A generalized quadratic estimate for random field nonstationarity †

Ethan Anderes[1,*] **and Joe Guinness**[2]

[1]*Department of Statistics, University of California, Davis CA 95616, USA. e-mail:* anderes@ucdavis.edu

[2]*Department of Statistics, North Carolina State University. e-mail:* jsguinne@ncsu.edu

Abstract: In this paper, we attempt to shed light on a new class of nonstationary random fields which exhibit, what we call, *local invariant nonstationarity*. We argue that the local invariant property has a special interaction with a new generalized quadratic estimate—also derived here—which extends an estimate originally developed for gravitational lensing of the Cosmic Microwave Background in Cosmology [10, 11]. The nature of this interaction not only encourages low estimation bias but also enables accurate (and fast) quantification of Frequentist mean square error quantification of the estimated nonstationarity. These quadratic estimates are interesting, in their own right, as they detect and estimate nonstationarity by probing correlation among Fourier frequencies, the absence of which is the characterizing feature of weak stationarity (by Bochner's Theorem). Moreover, this generalized quadratic estimate can be computed with a Fourier characterization that runs in $\mathcal{O}(n \log n)$ time when observing the field on a uniform grid of size n in $\mathbb{R}^d$. Finally, the work presented here partially addresses two other problems associated with the statistical theory of nonstationarity: 1) estimating the phase of a spatially varying modulated stationary random field and 2) identifying a larger class of nonstationary random fields which admit an extension of the quadratic estimator of gravitational lensing that extends the same attractive statistical properties.

Contents

*EA was partially supported by NSF CAREER grant DMS-1252795

†All the source code used for the simulations and graphics in this paper is publicly available through the on-line repository https://github.com/EthanAnderes/NonstationaryPhase.jl

1. Introduction

Many data sets in time series and spatial statistics show clear signs of nonstationarity [20, 6]. Despite the prevalence of nonstationary data, theory for understanding and estimating nonstationary random field models is still underdeveloped compared to what is known for stationary random fields. For example, there seems to be no consensus among statisticians as to the relative merits of various nonstationary models and their corresponding estimates found in the current literature (examples of such models can be found in [9, 21, 15, 5]). While powerful spectral techniques have been developed for deriving absolute continuity or orthogonality of stationary random fields [12] and for understanding the implications for spatial interpolation [22], we know quite little about such topics for nonstationary random fields. A further complicating matter is that even if the parametric form for the nonstationary data generating mechanism is known, difficulties associated with the inevitable increase in the number of unknown parameters and computational complexity can outweigh the benefits of fitting the true nonstationary model. The situation is far worse for spatial statistics—as compared to time series—where irregularity of spatial observation locations and large boundary effects can make estimation and modeling of nonstationarity more difficult.

In this paper, we attempt to shed light on a new class of nonstationary random fields which exhibit, what we call, *local invariant nonstationarity*. Formally defined in Section 2, a locally invariant nonstationary random field $\{Z(\boldsymbol{x})\colon \boldsymbol{x} \in \mathbb{R}^d\}$ has the feature that the covariance function can be written in the following form,

$$\operatorname{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) = K(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{y})),$$

where $K(\cdot,\cdot)\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{\theta}(\cdot)\colon \mathbb{R}^d \to \mathbb{R}^d$ is a vector field which characterizes the nonstationarity in $Z(\boldsymbol{x})$. We argue that the locally invariant property has a special interaction with a new generalized quadratic estimate (derived in Section 2.1) which extends an estimate originally developed for gravitational lensing of the Cosmic Microwave Background in Cosmology [10, 11]. The nature of this interaction not only encourages low estimation bias but also enables accurate (and fast) Frequentist mean square error quantification of the nonstationarity characterized by $\boldsymbol{\theta}(\boldsymbol{x})$. These quadratic estimates are interesting in their own right, as they detect and estimate nonstationarity by probing correlation among Fourier frequencies, the absence of which is the characterizing feature of weak stationarity (by Bochner's Theorem). Moreover, this generalized quadratic estimate can be computed with a Fourier characterization that runs in $\mathcal{O}(n \log n)$ time and $\mathcal{O}(n)$ storage when observing the field on a uniform grid of size n in $\mathbb{R}^d$. Finally, the work presented here partially addresses two unanswered questions that arise in two important bodies of work on nonstationarity: 1) estimating the spatially varying phase in the models analyzed by Dahlhaus [2, 3] and 2) how to extend the quadratic estimate of dark matter from the Cosmic Microwave Background to more general nonstationary random fields.

The seminal work of Dahlhaus in the time series literature [2, 3] is an example of a collection of results that hint at a more general statistical theory of nonstationarity. Dahlhaus develops asymptotic theory for a particular class of nonstationary random fields modeled by a spatially (or temporally) varying spectral density. These random fields were originally developed for time series [18, 19] and have the form

$$Z(\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{x}\cdot\boldsymbol{k}} A(\boldsymbol{k}, \boldsymbol{x}) \sqrt{C_{\boldsymbol{k}}} \frac{dW_{\boldsymbol{k}}}{(2\pi)^{d/2}} \tag{1}$$

where $C_{\boldsymbol{k}}$ is a spectral density, $dW_{\boldsymbol{k}}$ is an orthogonal increment random complex measure that satisfies $E|dW_{\boldsymbol{k}}|^2 = d\boldsymbol{k}$ (see [7] for details on random measures) and $A(\boldsymbol{k}, \boldsymbol{x})$ represents a spatial (or temporal) modulation of $\sqrt{C_{\boldsymbol{k}}}$. Dahlhaus proved consistency results for estimating the squared modulus $|A(\boldsymbol{k}, \boldsymbol{x})|^2$ when estimation is done by maximizing a weighted sum of local likelihood functions. Left unanswered, however, is the question of estimating the phase of $A(\boldsymbol{k}, \boldsymbol{x})$. In Section 3, we partly resolve this question by showing that a generalized quadratic estimate can be used to estimate a

pure phase modulation $A(\boldsymbol{k},\boldsymbol{x}) = \exp(iB(\boldsymbol{k},\boldsymbol{x}))$ where $B(\boldsymbol{k},\boldsymbol{x})$ is a function taking values in $\mathbb{R}$ and is a separable function of $\boldsymbol{x}$ and $\boldsymbol{k}$. Indeed, this nonstationary phase model has the local invariance property and, as such, can be accurately estimated (under certain conditions on $B(\boldsymbol{k},\boldsymbol{x})$) with the generalized quadratic estimate developed here.

Another important development in the statistical theory of nonstationarity comes from recent gravitational lensing studies of the Cosmic Microwave Background (CMB) [4, 24, 16, 23, 17]. In Cosmology, gravitational lensing describes the distortion of photon trajectories due to density fluctuations of intervening dark matter. These density fluctuations affect the CMB observations by introducing small nonstationarities in the original isotropic random field model of the CMB. The state-of-the-art estimator of lensing, the quadratic estimator developed by Hu and Okamoto [10, 11], has become an incredibly successful tool for probing the nature of dark matter, understanding cosmic structure and constraining cosmological parameters. What is so surprising about this estimate is that is has small bias. This is due to a delicate cancellation of terms in an infinite Taylor expansion of the lensing effect. Unfortunately there has been no clear explanation as to why this cancellation occurs and whether or not it exists in other models of nonstationarity. In this paper, we argue that this cancellation is due to the fact that the lensing-induced nonstationarity is locally invariant. Moreover, the generalized quadratic estimator developed here extends the lensing estimator to more general settings. Indeed, many of techniques we use to derive the generalized quadratic estimator are inspired by—and closely follow—those used by Hu and Okamoto [10, 11]. The point of this paper, in contrast to the work of Hu and Okamoto, is to identify the cause of the delicate Taylor series cancellation and extend the benefits of the lensing estimator to a larger class of nonstationary random fields available to general practitioners of spatial statistics.

The first part of this paper, given in Section 2, defines the *locally invariant* property and uses it to derive a corresponding generalized quadratic estimate, called the quadratic estimate hereafter, which has particularly low bias. These new estimates are global rather than local in nature and thus avoid complicating theoretical and practical issues related to bandwidth selection. Moreover, they are unique in that they work in the spectral domain by estimating cross correlation of the Fourier coefficients. In Section 2 we also derive analytic approximations for estimation variance and second order bias of the quadratic estimate. These approximations, and indeed the estimate itself, are often very accurate and have Fourier representations that yield fast computation.

The second part of this paper is given in Section 3. Here we analyze random fields which are characterized by a spatially varying spectral phase modulation of a stationary field, called nonstationary spectral phase random fields for the remainder of this paper. These models effectively generalize warping models, are locally invariant and, as such, are amenable to quadratic estimates. In Subsection 3.1 we characterize how the local spectral density of the nonstationary spectral phase model varies as a function of spatial location. The theory of optimal transport and the L_2-Wasserstein metric play an important role in this characterization and leads to a natural heuristic for quantifying estimation bias in terms of the Wasserstein geodesic cut locus (see Claim 3, Section 3.2 and Figure 3). We also present a set of simulations, based on nonstationary spectral phase random fields, which demonstrate the accuracy of both the quadratic estimate and our approximation to the mean squared sampling properties of the quadratic estimate.

1.1. Notation

For readability it will be advantageous to briefly summarize the notational conventions used throughout this paper, some of which are borrowed from Cosmology and somewhat nonstandard in the statistics literature. Variables taking values in $\mathbb{R}^d$ or $\mathbb{R}^{d\times d}$ will generally be written with bold font such as $\boldsymbol{x},\boldsymbol{y},\boldsymbol{k},\boldsymbol{\ell} \in \mathbb{R}^d$ or $\boldsymbol{A},\boldsymbol{M} \in \mathbb{R}^{d\times d}$. Indexing into vector or matrix coordinates are written with subscripts so that $\boldsymbol{x}_i \in \mathbb{R}$ denotes the i^{th} coordinate of $\boldsymbol{x} \in \mathbb{R}^d$, for example. When $\boldsymbol{z},\boldsymbol{w} \in \mathbb{C}^d$ we let the (non-Hermitian) dot product be denoted by $\boldsymbol{z}\cdot\boldsymbol{w} = \boldsymbol{z}_1\boldsymbol{w}_1 + \ldots + \boldsymbol{z}_d\boldsymbol{w}_d$.

Vector fields $\boldsymbol{\theta}(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^d$ are also written bold so that $\boldsymbol{\theta}(\boldsymbol{x}) = (\boldsymbol{\theta}_1(\boldsymbol{x}),\ldots,\boldsymbol{\theta}_d(\boldsymbol{x}))^T$ where $\boldsymbol{\theta}_i(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$. The Fourier transform of $\boldsymbol{\theta}(\boldsymbol{x})$, for example, is applied coordinate-wise and written

$\boldsymbol{\theta}_{\boldsymbol{k}} = (\boldsymbol{\theta}_{1,\boldsymbol{k}}, \ldots, \boldsymbol{\theta}_{d,\boldsymbol{k}})^T$ where $\boldsymbol{\theta}_{j,\boldsymbol{k}}$ denotes the Fourier transform of $\boldsymbol{\theta}_j(\boldsymbol{x})$ and is defined as

$$\boldsymbol{\theta}_{j,\boldsymbol{k}} = \int e^{-i\boldsymbol{x}\cdot\boldsymbol{k}} \boldsymbol{\theta}_j(\boldsymbol{x}) \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}.$$

For a mean-zero stationary random field $\{Z(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$ the autocorrelation function is denoted $C^Z(\boldsymbol{x}-\boldsymbol{y}) := E(Z(\boldsymbol{x})Z(\boldsymbol{y})^*)$ and the spectral density is denoted $C^{ZZ}_{\boldsymbol{\ell}} = (2\pi)^{d/2} C^Z_{\boldsymbol{\ell}}$ where, via our convention, $C^Z_{\boldsymbol{\ell}}$ denotes the Fourier transform of $C^Z(\boldsymbol{x})$. Parenthetical superscripts are reserved for indexing functions (and *not* higher order derivatives). For example $f^{(1)}_{\boldsymbol{k}}, f^{(2)}_{\boldsymbol{k}}, f^{(3)}_{\boldsymbol{k}}, \ldots$ denotes a sequence of functions taking arguments $\boldsymbol{k} \in \mathbb{R}^d$ in the Fourier domain. This convention avoids indexing ambiguities and the subscript convention of the Fourier transform.

In the derivations that follow, one may assume all random fields have periodic boundary conditions on $(-L/2, L/2]^d$, for some large $L > 0$. This alleviates subtleties associated with the Fourier transform of non-periodic random fields defined on $\mathbb{R}^d$. However, extensions to non-periodic random fields can be made with an appropriate use of generalized random fields and generalized Fourier transforms. To incorporate the periodic case and the—possibly generalized—general case we use a single notation for the Fourier transform in both situations. For example when $f(\boldsymbol{x})$ is periodic on $(-L/2, L/2]^d$ the notation $\int e^{-i\boldsymbol{x}\cdot\boldsymbol{k}} f(\boldsymbol{x}) \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$ and $\int e^{i\boldsymbol{x}\cdot\boldsymbol{k}} f_{\boldsymbol{k}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}}$ should be interpreted as notationally equivalent to $\int_{-L/2}^{L/2} \cdots \int_{-L/2}^{L/2} e^{-i\boldsymbol{x}\cdot\boldsymbol{k}} f(\boldsymbol{x}) \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$ and $\sum_{\boldsymbol{k} \in \frac{2\pi}{L}\mathbb{Z}^d} e^{i\boldsymbol{x}\cdot\boldsymbol{k}} f_{\boldsymbol{k}} \frac{(2\pi/L)^d}{(2\pi)^{d/2}}$, respectively. A consequence of this convention is that, in the periodic case, one equates $d\boldsymbol{k}$ with $(2\pi/L)^d$ and, therefore, the Dirac delta function $\delta_{\boldsymbol{k}}$ becomes a regular function taking the value $1/d\boldsymbol{k}$ when $\boldsymbol{k} = \boldsymbol{0}$ and zero otherwise. In particular, if $Z(\boldsymbol{x})$ is a mean zero stationary random field with with periodic boundary conditions on $(-L/2, L/2]^d$ then $E(Z_{\boldsymbol{k}} Z^*_{\boldsymbol{\omega}}) = \delta_{\boldsymbol{k}-\boldsymbol{\omega}} C^{ZZ}_{\boldsymbol{k}}$ and $E(|Z_{\boldsymbol{k}}|^2) = \delta_{\boldsymbol{0}} C^{ZZ}_{\boldsymbol{k}}$.

2. Locally invariant nonstationary random fields

In this section we define a property of nonstationary random fields called *local invariance*. This property, along with a small set of generic model and observational assumptions, appears to be an important ingredient for quadratic estimates of nonstationarity to have low bias. Indeed, this is the main theme of the paper: *that the structure of local invariance encourages bias cancellation*. A secondary theme of this paper is that local invariance provides a flexible restriction to the class of all random field covariance functions whereby making generalized quadratic estimation available to a wide class of nonstationary random field applications.

Definition 1. *Let $C^{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})$ be a positive definite covariance function defined on $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ and parameterized by a vector field $\boldsymbol{\theta}(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^d$. Then $\boldsymbol{\theta}(\boldsymbol{x})$ is said to be a **local invariant for** $C^{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})$ if there exists a function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that*

$$C^{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}) = K\big(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{y})\big). \tag{2}$$

*Equivalently, $Z(\boldsymbol{x})$ **is locally invariant with respect to** $\boldsymbol{\theta}(\boldsymbol{x})$ if $Z(\boldsymbol{x})$ is a random field with covariance function $C^{\boldsymbol{\theta}}$ that satisfies (2).*

The name *local invariance* is intended to express the following fact: any region where $\boldsymbol{\theta}(\boldsymbol{x})$ is constant results in the same local stationary model. In particular, suppose $Z(\boldsymbol{x})$ is a nonstationary random field with covariance function $C^{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y})$ satisfying (2). If $\boldsymbol{\theta}(\boldsymbol{x})$ has no variation (i.e. is constant) over neighborhoods $\Omega_1 \subset \mathbb{R}^d$ and $\Omega_2 \subset \mathbb{R}^d$ then $Z(\boldsymbol{x})$ is locally stationary over Ω_1 and Ω_2 with the same local autocovariance function $K(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{0})$.

Examples of local invariant nonstationary random fields are easy to find. Indeed any warped random field of the form $Z(\boldsymbol{x} + \boldsymbol{\theta}(\boldsymbol{x}))$ is locally invariant with respect to $\boldsymbol{\theta}(\boldsymbol{x})$ when Z is a stationary random field. Another example of a locally invariant model, discussed later in Section 3, is the spatially varying spectral phase model given in (1) where $A(\boldsymbol{k}, \boldsymbol{x}) = \exp(i\boldsymbol{\theta}(\boldsymbol{x}) \cdot \boldsymbol{\eta}_{\boldsymbol{k}})$ and $\boldsymbol{\eta}_{\boldsymbol{k}}$ is a known

function mapping $\mathbb{R}^d$ into $\mathbb{R}^d$ that has odd symmetry about the origin. It is interesting to note that many of the asymptotic results for spatially (or temporally) varying spectral models study the estimation of $|A(\boldsymbol{k},\boldsymbol{x})|^2$ using a local periodogram [2] or a version or the preperiodogram [3]. Since $|A(\boldsymbol{k},\boldsymbol{x})|^2 = 1$ for the spectral phase model (1), these local periodograms to not immediately apply to the estimation of $A(\boldsymbol{k},\boldsymbol{x})$ defined in (1).

In what follows we derive a quadratic estimate of $\boldsymbol{\theta}(\boldsymbol{x})$ for locally invariant nonstationarity of the form given in Definition 1. The estimates are derived under the following observational scenario: a single realization of the nonstationary random field observed on a dense grid with additive stationary noise. These assumptions hold throughout the paper. We list them here to be completely explicit.

Assumption 1. *Let $Z(\boldsymbol{x})$ be a mean zero nonstationary Gaussian random field with local invariant nonstationary covariance function $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})$ satisfying Definition 1. The data field, denoted $Z^{obs}(\boldsymbol{x})$, is observed on a dense regular grid and has the form*

$$Z^{obs}(\boldsymbol{x}) = Z(\boldsymbol{x}) + N(\boldsymbol{x})$$

where $N(\boldsymbol{x})$ is a mean zero stationary Gaussian generalized random noise with spectral density $C^{NN}_{\boldsymbol{k}}$.

Our second assumption is that the local invariant vector field $\boldsymbol{\theta}(\boldsymbol{x})$ can be additionally characterized by an unknown scalar potential function $\phi(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$. This assumption simply reduces the amount of complexity necessary for developing a quadratic estimate of $\boldsymbol{\theta}(\boldsymbol{x})$ while still retaining enough modeling flexibility.

Assumption 2. *Suppose the vector field $\boldsymbol{\theta}(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^d$ is characterized by an unknown scalar potential field $\phi(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}$ along with a known vector field $\boldsymbol{\xi}_{\boldsymbol{k}}$ which satisfies*

$$\boldsymbol{\theta}_{\boldsymbol{k}} := \left(\boldsymbol{\xi}_{1,\boldsymbol{k}}\,\phi_{\boldsymbol{k}}\,,\ \ldots\,,\ \boldsymbol{\xi}_{d,\boldsymbol{k}}\,\phi_{\boldsymbol{k}}\right)^T.$$

For each $j \in \{1,\ldots,d\}$ the coordinate spectral multiplier $\boldsymbol{\xi}_{j,\boldsymbol{k}} : \mathbb{R}^d \to \mathbb{C}$ is assumed to be a Hermitian function of $\boldsymbol{k}$ so that $\boldsymbol{\xi}_{j,-\boldsymbol{k}} = \boldsymbol{\xi}^*_{j,\boldsymbol{k}}$.

Notice that the above scalar potential model includes the case that $\boldsymbol{\theta}(\boldsymbol{x}) = \nabla\phi(\boldsymbol{x})$, by setting $\boldsymbol{\xi}_{\boldsymbol{k}} = i\boldsymbol{k}$, and the case that $\boldsymbol{\theta}(\boldsymbol{x}) = (\phi(\boldsymbol{x}),\ldots,\phi(\boldsymbol{x}))$, by setting $\boldsymbol{\xi}_{\boldsymbol{k}} = (1,\ldots,1)$.

Assumption 3. *There exists a mean zero stationary Gaussian random field prior for the unknown scalar potential $\phi(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}$. Let $C^{\phi}(\boldsymbol{x}-\boldsymbol{y}) = E(\phi(\boldsymbol{x})\phi(\boldsymbol{y}))$ denote the autocovariance function for $\phi(\boldsymbol{x})$ and $C^{\phi\phi}_{\boldsymbol{k}}$ denote the corresponding spectral density for this prior.*

It is important to note that the prior in Assumption 3 is not used to generate a Bayesian posterior. A Bayesian posterior sampling methodology would be an exciting development but not the scope of the current paper. Instead, the prior is only used to approximate the marginal distribution of the data which, in turn, is used to optimize Fourier weights and to generate a variance approximation for the quadratic estimate of $\phi_{\boldsymbol{k}}$. Indeed the quadratic estimate, derived in Section 2.1, is defined to be an unbiased estimate of ϕ (up to first order) regardless of how $C^{\phi\phi}_{\boldsymbol{k}}$ is specified. The only effect of misspecification (or non-existence) of $C^{\phi\phi}_{\boldsymbol{k}}$ will be to generate an estimate which gives too much weight to unruly frequencies and to report a less accurate mean squared error when using the approximations developed in Section 2.3.

Assumptions 1, 2 and 3 are generic and intended to isolate a small set of assumptions for deriving an estimate with small bias. Bias is not universally guaranteed to be small but the generalized quadratic estimates, derived in the next section for local invariance models, often have surprisingly small estimation bias due to the local invariant property.

2.1. A generalized quadratic estimate of $\phi(\boldsymbol{x})$

Based on assumptions 1, 2 and 3 given in the previous section, the first step for deriving a generalized quadratic estimate of $\phi(\boldsymbol{x})$ is to linearly approximate $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})$, with a Taylor expansion in $\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y})$

as follows

$$C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = C^{(0)}(\boldsymbol{x}-\boldsymbol{y}) + \boldsymbol{C}^{(1)}(\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y})) + \mathcal{O}(\phi^2) + \mathcal{O}(\phi^3) + \cdots . \tag{3}$$

where $C^{(0)}\colon \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{C}^{(1)}\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfies $\boldsymbol{C}^{(1)}(-\boldsymbol{x}) = -\boldsymbol{C}^{(1)}(\boldsymbol{x})$. Now, truncating (3) to first order and applying Claim 4 of Appendix A gives the following linear approximation of the cross frequency covariance in the Fourier transform of $Z(\boldsymbol{x})$

$$E\big(Z_{\boldsymbol{k}+\boldsymbol{\ell}}Z_{-\boldsymbol{k}}\big) \approx \phi_{\boldsymbol{\ell}}\left(\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right). \tag{4}$$

Recall Bochner's Theorem (details can be found in [7]) which states that the random field $Z(\boldsymbol{x})$ is stationary if and only if $E\big(Z_{\boldsymbol{k}+\boldsymbol{\ell}}Z_{-\boldsymbol{k}}\big) = 0$ for every $\boldsymbol{\ell} \neq \boldsymbol{0}$. Therefore nontrivial covariance between $Z_{\boldsymbol{k}+\boldsymbol{\ell}}$ and $Z_{-\boldsymbol{k}}$ at some nonzero lag $\boldsymbol{\ell}$ provides a direct probe into the nonstationarity present in $Z(\boldsymbol{x})$. Equation (4) is, therefore, a translation of how local invariant nonstationarity relates to nonzero cross covariance in $Z_{\boldsymbol{k}}$. Using this translation, Claim 5 of Appendix A derives the quadratic estimate of $\phi_{\boldsymbol{\ell}}$. This estimate is effectively an inverse-variance weighted autocovariance estimate in the Fourier domain and is given by

$$\hat{\phi}_{\boldsymbol{\ell}} = A_{\boldsymbol{\ell}}\int \left(\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right)^* \frac{Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}}C^{ZZobs}_{\boldsymbol{k}}}\frac{d\boldsymbol{k}}{(2\pi)^{d/2}}. \tag{5}$$

In the above formula, $A_{\boldsymbol{\ell}}$ is a normalizing constant (see Claim 5 for an exact expression), $C^{ZZobs}_{\boldsymbol{k}}$ is defined to be $C^{ZZm}_{\boldsymbol{k}} + C^{NN}_{\boldsymbol{k}}$ where $C^{ZZm}_{\boldsymbol{k}}$ is the spectral density of $Z(\boldsymbol{x})$ marginalized over the prior for $\phi(\boldsymbol{x})$ given in Assumption 3. Notice that the prior only serves to optimize the weights in $\hat{\phi}_{\boldsymbol{\ell}}$. Indeed, one can easily avoid specifying $C^{\phi\phi}_{\boldsymbol{k}}$ by instead defining $C^{ZZm}_{\boldsymbol{k}}$ to be $(2\pi)^{d/2}C^{(0)}_{\boldsymbol{k}}$ where $C^{(0)}_{\boldsymbol{k}}$ denotes the Fourier transform of $C^{(0)}(\boldsymbol{x})$ from (3).

The normalizing constant $A_{\boldsymbol{\ell}}$ is completely determined by the weights used on the terms $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}$ through the requirement that $\hat{\phi}_{\boldsymbol{\ell}}$ be unbiased up to first order in ϕ, in particular, requiring that $E(\hat{\phi}_{\boldsymbol{\ell}}) = \phi_{\boldsymbol{\ell}} + \mathcal{O}(\phi^2) + \mathcal{O}(\phi^3) + \cdots$. This results in the following an analytic characterization for $A_{\boldsymbol{\ell}}$

$$A_{\boldsymbol{\ell}}\int \frac{\left|\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right|^2}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}}C^{ZZobs}_{\boldsymbol{k}}}\frac{d\boldsymbol{k}}{(2\pi)^{d/2}} = 1 + \mathcal{O}(\phi) + \mathcal{O}(\phi^2) + \cdots . \tag{6}$$

Notice also that one is free to manually change weights used on each term $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}$ in (5). This may be advantageous for optimizing the sampling behavior of $\hat{\phi}_{\boldsymbol{\ell}}$ to specific applications. For example one may want to down-weight corrupted frequencies in a particular experimental setting. In this case, however, the form of the normalizing constant $A_{\boldsymbol{\ell}}$ given in (6) will need to be adjusted accordingly.

One of the advantages of the estimator $\hat{\phi}_{\boldsymbol{\ell}}$, defined in (5), is that there exists a fast algorithm for computing $\hat{\phi}_{\boldsymbol{\ell}}$ at all frequencies $\boldsymbol{\ell}$ simultaneously by alternating pointwise operations in the Fourier domain and the pixel domain. Indeed (5) is equivalent to

$$\hat{\phi}_{\boldsymbol{\ell}} = A_{\boldsymbol{\ell}}\sum_{p=1}^{d}\boldsymbol{\xi}^*_{p,\boldsymbol{\ell}}\int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}}\mathscr{A}(\boldsymbol{x})\mathscr{B}_p(\boldsymbol{x})\frac{d\boldsymbol{x}}{(2\pi)^{d/2}} \tag{7}$$

where $\mathscr{A}_{\boldsymbol{\ell}} := Z^{obs}_{\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$ and $\mathscr{B}_{p,\boldsymbol{\ell}} := i2\,\mathrm{imag}(\boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}})Z^{obs}_{\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$ which can be computed in $\mathcal{O}(n\log(n))$ time (when observing Z^{obs} on a grid of size n) by a sequence of fast Fourier transforms, inverse fast Fourier transforms and pointwise operations.

The derivation above only depends on the local invariance property insofar as it is used to optimize the weights in (5) and the resulting normalizing constant given in (7). Indeed, exactly similar arguments can be made for deriving quadratic estimates of nonstationary models which are not locally

invariant, such as covariance functions of the form $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = K(\boldsymbol{x}-\boldsymbol{y}, \boldsymbol{\theta}(\boldsymbol{x}) + \boldsymbol{\theta}(\boldsymbol{y}))$ for example. The key difference is that the quadratic estimate $\hat{\phi}_{\boldsymbol{\ell}}$ for models which are *not* locally invariant tend to either have a large $\mathcal{O}(\phi^2) + \mathcal{O}(\phi^3) + \cdots$ bias, small signal to noise ratio, or have significant non-Gaussian estimation variability. Local invariant models, in contrast, encourage a significant amount of cancellation occurring within $\mathcal{O}(\phi^2)+\mathcal{O}(\phi^3)+\cdots$ so that bias is small even in the regime of moderately large signal to noise ratio. Moreover, small higher order terms provide a regime where accurate prediction of mean square estimation variability is accurately approximated with easily computable formulas for second moments of estimation variability. This is explored in more detail in Section 2.3 and in the simulation examples presented later.

2.2. The Hu and Okamoto lensing estimate as a special case of $\hat{\phi}_{\boldsymbol{\ell}}$

In this section we show that $\hat{\phi}_{\boldsymbol{\ell}}$, derived in the previous section, is a special case of the original quadratic estimate developed in [10, 11] for Cosmic Microwave Background gravitational lensing. Start by letting $Z(\boldsymbol{x}) = T(\boldsymbol{x} + \nabla\phi(\boldsymbol{x}))$ denote the lensed Cosmic Microwave Background and $\phi(\boldsymbol{x})$ denote the projected gravitational potential in the $\mathbb{R}^2$ flat sky approximation. In the original derivation [10, 11] a Taylor approximation is first applied to $T(\boldsymbol{x} + \nabla\phi(\boldsymbol{x}))$ as follows

$$Z(\boldsymbol{x}) \approx T(\boldsymbol{x}) + \nabla T(\boldsymbol{x}) \cdot \nabla\phi(\boldsymbol{x}) \tag{8}$$

The above linear model is then used to linearly approximate $Z(\boldsymbol{x})Z(\boldsymbol{y})$ by additionally discarding the term $\big(\nabla T(\boldsymbol{x}) \cdot \nabla\phi(\boldsymbol{x})\big)\big(\nabla T(\boldsymbol{y}) \cdot \nabla\phi(\boldsymbol{y})\big)$ which is quadratic in ϕ. Taking Fourier transforms, then an expected value, results in the following approximation

$$E(Z_{\boldsymbol{k}+\boldsymbol{\ell}}Z_{-\boldsymbol{k}}) \approx \frac{\phi_{\boldsymbol{\ell}}}{2\pi}\Big(\boldsymbol{\ell}\cdot(\boldsymbol{k}+\boldsymbol{\ell})C^{TT}_{\boldsymbol{k}+\boldsymbol{\ell}} - \boldsymbol{\ell}\cdot\boldsymbol{k}C^{TT}_{\boldsymbol{k}}\Big). \tag{9}$$

Notice that (9) is a special case of (4), and hence a special case of $\hat{\phi}_{\boldsymbol{\ell}}$ in (5), when setting $\boldsymbol{\xi}_{\boldsymbol{\ell}} = i\boldsymbol{\ell}$ and $\boldsymbol{C}^{(1)}_{\boldsymbol{k}} = \frac{i\boldsymbol{k}}{2\pi}C^{TT}_{\boldsymbol{k}}$.

It is important to notice a particular subtlety when analyzing the accuracy of (9) in terms of the magnitude of the discarded terms in the Taylor approximation (8). This subtlety can be illustrated by assuming the displacement $\nabla\phi(\boldsymbol{x})$ is extremely large and happens to be a *constant function of* $\boldsymbol{x}$. In this case one clearly has $Z(\boldsymbol{x}) \not\approx T(\boldsymbol{x})$, i.e. the zeroth order Taylor approximation completely breaks down. Yet, in a distributional sense, the zeroth order Taylor approximation is perfect since $Z(\boldsymbol{x})$ and $T(\boldsymbol{x})$ have the same finite dimensional distributions (by the fact that $T(\boldsymbol{x})$ is isotropic and $\nabla\phi(\boldsymbol{x})$ is assumed to be constant). Therefore one can not quantify the accuracy of (9) by a map level analysis of the individual discarded terms. In fact we propose that (9) is accurate, not because the map level Taylor approximation (8) is good (for which it is not), but rather because $\nabla\phi(\boldsymbol{x})$ is a locally invariant parameter and the corresponding nonstationary covariance

$$C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = E(Z(\boldsymbol{x})Z(\boldsymbol{y})) = C^T(\boldsymbol{x}-\boldsymbol{y}+\nabla\phi(\boldsymbol{x})-\nabla\phi(\boldsymbol{y}))$$

has an accurate Taylor approximation in $\nabla\phi(\boldsymbol{x}) - \nabla\phi(\boldsymbol{y})$, vis-à-vis (3).

2.3. Variance and bias analytic approximation

In Section 2.1 a fast formula was derived for computing the estimate $\hat{\phi}_{\boldsymbol{\ell}}$ when observing a single realization of $Z^{obs}(\boldsymbol{x})$. The speed at which $\hat{\phi}_{\boldsymbol{\ell}}$ can be computed on a dense observation grid makes it possible to perform large scale Monte Carlo analysis on $\hat{\phi}_{\boldsymbol{\ell}}$ in any experimental setting for which $Z^{obs}(\boldsymbol{x})$ can be easily simulated. In this section we complement a simulation-based method of uncertainty quantification by providing analytic approximations to variance and second order bias of $\hat{\phi}_{\boldsymbol{\ell}}$. These approximations are often very accurate and inherit a similar Fourier representation as (7) for fast computation.

By inspection of (5) one can consider $\hat{\phi}_{\boldsymbol{\ell}}$ as a function of the quadratic form $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{-\boldsymbol{k}}$, integrating over the variable $\boldsymbol{k}$. In what follows we will consider the sampling behavior of $\hat{\phi}_{\boldsymbol{\ell}}$ when replacing $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{-\boldsymbol{k}}$ by some other function $X_{\boldsymbol{k},\boldsymbol{\ell}}$ of two variables $\boldsymbol{k}, \boldsymbol{\ell} \in \mathbb{R}^d$. The following definition sets notation for this operation which is useful for denoting terms which are related to variance and bias of the estimator $\hat{\phi}_{\boldsymbol{\ell}}$ derived in subsections 2.3.1 and 2.3.2.

Definition 2. *For any function $X_{\boldsymbol{k},\boldsymbol{\ell}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ let $\hat{\phi}_{\boldsymbol{\ell}}\{X_{\boldsymbol{k},\boldsymbol{\ell}}\}$ denote the quadratic estimate $\hat{\phi}_{\boldsymbol{\ell}}$ defined in (5) but applied to the function $X_{\boldsymbol{k},\boldsymbol{\ell}}$ rather than $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{-\boldsymbol{k}}$. In particular $\hat{\phi}_{\boldsymbol{\ell}}\{X_{\boldsymbol{k},\boldsymbol{\ell}}\}$ is a function of $\boldsymbol{\ell}$ and satisfies*

$$\hat{\phi}_{\boldsymbol{\ell}}\{X_{\boldsymbol{k},\boldsymbol{\ell}}\} := A_{\boldsymbol{\ell}} \int \left(\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right)^* \frac{X_{\boldsymbol{k},\boldsymbol{\ell}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}}.$$

If, on the other hand, $X_{\boldsymbol{k}}$ and $Y_{\boldsymbol{k}}$ are both function of a single frequency argument $\boldsymbol{k} \in \mathbb{R}^d$ then we define

$$\hat{\phi}_{\boldsymbol{\ell}}\{X,Y\} := A_{\boldsymbol{\ell}} \int \left(\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right)^* \frac{X_{\boldsymbol{k}+\boldsymbol{\ell}} Y_{-\boldsymbol{k}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}}.$$

In the following two sections we derive approximations to the mean squared error and bias when using $\hat{\phi}_{\boldsymbol{\ell}}$ to estimate $\phi_{\boldsymbol{\ell}}$. This comes in the form of two functions $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ and $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ which represent approximations to the spectral density of variance and bias after marginalizing over the unknown $\phi_{\boldsymbol{\ell}}$ using the Gaussian random field prior given in Assumption 3.

2.3.1. Variance spectral density $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$

There are two main contributions to the variability in $\hat{\phi}_{\boldsymbol{\ell}}$. The first source of variability is due to the additive observational noise $N(\boldsymbol{x})$. The second source, sometimes called *shape noise* in Cosmology, is due to the baseline stationary fluctuations in $Z(\boldsymbol{x})$ characterized by the autocovariance function $K(\boldsymbol{x}-\boldsymbol{y}, \boldsymbol{0})$. The spectral density of this shape noise can be approximated by $(2\pi)^{d/2} C^{(0)}_{\boldsymbol{\ell}}$, which corresponds to the zero$^{\text{th}}$ order approximation in (3), or by $C^{ZZm}_{\boldsymbol{\ell}}$ which denotes the spectral density of $Z(\boldsymbol{x})$ marginalized over the prior from Assumption 3. Both these approximations can be used, within the formulas derived below, to give accurate approximations to the mean squared variability in $\hat{\phi}_{\boldsymbol{\ell}}$. However, in the cosmology literature on gravitational lensing, the marginal stationary model for $Z(\boldsymbol{x})$ is typically used, rather than $(2\pi)^{d/2} C^{(0)}_{\boldsymbol{\ell}}$, for shape noise quantification.

To derive $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ first let $X(\boldsymbol{x})$ denote the mean zero stationary Gaussian random field which models the sum of the observational noise $N(\boldsymbol{x})$ and the shape noise discussed in the previous paragraph. By propagating the random field X through the quadratic estimate one obtains an estimate of variability of $\hat{\phi}_{\boldsymbol{\ell}}$ around its expected value. In particular

$$E\big(\big[\hat{\phi}_{\boldsymbol{\ell}} - E(\hat{\phi}_{\boldsymbol{\ell}})\big]\big[\hat{\phi}_{\boldsymbol{\ell}'} - E(\hat{\phi}_{\boldsymbol{\ell}'})\big]^*\big) \approx E\big(\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}\,\hat{\phi}_{\boldsymbol{\ell}'}\{X,X\}^*\big) =: \delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'} C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}} \tag{10}$$

where the existence of the spectral density $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ is guaranteed by the fact that $X(\boldsymbol{x})$ is stationary so that $\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}$ is stationary in the pixel domain (see Claim 6). Depending on which approximation one uses for the baseline stationary fluctuations in $Z(\boldsymbol{x})$, the spectral density of $X(\boldsymbol{x})$ can be defined in one of the following two ways

$$C^{XX}_{\boldsymbol{\ell}} = \begin{cases} C^{NN}_{\boldsymbol{\ell}} + (2\pi)^{d/2} C^{(0)}_{\boldsymbol{\ell}}, & \text{option 1;} \\ C^{NN}_{\boldsymbol{\ell}} + C^{ZZm}_{\boldsymbol{\ell}}, & \text{option 2.} \end{cases} \tag{11}$$

Depending on what definition one uses for $C^{XX}_{\boldsymbol{\ell}}$, Claim 6 establishes that

$$C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}} = 2A^2_{\boldsymbol{\ell}} \int \left|\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\right|^2 \frac{C^{XX}_{\boldsymbol{k}+\boldsymbol{\ell}}}{(C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}})^2} \frac{C^{XX}_{\boldsymbol{k}}}{(C^{ZZobs}_{\boldsymbol{k}})^2} \frac{d\boldsymbol{k}}{(2\pi)^d}. \tag{12}$$

In certain situations the right hand side of (12) can be simplified. Recall that in the definition of $\hat{\phi}_{\boldsymbol{\ell}}$, one has two options for defining $C_{\boldsymbol{\ell}}^{ZZobs}$, either $C_{\boldsymbol{\ell}}^{NN} + (2\pi)^{d/2} C_{\boldsymbol{\ell}}^{(0)}$ or $C_{\boldsymbol{\ell}}^{NN} + C_{\boldsymbol{\ell}}^{ZZm}$, depending on if one wants to use the prior spectral density $C_{\boldsymbol{\ell}}^{\phi\phi}$ for optimizing the quadratic estimate weights. If the choice of $C_{\boldsymbol{\ell}}^{ZZobs}$ matches the choice of $C_{\boldsymbol{\ell}}^{XX}$ then one obtains the following simplification of (12)

$$C_{\boldsymbol{\ell}}^{\text{var }\hat{\phi}} = 2(2\pi)^{-d/2} A_{\boldsymbol{\ell}}. \tag{13}$$

It should be noted that when marginalizing over the prior given in Assumption 3 the process $Z(\boldsymbol{x})$ becomes stationary but non-Gaussian. On the other hand, when conditioning on ϕ, the process $Z(\boldsymbol{x})$ is Gaussian but nonstationary. Therefore, when using option 2 in equation (11) to model $C_{\boldsymbol{\ell}}^{\text{var }\hat{\phi}}$, the approximation in (10) includes a Gaussian approximation to $X(\boldsymbol{x})$. Finally, we mention that Claim 6 also gives a Fourier based characterization for fast computation of $C_{\boldsymbol{\ell}}^{\text{var }\hat{\phi}}$.

2.3.2. Bias spectral density $C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}}$

The higher order terms $\mathcal{O}(\phi^n)$ in (3) are the exclusive source of bias in the quadratic estimate. The relation between $\mathcal{O}(\phi^n)$ and estimation bias can be written as follows

$$E(\hat{\phi}_{\boldsymbol{\ell}}|\phi) - \phi_{\boldsymbol{\ell}} = \hat{\phi}_{\boldsymbol{\ell}}\{\mathcal{O}(\phi^2)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}}\} + \hat{\phi}_{\boldsymbol{\ell}}\{\mathcal{O}(\phi^3)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}}\} + \cdots$$

where $\mathcal{O}(\phi^n)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}}$ is defined to be Fourier transform of $\mathcal{O}(\phi^n)(\boldsymbol{x},\boldsymbol{y})$, defined in (3), and evaluated at frequencies $\boldsymbol{k}+\boldsymbol{\ell}$ and $-\boldsymbol{k}$, respectively. For the local invariant models we consider here, the second order bias term has the following form

$$\mathcal{O}(\phi^2)(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{y}))^T \boldsymbol{C}^{(2)}(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{y})) \tag{14}$$

where $\boldsymbol{C}^{(2)}\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a symmetric function about the origin. This expression makes it clear how local invariant nonstationarity encourages low quadratic estimation bias. When the observational noise level is small, the high frequency fluctuations in the observations $Z^{obs}(\boldsymbol{x})$ are more influential to the quadratic estimate. At these small scales the smoothness of $\boldsymbol{\theta}(\boldsymbol{x})$ and the function $\boldsymbol{C}^{(2)}(\boldsymbol{x}-\boldsymbol{y})$ will attenuate the influence of $\mathcal{O}(\phi^2)$ when propagated through $\hat{\phi}_{\boldsymbol{\ell}}$.

For remainder of this section we analyze how the second order term (14) propagates to second order bias in the quadratic estimate, denoted $\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}}$. Claim 7 in the Appendix gives the following expression for $\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}}$

$$\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}} = \hat{\phi}_{\boldsymbol{\ell}}\{\mathcal{O}(\phi^2)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}}\} = 2\sum_{p,q=1}^{d} \int \boldsymbol{\theta}_{p,\boldsymbol{\omega}}\boldsymbol{\theta}_{q,\boldsymbol{\ell}-\boldsymbol{\omega}}\, \hat{\phi}_{\boldsymbol{\ell}}\{\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\}\frac{d\boldsymbol{\omega}}{(2\pi)^{d/2}}. \tag{15}$$

Moreover the marginal expected value of this bias term satisfies $E\big(\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}}\big) = 0$ when $\boldsymbol{\ell} \neq 0$. Therefore, to quantify the marginal magnitude of the second order bias one must use the variance of (15). This is done in Claim 7 and establishes that when $\boldsymbol{\theta}(\boldsymbol{x})$ is mean zero Gaussian random field with spectral density matrix $C_{\boldsymbol{\ell}}^{\boldsymbol{\theta}\boldsymbol{\theta}}$ the corresponding spectral density for $\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}}$, denoted $C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}}$, satisfies

$$C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}} = 4\sum_{p,q,p',q'=1}^{d} \int \Big(C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,p',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,q',\boldsymbol{\ell}-\boldsymbol{\omega}} + C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,q',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,p',\boldsymbol{\ell}-\boldsymbol{\omega}}\Big) \times \hat{\phi}_{\boldsymbol{\ell}}\{\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\}\hat{\phi}_{\boldsymbol{\ell}}\{\boldsymbol{C}^{(2)}_{p',q',\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p',q',\boldsymbol{k}+\boldsymbol{\omega}}\}^* \frac{d\boldsymbol{\omega}}{(2\pi)^d} \tag{16}$$

when $\boldsymbol{\ell} \neq \mathbf{0}$. Notice that $\hat{\phi}_{\boldsymbol{\ell}}^{\text{bias}}$ equals the exact, map level, bias contribution from the second order term $\mathcal{O}(\phi^2)$. Therefore the statement that $C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}}$ is an approximation to the second order bias only

refers to the fact that it marginally quantifies the impact of the second order term $\mathcal{O}(\phi^2)$ rather than the all order bias $\mathcal{O}(\phi^2) + \mathcal{O}(\phi^3) + \cdots$.

In contrast to $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$, which can be computed quickly using forward and inverse Fourier transformations, the calculation of $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ appears to require explicit looping over $\boldsymbol{\omega}$ for each $\boldsymbol{\ell}$. This is problematic when $Z(\boldsymbol{x})$ is observed on a high dimensional dense grid. However, there is an approximation to $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ which is both fast and yields excellent numerical accuracy for frequencies $\boldsymbol{\ell}$ with small to moderate magnitude. The approximation is derived with a second order Taylor approximation $\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}} \approx \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} + \nabla \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} \boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\omega}^T \nabla^2 \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} \boldsymbol{\omega}$ so that

$$\hat{\phi}_{\boldsymbol{\ell}}\big\{\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\big\} \approx -\hat{\phi}_{\boldsymbol{\ell}}\big\{\nabla \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}}\big\}\,\boldsymbol{\omega} - \tfrac{1}{2}\boldsymbol{\omega}^T\,\hat{\phi}_{\boldsymbol{\ell}}\big\{\nabla^2 \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}}\big\}\,\boldsymbol{\omega}. \tag{17}$$

The advantage being that $\hat{\phi}_{\boldsymbol{\ell}}\big\{\nabla \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}}\big\}$ and $\hat{\phi}_{\boldsymbol{\ell}}\big\{\nabla^2 \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}}\big\}$ only need to be computed once and can therefore be factored out of the integral (16). The factored integral is then recognized as a convolution which can be quickly computed using forward and inverse Fourier transforms. The quality of the approximation to $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ is illustrated in Section 3.3 where simulations are done on a sufficiently coarse grid to allow a comparison of both $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ and the fast approximation. In Section 3.4, however, simulations are done on a two dimensional grid which is dense enough to necessitate the fast approximation to $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$.

2.4. An illustration of the bias reduction due to local invariance

In this section we give an example of two stochastic processes with nearly the same values of $\boldsymbol{C}^{(0)}_{\boldsymbol{k}}, \boldsymbol{C}^{(1)}_{\boldsymbol{k}}$ and $\boldsymbol{C}^{(2)}_{\boldsymbol{k}}$, discussed above, but where one process is not locally invariant. A quadratic estimate of nonstationarity is derived for both models and the resulting bias of each is compared. The main conclusion is that a small deviation from the locally invariant structure results in a bias that is orders of magnitude larger than what is found in the local invariant model.

Consider the following two periodic nonstationary stochastic processes[1] on $[-\pi, \pi)$

$$Z(t) := \int e^{itk} e^{i\phi(t)k} \sqrt{C_k} \frac{dB_k}{\sqrt{2\pi}}, \qquad \widetilde{Z}(t) := \int e^{itk} e^{\phi(t)|k|} \sqrt{C_k} \frac{dB_k}{\sqrt{2\pi}} \tag{18}$$

where dB_k is complex Gaussian white noise, C_k is the Matérn spectral density with parameters $\nu = 2, \rho = 0.025, \sigma = 1$ (using parameterization given in equation (33) of [22]), $C_k^{\phi\phi}$ is the Matérn spectral density with parameters $\nu = 3, \rho = 2\pi/10, \sigma = 0.03$. Notice that $Z(t)$ has a local invariant nonstationarity, whereas $\widetilde{Z}(t)$ does not. Indeed the analog to expansion (3) for the two covariance structures is given by

$$\begin{aligned} \operatorname{cov}(Z(t), Z(s)) &= C^{(0)}(t-s) + \big(\phi(t) - \phi(s)\big)\ C^{(1)}(t-s) \\ &\quad + \big(\phi(t) - \phi(s)\big)^2 C^{(2)}(t-s) + \mathcal{O}(\phi^3) \end{aligned} \tag{19}$$

$$\begin{aligned} \operatorname{cov}(\widetilde{Z}(t), \widetilde{Z}(s)) &= C^{(0)}(t-s) + \big(\phi(t) + \phi(s)\big)\ \widetilde{C}^{(1)}(t-s) \\ &\quad + \big(\phi(t) + \phi(s)\big)^2 \widetilde{C}^{(2)}(t-s) + \mathcal{O}(\phi^3) \end{aligned} \tag{20}$$

where $\widetilde{C}_k^{(1)}$ and $\widetilde{C}_k^{(2)}$ are related to the corresponding local invariant terms as follows

$$\widetilde{C}_k^{(1)} := |C_k^{(1)}|, \qquad \widetilde{C}_k^{(2)} := -C_k^{(2)}. \tag{21}$$

[1] We follow our notational convention and use non-bold symbols in this section to indicate scalar quantities for $d = 1$. Moreover, due to the periodic nature of $Z(t)$ and $\tilde{Z}(t)$, our notation dictates that the "integrals" in (18) are actually denoting infinite series so that $\int \equiv \sum_{k\in\mathbb{Z}}$ in this case.

The quadratic estimate $\hat{\phi}_\ell$ based on the observed local invariant process $Z^{obs}(t) = Z(t)$, without observational noise (so that $C_\ell^{NN} \equiv 0$), is defined by (5). To construct a quadratic estimate of ϕ_ℓ based on observations $\widetilde{Z}^{obs}(t) = \widetilde{Z}(t)$ first notice that one can use the expansion (20) to derive the following approximation

$$E\left(\widetilde{Z}_{k+\ell}\widetilde{Z}_{-k}\right) \approx \phi_\ell \left(\widetilde{C}_k^{(1)} + \widetilde{C}_{k+\ell}^{(1)}\right).$$

This is similar to (4) with the exception of one sign change necessary to accommodate the non local invariant structure in $\widetilde{Z}$. The above approximation can now be used to define the following quadratic estimate of ϕ_ℓ, denoted $\widetilde{\phi}_{\boldsymbol{\ell}}$, from observations $\widetilde{Z}^{obs}(t)$

$$\widetilde{\phi}_\ell := \widetilde{A}_\ell \int \left(\widetilde{C}_k^{(1)} + \widetilde{C}_{k+\ell}^{(1)}\right)^* \frac{\widetilde{Z}_{k+\ell}^{obs}\widetilde{Z}_{-k}^{obs}}{\widetilde{C}_{k+\ell}^{ZZobs}\widetilde{C}_k^{ZZobs}} \frac{dk}{\sqrt{2\pi}}$$

where $\widetilde{A}_\ell$ is defined just as in (6) with the exception that the minus sign is switched to a plus sign. Moreover, the approximations given in Sections 2.3.1 and 2.3.2 can be similarly modified—just changing the negative sign in (15) and in the definition of $C_\ell^{\text{var }\hat{\phi}}$—to produce analogous approximations for the variance and bias of $\widetilde{\phi}_\ell$, denoted $C_\ell^{\text{var }\widetilde{\phi}}$ and $C_\ell^{\text{bias }\widetilde{\phi}}$ respectively.

Figure 1 shows the second order bias and variance approximation for the local invariant estimate $\hat{\phi}_\ell$ (**shown at left**) compared to the non local invariant estimate $\widetilde{\phi}_\ell$ (**shown at right**). The left plot shows $\ell^2 C_\ell^{\text{var }\hat{\phi}}$ and $\ell^2 C_\ell^{\text{bias }\hat{\phi}}$ (**solid-green** and **dashed-red**, respectively) whereas the right plot shows $\ell^2 C_\ell^{\text{var }\widetilde{\phi}}$ and $\ell^2 C_\ell^{\text{bias }\widetilde{\phi}}$ (**solid-green** and **dashed-red**, respectively). Both plots use the same axis range and additionally show the signal spectral density $\ell^2 C_\ell^{\phi\phi}$ (**dotted line**) for comparison of the respective signal to noise ratios. Note that all spectral densities shown are multiplied by ℓ^2 to improve the visualization of the high frequency power.

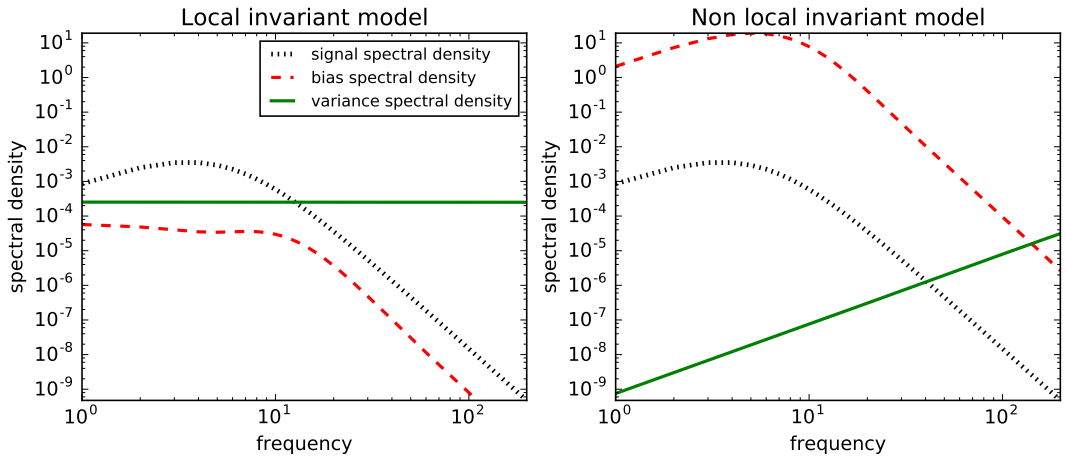

FIG 1. *These plots are intended to give an illustration of the bias attenuation effect of the local invariant covariance structure (the specifics of the models are given in Section 2.4). The plot* ***at left*** *corresponds to the local invariant model and show estimation variance* $\ell^2 C_\ell^{var\ \hat{\phi}}$ *(**solid-green**), bias* $\ell^2 C_\ell^{bias\ \hat{\phi}}$ *(**dashed-red**) and the signal* $\ell^2 C_\ell^{\phi\phi}$ *(**dotted-black**). In contrast the plot* ***at right*** *shows their non local invariant counterparts. Notice that the locally invariant second order bias is orders of magnitude smaller than the corresponding non local invariant bias, even though the functions* $C_k^{(1)}$ *and* $C_k^{(2)}$ *are similar to their non local invariant counterparts* $\widetilde{C}_k^{(1)}$ *and* $\widetilde{C}_k^{(2)}$.

The main conclusion from Figure 1 is that, even though the functions $\widetilde{C}_k^{(1)}$ and $\widetilde{C}_k^{(2)}$ are very similar to their locally invariant counterparts, the local invariant model results in quadratic estimation

bias that is orders of magnitude smaller than the corresponding non local invariant estimation bias. Indeed at most small frequencies ℓ one has

$$C_\ell^{\text{bias } \hat{\phi}} \ll \min(C_\ell^{\phi\phi}, C_\ell^{\text{var } \hat{\phi}})$$
$$C_\ell^{\text{bias } \widetilde{\phi}} \gg \max(C_\ell^{\phi\phi}, C_\ell^{\text{var } \widetilde{\phi}}).$$

Figure 1 also shows that for small ℓ, the signal to noise ratio $C_\ell^{\phi\phi}/C_\ell^{\text{var } \hat{\phi}}$ is large enough to suggest $\hat{\phi}(t)$ will be successful at tracking the large scale features of $\phi(t)$. Moreover, even at large ℓ, where the signal to noise ratio for $\hat{\phi}_\ell$ is small, the fact that the bias is sub-dominant to the signal implies that detection of $C_\ell^{\phi\phi}$ is possible by averaging over a large number of frequencies to attenuate the impact of $C_\ell^{\text{var } \hat{\phi}}$.

3. Nonstationary spectral phase model

In this section we specialize the results of the previous section to random fields which have a particular form: nonstationary spectral phase random fields. These models are locally invariant (c.f. Definition 1) and can be viewed as generalized warping models which are amenable to quadratic estimates. The general formulas for $\hat{\phi}_{\boldsymbol{\ell}}$, $C_{\boldsymbol{\ell}}^{\text{var } \hat{\phi}}$ and $C_{\boldsymbol{\ell}}^{\text{bias } \hat{\phi}}$, given in the previous section, are derived for the nonstationary spectral phase model to yield exact formulas. In subsection 3.1 we present a characterization of the local spectral densities for nonstationary spectral phase models using the theory of optimal transport and the L_2-Wasserstein metric. In subsection 3.2 we present informal guidance for specifying some of the modeling parameters of the nonstationary models presented here. Finally, in sections 3.3 and 3.4, we present a set of simulations which demonstrate the quadratic estimate and the accuracy of our approximation to the mean squared sampling properties.

Definition 3. *A real random field $Z(\boldsymbol{x})$ on $\mathbb{R}^d$ is said to be a **nonstationary spectral phase random field** if there exists functions $C_{\boldsymbol{k}} : \mathbb{R}^d \to \mathbb{R}^+$, $\boldsymbol{\theta}(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}^d$ and $\boldsymbol{\eta}_{\boldsymbol{k}} : \mathbb{R}^d \to \mathbb{R}^d$ such that $C_{\boldsymbol{k}}$ has finite $L_1(\mathbb{R}^d)$ norm, $C_{-\boldsymbol{k}} = C_{\boldsymbol{k}}$, $\boldsymbol{\eta}_{-\boldsymbol{k}} = -\boldsymbol{\eta}_{\boldsymbol{k}}$ and*

$$Z(\boldsymbol{x}) = \int \exp\left(i\boldsymbol{x}\cdot\boldsymbol{k}\right) \exp\left(i\boldsymbol{\theta}(\boldsymbol{x})\cdot\boldsymbol{\eta}_{\boldsymbol{k}}\right) \sqrt{C_{\boldsymbol{k}}} \frac{dW_{\boldsymbol{k}}}{(2\pi)^{d/2}} \tag{22}$$

where $dW_{\boldsymbol{k}}$ denotes a complex Gaussian white noise random measure on $\mathbb{R}^d$ which satisfies $E|dW_{\boldsymbol{k}}|^2 = d\boldsymbol{k}$.

The covariance function $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) := \text{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y}))$ for $Z(\boldsymbol{x})$ defined by (22), conditioning on $\boldsymbol{\eta}_{\boldsymbol{k}}$ and $\boldsymbol{\theta}(\boldsymbol{x})$, can be computed as follows

$$C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = \int \exp\left(i(\boldsymbol{x}-\boldsymbol{y})\cdot\boldsymbol{k}\right) \exp\left(i(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y}))\cdot\boldsymbol{\eta}_{\boldsymbol{k}}\right) C_{\boldsymbol{k}} \frac{d\boldsymbol{k}}{(2\pi)^d}. \tag{23}$$

The fact that $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})$ can be written as a function of $\boldsymbol{x}-\boldsymbol{y}$ and $\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y})$ implies that $Z(\boldsymbol{x})$ satisfies Definition 1 and is therefore locally invariant with respect to the nonstationarity characterized by $\boldsymbol{\theta}(\boldsymbol{x})$. By assuming $\boldsymbol{\eta}_{\boldsymbol{k}}$ is known and $\boldsymbol{\theta}(\boldsymbol{x})$ is characterized by a scalar potential $\phi(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ (see Assumption 2) the results of Section 2.1 can be applied to generate a quadratic estimate $\hat{\phi}_{\boldsymbol{\ell}}$ based on a observing a single realization of $Z(\boldsymbol{x})$ with additive stationary noise. Indeed, by expanding $\exp(i(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y}))\cdot\boldsymbol{\eta}_{\boldsymbol{k}})$ in (23), to second order, one obtains the following expression for the terms in (3) and (14)

$$C_{\boldsymbol{k}}^{(0)} = \frac{C_{\boldsymbol{k}}}{(2\pi)^{d/2}}, \qquad \boldsymbol{C}_{\boldsymbol{k}}^{(1)} = \frac{i\boldsymbol{\eta}_{\boldsymbol{k}} C_{\boldsymbol{k}}}{(2\pi)^{d/2}}, \qquad \boldsymbol{C}_{\boldsymbol{k}}^{(2)} = -\frac{\boldsymbol{\eta}_{\boldsymbol{k}}\boldsymbol{\eta}_{\boldsymbol{k}}^T C_{\boldsymbol{k}}}{2(2\pi)^{d/2}}. \tag{24}$$

The last ingredient needed for computing $\hat{\phi}_{\boldsymbol{\ell}}$, $C_{\boldsymbol{\ell}}^{\text{var }\hat{\phi}}$ and $C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}}$ is an expression for the marginal spectral density of the observations $Z(\boldsymbol{x})+N(\boldsymbol{x})$, denoted $C_{\boldsymbol{k}}^{ZZobs}$ in Section 2.1. Notice that Assumption 3 guarantees that $\boldsymbol{\theta}(\boldsymbol{x})$ is a stationary mean zero Gaussian random field. Therefore

$$E\big(\exp(i(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y}))\cdot\boldsymbol{\eta}_{\boldsymbol{k}})\big)=\exp\big(-\tfrac{1}{2}\boldsymbol{\eta}_{\boldsymbol{k}}^T\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{y})\boldsymbol{\eta}_{\boldsymbol{k}}\big)$$

where $\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{y})$ is the covariance matrix of $\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y})$. This implies that the marginal spectral density of the observations has the form

$$C_{\boldsymbol{k}}^{ZZobs}=(2\pi)^{d/2}C_{\boldsymbol{k}}^{Zm}+C_{\boldsymbol{k}}^{NN} \tag{25}$$

where $C_{\boldsymbol{k}}^{Zm}$ is the Fourier transform of the marginal auto covariance of $Z(\boldsymbol{x})$ and is given by

$$C^{Zm}(\boldsymbol{x}-\boldsymbol{y})=\int\exp(i(\boldsymbol{x}-\boldsymbol{y})\cdot\boldsymbol{k})\exp\big(-\tfrac{1}{2}\boldsymbol{\eta}_{\boldsymbol{k}}^T\boldsymbol{\Sigma}(\boldsymbol{x}-\boldsymbol{y})\boldsymbol{\eta}_{\boldsymbol{k}}\big)\,C_{\boldsymbol{k}}\frac{d\boldsymbol{k}}{(2\pi)^d}.$$

The above expressions for $C_{\boldsymbol{k}}^{ZZobs}$, $\boldsymbol{C}_{\boldsymbol{k}}^{(1)}$ and $\boldsymbol{C}_{\boldsymbol{k}}^{(2)}$ completely define the quadratic estimate $\hat{\phi}_{\boldsymbol{\ell}}$, the second order bias spectral density $C_{\boldsymbol{\ell}}^{\text{bias }\hat{\phi}}$ and the approximation to estimation variance characterized by $C_{\boldsymbol{\ell}}^{\text{var }\hat{\phi}}$ derived in Section 2.1.

3.1. Locally attainable spectral densities

In this section we investigate the set of possible local spectral densities, for different values of the nonstationary function $\boldsymbol{\theta}(\boldsymbol{x})$, under the nonstationary phase random field model of $Z(\boldsymbol{x})$. Since the nonstationarity in $Z(\boldsymbol{x})$ is exclusively due to local variation of a spectral phase, there is significant restriction on how local spectral densities vary with $\boldsymbol{x}$. For example, one can easily see that all the local spectral densities of Z must have the same integral (so that the pointwise variance of $Z(\boldsymbol{x})$ is a constant function of $\boldsymbol{x}$). In what follows we characterize further restrictions and, in doing so, identify a second source of estimation bias due to the curved nature of the local spectral models. The theory of optimal transport and the L_2-Wasserstein metric play an important role in the characterization of local stationary models. We refer the reader to the excellent book [25] for an overview of the subject.

The local invariance property of all nonstationary phase models implies that the local distribution of $Z(\boldsymbol{x})$ is invariant to changes in the magnitude of $\boldsymbol{\theta}(\boldsymbol{x})$. However, the local behavior of $Z(\boldsymbol{x})$ *is* sensitive to the gradient of $\boldsymbol{\theta}(\boldsymbol{x})$. In particular, suppose $\boldsymbol{\theta}(\boldsymbol{x})$ has the form $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ where $\boldsymbol{A}\in\mathbb{R}^{d\times d}$ and $\boldsymbol{b}\in\mathbb{R}^d$. In this case the covariance function $C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})$ is still invariant to changes in $\boldsymbol{b}\in\mathbb{R}^d$ but is sensitive to changes in $\boldsymbol{A}\in\mathbb{R}^{d\times d}$. Indeed assuming $\boldsymbol{\theta}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ one has

$$C^{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})=\int\exp\big(i(\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{k}+\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\big)\,C_{\boldsymbol{k}}\frac{d\boldsymbol{k}}{(2\pi)^d}=\int\exp\big(i(\boldsymbol{x}-\boldsymbol{y})\cdot\boldsymbol{\omega}\big)\frac{d\lambda(\boldsymbol{\omega})}{(2\pi)^d} \tag{26}$$

where λ is the spectral measure obtained by a change of variables $\boldsymbol{\omega}=\boldsymbol{k}+\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}}$ (under appropriate measurability assumptions on $\boldsymbol{\eta}_{\boldsymbol{k}}$) Therefore when $\boldsymbol{\theta}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ in a local neighborhood about $\boldsymbol{x}$ the process $Z(\boldsymbol{x})$ becomes locally stationary with local spectral measure given by λ.

The optimal transport literature uses the notation $\boldsymbol{v}\sharp\lambda(B):=\lambda(\boldsymbol{v}^{-1}(B))$ to denote the push forward of a measure λ on $\mathbb{R}^d$ under a measurable transformation $\boldsymbol{v}_{\boldsymbol{k}}:\mathbb{R}^d\to\mathbb{R}^d$. For our needs it will be useful to extend this definition to spectral measures λ which have a spectral density $C_{\boldsymbol{k}}$ with respect to Lebesgue measure on $\mathbb{R}^d$. In particular we let $\boldsymbol{v}_{\boldsymbol{k}}\sharp C_{\boldsymbol{k}}$ denote the push forward of the measure $C_{\boldsymbol{k}}d\boldsymbol{k}$ under the transformation $\boldsymbol{v}_{\boldsymbol{k}}$. This notation allows one to easily express the spectral measure λ in (26) as

$$\lambda=(\boldsymbol{k}+\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}$$

and, in doing so, creates a succinct notation for the collection of locally attainable spectral measures defined as follows.

Definition 4. *Suppose $Z(\boldsymbol{x})$ is a nonstationary phase random field on $\mathbb{R}^d$ satisfying Definition 3. **The family of locally attainable spectral measures for** $Z(\boldsymbol{x})$ is defined to be the collection of measures*

$$\mathscr{C}^{C,\boldsymbol{\eta}} := \left\{ (\boldsymbol{k} + \boldsymbol{A}^T \boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} : \boldsymbol{A} \in \mathbb{R}^{d \times d} \right\}$$

so that for each $\lambda \in \mathscr{C}^{C,\boldsymbol{\eta}}$ there exists a matrix $\boldsymbol{A}$ such that when $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ the random field $Z(\boldsymbol{x})$ becomes stationary with spectral measure λ.

Notice that, depending on $\boldsymbol{\xi}$, there may a restriction on the possible matrices $\boldsymbol{A}$ which satisfy $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. This will further limit the set of attainable local spectral densities but is not included in the definition of $\mathscr{C}^{C,\boldsymbol{\eta}}$. The role of $\boldsymbol{\xi}$, in terms of modeling $Z(\boldsymbol{x})$, is discussed in Section 3.2 below.

The following claim shows that given any two spectral densities $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$, with the same integral and finite second moments, there exists a nonstationary phase random field $Z(\boldsymbol{x})$ which has both $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ as locally attainable spectral densities (i.e. $C_{\boldsymbol{k}}, \tilde{C}_{\boldsymbol{k}} \in \mathscr{C}^{C,\boldsymbol{\eta}}$). Moreover, each measure contained in the L_2-Wasserstein geodesic connecting $C_{\boldsymbol{k}}d\boldsymbol{k}$ to $\tilde{C}_{\boldsymbol{k}}d\boldsymbol{k}$ is also locally attainable by $Z(\boldsymbol{x})$.

Claim 1 (Any pair of spectral densities with the same integral are attainable). *Let $d \geq 1$ be an integer, $t_0 > 0$ be a real number and $C_{\boldsymbol{k}}$, $\tilde{C}_{\boldsymbol{k}}$ be two spectral densities on $\mathbb{R}^d$ with finite second moments such that $\sigma^2 = \int_{\mathbb{R}^d} C_{\boldsymbol{k}} d\boldsymbol{k} = \int_{\mathbb{R}^d} \tilde{C}_{\boldsymbol{k}} d\boldsymbol{k}$. Then there exists a vector field $\boldsymbol{\eta}_{\boldsymbol{k}} \colon \mathbb{R}^d \to \mathbb{R}^d$ which is $L_2(\mathbb{R}^d)$ integrable with respect to $C_{\boldsymbol{k}}d\boldsymbol{k}$ and generates a one dimensional curve of spectral measures $\{\lambda^{(t)} : t \in [0, t_0]\}$, defined by*

$$\lambda^{(t)} := (\boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} \tag{27}$$

with endpoints $\lambda^{(0)} = C_{\boldsymbol{k}}d\boldsymbol{k}$ and $\lambda^{(t_0)} = \tilde{C}_{\boldsymbol{k}}d\boldsymbol{k}$, such that $\{\lambda^{(t)} : t \in [0, t_0]\} \subset \mathscr{C}^{C,\boldsymbol{\eta}}$. In particular there exists a nonstationary phase random field model for which $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ are both locally attainable. Moreover, $\{\lambda^{(t)} : t \in [0, t_0]\}$ is a L_2-Wasserstein geodesic path within the class of absolutely continuous spectral measures (with total mass σ^2 and finite second moments) and each measure $\lambda^{(t)}$ has a density $C^{(t)}_{\boldsymbol{k}}$ with respect to Lebesgue measure on $\mathbb{R}^d$ which (weakly) satisfies

$$\partial_t C^{(t)}_{\boldsymbol{k}} + \operatorname{div}(\boldsymbol{\eta}^{(t)}_{\boldsymbol{k}} C^{(t)}_{\boldsymbol{k}}) = 0 \tag{28}$$

for all $t \in [0, t_0]$ where $\boldsymbol{\eta}^{(t)}_{\boldsymbol{k}} := \boldsymbol{\eta}_{T_t^{-1}(\boldsymbol{k})}$ and $T_t(\boldsymbol{k}) := \boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}}$.

Proof. By standard optimal transport theory (see [25] for example) the assumptions on $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ guarantee the existence of a convex function $\psi_{\boldsymbol{k}} : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\psi_{\boldsymbol{k}}$ is the optimal transport from $C_{\boldsymbol{k}}d\boldsymbol{k}$ to $\tilde{C}_{\boldsymbol{k}}d\boldsymbol{k}$. Let $\boldsymbol{\eta}_{\boldsymbol{k}} := \frac{1}{t_0}(\nabla\psi_{\boldsymbol{k}} - \boldsymbol{k})$ so that

$$\lambda^{(t)} := (\boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} = \left((1 - \tfrac{t}{t_0})\boldsymbol{k} + \tfrac{t}{t_0}\nabla\psi_{\boldsymbol{k}}\right) \sharp C_{\boldsymbol{k}}. \tag{29}$$

The particular form of the right hand side of (29) implies each measure $\lambda^{(t)}$ has a density $C^{(t)}_{\boldsymbol{k}}$ with respect to Lebesgue measure and the path of measures $\{\lambda^{(t)} \colon t \in [0, t_0]\}$ forms an L_2-Wasserstein geodesic with endpoints $C_{\boldsymbol{k}}d\boldsymbol{k}$ and $\tilde{C}_{\boldsymbol{k}}d\boldsymbol{k}$ at $t = 0$ and $t = t_0$ respectively (see Proposition 5.9 in [25]). Moreover, $C^{(t)}_{\boldsymbol{k}}$ weakly satisfies (28) by Theorem 5.34 of [25] and the fact that $C^{(t)}_{\boldsymbol{k}} = T_t \sharp C^{(0)}_{\boldsymbol{k}}$ where $\partial_t T_t(\boldsymbol{k}) = \boldsymbol{\eta}_{\boldsymbol{k}}$. By setting $\boldsymbol{A} = t\boldsymbol{I}_d$ in (26) one has

$$\int \exp\left(i(\boldsymbol{x} - \boldsymbol{y})\cdot(\boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}})\right) C_{\boldsymbol{k}} \frac{d\boldsymbol{k}}{(2\pi)^d} = \int \exp\left(i(\boldsymbol{x} - \boldsymbol{y})\cdot\boldsymbol{\omega}\right) C^{(t)}_{\boldsymbol{\omega}} \frac{d\boldsymbol{\omega}}{(2\pi)^d}$$

which implies that for each $t \in [0, t_0]$ the measure $\lambda^{(t)}$ is a locally attainable spectral measure. □

Isotropic spectral densities are an important special case for many statistical applications. The following claim allows considerable simplification for the construction of the vector field $\boldsymbol{\eta}_{\boldsymbol{k}}$ guaranteed by Claim 1.

Claim 2 (Special case for isotropic spectral densities). *Let $d \geq 1$ be an integer, $t_0 > 0$ be a real number and $C_{|\boldsymbol{k}|}$, $\tilde{C}_{|\boldsymbol{k}|}$ be two isotropic spectral densities on $\mathbb{R}^d$ with finite second moments and total mass σ^2. Define*

$$\boldsymbol{\eta}_{\boldsymbol{k}} := \frac{1}{t_0}\Big(\tilde{F}^{-1} \circ F(|\boldsymbol{k}|)\frac{\boldsymbol{k}}{|\boldsymbol{k}|} - \boldsymbol{k}\Big) \tag{30}$$

where $F(r) := \frac{2\pi^{d/2}}{\sigma^2\Gamma(d/2)}\int_0^r \xi^{d-1}C_\xi d\xi$ and $\tilde{F}(r) := \frac{2\pi^{d/2}}{\sigma^2\Gamma(d/2)}\int_0^r \xi^{d-1}\tilde{C}_\xi d\xi$. Then for all $t \in [0, t_0]$, $\boldsymbol{\eta}_{\boldsymbol{k}}$ generates the spectral measures $\lambda^{(t)} := (\boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{|\boldsymbol{k}|}$ defined in Claim 1. In particular, $\{\lambda^{(t)} : t \in [0, t_0]\}$ forms a L_2-Wasserstein geodesic path of locally attainable spectral densities in $\mathscr{C}^{C,\boldsymbol{\eta}}$, with endpoints $\lambda^{(0)} = C_{|\boldsymbol{k}|}d\boldsymbol{k}$ and $\lambda^{(t_0)} = \tilde{C}_{|\boldsymbol{k}|}d\boldsymbol{k}$, where $\boldsymbol{k} + t\boldsymbol{\eta}_{\boldsymbol{k}}$ is the optimal transport from $\lambda^{(0)}$ to $\lambda^{(t)}$.

Proof. By the proof of Claim 1 it will be sufficient to show that $\tilde{F}^{-1}\circ F(|\boldsymbol{k}|)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$ is the optimal transport from $C_{|\boldsymbol{k}|}$ to $\tilde{C}_{|\boldsymbol{k}|}$. Let $\boldsymbol{K}$ and $\tilde{\boldsymbol{K}}$ be random vectors in $\mathbb{R}^d$ with densities $C_{|\boldsymbol{k}|}/\sigma^2$ and $\tilde{C}_{|\boldsymbol{k}|}/\sigma^2$, respectively. By the distributional rotational symmetry of $\boldsymbol{K}$ there exists a convex $\psi(r) : \mathbb{R}^+ \to \mathbb{R}$ such that $\nabla(\psi(|\boldsymbol{k}|)) = \psi'(|\boldsymbol{k}|)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$ is the optimal transport from $\mathscr{L}\boldsymbol{K}$ to $\mathscr{L}\tilde{\boldsymbol{K}}$ (where $\mathscr{L}\boldsymbol{K}$ and $\mathscr{L}\tilde{\boldsymbol{K}}$ denotes the law, i.e. probability distribution, of $\boldsymbol{K}$). Also notice that $\psi'(r)$ is the optimal transport from $\mathscr{L}|\boldsymbol{K}|$ to $\mathscr{L}|\tilde{\boldsymbol{K}}|$ since $\psi(r)$ is convex and

$$P(\psi'(|\boldsymbol{K}|) \leq r) = P\big(|\psi'(|\boldsymbol{K}|)\tfrac{\boldsymbol{K}}{|\boldsymbol{K}|}| \leq r\big) = P(|\tilde{\boldsymbol{K}}| \leq r).$$

The optimal transport between two univariate random variables (see [25]) is given by the composition of the quantile function (of the target measure) and the cumulative distribution function (of the base measure). Therefore the optimal transport from $\mathscr{L}|\boldsymbol{K}|$ to $\mathscr{L}|\tilde{\boldsymbol{K}}|$ is given by $\tilde{F}^{-1} \circ F(r)$ where $F(r) = \frac{2\pi^{d/2}}{\sigma^2\Gamma(d/2)}\int_0^r \xi^{d-1}C_\xi d\xi$ and $\tilde{F}(r) = \frac{2\pi^{d/2}}{\sigma^2\Gamma(d/2)}\int_0^r \xi^{d-1}\tilde{C}_\xi d\xi$ are the cumulative distribution functions of $|\boldsymbol{K}|$ and $|\tilde{\boldsymbol{K}}|$, respectively. By the uniqueness of optimal transports one has $\psi'(r) = \tilde{F}^{-1}\circ F(r)$ and therefore

$$\nabla(\psi(|\boldsymbol{k}|)) = \psi'(|\boldsymbol{k}|)\frac{\boldsymbol{k}}{|\boldsymbol{k}|} = \tilde{F}^{-1} \circ F(|\boldsymbol{k}|)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$$

is the optimal transport from $\mathcal{L}\boldsymbol{K}$ to $\mathcal{L}\tilde{\boldsymbol{K}}$, as was to be shown. □

As a corollary to the above theorem one can obtain quasi-closed form solutions for $\boldsymbol{\eta}_{\boldsymbol{k}}$ when the spectral densities $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ are both Matérn spectral densities with the same integral and with finite second moments. The form of $\boldsymbol{\eta}_{\boldsymbol{k}}$, in this case, can be computed using the incomplete beta function and the quantiles of beta random variables (which is not technically given in closed form but for which simple Newton methods are guaranteed to converge, see [8]).

Corollary 1 (Optimal transports between Matérn spectral densities). *Let $d \geq 1$ be an integer and $t_0, \nu, \tilde{\nu}, \rho, \tilde{\rho}, \sigma^2 > 0$ be real numbers such that $\nu, \tilde{\nu} > 1$. If*

$$C_{|\boldsymbol{k}|} = \sigma^2 2^d \pi^{d/2}\frac{\Gamma(\nu + d/2)}{\Gamma(\nu)}\left(\frac{4\nu}{\rho^2}\right)^{\nu}\left(\frac{4\nu}{\rho^2} + |\boldsymbol{k}|^2\right)^{-\nu-d/2} \tag{31}$$

$$\tilde{C}_{|\boldsymbol{k}|} = \sigma^2 2^d \pi^{d/2}\frac{\Gamma(\tilde{\nu} + d/2)}{\Gamma(\tilde{\nu})}\left(\frac{4\tilde{\nu}}{\tilde{\rho}^2}\right)^{\tilde{\nu}}\left(\frac{4\tilde{\nu}}{\tilde{\rho}^2} + |\boldsymbol{k}|^2\right)^{-\tilde{\nu}-d/2} \tag{32}$$

then F and $\tilde{F}^{-1}$, defined in Claim 2, are given by

$$F(r) = I_{r^2/(4\nu/\rho^2 + r^2)}(d/2, \nu) \tag{33}$$

$$\tilde{F}^{-1}(u) = \left(\frac{4\tilde{\nu}}{\tilde{\rho}^2}\right)^{1/2}\left(\frac{1}{Q_u(d/2, \tilde{\nu})} - 1\right)^{-1/2} \tag{34}$$

where $I_x(p, q)$ is the incomplete beta function and $Q_u(p, q)$ is the quantile function for a univariate Beta(p, q) random variable evaluated at $u \in (0, 1)$.

Proof. First notice that $(2\pi)^d\sigma^2 = \int_{\mathbb{R}^d} C_{|\boldsymbol{k}|}d\boldsymbol{k} = \int_{\mathbb{R}^d} \tilde{C}_{|\boldsymbol{k}|}d\boldsymbol{k}$ and the constraints $\nu, \tilde{\nu} > 1$ are sufficient to ensure $C_{|\boldsymbol{k}|}$ and $\tilde{C}_{|\boldsymbol{k}|}$ have finite second moments. Therefore Claim 2 applies. For any $\nu, a > 0$, the change of variables $x = y^2/(a+y^2)$ gives

$$\int_0^r \frac{y^{d-1}}{(a+y^2)^{\nu+d/2}}dy = \frac{1}{2a^\nu}\int_0^{\frac{r^2}{a+r^2}} (1-x)^{\nu-1}x^{d/2-1}dx = \frac{B(d/2,\nu)}{2a^\nu} I_{r^2/(a+r^2)}(d/2,\nu).$$

Therefore $F(r) := \frac{2\pi^{d/2}}{\sigma^2(2\pi)^d\Gamma(d/2)}\int_0^r \xi^{d-1}C_\xi d\xi = I_{r^2/(4\nu/\rho^2+r^2)}(d/2,\nu)$ and similarly for $\tilde{F}(r)$. This immediately gives (33) an (34). □

The Wasserstein structure of the locally attainable spectral models gives a convenient geometric picture for potential difficulties when estimating local spectra in the nonstationary phase model. For example, the space of probability distributions (with finite second moments) has positive curvature (in the sense of Aleksandrov's notion of metric curvature) under the L_2-Wasserstein metric [1]. A less precise mathematical illustration of this is the fact is that the two locally attainable spectra, $(\boldsymbol{k} + \boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}$ and $(\boldsymbol{k} - \boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}$, become asymptotically indistinguishable as the entries $\boldsymbol{A}$ become arbitrarily large. Indeed, the probability distributions of $\boldsymbol{K} + \boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{K}}$ and $\boldsymbol{K} - \boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{K}}$ are similar when $\boldsymbol{K}$ is a random vector with unnormalized density $C_{\boldsymbol{k}}$ and the magnitude of the entries of $\boldsymbol{A}$ are large (since $\boldsymbol{\eta}_{\boldsymbol{k}}$ has odd symmetry and $C_{\boldsymbol{k}}$ has even symmetric). One implication of this asymptotic non-identifiability is that estimates of the local spectra can break down when $\boldsymbol{\theta}(\boldsymbol{x})$ has large local linear fluctuations (i.e. when the entries of $\boldsymbol{A}$ are large) so that the two local models $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and $\boldsymbol{\theta}(\boldsymbol{x}) = -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ are difficult to distinguish from the data.

A possible numerical tool for probing this breakdown is to analyze the so called ***cut locus*** associated with L_2-Wasserstein geodesics. Informally, the cut locus quantifies the maximal extent of geodesic paths emanating from some point. The cut locus of the geodesics emanating from $C_{\boldsymbol{k}}d\boldsymbol{k}$ can serve as a proxy for when curvature effects will corrupt local estimation of $\boldsymbol{\theta}(\boldsymbol{x})$ due to the similarity between the two local spectra characterized by the local linear models $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and $\boldsymbol{\theta}(\boldsymbol{x}) = -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$. If the cut locus starting at $C_{\boldsymbol{k}}d\boldsymbol{k}$ is far from $C_{\boldsymbol{k}}d\boldsymbol{k}$ this implies the geodesic paths are long and the entries of $\boldsymbol{A}$ need to be much larger for curvature difficulties to arise. In particular, fix $\boldsymbol{A}$ and consider the locally attainable models $(\boldsymbol{k} + c\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}$ indexed by $c \in \mathbb{R}$. If there exists a maximal $c_0 > 0$ such that $\{(\boldsymbol{k}+c\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}} \colon c \in [0, c_0]\}$ and $\{(\boldsymbol{k}+c\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}} \colon c \in [-c_0, 0]\}$ are both L_2-Wasserstein geodesic, then nonstationary local linear models of the form $\theta(\boldsymbol{x}) = \pm c\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ are less exposed to curvature effects when $|c| \ll c_0$. Claim 3, below, allows one to numerically compute the maximal such cutoff c_0 which characterizes, what we call, the *symmetric two sided cut locus*.

Claim 3. *Let $d \geq 1$ be an integer, $c_0 > 0$ be a real number, $\boldsymbol{A} \in \mathbb{R}^d$, $C_{\boldsymbol{k}}$ is a spectral density on $\mathbb{R}^d$ with finite second moments and $\boldsymbol{\eta}_{\boldsymbol{k}} \colon \mathbb{R}^d \to \mathbb{R}^d$ which is $L_2(\mathbb{R}^d)$ integrable with respect to $C_{\boldsymbol{k}}d\boldsymbol{k}$. Suppose both $\boldsymbol{k} - c_0\boldsymbol{A}\boldsymbol{\eta}_{\boldsymbol{k}}$ and $\boldsymbol{k} + c_0\boldsymbol{A}\boldsymbol{\eta}_{\boldsymbol{k}}$ are C^1 diffeomorphisms which are gradients of convex functions. Then both $\{(\boldsymbol{k} - c\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}} \colon c \in [0, c_0]\}$ and $\{(\boldsymbol{k} + c\boldsymbol{A}^T\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}} \colon c \in [0, c_0]\}$ are paths of absolutely continuous measures which are also L_2-Wasserstein geodesics.*

Proof. Let $\phi_{\boldsymbol{k}}$ and $\psi_{\boldsymbol{k}}$ be convex functions defined on $\mathbb{R}^d$ such that $\nabla\phi_{\boldsymbol{k}} = \boldsymbol{k} - c_0\boldsymbol{\eta}_{\boldsymbol{k}}$ and $\nabla\psi_{\boldsymbol{k}} = \boldsymbol{k} + c_0\boldsymbol{\eta}_{\boldsymbol{k}}$. By the diffeomorphic assumption on $\boldsymbol{k} - c_0\boldsymbol{\eta}_{\boldsymbol{k}}$ and $\boldsymbol{k} + c_0\boldsymbol{\eta}_{\boldsymbol{k}}$ there exists two spectral densities $C_{\boldsymbol{k}}^{(-c_0)}$ and $C_{\boldsymbol{k}}^{(c_0)}$ which satisfy

$$C_{\boldsymbol{k}}^{(-c_0)}d\boldsymbol{k} = (\boldsymbol{k} - c_0\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}$$
$$C_{\boldsymbol{k}}^{(c_0)}d\boldsymbol{k} = (\boldsymbol{k} + c_0\boldsymbol{\eta}_{\boldsymbol{k}})\sharp C_{\boldsymbol{k}}.$$

Now for any $c \in [0, c_0]$ one has

$$\boldsymbol{k} - c\boldsymbol{\eta}_{\boldsymbol{k}} = \big(1 - \tfrac{c}{c_0}\big)\boldsymbol{k} + \tfrac{c}{c_0}\nabla\phi_{\boldsymbol{k}} \tag{35}$$
$$\boldsymbol{k} + c\boldsymbol{\eta}_{\boldsymbol{k}} = \big(1 - \tfrac{c}{c_0}\big)\boldsymbol{k} + \tfrac{c}{c_0}\nabla\psi_{\boldsymbol{k}}. \tag{36}$$

The explicit form of the right hand side of (35) and (36) imply $\{(\boldsymbol{k} - c\boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} : c \in [0, c_0]\}$ and $\{(\boldsymbol{k} + c\boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} : c \in [0, c_0]\}$ are L_2-Wasserstein geodesics of absolutely continuous measures, connecting $C_{\boldsymbol{k}}^{(-c_0)} d\boldsymbol{k}$ to $C_{\boldsymbol{k}} d\boldsymbol{k}$ and $C_{\boldsymbol{k}} d\boldsymbol{k}$ to $C_{\boldsymbol{k}}^{(c_0)} d\boldsymbol{k}$ respectively (by Proposition 5.9 in [25]). □

3.2. Modeling $\boldsymbol{\xi}_{\boldsymbol{k}}, C_{\boldsymbol{k}}$ and $\boldsymbol{\eta}_{\boldsymbol{k}}$

In this section we give some informal guidance for specifying $\boldsymbol{\xi}_{\boldsymbol{k}}, C_{\boldsymbol{k}}$ and $\boldsymbol{\eta}_{\boldsymbol{k}}$ in the nonstationary spectral phase model (see Definition 3). Since the quadratic estimate is adept at detecting small departures from stationarity—partly due to the accuracy of the variance calculations and the speed at which the quadratic estimate can be simulated under a null stationary model—we focus on the situation where the statistician wants to estimate potential nonstationary extensions of a base stationary model. Further details of this approach can be found in the simulation sections 3.3 and 3.4.

Modeling $C_{\boldsymbol{k}}$ and $\boldsymbol{\eta}_{\boldsymbol{k}}$. In sections 3.3 and 3.4 we model $\boldsymbol{\eta}_{\boldsymbol{k}}$ implicitly by specifying two spectral densities $C_{\boldsymbol{k}}, \tilde{C}_{\boldsymbol{k}}$ and require that they both be locally attainable in the nonstationary random field model (note: $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ must have the same $L_1(\mathbb{R}^d)$ integral). In other words we construct a vector field $\boldsymbol{\eta}_{\boldsymbol{k}}$ from $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ by requiring $C_{\boldsymbol{k}}, \tilde{C}_{\boldsymbol{k}} \in \mathscr{C}^{C,\boldsymbol{\eta}}$. The results of Section 3.1 show this is possible by setting

$$\boldsymbol{\eta}_{\boldsymbol{k}} := \frac{1}{t_0}(\nabla \psi_{\boldsymbol{k}} - \boldsymbol{k}) \tag{37}$$

where $\nabla \psi_{\boldsymbol{k}}$ is the optimal transport from $C_{\boldsymbol{k}}$ to $\tilde{C}_{\boldsymbol{k}}$ and $t_0 > 0$ is determined by the desired physical units of $\boldsymbol{\theta}(\boldsymbol{x})$ or is set to balance the bias and variance of the quadratic estimate (more on this in the next paragraph). Corollary 1 seems particularity useful for this approach in that $C_{\boldsymbol{k}}$ can be determined by an overall Matérn fit and $\tilde{C}_{\boldsymbol{k}}$ can be defined by perturbing the Matérn parameters in a direction of interest. For example, consider the case where one is interested in detecting nonstationarity arising from spatial variation in the Matérn smoothness parameter ν. Using the notation given in Corollary 1 one could estimate σ^2, ρ and ν (the parameters of $C_{\boldsymbol{k}}$) by an overall stationary fit, then define $\tilde{\nu} := \nu + \epsilon$ and $\tilde{\rho} := \rho$ (the parameters of $\tilde{C}_{\boldsymbol{k}}$) for some $\epsilon \in \mathbb{R}$.

Generally larger values of t_0 or smaller values of ϵ will increase estimation variance and decrease estimation bias. We do not yet have a coherent story for the precise nature the dependence of bias and variance as a function of t_0 and ϵ. However, the cut locus of the L_2-Wasserstein geodesics emanating from $C_{\boldsymbol{k}}$ (discussed at the end of Section 3.1) can be a useful tool for probing this dependence. For example, given t_0 and ϵ one can compute the maximal c_0 which satisfies the antecedent of Claim 3. This maximal c_0 can be thought of a type of *symmetric two-sided cut locus* and loosely serves to characterize an upper bound on the magnitude of the entries of $\boldsymbol{A}$, beyond which bias is likely to dominate. This will be explored in more detail in Section 3.3 as a diagnostic tool for determining values of t_0 and ϵ that result in large quadratic estimation bias.

Modeling $\boldsymbol{\xi}_{\boldsymbol{k}}$. In some cases the spectral multiplier $\boldsymbol{\xi}_{\boldsymbol{k}}$ will be constrained by the physics of a particular application. An example of such a constraint is that $\boldsymbol{\theta}(\boldsymbol{x})$ be required to be curl free or divergence free. Indeed, a curl free constraint is active in the gravitational lensing problem and is enforced by setting $\boldsymbol{\xi}_{\boldsymbol{k}} = i\boldsymbol{k}$. In the absence of such constraints one can potentially use $\boldsymbol{\xi}_{\boldsymbol{k}}$ to restrict the possible matrices $\boldsymbol{A}$ which parameterize the locally attainable spectral models $\mathscr{C}^{C,\boldsymbol{\eta}} = \{(\boldsymbol{k} + \boldsymbol{A}^T \boldsymbol{\eta}_{\boldsymbol{k}}) \sharp C_{\boldsymbol{k}} : \boldsymbol{A} \in \mathbb{R}^{d \times d}\}$. If $\xi_{\boldsymbol{k}} := i\boldsymbol{k}$, for example, then $\boldsymbol{A}$ must be of the form $(\partial_{\boldsymbol{x}_p} \partial_{\boldsymbol{x}_q} \phi(\boldsymbol{x}))_{p,q=1}^d = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ where $\boldsymbol{U}$ is a rotation matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with real entries.

3.3. Nonstationary phase example $d = 1$

In this section we present a simulation example to illustrate the quadratic estimate of $\boldsymbol{\theta}(\boldsymbol{x})$, or equivalently the potential $\phi(\boldsymbol{x})$, when observing a single realization of a nonstationary phase random field

$Z(\boldsymbol{x})$ in dimension $d = 1$ (c.f. Definition 3). According to our notational conventions, vector quantities such as $\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{\xi_k}, \boldsymbol{\eta_k}, \boldsymbol{\theta}(\boldsymbol{x})$ are replaced with non-bold scalar notation $x, k, \xi_k, \eta_k, \theta(x)$ to indicate scalar quantities for $d = 1$. There are multiple points we hope to convey with this example. The first is that the quadratic estimate $\hat{\phi}_\ell$, constructed to detect a spatially varying smoothness parameter, is fast and accurate. The second point is that $C_\ell^{\text{var}\,\hat{\phi}}$ and $C_\ell^{\text{bias}\,\hat{\phi}}$ accurately quantify the empirical variance and bias of $\hat{\phi}_\ell$. A third point is that the fast approximation to $C_\ell^{\text{bias}\,\hat{\phi}}$, discussed in the last paragraph of Section 2.3.2, is accurate over a wide range of wave numbers. Finally we illustrate qualitative features of the estimation bias which results when $\theta'(x)$ is large enough to exceed the symmetric two-sided cut locus discussed in Section 3.2 (c.f. Claim 3).

Following the modeling approach outlined in Section 3.2 we first define C_k to be the Matérn spectral density given in (31) with parameters $(\nu, \rho, \sigma^2) := (2, 0.05, 1)$. Now η_k is defined implicitly by specifying a second spectral density $\tilde{C}_k$ that is required to be locally attainable within the same nonstationary phase model for $Z(x)$. $\tilde{C}_k$ is defined to be the Matérn spectral density given in (32) with parameters $(\tilde{\nu}, \tilde{\rho}, \sigma^2) := (2.1, 0.05, 1)$. The variance parameter σ^2 is the same for both C_k and $\tilde{C}_k$ as per the necessary requirement for locally attainable spectral densities. Notice that the only difference between the two Matérn models is the fractional smoothness parameter which is set to model nonstationarity in the local smoothness in $Z(x)$. The parameter t_0 used in (37) to determine η_k is set to 1.5 for Figure 2 and $1.5/7$ for Figure 3.

A single ground truth potential $\phi(x)$ is used throughout this section and was simulated from a mean zero stationary Gaussian process with Matérn parameters $(\nu, \rho, \sigma^2) := (5, 1.5, 15^2/(2\pi)^4)$. The spectral multiplier ξ_k is set to ik so that $\theta(x) = \phi'(x)$. The derivative $\theta'(x)$, in particular $\phi''(x)$, is shown in blue in the middle plot of figures 2 and 3. The quantities $C_\ell^{(0)}, C_\ell^{(1)}, C_\ell^{(2)}$ and C_ℓ^{ZZobs} used to generate $\hat{\phi}_\ell$, $C_\ell^{\text{var}\,\hat{\phi}}$ and $C_\ell^{\text{bias}\,\hat{\phi}}$ are determined by (24) and (25). Finally, to avoid potential aliasing issues in the simulated data $Z^{obs}(x)$, the quadratic estimate is set to ignore 10% of Fourier coefficients which are nearest the Nyquist limit by truncating the weights corresponding to those frequency pairs.

The process $Z(x)$ in this section is defined on $[-5, 5)$ with periodic boundary conditions. The observed process $Z^{obs}(x)$ is simulated without additive noise on 10^4 evenly spaced observation locations in $[-5, 5)$. A simple discrete Riemann sum approximation, at each observed x, was used to approximate to the integral (22). In general, this type of approximation will result in aliasing errors. Generating a distributionally exact simulation of $Z(x)$, without any approximation, appears to be an open problem. It is not yet clear what impact the aliasing errors, present in our simulation, have on the quadratic estimate. However, we found little empirical difference in the performance of the quadratic estimate when reducing the aliasing errors by increasing the frequency upper limit used for the Riemann sum approximation.

Figure 2 shows the results of our simulation when t_0 is set to 1.5. The **top plot** shows a simulation of the nonstationary phase process $Z(x)$. The **blue line** in the **middle plot** shows $\theta'(x)$ along with 5 realizations of the quadratic estimate $\hat{\theta}'(x)$, shown in **grey**, each one applied to an independent realization of $Z(x)$ with the same $\phi(x)$. The **dashed line** in the **middle plot** shows an empirical estimate to $E(\hat{\theta}'(x)|\phi)$ based on averaging the quadratic estimate applied to 100 independent realizations $Z(x)$ all simulated with the same nonstationary potential $\phi(x)$. On average, computing these 100 quadratic estimates (each based on 10^4 observations) took 0.008 seconds on a 2013 MacBook Pro with a 2.3 GHz Intel Core i7 CPU. This illustrates that the quadratic estimate can be computed extremely fast on a dense set of observations. Notice also the estimate is accurate with respect to both variance and bias. Indeed, by comparing signal spectral density $\ell^2 C_\ell^{\phi\phi}$ (**dotted black line** in **both bottom plots**) with $\ell^2 C_\ell^{\text{var}\,\hat{\phi}}$ and $\ell^2 C_\ell^{\text{bias}\,\hat{\phi}}$ (**green and red lines** respectively) one can see that the signal-to-noise ratio for estimation accuracy per-frequency is significantly greater than 1 for a large range of wavenumbers. The **bottom two plots** in Figure 2 show the accuracy of the analytic approximations $\ell^2 C_\ell^{\text{var}\,\hat{\phi}}$ and $\ell^2 C_\ell^{\text{bias}\,\hat{\phi}}$ for quantifying the empirical variance and bias (**green and red dots** respectively) computed

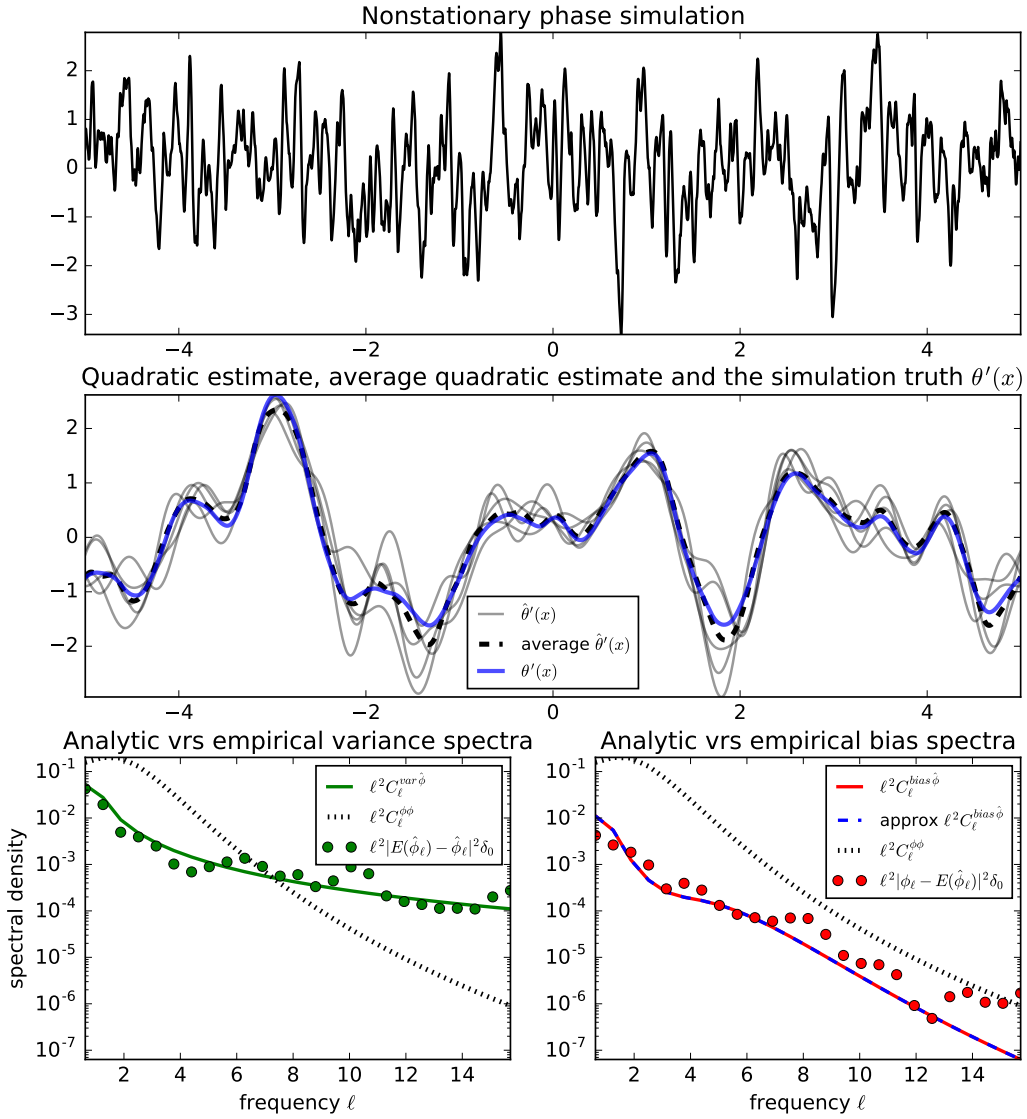

FIG 2. *This figure shows a simulation of a nonstationary phase random field $Z(x)$ (**top**), the quadratic estimate of $\theta'(x)$ (**middle**) and the spectral characterizations of estimation variance (**bottom-left**) and estimation bias (**bottom-right**). The details of the simulation are given in Section 3.3. In the **middle plot** the **blue line** shows the value of $\theta'(x) = \phi''(x)$ which characterizes the nonstationarity in $Z(x)$ through Definition 3, the **grey lines** show different quadratic estimates $\hat{\theta}'(x)$ each one applied to an independent realizations of $Z(x)$ with the same $\theta(x)$ and the **dashed line** shows the empirical average of $\hat{\theta}'(x)$ over 100 such realizations. In the **bottom-left plot** The main purpose of this simulation is intended to illustrate the accuracy of the quadratic estimate and the ability of $C_{\boldsymbol{\ell}}^{var\,\hat{\phi}}$ and $C_{\boldsymbol{\ell}}^{bias\,\hat{\phi}}$ to approximate the empirical variance and bias of the estimate. A secondary goal of this figure is to also show that the fast approximation to $C_{\boldsymbol{\ell}}^{bias\,\hat{\phi}}$ is very accurate over a wide range of small wave numbers.*

from the 100 realizations of $\hat{\theta}'(x)$. The computation of $C_\ell^{\text{var}\,\hat{\phi}}$ took 0.098 seconds. The fast approximation to $\ell^2 C_\ell^{\text{bias}\,\hat{\phi}}$ is plotted with the **dashed blue line** in the **bottom right plot**. This approximation can be seen to be very accurate, nearly indistinguishable from the **red line**, and took only 0.308 seconds to compute (compared to 99.44 seconds for computing $C_\ell^{\text{bias}\,\hat{\phi}}$ directly).

The only difference between Figure 2 and Figure 3 is that t_0 is reduced from 1.5 to $1.5/7$. This has the effect of shrinking the two-sided cut locus (discussed in Section 3.2). This is equivalent to scaling $\phi(x)$ by a factor of 7 which has the effect of dramatically increasing the bias in the quadratic estimate. Indeed, the main point of Figure 3 is to illustrate the qualitative features of the quadratic estimate bias when $\phi(x)$ is too large for the linear approximation in (3) to hold. Indeed, when the magnitude of the true $\theta'(x)$ exceeds the two-sided cut locus, the estimate $\hat{\theta}'(x)$ transitions from a low bias estimate to a bias dominated one as seen in the **middle plot**. This is presumably due to the ability of the two-sided cut locus to identify when the geodesic path of local spectral densities begins to curl in on itself, creating an ill-posed inversion from observed local spectral density to estimate $\phi(x)$.

3.4. Nonstationary phase example $d = 2$

In this section we perform a simulation example to illustrate the quadratic estimate applied to a nonstationary phase random field $Z(\boldsymbol{x})$ in dimension $d = 2$. Besides the increase of dimension, there are two main differences in this simulation as compared to the simulation given in Section 3.3. The first difference is that the spectral multiplier $\boldsymbol{\xi}_{\boldsymbol{k}}$ is set to $(i\boldsymbol{k}_2, -i\boldsymbol{k}_1)^T$ where $\boldsymbol{k} = (\boldsymbol{k}_1, \boldsymbol{k}_2)$. Therefore $\boldsymbol{\theta}(\boldsymbol{x}) = (\partial_{\boldsymbol{x}_2}\phi(\boldsymbol{x}), -\partial_{\boldsymbol{x}_1}\phi(\boldsymbol{x}))^T$ is a divergent free vector field. The second main difference is that the two spectral densities $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$, defined by (31) and (32), have different Matérn parameter values as those used in Section 3.3. The parameter values for $C_{\boldsymbol{k}}$ are given by $(\nu, \rho, \sigma^2) := (1.5, 0.015, 1)$ and the parameter values for $\tilde{C}_{\boldsymbol{k}}$ are given by $(\tilde{\nu}, \tilde{\rho}, \sigma^2) := (1.7, 0.014, 1)$. Recall that $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ are used to generate $\boldsymbol{\eta}_{\boldsymbol{k}}$ (c.f. Section 3.2) by requiring both $C_{\boldsymbol{k}}$ and $\tilde{C}_{\boldsymbol{k}}$ be locally attainable spectral models in $Z(\boldsymbol{x})$. Therefore the corresponding quadratic estimate is tuned to detect not only a variation in the smoothness of $Z(\boldsymbol{x})$ but also a corresponding scale change, where the correspondence is related inversely (an increase in smoothness corresponding to a reduction of spatial scale and vice versa).

For this simulation example, the process $Z(\boldsymbol{x})$ is defined on $[-\pi, \pi)^2$ with periodic boundary conditions and the observed process $Z^{obs}(\boldsymbol{x})$ is generated without additive noise on a evenly spaced grid of size 400×400. Just as in Section 3.3, a simple discrete Riemann sum approximation, at each observed $\boldsymbol{x}$, was used to approximate to the integral (22). The ground truth potential $\phi(\boldsymbol{x})$, used to generate the nonstationarity in $Z(\boldsymbol{x})$, is simulated from a mean zero stationary Gaussian process with Matérn parameters $(\nu, \rho, \sigma^2) := (5, 0.3\pi, 0.5^2)$. Just as in Section 3.3 the quantities $C_{\boldsymbol{\ell}}^{(0)}, C_{\boldsymbol{\ell}}^{(1)}, C_{\boldsymbol{\ell}}^{(2)}$ and $C_{\boldsymbol{\ell}}^{ZZ^{obs}}$ used to generate $\hat{\phi}_{\boldsymbol{\ell}}$, $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ and $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ are determined by (24) and (25). Finally, the parameter t_0 used in (37) to determine $\boldsymbol{\eta}_k$ is set to 1.5.

Figure 4 graphically summarizes the simulation results. The **top left image** shows the quadratic estimate $\hat{\phi}(\boldsymbol{x})$ and the **top right image** shows the ground truth $\phi(\boldsymbol{x})$. These top images are intended to illustrate the high accuracy of the estimate. The **bottom right image** shows the data $Z^{obs}(\boldsymbol{x})$ used to generate the estimate $\hat{\phi}(\boldsymbol{x})$. The **bottom left plot** shows the radial profile of $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ (**solid green**), $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{\phi\phi}$ (**dotted black**) along with the fast approximation to $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ (**dashed blue**) and the corresponding radially averaged empirical mean squared error per wavenumber (**dotted green**). The computation of $\hat{\phi}_{\boldsymbol{\ell}}$ and $C_{\boldsymbol{\ell}}^{\text{var}\,\hat{\phi}}$ took 0.25 seconds and 0.23 seconds to compute, respectively. The fast approximation to $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ took 77 seconds (the exact value of $C_{\boldsymbol{\ell}}^{\text{bias}\,\hat{\phi}}$ is not computed in this case since the imputation is intensive and takes on the order of hours in our implementation and is not shown).

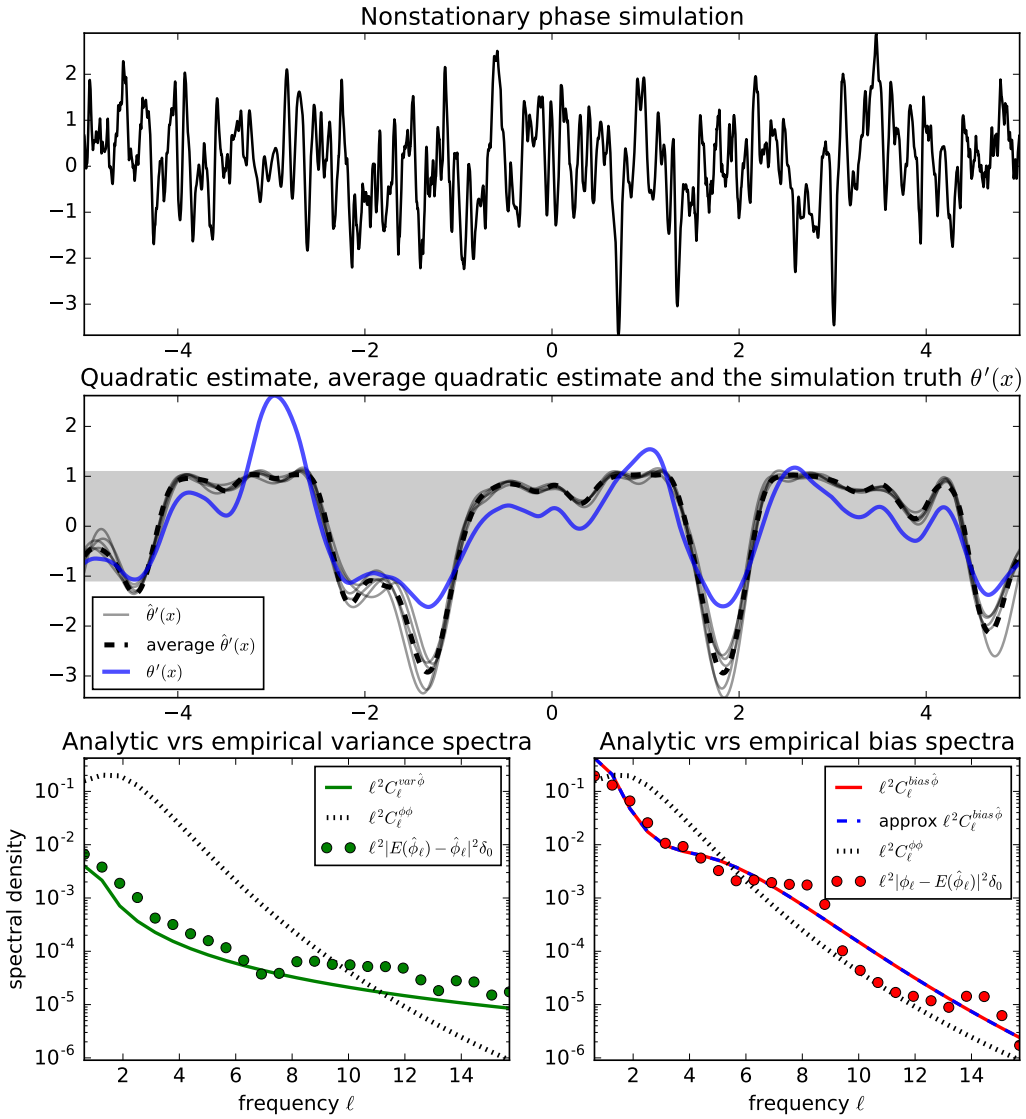

FIG 3. *An illustration of the quadratic estimate bias which results when $\theta'(x)$ is large enough to exceed the* ***symmetric two-sided cut locus*** *discussed Section 3.2 (c.f. Claim 3). The* ***shaded region*** *shown in the* ***middle plot*** *corresponds to the interior of the two-sided cut locus. When the true $\theta'(x)$ (shown in* ***blue****) exits the symmetric two-sided cut locus, the quadratic estimate suffers from large bias, attenuating for negative $\theta'(x)$ and amplifying for positive $\theta'(x)$. Note that the only difference between this figure and Figure 2 is the parameter t_0 (c.f. Section 3.2) which was reduced by a factor of $1/7$. This has the effect of shrinking the symmetric two-sided cut locus. All other parameters, including the random seed, are the same.*

4. Discusssion

Part of the motivation for this paper is an attempt to construct an extended class of nonstationary random fields, and a corresponding generalized quadratic estimate, which share the same attractive statistical properties of an estimate originally developed for gravitational lensing studies of the Cosmic Microwave Background [10, 11]. In doing so we have identified a particular form of nonstationarity, we call *local invariance*, which encourages a delicate cancellation of estimation bias. This local invariant property—we believe—is the main source of what makes the gravitational lensing estimates so successful. Indeed, the generalized quadratic estimate, derived in Section 2, shares many of the same attractive statistical features as the original gravitational lensing estimate: it is particularly adept at detecting small departures from stationarity and allows fast, accurate quantification of mean square sampling properties. In Section 3 we focus on a particular subclass of locally invariant nonstationary random fields which are given by a spatially varying spectral phase modulation of a stationary random field. In this work, the theory of optimal transport and the L_2-Wasserstein metric play a major role in characterizing the behavior of the set of possible local spectral densities under these models and leads to a natural heuristic for quantifying estimation bias in terms of the Wasserstein geodesic cut locus (see Claim 3, Section 3.2 and Figure 3).

One of the byproducts of this paper is the understanding that a nonstationary spectral phase can be estimated by analyzing the correlation among the Fourier coefficients of the nonstationary random field $Z(\boldsymbol{x})$. This was illustrated in Section 3 using a quadratic estimate to reconstruct a spatially varying spectral phase modulation of a stationary random field. Left unanswered, however, is the question of how one simultaneously estimates both the phase and the magnitude of the spectral modulation $A(\boldsymbol{k}, \boldsymbol{x})$ in model (1). It appears this line of research has the potential to merge the seminal work of Dahlhaus [2, 3] with the generalized quadratic estimate, presented here, for nonstationary estimation within a broad class of nonstationary random fields.

It is also important to mention the fact that we have derived our results under the rather idealized assumption that the observations locations form a dense regular grid and $Z(\boldsymbol{x})$ has periodic boundary conditions. Extensions to more realistic experimental conditions are not in the scope of this paper but are clearly important for real life applications. The situation is not hopeless, however, since these same features are ubiquitous in measurements of the Cosmic Microwave Background. Despite this, Cosmologists have devised methods which turn the idealized quadratic estimate into a pragmatic statistical tool for probing gravitational lensing (see [14, 24, 16, 17], for example). This suggests there exist analogous methods which can make the generalized quadratic estimate available to more general observational scenarios.

We finish with a discussion of Assumption 2 that stipulates $\boldsymbol{\theta}(\boldsymbol{x})$ be characterized by a scalar potential $\phi(\boldsymbol{x})$. It is yet unclear how one generalizes this assumption, especially in the case where $\boldsymbol{\theta}(\boldsymbol{x})$ maps into a higher dimensional space $\mathbb{R}^m$ for $m > d$. Notice that by considering a general $\boldsymbol{\theta}(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^m$ one may redefine $\boldsymbol{\theta}(\boldsymbol{x})$ by absorbing (i.e. concatenating) the observation locations $\boldsymbol{x}$ into extra coordinates of $\boldsymbol{\theta}(\boldsymbol{x})$. This generalization enables one to replace the local invariant condition $\mathrm{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) = K(\boldsymbol{x} - \boldsymbol{y}, \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(y))$ with the more general condition

$$\mathrm{cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) = K(\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(y)). \tag{38}$$

Random fields $Z(\boldsymbol{x})$, which have a covariance function of the form (38), are simply traces of stationary random fields defined on the higher dimensional space $\mathbb{R}^m$, restricted to the d-dimensional parameterized surface $\{\boldsymbol{\theta}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d\}$. Viewed from this perspective, it appears plausible that there exists a deeper, more geometric, picture of local invariance and quadratic estimation. It is not yet clear whether or not this viewpoint is useful, but it is tempting to imagine that the generalized quadratic estimate is simply a manifold embedding estimate in disguise. If such a development bears theoretical fruit, it would be a major step in the direction of a unified statistical theory of nonstationary random fields.

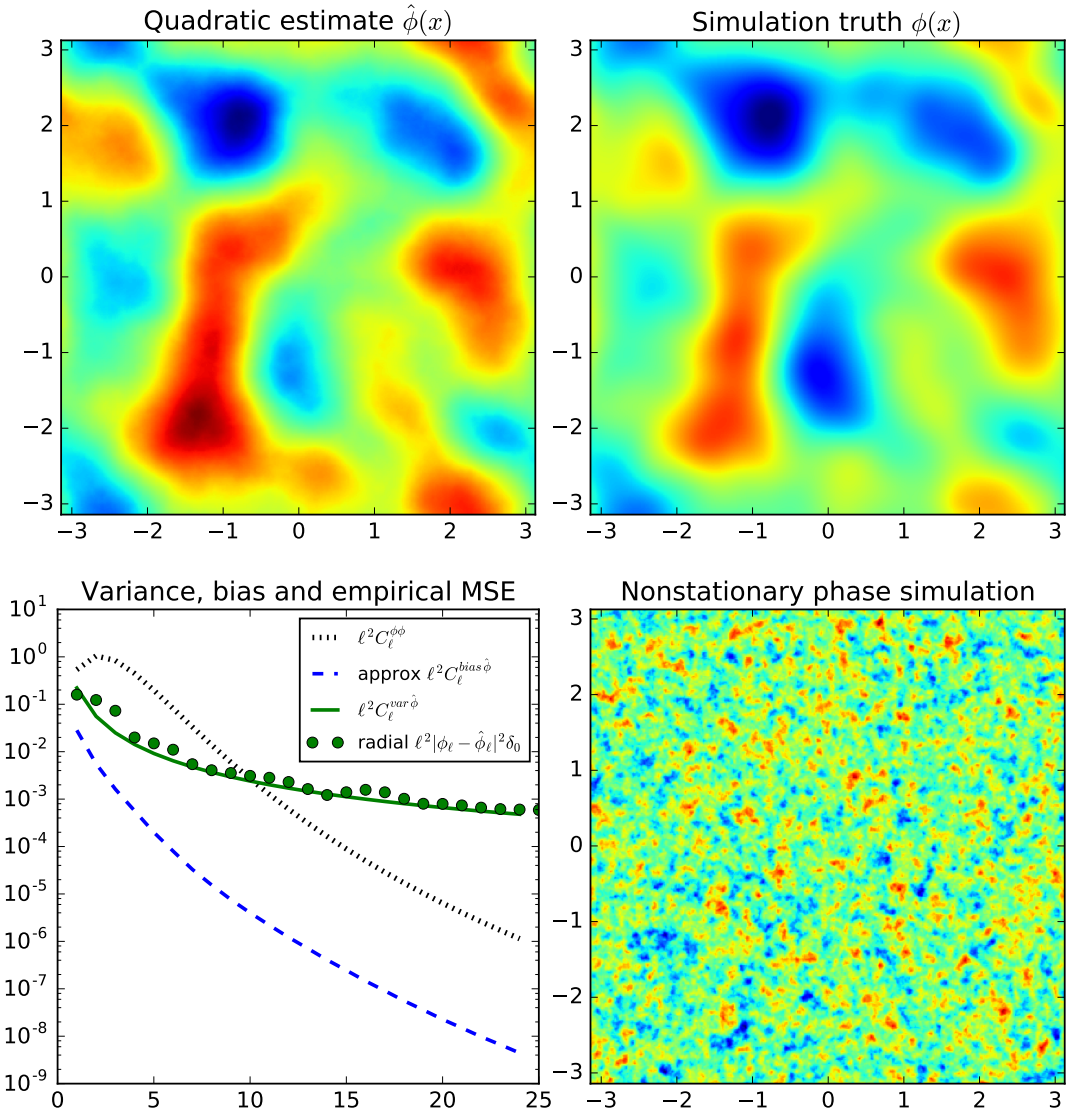

FIG 4. *An illustration of the quadratic estimate in $d = 2$ for a nonstationary phase random field $Z(\boldsymbol{x})$. The model for $Z(\boldsymbol{x})$ is defined to have local variation in both the local smoothness of $Z(\boldsymbol{x})$ and a local range parameter, where an increase of local smoothness corresponds reduction of local range and vice versa. The* ***top left image*** *shows the quadratic estimate $\hat{\phi}(\boldsymbol{x})$ with the ground truth $\phi(\boldsymbol{x})$ shown in the* ***top right image*** *and the data shown in the* ***bottom right image****. The* ***bottom left plot*** *shows the radial profile of $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{var\,\hat{\phi}}$ (**solid green**), $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{\phi\phi}$ (**dotted black**) along with the fast approximation to $|\boldsymbol{\ell}|^2 C_{\boldsymbol{\ell}}^{bias\,\hat{\phi}}$ (**dashed blue**) and the corresponding radially averaged empirical mean squared error per wavenumber (**dotted green**). See Section 3.4 for further simulation details.*

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[2] R. Dahlhaus. Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37, 1997.

[3] R. Dahlhaus. A likelihood approximation for locally stationary processes. *Annals of Statistics*, pages 1762–1794, 2000.

[4] S. Das et al. Detection of the power spectrum of cosmic microwave background lensing by the Atacama cosmology telescope. *Physical Review Letters*, 107(2):021301, 2011.

[5] G. Fuglstad, F. Lindgren, D. Simpson, and H. Rue. Exploring a new class of non-stationary spatial gaussian random fields with varying local anisotropy. *Statistica Sinica*, 25(1):115–133, 2015.

[6] G. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, 14:505–531, 2015.

[7] I. Gikhman and A. Skorokhod. *The Theory of Stochastic Processes I*. Classics in Mathematics. Springer Berlin Heidelberg, 2015.

[8] G. Giner and G. Smyth. A monotonically convergent Newton iteration for the quantiles of any unimodal distribution, with application to the inverse Gaussian distribution. 2014.

[9] T. Hsing, T. Brown, and B. Thelen. Local intrinsic stationarity and its inference. *Annals of Statistics*.

[10] W. Hu. Mapping the dark matter through the cosmic microwave background damping tail. *The Astrophysical Journal Letters*, 557(2):L79, 2001.

[11] W. Hu and T. Okamoto. Mass reconstruction with cosmic microwave background polarization. *The Astrophysical Journal*, 574(2):566, 2002.

[12] I. Ibragimov and Yurii A. Rozanov. *Gaussian random processes*, volume 9. Springer Science & Business Media, 2012.

[13] L. Isserlis. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika*, 11(3):185–190, 1916.

[14] T. Namikawa, D. Hanson, and R. Takahashi. Bias-hardened CMB lensing. *Monthly Notices of the Royal Astronomical Society*, 431(1):609–620, 2013.

[15] C. Paciorek and M. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506, 2006.

[16] Planck Collaboration. Planck 2013 results. XVII. Gravitational lensing by large-scale structure. *Astronomy and Astrophysics*, 571:A17, November 2014.

[17] Planck Collaboration. Planck 2015 results. XV. Gravitational lensing. *ArXiv e-prints*, February 2015.

[18] M. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237, 1965.

[19] M. Priestley. *Spectral analysis and time series*, volume 2. Academic press, 1981.

[20] P. Sampson. Constructions for nonstationary spatial processes. *Handbook of Spatial Statistics*, pages 119–130, 2010.

[21] P. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

[22] M. Stein. *Interpolation of spatial data: some theory for Kriging*. Springer Science & Business Media, 2012.

[23] The Polarbear Collaboration. A Measurement of the Cosmic Microwave Background B-mode Polarization Power Spectrum at Sub-degree Scales with POLARBEAR. *The Astrophysical Journal*, 794:171, October 2014.

[24] A. Van Engelen et al. A measurement of gravitational lensing of the microwave background using south pole telescope data. *The Astrophysical Journal*, 756(2):142, 2012.

[25] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[26] G. Wick. The evaluation of the collision matrix. *Physical review*, 80(2):268, 1950.

Appendix A: Detailed derivations

Claim 4. *Let $\boldsymbol{\theta}(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^d$ be a vector field and $Z(\boldsymbol{x})$ be a random field which satisfies $E\big(Z(\boldsymbol{x})Z(\boldsymbol{y})|\boldsymbol{\theta}(\cdot)\big) = C^{(0)}(\boldsymbol{x}-\boldsymbol{y}) + \boldsymbol{C}^{(1)}(\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y})) + \mathcal{O}(\boldsymbol{\theta}^2)$. Then*

$$E(Z_{\boldsymbol{k}+\boldsymbol{\ell}}Z_{-\boldsymbol{k}}|\boldsymbol{\theta}(\cdot)) = \boldsymbol{\theta}_{\boldsymbol{\ell}}\cdot\big(\boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\big) + \mathcal{O}(\boldsymbol{\theta}^2) \tag{39}$$

when $\boldsymbol{\ell}\neq 0$.

Proof. The Fourier transform $C^{(0)}(\boldsymbol{x}-\boldsymbol{y})$, with respect to $\boldsymbol{x}$ and $\boldsymbol{y}$, gives $\delta_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2}(2\pi)^{d/2}C^{(0)}_{\boldsymbol{\ell}_1}$. Similarly, the Fourier transform of $\boldsymbol{\theta}(\boldsymbol{x})\cdot\boldsymbol{C}^{(1)}(\boldsymbol{x}-\boldsymbol{y})$ and $-\boldsymbol{\theta}(\boldsymbol{y})\cdot\boldsymbol{C}^{(1)}(\boldsymbol{x}-\boldsymbol{y})$, with respect to $\boldsymbol{x}$ and $\boldsymbol{y}$, gives $\boldsymbol{\theta}_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2}\cdot\boldsymbol{C}^{(1)}_{-\boldsymbol{\ell}_2}$ and $-\boldsymbol{\theta}_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2}\cdot\boldsymbol{C}^{(1)}_{\boldsymbol{\ell}_1}$ respectively. Summing these three terms gives

$$E\big(Z_{\boldsymbol{\ell}_1}Z_{\boldsymbol{\ell}_2}|\boldsymbol{\theta}(\cdot)\big) = (2\pi)^{d/2}C^{(0)}_{\boldsymbol{\ell}_1}\delta_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2} + \boldsymbol{\theta}_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2}\cdot\big(\boldsymbol{C}^{(1)}_{-\boldsymbol{\ell}_2} - \boldsymbol{C}^{(1)}_{\boldsymbol{\ell}_1}\big) + \mathcal{O}(\boldsymbol{\theta}^2).$$

Replacing $\boldsymbol{\ell}_1$ with $\boldsymbol{k}+\boldsymbol{\ell}$ and $\boldsymbol{\ell}_2$ with $-\boldsymbol{k}$ finishes the derivation. □

Claim 5. *Suppose Assumptions 1, 2 and 3 hold. Then the first order unbiased quadratic estimate of $\phi_{\boldsymbol{\ell}}$, which corresponds to an approximate inverse variance weighted averaging of $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}$, has the form*

$$\hat{\phi}_{\boldsymbol{\ell}} = A_{\boldsymbol{\ell}}\sum_{p=1}^{d}\boldsymbol{\xi}^*_{p,\boldsymbol{\ell}}\int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}}\mathscr{A}(\boldsymbol{x})\mathscr{B}_p(\boldsymbol{x})\frac{d\boldsymbol{x}}{(2\pi)^{d/2}} \tag{40}$$

where $\mathscr{A}_{\boldsymbol{\ell}} := Z^{obs}_{\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{B}_{p,\boldsymbol{\ell}} := i2\,imag(\boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}})Z^{obs}_{\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$ and the normalizing constant $A_{\boldsymbol{\ell}}$ is given by

$$A^{-1}_{\boldsymbol{\ell}} = \sum_{p,q=1}^{d}\boldsymbol{\xi}_{p,\boldsymbol{\ell}}\boldsymbol{\xi}^*_{q,\boldsymbol{\ell}}\int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}}\left[2\mathscr{A}_{p,q}(\boldsymbol{x})\mathscr{B}(\boldsymbol{x}) - \mathscr{C}_p(\boldsymbol{x})\mathscr{C}_q(\boldsymbol{x}) - \mathscr{D}_p(\boldsymbol{x})\mathscr{D}_q(\boldsymbol{x})\right]\frac{d\boldsymbol{x}}{(2\pi)^{d/2}} \tag{41}$$

where $\mathscr{A}_{p,q,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}}\boldsymbol{C}^{(1)^*}_{q,\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{B}_{\boldsymbol{\ell}} := 1/C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{C}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$ and $\mathscr{D}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{\ell}}/C^{ZZobs}_{\boldsymbol{\ell}}$.

Proof. By Claim 4 we have that

$$E\big(Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}\big) = \phi_{\boldsymbol{\ell}}\left(\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}\right) + \mathcal{O}(\phi^2) \tag{42}$$

when $\boldsymbol{\ell}\neq 0$ where $\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}$. Therefore the quadratic estimate, as a weighted average of the first order unbiased terms $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}/(\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}})$, can be written in the form

$$\hat{\phi}_{\boldsymbol{\ell}} = \int w_{\boldsymbol{k},\boldsymbol{\ell}}\frac{Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}}{\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}}\frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \tag{43}$$

where $w_{\boldsymbol{k},\boldsymbol{\ell}}\geq 0$ are normalized so that $\hat{\phi}_{\boldsymbol{\ell}}$ has expected value $\phi_{\boldsymbol{\ell}}+\mathcal{O}(\phi^2)$ using (42). Assuming $\boldsymbol{\ell}\neq 0$, the Gaussian part of the variance of $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}/(\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}})$ can be computed as follows

$$\begin{aligned}
\operatorname{var}\left(\frac{Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}}{\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}}\right) &= \frac{1}{|\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}|^2}\Big[E(Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}}Z^{obs}_{-\boldsymbol{k}-\boldsymbol{\ell}}Z^{obs}_{\boldsymbol{k}}) - E(Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}})E(Z^{obs}_{-\boldsymbol{k}-\boldsymbol{\ell}}Z^{obs}_{\boldsymbol{k}})\Big]\\
&\approx \frac{1}{|\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}|^2}\underbrace{\Big[E(Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{-\boldsymbol{k}-\boldsymbol{\ell}})E(Z^{obs}_{\boldsymbol{k}}Z^{obs}_{-\boldsymbol{k}}) + E(Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}Z^{obs}_{\boldsymbol{k}})E(Z^{obs}_{-\boldsymbol{k}}Z^{obs}_{-\boldsymbol{k}-\boldsymbol{\ell}})\Big]}_{\text{only keeping the Gaussian part of the trispectrum}}\\
&= \frac{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}}C^{ZZobs}_{\boldsymbol{k}}}{|\boldsymbol{\xi}_{\boldsymbol{\ell}}\cdot\boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}|^2}\left[\delta^2_{\boldsymbol{0}} + \delta^2_{2\boldsymbol{k}+\boldsymbol{\ell}}\right].
\end{aligned}$$

where $C_{\boldsymbol{\ell}}^{ZZobs}$ denotes the spectral density of $Z^{obs}(\boldsymbol{x})$ marginalized over $\phi(\cdot)$. If we ignore the term $\delta^2_{2\boldsymbol{k}+\boldsymbol{\ell}}$, which only activates at the point $\boldsymbol{k} = -\boldsymbol{\ell}/2$, then by defining $w_{\boldsymbol{k},\boldsymbol{\ell}}$ in (43) to be proportional to the approximate inverse (Gaussian part of the) variance of $Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{-\boldsymbol{k}} / (\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}})$ one has

$$\begin{aligned}
\hat{\phi}_{\boldsymbol{\ell}} &= A_{\boldsymbol{\ell}} \int \frac{|\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}|^2}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{-\boldsymbol{k}}}{\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \sum_{p=1}^{d} \boldsymbol{\xi}^*_{p,\boldsymbol{\ell}} \int \Big[\frac{Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}}} \frac{\boldsymbol{C}^{(1)}_{p,-\boldsymbol{k}} Z^{obs}_{-\boldsymbol{k}}}{C^{ZZobs}_{-\boldsymbol{k}}} - \frac{\boldsymbol{C}^{(1)^*}_{p,\boldsymbol{k}+\boldsymbol{\ell}} Z^{obs}_{\boldsymbol{k}+\boldsymbol{\ell}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}}} \frac{Z^{obs}_{-\boldsymbol{k}}}{C^{ZZobs}_{-\boldsymbol{k}}} \Big] \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \sum_{p=1}^{d} \boldsymbol{\xi}^*_{p,\boldsymbol{\ell}} \int \left[\mathscr{A}_{\boldsymbol{k}+\boldsymbol{\ell}} \mathscr{D}_{p,-\boldsymbol{k}} - \mathscr{C}_{p,\boldsymbol{k}+\boldsymbol{\ell}} \mathscr{A}_{-\boldsymbol{k}} \right] \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \sum_{p=1}^{d} \boldsymbol{\xi}^*_{p,\boldsymbol{\ell}} \int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}} \left[\mathscr{A}(\boldsymbol{x}) \mathscr{D}_p(\boldsymbol{x}) - \mathscr{C}_p(\boldsymbol{x}) \mathscr{A}(\boldsymbol{x}) \right] \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}
\end{aligned} \tag{44}$$

where $\mathscr{A}_{\boldsymbol{\ell}} := Z^{obs}_{\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{D}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} Z^{obs}_{\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$ and $\mathscr{C}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{\ell}} Z^{obs}_{\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$. Notice that the Fourier transform of $\mathscr{D}_p(\boldsymbol{x}) - \mathscr{C}_p(\boldsymbol{x})$ can be simplified as follows

$$\mathscr{D}_{p,\boldsymbol{\ell}} - \mathscr{C}_{p,\boldsymbol{\ell}} = \left[\boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} - \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{\ell}} \right] \frac{Z^{obs}_{\boldsymbol{\ell}}}{C^{ZZobs}_{\boldsymbol{\ell}}} = i2\, imag(\boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}}) \frac{Z^{obs}_{\boldsymbol{\ell}}}{C^{ZZobs}_{\boldsymbol{\ell}}}.$$

This gives (40) as was to be shown.

The normalizing constant $A_{\boldsymbol{\ell}}$ is defined so that the right hand of (44) is unbiased (up to first order). Utilizing (42) this unbiased constraint is written as follows

$$\begin{aligned}
1 &= A_{\boldsymbol{\ell}} \int \frac{|\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{f}_{\boldsymbol{k},\boldsymbol{\ell}}|^2}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \int \frac{\left| \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} \right|^2}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \sum_{p,q=1}^{d} \boldsymbol{\xi}_{p,\boldsymbol{\ell}} \boldsymbol{\xi}^*_{q,\boldsymbol{\ell}} \int \frac{\boldsymbol{C}^{(1)}_{p,\boldsymbol{k}+\boldsymbol{\ell}} \boldsymbol{C}^{(1)^*}_{q,\boldsymbol{k}+\boldsymbol{\ell}} + \boldsymbol{C}^{(1)}_{p,\boldsymbol{k}} \boldsymbol{C}^{(1)^*}_{q,\boldsymbol{k}} - \boldsymbol{C}^{(1)}_{p,\boldsymbol{k}+\boldsymbol{\ell}} \boldsymbol{C}^{(1)^*}_{q,\boldsymbol{k}} - \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{k}+\boldsymbol{\ell}} \boldsymbol{C}^{(1)}_{q,\boldsymbol{k}}}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}}} \frac{d\boldsymbol{k}}{(2\pi)^{d/2}} \\
&= A_{\boldsymbol{\ell}} \sum_{p,q=1}^{d} \boldsymbol{\xi}_{p,\boldsymbol{\ell}} \boldsymbol{\xi}^*_{q,\boldsymbol{\ell}} \int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}} \left[2\mathscr{A}_{p,q}(\boldsymbol{x}) \mathscr{B}(\boldsymbol{x}) - \mathscr{C}_p(\boldsymbol{x}) \mathscr{C}_q(\boldsymbol{x}) - \mathscr{D}_p(\boldsymbol{x}) \mathscr{D}_q(\boldsymbol{x}) \right] \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}
\end{aligned} \tag{45}$$

where $\mathscr{A}_{p,q,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} \boldsymbol{C}^{(1)^*}_{q,\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{B}_{\boldsymbol{\ell}} := 1/C^{ZZobs}_{\boldsymbol{\ell}}$, $\mathscr{C}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$ and $\mathscr{D}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$. □

Claim 6 (Estimation variance). *Suppose $X(\boldsymbol{x})$ is a mean-zero Gaussian random field with spectral density given by $C^{XX}_{\boldsymbol{\ell}}$. Then the spectral density of $\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}$ (c.f. Definition 2), which satisfies $\delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'} C^{\text{var}\,\hat{\phi}}_{\boldsymbol{\ell}} = E\big(\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}\, \hat{\phi}_{\boldsymbol{\ell}'}\{X,X\}^*\big)$, is given as follows*

$$C^{\text{var}\,\hat{\phi}}_{\boldsymbol{\ell}} = 2A^2_{\boldsymbol{\ell}} \int \left| \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}} \right|^2 \frac{C^{XX}_{\boldsymbol{k}+\boldsymbol{\ell}}}{(C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}})^2} \frac{C^{XX}_{\boldsymbol{k}}}{(C^{ZZobs}_{\boldsymbol{k}})^2} \frac{d\boldsymbol{k}}{(2\pi)^d} \tag{46}$$

$$= 2A^2_{\boldsymbol{\ell}} \sum_{p,q=1}^{d} \boldsymbol{\xi}_{p,\boldsymbol{\ell}} \boldsymbol{\xi}^*_{q,\boldsymbol{\ell}} \int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}} \left[2\mathscr{A}_{p,q}(\boldsymbol{x}) \mathscr{B}(\boldsymbol{x}) - \mathscr{C}_p(\boldsymbol{x}) \mathscr{C}_q(\boldsymbol{x}) - \mathscr{D}_p(\boldsymbol{x}) \mathscr{D}_q(\boldsymbol{x}) \right] \frac{d\boldsymbol{x}}{(2\pi)^d} \tag{47}$$

for all $\boldsymbol{\ell} \neq 0$ *where* $\mathscr{A}_{p,q,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} \boldsymbol{C}^{(1)^*}_{q,\boldsymbol{\ell}} C^{XX}_{\boldsymbol{\ell}} / (C^{ZZobs}_{\boldsymbol{\ell}})^2$, $\mathscr{B}_{\ell} := C^{XX}_{\boldsymbol{\ell}} / (C^{ZZobs}_{\boldsymbol{\ell}})^2$, $\mathscr{C}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)}_{p,\boldsymbol{\ell}} C^{XX}_{\boldsymbol{\ell}} / (C^{ZZobs}_{\boldsymbol{\ell}})^2$ *and* $\mathscr{D}_{p,\boldsymbol{\ell}} := \boldsymbol{C}^{(1)^*}_{p,\boldsymbol{\ell}} C^{XX}_{\boldsymbol{\ell}} / (C^{ZZobs}_{\boldsymbol{\ell}})^2$. *Moreover, if* $C^{XX}_{\boldsymbol{\ell}} = C^{ZZobs}_{\boldsymbol{\ell}}$ *one obtains*

$$C^{\text{var}\,\hat{\phi}}_{\boldsymbol{\ell}} = 2(2\pi)^{-d/2} A_{\boldsymbol{\ell}}. \tag{48}$$

Proof. Recall Definition 2 which states that $\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}$ denotes the quadratic estimate applied to data $X(\boldsymbol{x})$. Therefore

$$E\big(\hat{\phi}_{\boldsymbol{\ell}}\{X,X\}\,\hat{\phi}_{\boldsymbol{\ell}'}\{X,X\}^*\big) = A_{\boldsymbol{\ell}} A_{\boldsymbol{\ell}'} \iint \Big(\boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}+\boldsymbol{\ell}}\Big)^* \Big(\boldsymbol{\xi}_{\boldsymbol{\ell}'} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}'} - \boldsymbol{\xi}_{\boldsymbol{\ell}'} \cdot \boldsymbol{C}^{(1)}_{\boldsymbol{k}'+\boldsymbol{\ell}'}\Big) \times \frac{E\big(X_{\boldsymbol{k}+\boldsymbol{\ell}} X_{-\boldsymbol{k}} X_{-\boldsymbol{k}'-\boldsymbol{\ell}'} X_{\boldsymbol{k}'}\big)}{C^{ZZobs}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{ZZobs}_{\boldsymbol{k}} C^{ZZobs}_{\boldsymbol{k}'+\boldsymbol{\ell}'} C^{ZZobs}_{\boldsymbol{k}'}} \frac{d\boldsymbol{k} d\boldsymbol{k}'}{(2\pi)^d}. \tag{49}$$

Now expanding the above fourth moment, using Wick's theorem (also called Isserlis's Theorem) [26, 13] and the Gaussianity of $X(\boldsymbol{x})$, one obtains

$$\begin{aligned} E\big(X_{\boldsymbol{k}+\boldsymbol{\ell}} X_{-\boldsymbol{k}} X_{-\boldsymbol{k}'-\boldsymbol{\ell}'} X_{\boldsymbol{k}'}\big) &= C^{XX}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{XX}_{-\boldsymbol{k}'} \big(\delta_{\boldsymbol{\ell}}\delta_{\boldsymbol{\ell}'} + \delta_{\boldsymbol{k}-\boldsymbol{k}'+\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}-\boldsymbol{k}'} + \delta_{\boldsymbol{k}+\boldsymbol{k}'+\boldsymbol{\ell}}\delta_{\boldsymbol{k}+\boldsymbol{k}'+\boldsymbol{\ell}'}\big) \\ &= C^{XX}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{XX}_{-\boldsymbol{k}'} \big(\delta_{\boldsymbol{\ell}}\delta_{\boldsymbol{\ell}'} + \delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}-\boldsymbol{k}'} + \delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}+\boldsymbol{k}'+\boldsymbol{\ell}'}\big). \end{aligned}$$

The term $\delta_{\boldsymbol{\ell}}\delta_{\boldsymbol{\ell}'}$ is only nonzero when $\boldsymbol{\ell} = \boldsymbol{\ell}' = 0$. Furthermore, by a change of variables, one can see that effect of the terms $\delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}-\boldsymbol{k}'}$ and $\delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}+\boldsymbol{k}'+\boldsymbol{\ell}'}$ in (49) are identical. Therefore assuming $\boldsymbol{\ell} \neq 0$ and replacing $E\big(X_{\boldsymbol{k}+\boldsymbol{\ell}} X_{-\boldsymbol{k}} X_{-\boldsymbol{k}'-\boldsymbol{\ell}'} X_{\boldsymbol{k}'}\big)$ in (49) with $2C^{XX}_{\boldsymbol{k}+\boldsymbol{\ell}} C^{XX}_{-\boldsymbol{k}'} \delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'}\delta_{\boldsymbol{k}-\boldsymbol{k}'}$ gives (46). A similar method to the derivation for (45) can be used to establish (47). Finally, if one replaces $C^{XX}_{\boldsymbol{k}}$ in (46) with the marginal spectral density $C^{ZZobs}_{\boldsymbol{k}}$ then using (45) one obtains $C^{\text{var}\,\hat{\phi}}_{\boldsymbol{\ell}} = 2(2\pi)^{-d/2} A_{\boldsymbol{\ell}}$. □

Claim 7 (Estimation bias). *Let* $\mathcal{O}(\phi^2)(\boldsymbol{x},\boldsymbol{y}) := (\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{y}))^T \boldsymbol{C}^{(2)}(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\boldsymbol{y}))$ *where* $\boldsymbol{C}^{(2)}(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *and* $\boldsymbol{C}^{(2)}(-\boldsymbol{x}) = \boldsymbol{C}^{(2)}(\boldsymbol{x})$. *Then*

$$\mathcal{O}(\phi^2)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}} = \sum_{p,q=1}^{d} \int \boldsymbol{\theta}_{p,\boldsymbol{\omega}} \boldsymbol{\theta}_{q,\boldsymbol{\ell}-\boldsymbol{\omega}} \Big(\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} + \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\ell}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\ell}-\boldsymbol{\omega}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\Big) \frac{d\boldsymbol{\omega}}{(2\pi)^{d/2}}. \tag{50}$$

Define $\hat{\phi}^{bias}_{\boldsymbol{\ell}} := \hat{\phi}_{\boldsymbol{\ell}}\big\{\mathcal{O}(\phi^2)_{\boldsymbol{k}+\boldsymbol{\ell},-\boldsymbol{k}}\big\}$ *(see Definition 2), where the quadratic estimate* $\hat{\phi}_{\boldsymbol{\ell}}$ *satisfies results all the assumptions given in Claim 5. Then*

$$\hat{\phi}^{bias}_{\boldsymbol{\ell}} = 2 \sum_{p,q=1}^{d} \int \boldsymbol{\theta}_{p,\boldsymbol{\omega}} \boldsymbol{\theta}_{q,\boldsymbol{\ell}-\boldsymbol{\omega}}\, \hat{\phi}_{\boldsymbol{\ell}}\Big\{\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\Big\} \frac{d\boldsymbol{\omega}}{(2\pi)^{d/2}} \tag{51}$$

and the spectral density of $\hat{\phi}^{bias}_{\boldsymbol{\ell}}$*, under a mean zero Gaussian random field model for* $\boldsymbol{\theta}(\boldsymbol{x})$ *with spectral density matrix* $C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{\boldsymbol{\ell}}$*, satisfies*

$$E\big(\hat{\phi}^{bias}_{\boldsymbol{\ell}} \hat{\phi}^{bias^*}_{\boldsymbol{\ell}'}\big) = 4\delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'} \sum_{p,q,p',q'=1}^{d} \int \Big(C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,p',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,q',\boldsymbol{\ell}-\boldsymbol{\omega}} + C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,q',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,p',\boldsymbol{\ell}-\boldsymbol{\omega}}\Big) \times \hat{\phi}_{\boldsymbol{\ell}}\Big\{\boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}}\Big\} \hat{\phi}_{\boldsymbol{\ell}}\Big\{\boldsymbol{C}^{(2)}_{p',q',\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p',q',\boldsymbol{k}+\boldsymbol{\omega}}\Big\}^* \frac{d\boldsymbol{\omega}}{(2\pi)^d} \tag{52}$$

when $\boldsymbol{\ell} \neq 0$ *or* $\boldsymbol{\ell}' \neq 0$.

Proof. Notice that for any $p, q \in \{1, \dots, d\}$ if one defines $B_{p,q}(\boldsymbol{x}, \boldsymbol{y}) := (\boldsymbol{\theta}_p(\boldsymbol{x}) - \boldsymbol{\theta}_p(\boldsymbol{y}))(\boldsymbol{\theta}_q(\boldsymbol{x}) - \boldsymbol{\theta}_q(\boldsymbol{y}))$ then the Fourier transform of $B_{p,q}(\boldsymbol{x}, \boldsymbol{y})$ with respect to $(\boldsymbol{x}, \boldsymbol{y})$, evaluated at frequency vector $(\boldsymbol{k}, \boldsymbol{k}')$, is given by

$$B_{p,q,\boldsymbol{k},\boldsymbol{k}'} = \int \boldsymbol{\theta}_{p,\boldsymbol{\omega}} \Big(\boldsymbol{\theta}_{q,\boldsymbol{k}-\boldsymbol{\omega}} \delta_{-\boldsymbol{k}'} + \boldsymbol{\theta}_{q,\boldsymbol{k}'-\boldsymbol{\omega}} \delta_{-\boldsymbol{k}} \delta_{\boldsymbol{\omega}+\boldsymbol{\omega}'-\boldsymbol{k}'} - \boldsymbol{\theta}_{q,\boldsymbol{k}'} \delta_{\boldsymbol{\omega}-\boldsymbol{k}} \delta_{\boldsymbol{\omega}'-\boldsymbol{k}'} - \boldsymbol{\theta}_{q,\boldsymbol{k}} \delta_{\boldsymbol{\omega}'-\boldsymbol{k}} \delta_{\boldsymbol{\omega}-\boldsymbol{k}'} \Big) d\boldsymbol{\omega}.$$

Using the fact that Fourier transform of $\boldsymbol{C}^{(2)}_{p,q}(\boldsymbol{x} - \boldsymbol{y})$ equals $(2\pi)^{d/2} \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} \delta_{\boldsymbol{k}+\boldsymbol{k}'}$ one obtains

$$\begin{aligned}
\mathcal{O}(\phi^2)_{\boldsymbol{k},\boldsymbol{k}'} &= \sum_{p,q=1}^{d} \Big[\underbrace{\boldsymbol{C}^{(2)}_{p,q}(\boldsymbol{x} - \boldsymbol{y})}_{A_{p,q}(\boldsymbol{x},\boldsymbol{y})} \underbrace{(\boldsymbol{\theta}_p(\boldsymbol{x}) - \boldsymbol{\theta}_p(\boldsymbol{y}))(\boldsymbol{\theta}_q(\boldsymbol{x}) - \boldsymbol{\theta}_q(\boldsymbol{y}))}_{B_{p,q}(\boldsymbol{x},\boldsymbol{y})} \Big]_{\boldsymbol{k},\boldsymbol{k}'} \\
&= \sum_{p,q=1}^{d} \iint A_{p,q,\boldsymbol{z}+\boldsymbol{k},\boldsymbol{z}'+\boldsymbol{k}'} B_{p,q,-\boldsymbol{z},-\boldsymbol{z}'} \frac{d\boldsymbol{z}\, d\boldsymbol{z}'}{(2\pi)^d} \\
&= \sum_{p,q=1}^{d} \int \boldsymbol{\theta}_{p,\boldsymbol{\omega}} \boldsymbol{\theta}_{q,\boldsymbol{k}+\boldsymbol{k}'-\boldsymbol{\omega}} \Big(\boldsymbol{C}^{(2)}_{p,q,-\boldsymbol{k}'} + \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}-\boldsymbol{\omega}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{\omega}-\boldsymbol{k}'} \Big) \frac{d\boldsymbol{\omega}}{(2\pi)^{d/2}}. \qquad (53)
\end{aligned}$$

Making the substitution $\boldsymbol{k}' \to -\boldsymbol{k}$ and $\boldsymbol{k} \to \boldsymbol{k} + \boldsymbol{\ell}$ in (53) proves (50). Equation (51) immediately follows from the fact that

$$\hat{\phi}_{\boldsymbol{\ell}} \Big\{ \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\ell}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\ell}-\boldsymbol{\omega}} \Big\} = \hat{\phi}_{\boldsymbol{\ell}} \Big\{ \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}} - \boldsymbol{C}^{(2)}_{p,q,\boldsymbol{k}+\boldsymbol{\omega}} \Big\}$$

which is established by utilizing the three properties $\boldsymbol{C}^{(2)}_{-\boldsymbol{k}} = \boldsymbol{C}^{(2)}_{\boldsymbol{k}} \in \mathbb{R}^{d \times d}$, $\boldsymbol{C}^{(1)^*}_{-\boldsymbol{k}} = \boldsymbol{C}^{(1)}_{\boldsymbol{k}}$ and $\boldsymbol{C}^{(1)}_{-\boldsymbol{k}} = -\boldsymbol{C}^{(1)}_{\boldsymbol{k}}$ along with the change of variables $\tilde{\boldsymbol{k}} = -\boldsymbol{k} - \boldsymbol{\ell}$. Finally, using Wick's theorem for Gaussian $\boldsymbol{\theta}(\boldsymbol{x})$ gives

$$E\big(\boldsymbol{\theta}_{p,\boldsymbol{\omega}} \boldsymbol{\theta}_{q,\boldsymbol{\ell}-\boldsymbol{\omega}} \boldsymbol{\theta}^*_{p',\boldsymbol{\omega}'} \boldsymbol{\theta}^*_{q',\boldsymbol{\ell}'-\boldsymbol{\omega}'}\big) = \delta_{\boldsymbol{\ell}-\boldsymbol{\ell}'} \big(C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,p',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,q',\boldsymbol{\ell}-\boldsymbol{\omega}} \delta_{\boldsymbol{\ell}-\boldsymbol{\omega}'-\boldsymbol{\omega}} + C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{p,q',\boldsymbol{\omega}} C^{\boldsymbol{\theta}\boldsymbol{\theta}}_{q,p',\boldsymbol{\ell}-\boldsymbol{\omega}} \delta_{\boldsymbol{\omega}'+\boldsymbol{\omega}} \big) \qquad (54)$$

when $\boldsymbol{\ell} \neq 0$ or $\boldsymbol{\ell}' \neq 0$. Expanding the quadratic $E\big(\hat{\phi}^{\text{bias}}_{\boldsymbol{\ell}} \hat{\phi}^{\text{bias}^*}_{\boldsymbol{\ell}'}\big)$, applying Fubini and (54) then gives (52) as was to be shown. □