



# Fast adaptive Fourier integration for spectral densities of Gaussian processes

Paul G. Beckman<sup>1</sup> · Christopher J. Geoga<sup>2</sup>

Received: 15 April 2024 / Accepted: 4 October 2024 / Published online: 11 November 2024  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

The specification of a covariance function is of paramount importance when employing Gaussian process models, but the requirement of positive definiteness severely limits those used in practice. Designing flexible stationary covariance functions is, however, straightforward in the spectral domain, where one needs only to supply a positive and symmetric spectral density. In this work, we introduce an adaptive integration framework for efficiently and accurately evaluating covariance functions and their derivatives at irregular locations directly from *any* continuous, integrable spectral density. In order to make this approach computationally tractable, we employ high-order panel quadrature, the nonuniform fast Fourier transform, and a Nyquist-informed panel selection heuristic, and derive novel algebraic truncation error bounds which are used to monitor convergence. As a result, we demonstrate several orders of magnitude speedup compared to naive uniform quadrature approaches, allowing us to evaluate covariance functions from slowly decaying, singular spectral densities at millions of locations to a user-specified tolerance in seconds on a laptop. We then apply our methodology to perform gradient-based maximum likelihood estimation using a previously numerically infeasible long-memory spectral model for wind velocities below the atmospheric boundary layer.

**Keywords** Gaussian process · Covariance function · Quadrature · Nonuniform fast Fourier transform · Long-memory process

## 1 Introduction

Gaussian process (GP) models are ubiquitous in many statistical settings. They provide a flexible method of interpolating data which incorporates dependence structure, and yield model-implied second-order information that can be used for uncertainty quantification. Let  $Z(x)$  be a GP with mean function  $\mathbb{E}Z(x) \equiv 0$  and positive definite covariance function

$$\text{Cov}(Z(x), Z(x')) = K_{\theta}(x, x'), \quad (1.1)$$

PGB is supported in part by the Office of Naval Research under award #N00014-21-1-2383 and by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0022158.

✉ Paul G. Beckman  
paul.beckman@cims.nyu.edu

<sup>1</sup> Courant Institute, New York University, New York, NY 10012, USA

<sup>2</sup> Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA

where  $K_{\theta}$  is some parametric family of positive definite covariance functions indexed by parameters  $\theta$ . Consider observing  $Z$  at locations  $x_1, \dots, x_n \in \mathbb{R}$  corresponding to measurements  $z_i := Z(x_i)$ . Then the vector of observations  $\mathbf{z} := [z_1, \dots, z_n] \in \mathbb{R}^n$  is distributed as

$$\mathbf{z} \sim N(\mathbf{0}, \Sigma_{\theta}) \quad (1.2)$$

where the covariance matrix is given entry-wise by  $(\Sigma_{\theta})_{ij} := K_{\theta}(x_i, x_j)$ . For the duration of this work, we will make the additional simplifying assumption that the process  $Z$  is *stationary*, so that  $K_{\theta}(x, x') = K_{\theta}(x - x')$ , which implies that the process  $Z(x)$  is translation-invariant. This assumption limits the types of dependence structure one can express in exchange for significant theoretical and computational benefits. We refer readers to Stein (1999) for a more detailed introduction and discussion of GPs, covariance functions, and stationarity.

Once a parametric family of covariance models has been specified for some dataset  $\mathbf{z}$ , practitioners often need to fit the model parameters  $\theta$ , after which they can perform interpolation and other downstream tasks using the law specified

by the fitted parameters. In order to estimate the parameters  $\theta$ , one can compute the maximum likelihood estimator (MLE), denoted  $\hat{\theta}$ , which minimizes the Gaussian negative log-likelihood

$$-2\ell(\theta) := \log |\Sigma_{\theta}| + z^{\top} \Sigma_{\theta}^{-1} z + n \log(2\pi). \quad (1.3)$$

Naturally, evaluating the log-likelihood in this form requires evaluation of  $K_{\theta}$  in order to construct  $\Sigma_{\theta}$ . If  $K_{\theta}$  is available in closed form, this is of course no issue. Common examples of covariance functions in this category are the Matérn model with  $K_{\theta}(r) \propto r^{\nu} \mathcal{K}_{\nu}(r)$  and  $\mathcal{K}_{\nu}$  the modified second-kind Bessel function (Olver 2010), and its special cases  $K_{\theta}(r) \propto \exp(-|r|)$  for  $\nu = 1/2$  and  $K_{\theta}(r) \propto \exp(-|r|^2)$  for  $\nu \rightarrow \infty$ .

Unfortunately, as has been noted many times in the literature, positive definite covariance functions that can be expressed in closed form are fairly difficult to construct mathematically, and thus few are known and used in practice. This severely limits the expressiveness of Gaussian process models available to practitioners. For stationary processes, however, one can alternatively construct valid covariance models by specifying their Fourier transform. Bochner's theorem states that

$$K_{\theta}(r) := \mathcal{F}\{S_{\theta}\}(r) := \int_{-\infty}^{\infty} S_{\theta}(\omega) \exp(2\pi i \omega r) d\omega \quad (1.4)$$

is a positive definite function for any integrable positive function  $S_{\theta}$ , which we refer to as a *spectral density*. Closed form positive spectral densities are much more easily constructed and parameterized than positive definite covariance functions. Thus from a modeling flexibility perspective it can be desirable to build and fit models via their spectral density.

Many methods for working in the spectral domain in a variety of special cases already exist in the literature. For gridded data in one or two dimensions, Whittle-type approximations are popular and fast to compute, but may introduce severe bias without special care (Whittle 1963; Sykulski et al. 2019; Grainger et al. 2021). For spatio-temporal processes that are regular in at least one dimension, "half-spectral" models have been used to generate flexible kernels (Cressie and Huang 1999; Stein 2005; Horrell and Stein 2017; Geoga et al. 2021). For gridded data in two dimensions with missing values, Guinness (2019) provides an imputation method for estimating spectral densities via the EM algorithm. More generally applicable is the "Random Fourier Features" paradigm (Rahimi and Recht 2007), which can be viewed as a Monte Carlo approximation to the Fourier integral (1.4). A number of methods have also been proposed which are equivalent to truncating the Fourier integral and applying the trapezoidal rule on a finite interval, including the "Regular Fourier Features" (Hensman et al. 2018), "Equispaced Fourier Gaussian Process" frameworks

(Greengard et al. 2022). A related method uses generalized Gaussian quadrature to approximate (1.4) for families of kernels (Greengard 2021), but has not been extended to singular or slowly-decaying spectral densities, which are the focus of this work. Another close analog to our approach is the method of Im et al. (2007), which uses spline bases to flexibly capture spectral densities at low frequencies and algebraic tails at higher frequencies.

Working in the spectral domain is often motivated as much by efficient computation as it is by flexible modeling, and thus many of the above methods do not directly provide high accuracy kernel evaluations, focusing instead on the problem of efficiently inverting  $\Sigma_{\theta}$ . The method we introduce here fills this gap, providing an approximation-agnostic tool for performing maximum likelihood estimation from a general spectral density  $S_{\theta}$  with fully irregularly sampled data. In this work, we demonstrate that panel Gaussian integration of the Fourier integral allows the accurate and efficient computation of covariances  $K(r)$  and their derivatives  $\frac{\partial}{\partial \theta_j} K_{\theta}(r)$  from any continuous, integrable spectral density  $S_{\theta}$ . This in turn facilitates gradient-based maximum likelihood estimation directly on any parameterization of the spectral density. By taking advantage of modern nonuniform fast Fourier transform (NUFFT) (Dutt and Rokhlin 1993; Barnett et al. 2019) and automatic differentiation (AD) (Griewank and Walther 2008) methods, along with careful analysis of the tail behavior of the spectral density, we can compute  $K_{\theta}(r)$  and its derivatives with respect to parameters to  $\epsilon = 10^{-12}$  accuracy even for slowly decaying spectral densities at millions of inter-observation distances  $r$  in seconds on a standard laptop. Our free and open source Julia code is available at <https://github.com/pbeckman/SpectralKernels.jl>.

While the possibilities for functional forms of spectral densities are endless and the machinery described here is generally applicable, in this work we study in detail the simple extension given by incorporating an integrable singularity into a bounded spectral density  $S_{\theta}(\omega)$ , resulting in the new model  $|\omega|^{-\alpha} S_{\theta}(\omega)$ . Such extensions correspond to "long memory" processes, for which we derive some theoretical properties, overcome numerical challenges associated with evaluating their covariance functions, and demonstrate their practical value by fitting a singular model to Doppler LiDAR wind velocity profiles. For the remainder of this paper we will often suppress the dependence of  $\Sigma$ ,  $K$ , and  $S$  on  $\theta$  for notational clarity.

## 2 Methodology

We are concerned here only with real-valued covariance functions  $K$ , and therefore assume  $S$  is an even function.

This results in the simplification

$$K(r) = 2 \int_0^\infty S(\omega) \cos(2\pi\omega r) d\omega. \tag{2.1}$$

In order to evaluate  $K(r)$  by directly computing (2.1), one must choose a quadrature rule. As the spectral density  $S$  is assumed to be integrable, it must decay sufficiently fast for large  $\omega$ , so that one can truncate the infinite interval  $[0, \infty)$  at some point  $b$  and integrate only on the finite interval  $[0, b]$ . This gives

$$\begin{aligned} K(r) &\approx 2 \int_0^b S(\omega) \cos(2\pi\omega r) d\omega \\ &\approx 2 \sum_{j=1}^m \gamma_j S(\omega_j) \cos(2\pi\omega_j r), \end{aligned} \tag{2.2}$$

where  $\omega_j$  and  $\gamma_j$  are the nodes and weights of the chosen quadrature rule. The simplest choice of quadrature is the trapezoidal rule, which uses equispaced points  $\omega_j = (j-1)h$  with weights  $\gamma_1 = \gamma_m = \frac{h}{2}$  and  $\gamma_j = h$  for  $j = 2, \dots, m-1$ , where the grid spacing is  $h := \frac{b}{m-1}$ . The aliasing and truncation errors when using the trapezoidal rule are treated in detail for the squared exponential and Matérn models in Barnett et al. (2023).

For known spectral densities  $S$  with sufficiently fast decay, this can be a highly accurate quadrature. However, it has two limitations. First, for more flexible and complex spectral densities  $S$ , the analysis used in Barnett et al. (2023) to choose the grid spacing  $h$  necessary to resolve  $S$  becomes difficult, and must be done for each new parametric family  $S$ . Second, for small  $r$  we may need to take both a small  $h$  to control the quadrature error, as well as a large  $b$  to control the truncation error when integrating slowly decaying spectral densities  $S$  such as the commonly used Matérn model Stein (1999)

$$S_\theta(\omega) = \phi^2(\rho^2 + \omega^2)^{-\nu-\frac{1}{2}}, \quad \theta := \{\phi, \rho, \nu\} \tag{2.3}$$

with small values of the smoothness parameter  $\nu$ , e.g.  $\nu = 1/2$ . This can result in a number of quadrature nodes  $m$  which is prohibitively large from a computational standpoint. In the remainder of this section, we demonstrate that panel Gaussian integration of the Fourier integral allows us to overcome both of these limitations.

### 2.1 Panel integration of the Fourier integral

If one wishes to integrate (2.1) for a large range of  $r$  with uniform accuracy, using a dense quadrature rule for  $\omega \in [0, b]$  for some large  $b$  might appear to be an unavoidable cost. However, as we demonstrate numerically in this section and

theoretically in Sect. 2.3.2, truncation error in  $K(r)$  is concentrated near the origin in  $r$ -space. Intuitively, this is because the high frequency information contained in the tails of the spectral density  $S$  has a much greater impact on covariances between observations which are close together. This intuition will be made precise shortly by Theorem 1.

The localization of error near  $r = 0$  suggests that we can compute the Fourier integral on the interval  $\omega \in [0, b]$ , then iteratively add the contribution of the Fourier integral on successive intervals in  $\omega$ -space for *only* those  $r$ 's nearest the origin in  $r$ -space. We will refer to each interval  $[a, b]$  in  $\omega$ -space as a *panel*. Define the exact and discretized panel integrals

$$I_{[a,b]}(r) := \int_a^b S(\omega) \cos(2\pi\omega r) d\omega \tag{2.4}$$

$$\tilde{I}_{[a,b]}^{(m)}(r) := \sum_{j=1}^m \gamma_j S(\omega_j) \cos(2\pi\omega_j r) \tag{2.5}$$

where  $\omega_j$  and  $\gamma_j$  are nodes and weights of an  $m$ -point quadrature rule on  $[a, b]$ . The error

$$E_{[a,b]}^{(m)}(r) := \left| I_{[a,b]}(r) - \tilde{I}_{[a,b]}^{(m)}(r) \right| \tag{2.6}$$

in the  $m$ -point trapezoidal rule is only order  $\mathcal{O}(m^{-2})$  when the integrand is non-periodic, as is the case in this panel integral setting. However, the  $m$ -point Gauss–Legendre rule, which integrates polynomials of degree  $2m - 1$  exactly, is generally much more accurate for smooth, non-periodic functions. In particular, for  $f \in \mathcal{C}^s([a, b])$  with mild additional smoothness assumptions, an  $m$ -order Gauss–Legendre quadrature rule will have error of order  $\mathcal{O}(m^{-2s-1})$  (Trefethen 2019, Theorem 19.4). Gauss–Legendre quadrature is thus amenable to panel integration and adaptivity, which allows us to accurately discretize Fourier integrals for any continuous, integrable spectral density. Adaptive quadrature will be treated in greater detail in Sect. 2.3.1.

By the Nyquist-Shannon sampling theorem, one requires at least  $m = 2r \cdot (b - a)$  values to completely determine a function  $f(\omega)$  on the interval  $\omega \in [a, b]$  if it has bandlimit  $r$ , that is, if  $\mathcal{F}\{f\}(t) \equiv 0$  for all  $|t| > r$ . Therefore, for a desired accuracy  $\delta$ , a given  $m$ , and sufficiently smooth  $S$ , we expect an  $\mathcal{O}(m)$ -node Gauss–Legendre rule to compute the Fourier integral with integrand  $f(\omega) = S(\omega) \cos(2\pi\omega r)$  to within accuracy  $\delta$  on any interval of length  $b - a = \frac{m}{2r}$ . We use this heuristic to choose the next panel  $[a, b]$  in  $\omega$ -space, where  $r$  is taken to be the largest distance for which  $K(r)$  has not yet converged. Then all  $r' < r$  result in less oscillatory integrands, which are therefore also accurately resolved by the  $m$ -point rule.

It is worth noting that this Nyquist-informed heuristic for panel selection is not tight in either direction. If  $S$  contains

sharp features, then the integrand  $f(\omega) = S(\omega) \cos(2\pi\omega r)$  will not have bandlimit  $r$ , and thus more quadrature nodes may be required to achieve the desired accuracy  $\delta$ . Conversely, if  $S$  is very smooth, then the effect of aliasing may be below the desired accuracy  $\delta$ , and thus fewer quadrature nodes may still yield a adequately accurate result despite not fully resolving the integrand. However, for sufficiently large  $m$  we find that this Nyquist-based choice typically computes the Fourier integral to double precision with neither significant redundant oversampling nor the need for further refinement where  $S$  is smooth.

The novelty and efficiency of our approach stems from the two related mechanisms discussed above. First, we can reduce the number of  $r$  values to be computed after adding each panel Fourier integral, as the largest  $r$ 's have converged. This reduces the number of points  $n$  at which we must evaluate the sum (2.5). Second, because  $\cos(2\pi\omega r)$  is less oscillatory in  $\omega$  for smaller  $r$ , as the largest  $r$ 's converge we can take increasingly large panels while keeping the number of oscillations per panels constant. Therefore we use the same  $m$ -point quadrature rule to accurately resolve the integrand. This can significantly reduce the total number of quadrature nodes used to evaluate the Fourier integral when compared to a more uniform quadrature scheme. Figure 1 provides a visual example of this panel growth as  $K(r)$  is resolved for the largest  $r$ 's which reduces the highest remaining Nyquist frequency. In particular, note that the spacing between quadrature nodes increases by almost two orders of magnitude between the first and fourth panels, corresponding to a proportional reduction in computational effort relative to using, for example, the trapezoidal rule to integrate the same interval in  $\omega$ -space. This trend only continues as we progress in  $\omega$ -space, adaptively generating a highly non-uniform quadrature rule with increasingly sparse nodes as  $\omega$  increases. The resulting quadrature provides orders-of-magnitude speedups over alternatives, as we demonstrate in Sect. 3.3.

Computing the sum (2.5) directly for  $n$  distances with an  $m$ -point quadrature rule has  $\mathcal{O}(nm)$  complexity for each panel. This ostensibly introduces a major tension in the computation of these panel integrals: higher-order quadrature rules enable larger panels and faster convergence, but particularly for large data sizes they lead to enormously burdensome computations if done directly. In this next section, we discuss how the NUFFT can be used to relieve this tension by reducing the cost of each panel integral to  $\mathcal{O}(m + n \log n)$  — a speedup which is imperative to making Fourier integration computationally tractable.

**Remark 1** The idea of using the same number of quadrature nodes  $m$  to discretize Fourier integrals for which the space-frequency product  $r \cdot (b - a)$  is constant is reminiscent of the complementary low-rank property used in butterfly algo-

rithms O'Neil et al. (2010); Li and Yang (2017). One could view the present work as a sparsified butterfly-like algorithm for which the error control is designed specifically for the problem of adaptively computing continuous Fourier transforms of slowly decaying functions.

## 2.2 Acceleration with the nonuniform fast Fourier transform

As each panel Fourier integral (2.5) is a sum of cosines with nonuniform frequencies  $\omega_1, \dots, \omega_m$  evaluated at nonuniform distances  $r_1, \dots, r_n$ , it can be computed as the real part of the exponential sum

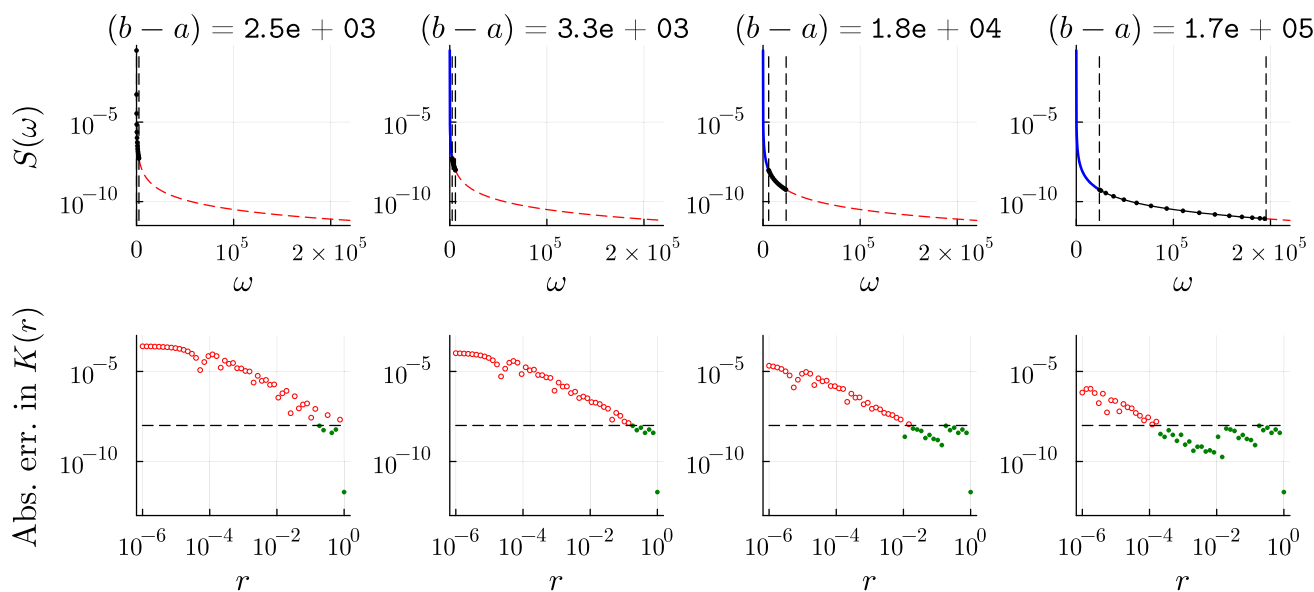
$$f_k = \sum_{j=1}^m \gamma_j S(\omega_j) \exp(2\pi i \omega_j r_k) \quad k = 1, \dots, n \quad (2.7)$$

so that  $\tilde{I}_{[a,b]}^{(m)}(r_k) = \Re(f_k)$ . Equivalently, the computation of the Fourier integral for each panel  $[a, b]$  in  $\omega$ -space at all unconverged distances  $r$  can be viewed as a matrix–vector product of a dense  $n \times m$  nonuniform discrete Fourier matrix with a vector of  $m$  evaluations of the spectral density along with appropriate quadrature weights given by

$$\begin{bmatrix} \exp(2\pi i \omega_1 r_1) & \cdots & \exp(2\pi i \omega_m r_1) \\ \vdots & \ddots & \vdots \\ \exp(2\pi i \omega_1 r_n) & \cdots & \exp(2\pi i \omega_m r_n) \end{bmatrix} \begin{bmatrix} \gamma_1 S(\omega_1) \\ \vdots \\ \gamma_m S(\omega_m) \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}. \quad (2.8)$$

Computing this dense matrix–vector product directly has a cost of  $\mathcal{O}(nm)$ . However, the sum (2.7) and the matrix–vector product (2.8) are equivalent views of exactly a “type 3” nonuniform to nonuniform discrete Fourier transform, which can be evaluated to accuracy  $\delta$  in  $\mathcal{O}(m + n \log n)$  complexity using the NUFFT. Most existing NUFFT algorithms work by convolving the input data with some “spreading” function, performing an equispaced FFT on a fine grid, and deconvolving to obtain values at the desired output locations. The development of efficient NUFFT libraries with tunable accuracy guarantees has been the subject of significant research in the past few decades. See for example Barnett et al. (2019); Greengard et al. (2007); Dutt and Rokhlin (1993).

In the present context of computing pointwise covariances from the spectral density, the improved scaling of the NUFFT when compared to direct summation is essential to facilitating the use of larger, more accurate quadrature rules to compute the values of each panel integral at a greater number of distances. The dramatic computational impact of the NUFFT will be studied in greater detail in Sect. 3.3, where we observe orders of magnitude speedup for practical data sizes.



**Fig. 1** Panel integration of a Matérn spectral density with  $\nu = \frac{1}{2}$ ,  $\rho = 1$ , and  $\phi$  chosen so that  $K(0) = 1$ . The top row shows various panels of the spectral density being integrated, and each corresponding plot in the bottom row shows the absolute truncation error in the kernel after the

panel integral above has been added. A small subset of the quadrature nodes are shown in each panel. A tolerance of  $\epsilon = 10^{-8}$  and  $m = 5,000$  nodes per panel are used

**Remark 2** The number of points  $m$  in each NUFFT can be tuned to improve performance. Generally, using large NUFFTs helps to amortize setup costs and take advantage of multi-threading. However, ill-conditioning and round-off errors limit the accuracy of the NUFFT to about  $10^{-9}$  for large inputs (Barnett et al. 2019, Remark 9). So if higher accuracy is needed, smaller NUFFTs must be used at the expense of a larger pre-factor.

### 2.3 Error estimation

As we saw in Sect. 2.1, the key idea behind fast panel integration of Fourier integrals is to use panels of increasing size in  $\omega$ -space. This is made possible by choosing a working tolerance  $\delta$  and sequentially integrating panels in  $\omega$ -space for *only* those distances  $r$  at which  $K(r)$  has not yet been resolved to within  $\delta$  accuracy. Determining when  $K(r)$  is adequately resolved for each  $r$  requires a careful analysis of multiple sources of error, and naive stopping criteria often result in the loss of several digits of accuracy in the computed integral.

Recall that for any valid covariance function  $K$ , one has that  $K(0) \geq |K(r)|$  for all  $r > 0$ . Then for any set of locations  $x_1, \dots, x_n$ , computing each entry in an approximate covariance matrix  $\tilde{\Sigma}_{ij} := \tilde{K}(x_i - x_j)$  such that the pointwise error relative to  $K(0)$  is controlled  $|K(r) - \tilde{K}(r)| / K(0) < \epsilon$  for all  $r$ , consequently bounds the relative max-norm error in  $\tilde{\Sigma}$

$$\frac{\|\Sigma - \tilde{\Sigma}\|_{\max}}{\|\Sigma\|_{\max}} = \frac{\max_{1 \leq i, j \leq n} |K(x_i - x_j) - \tilde{K}(x_i - x_j)|}{K(0)} < \epsilon. \tag{2.9}$$

By the equivalence of norms, controlling the relative max-norm error also controls the relative Frobenius and spectral norm errors, up to constants which may depend on  $n$ .

We emphasize that the pointwise error relative to  $K(0)$  is equivalent to neither the relative *nor* the absolute pointwise error in each  $K(r)$ . If  $K(r) = 1e-12$ , then *relative* error  $|K(r) - \tilde{K}(r)| / |K(r)| < 1e-8$  would require cancellation in the sum of panel integrals to 20 digits, which is impossible in general in double precision. Conversely, if  $K(r) = 1e12$ , then *absolute* error  $|K(r) - \tilde{K}(r)| < 1e-8$  would require 20 correct digits, which is again impossible in double precision. These are standard and well-known limitations of adaptive integration in finite precision arithmetic, and neither bounding the relative nor absolute pointwise error by  $\epsilon$  are necessary conditions for bounding the resulting error in log-likelihood evaluation by  $\epsilon$ .

As our final goal is to evaluate the log-likelihood, we argue that controlling relative norm errors in  $\tilde{\Sigma}$  is a relevant metric, as small pointwise errors in  $\Sigma$  will result — at least asymptotically — in small log-likelihood errors because the log-likelihood is a smooth function of the entries of the covariance matrix  $\Sigma$  for fixed data  $\mathbf{y}$ . However, other choices of error metric are of course possible, and we emphasize that the connection between pointwise kernel accuracy and the

accuracy of the resulting log-likelihood is not direct. If the true covariance matrix is poorly conditioned, then the log-likelihood may be computed to significantly fewer than 16 digits of accuracy purely due to fundamental limitations of double-precision arithmetic, *even if every kernel evaluation is accurate to machine precision*. See Section S.1 in the supplementary material for a brief numerical study of the impact of pointwise errors and ill-conditioning on log-likelihood accuracy. Since this work is focused on kernel evaluation, we leave a more thorough error analysis of the Gaussian log-likelihood to future work. Returning to the problem of error estimation within our method, we note that there are two sources of error at play which we seek to bound above by  $\delta$  — quadrature error and truncation error — which we now treat individually.

### 2.3.1 Quadrature error

To control the quadrature error, we use a straightforward adaptive approach. As is standard practice in adaptive integration, we estimate the error in the discretized integral by comparing the result to that obtained by a higher order quadrature rule. For this purpose we use a  $2m$ -point Gauss–Legendre rule. The error in the  $m$ -point rule on  $[a, b]$  is then approximated as

$$\begin{aligned} \tilde{E}_{[a,b]}^{(m)}(r) &:= \left| \tilde{I}_{[a,b]}^{(2m)}(r) - \tilde{I}_{[a,b]}^{(m)}(r) \right| \\ &\approx \left| I_{[a,b]}(r) - \tilde{I}_{[a,b]}^{(m)}(r) \right|. \end{aligned} \tag{2.10}$$

For a given panel, if  $\tilde{E}_{[a,b]}^{(m)}(r) > \delta$  for any unconverged distance  $r$ , we divide  $[a, b]$  in two and repeat this procedure separately on  $[a, (a + b)/2]$  and  $[(a + b)/2, b]$ . We deem these subpanels converged when both  $\tilde{E}_{[a,(a+b)/2]}^{(m)}(r) < \delta/2$  and  $\tilde{E}_{[(a+b)/2,b]}^{(m)}(r) < \delta/2$ , so that the total error on  $[a, b]$  remains bounded by  $\delta$  for all unconverged  $r$ 's. If the error remains large for any  $r$  in one or both of these subpanels, we continue this dyadic refinement and proportional tolerance reduction until we obtain an approximation  $\tilde{I}_{[a,b]}(r)$  as a sum of subpanel integrals such that the total quadrature error is bounded by  $\delta$ . This is again standard practice for adaptive integration, and can be performed to high accuracy. See for example Gonnet (2012).

### 2.3.2 Truncation error

Controlling the truncation error, given by

$$E_{\text{trunc}}(b, r) := \int_b^\infty S(\omega) \cos(2\pi \omega r) d\omega, \tag{2.11}$$

is a more subtle issue. As discussed in Sect. 2.1, we iteratively integrate panels from zero to infinity in spectral space. The remaining question is how to determine for each  $r$  when we have integrated a sufficient interval  $[0, b]$  in spectral space so that  $|E_{\text{trunc}}(b, r)| < \delta$  and we can cease adding new panels.

One could check that the contribution of the current panel is less than the tolerance  $\delta$ , that is  $\tilde{I}_{[a,b]}^{(m)}(r) < \delta$ . This is a necessary but *not* a sufficient condition for  $E_{\text{trunc}}(b, r) < \delta$ . In practice, for spectral densities which decay exponentially this condition results in a negligible loss of accuracy. However, for spectral densities with slow algebraic decay, terminating integration when  $\tilde{I}_{[a,b]}^{(m)}(r) < \delta$  may result in errors significantly greater than  $\delta$ , as one would be truncating many panels whose contributions would each be  $\mathcal{O}(\delta)$ .

Therefore, for better error control, we consider the exponent  $\beta$  and constant  $c$  in the tails of the spectral density such that  $S(\omega) \sim c\omega^{-\beta}$  as  $r \rightarrow \infty$ . For many spectral densities  $c$  and  $\beta$  can be derived analytically, often by straightforward means. Otherwise, these coefficients can be estimated using linear least squares in log space. With these values, we can compute the truncation error analytically for the resulting power law tail, which gives

$$|E_{\text{trunc}}(b, r)| \approx \left| \int_b^\infty c\omega^{-\beta} \cos(2\pi \omega r) d\omega \right| \tag{2.12}$$

$$= -c(2\pi r)^{\beta-1} \Re\left( (-i)^{\beta-1} \Gamma(-\beta + 1, -2\pi i br) \right), \tag{2.13}$$

where  $\Gamma(s, z) := \int_z^\infty t^{s-1} \exp(-t) dt$ , is the upper incomplete Gamma function (Olver 2010). There exist a number of libraries to numerically evaluate this special function. However, it is helpful to have a simple, tight, and easily invertible algebraic upper bound on this truncation error. We now derive such a bound as a consequence of the following lemma, whose proof is given in the appendix.

**Lemma 1** For any  $s, y > 0$ ,

$$|\Gamma(-s, iy)| \leq \min\left( y^{-s-1}, \frac{y^{-s}}{s} \right). \tag{2.14}$$

**Theorem 1** For any  $\beta > 1$  and any  $b, r > 0$ ,

$$\begin{aligned} &\left| \int_b^\infty \omega^{-\beta} \exp(-2\pi i \omega r) d\omega \right| \\ &\leq \min\left( \frac{1}{\beta - 1} b^{-\beta+1}, \frac{1}{2\pi r} b^{-\beta} \right), \end{aligned} \tag{2.15}$$

where the first term gives a tighter bound when  $br \leq \frac{\beta-1}{2\pi}$ , and the second term otherwise.

**Proof** Taking the change of variables  $t = 2\pi i\omega r$ , we obtain

$$\left| \int_b^\infty \omega^{-\beta} \exp(-2\pi i\omega r) d\omega \right| = \left| (2\pi i r)^{\beta-1} \int_{2\pi i b r}^\infty t^{\beta-1} \exp(-t) dt \right| = (2\pi r)^{\beta-1} |\Gamma(-\beta + 1, 2\pi i b r)|.$$

Applying Lemma 1 gives the desired result. □

This result indicates that when the product  $br$  is small (for close together observations or limited integration domains in spectral space), the truncation error in the Fourier integral behaves like the truncation error in the non-oscillatory integral. But when  $br$  is large (for far apart observations or large integration domains in spectral space), the truncation error decays faster by an additional power of  $b$ , with a constant that decreases for larger  $r$ . See Fig. 2. This makes precise the intuition of Fig. 1 that for a fixed integration domain, the truncation error for larger distances  $r$  is lower.

One can therefore use the analytic formula (2.12) or the simpler algebraic upper bound (2.15) to control the truncation error. In conjunction with the adaptive integration and quadrature error estimation described in Sect. 2.3.1, we obtain accurate estimates of the total error in the computed Fourier integral.

### 2.4 Power law singularities at the origin

Long-memory Gaussian processes can be characterized by having an integrable singularity in their spectral density  $S$  at the origin. Such models typically lack closed form expressions for the resulting kernel  $K$ , and even when such expressions are available, they are often difficult to compute numerically. In this work, we will focus on a modification of the Matérn family that we will call a “singular” Matérn model, which was first proposed in Porcu and Stein (2012) and corresponds to a spectral density given by

$$S_\theta(\omega) = \phi^2 |\omega|^{-\alpha} (\rho^2 + \omega^2)^{-\nu - \frac{1}{2}}, \quad \theta := \{\phi, \alpha, \rho, \nu\} \tag{2.16}$$

with  $0 \leq \alpha < 1$ . As will be discussed in detail in the next section, the corresponding covariance function does technically have a closed form representation, but it is exceptionally challenging to evaluate numerically.

In contrast, such spectral densities with power law singularities are treated easily by the panel Gaussian quadrature framework presented here. While directly applying a Gauss–Legendre quadrature to the integral

$$K(r) = 2 \int_0^b \omega^{-\alpha} S(\omega) \cos(2\pi \omega r) d\omega \tag{2.17}$$

results in low accuracy, Gauss–Jacobi quadratures are designed specifically to treat the singularity  $\omega^{-\alpha}$  accurately, and can be computed quickly and accurately (Glaser et al. 2007; Hale and Townsend 2013). Simply using a Gauss–Jacobi rule on the first panel followed by Gauss–Legendre rules on all remaining panels allows accurate computation of singular Fourier integrals of this form. For panels that do not contain the origin, the  $\omega^{-\alpha}$  term may safely be included in the Fourier integrand, and the standard Gauss–Legendre rule gives high accuracy.

### 2.5 Computing derivatives of the kernel

In order to take advantage of gradient-based optimizers for maximum likelihood estimation, one must compute derivatives  $\frac{\partial}{\partial \theta_j} K(r)$ . Assuming that the parametric family  $S_\theta$  is sufficiently well-behaved as to allow exchanging differentiation and integration, we have

$$\frac{\partial K_\theta(r)}{\partial \theta_j} = 2 \int_0^\infty \omega^{-\alpha} \frac{\partial S(\omega)}{\partial \theta_j} \cos(2\pi \omega r) d\omega, \tag{2.18}$$

which can be computed using the same framework that is used to evaluate  $K_\theta$  itself.

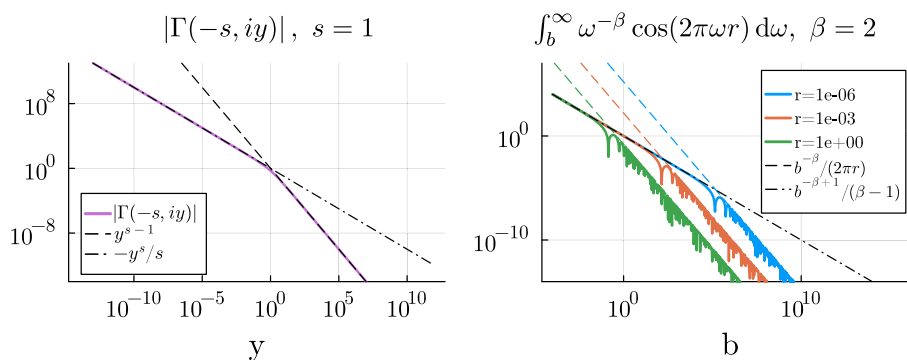
While differentiating even standard covariance functions such as the Matérn with respect to kernel parameters can be challenging (Geoga et al. 2023), most common spectral densities are very simple to differentiate. The computation of the partial derivatives  $\frac{\partial}{\partial \theta_j} S_\theta(\omega)$  can be done easily using automatic differentiation (AD) (Griewank and Walther 2008), which operates at the code level to programmatically generate derivatives of a given function. As a result, end users can simply write any parametric spectral density  $S_\theta$  that they would like, and the derivatives of  $K_\theta$  will be obtained automatically using our software.

This ease of differentiation can be extended to singular spectral densities. To compute the derivative of the kernel  $K$  with respect to the singularity parameter  $\alpha$ , given by

$$\frac{\partial K(r)}{\partial \alpha} = -2 \int_0^\infty \omega^{-\alpha} \log(\omega) S(\omega) \cos(2\pi \omega r) d\omega, \tag{2.19}$$

we require a method for accurately integrating the  $\omega^{-\alpha} \log(\omega)$  singularity on the panel  $[0, b]$  containing the origin. Ignoring the  $\log(\omega)$  singularity and applying a Gauss–Jacobi rule results in low accuracy, especially as  $\alpha$  approaches 1 and the weights become relatively large and positive near zero.

**Fig. 2** The incomplete Gamma function bound from Lemma 1 (left) and corresponding bound on the power law truncation error given by Theorem 1 for various  $r$  (right)



However, if we apply integration by parts

$$\begin{aligned}
 & (1 - \alpha) \int_0^b \omega^{-\alpha} \log(\omega) S(\omega) \cos(2\pi\omega r) d\omega \\
 &= b^{1-\alpha} \log(b) S(b) \cos(2\pi br) \\
 &\quad - \int_0^b \omega^{-\alpha} (S(\omega) + \omega \log(\omega) S'(\omega)) \cos(2\pi\omega r) d\omega \\
 &\quad + 2\pi r \int_0^b \omega^{1-\alpha} \log(\omega) S(\omega) \sin(2\pi\omega r) d\omega,
 \end{aligned}
 \tag{2.20}$$

we see that the resulting integrals involve only the power law singularity  $\omega^{-\alpha}$ , and that the log singularity has been removed. As  $S$  is a closed form function provided by the user,  $S'(\omega)$  can be obtained analytically or using AD, and we can evaluate the above expression using the same Gauss-Jacobi rule that is employed to compute kernel values. This requires only two NUFFTs, and avoids any additional cost to compute specialized quadratures for the  $\omega^{-\alpha} \log(\omega)$  singularity. For panels which do not contain the origin, the  $\omega^{-\alpha} \log \omega$  term can be included in the integrand, and the Gauss-Legendre rule gives high accuracy as before. Using this strategy and incorporating it into a custom rule for the AD engine again means that this derivative can be obtained entirely programmatically without any end-user intervention.

### 3 Numerical results

Before employing the adaptive integration method described above to compute kernel values in a maximum likelihood estimation context with real data, we provide several demonstrations. First, we give examples of statistically interesting models which can be written easily in the spectral domain but for which no closed form expression exists for the corresponding covariance function. Next, we show that our method can accurately evaluate the singular Matérn covariance function even when existing alternative numerical methods fail. We close with a runtime comparison which illustrates that our adaptive NUFFT-accelerated Gaussian

quadrature scheme is necessary to efficiently obtain high accuracy kernel evaluations.

The online supplement to this work contains several additional numerical demonstrations and examples. In particular, it provides a demonstration of the sharpness of the error control described in Sect. 2.3 applied to the standard Matérn model, as well as further numerical experiments involving the singular Matérn model that validate the correctness of our method and compare it to alternatives.

#### 3.1 Designing new spectral densities

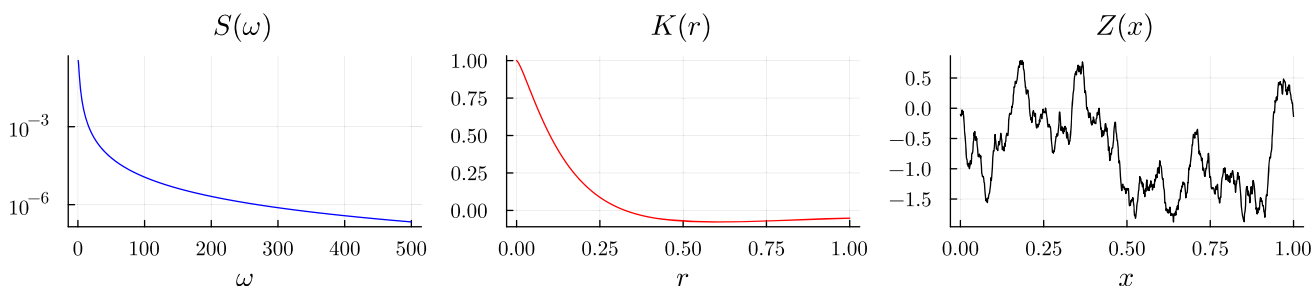
As a motivating example to illustrate how easily practitioners can write models in the spectral domain, we study a generalization of the standard Matérn model  $K(r) \propto r^\nu \mathcal{K}_\nu(r)$ . Two known limitations of the Matérn model are its exponential decay as  $r \rightarrow \infty$  for any finite  $\nu$ , and its inability to take on negative values. While loosening these restrictions in “kernel space” is challenging, it is trivial to write spectral densities whose corresponding kernels move beyond these limitations. Consider, for example, the model

$$\begin{aligned}
 S_\theta(\omega) &= \phi^2 (\lambda + (1 - \lambda) |\omega|^\gamma) (\rho^2 + |\omega|^\tau)^{-\nu - \frac{1}{2}}, \\
 \theta &:= \{\phi, \lambda, \gamma, \rho, \tau, \nu\},
 \end{aligned}
 \tag{3.1}$$

with  $\lambda \in [0, 1]$ ,  $\tau \in (0, 2]$ , and the constraint  $\tau(\nu + 1/2) - \gamma > 1$  for integrability. In the case  $\tau = 2$  this is precisely a standard Matérn model plus a fractional derivative of a standard Matérn model. For  $\tau < 2$  the function  $S_\theta$  will not be smooth at the origin, so the corresponding kernel will decay more slowly than the exponential rate of the standard Matérn. For small values of  $\lambda$  and large enough values of  $\gamma$ , the corresponding kernel can also take negative values. Figure 3 shows an example of this model.

To further highlight the broad class of novel models which can be specified through their spectral densities and fit using our method, we consider two more examples. The first is an “oscillatory” Matérn spectral density given by

$$S_\theta(\omega) = \phi^2 (\rho^2 + \omega^2)^{-\nu - \frac{1}{2}} (1 - \exp(-\lambda|\omega|) \sin(\gamma|\omega|)),$$



**Fig. 3** The generalized Matérn spectral density (3.1) (left), its corresponding covariance function (center), and a sample path from the process (right)

$$\theta := \{\phi, \rho, \nu, \lambda, \gamma\}. \tag{3.2}$$

This model promotes oscillatory behavior through strong negative then positive kernel values near the origin, while still maintaining full control over the mean-square differentiability of the process through the parameter  $\nu$ .

The final model we consider here is a semi-parametric long-memory model given by

$$S_{\theta}(\omega) = \phi^2 |\omega|^{-\alpha} \exp \left\{ -\lambda |\omega| + \sum_{k=0}^K c_k T_k \left( \frac{|\omega| - \rho}{|\omega| + \rho} \right) \right\},$$

$$\theta := \{\phi, \alpha, \lambda, \rho, c_0, \dots, c_K\}, \tag{3.3}$$

where  $T_k$  is the Chebyshev polynomial of order  $k$ . This model is in the spirit of semi-parametric ideas like the popular *spectral mixture kernel* (Wilson and Adams 2013) in that it avoids directly specifying the functional form of the spectral density. The singularity parameter here again gives the model the flexibility to capture slowly decaying tails in the covariance function, and the arbitrary number of orthogonal polynomials affecting frequencies near the origin can yield flexible kernels. Both of these models are shown in the spectral domain, covariance domain, and with an example sample path in Fig. 4.

### 3.2 Singular Matérn model

As previously mentioned, the singular Matérn model (2.16) is a useful tool for validating our approach and demonstrating its effectiveness for spectral densities with origin singularities. The Fourier transform of (2.16) is given by

$$K(r) = 2 \int_0^{\infty} |\omega|^{-\alpha} \phi (\rho^2 + \omega^2)^{-\nu-1/2} \cos(2\pi \omega r) d\omega$$

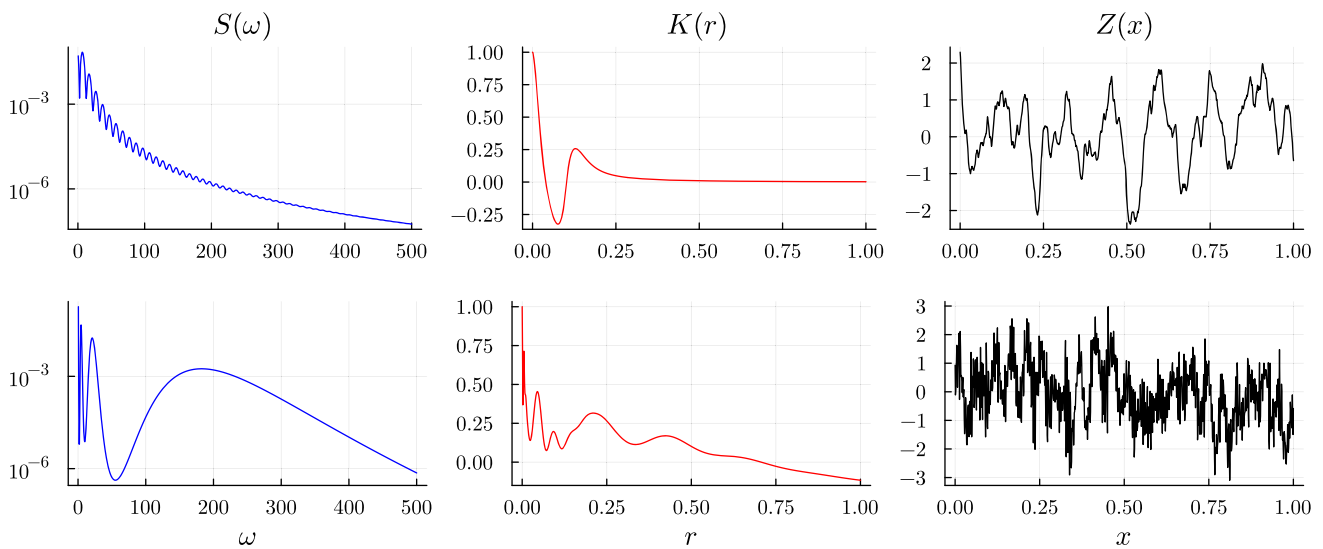
$$= \frac{\phi}{\rho^{2\nu} (2\pi r)^{-\alpha} \Gamma(\nu + \frac{1}{2})} \cdot$$

$$\left[ (2\pi r \rho)^{-\alpha} \Gamma(\frac{2\nu+\alpha}{2}) \Gamma(\frac{1-\alpha}{2}) \cdot$$

$${}_1F_2\left(\frac{1-\alpha}{2}, \left\{\frac{1}{2}, \frac{2-2\nu-\alpha}{2}\right\}, (\rho\pi r)^2\right) \right.$$

$$+ 2(2\pi r \rho)^{2\nu} \cos\left(\frac{(2\nu+\alpha)\pi}{2}\right) \Gamma(\frac{2\nu+1}{2}) \Gamma(-\alpha - 2\nu) \cdot$$

$$\left. {}_1F_2\left(\frac{2\nu+1}{2}, \left\{\frac{1+2\nu+\alpha}{2}, \frac{2+2\nu+\alpha}{2}\right\}, (\rho\pi r)^2\right) \right], \tag{3.4}$$



**Fig. 4** The top row shows the “oscillatory” Matérn spectral density (3.2) (left), its corresponding covariance function (center), and a sample path from the process (right). The bottom row displays analogous quantities for the semi-parametric long-memory model (3.3)

where  $\Gamma$  is the Gamma function and  ${}_1F_2$  is a generalized hypergeometric function (Olver 2010). Evaluating this  $K(r)$  accurately in double precision is extremely challenging, as the two terms inside the brackets are rapidly growing numbers of opposite signs which approach each other in absolute value and cancel as  $r$  grows. In exact arithmetic this is not a problem, but it poses a serious numerical issue in finite precision. Consider the case where both of the above terms have absolute value of order  $10^{20}$ , and a routine for evaluating any of the constituent special functions incurs a numerical relative error of even  $10^{-15}$ . Such small relative errors are inevitable, as double precision floating point numbers store only roughly 16 relative digits. Then an addition which should result in exact cancellation to zero could give a value of size  $10^{-15} \times 10^{20} \approx 10^5$  instead. Considering how quickly the inner terms grow, one reaches this regime for small  $r$  even with non-pathological choices of  $\phi$ ,  $\rho$ , and  $\nu$ .

To illustrate the challenge of evaluating the singular Matérn covariance function (3.4), we compare three methods for computing  $K(r)$ : one using a double-precision library for all special functions, one using the extended precision mathematical library Arb (Johansson 2017) for all special functions with 10,000 bits of precision, and our method.

Figure 5 shows the evaluated kernel on a regular grid of points in  $[0, 1]$  for a choice of  $\nu = 2.1$ ,  $\alpha = 0.3$ , several  $\rho$  values, and in all cases a  $\phi$  such that  $K(0) = 1$ , which is a “best case” numerical choice. For the case  $\rho = 2$ , Fig. 5 shows good visual agreement between the three methods. But as  $\rho$  is increased, we can see that both direct evaluation methods clearly fail to achieve even a single correct digit by  $r = 0.8$ . For  $\rho = 10$ , which is again still not a pathological parameter choice, we see that both direct evaluation methods are

so affected by roundoff errors ruining cancellation that they provide  $K(1)$  to be on the order of  $1 \times 10^{14}$ . As an additional indication of correctness even when direct methods cannot be used for validation, the online supplement provides summary tables that show that the kernel values obtained with our quadrature-based method correctly provide positive-definite matrices for all tested parameter values.

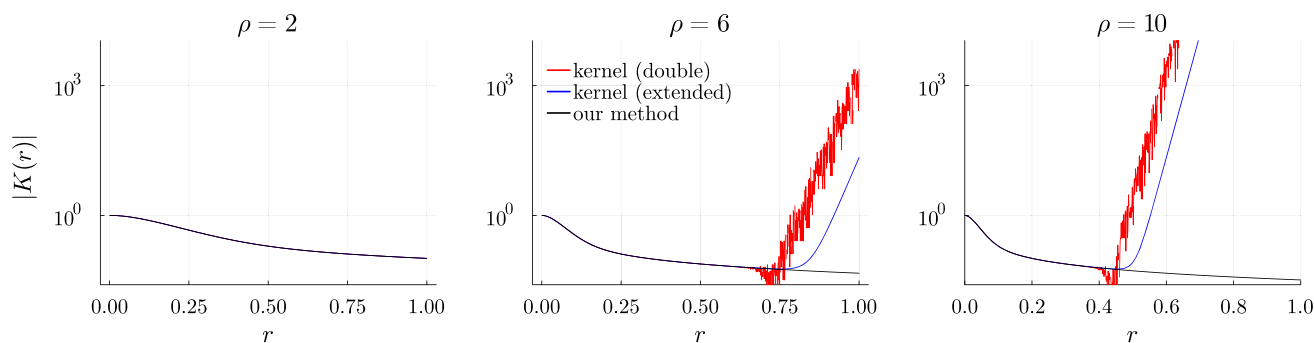
The difficulty of accurately evaluating  $K(r)$  for large  $r$  is especially problematic because long-memory models are most commonly applied to data which are strongly dependent on long time horizons — and it is precisely these cases in which evaluating the covariance function for large  $r$  is most relevant. This is because singular spectral densities can be used to build covariance functions that decay exceptionally slowly. As the following theorem demonstrates, singular spectral densities can provide covariance functions that are square-integrable but not integrable, or for  $\alpha > 1/2$  not even square-integrable.

**Theorem 2** *Let  $S \in C^2(\mathbb{R}) \cap L^1(\mathbb{R})$  be a bounded, symmetric, positive spectral density with  $S', S'' \in L^1(\mathbb{R})$ . Then for all  $\alpha \in [0, 1)$*

$$K_\alpha(r) := 2 \int_0^\infty |\omega|^{-\alpha} S(\omega) \cos(2\pi\omega r) d\omega \sim r^{-1+\alpha} \text{ as } r \rightarrow \infty.$$

We note that the requirement  $S \in C^2(\mathbb{R})$  is far from a necessary condition, and this asymptotic behavior is a very general phenomenon. For example  $S(\omega) = \exp(-|\omega|)$  yields the covariance function

$$K_\alpha(r) = \frac{2}{\Gamma(1-\alpha)} \left(1 + (2\pi r)^2\right)^{-\frac{1}{2}(1-\alpha)}$$



**Fig. 5** A comparison of the three methods for evaluating the covariance function given in Eq. (3.4): the double-precision direct kernel routine (red), the extended precision direct kernel routine (blue), and our Fourier quadrature routine (black)

$$\cos\left((1 - \alpha) \tan^{-1}(2\pi |r|)\right) \sim r^{-1+\alpha} \tag{3.5}$$

as  $r \rightarrow \infty$  despite the fact that  $S$  is not even  $C^1(\mathbb{R})$ . In such cases where  $S$  is monotonic on  $[0, \infty)$ , one can apply more general Tauberian results from the theory of slowly varying functions to obtain the same asymptotic decay (Bingham et al. 1989, Theorem 4.10.3).

**Remark 3** It is possible that one could use an expansion of the hypergeometric function in conjunction with relevant special function identities to derive an evaluation scheme for (3.4) which avoids the numerical instability of adding very large floats. However, doing so efficiently and accurately would represent a significant research endeavor, and our method can evaluate a broad class of such covariance functions without requiring further kernel-specific effort.

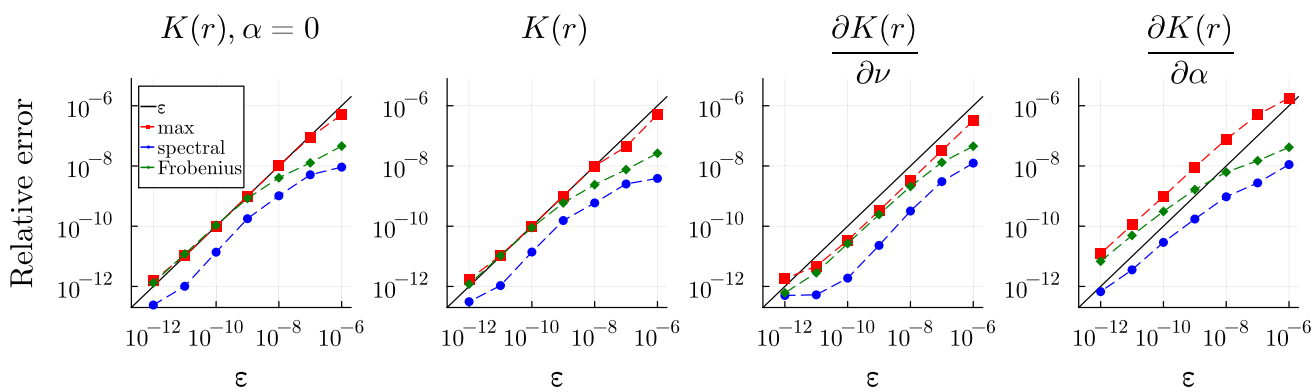
### 3.3 Performance for dense covariance matrix construction

As discussed in Sect. 2.3, by controlling the pointwise error relative to  $K(0)$  we have bounded the relative max-norm error in any covariance matrix with entries computed using our method. In Fig. 6, we demonstrate agreement between the user-specified tolerance  $\epsilon$  and the relative error in  $\Sigma$  in various matrix norms with  $N = 1000$  random observation locations  $x_1, \dots, x_N \stackrel{i.i.d.}{\sim} \text{Unif}([0, 1])$  for a slowly-decaying singular Matérn model with  $\nu = 0.51$ ,  $\rho = 0.5$ , and  $\alpha = 0.1$ , with  $\phi$  chosen so that  $K(0) = 1$ . We also show analogous errors in the derivative matrices  $\frac{\partial}{\partial \theta} \Sigma$  computed using AD for parameters  $\theta \in \{\nu, \alpha\}$ , for which the derivatives are most numerically challenging.

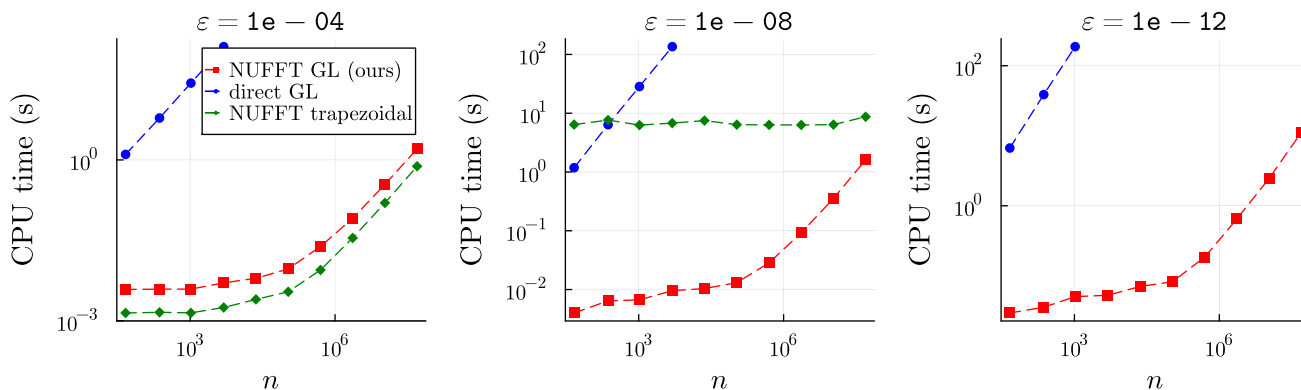
Having confirmed the accuracy of our scheme, we now study its computational cost, and demonstrate that both the NUFFT and our Nyquist-based heuristic for increasing panel length are necessary for computational efficiency. For various  $N$  ranging from 10 to 10,000, we form the dense covariance matrix for  $N$  uniform random observation locations in  $[0, 1]$  as above and time three approaches. First, we run our adap-

tive Gauss–Legendre method and time only the evaluation of the necessary NUFFTs, excluding the computation of error estimates. This provides a fair comparison with the trapezoidal rule, which is non-adaptive. Second, we repeat this with direct Fourier sums in place of the NUFFT. Finally, we use (Barnett et al. 2023, Corollary 6) to determine the grid spacing  $h$  and number of quadrature nodes  $m$  in a trapezoidal rule necessary to obtain each tolerance, and time the “type 2” NUFFT needed to compute kernel values from this quadrature rule. Figure 7 shows the resulting runtimes using an 8 core Apple M1 Pro CPU with 32GB of memory.

There are a two significant conclusions to be gathered from these results. The most obvious is the paramount importance of the NUFFT. The difference between  $\mathcal{O}(nm)$  and  $\mathcal{O}(m + n \log n)$  complexity for direct summation and the NUFFT respectively gives an orders of magnitude speedup, without which one can afford to compute only very few kernel values. The other conclusion is the necessity of our Nyquist-based heuristic for obtaining high accuracy kernel values. For tolerance  $\epsilon = 10^{-8}$ , the trapezoidal rule requires both small  $h$  in order to resolve the spectral density near the origin, as well as large  $m$  to control the truncation error. This results in a very large NUFFT of size  $m \approx 10^8$ . The  $\mathcal{O}(m)$  “spreading” step in the NUFFT thus becomes the dominant cost for all tested  $n$ , leading to a nearly constant cost which is orders of magnitude slower than our adaptive quadrature for moderate  $n$ . For  $\epsilon = 10^{-12}$ , the necessary number of equispaced trapezoidal nodes is  $m \approx 10^{12}$ , and results in memory issues. As our adaptive scheme uses panels with  $m = 2^{16}$  nodes, the  $\mathcal{O}(m)$  “spreading” step is only the bottleneck up to  $n = 10^5$  or so, after which the  $\mathcal{O}(n \log n)$  equispaced FFT which is evaluated within the NUFFT routine becomes the dominant cost, and we see quasilinear scaling with  $n$ . Therefore our method still runs in seconds, even when computing the kernel at approximately 50 million distances to 12 digit accuracy.



**Fig. 6** Comparison of the relative error  $\|A - \tilde{A}\| / \|A\|$  with the requested tolerance  $\epsilon$  for a variety of matrix norms, where  $A = \Sigma$  (first and second plots) or  $A = \frac{\partial}{\partial \theta} \Sigma$  (third and fourth plots). In the first plot,  $\alpha$  is set to zero to test the non-singular case, and all other parameters remain unchanged



**Fig. 7** A comparison of the runtime cost of computing kernel values from a Matérn model with  $\nu = 0.55$ ,  $\alpha = 0.5$ , and  $\phi$  chosen so that  $K(0) = 1$ . For three tolerances  $\epsilon$  we plot Gauss–Legendre quadrature with direct summation (blue), global trapezoidal quadrature with the NUFFT (green), and Gauss–Legendre quadrature with the NUFFT (red)

ture with direct summation (blue), global trapezoidal quadrature with the NUFFT (green), and Gauss–Legendre quadrature with the NUFFT (red)

### 4 Application to doppler LiDAR data

In this section we demonstrate the practical value of our framework with an application to high-frequency vertical wind profiles. The US Department of Energy’s Atmospheric Radiation Measurement (ARM) program offers a large collection of freely available measurements collected at field sites across the country, and in this work we look at the Doppler LiDAR-based vertical wind profiles made at the main field site in the Southern Great Plains (SGP) collection (Muradyan and Coulter 2020; Newsom 2012). These measurements of vertical profiles are made at a temporal resolution of approximately one second and a spatial resolution of 30m, thus providing particularly high resolution in time. However, aside from some special small segments across the several years’ worth of data, the measurements have frequent interruptions due to horizontal sweeps made by the sensor, occasional long pauses for various reasons, or other momentary instrument-based delays. For this reason, unless one focuses on very narrow time intervals of approximately 12–13 minutes or is uninterested in studying high-frequency

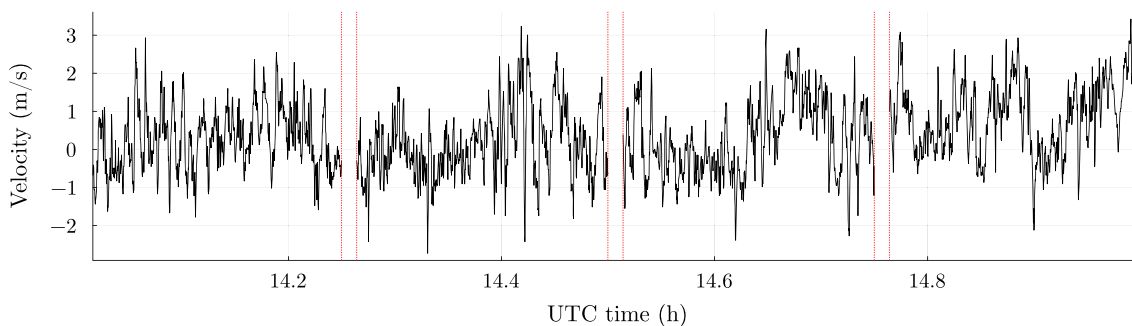
structure of the process, approaches like ours for continuous-time models which are applicable to irregularly sampled data become necessary.

#### 4.1 Preliminary analysis

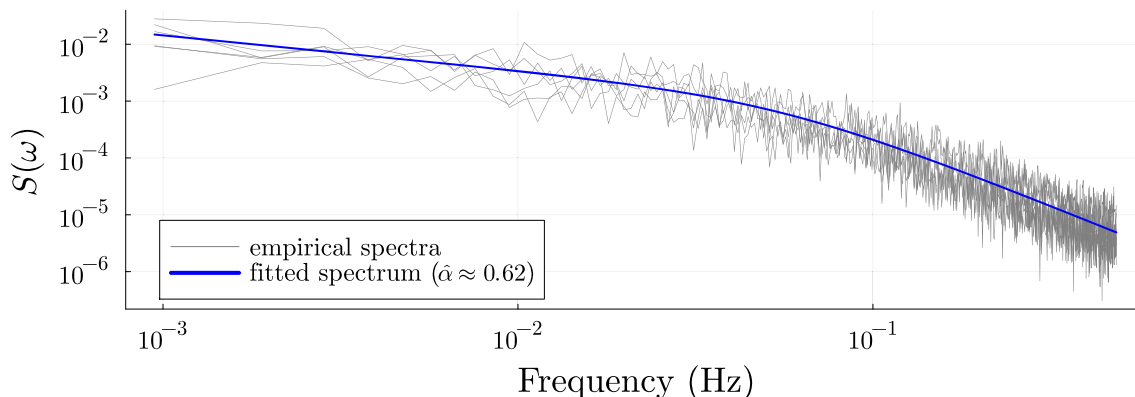
In this work, we investigate the same six days that were studied in Geoga et al. (2021), but by lifting the limitation of gridded data, we can examine a full hour of data on each of the days. Figure 8 shows an example of one hour of measurements made at the altitude of 240m.

A practitioner looking to model this data may first compute a Whittle-type estimator (Whittle 1951) as an exploratory tool. Figure 9 shows the result of Whittle estimation that is performed by treating each of the six one-hour segments of data as i.i.d. samples, breaking them into four segments based on the largest gaps and ignoring the smaller irregularities, and computing one averaged periodogram per day.

The Whittle MLE for this data implies a strong singularity of  $\hat{\alpha} \approx 0.62$ , corresponding to a process with significant long-range dependence even when segments are constrained



**Fig. 8** The Doppler LIDAR data from 1400 to 1500 UTC on June 03, 2015, at an altitude of 240m. Vertical lines are given to emphasize the largest three gaps in measurements



**Fig. 9** A Whittle MLE for the singular Matérn model computed treating the six days of LiDAR data as i.i.d. samples and doing rudimentary gap and sample irregularity handling

to 12 – 13 minutes, which limits the lowest observable frequencies.

### 4.2 Parameter estimation methodology

Motivated by this exploratory analysis, in this section we individually fit each of the six days of data as a continuous-time process using the singular Matérn model. In all cases, the MLE indicates a strong singularity comparable to the one obtained by the Whittle estimator. We additionally fit a standard Matérn model to compare log-likelihoods, and in all cases the log-likelihood of the data is materially improved by adding the singularity. In the body of this manuscript we focus on discussing one day of data, but the supplementary material provides full diagnostics for all six days that were fitted.

All estimates discussed below and in the supplement were obtained using the Artelys KNITRO optimizer (Byrd et al. 2006) with the sequential quadratic programming (SQP) algorithm. In this spirit of Fisher scoring, we use the *expected* Fisher information matrix, given by

$$\mathcal{I}(\theta)_{jk} = \frac{1}{2} \text{tr} \left( \Sigma_{\theta}^{-1} \left[ \frac{\partial}{\partial \theta_j} \Sigma_{\theta} \right] \Sigma_{\theta}^{-1} \left[ \frac{\partial}{\partial \theta_k} \Sigma_{\theta} \right] \right),$$

as a proxy for the Hessian of the log-likelihood. This matrix has the benefit of being computable with only first derivatives of the covariance function, and has meaningfully better performance than a general-purpose BFGS approximation (Geoga et al. 2020; Guinness 2021; Geoga et al. 2021; Beckman et al. 2023).

### 4.3 Results

We now discuss in detail the estimation results for the data on June 03, one of the six days that was studied. Analogous figures and tables for the other five days can be found in the supplemental material. The results from fitting singular and standard Matérn models with a nugget are summarized by point estimates and terminal likelihood values in Table 1. Along with point estimates, this table provides standard deviations implied by the expected Fisher information matrix. These implied uncertainties are often informative, but should be interpreted with care because the regularity conditions under which the expected Fisher matrix converges to the asymptotic precision of the MLE do not hold in many spatial modeling settings (Stein 1999).

The most obvious observation from Table 1 is that the singular Matérn model has a materially better log-likelihood

**Table 1** MLE estimates and terminal negative log-likelihoods for the standard Matérn (fixed  $\alpha = 0$ ) and the singular Matérn model for data on June 03.

	Fixed $\alpha = 0$	Fitted $\alpha$
$\ell(\hat{\theta})$	-1898.42	-1948.35
$\phi$	25.72 (4.182)	205.3 (95.429)
$\rho$	100.0 (7.854)	351.2 (45.705)
$\nu$	0.7947 (0.023)	0.6908 (0.058)
$\alpha$	-	0.775 (0.072)

When possible, expected Fisher matrix-implied standard deviations are provided in parentheses

than the standard Matérn model. A more scientifically interesting observation pertains to the implied smoothness of the process: noting that

$$\phi^2 |\omega|^{-\alpha} (\rho^2 + \omega^2)^{-\nu - \frac{1}{2}} \sim \omega^{-\alpha - 2\nu - 1}$$

at high frequencies, we see that the implied decay rate of the singular Matérn spectral density using the estimated  $(\hat{\nu}, \hat{\alpha}) \approx (0.69, 0.78)$  is  $\hat{\beta} = \hat{\alpha} + 2\hat{\nu} + 1 \approx 3.06$ . Recall that a process with spectral density  $S$  is mean-square differentiable if and only if  $\int \omega^2 S(\omega) < \infty$ , which in this notation is equivalent to  $\beta > 3$ . Therefore, under the singular model where  $\alpha$  is estimated, the process is mean-square differentiable, whereas in the standard Matérn model the estimate of  $\hat{\nu} = 0.79$  gives the decay  $\hat{\beta} = 2\hat{\nu} + 1 \approx 2.58$  which implies that it is not. The differentiability of these measurements below the atmospheric boundary layer (ABL) where convective forces are dominant and the process exhibits chaotic behavior has been a question of interest in several prior applications (Geoga et al. 2021; Geoga and Stein 2023), with parameter estimates often being very borderline. The differentiability under the singular Matérn model is an indication that the long-range dependence parameter  $\alpha$  may be valuable in disentangling low-frequency and high-frequency behavior in such processes, and agrees with the more recent work that uses continuous-time models in kernel-space (Geoga and Stein 2023).

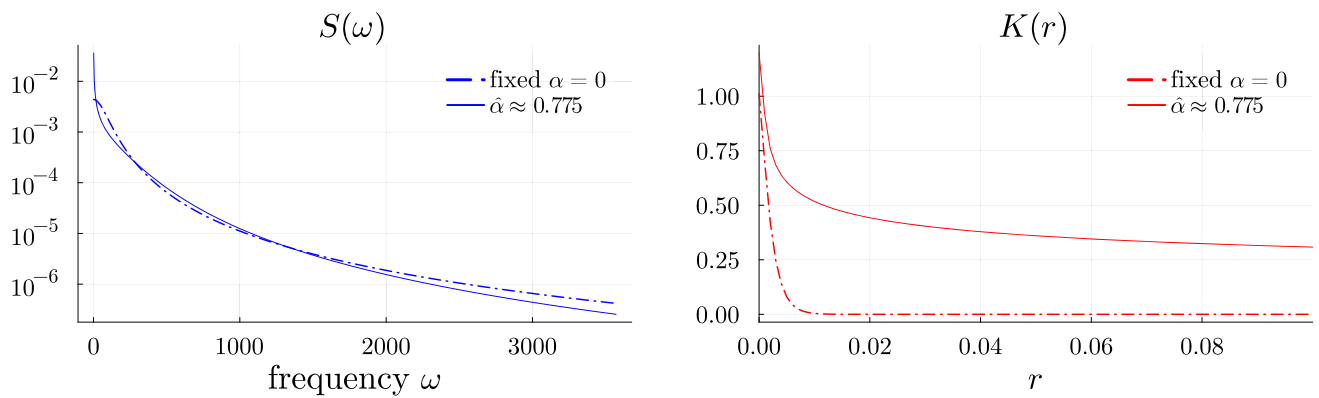
A related observation that is particularly interesting from a theoretical perspective is that, to the degree that the expected Fisher information matrices can be trusted to serve as proxies for the precision of the MLE, the smoothness and singularity parameters  $\nu$  and  $\alpha$  are jointly resolved reasonably well. Table 1 shows that the implied uncertainties for both of those parameters are small, whereas the parameters  $\phi$  and  $\rho$  which now most affect moderate frequencies are much less well-resolved. It is well-known that in most fixed-domain asymptotic regimes in fewer than four dimensions, no individual parameter besides the smoothness  $\nu$  in the Matérn model can be estimated consistently (Ibragimov and Rozanov

1978; Stein 1999; Zhang 2004), and in light of this observation the high uncertainty in  $\phi$  and  $\rho$  is not surprising. For the singular Matérn model in particular, however,  $\rho$  serves the important purpose of making  $S$  bounded at the origin, so that the singularity parameter  $\alpha$  can be disentangled from the effect of  $\nu$  on the tail decay in the model  $|\omega|^{-\alpha} S(\omega)$ .

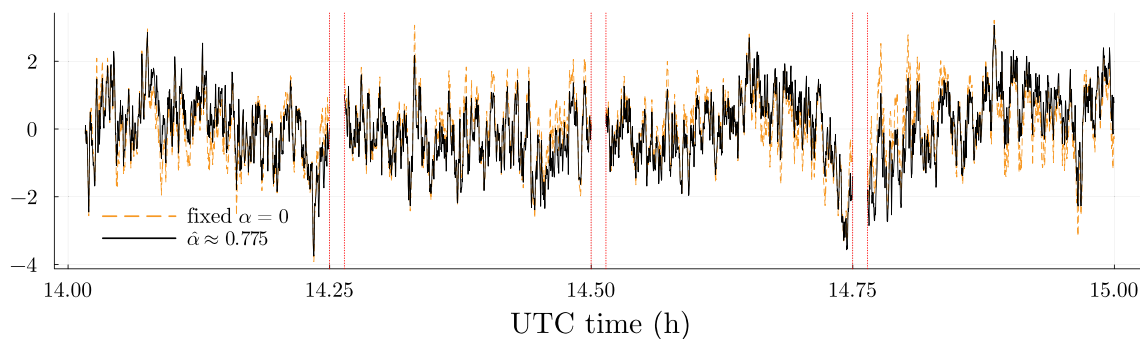
To understand what these estimates imply about the spectral densities and covariances of the processes, Fig. 10 shows the MLE-implied spectra and kernel values from both models. As can be seen, the singular model moves much of the spectral mass into the singularity near the origin, and the more rapid decay in the tails of the singular Matérn spectral density is clearly visible. The plot of the implied kernel values in the center of Fig. 10 shows a more dramatic difference, as the estimated standard Matérn kernel is highly concentrated at the origin, while the singular Matérn kernel displays slow decay. Finally, Fig. 11 shows sample paths from the two MLE-implied models computed with the same white noise forcing. As Table 1 and Fig. 10 would suggest, the additional high-frequency information in the standard Matérn model is quite prominent.

## 5 Discussion

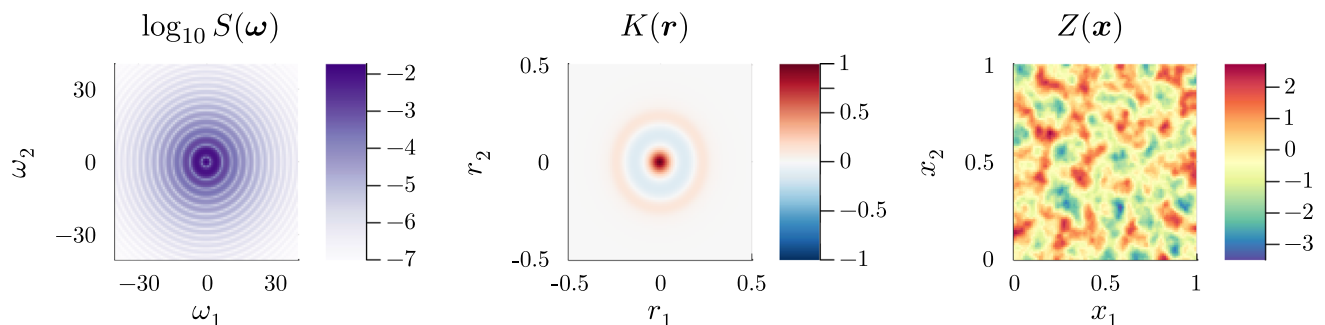
In this work we introduce a numerical method for efficiently and accurately evaluating the Fourier transform of spectral densities, even those that are barely integrable due to singularities at the origin or slowly decaying tails. Making this strategy computational practical requires several technical observations. The first and most crucial is that high-order quadrature rules can significantly accelerate convergence. The use of the NUFFT is critical for this purpose, as it reduces the computational cost to quasilinear complexity in both the number of quadrature nodes  $m$  and the number of inter-observation distances  $n$ . The second vital observation is that the covariance function  $K$  is generally resolved to a given tolerance  $\epsilon$  more quickly at distances  $r$  that are well-separated from the origin. Therefore one can adaptively increase the width of panels being integrated as one reduces the highest frequency oscillations in the integrands remaining to be resolved. This Nyquist-based observation is essential to overcoming the difficulties of slowly decaying spectral densities, which may take orders of magnitude longer to converge if one were to use a more uniform quadrature rule. Finally, the design of an efficient but precise stopping criterion based on both truncation error *and* panel contribution is imperative to keeping the routine performant but accurate for a wide variety of spectral densities. Combining all of these observations and the additional technical discussions provided in Sect. 2, one can evaluate covariance functions specified by these spectral densities at tens of millions of locations in seconds on a laptop.



**Fig. 10** Left: Model-implied spectral densities for the data on June 03 and the parameter estimates in Table 1. Right: the corresponding model-implied kernels



**Fig. 11** Sample paths with identical white noise forcing of the processes implied by the standard Matérn MLE (orange dashes) and the singular Matérn MLE (solid black), showing the noticeable affect of the more rapid spectral decay implied by the singular Matérn model



**Fig. 12** Log of spectral density (left), corresponding covariance function (center), and sample from the process (right) using the isotropic “oscillatory” Matérn model (3.2) in two dimensions

The cost of using this method to compute kernel values will naturally be significantly higher than direct kernel evaluation when a simple closed form exists, for example the standard Matérn model. Although the evaluation of  $\mathcal{K}_\nu$  makes the Matérn covariance numerically challenging, it can be computed to nearly machine precision in less than a microsecond due to advances in modern special function libraries (Geoga et al. 2023). Therefore, in such cases where efficient special function implementations can be used, direct kernel evaluation will likely be both faster and more accurate than even the  $\epsilon = 10^{-12}$  given by our method. Yet

for the vast majority of spectral densities whose Fourier transforms are completely unavailable in closed form, our method is a robust, broadly applicable option for obtaining direct kernel evaluations for irregularly sampled data which can be composed with a practitioner’s estimation method of choice. The code and details of the integration strategy can surely be improved, further reducing the overall cost of this approach—but even now it makes many modeling choices available that were previously effectively impossible. While the LiDAR application in this work is focused on the singular Matérn model, we again remind the reader that the

objective of this framework is to empower practitioners to work with any continuous, integrable spectral density. The extremely slow decay of a long-memory covariance function is of course not always desirable, and there are many other new ways to add valuable degrees of freedom to parametric families of spectral densities. It is our hope that this method and the accompanying software will inspire others to explore and discover new parametric families of models useful to their own contexts.

We emphasize that our framework can be used in conjunction with a number of existing methods for large scale and nonstationary GPs. First, because it requires only a spectral density and a list of distances  $r_j$  at which to evaluate the corresponding covariance function, our method can be used with any approximation strategy for which these distances are specified in advance. Examples include Vecchia approximation methods (Vecchia 1988; Stein et al. 2004; Katzfuss and Guinness 2021) and entry-based rank-structured covariance matrix approximations (Chen and Stein 2023; Geoga et al. 2020; Beckman et al. 2023). Second, we note that while our approach is of course intrinsically limited to evaluating stationary covariance functions, it is compatible with a number of methods which construct nonstationary models from stationary ones, including warping (Sampson and Guttorp 1992) and convolutional frameworks (Paciorek and Schervish 2006).

Finally, it is natural to ask how this methodology can be extended to multiple dimensions. For any isotropic spectral density  $S(\omega) = S(\|\omega\|)$  in  $\mathbb{R}^d$  one can integrate out the radial variables, resulting in the covariance function

$$K(r) = \frac{2\pi}{r^{\frac{d}{2}-1}} \int_0^\infty S(\omega) J_{\frac{d}{2}-1}(2\pi\omega r) \omega^{\frac{d}{2}} d\omega \tag{5.1}$$

where  $J_\nu$  is the Bessel function of the first kind of order  $\nu$ . We can compute this integral using the one-dimensional adaptive integration framework just described, with the only modification being that the NUFFT is replaced by a nonuniform fast Hankel transform. While there exist a number of fast Hankel transform algorithms (Cree and Bones 1993; Townsend 2015), we are not aware of a fully nonuniform method in which  $r$  and  $\omega$  can be chosen freely to accommodate the use of Gaussian quadrature rules for irregularly sampled data. Without this fast transform, we are restricted to direct  $\mathcal{O}(nm)$  summation to compute each panel integral. Figure 12 shows an example of a 2D isotropic kernel computed with direct summation. The case of anisotropic spectral densities is more complicated, and requires multi-dimensional quadrature, but the modeling options it allows would be enormous.

### Appendix: Proofs

**Proof of Lemma 1** Using the integral representation 8.6.4 from Olver (2010),

$$|\Gamma(-s, iy)| = \left| \frac{(iy)^{-s} \exp(-iy)}{\Gamma(s+1)} \int_0^\infty \frac{t^s \exp(-t)}{t+iy} dt \right| \tag{A.2}$$

$$\leq \frac{y^{-s}}{\Gamma(s+1)} \int_0^\infty \frac{t^s \exp(-t)}{|t+iy|} dt \tag{A.3}$$

$$\leq \frac{y^{-s}}{\Gamma(s+1)} \int_0^\infty \frac{t^s \exp(-t)}{y} dt \tag{A.4}$$

$$= \frac{y^{-s-1}}{\Gamma(s+1)} \int_0^\infty t^s \exp(-t) dt \tag{A.5}$$

$$= y^{-s-1}, \tag{A.6}$$

where  $\Gamma(z) := \Gamma(0, z)$  is the usual gamma function. By a similar argument, we obtain

$$|\Gamma(-s, iy)| \leq \frac{y^{-s}}{\Gamma(s+1)} \int_0^\infty \frac{t^s \exp(-t)}{t} dt \tag{A.7}$$

$$= \frac{y^{-s}}{\Gamma(s+1)} \Gamma(s) \tag{A.8}$$

$$= \frac{y^{-s}}{s} \tag{A.9}$$

using the fundamental property of the Gamma function  $\Gamma(z+1) = z\Gamma(z)$ .  $\square$

**Proof of Theorem 2** Define the covariance function  $K := \mathcal{F}^{-1}\{S\}$  corresponding to  $S$ . As  $S$  is integrable,  $K$  is well defined and bounded

$$|K(r)| \leq \int_{-\infty}^\infty S(\omega) |\cos(2\pi\omega r)| d\omega = \|S\|_{L^1(\mathbb{R})} < \infty. \tag{A.10}$$

As  $S \in C^2(\mathbb{R})$  and  $S', S'' \in L^1(\mathbb{R})$ , integrating by parts twice and applying the Riemann-Lebesgue lemma gives the standard decay rate  $K(r) = o(r^{-2})$  for  $r \rightarrow \infty$ . Therefore  $K$  is integrable with

$$M := \int_{-\infty}^\infty K(t) dt < \infty. \tag{A.11}$$

Using the fact that

$$\mathcal{F}^{-1}\{|\cdot|^{-\alpha}\}(r) = (2\pi)^\alpha \frac{\Gamma(1-\alpha)}{\pi} \sin\left(\frac{\pi\alpha}{2}\right) |r|^{-1+\alpha} \tag{A.12}$$

in the distributional sense and applying the convolution theorem, we obtain

$$K_\alpha(r) = \mathcal{F}^{-1}\{|\cdot|^{-\alpha} S(\cdot)\}(r) \tag{A.13}$$

$$= \left( \mathcal{F}^{-1}\{|\cdot|^{-\alpha}\} * \mathcal{F}^{-1}\{S\} \right)(r) \tag{A.14}$$

$$\propto \left( |\cdot|^{-1+\alpha} * K \right)(r) \tag{A.15}$$

$$= \int_{-\infty}^{\infty} |t|^{-1+\alpha} K(r-t) dt. \tag{A.16}$$

Take  $0 < c < (1 - \alpha)^{\frac{1}{2-\alpha}} < 1$ . First consider the upper tail of the integral (A.16). Taylor expanding around  $t = r$  gives

$$\begin{aligned} & \int_{cr}^{\infty} t^{-1+\alpha} K(r-t) dt \\ &= \int_{cr}^{\infty} \left( r^{-1+\alpha} - (1-\alpha)\xi^{-2+\alpha}(t-r) \right) K(r-t) dt \end{aligned} \tag{A.17}$$

for some  $\xi$  between  $r$  and  $t$ . This results in two terms. As  $K$  is integrable, the first term gives the desired asymptotic behavior

$$\begin{aligned} r^{-1+\alpha} \int_{cr}^{\infty} K(r-t) dt &= r^{-1+\alpha} \left( M - \int_{(1-c)r}^{\infty} K(u) du \right) \\ &\sim r^{-1+\alpha}. \end{aligned} \tag{A.18}$$

As  $\xi \geq cr$  and  $K(r) = o(r^{-2})$  for  $r \rightarrow \infty$ , the second term decays with at least this asymptotic rate

$$(1-\alpha) \int_{cr}^{\infty} \xi^{-2+\alpha}(t-r) K(r-t) dt \lesssim r^{-1+\alpha} \tag{A.19}$$

and by our choice of  $c$ , is strictly smaller in magnitude than (A.18), which avoids cancellation. Next consider the lower tail of (A.16)

$$\begin{aligned} \int_{-\infty}^{-cr} (-t)^{-1+\alpha} K(r-t) dt &\leq r^{-1+\alpha} \int_{(1+c)r}^{\infty} K(u) du \\ &\lesssim r^{-2+\alpha} \end{aligned} \tag{A.20}$$

again due to the fact that  $K(r) = o(r^{-2})$  for  $r \rightarrow \infty$ . Finally consider the central term in (A.16). Define the interval  $I_r := [(1-c)r, (1+c)r]$ . Then we have

$$\int_{-cr}^{cr} t^{-1+\alpha} K(r-t) dt \leq \|K\|_{L^\infty(I_r)} \frac{2}{\alpha} (cr)^\alpha \lesssim r^{-2+\alpha} \tag{A.21}$$

as  $\|K\|_{L^\infty(I_r)} = o(r^{-2})$  for  $r \rightarrow \infty$ . □

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11222-024-10519-0>.

**Acknowledgements** The authors would like to thank Michael L. Stein and Michael O’Neil for their helpful suggestions and feedback

**Declarations**

**Conflict of interest** The authors report no conflict of interest.

**References**

Barnett, A., Greengard, P., Rachh, M.: Uniform approximation of common gaussian process kernels using equispaced Fourier grids. (2023). arXiv preprint [arXiv:2305.11065](https://arxiv.org/abs/2305.11065)

Barnett, A., Magland, J., af Klinteberg, L.: A parallel nonuniform fast Fourier transform library based on an “exponential of semicircle” kernel. *SIAM Journal on Scientific Computing* **41**(5), C479–C504 (2019)

Beckman, P., Geoga, C., Stein, M., Anitescu, M.: Scalable computations for nonstationary Gaussian processes. *Stat. Comput.* **33**(4), 84 (2023)

Bingham, N., Goldie, C., Teugels, J.: *Regular variation*. Number 27. Cambridge university press (1989)

Byrd, R., Nocedal, J., Waltz, R.: Knitro: an integrated package for nonlinear optimization. In: Di Pillo, G., Roma, M. (eds.) *Large-Scale Nonlinear Optimization, Nonconvex Optimization and Its Applications*, pp. 35–59. Springer, US, Boston, MA (2006)

Chen, J., Stein, M.: Linear-cost covariance functions for Gaussian random fields. *J. Am. Stat. Assoc.* **118**(541), 147–164 (2023)

Cree, M.J., Bones, P.J.: Algorithms to numerically evaluate the Hankel transform. *Computers & Mathematics with Applications* **26**(1), 1–12 (1993)

Cressie, N., Huang, H.: Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Am. Stat. Assoc.* **94**(448), 1330–1339 (1999)

Dutt, A., Rokhlin, V.: Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* **14**(6), 1368–1393 (1993)

Geoga, C., Anitescu, M., Stein, M.: Scalable Gaussian process computations using hierarchical matrices. *J. Comput. Graph. Stat.* **29**(2), 227–237 (2020)

Geoga, C., Anitescu, M., Stein, M.: Flexible nonstationary spatiotemporal modeling of high-frequency monitoring data. *Environmetrics* **32**(5), e2670 (2021)

Geoga, C., Marin, O., Schanen, M., Stein, M.: Fitting Matérn smoothness parameters using automatic differentiation. *Stat. Comput.* **33**(2), 48 (2023)

Geoga, C., Stein, M.: A scalable method to exploit screening in Gaussian process models with noise. *Journal of Computational and Graphical Statistics*, (just-accepted):1–19 (2023)

Glaser, A., Liu, X., Rokhlin, V.: A fast algorithm for the calculation of the roots of special functions. *SIAM J. Sci. Comput.* **29**(4), 1420–1438 (2007)

Gonnet, P.: A review of error estimation in adaptive quadrature. *ACM Computing Surveys (CSUR)* **44**(4), 1–36 (2012)

- Grainger, J., Sykulski, A., Jonathan, P., Ewans, K.: Estimating the parameters of ocean wave spectra. *Ocean Eng.* **229**, 108934 (2021)
- Greengard, L., Lee, J., Inati, S.: The fast sinc transform and image reconstruction from nonuniform samples in k-space. *Communications in Applied Mathematics and Computational Science* **1**(1), 121–131 (2007)
- Greengard, P.: Efficient fourier representations of families of gaussian processes. (2021). arXiv preprint [arXiv:2109.14081](https://arxiv.org/abs/2109.14081)
- Greengard, P., Rachh, M., Barnett, A.: Equispaced Fourier representations for efficient Gaussian process regression from a billion data points. (2022). arXiv preprint [arXiv:2210.10210](https://arxiv.org/abs/2210.10210)
- Griewank, A., Walther, A.: *Evaluating derivatives: principles and techniques of algorithmic differentiation*. SIAM (2008)
- Guinness, J.: Spectral density estimation for random fields via periodic embeddings. *Biometrika* **106**(2), 267–286 (2019)
- Guinness, J.: Gaussian process learning via Fisher scoring of Vecchia's approximation. *Stat. Comput.* **31**(3), 25 (2021)
- Hale, N., Townsend, A.: Fast and accurate computation of Gauss-Legendre and Gauss-Jacobi quadrature nodes and weights. *SIAM J. Sci. Comput.* **35**(2), A652–A674 (2013)
- Hensman, J., Durrande, N., Solin, A.: Variational Fourier features for Gaussian processes. *J. Mach. Learn. Res.* **18**(151), 1–52 (2018)
- Horrell, M., Stein, M.: Half-spectral space-time covariance models. *Spatial. Statistics* **19**, 90–100 (2017)
- Ibragimov, I., Rozanov, Y.: *Gaussian random processes*. Springer-Verlag, New York, Stochastic Modelling and Applied Probability (1978)
- Im, H., Stein, M., Zhu, Z.: Semiparametric estimation of spectral density with irregular observations. *J. Am. Stat. Assoc.* **102**(478), 726–735 (2007)
- Johansson, F.: Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE T Comput* **66**, 1281–1292 (2017)
- Katzfuss, M., Guinness, J.: A general framework for Vecchia approximations of Gaussian processes (2021)
- Li, Y., Yang, H.: Interpolative butterfly factorization. *SIAM J. Sci. Comput.* **39**(2), A503–A531 (2017)
- Muradyan, P., Coulter, R.: *Radar Wind Profiler (RWP) and Radio Acoustic Sounding System (RASS) Instrument Handbook (DOE/SC-ARM-TR-044)*. DOE Office of Science Atmospheric Radiation Measurement (ARM) Program, United States (2020)
- Newsom, R.: Doppler lidar (DL) handbook. Technical Report DOE/SC-ARM/TR-101, DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States) (2012)
- Olver, F.: *NIST handbook of mathematical functions* hardback and CD-ROM. Cambridge University Press (2010)
- O'Neil, M., Woolfe, F., Rokhlin, V.: An algorithm for the rapid evaluation of special function transforms. *Appl. Comput. Harmon. Anal.* **28**(2), 203–226 (2010)
- Paciorek, C. J., Schervish, M. J.: Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, 17(5):483–506 (2006)
- Porcu, E., Stein, M.: On some local, global and regularity behaviour of some classes of covariance functions. In *Advances and challenges in space-time modelling of natural events*, pages 221–238. Springer (2012)
- Rahimi, A., Recht, B.: Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20 (2007)
- Sampson, P., Guttorp, P.: Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Stat. Assoc.* **87**(417), 108–119 (1992)
- Stein, M.: *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media (1999)
- Stein, M.: Statistical methods for regular monitoring data. *J. R. Stat. Soc. Ser. B Stat Methodol.* **67**(5), 667–687 (2005)
- Stein, M., Chi, Z., Welty, L.: Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B Stat Methodol.* **66**(2), 275–296 (2004)
- Sykulski, A., Olhede, S., Guillaumin, A., Lilly, J., Early, J.: The debiased whittle likelihood. *Biometrika* **106**(2), 251–266 (2019)
- Townsend, A.: A fast analysis-based discrete Hankel transform using asymptotic expansions. *SIAM J. Numer. Anal.* **53**(4), 1897–1917 (2015)
- Trefethen, L.: *Approximation theory and approximation practice, extended edition*. SIAM (2019)
- Vecchia, A.: Estimation and model identification for continuous spatial processes. *J. R. Stat. Soc. Ser. B Stat Methodol.* **50**(2), 297–312 (1988)
- Whittle, P.: Hypothesis testing in time series analysis (1951)
- Whittle, P.: Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute* **40**(2), 974–994 (1963)
- Wilson, A., Adams, R.: Gaussian process kernels for pattern discovery and extrapolation. In *International conference on machine learning*, pages 1067–1075. PMLR (2013)
- Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Stat. Assoc.* **99**(465), 250–261 (2004)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.