}P_\tau(u_0)$ does not exist. We briefly alluded to this problem immediately following the statement of Theorem 3.2 in Section 3. However, the following lemma which invokes the inverse function $\nabla S_x^{(-1)}(u):\overline{\mathcal{N}_\eta(u_0)}\to\overline{\mathcal{N}_\eta(x)}$ —defined in Lemma 2.2 where x is written as a function of u—provides a simple way to nullify the cross phase factors. Note that since each $\nabla S_x^{(-1)}$ is a bijection, $N(u)$ doesn't vary over $\overline{\mathcal{N}_\eta(u_0)}$. Pursuant to Lemma 2.2, the Hessian signatures $\sigma_{x_k(u)}$ and $\sigma_{x_l(u)}$ also remain constant over $\overline{\mathcal{N}_\eta(u_0)}$.

Lemma 3.4. [*Cross Factor Nullifier Lemma*] *The integral of the cross term in the second line of* (3.13) *over the closed region* $\overline{\mathcal{N}_\eta(u_0)}$ *approaches zero as* $\tau\to 0$, *i.e.,* $\forall k\neq l$

$$\lim_{\tau\to 0}\int_{\mathcal{N}_\eta(u_0)}\frac{\exp\left\{\dfrac{i}{\tau}\left[\Psi(x_k(u);u)-\Psi(x_l(u);u)\right]\right\}}{\left|\det\left(\mathcal{H}_{x_k(u)}\right)\right|^{\frac{1}{2}}\left|\det\left(\mathcal{H}_{x_l(u)}\right)\right|^{\frac{1}{2}}}\,\mathrm{d}u = 0. \qquad (3.15)$$

The proof is given in Appendix A. Combining (3.14) and (3.15) yields

$$\lim_{\tau\to 0}\int_{\mathcal{N}_\eta(u_0)}P_\tau(u)\,\mathrm{d}u = \frac{1}{\mu(\Omega)}\int_{\mathcal{N}_\eta(u_0)}\sum_{k=1}^{N(u)}\frac{1}{\left|\det\left(\mathcal{H}_{x_k(u)}\right)\right|}\,\mathrm{d}u = \int_{\mathcal{N}_\eta(u_0)}P(u)\,\mathrm{d}u. \quad (3.16)$$

Equation (3.16) demonstrates the equivalence of the cumulative distributions corresponding to the densities $P_\tau(u)$ and $P(u)$ when integrated over *any* sufficiently small neighborhood $\mathcal{N}_\eta(u_0)$ of radius η. To recover the density $P(u)$, we let $\alpha < \eta$ and take the limit with respect to α.

Taking a mild digression from the main theme of this paper, in the next section (Section 4), we build an informal bridge between the commonly used characteristic function formulation for computing densities and our wave function method. The motivation behind this section is merely to provide an intuitive reason behind our Theorem 3.2, where we directly manipulate the power spectrum of $\phi(x)=\exp\left(\dfrac{iS(x)}{\tau}\right)$ into the characteristic function formulation stated in (4.2), circumventing the need for the closed-form expression of the density function $P(u)$ given in (2.4). We request the reader to bear in mind the following cautionary note. What we show below *cannot* be treated as a formal proof of Theorem 3.2. Our attempt here merely provides a mathematically intuitive justification for establishing the equivalence between the power spectrum and the characteristic function formulations and thereby to the density function $P(u)$. On the basis of the reasons described therein, we strongly believe that the mechanism of stationary phase is essential to formally prove our main theorem (Theorem 3.2). It is best to treat the wave function and the cha-

racteristic function methods as two *different* approaches for estimating the probability density functions and *not* reformulations of each other. To press this point home, we also comment on the computational complexity of the wave function and the characteristic function methods at the end of the next section.

4. Relation between the Characteristic Function and Power Spectrum Formulations of the Gradient Density

The characteristic function $\psi_Y(w)$ for the random variable $Y = \nabla S(X)$ is defined as the expected value of $\exp(iw \cdot Y)$, namely

$$\psi_Y(w) \equiv E\left[\exp(iw \cdot Y)\right] = \frac{1}{\mu(\Omega)} \int_\Omega \exp(iw \cdot \nabla S(x)) dx. \tag{4.1}$$

Here $\dfrac{1}{\mu(\Omega)}$ denotes the density of the uniformly distributed random variable X on Ω.

The inverse Fourier transform of a characteristic function also serves as the density function of the random variable under consideration [30]. In other words, the density function $P(u)$ of the random variable Y can be obtained via

$$
\begin{aligned}
P(u) &= \frac{1}{(2\pi)^d} \int \psi_Y(\omega) \exp(-iw \cdot u) dw \\
&= \frac{1}{(2\pi)^d \mu(\Omega)} \iint_\Omega \exp(iw \cdot [\nabla S(x) - u]) dx dw.
\end{aligned}
\tag{4.2}
$$

Having set the stage, we can now proceed to highlight the close relationship between the characteristic function formulation of the density and our formulation arising from the power spectrum. For simplicity, we choose to consider a region Ω that is the product of closed intervals, $\Omega = \prod_{i=1}^{d}[a_i, b_i]$. Based on the expression for the scaled Fourier transform $F_\tau(u)$ in (3.1), the power spectrum $P_\tau(u)$ is given by

$$P_\tau(u) = \frac{1}{(2\pi\tau)^d \mu(\Omega)} \int_\Omega \int_\Omega \exp\left\{\frac{i}{\tau}\left[S(x) - S(y) - (u \cdot (x - y))\right]\right\} dx dy.$$

Define the following change of variables,

$$\omega = \frac{x - y}{\tau}, v = \frac{x + y}{2}.$$

Then $x = v + \dfrac{\omega\tau}{2}, y = v - \dfrac{\omega\tau}{2}$ and the integral limits for ω and v are given by

$$W = \prod_{i=1}^{d} \left[\frac{a_i - b_i}{\tau}, \frac{b_i - a_i}{\tau}\right]$$

$$V_\omega = \prod_{i=1}^{d} \left[a_i + \frac{|\omega_i|\tau}{2}, b_i - \frac{|\omega_i|\tau}{2}\right]$$

where ω_i is the ith component of $\boldsymbol{\omega}$. Note that the Jacobian of this transformation is τ^d. Now we may write the integral $P_\tau(\boldsymbol{u})$ in terms of these new variables as

$$P_\tau(\boldsymbol{u}) = \frac{1}{(2\pi)^d \mu(\Omega)} \int_W \xi(\boldsymbol{\omega}, \boldsymbol{u}) \, d\boldsymbol{\omega} \qquad (4.3)$$

where

$$\xi(\boldsymbol{\omega}, \boldsymbol{u}) = \int_{V_{\boldsymbol{\omega}}} \exp\left\{ \frac{i}{\tau}\left[S\left(\boldsymbol{v} + \frac{\tau\boldsymbol{\omega}}{2}\right) - S\left(\boldsymbol{v} - \frac{\tau\boldsymbol{\omega}}{2}\right) \right] \right\} \exp(-i\boldsymbol{u}\cdot\boldsymbol{\omega}) \, d\boldsymbol{v}. \qquad (4.4)$$

The mean value theorem applied to $S\left(\boldsymbol{v} + \dfrac{\tau\boldsymbol{\omega}}{2}\right) - S\left(\boldsymbol{v} - \dfrac{\tau\boldsymbol{\omega}}{2}\right)$ yields

$$\xi(\boldsymbol{\omega}, \boldsymbol{u}) = \int_{V_{\boldsymbol{\omega}}} \exp\left(i\boldsymbol{\omega}\cdot\left[\nabla S(z(\boldsymbol{v}, \boldsymbol{\omega})) - \boldsymbol{u} \right] \right) d\boldsymbol{v}, \qquad (4.5)$$

where

$$z(\boldsymbol{v}, \boldsymbol{\omega}) \equiv c\left(\boldsymbol{v} + \frac{\tau\boldsymbol{\omega}}{2}\right) + (1-c)\left(\boldsymbol{v} - \frac{\tau\boldsymbol{\omega}}{2}\right)$$

with $c \in [0,1]$. When $\boldsymbol{\omega}$ is fixed and $\tau \to 0$, $z(\boldsymbol{v}, \boldsymbol{\omega}) \to \boldsymbol{v}$ and so for small values of τ we get

$$\xi(\boldsymbol{\omega}, \boldsymbol{u}) \approx \int_{n \in V_{\boldsymbol{\omega}}} \exp\left(i\boldsymbol{\omega}\cdot\left[\nabla S(\boldsymbol{v}) - \boldsymbol{u} \right] \right) d\boldsymbol{v}. \qquad (4.6)$$

Again we would like to drive the following point home. We *do not* claim that we have formally proved the above approximation. On the contrary, we believe that it might be an onerous task to do so as the mean value theorem point z in (4.5) is unknown and the integration limits for $\boldsymbol{v}$ directly depend on τ. The approximation is stated with the sole purpose of providing an intuitive reason for our theorem (Theorem 3.2) and to provide a clear link between the characteristic function and wave function methods for density estimation.

Furthermore, note that the integral range for $\boldsymbol{\omega}$ depends on τ and so when $\boldsymbol{\omega} = O\left(\dfrac{1}{\tau}\right)$, $\boldsymbol{\omega}\tau \nrightarrow 0$ as $\tau \to 0$ and hence the above approximation for $\xi(\boldsymbol{\omega}, \boldsymbol{u})$ in (4.6) might seem to break down. To evade this seemingly ominous problem, we manipulate the domain of integration for $\boldsymbol{\omega}$ as follows. Fix an $\epsilon \in (0,1)$ and let

$$W = W^\infty \cup W^\tau = \left(W \setminus \prod_{i=1}^d [-M_i, M_i] \right) \cup \prod_{i=1}^d [-M_i, M_i]$$

where

$$M_i \equiv (b_i - a_i)\tau^{\epsilon-1}. \qquad (4.7)$$

By defining M_i as above, note that in W^τ, $\boldsymbol{\omega}$ is deliberately made to be $O(\tau^{\epsilon-1})$ and hence $\boldsymbol{\omega}\tau \to 0$ as $\tau \to 0$. Hence the approximation for $\xi(\boldsymbol{\omega}, \boldsymbol{u})$ in (4.6) might hold for this integral range of $\boldsymbol{\omega}$. For consideration of $\boldsymbol{\omega} \in W^\infty$, recall that Theorem 3.2 requires the power spectrum $P_\tau(\boldsymbol{u})$ to be in-

tegrated over a small neighborhood $\mathcal{N}_\alpha(\boldsymbol{u}_0)$ around $\boldsymbol{u}_0$. By using the true expression for $\xi(\boldsymbol{\omega},\boldsymbol{u})$ from (4.4) and performing the integral for $\boldsymbol{u}$ prior to $\boldsymbol{\omega}$ and $\boldsymbol{v}$, we get

$$\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}\int_{W^\infty}\xi(\boldsymbol{\omega},\boldsymbol{u})\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{u}$$

$$=\int_{W^\infty}\int_{V_{\boldsymbol{\omega}}}\exp\left\{i\boldsymbol{\omega}\cdot\left[\nabla S(z(\boldsymbol{v},\boldsymbol{\omega}))\right]\right\}\left[\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}\exp(-i\boldsymbol{\omega}\cdot\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\right]\mathrm{d}\boldsymbol{v}\,\mathrm{d}\boldsymbol{\omega}.$$

Since both M_i in (4.7) and the lower and the upper limits for ω_i, namely $\pm\dfrac{b_i-a_i}{\tau}$ respectively approach ∞ as $\tau\to 0$, the Riemann-Lebesgue lemma [21] guarantees that $\forall\boldsymbol{\omega}\in W^\infty$, the integral

$$\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}\exp(-i\boldsymbol{\omega}\cdot\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

approaches zero as $\tau\to 0$. Hence for small values of τ, we can expect the integral over W^τ to dominate over the other. This leads to the following approximation,

$$\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}P_\tau(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}\approx\frac{1}{(2\pi)^d\,\mu(\Omega)}\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}\int_{W^\tau}\xi(\boldsymbol{\omega},\boldsymbol{u})\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{u}, \tag{4.8}$$

as τ approaches zero. Combining the above approximation with the approximation for $\xi(\boldsymbol{\omega},\boldsymbol{u})$ given in (4.6) and noting that the integral domain for $\boldsymbol{\omega}$ and $\boldsymbol{v}$ approaches $\mathbb{R}^d$ and Ω respectively as $\tau\to 0$, the integral of the power spectrum $P_\tau(u)$ over the neighborhood $\mathcal{N}_\alpha(\boldsymbol{u}_0)$ at small values of τ in (4.3) can be approximated by

$$\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}P_\tau(u)\,\mathrm{d}u\approx\frac{1}{(2\pi)^d\,\mu(\Omega)}\int_{\mathcal{N}_\alpha(\boldsymbol{u}_0)}\iint_\Omega\exp\left(i\boldsymbol{\omega}\cdot\left[\nabla S(\boldsymbol{v})-\boldsymbol{u}\right]\right)\mathrm{d}\boldsymbol{v}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{u}.$$

This form *exactly* coincides with the expression given in (4.2) obtained through the characteristic function formulation.

The approximations given in (4.6) and (4.8) cannot be proven easily as they involve limits of integration which directly depend on τ. Furthermore, the mean value theorem point $z(\boldsymbol{v},\boldsymbol{\omega})$ in (4.5) is arbitrary and *cannot be determined beforehand for a given value of τ*. The difficulties faced here emphasize the need for the method of stationary phase to formally prove Theorem 3.2.

As we remarked before, the characteristic function and our wave function methods should not be treated as mere reformulations of each other. This distinction is further emphasized when we find our method to be computationally more efficient than the characteristic function approach in the finite sample-set scenario where we estimate the gradient density from N samples of the function *S*. Given these N sample values $\hat{S}$ and their gradients $\nabla\hat{S}$, the characteristic function defined in (4.1) needs to be computed for N integral values of $\boldsymbol{\omega}$. Each value of $\boldsymbol{\omega}$ requires summation over the N sampled values of $\exp(i\boldsymbol{\omega}\nabla S(\boldsymbol{x}))$. Hence the total time required to determine the characteristic function is $O(N^2)$. The joint density function of the gradient is obtained via the inverse

Fourier transform of the characteristic function, which is an $O(N \log N)$ operation. The overall time complexity is therefore $O(N^2)$. In our wave function method the Fourier transform of $\exp\left(\dfrac{i\hat{S}(\boldsymbol{x})}{\tau}\right)$ at a given value of τ can be computed in $O(N \log N)$ and the subsequent squaring operation to obtain the power spectrum can be performed in $O(N)$. Hence the density function can be determined in $O(N \log N)$, which is more efficient when compared to the $O(N^2)$ complexity of the characteristic function approach.

5. Experimental Evidence in 2D

We would like to emphasize that our wave function method for computing the gradient density is very fast and straightforward to implement as it requires computation of a single Fourier transform. We ran multiple simulations on many different types of functions to assess the efficacy of our wave function method. Below we show comparisons with the standard histogramming technique where the functions were sampled on a regular grid between $[-0.125, 0.125] \times [-0.125, 0.125]$ at a grid spacing of $\dfrac{1}{2^{13}}$. For the sake of convenience, we normalized the functions such that the maximum gradient magnitude value $(\|\nabla S\|)$ is 1. Using the sampled values $\hat{S}$, we first computed the fast Fourier transform of $\exp\left(\dfrac{i\hat{S}}{\tau}\right)$ at $\tau = 0.00004$, then computed the power spectrum followed by normalization to obtain the joint gradient density. We also computed the discrete derivative of S at the grid locations to obtain the gradient $\nabla S = \left(\hat{S}_{x_1}, \hat{S}_{x_2}\right)$ and then determined the gradient density by histogramming. For better visualization, we marginalized the density along the radial and the orientation directions. The plots shown in Figure 1 provide visual, empirical evidence corroborating our theorem. Notice the near-perfect match between the gradient densities computed via standard histogramming and our wave function method. The accuracy of the density marginalized along the orientations further strengthens our claim made in Section 1 about the wave function method serving as a reliable estimator for the histogram of oriented gradients (HOG). In Figure 2, we verify the convergence of our estimated density to the true density as $\tau \to 0$ in accordance with Theorem 3.2.

6. Conclusions

Observe that the integrals

$$I_\tau(\boldsymbol{u}_0) = \int_{\mathcal{N}_\eta(\boldsymbol{u}_0)} P_\tau(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}, \quad I(\boldsymbol{u}_0) = \int_{\mathcal{N}_\eta(\boldsymbol{u}_0)} P(\boldsymbol{u}) \, \mathrm{d}\boldsymbol{u}$$

give the interval measures of the density functions P_τ and P respectively. Theorem 3.2 states that at small values of τ, both the interval measures are approximately equal, with the difference between them being $O(\sqrt{\tau})$ which converges to zero as $\tau \to 0$. Recall that by definition, P_τ is the normalized power
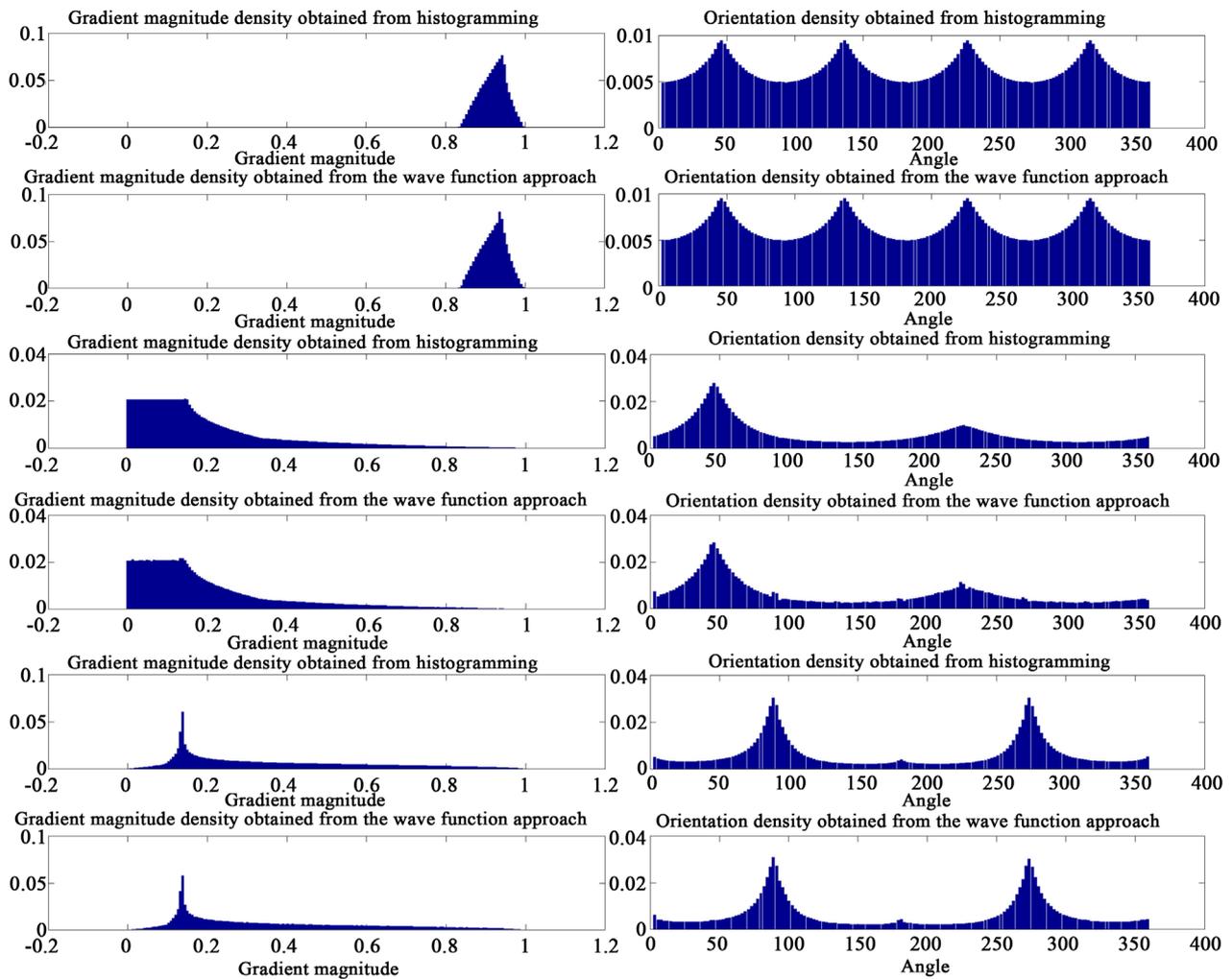
Figure 1. Comparison results. 1) Left: Gradient magnitude density, 2) Right: Gradient orientation density. In each sub-figure, we plot the density function obtained from histogramming and the wave function method at the top and bottom respectively.

spectrum of the wave function $\phi(x) = \exp\left(\dfrac{iS(x)}{\tau}\right)$. Hence we conclude that the power spectrum of $\phi(x)$ can potentially serve as a *joint density estimator* for the gradient of S at small values of τ, where the frequencies act as gradient histogram bins. We also built an informal bridge between our wave function method and the characteristic function approach for estimating probability densities, by directly trying to recast the former expression into the latter. The difficulties faced in relating the two approaches reinforce the stationary phase method as a powerful tool to formally prove Theorem 3.2. Our earlier result proved in [20], where we employ the stationary phase method to compute the gradient density of Euclidean distance functions in two dimensions, is now generalized in Theorem 3.2 which establishes a similar gradient density estimation result for arbitrary smooth functions in arbitrary finite dimensions.

As mentioned earlier, in [23] we have established error bounds in one dimension for the practical finite sample-set setting, wherein the gradient density is
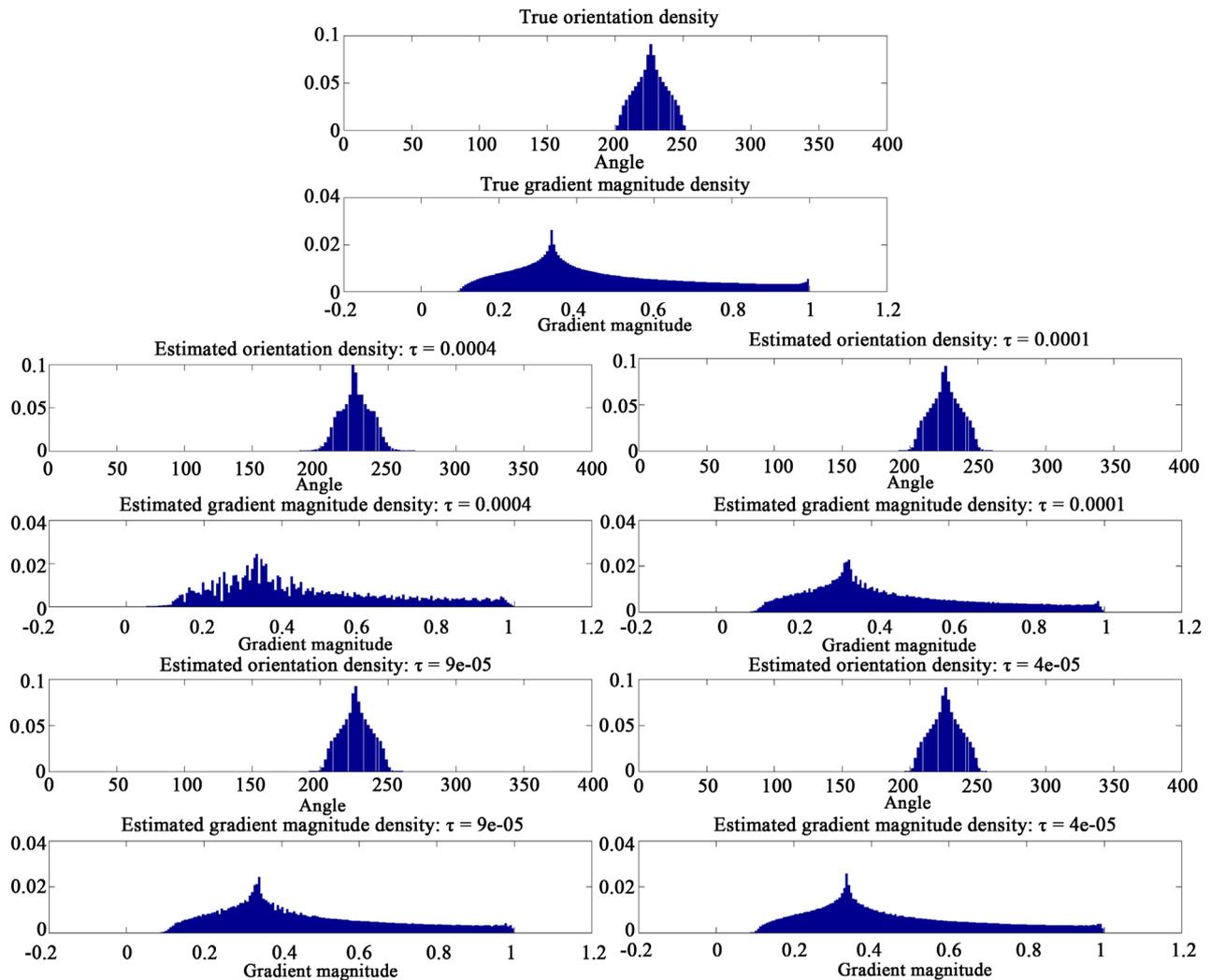
Figure 2. Convergence of our wave function method as $\tau \to 0$. The value of τ is steadily decreased from right to left, top to bottom.

estimated from a finite, discrete set of samples, instead of assuming that the function is fully described over a compact set Ω. In the future, we plan to extend this work and derive similar *finite sample error bounds* for gradient density estimation in arbitrary higher dimensions.

Support

This research work benefited from the support of the AIRBUS Group Corporate Foundation Chair in Mathematics of Complex Systems established in ICTS-TIFR.

This research work benefited from the support of NSF IIS 1743050.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Parzen, E. (1962) On the Estimation of a Probability Density Function and the Mode. *The Annals of Mathematical Statistics*, **33**, 1065-1076. https://doi.org/10.1214/aoms/1177704472

[2] Rosenblatt, M. (1956) Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, **33**, 832-837. https://doi.org/10.1214/aoms/1177728190

[3] Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC, New York. https://doi.org/10.1007/978-1-4899-3324-9

[4] Bishop, C.M. (2006) Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, New York.

[5] Jones, D.S. and Kline, M. (1958) Asymptotic Expansions of Multiple Integrals and the Method of Stationary Phase. *Journal of Mathematical Physics*, **37**, 1-28. https://doi.org/10.1002/sapm19583711

[6] Cooke, J.C. (1982) Stationary Phase in Two Dimensions. *IMA Journal of Applied Mathematics*, **29**, 25-37. https://doi.org/10.1093/imamat/29.1.25

[7] McClure, J.P. and Wong, R. (1991) Two-Dimensional Stationary Phase Approximation: Stationary Point at a Corner. *SIAM Journal on Mathematical Analysis*, **22**, 500-523. https://doi.org/10.1137/0522032

[8] McClure, J.P. and Wong, R. (1997) Justification of the Stationary Phase Approximation in Time-Domain Asymptotics. *Proceedings: Mathematical, Physical and Engineering Sciences*, **453**, 1019-1031. https://doi.org/10.1098/rspa.1997.0057

[9] Wong, R. (1989) Asymptotic Approximations of Integrals. Academic Press, New York.

[10] Wong, R. and McClure, J.P. (1981) On a Method of Asymptotic Evaluation of Multiple Integrals. *Mathematics of Computation*, **37**, 509-521. https://doi.org/10.2307/2007443

[11] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 20-26 June 2005, San Diego, 886-893. https://doi.org/10.1109/CVPR.2005.177

[12] Zhu, Q., Yeh, M.-C., Cheng, K.-T. and Avidan, S. (2006) Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. *IEEE Conference on Computer Vision and Pattern Recognition*, New York, 17-22 June 2006, 1491-1498. https://doi.org/10.1109/CVPR.2006.119

[13] Suard, F., Rakotomamonjy, A. and Bensrhair, A. (2006) Pedestrian Detection Using Infrared Images and Histograms of Oriented Gradients. *IEEE Conference on Intelligent Vehicles*, Tokyo, 13-15 June 2006, 206-212. https://doi.org/10.1109/IVS.2006.1689629

[14] Bertozzi, M., Broggi, A., Del Rose, M., Felisa, M., Rakotomamonjy, A. and Suard, F. (2007) A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier. *IEEE Conference on Intelligent Transportation Systems*, Seattle, 30 September-3 October 2007, 143-148. https://doi.org/10.1109/ITSC.2007.4357692

[15] Vapnik, V.N. (1998) Statistical Learning Theory. Wiley-Interscience, New York.

[16] Hu, R. and Collomosse, J. (2013) A Performance Evaluation of Gradient Field HOG Descriptor for Sketch Based Image Retrieval. *Computer Vision and Image Understanding*, **117**, 790-806. https://doi.org/10.1016/j.cviu.2013.02.005

[17] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F., Buell, D.A., *et al*. (2019) Quantum Supremacy Using a Programmable Superconducting Processor. *Nature*, **574**, 505-510. https://doi.org/10.1038/s41586-019-1666-5

[18] Boixo, S., Isakov, S.V., Smelyanskiy, V.N., Babbush, R., Ding, N., Jiang, Z., Bremner, M.J., Martinis, J.M. and Neven, H. (2018) Characterizing Quantum Supremacy in Near-Term Devices. *Nature Physics*, **14**, 595-600. https://doi.org/10.1038/s41567-018-0124-x

[19] Markov, I.G., Fatima, A., Isakov, S.V. and Boixo, S. (2018) Quantum Supremacy Is Both Closer and Farther than It Appears. arXiv:1807.10749 [quant-ph]

[20] Gurumoorthy, K.S. and Rangarajan, A. (2012) Distance Transform Gradient Density Estimation Using the Stationary Phase Approximation. *SIAM Journal on Mathematical Analysis*, **44**, 4250-4273. https://doi.org/10.1137/110859336

[21] Bracewell, R.N. (1999) The Fourier Transform and Its Applications. 3rd Edition, McGraw-Hill, New York.

[22] Fukunaga, K. and Hostetler, L. (1975) The Estimation of the Gradient of a Density Function with Applications in Pattern Recognition. *IEEE Transactions on Information Theory*, **21**, 32-40. https://doi.org/10.1109/TIT.1975.1055330

[23] Gurumoorthy, K.S. and Rangarajan, A. (2014) Error Bounds for Gradient Density Estimation Computed from a Finite Sample Set Using the Method of Stationary Phase. arXiv:1404.1147 [stat.CO]

[24] Scott, D. (1979) On Optimal and Data-Based Histograms. *Biometrika*, **66**, 605-610. https://doi.org/10.1093/biomet/66.3.605

[25] Cencov, N.N. (1962) Estimation of an Unknown Distribution Density from Observations. *Soviet Mathematics*, **3**, 1559-1562.

[26] Wahba, G. (1975) Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation. *Annals of Statistics*, **3**, 15-29. https://doi.org/10.1214/aos/1176342997

[27] Griffiths, D.J. (2005) Introduction to Quantum Mechanics. 2nd Edition, Prentice Hall, Upper Saddle River.

[28] Feynman, R.P. and Hibbs, A.R. (2010) Quantum Mechanics and Path Integrals: Emended Edition. Dover Books on Physics. Dover, Mineola.

[29] Goldstein, H., Poole, C.P. and Safko, J.L. (2001) Classical Mechanics. 3rd Edition, Addison Wesley, Boston. https://doi.org/10.1119/1.1484149

[30] Billingsley, P. (1995) Probability and Measure. 3rd Edition, Wiley-Interscience, New York.

[31] Rudin, W. (1976) Principles of Mathematical Analysis. 3rd Edition, McGraw-Hill, New York.

Appendix A. Proof of Lemmas

1) Proof of Finiteness Lemma

Proof. We prove the result by contradiction. Observe that $\mathcal{A}_u$ is a subset of the compact set Ω. If $\mathcal{A}_u$ is not finite, then by Theorem (2.37) in [31], $\mathcal{A}_u$ has a limit point $x_0 \in \Omega$. If $x_0 \in \partial\Omega$, then $u \in \mathcal{C}$ giving a contradiction. Otherwise, consider a sequence $\{x_n\}_{n=1}^{\infty}$, with each $x_n \in \mathcal{A}_u$, converging to x_0. Since $\nabla S(x_n) = u$ for all *n*, from continuity it follows that $\nabla S(x_0) = u$ and hence $x_0 \in \mathcal{A}_u$. Let $p_n \equiv x_n - x_0$ and $h_n \equiv \dfrac{p_n}{\|p_n\|}$. Then

$$\lim_{n \to \infty} \frac{\nabla S(x_n) - \nabla S(x_0) - \mathcal{H}_{x_0} p_n}{\|p_n\|} = 0,$$

where the linear operator $\mathcal{H}_{x_0}$ is the *Hessian* of S at x_0 (obtained from the set of derivatives of the vector field $\nabla S : \mathbb{R}^d \to \mathbb{R}^d$ at the location x_0). As $\nabla S(x_n) = \nabla S(x_0) = u$ and $\mathcal{H}_{x_0}$ is linear, we get

$$\lim_{n \to \infty} \mathcal{H}_{x_0} h_n = 0.$$

Since h_n is defined above to be a unit vector, it follows that $\mathcal{H}_{x_0}$ is rank deficient and $\det(\mathcal{H}_{x_0}) = 0$. Hence $x_0 \in \mathcal{B}$ and $u \in \mathcal{C}$ resulting in a contradiction.

2) Proof of Neighborhood Lemma

Proof. Observe that the set $\mathcal{B}$ defined in (2.2) is closed because if x_0 is a limit point of $\mathcal{B}$, from the continuity of the determinant function we have $\det(\mathcal{H}_{x_0}) = 0$ and hence $x_0 \in \mathcal{B}$. Being a bounded subset of Ω, $\mathcal{B}$ is also compact. As $\partial\Omega$ is also compact and ∇S is continuous, $\mathcal{C}$ is compact and hence $\mathbb{R}^d - \mathcal{C}$ is open. Then for $u_0 \notin \mathcal{C}$, there exists an open neighborhood $\mathcal{N}_r(u_0)$ for some $r > 0$ around u_0 such that $\mathcal{N}_r(u_0) \cap \mathcal{C} = \varnothing$. By letting $\eta = \dfrac{r}{2}$, we get the required closed neighborhood $\overline{\mathcal{N}_\eta(u_0)} \subset \mathcal{N}_r(u_0)$ containing u_0.

Since $\det(\mathcal{H}_x) \neq 0, \forall x \in \mathcal{A}_{u_0}$, points 1, 2 and 3 of this lemma follow directly from the inverse function theorem. As $|\mathcal{A}_{u_0}|$ is finite by Lemma 2.1, the closed neighborhood $\overline{\mathcal{N}_\eta(u_0)}$ can be chosen independently of $x \in \mathcal{A}_{u_0}$ so that points 1 and 3 are satisfied $\forall x \in \mathcal{A}_{u_0}$. In order to prove point 4, note that the eigenvalues of $\mathcal{H}_x$ are all *non-zero* and vary continuously for $x \in \overline{\mathcal{N}_\eta(x)}$. As the eigenvalues never cross zero, they retain their sign and so the signature of the Hessian stays fixed.

3) Proof of Density Lemma

Proof. Since the random variable *X* is assumed to have a uniform distribution on Ω, its density at every location $x \in \Omega$ equals $\dfrac{1}{\mu(\Omega)}$. Recall that the random variable *Y* is obtained via a random variable transformation from *X*, using the function ∇S. The Jacobian of ∇S at a location $x \in \Omega$ equals the Hessian $\mathcal{H}_x$ of the function *S* at x. Barring the set $\mathcal{C}$ corresponding to the

union of the image (under ∇S) of the set of points $\mathcal{B}$ (where the Hessian vanishes) and the boundary $\partial\Omega$, the density of Y exists on $\boldsymbol{u} \in \mathbb{R}^d - \mathcal{C}$ and is given by (2.4). Please see well known sources such as [30] for a detailed explanation.

For the sake of completeness we explicitly prove the well-known result stated in Integral Lemma 3.1.

4) Proof of Integral Lemma

Proof. Define a function $H(\boldsymbol{x})$ by

$$H(\boldsymbol{x}) \equiv \begin{cases} 1: & \text{if } \boldsymbol{x} \in \Omega; \\ 0: & \text{otherwise.} \end{cases}$$

Let $f(\boldsymbol{x}) = H(\boldsymbol{x}) \exp\left(\dfrac{iS(\boldsymbol{x})}{\tau}\right)$. Then,

$$F_\tau(\boldsymbol{u}) = \frac{1}{(2\pi\tau)^{\frac{d}{2}} \mu(\Omega)^{\frac{1}{2}}} \int f(\boldsymbol{x}) \exp\left(\frac{-i(\boldsymbol{u}\cdot\boldsymbol{x})}{\tau}\right) \mathrm{d}\boldsymbol{x}.$$

Letting $\boldsymbol{v} = \dfrac{\boldsymbol{u}}{\tau}$ and $G(\boldsymbol{v}) = F_\tau(\boldsymbol{u})$, we get

$$(\tau)^{\frac{d}{2}} \mu(\Omega)^{\frac{1}{2}} G(\boldsymbol{v}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int f(\boldsymbol{x}) \exp(-i\boldsymbol{v}\cdot\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

As f is ℓ^1 integrable, by Parseval's Theorem (see [21]) we have

$$\int |f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = \tau^d \mu(\Omega) \int |F_\tau(\boldsymbol{v}\tau)|^2 \mathrm{d}\boldsymbol{v} = \mu(\Omega) \int |F_\tau(\boldsymbol{u})|^2 \mathrm{d}\boldsymbol{u}.$$

By noting that

$$\int |f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = \int_\Omega \left|\exp\left(\frac{iS(\boldsymbol{x})}{\tau}\right)\right|^2 \mathrm{d}\boldsymbol{x} = \mu(\Omega),$$

the result follows.

5) Proof of Cross Factor Nullifier Lemma

Proof. Let $p_{k,l}(\boldsymbol{u})$ denote the phase of the exponential in the cross term (excluding the terms with constant signatures), *i.e.*,

$$\begin{aligned} p_{k,l}(\boldsymbol{u}) &= \Psi(\boldsymbol{x}_k(\boldsymbol{u}); \boldsymbol{u}) - \Psi(\boldsymbol{x}_l(\boldsymbol{u}); \boldsymbol{u}) \\ &= S(\boldsymbol{x}_k(\boldsymbol{u})) - S(\boldsymbol{x}_l(\boldsymbol{u})) - \boldsymbol{u}\cdot(\boldsymbol{x}_k(\boldsymbol{u}) - \boldsymbol{x}_l(\boldsymbol{u})). \end{aligned} \tag{A.1}$$

Its gradient with respect to $\boldsymbol{u}$ equals

$$\nabla p_{k,l}(\boldsymbol{u}) = J_{\boldsymbol{x}_k}\left[\nabla S(\boldsymbol{x}_k(\boldsymbol{u})) - \boldsymbol{u}\right] - J_{\boldsymbol{x}_l}\left[\nabla S(\boldsymbol{x}_l(\boldsymbol{u})) - \boldsymbol{u}\right] - \boldsymbol{x}_k(\boldsymbol{u}) + \boldsymbol{x}_l(\boldsymbol{u})$$

where $J_{\boldsymbol{x}_k}$ is the Jacobian of $\boldsymbol{x}(\boldsymbol{u})$ at $\boldsymbol{x}_k$ whose $(i,j)^{\text{th}}$ term equals $\dfrac{\partial x_j}{\partial u_i}$ (with a similar expression for $J_{\boldsymbol{x}_l}$). Since $\nabla S(\boldsymbol{x}_k(\boldsymbol{u})) = \nabla S(\boldsymbol{x}_l(\boldsymbol{u})) = \boldsymbol{u}$, we get $\nabla p_{k,l}(\boldsymbol{u}) = \boldsymbol{x}_l(\boldsymbol{u}) - \boldsymbol{x}_k(\boldsymbol{u}) \neq 0$. This means that the phase function of the exponential in the statement of the lemma is *non-stationary* and hence *does not contain any stationary points of the first kind.* Let

$$q_{k,l}(\boldsymbol{u}) = \frac{1}{\sqrt{\left|\det\left(\mathcal{H}_{x_k(\boldsymbol{u})}\right)\right|}\sqrt{\left|\det\left(\mathcal{H}_{x_l(\boldsymbol{u})}\right)\right|}}. \tag{A.2}$$

Since $\nabla p_{k,l} \neq 0$, consider the vector field $\boldsymbol{f}_{k,l}(\boldsymbol{u}) = \dfrac{\nabla p_{k,l}(\boldsymbol{u})}{\left\|\nabla p_{k,l}(\boldsymbol{u})\right\|^2} q_{k,l}(\boldsymbol{u})$ and as before note that

$$
\begin{aligned}
&\exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right) q_{k,l}(\boldsymbol{u}) \\
&= i\tau\left[\nabla \cdot \boldsymbol{f}_{k,l}(\boldsymbol{u})\right]\exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right) - i\tau\nabla \cdot \left[\boldsymbol{f}_{k,l}(\boldsymbol{u})\exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right)\right]
\end{aligned} \tag{A.3}
$$

where $\nabla \cdot$ is the divergence operator. Inserting (A.3) in the second line of (3.13), integrating over $\overline{\mathcal{N}_\eta(\boldsymbol{u}_0)}$, and applying the divergence theorem we get

$$
\begin{aligned}
&\int_{\overline{\mathcal{N}_\eta(\boldsymbol{u}_0)}} \exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right) q_{k,l}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \\
&= i\tau \int_{\overline{\mathcal{N}_\eta(\boldsymbol{u}_0)}} \left[\nabla \cdot \boldsymbol{f}_{k,l}(\boldsymbol{u})\right]\exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right)\mathrm{d}\boldsymbol{u} \\
&\quad - i\tau \int_{\partial \mathcal{N}_\eta(\boldsymbol{u}_0)} \left(\boldsymbol{f}_{k,l} \cdot \boldsymbol{n}\right)\exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u}(\boldsymbol{v}))\right)\mathrm{d}\boldsymbol{v}.
\end{aligned} \tag{A.4}
$$

Here $\boldsymbol{n}$ is the unit outward normal to the positively oriented boundary $\overline{\partial \mathcal{N}_\eta(\boldsymbol{u}_0)}$ parameterized by $\boldsymbol{v}$. In the right side of (A.4), notice that all terms inside the integral are bounded. The factor τ outside the integral ensures that

$$\lim_{\tau \to 0} \int_{\overline{\mathcal{N}_\eta(\boldsymbol{u}_0)}} \exp\left(\frac{i}{\tau} p_{k,l}(\boldsymbol{u})\right) q_{k,l}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} = 0.$$

Appendix B. Well-Behaved Function on the Boundary

One of the foremost requirements for Theorem 3.2 to be valid is that the function $\Psi(\boldsymbol{x};\boldsymbol{u}) = S(\boldsymbol{x}) - \boldsymbol{u} \cdot \boldsymbol{x}$ have a *finite* number of stationary points of the second kind on the boundary for almost all $\boldsymbol{u}$. The stationary points of the second kind are the critical points on the boundary $\Gamma = \partial\Omega$ where a level curve $\Psi(\boldsymbol{x};\boldsymbol{u}) = c$ touches Γ for some constant c [9] [10]. Contributions from the second kind are generally $O\left(\tau^{\frac{d+1}{2}}\right)$, but an infinite number of them could produce a combined effect of $O\left(\tau^{\frac{d}{2}}\right)$, tantamount to a stationary point of the first kind [9]. If so, we need to account for the contribution from the boundary which could in effect invalidate our theorem and therefore our entire approach. However, the condition for the infinite occurrence of stationary points of the second kind is so restrictive that for all practical purposes they can be ignored. If the given function S is well-behaved on the boundary in the sense explained below, these thorny issues can be sidestepped. Furthermore, as we will be integrating over $\boldsymbol{u}$ to remove the cross-phase factors, it suffices that the aforementioned finiteness condition be satisfied for *almost* all $\boldsymbol{u}$ instead of for all $\boldsymbol{u}$.

Let the location $x \in \Gamma$ be parameterized by the variable y, *i.e.*, $x(y)$. Let $Q(x)$ denote the Jacobian matrix of $x(y)$ whose $(i, j)^{\text{th}}$ entry is given by

$$Q_{ij}(x) = \frac{\partial x_j}{\partial y_i}.$$

Stationary points of the second kind occur at locations x where $\nabla \Psi(x(y); u) = 0$, which translates to

$$Q(x)(\nabla S - u) = 0. \tag{B.1}$$

This leads us to define the notion of a well-behaved function on the boundary.

Definition: A function S is said to be well-behaved on the boundary provided (B.1) is satisfied only at a *finite* number of boundary locations for almost all $u \in \mathcal{C}$.

The definition immediately raises the following questions: 1) *Why is the assumption of a well behaved S weak*? and 2) *Can the well-behaved condition imposed on S be easily satisfied in all practical scenarios*? Recall that the finiteness of premise (B.1) entirely depends on the behavior of the function S on the boundary Γ. Scenarios can be manually handcrafted where the finiteness assumption is violated and (B.1) is forced to be satisfied at all locations. Hence it is meaningful to ask: *What stringent conditions are required to incur an infinite number of stationary points on the boundary?* We would like to convince the reader that in all practical scenarios, S will contain only a finite number of stationary points on the boundary and hence it is befitting to assume that the function S is well-behaved on the boundary. The reader should bear in mind that our explanation here is *not* a formal proof but an intuitive reasoning of why the well-behaved condition imposed on S is reasonable.

To streamline our discussion, we consider the special case where the boundary Γ is composed of a sequence of hyper-planes as any smooth boundary can be approximated to a given degree of accuracy by a finite number of hyper-planes. On any given hyperplane, $Q(x)$ remains fixed. Recall that from the outset, we omit the set $\mathcal{C}$ (*i.e.*, $u \notin \mathcal{C}$) which includes the image under ∇S of the boundary $\Gamma = \partial \Omega$. Hence $\nabla S \neq u$ for any point $x \in \Gamma$ for $u \notin \mathcal{C}$. Since the rank of Q is $d - 1$ and $\nabla S - u$ is required to be orthogonal to all the $d - 1$ rows of Q for condition 33 to hold, $\nabla S - u$ is confined to a 1-D subspace. So if we enforce ∇S to vary smoothly on the hyperplane and not be constant, we can circumvent the occurrence of an infinite number of stationary points of the second kind for all u. Additionally, we can safely disregard the characteristics of the function S at the intersection of these hyperplanes as they form a set of measure zero. To press this point home, we now formulate the worst possible scenario where ∇S is a constant vector t. Let $\mathcal{D}$ denote a portion of Γ where $\nabla S = t$. Let $u = u_0$ and $u = u_1$ result in infinite number of stationary points of the second kind on $\mathcal{D}$. As $\nabla S - u$ is limited to a 1-D subspace, we must have $t - u_1 = \lambda(t - u_0)$ for some $\lambda \neq 0$, *i.e.*, $u_1 = (1 - \lambda)t + \lambda u_0$. So in any given region of Γ, there is at most a 1-D subspace (measure zero) of u

which results in an infinite number of stationary points of the second kind in that region. Our well-behaved condition is then equivalent to assuming that the number of planar regions on the boundary where ∇S is constant is finite.

The boundary condition is best exemplified with a 2D example. Consider a line segment on the boundary $x_2 = mx_1 + b$. Without loss of generality, assume the parameterization $y = x_1$. Then $Q = \begin{bmatrix} 1 \\ m \end{bmatrix}$. Equation (B.1) can be interpreted as $S_1 + mS_2 = u_1 + mu_2$ where, $S_i = \dfrac{\partial S}{\partial x_i}$. So if we plot the sum $S_1 + mS_2$ for points along the line, the requirement reduces to the function $S_1 + mS_2$ not oscillating an infinite number of times around an infinite number of ordinate locations $u_1 + mu_2$. It is easy to see that the imposed condition is indeed weak and is satisfied by almost all smooth functions. Consequently, we can affirmatively conclude that the enforced well-behaved constraint (B) does not impede the usefulness and application of our wave function method for estimating the joint gradient densities of smooth functions.