On integral theorems and their statistical properties

Nhat Ho¹ and Stephen G. Walker¹,²

¹Department of Statistics and Data Sciences, and ²Department of Mathematics, University of Texas at Austin, Austin, TX, USA

We introduce a class of integral theorems based on cyclic functions and Riemann sums approximating integrals. The Fourier integral theorem, derived as a combination of a transform and inverse transform, arises as a special case. The integral theorems provide natural estimators of density functions via Monte Carlo methods. Assessment of the quality of the density estimators can be used to obtain optimal cyclic functions, alternatives to the sin function, which minimize square integrals. Our proof techniques rely on a variational approach in ordinary differential equations and the Cauchy residue theorem in complex analysis.

1. Introduction

The Fourier integral theorem, see for example [1,2], is a remarkable result. For all real integrable and continuous function $m : \mathbb{R}^d \rightarrow \mathbb{R}$, it yields

$$m(y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \cos(s^T(y - x)) m(x) \, dx \, ds. \quad (1.1)$$

The derivation of this result is from a combination of the Fourier and inverse Fourier transforms. As far as we are aware, it is only the cos function that yields such an integral theorem, a point supported by the paper [3].

If we write equation (1.1) as $m(y) = \int_{\mathbb{R}^d} \psi(y - x) m(x) \, dx$, we see that we have the opportunity for obtaining a natural Monte Carlo estimator of $m(y)$ using a sample from $m(\cdot)$, provided it is integrable on $\mathbb{R}^d$.

The relevance of the Fourier integral theorem towards Monte Carlo estimators was observed by [4], and used in several statistics and machine learning applications, such as multivariate density estimation, non-parametric mode clustering and modal regression, quantile regression and
generative model. The methodological benefits of the Fourier integral theorem come from rewriting equation (1.1) as

\[ m(y) = \lim_{R \to \infty} \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^d \frac{\sin(R(y_j - x_j))}{y_j - x_j} m(x) \, dx, \]  

(1.2)

where \( y = (y_1, \ldots, y_d) \) and \( x = (x_1, \ldots, x_d) \). To see how (1.2) follows from (1.1), at least when \( d = 1 \), we can write the first equation as

\[ \frac{1}{\pi} \lim_{R \to \infty} \int_0^R \int_{\mathbb{R}} \cos(s(y - x)) \, m(x) \, ds \, dx \]

and now implement the \( s \) integral. Equation (1.2) contains an important insight: even though we have certain dependent structures in \( m(x) \), by taking the products of (independent) sin functions, the Monte Carlo estimators are able to capture the dependent structure. This eliminates the cumbersome and delicate procedure of choosing a covariance matrix to guarantee good practical performance of estimators based on, e.g. multi-variate Gaussian kernels [5–7]. Further, it clearly extends the theoretical work done in the one-dimensional case by [8,9].

To motivate the theory developed in the paper, here we present a statistical application for the integral theorem. We consider the density function \( p(y | \theta) = \int N(y | 0, \nu^2) \, dG(v | 2, \theta) \) with \( \theta \) unknown and to be estimated. Here, \( G \) represents the Gamma distribution. The integral to obtain a likelihood function for \( \theta \) is not tractable; it is of the form \( \int v^{-1} e^{-1/2 \nu^2} e^{-v \theta} \, dv \). However, it is tractable for \( y = 0 \). This motivates the use of the integral theorem to estimate the density at \( p(0 | \theta_0) \) and hence to estimate \( \theta \).

Suppose \( y_{1:n} \) are a sample from the model with true value \( \theta_0 \). Then based on the Fourier integral theorem the estimate of \( p(0 | \theta_0) \) is given by: \( \hat{p}(0 | \theta_0) = n(\pi)^{-1} \sum_{i=1}^n \sin(Ry_i)/y_i \) and the correct value of \( p(0 | \theta) = \theta/\sqrt{2\pi} \). Therefore, the estimator of \( \theta \) is \( \hat{\theta} = \hat{p}(0 | \theta_0)/\sqrt{2\pi} \). By way of illustration, we took \( n = 100 \) with \( \theta_0 = 4 \). The choice of \( R \) was 50 and we repeated the experiments 1000 times. A histogram of the 100 estimators is presented in figure 1. The mean value is 3.75.

Clearly, this example is a simple case, and the \( \theta \) could be estimated using numerical routines such as an expectation-maximization (EM) algorithm [10]. Nevertheless, even an EM algorithm is difficult to implement in this case. Further examples of the integral theorem appear in [11] where it is used to compute Bayesian marginal likelihoods.

Another demonstration of the integral theorem highlights the lack of need for a covariance matrix when estimating with a bivariate sample. We take \( f(x, y) \) to be a banana-shaped density; see figure 2. We use these samples to estimate \( f(0, 0) \), the true value of which is 0.347. In figure 3, we show the histogram of 100 experiments using the Fourier integral theorem (upper figure).
Figure 2. Sample from bivariate banana-shaped density.

Figure 3. Histogram of estimators of $f(0, 0)$ using Fourier integral theorem (a) and an independent Gaussian kernel density (b).

and an independent Gaussian kernel density estimator using the recommended bandwidth for the appropriate sample size. In particular, for $n = 100000$, we take $R = 20$ and the bandwidth for each Gaussian kernel to be Silverman’s rule of thumb, e.g. $h_x = n^{-1/6} \sqrt{S_{xx}}$, where $S_{xx}$ is the sample variance of the $x$ values. The histogram from the Fourier integral theorem accurately picks out the true value for $f(0, 0)$, in fact the mean is precisely 0.347, whereas the kernel density estimator is clearly biased. This shows that whereas a covariance matrix is not required for the Fourier integral theorem, it is recommended for the Gaussian kernel density [5–7,12]. The reason is that the Fourier estimator is using an unbiased Monte Carlo sampler whereas the Gaussian kernel density estimator is not.

These illustrations motivate the need for a study of not just the Fourier integral theorem, but integral theorems in general. The aim in this paper is to introduce a general class of integral theorems to be used as Monte Carlo estimators with applications in statistics and machine learning. The reasoning is as follows: for different function $m$ there will possibly be alternative integral theorems that provide better Monte Carlo estimators. Indeed, we do find new integral theorems that provide superior Monte Carlo estimators.
We do not come at this new class of integral theorems using transforms and inverse transforms, but rather use the novel combination of two well-known objects: a cyclic function that integrates to 0 over each cyclic interval, and a Riemann sum approximation to an integral.

In particular, we define almost everywhere differentiable cyclic functions $\phi : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\int_0^{2\pi} \phi(x) \, dx = 0 \quad \text{and} \quad \int_{\mathbb{R}} \frac{\phi(x)}{x} \, dx = 1. \tag{1.3}$$

The Fourier integral theorem corresponds to

$$\phi(x) = \frac{1}{\pi} \sin x. \tag{1.4}$$

Then, via Riemann sums approximating integrals, we demonstrate that

$$m(y) = \lim_{R \to \infty} \int_{\mathbb{R}} \prod_{j=1}^{d} \phi(R(y_j - x_j)) \frac{y_j - x_j}{m(x) \, dx}. \tag{1.5}$$

Similar to the Fourier integral theorem, (1.2), the general integral theorem in equation (1.5) is also able to automatically preserve the dependence structures in the function $m$. With our finding of large classes of integral functions, the question posed is which kernel $\phi$ has optimal properties in terms of estimation.

In this work, we specifically answer this question in the context of the kernel density estimation problem [8,12–21]. Indeed, the kernel density estimator based on equation (1.5) would be given by

$$\hat{m}_R(y) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \phi(R(y_j - X_{ij})) \frac{y_j - X_{ij}}{m(x) \, dx}. \tag{1.6}$$

where $X_1, \ldots, X_n \in \mathbb{R}^d$ represent the sample from the density function $m$ and a finite $R$ is required for the smoothing. We study upper bounds for bias of the estimator $\hat{m}_R$ based on $R$ and the sample size $n$ in theorem 2.2 and corollary 2.3.

In order to find the optimal kernel $\phi$, we use asymptotic mean integrated square error, which is a key property determining the quality of an estimator; see for example [5,22]. To ease the findings, we specifically consider the univariate settings, i.e. $d = 1$ and use the following two terms stemming from that error to determine the optimal kernel: the term

$$\int_{\mathbb{R}} \left( \frac{\phi(x)}{x} \right)^2 \, dx, \tag{1.7}$$

which provides an upper bound on the variance of density estimator $\hat{m}_R$; and the term

$$\int_{\mathbb{R}} \left( \frac{\phi(x)}{x} \right)^2 m(x) \, dx, \tag{1.8}$$

which yields more precise asymptotic behaviours of the variance of the density estimator $\hat{m}_R$ than that from equation (1.7). To support the idea of only considering the variances, we show that the bias does not depend on $\phi$. We demonstrate that by minimizing the first term (1.7) subject to the constraints (1.3), the optimal kernel $\phi$ is the sin function (1.4) in the Fourier integral theorem. This is achieved via a variational approach in ordinary differential equations. On the other hand, by using the Cauchy residue theorem in complex analysis, we prove that the optimal kernel is not the sin kernel when minimizing the second term (1.8) subject to the constraints (1.3). It also demonstrates the usefulness in finding other integral theorems to the Fourier integral theorem.

The organization of the paper is as follows. In §2, we first revisit the Fourier integral theorem and establish the bias of its density estimator via Riemann sums approximating integrals. Then, using the insight from that theorem, we introduce a general class of integral theorems that possess similar approximation errors. After deriving our class of integral theorems, in §3, we study optimal kernels that minimize either the problem (1.7) or the problem (1.8) subject to the constraints (1.3). Finally, we conclude the paper with a few discussions in §4.
2. Integral theorems

We first study the bias of the kernel density estimator (1.6) or equivalently approximation property of the Fourier integrals theorem via the Riemann sums approximating integral theorem in §2a. Then, using the insight from that result, we introduce a general class of integral theorem that possesses similar approximation behaviour to the Fourier integral theorem, in §2b.

(a) The Fourier integral theorem revisited

Before going into the details of the general integral theorem, we reconsider the approximation property of the Fourier integral theorem. In [4], the authors use the tail behaviour of the Fourier transform of the function \( m(\cdot) \) to characterize an approximation error of the Fourier integral theorem when truncating one of the integrals. However, the technique in the proof is inherently based on properties of the sin kernel and is non-trivial to extend to other choices of cyclic function; examples of such functions are provided in §2b.

In this paper, we provide insight into the approximation error of the Fourier integral theorem via the Riemann sums approximating integral theorem. In particular, we would like to show that when the function \( m(\cdot) \) is differentiable, uniformly continuous up to the \( k \)-th order derivative of \( g \), the integrals 

\[
\int_{\mathbb{R}^d} \prod_{j=1}^d \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} m(x) \, dx.
\]

are finite for all \( 0 \leq k \leq K \). To obtain the bound, we start with the following definition of the class of univariate functions that we use throughout our study.

Definition 2.1. The univariate function \( f(\cdot) \) is said to belong to the class \( T^K(\mathbb{R}) \) if for any \( y \in \mathbb{R} \), the function \( g(x) = (f(x) - f(y))/(x - y) \) satisfies the following conditions:

1. The function \( g \) is differentiable, uniformly continuous up to the \( K \)-th order, and the limits

\[
\lim_{|x| \to +\infty} |g^{(k)}(x)| = 0 \text{ for any } 0 \leq k \leq K
\]

where \( g^{(k)}(\cdot) \) denotes the \( k \)-th order derivative of \( g \);

2. The integrals \( \int_{\mathbb{R}} |g^{(k)}(x)| \, dx \) are finite for all \( 0 \leq k \leq K \).

Note that, for the function \( g \) in definition 2.1, for any \( y \in \mathbb{R} \) when \( x = y \), we choose \( g(y) = f^{(1)}(y) \).

Based on definition 2.1, we now state the following result.

Theorem 2.2. Assume that the univariate functions \( m_j \in T^{K_j}(\mathbb{R}) \) for any \( 1 \leq j \leq d \) where \( K_1, \ldots, K_d \) are given positive integer numbers. Then, if we have \( m(x) = \prod_{j=1}^d m_j(x_j) \) or \( m(x) = \sum_{j=1}^d m_j(x_j) \) for any \( x = (x_1, \ldots, x_d) \), there exist universal constants \( C \) and \( \tilde{C} \) depending on \( d \) such that as long as \( R \geq C \) we obtain 

\[
|m_R(y) - m(y)| \leq \tilde{C}/R^K, \text{ where } K = \min_{1 \leq j \leq d} (K_j).
\]

Proof. We first prove the result when \( d = 1 \). In particular, we would like to show that when the function \( m \in T^K(\mathbb{R}) \), there exists a universal constant \( C > 0 \) such that we have 

\[
|m_R(y) - m(y)| \leq C/R^K.
\]

In fact, from the definition of \( m_R(y) \) in equation (2.1), we have

\[
|m_R(y) - m(y)| = \left| \frac{1}{\pi} \int_{\mathbb{R}} \frac{\sin(R(y - x))}{(y - x)} (m(x) - m(y)) \, dx \right|.
\]
For simplicity of the presentation, for any $y \in \mathbb{R}$, we write $g(x) = (m(x) - m(y))/(x - y)$ for all $x \in \mathbb{R}$. Then, we can rewrite the above equality as

$$|m_R(y) - m(y)| = \left| \frac{1}{\pi} \int_\mathbb{R} \sin(R(y - x))g(x) \, dx \right|$$

$$= \left| \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \int_{y+(2\pi(k+1)/R)}^{y+(2\pi k/R)} \sin(R(y - x))g(x) \, dx \right|.$$ 

Invoking the change of variables $x = y + ((t + 2\pi k)/R)$, the above equation becomes

$$|m_R(y) - m(y)| = \left| \frac{1}{\pi R} \sum_{k=-\infty}^{\infty} \int_{[0,2\pi]} \sin(t) \cdot g \left( y + \frac{t + 2\pi k}{R} \right) \, dt \right|.$$ 

(2.2)

Since $m \in T^K(\mathbb{R})$, the function $g$ is differentiable up to the Kth order. Therefore, using a Taylor expansion up to the Kth order, leads to $g(y + (t + 2\pi k)/R)$ equal to

$$\sum_{\alpha \leq K-1} \frac{t^\alpha}{R^\alpha \alpha!} g^{(\alpha)} \left( y + \frac{2\pi k}{R} \right) + \frac{t^K}{R^K(K-1)!} \int_0^1 (1 - \xi)^{K-1} g^{(K)} \left( y + \frac{2\pi k}{R} + \frac{\xi t}{R} \right) \, d\xi.$$ 

Plugging the above Taylor expansion into equation (2.2), we have

$$|m_R(y) - m(y)| = \left| \frac{1}{\pi R} \sum_{\ell=0}^{K} A_\ell \right| \leq \frac{1}{\pi R} \sum_{\ell=0}^{K} |A_\ell|,$$ 

(2.3)

where, for $\ell \in \{0, 1, \ldots, K-1\}$, we define

$$A_\ell = \frac{1}{R^\ell} \int_{[0,2\pi]} \left( \frac{t^\ell}{\ell!} \sin(t) \right) \, dt \left( \sum_{k=-\infty}^{\infty} g^{(\ell)} \left( y + \frac{2\pi k}{R} \right) \right)$$

and

$$A_K = \sum_{k=-\infty}^{\infty} \int_{[0,2\pi]} \sin(t) \left( \frac{t^K}{R^K(K-1)!} \int_0^1 (1 - \xi)^{K-1} g^{(K)} \left( y + \frac{2\pi k}{R} + \frac{\xi t}{R} \right) \, d\xi \right) \, dt.$$ 

We now find a bound for $|A_\ell|$ for $\ell \in \{0, 1, \ldots, K-1\}$; we will demonstrate that

$$\left| \sum_{k=-\infty}^{\infty} g^{(\ell)} \left( y + \frac{2\pi k}{R} \right) \right| \leq \frac{C}{R^{K-(\ell+1)}}, \text{ for all } \ell \in \{0, 1, \ldots, K-1\}$$ 

(2.4)

where C is some universal constant. To obtain these bounds, we will use an inductive argument on $\ell$. We first start with $\ell = K - 1$. In fact, we have

$$\left| \sum_{k=-\infty}^{\infty} g^{(K-1)} \left( y + \frac{2\pi k}{R} \right) \left( \frac{(2\pi)}{R} \right) - \int_\mathbb{R} g^{(K-1)}(x) \, dx \right|$$

$$= \left| \sum_{k=-\infty}^{\infty} \int_{y+(2\pi(k+1)/R)}^{y+(2\pi k/R)} g^{(K-1)}(x) - g^{(K-1)} \left( y + \frac{2\pi k}{R} \right) \, dx \right|$$

$$\leq \sum_{k=-\infty}^{\infty} \int_{y+(2\pi(k+1)/R)}^{y+(2\pi k/R)} g^{(K-1)}(x) - g^{(K-1)} \left( y + \frac{2\pi k}{R} \right) \, dx.$$ 

An application of Taylor expansion leads to

$$g^{(K-1)}(x) = g^{(K-1)} \left( y + \frac{2\pi k}{R} \right) + \left( x - y - \frac{2\pi k}{R} \right)$$

$$\times \int_0^1 g^{(K)} \left( (1 - \xi) \left( y + \frac{2\pi k}{R} \right) + \xi x \right) \, d\xi.$$
Now for any $\xi \in [0, 1]$ and $x \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]$, we have

\[ |g^{(K)}((1 - \xi)(y + 2\pi k/R) + \xi x)| \leq \sup_{t \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]} |g^{(K)}(t)|. \]

Collecting the above results, we find that

\[ \sum_{k = -\infty}^{\infty} \int_{y + (2\pi (k + 1)/R)}^{y + (2\pi k/R)} |g^{(K-1)}(x) - g^{(K-1)}(y + (2\pi k/R))| \, dx \leq \sum_{k = -\infty}^{\infty} \int_{y + (2\pi (k + 1)/R)}^{y + (2\pi k/R)} \left( x - y - \frac{2\pi k}{R} \right) \, dx \sup_{t \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]} |g^{(K)}(t)| \]

\[ = \frac{2\pi^2}{R^2} \sum_{k = -\infty}^{\infty} \sup_{t \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]} |g^{(K)}(t)|. \]

Using a Riemann sums approximating integrals theorem, we have

\[ \lim_{R \to \infty} \sum_{k = -\infty}^{\infty} \sup_{t \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]} |g^{(K)}(t)| \frac{2\pi}{R} = \int_{\mathbb{R}} |g^{(K)}(x)| \, dx < \infty, \]

where the finite value of the integral is due to the assumption that $m \in \mathcal{T}^K(\mathbb{R})$. Furthermore, the above limit is uniform in terms of $y$ as $g^{(K)}$ is uniformly continuous. Collecting the above results, there exists a universal constant $C$ such that as long as $R \geq C$, the following inequality holds:

\[ \frac{2\pi^2}{R^2} \sum_{k = -\infty}^{\infty} \sup_{t \in [y + (2\pi k/R), y + (2\pi (k + 1)/R)]} |g^{(K)}(t)| \leq \frac{C_1}{R}, \]

where $C_1$ is some universal constant. Combining all of the previous results, we obtain

\[ \left| \sum_{k = -\infty}^{\infty} g^{(K-1)}(y + \frac{2\pi k}{R}) \frac{(2\pi)}{R} - \int_{\mathbb{R}} g^{(K-1)}(x) \, dx \right| \leq \frac{C}{R}. \]

Since $m \in \mathcal{T}^K(\mathbb{R})$, using integration by parts, we get $\int_{\mathbb{R}} g^{(K-1)}(x) \, dx = 0$. Therefore, we obtain the conclusion of equation (2.4) when $\ell = K - 1$.

Now assume that the conclusion of equation (2.4) holds for $1 \leq \ell \leq K - 1$. We will prove that the conclusion also holds for $\ell - 1$. With a similar argument to the setting $\ell = K - 1$, we obtain

\[ \left| \sum_{k = -\infty}^{\infty} g^{(\ell)}(y + \frac{2\pi k}{R}) \frac{(2\pi)}{R} - \int_{\mathbb{R}} g^{(\ell)}(x) \, dx \right| \]

\[ = \left| \sum_{k = -\infty}^{\infty} \int_{y+(2\pi (k+1)/R)}^{y+(2\pi k/R)} \left( g^{(\ell)}(x) - g^{(\ell)}(y + \frac{2\pi k}{R}) \right) \, dx \right|. \tag{2.5} \]

Using a Taylor expansion, we have

\[ g^{(\ell)}(x) = g^{(\ell)}(y + \frac{2\pi k}{R}) + \sum_{\alpha \leq K-1-\ell} \frac{(x - y - (2\pi k/R))^\alpha}{\alpha!} g^{(\ell + \alpha)}(y + \frac{2\pi k}{R}) \]

\[ + \frac{(x - y - (2\pi k/R))^{K-1}}{(K - \ell - 1)!} \int_0^1 (1 - \xi)^{K-\ell-1} \xi^{K-1} (1 - \xi) \left( y + \frac{2\pi k}{R} + \xi x \right) \, d\xi. \]

Plugging the above Taylor expansion into equation (2.5), we find that

\[ \sum_{k = -\infty}^{\infty} \int_{y+(2\pi (k+1)/R)}^{y+(2\pi k/R)} \left| g^{(\ell)}(x) - g^{(\ell)}(y + \frac{2\pi k}{R}) \right| \, dx \leq S_1 + S_2, \]
where $S_1$ and $S_2$ are defined as follows:

$$S_1 = \sum_{\alpha \leq K-1-\ell} \sum_{k=-\infty}^{\infty} \left( \int_{y+(2\pi k/R)}^{y+(2\pi (k+1)/R)} \frac{(x-y-(2\pi k/R))^\alpha}{\alpha!} \, dx \right) g^{(\ell+\alpha)} \left( y + \frac{2\pi k}{R} \right)$$

$$= \sum_{\alpha \leq K-1-\ell} \sum_{k=-\infty}^{\infty} \frac{(2\pi)^{\alpha+1}}{R^{\alpha+1}(\alpha+1)!} g^{(\ell+\alpha)} \left( y + \frac{2\pi k}{R} \right) |

$$S_2 = \sum_{k=-\infty}^{\infty} \left( \int_{y+(2\pi k/R)}^{y+(2\pi (k+1)/R)} \frac{(x-y-(2\pi k/R))^{K-\ell}}{(K-\ell-1)!} \right) x \left( 1 - \frac{2\pi k}{R} \right) d\xi \right) dx \right) dx

\leq \sum_{k=-\infty}^{\infty} \sup_{t \in [y+(2\pi k/R), y+(2\pi (k+1)/R)]} \left| g^{(K)}(t) \right| \frac{2\pi K^{K-\ell+1}}{R^{K-\ell+1}(K-\ell+1)!} \left| g^{(K)}(t) \right| \leq C_3 R^{K-\ell},

when $R$ is sufficiently large where $C_3$ is some universal constant.

An application of the triangle inequality and the hypothesis of the induction argument shows that

$$S_1 \leq \sum_{\alpha \leq K-1-\ell} \frac{(2\pi)^{\alpha+1}}{R^{\alpha+1}(\alpha+1)!} \sum_{k=-\infty}^{\infty} g^{(\ell+\alpha)} \left( y + \frac{2\pi k}{R} \right) \leq C_2 R^{K-\ell},$$

where $C_2$ is some universal constant.

Similar to the setting $\ell = K - 1$, an application of the Riemann sums approximating integrals theorem leads to

$$S_2 \leq (K-\ell) \frac{2\pi^{K-\ell+1}}{R^{K-\ell+1}(K-\ell+1)!} \sum_{k=-\infty}^{\infty} \sup_{t \in [y+(2\pi k/R), y+(2\pi (k+1)/R)]} \left| g^{(K)}(t) \right| \leq C_3 R^{K-\ell},$$

for some universal constant $C_3$. As $\int_{-\infty}^{\infty} g^{(\ell)}(x) \, dx = 0$, the above inequality leads to the conclusion of equation (2.4) for $1 \leq \ell \leq K - 1$. As a consequence, we obtain the conclusion of equation (2.4) for all $\ell \in \{0, 1, \ldots, K - 1\}$.

Given equation (2.4), an application of the triangle inequality leads to

$$|A_\ell| \leq \frac{1}{R^{\ell}} \left| \int_{[0,2\pi]} \frac{t^\ell \sin(t)}{\ell!} \, dt \right| \sum_{k=-\infty}^{\infty} g^{(\ell)} \left( y + \frac{2\pi k}{R} \right) \leq \frac{\widetilde{C}}{R^{K-1}} \quad (2.6)$$

for all $0 \leq \ell \leq K - 1$.

We now find a bound for $|A_K|$. A direct application of the triangle inequality leads to the following bound of $|A_K|$:

$$|A_K| \leq \int_{[0,2\pi]} \frac{|\sin(t)|^K}{R^{K(K-1)!}} \, dt \left( \sum_{k=-\infty}^{\infty} \int_{0}^{1} (1 - \xi)^{K-1} \left| g^{(K)} \left( y + \frac{2\pi k}{R} + \frac{\xi t}{R} \right) \right| \, d\xi \right).$$
For any $\xi \in [0, 1]$ and $t \in [0, 2\pi)$, we have
\[
\left| g^{(K)} \left( y + \frac{2\pi k}{R} + \frac{\xi t}{R} \right) \right| \leq \sup_{x \in [y+(2\pi k/R), y+(2\pi(k+1)/R)]} |g^{(K)}(x)|.
\]

Putting the above inequalities together, we find that
\[
|A_K| \leq \int_{[0,2\pi)} \left| \frac{\sin(t) t^K}{R^K K!} \right| dt \left( \sum_{k=-\infty}^{\infty} \sup_{x \in [y+(2\pi k/R), y+(2\pi(k+1)/R)]} |g^{(K)}(x)| \right).
\]

From the Riemann sums approximating integrals theorem, we obtain
\[
\frac{(2\pi)}{R^{1-K}|A_K|} \leq C_1 \int_{[0,2\pi)} \left| \frac{\sin(t) t^K}{R^K K!} \right| dt,
\]
where $C_1$ is some universal constant. Collecting the above results, we conclude that $|A_K| \leq \tilde{C}/R^{K-1}$ for some constant $\tilde{C}$. Putting this bound and (2.6) into equation (2.3), we obtain the conclusion of the theorem when $d = 1$.

We now provide the proof for a general dimension $d$. When $m(x) = \sum_{j=1}^{d} m_j(x_j)$ for any $x = (x_1, \ldots, x_d)$: From the definition of $m_R(y)$ in equation (2.1), we have
\[
|m_R(y) - m(y)| = \left| \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} (m(x) - m(y)) \, dx \right|
\]
\[
= \left| \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} \left( \sum_{j=1}^{d} m_j(x_j) - \sum_{j=1}^{d} m_j(y_j) \right) \, dx \right|
\]
\[
\leq \sum_{j=1}^{d} \left| \frac{1}{\pi} \int_{\mathbb{R}} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} (m_j(x_j) - m_j(y_j)) \, dx \right|.
\]

Since $m_j \in T^K_j(\mathbb{R})$ for $1 \leq j \leq d$, an application of the result of theorem 2.2 when $d = 1$ leads to
\[
\left| \frac{1}{\pi} \int_{\mathbb{R}} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} (m_j(x_j) - m_j(y_j)) \, dx \right| \leq \frac{C_j}{R^K_j},
\]
where $C_j$ are universal constants. Putting the above results together, we obtain the conclusion of the theorem when $m$ is the summation of the functions $m_1, m_2, \ldots, m_d$.

When $m(x) = \prod_{j=1}^{d} m_j(x_j)$ for any $x = (x_1, \ldots, x_d)$: Similar to the argument when $m$ is the summation of $m_1, m_2, \ldots, m_d$, we have $|m_R(y) - m(y)|$ equal to
\[
\left| \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} \left( \prod_{j=1}^{d} m_j(x_j) - \prod_{j=1}^{d} m_j(y_j) \right) \, dx \right|
\]
\[
= \left| \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} \left( \sum_{\ell=0}^{d-1} \prod_{j=1}^{\ell} m_j(y_j) \prod_{j=1}^{d-\ell} m_j(x_j) - \prod_{j=1}^{d} m_j(y_j) \right) \, dx \right|
\]
\[
\leq \sum_{\ell=0}^{d-1} \left| \frac{1}{\pi^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \frac{\sin(R(y_j - x_j))}{(y_j - x_j)} \left( \prod_{j=1}^{\ell} m_j(y_j) \prod_{j=1}^{d-\ell} m_j(x_j) - \prod_{j=1}^{d} m_j(y_j) \right) \, dx \right|
\]
\[
\leq C \sum_{\ell=0}^{d-1} \left| \frac{1}{\pi} \int_{\mathbb{R}} \frac{\sin(R(y_{\ell+1} - x_{\ell+1}))}{(y_{\ell+1} - x_{\ell+1})} (m_{\ell+1}(x_{\ell+1}) - m_{\ell+1}(y_{\ell+1})) \, dx_{\ell+1} \right|,
\]
where $C$ is some universal constant. Using the above bound and the result in one dimension of theorem 2.2 for $m_1, m_2, \ldots, m_d$, we obtain the conclusion of the theorem when $m$ is the product of these functions.
To appreciate the proof, we demonstrate the key idea in the one-dimensional case. Here,

\[ m_R(y) - m(y) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin(R(y - x))}{y - x} (m(x) - m(y)) \, dx, \]

which we write as \( m_R(y) - m(y) = 1/\pi \int_{-\infty}^{+\infty} \sin(R(x - y)) \, g(x) \, dx \), where \( g(x) = (m(x) - m(y))/(x - y) \). We set \( y = 0 \) with no loss of generality, so \( m_R(y) - m(y) = 1/\pi \int_{-\infty}^{+\infty} \sin(z) \, e^g(z \epsilon) \, dz \), where \( \epsilon = 1/R \). Now due to the cyclic behaviour of the \( \sin \) function we can write this as \( m_R(y) - m(y) = 1/\pi \int_{0}^{2\pi} \sin(t) \sum_{k=-\infty}^{+\infty} e^g(t + 2\pi k) \, dt \). The term \( \sum_{k=-\infty}^{+\infty} e^g(t + 2\pi k) \) is a Riemann sum approximation to an integral which converges to a constant, for all \( t \), as \( \epsilon \to 0 \). The overall convergence to 0 is then a consequence of \( \int_{0}^{2\pi} \sin t \, dt = 0 \). Hence, it is how the Riemann sum converges to a constant which determines the speed at which \( m_R(y) - m(y) \to 0 \).

(b) General integral theorem

It is interesting to note that the \( \sin \) function in the Fourier integral theorem could be replaced by any cyclic function which integrates to 0 over the cyclic interval. In particular, we consider the following general form of integral theorem:

\[ m_{R,\phi}(y) = \int_{\mathbb{R}^d} \prod_{j=1}^{d} \phi(R(y_j - x_j)) \, m(x) \, dx, \tag{2.7} \]

where the univariate function \( \phi \) is a cyclic function on \((0, 2\pi)\). Using the proof technique of theorem 2.2 and the assumptions with function \( m \) in that theorem, we also obtain the following approximation error of the general integral theorem:

Corollary 2.3. Assume that the kernel \( \phi \) satisfies the constraints (1.3) and the function \( m \) satisfies the assumptions in theorem 2.2. Then, there exist universal constants \( C \) and \( \bar{C} \) depending on \( d \) and the function \( \phi \) such that when \( R \geq C \) we have \(|m_{R,\phi}(y) - m(y)| \leq C/R^k \), where \( K \) is defined as in theorem 2.2.

Therefore, we have a general class of integral theorems that possesses similar approximation errors as that of the Fourier integral theorem. Furthermore, the general integral theorems are also able to automatically maintain the dependence structures in function \( m \).

We now discuss some examples of function \( \phi \) that have connection to Haar wavelet and splines.

Example 2.4 (Haar wavelet integral theorem). We consider the piece-wise linear function

\[ \phi_{\text{Haar}}(x) = \begin{cases} \frac{(x - 2k\pi)^2}{C_1 \pi}, & \text{if } x \in \left(\left(2k - \frac{1}{2}\right)\pi, \left(2k + \frac{1}{2}\right)\pi\right) \\ \frac{(2k + 1)\pi - x)^2}{C_1 \pi}, & \text{if } x \in \left(\left(2k + \frac{1}{2}\right)\pi, \left(2k + \frac{3}{2}\right)\pi\right) \end{cases}, \]

where

\[ C_1 = \sum_{k=-\infty}^{\infty} \left(2(2k + 1)\log \left(\frac{4k + 3}{4k + 1}\right) - 4k \log \left(\frac{4k + 1}{4k - 1}\right)\right), \]

and

\[ \phi'_{\text{Haar}}(x) = \begin{cases} \frac{2}{C_1 \pi} & x \in \left(\left(2k - \frac{1}{2}\right)\pi, \left(2k + \frac{1}{2}\right)\pi\right) \\ -\frac{2}{C_1 \pi} & x \in \left(\left(2k + \frac{1}{2}\right)\pi, \left(2k + \frac{3}{2}\right)\pi\right). \end{cases} \]

It demonstrates that the derivative of the function \( \phi \) is the Haar wavelet function, which can be named the ‘Haar wavelet integral theorem’ for this choice of \( \phi_{\text{Haar}} \).


**Example 2.5 (Spline integral theorem).** Here, we take into account the piece-wise quadratic function

\[
\phi_{spline}(x) = \begin{cases} 
\frac{4(x - 2k\pi)(x - (2k - 1)\pi)}{C_2\pi^2}, & \text{if } x \in ((2k - 1)\pi, 2k\pi) \\
\frac{4(x - 2k\pi)((2k + 1)\pi - x)}{C_2\pi^2}, & \text{if } x \in (2k\pi, (2k + 1)\pi)
\end{cases},
\]

where

\[
C_2 = \sum_{k=-\infty}^{\infty} \left( 8k(2k - 1) \log \left( \frac{2k}{2k - 1} \right) - 8k(2k + 1) \log \left( \frac{2k + 1}{2k} \right) \right).
\]

A direct calculation shows that

\[
\phi'_{spline}(x) = \begin{cases} 
\frac{8x - 4(4k - 1)\pi}{C_2\pi^2}, & x \in ((2k - 1)\pi, 2k\pi)
\\
\frac{-8x + 4(4k + 1)\pi}{C_2\pi^2}, & x \in (2k\pi, (2k + 1)\pi)
\end{cases},
\]

Therefore, the first derivative of \(\phi_{spline}\) is a piece-wise linear function. The particular form of \(\phi'_{spline}\) justifies the spline integral theorem for this choice of kernel function \(\phi_{spline}\).

Finally, we would like to highlight that the Haar wavelet and spline integral theorems are just two instances of the integral theorems. In general, the class of cyclic function \(\phi\) satisfying an integral theorem is vast.

### 3. Optimal functions

In this section, we discuss optimal functions \(\phi\) from the integral theorem with respect to the kernel density estimation problem. To ease the findings, we specifically consider the univariate settings, namely, \(d = 1\). Subject to the constraints in equation (1.3), as mentioned in the Introduction, we consider minimizing either the problem (1.7) or the problem (1.8). We show that the sin function minimizes \(\int_{\mathbb{R}} (\phi(x)/x)^2 \, dx\) subject to constraints in equation (1.3), whereas this is not the case when we introduce a density function, i.e. the aim now being to minimize \(\int_{\mathbb{R}} (\phi(x)/x)^2 \, m(x) \, dx\) for some density function \(m\), which yields more precise asymptotic behaviours of the variance of density estimator \(\hat{m}_R\) in equation (1.6).

Before proceeding, we briefly explain why the bias is not relevant, as it only depends on \(m\). To see this, we can write the bias as

\[
R^{-1} \int \phi(s) \left\{ \frac{m(y + s/R) - m(y)}{s/R} \right\} \, ds.
\]

How this goes to zero as \(R \to \infty\) depends solely on the \((m(y + s/R) - m(y))/s\) term and how it converges to \(m'(y)\); recall the \(\phi\) is cyclic and integrates to zero over every interval of the type \((2\pi k, 2\pi (k + 1))\).

#### (a) The sin function

As we have mentioned, a direct application of the general integral theorem is for a Monte Carlo estimator of density functions: \([4]\). The bias of the estimator \(\hat{m}_R\), namely, \(E[\hat{m}_R(y)] - m_R(y)\), has been established in corollary 2.3. A natural question to ask is the form of optimal kernel \(\phi\) that leads to a good quality of density estimator \(\hat{m}_R\). To answer that question, we use asymptotic mean integrated square error, see \([5,22]\), which is equivalent to finding the optimal kernel \(\phi\) which leads to a small variance for the estimator \(\hat{m}_R\). A simple calculation shows that

\[
\text{Var}(\hat{m}_R(y)) = \frac{1}{n} \text{Var} \left( \frac{\phi(R(y - X))}{(y - X)} \right),
\]

for any \(y \in \mathbb{R}\) where the outer variance is taken with respect to random variable \(X\) following density function \(m\). With an assumption that \(||m||_\infty < \infty\), we can upper bound the variance of
\( \hat{m}_R(y) \) as follows:

\[
\text{Var}(\hat{m}_R(y)) \leq \frac{R||m||_{\infty}}{n} \int_{-\infty}^{\infty} \frac{\phi^2(x)}{x^2} \, dx. \tag{3.2}
\]

The integral in the upper bound in equation (3.2) is convenient as it involves \( m \). It indicates that the optimal kernel \( \phi \) minimizing that integral will be independent of \( m \), which also yields good insight into the behaviour of the variance of the density estimator \( \hat{m}_R \) for all \( y \) and \( m \). Therefore, we consider minimizing the upper bound (3.2) with respect to the constraints that \( \phi \) is almost surely differentiable cyclic function in \((0, 2\pi)\) and satisfies the constraints (1.3). It is equivalent to solving the following objective function \( \min_{\phi} \int_{-\infty}^{\infty} \phi^2(x)/x^2 \, dx \), such that \( \phi \) satisfies (1.3). This is the objective function (1.7) mentioned in the Introduction.

To study the optimal function \( \phi \) that satisfy these constraints, we define the following functions: 
\[
\alpha(x) := \sum_{k=-\infty}^{\infty} (x + 2k\pi)^{-1} \quad \text{and} \quad \beta(x) = \sum_{k=-\infty}^{\infty} (x + 2k\pi)^2 \text{ for any } x \in (0, 2\pi).
\]
These will be useful in the proof to the forthcoming theorem 3.1. Now, we would like to prove that

\[
\alpha(x) = \frac{1}{2 \tan(x/2)} \quad \text{and} \quad \beta(x) = -\frac{1}{2 \sin^2(x/2)}. \tag{3.3}
\]

In fact, from the infinite product representation of the sin function, we have

\[
\frac{\sin(\pi t)}{\pi t} = \prod_{k=1}^{\infty} \left(1 - \frac{t^2}{k^2}\right), \tag{3.4}
\]

for any \( t \in (0, 1) \). By taking the logarithm of both sides of the equation and taking the derivative with respect to \( x \), we obtain that

\[
\frac{\pi \cos(\pi t)}{\sin(\pi t)} - \frac{1}{t} = -\sum_{k=1}^{\infty} \frac{2t}{k^2 - t^2}.
\]

By using the change of variable changing \( t = x/(2\pi) \), we obtain the conclusion that \( 2\alpha(x) = \cot(x/2) \). The form of \( \beta(x) \) can be obtained directly by taking the derivative of \( \alpha(x) \). Therefore, we obtain the conclusion of claim (3.3).

Now, we state our main result for the optimal kernel \( \phi \) solving the objective function (1.7).

**Theorem 3.1.** The optimal cyclic and almost everywhere differentiable function \( \phi \) that solves the objective function (1.7) subject to constraints (1.3) is \( \phi(x) = \sin(x)/\pi \) for all \( x \in (0, 2\pi) \).

Interestingly, if we consider a truncation of the sin function, which corresponds to the optimal kernel \( \phi \) for solving objective function (1.7), at \( k = 1 \) in equation (3.4), we obtain the Epanechnikov kernel, \([14, 23]\), which is given by \( k_{\text{Epa}}(x) = 3(1 - x^2)/4 \), for \( x \in (-1, 1) \) and \( 0 \) otherwise. This kernel had been shown to have optimal efficiency among non-negative kernels that are differentiable up to the second order: \([24]\). Direct calculation shows that \( \int_{\mathbb{R}} k_{\text{Epa}}^2(x) \, dx = 3/5 \), which is greater than \( \int_{\mathbb{R}} \sin^2(x)/(\pi^2 x^2) \, dx = 1/\pi \). Therefore, if we use the term (1.7) as an indication for the quality of our variance, the Epanechnikov kernel is not better than the sin kernel from the Fourier integral theorem. It also aligns with an observation from \([24]\) that we can construct better kernels, which can take negative values, than the Epanechnikov kernel without restricting to only the non-negative kernels.

**Proof.** Since the function \( \phi \) is cyclic in \((0, 2\pi)\), we have

\[
\int_{-\infty}^{\infty} \frac{\phi(x)}{x} \, dx = \sum_{k=-\infty}^{\infty} \int_{2k\pi}^{(2k+1)\pi} \frac{\phi(x)}{x} \, dx = \sum_{k=-\infty}^{\infty} \int_{0}^{2\pi} \frac{\phi(x + 2k\pi)}{x + 2k\pi} \, dx
\]

\[
= \int_{0}^{2\pi} \phi(x) \alpha(x) \, dx.
\]
Similarly, we also obtain that \( \int_{-\infty}^{\infty} (\phi^2(x)/x^2) \, dx = \int_{0}^{2\pi} \phi^2(x)\beta(x) \, dx \). Given these equations, the original problem can be rewritten as follows:

\[
\min_{\phi} \int_{0}^{2\pi} \phi^2(x)\beta(x) \, dx, \tag{3.5}
\]

such that \( \int_{0}^{2\pi} \phi(x) \, dx = 0 \) and \( \int_{0}^{2\pi} \phi(x)\alpha(x) \, dx = 1 \). The Lagrangian function corresponding to the objective function (3.5) takes the form

\[
\mathcal{L}(\phi) = \int_{0}^{2\pi} \phi^2(x)\beta(x) \, dx - \lambda_1 \int_{0}^{2\pi} \phi(x) \, dx - \lambda_2 \left( \int_{0}^{2\pi} \phi(x)\alpha(x) \, dx - 1 \right).
\]

Since \( \phi \) is almost surely differentiable, the function \( \mathcal{L} \) can be rewritten as follows:

\[
\mathcal{L}(\phi) = \phi^2(2\pi)B(2\pi) - \phi^2(0)B(0) - \int_{0}^{2\pi} 2\phi(x)\phi'(x)B(x) \, dx - \lambda_1 \int_{0}^{2\pi} \phi(x) \, dx \\
- \lambda_2 \left( \int_{0}^{2\pi} \phi(x)\alpha(x) \, dx - 1 \right) \\
= -\int_{0}^{2\pi} \phi(x) \left( 2\phi'(x)B(x) + \lambda_1 + \lambda_2\alpha(x) \right) \, dx - C = \int_{0}^{2\pi} F(x) \, dx - C,
\]

where \( B \) is the function such that \( B'(x) = \beta(x) \) and \( C = \lambda_2 + \phi^2(2\pi)B(2\pi) - \phi^2(0)B(0) \). To find \( \phi \) that minimizes the function \( \mathcal{L} \), we use the Euler–Lagrange equation, see for example [25], which entails that \( \partial F/\partial \phi - \partial/\partial x(\partial F/\partial \phi') = 0 \). This equation leads to \( \phi(x) = -(\lambda_1 + \lambda_2\alpha(x))/(2\beta(x)) \), where \( \lambda_1 \) and \( \lambda_2 \) can be determined by solving the conditions (1.3). Given the form of optimal \( \phi \) and the forms of \( \alpha(x) \) and \( \beta(x) \) in equation (2.2), the first condition in equation (1.3) leads to \( \int_{0}^{2\pi} (2\lambda_1 \sin^2(x/2) + \lambda_2 \sin(x/2) \cos(x/2)) \, dx = 0 \). This yields \( \lambda_1 = 0 \). The second condition in equation (1.3) indicates that \( \int_{0}^{2\pi} \lambda_2 \cos^2(x/2) \, dx = 4 \), which leads to \( \lambda_2 = \pi/4 \). Therefore, we have the optimal kernel \( \phi(x) = \sin(x)/\pi \) for all \( x \in (0, 2\pi) \). As a consequence, we obtain the conclusion of the theorem.

\[\blacksquare\]

(b) Optimal function with Cauchy density

In the previous section, we saw that the sin function is optimal for minimizing \( \int_{R} (\phi(x)/x)^2 \, dx \) subject to the constraints (1.3), namely, \( \int_{R} (\phi(x)/x) \, dx = 1 \) and \( \int_{0}^{2\pi} \phi(x) \, dx = 0 \) and \( \phi \) being a cyclic function on \( (0, 2\pi) \). That objective function stems from the upper bound (3.2) on the variance of density estimator \( \hat{m}_R \).

In this section, we demonstrate the reason in finding other integral theorems to Fourier’s by finding optimal function \( \phi \), also satisfying the constraints (1.3), which minimizes the leading term from the variance of \( \hat{m}_R \), which is given by: \( \int_{R} (\phi(x)/x)^2 \, m(x) \, dx \) for some density function \( m \) on \( R \). Since the above integral captures the leading term of the variance of \( \hat{m}_R \), it gives more precise asymptotic behaviour of the variance than that of the term (1.7). Furthermore, as that integral involves the density function \( m \), it indicates that the optimal kernel \( \phi \) also depends on \( m \). To illustrate our findings, we specifically consider the setting when \( m \) is the Cauchy density; i.e. \( m(x) = \pi^{-1}/(1 + x^2) \) for all \( x \), which would be useful when modelling heavy tailed distributions and moreover we are able to solve the relevant equations.

The proof idea for obtaining the optimal kernel \( \phi \) is the same as in that of theorem 3.1. The only thing that is different is that we need to find a new function \( \beta \), which we refer to as \( \beta_{\text{Cauchy}} \); given now by

\[
\beta_{\text{Cauchy}}(x) = \sum_{m=-\infty}^{\infty} \frac{1}{(x + 2\pi m)^2} \frac{1}{(1 + (x + 2\pi m)^2)}.
\]
To find the closed-form expression of $\beta_{\text{Cauchy}}(x)$, we use contour integration from complex analysis (e.g. [26]); so consider

$$I_R = \int_{\gamma_R} \frac{\cot(z/2) \, dz}{(x + z)^2(1 + (x + z)^2)},$$

where $\gamma_R$ is a circle in the complex plane of radius $R$ around the origin. The simple poles occur at $z = 2\pi m$ for all integers $m$, giving a total residue of $2\beta(x)$, since the relevant coefficient in the Laurent expansion of $\cot(z/2)$ is 2; also at $z = x \pm i$ for which the residues are $-\cot(\frac{1}{2}(x \pm i))$. There is a double pole at $z = -x$ for which the residue is the first derivative of $f(z) = \cot(z/2)(1 + (x + z)^2)^{-1}$ evaluated at $z = -x$. From direct calculation, this term is $\frac{1}{2} \sin^2(x/2)$.

Now using the Cauchy residue theorem and noting that $I_R \to 0$ as $R \to \infty$, and expanding $\cot((x \pm i)/2)$, we obtain

$$\beta_{\text{Cauchy}}(x) = \frac{1}{4} \sin^2(x/2) - \frac{1}{2} \coth(1/2) \frac{1 + \cot^2(x/2)}{\coth^2(1/2) + \cot^2(x/2)}.$$

As shown in the proof of theorem 3.1, the optimal function $\phi$ is of the form

$$\phi_{\text{Cauchy}}(x) = \frac{\lambda_1 + \lambda_2 \alpha(x)}{\beta_{\text{Cauchy}}(x)},$$

(3.6)

where recall that $\alpha(x) = \cot(x/2)$, with the $\lambda_1, \lambda_2$ values being now able to capture the coefficient of $\frac{1}{4}$. Different from theorem 3.1, we do not have closed-form expressions for $\lambda_1$ and $\lambda_2$. However, numerical integration can be used to determine the $\lambda_1, \lambda_2$ values to meet the constraints (1.3), and this yields $\lambda_1 = 0$ and $\lambda_2 = 0.118$. A picture of the optimal function $\phi_{\text{Cauchy}}$ is given in figure 4.

We note that such a result is not restricted to only the setting when the samples are generated from the Cauchy distribution; actually, when we have samples from a Gaussian distribution, using the kernel $\phi_{\text{Cauchy}}$ also yields slightly better variance than the sin kernel from the Fourier integral theorem. We leave a detailed investigation of the benefit of $\phi_{\text{Cauchy}}$ over the sin kernel for general settings of density estimation problem in future work.

For finding the optimal cyclic function for the general density function $m$, we would need to be able to find the function $\sum_{k=-\infty}^{k=\infty} m(x + 2\pi k)(x + 2\pi k)^{1/2}$. This can always be tackled using contour integration though is not straightforward in many cases. Numerical solutions are obviously easy to get. For example, for the Gaussian density, the $\phi$ function (up to proportionality) appears in figure 5, but as yet we have been unable to find an explicit solution.
4. Discussion

In this paper, we have introduced a general class of integral theorems. These provide natural Monte Carlo density estimators, which can automatically preserve the dependence structure of a dataset. In the univariate density estimation setting, we demonstrate that the Fourier integral theorem is optimal when we minimize a square integral in equation (1.7); a term that indicates a good variance of the density estimator. To show the benefit of a general class of integral theorems, we also consider optimal kernels that minimize the term (1.8), which provides a more precise nature of the variance of the kernel density estimator. Our study shows that the optimal kernel for alternative objective function are generally not the sin kernel.

Here, we discuss a few future directions of research. First, we have only obtained the optimal kernels in our general class of integral theorems for density estimation. It is also important to study optimal kernels for other statistical estimation tasks, such as non-parametric (modal) regression [18, 24, 27] and mode clustering [28–30]. Second, the work of [31] proposes using double Fourier transforms to approximate a non-parametric function that can capture both the correlation of words in each sequence and the correlation of sequences in natural language processing tasks. Given our study with the general integral theorems, it is of interest to investigate whether we can develop the general notion of double Fourier transforms in the similar way as we do for the integral theorems and whether the choice of double Fourier transforms is optimal for estimating the non-parametric function arising in natural language processing tasks.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors’ contributions. N.H.: investigation, methodology, writing—review and editing; S.W.: formal analysis, investigation, methodology, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. The authors are grateful for helpful comments and suggestions from two referees.

References

1. Wiener N. 1933 The fourier integral and certain of its applications. Cambridge, UK: Cambridge University Press.
2. Bochner S. 1959 Lectures on fourier integrals. Princeton, NJ: Princeton University Press.
3. Fowler RH. 1921 A simple extension of Fourier’s integral theorem and some physical applications, in particular to the theory of quanta. *Proc. R. Soc. Lond. A* **99**, 462–471. (doi:10.1098/rspa.1921.0061)

4. Ho N, Walker S. 2021 Multivariate smoothing via the Fourier integral theorem and Fourier kernel. *J. Mach. Learn. Res.* Submitted.

5. Wand M. 1992 Error analysis for general multivariate kernel estimators. *J. Nonparametr. Stat.* **2**, 1–15. (doi:10.1080/10485259208832538)

6. Staniswalis J, Messer K, Finston D. 1993 Kernel estimators for multivariate regression. *J. Nonparametr. Stat.* **3**, 103–121. (doi:10.1080/10485259308832575)

7. Chacon J, Duong T. 2018 *Multivariate kernel smoothing and its applications*. New York, NY: CRC Press.

8. Parzen E. 1962 On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065–1076. (doi:10.1214/aoms/1177728190)

9. Davis K. 1975 Mean square error properties of density estimates. *Ann. Stat.* **3**, 1025–1030. (doi:10.1214/aos/1176343207)

10. Dempster AP, Laird NM, Rubin DB. 1977 Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**, 1–38. (doi:10.1111/j.2517-6161.1977.tb01600.x)

11. Rotiroti F, Walker S. 2022 Computing marginal likelihoods via the Fourier integral theorem and pointwise estimation of posterior densities. *Stat. Comput.* **32**, 67. (doi:10.1007/s11222-022-10131-0)

12. Wand M, Jones M. 1993 Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Am. Stat. Assoc.* **88**, 520–528. (doi:10.1080/01621459.1993.10476303)

13. Rosenblatt M. 1956 Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27**, 832–837. (doi:10.1214/aoms/1177728190)

14. Epanechnikov VA. 1969 Non-parametric estimation of a multivariate probability density. *Theory Probab. Appl.* **14**, 153–158. (doi:11.1371/1114019)

15. Yakowitz SJ. 1985 Nonparametric density estimation, prediction, and regression for Markov sequences. *J. Am. Stat. Assoc.* **80**, 215–221. (doi:10.1080/01621459.1985.10477164)

16. Müller HG. 1984 Smooth optimum kernel estimators of densities, regression curve and modes. *Ann. Stat.* **12**, 766–774. (doi:10.1214/aos/1176348768)

17. Wasserman L. 2006 *All of nonparametric statistics*. New York, NY: Springer.

18. Botev ZI, Grotowski JF, Kroese DP. 2010 Kernel density estimation via diffusion. *Ann. Stat.* **38**, 2916–2957. (doi:10.1214/10-AOS799)

19. Gine E, Nickl R. 2010 Confidence bands in density estimation. *Ann. Stat.* **38**, 1122–1170. (doi:10.1214/09-AOS738)

20. Jiang H. 2017 Uniform convergence rates for kernel density estimation. In *Proceedings of the 34th ICML, Sydney*. Proceedings of Machine Learning Research 70.

21. Wand M. 1994 Fast computation of multivariate kernel estimators. *J. Comput. Graph. Stat.* **3**, 433–445. (doi:10.1080/10618600.1994.10474656)

22. Mueller HG. 1984 Smooth optimum kernel estimators of densities, regression curve and modes. *Ann. Stat.* **12**, 766–774. (doi:10.1214/10-AOS799)

23. Terrell GR, Scott DW. 1992 Variable kernel density estimation. *Ann. Stat.* **20**, 1236–1265. (doi:10.1214/105040100000000300)

24. Azzalini A, Torelli N. 2007 Clustering via nonparametric density estimation. *Stat. Comput.* **17**, 71–80. (doi:10.1007/s11222-006-9010-y)

25. Chang J, Duong T. 2013 Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electron. J. Stat.* **7**, 499–532. (doi:10.1214/13-EJS781)

26. Chacon J, Duong T. 2015 A population background for nonparametric density-based clustering. *Stat. Sci.* **30**, 518–532. (doi:10.1214/15-STSC526)

27. Lee-Thorp J, Ainslie J, Eckstein I, Ontañón S. 2021 FNet: mixing tokens with Fourier transforms. (https://arxiv.org/abs/2105.03824)