

## Wavelet invariants for statistically robust multi-reference alignment

MATTHEW HIRN

Department of Computational Mathematics, Science and Engineering, Department of Mathematics  
and Center for Quantum Computing, Science and Engineering,  
Michigan State University, East Lansing, MI 48824

AND

ANNA LITTLE<sup>†</sup>

Department of Computational Mathematics, Science and Engineering,  
Michigan State University, East Lansing, MI 48824

<sup>†</sup>Corresponding author. Email: littl119@msu.edu

[Received on 9 October 2019; revised on 21 May 2020; accepted on 9 June 2020]

We propose a nonlinear, wavelet-based signal representation that is translation invariant and robust to both additive noise and random dilations. Motivated by the multi-reference alignment problem and generalizations thereof, we analyze the statistical properties of this representation given a large number of independent corruptions of a target signal. We prove the nonlinear wavelet-based representation uniquely defines the power spectrum but allows for an unbiasing procedure that cannot be directly applied to the power spectrum. After unbiasing the representation to remove the effects of the additive noise and random dilations, we recover an approximation of the power spectrum by solving a convex optimization problem, and thus reduce to a phase retrieval problem. Extensive numerical experiments demonstrate the statistical robustness of this approximation procedure.

**Keywords:** multi-reference alignment; method of invariants; wavelets; signal processing; wavelet scattering transform.

### 1. Introduction

The goal in classic multi-reference alignment (MRA) is to recover a hidden signal  $f : \mathbb{R} \rightarrow \mathbb{R}$  from a collection of noisy measurements. Specifically, the following data model is assumed.

**MODEL 1** (Classic MRA). The *classic MRA data model* consists of  $M$  independent observations of a compactly supported, real-valued signal  $f \in \mathbf{L}^2(\mathbb{R})$ :

$$y_j(x) = f(x - t_j) + \varepsilon_j(x), \quad 1 \leq j \leq M, \quad (1.1)$$

where:

- (i)  $\text{supp}(y_j) \subseteq \left[ -\frac{1}{2}, \frac{1}{2} \right]$  for  $1 \leq j \leq M$ .
- (ii)  $\{t_j\}_{j=1}^M$  are independent samples of a random variable  $t \in \mathbb{R}$ .

- (iii)  $\{\varepsilon_j(x)\}_{j=1}^M$  are independent white noise processes on  $[-\frac{1}{2}, \frac{1}{2}]$  with variance  $\sigma^2$ .

The signal is thus subjected to both random translation and additive noise. The MRA problem arises in numerous applications, including structural biology [32,64,65,70,71,79], single cell genomic sequencing [51], radar [43,85], crystalline simulations [76], image registration [18,40,69] and signal processing [85]. It is a simplified model relevant for cryo-electron microscopy (cryo-EM), an imaging technique for molecules that achieves near atomic resolution [11,14,75]. In this application one seeks to recover a three-dimensional reconstruction of the molecule from many noisy two-dimensional images/projections [41]. Although MRA ignores the tomographic projection of cryo-EM, investigation of the simplified model provides important insights. For example, [5,66] investigate the optimal sample complexity for MRA and demonstrate that  $M = \Theta(\sigma^6)$  is required to fully recover  $f$  in the low signal-to-noise regime when the translation distribution is periodic; this optimal sample complexity is the same for cryo-EM [7,82]. Recent work has established an improved sample complexity of  $M = \Theta(\sigma^4)$  for MRA when the translation distribution is aperiodic [1], and this rate has been shown to also hold in the more complicated setting of cryo-EM, if the viewing angles are non-uniformly distributed [72]. Problems closely related to Model 1 include the heterogenous MRA problem, where the unknown signal  $f$  is replaced with a template of  $k$  unknown signals  $f_1, \dots, f_k$  [16,54,66,77], as well as multi-reference factor analysis, where the underlying (random) signal follows a low-rank factor model and one seeks to recover its covariance matrix [50].

Approaches for solving MRA generally fall into two categories: *synchronization methods* and methods that estimate the signal directly, i.e. without estimating nuisance parameters. Synchronization methods attempt to recover the signal by aligning the translations and then averaging. They include methods based on *angular synchronization* [8,15,24,67,73,84], where for each pair of signals the best pairwise shift is computed and then the translations are estimated from this pairwise information [6], and *semi-definite programming* [4,9,10,25], which approximates the quasi-maximum likelihood estimator of the shifts by relaxing a non-convex rank constraint. However, these methods fail in the low signal-to-noise regime. Methods that estimate the signal directly include both the *method of moments* [44,48,72] and *expectation maximization, or EM-type, algorithms* [1,30]; a number of EM-type algorithms have also been developed for the more complicated cryo-EM problem [33,68]. An important special case of the method of moments is the *method of invariants*, which seeks to recover  $f$  by computing translation invariant features, and thus avoids aligning the translations. However, the task is a difficult one, as a complete representation is needed to recover the signal, and yet the representation may be difficult to invert and corrupted by statistical bias. Generally, the signal is recovered from translation invariant moments, which are estimated in the Fourier domain [29,44]. Recent work [5,14] utilizes such Fourier invariants (mean, power spectrum and bispectrum) and recovers  $\hat{f}$  by solving a non-convex optimization problem on the manifold of phases.

Classic MRA however fails to capture many of the biological phenomena arising in molecular imaging, such as the random rotations of the molecules and the tomographic projection associated with the imaging of three-dimensional objects. Another shortcoming is that the model fails to capture the dynamics that arise from flexible regions in macromolecular structures. These flexible regions are very important in structural biology, for example in understanding molecular interactions [36,39,52,53] and molecular recognition of epigenetic regulators of histone tails [17,31,58]. The large-scale dynamics of these regions makes imaging challenging [81], and thus sample preparation in cryo-EM generally seeks to minimize these dynamics by focusing on well-folded macromolecules frozen in vitreous ice [63]. However, this ‘may severely impact... the nature of the intrinsic dynamics and interactions displayed by macromolecules’ [63]. Although modern cryo-EM is making great strides in understanding flexible

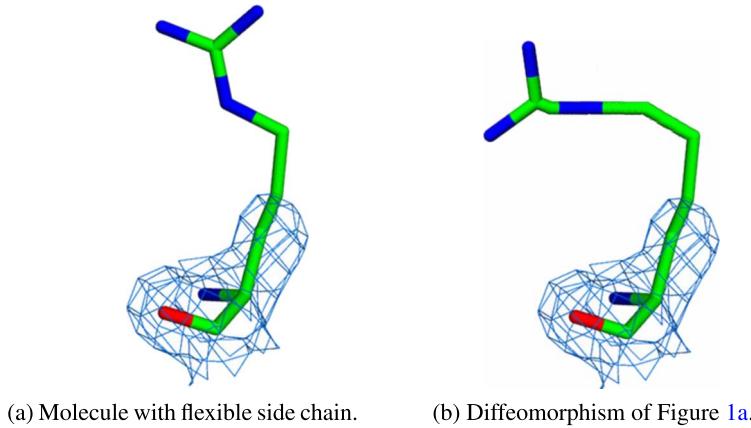


FIG. 1. Dynamics arising from flexible regions in macromolecular structures [63].

systems [3,37,38,59], formulating models that are more capable of capturing the motions associated with the flexible regions of macromolecules could open the door to applying cryo-EM more broadly, i.e. to less well-folded macromolecules. Mathematically, the motion of the flexible region can be modeled as a diffeomorphism. See Fig. 1, which shows a molecule with a flexible side chain (1(a)) and a diffeomorphism resulting from movement of the flexible region (1(b)). Figure 1(a) is taken from [63], and Fig. 1(b) was obtained by deforming it.

This article thus generalizes the classic MRA problem to include a random diffeomorphism. Specifically, we consider recovering a hidden signal  $f : \mathbb{R} \rightarrow \mathbb{R}$  from

$$y_j(x) = L_{\tau_j} f(x - t_j) + \varepsilon_j(x) \quad , \quad 1 \leq j \leq M,$$

where  $L_\tau$  is a dilation operator that dilates by a factor of  $(1 - \tau)$ . The dilation operator  $L_\tau$  is a simplified model for more general diffeomorphisms  $L_\zeta f(x) = f(\zeta(x))$ , since in the simplest case when  $\zeta(x)$  is affine,  $L_\zeta$  simply translates and dilates  $f$  (see Section 2.1). Dilations are also relevant for the analysis of time-warped audio signals, which can arise from the Doppler effect and in speech processing and bioacoustics. For example, [60–62] consider a stationary random signal  $f(x)$ , which is *time-warped*, i.e.  $D_\zeta f(x) = \sqrt{\zeta'(x)} f(\zeta(x))$ , and use a maximum likelihood approach to estimate  $\zeta$ . In [27,28], a similar stochastic time warping model is analyzed using wavelet based techniques. The noisy dilation MRA model considered here corresponds to the simplest case of time-warping, when  $\zeta$  is an affine function. This special case is in fact very important in imaging applications [22,23,46,57,69,80], where it is critical to compute features which are scale invariant, as objects are naturally dilated by the ‘zoom’ of an image.

A new approach is needed to solve this more general MRA problem, as Fourier invariants will fail, being unstable to the action of diffeomorphisms, including dilations. The instability occurs in the high frequencies, where even a small diffeomorphism can significantly alter the Fourier modes. We instead propose  $L^2(\mathbb{R})$  wavelet coefficient norms as invariants, using a continuous wavelet transform. This approach is inspired by the invariant scattering representation of [56], which is provably stable to the actions of small diffeomorphisms. However, here we replace local averages of the modulus of the wavelet coefficients with global averages (i.e. integrations) of the modulus squared, thus providing rigid invariants that can be statistically unbiased. Similar invariant coefficients have been utilized in

a number of applications including predicting molecular properties [34,35] and quantum chemical energies [45], and in microcanonical ensemble models for texture synthesis [19]. Recent work [42] has also generalized such coefficients to graphs.

### 1.1 Notation

The Fourier transform of a signal  $f \in \mathbf{L}^1(\mathbb{R})$  is

$$\widehat{f}(\omega) = \int f(x) e^{-ix\omega} dx.$$

We remind the reader that compactly supported  $\mathbf{L}^2(\mathbb{R})$  functions are in  $\mathbf{L}^1(\mathbb{R})$ . The power spectrum is the nonlinear transform  $P : \mathbf{L}^2(\mathbb{R}) \rightarrow \mathbf{L}^1(\mathbb{R})$  that maps  $f$  to

$$(Pf)(\omega) = |\widehat{f}(\omega)|^2, \quad \omega \in \mathbb{R}.$$

We denote  $f(x) \leq Cg(x)$  for some absolute constant  $C$  by  $f(x) \lesssim g(x)$ . We also write  $f(x) = O(g(x))$  if  $|f(x)| \leq Cg(x)$  for all  $x \geq x_0$  for some constants  $x_0, C > 0$ ;  $f(x) = o(g(x))$  denotes  $f(x)/g(x) \rightarrow 0$  as  $x \rightarrow \infty$ ;  $f(x) = \Theta(g(x))$  denotes  $C_1g(x) \leq |f(x)| \leq C_2g(x)$  for all  $x \geq x_0$  for some constants  $x_0, C_1, C_2 > 0$ . The minimum of  $a$  and  $b$  is denoted  $a \wedge b$ , and the maximum by  $a \vee b$ .

## 2. MRA models and the method of invariants

Standard MRA models are generalized to models that include deformations of the underlying signal in Section 2.1. Section 2.2 reviews power spectrum invariants and introduces  $\mathbf{L}^2(\mathbb{R})$  wavelet coefficient invariants. Theorem 2.4 proves wavelet coefficient invariants computed with a continuous wavelet transform and a suitable mother wavelet are equivalent to the power spectrum, showing there is no information loss in the transition from one representation to the other.

### 2.1 MRA data models

A standard MRA scenario considers the problem of recovering a signal  $f \in \mathbf{L}^2(\mathbb{R})$  in which one observes random translations of the signal, each of which is corrupted by additive noise. The problem is particularly difficult when the signal-to-noise ratio (SNR) is low, as registration methods become intractable. In [5,13,14,16,54,74] the authors propose a method using Fourier-based invariants, which are invariant to translations and thus eliminate the need to register signals.

A more general MRA scenario incorporates random deformations of the signal  $f$ , which could be used to model underlying physical variability that is not captured by rigid transformations and additive noise models. For example [4,7] consider a discrete signal  $f$  corrupted by an arbitrary group action, [47,85] consider random deformations arising in RADAR and [2] considers a generalization of MRA where signals are rescaled by random constants. Another natural mathematical model is small, random diffeomorphisms, which leads to observations of the form

$$y_j(x) = L_\xi f(x - t_j) + \varepsilon_j(x), \quad 1 \leq j \leq M, \quad (2.1)$$

where  $\zeta_j \in \mathbf{C}^1(\mathbb{R})$  is a random diffeomorphism,  $t_j \in \mathbb{R}$  is a random translation and the signals  $\varepsilon_j(x)$  are independent white noise random processes. The transform  $L_\zeta$  is the action of the diffeomorphism  $\zeta$  on  $f$ ,

$$L_\zeta f(x) = f(\zeta(x)).$$

If  $\|(\zeta^{-1})'\|_\infty < \infty$ , then one can verify  $L_\zeta : \mathbf{L}^2(\mathbb{R}) \rightarrow \mathbf{L}^2(\mathbb{R})$ .

One of the keys to the Fourier invariant approach of [5, 13, 14, 16, 54, 74] is the authors can unbias the Fourier invariants of the noisy signals, thus allowing them to devise an unbiased estimator of the Fourier invariants of the signal  $f$  (or a mixture of signals in the heterogeneous MRA case). For the diffeomorphism model (2.1) this would require developing a procedure for unbiasing the (Fourier) invariants of  $\{y_j\}_{j=1}^M$  against both additive noise and random diffeomorphisms.

In order to get a handle on the difficulties associated with the proposed diffeomorphism model, in this paper we consider random dilations of the signal  $f$ , which corresponds to restricting the diffeomorphism to be of the form

$$\zeta(x) = \frac{x}{1-\tau}, \quad |\tau| \leq 1/2.$$

Specifically, we assume the following noisy dilation MRA model.

**MODEL 2** (Noisy dilation MRA data model). The *noisy dilation MRA data model* consists of  $M$  independent observations of a compactly supported, real-valued signal  $f \in \mathbf{L}^2(\mathbb{R})$ :

$$y_j(x) = L_\tau f(x - t_j) + \varepsilon_j(x), \quad 1 \leq j \leq M, \quad (2.2)$$

where  $L_\tau$  is an  $\mathbf{L}^1(\mathbb{R})$  normalized dilation operator,

$$L_\tau f(x) = (1-\tau)^{-1} f\left((1-\tau)^{-1}x\right).$$

In addition, we assume the following:

- (i)  $\text{supp}(y_j) \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]$  for  $1 \leq j \leq M$ .
- (ii)  $\{t_j\}_{j=1}^M$  are independent samples of a random variable  $t \in \mathbb{R}$ .
- (iii)  $\{\tau_j\}_{j=1}^M$  are independent samples of a bounded, symmetric random variable  $\tau$  satisfying

$$\tau \in \mathbb{R}, \quad \mathbb{E}(\tau) = 0, \quad \text{Var}(\tau) = \eta^2, \quad |\tau| \leq 1/2.$$

- (iv)  $\{\varepsilon_j(x)\}_{j=1}^M$  are independent white noise processes on  $\left[-\frac{1}{2}, \frac{1}{2}\right]$  with variance  $\sigma^2$ .

**REMARK 2.1** The interval  $\left[-\frac{1}{2}, \frac{1}{2}\right]$  is arbitrary and can be replaced with any interval of length 1. In addition, the spatial box size is arbitrary, i.e.  $\left[-\frac{1}{2}, \frac{1}{2}\right]$ , can be replaced with  $\left[-\frac{N}{2}, \frac{N}{2}\right]$ . All results still hold with  $\sigma\sqrt{N}$  replacing  $\sigma$  wherever it appears.

Thus, the hidden signal  $f$  is supported on an interval of length 1, and we observe  $M$  independent instances of the signal that have been randomly translated, randomly dilated and corrupted by additive

white noise. We assume the hidden signal is real, but the proposed methods can also handle complex valued signals with minor modifications. Recall  $\varepsilon(x)$  is a white noise process if  $\varepsilon(x) = dB_x$ , i.e. it is the derivative of a Brownian motion with variance  $\sigma^2$ .

While the noisy dilation MRA model does not capture the full richness of the diffeomorphism model, it already presents significant mathematical difficulties. Indeed, as we show in Section 5, Fourier invariants, specifically the power spectrum, cannot be used to form accurate estimators under the action of dilations and random additive noise. The reason is that Fourier measurements are not stable to the action of small dilations (measured here by  $|\tau|$ ), since the displacement of  $\widehat{L_\tau f}(\omega)$  relative to  $\widehat{f}(\omega)$  depends on  $|\omega|$ . Intuitively, high-frequency modes are unstable, and yet high frequencies are often critical; for example removing high frequencies increases the sample complexity needed to distinguish between signals in a heterogeneous MRA model [5]. We thus replace Fourier-based invariants with wavelet coefficient invariants, which are defined in Section 2.2. As we show the wavelet invariants of the signal  $f$  can be accurately estimated from wavelet invariants of the noisy signals  $\{y_j\}_{j=1}^M$ , with no information loss relative to the power spectrum of  $f$ .

For future reference we also define the following dilation MRA model, which includes random translations and random dilations but no additive noise. Thus, Models 1 and 3 are both special cases of Model 2.

**MODEL 3** (Dilation MRA data model). The *dilation MRA data model* consists of  $M$  independent observations of a compactly supported, real-valued signal  $f \in \mathbf{L}^2(\mathbb{R})$ :

$$y_j(x) = L_{\tau_j} f(x - t_j), \quad 1 \leq j \leq M, \quad (2.3)$$

where  $L_\tau$  is an  $\mathbf{L}^1(\mathbb{R})$  normalized dilation operator,

$$L_\tau f(x) = (1 - \tau)^{-1} f\left((1 - \tau)^{-1} x\right).$$

In addition, we assume (i)–(iii) of Model 2.

## 2.2 Method of invariants

We now discuss how invariant representations can be used to solve MRA data models and introduce the wavelet invariants used in this article.

**2.2.1 Motivation and related work** Let  $T_t f(x) = f(x - t)$  denote the operator that translates by  $t$  acting on a signal  $f$ . Invariant measurement models seek a representation  $\Phi(f) \in \mathcal{B}$  in a Banach space  $\mathcal{B}$  such that

$$\Phi(T_t f) = \Phi(f), \quad \forall t \in \mathbb{R}. \quad (2.4)$$

In MRA problems, one additionally requires that

$$\Phi(f) = \Phi(g) \iff g = T_t f \text{ for some } t \in \mathbb{R}. \quad (2.5)$$

The first condition (2.4) removes the need to align random translations of the signal  $f$ , whereas the second condition (2.5) ensures that if one can estimate  $\Phi(f)$  from the collection  $\{\Phi(y_j)\}_{j=1}^M$ , then one

can recover an estimate of  $f$  (up to translation) by solving

$$f^* = \arg \inf_{g \in \mathbf{L}^1 \cap \mathbf{L}^2(\mathbb{R})} \|\Phi(g) - \Phi(f)\|_{\mathcal{B}}, \quad (2.6)$$

where  $\|\cdot\|_{\mathcal{B}}$  is the Banach space norm.

When the observed signals  $\{y_j\}_{j=1}^M$  are corrupted by more than just a random translation, though, as in Model 2, estimating  $\Phi(f)$  from  $\{\Phi(y_j)\}_{j=1}^M$  is not always straightforward. Indeed, one would like to compute

$$\bar{\Phi}_M(f) = \frac{1}{M} \sum_{j=1}^M \Phi(y_j), \quad (2.7)$$

but the quantity  $\bar{\Phi}_M(f)$  is not always an unbiased estimator of  $\Phi(f)$ , meaning that  $\lim_{M \rightarrow \infty} \bar{\Phi}_M(f) \neq \Phi(f)$ . In order to circumvent this issue, one must select a representation  $\Phi$  such that

$$\mathbb{E} \Phi(y_j) = \Phi(f) + b_{\Phi}(f, \mathcal{M}), \quad (2.8)$$

where  $b_{\Phi}(f, \mathcal{M})$  is a bias term depending on the choice of  $\Phi, f$ , and the signal corruption model  $\mathcal{M}$ . If (2.8) holds and if we can compute a  $\tilde{b}$  such that  $\mathbb{E} \tilde{b}_{\Phi}(y_j, \mathcal{M}) = b_{\Phi}(f, \mathcal{M}) + \delta$  for  $|b_{\Phi}(f, \mathcal{M})| \gg |\delta|$ , then one can amend (2.7) to reduce the bias

$$\tilde{\Phi}_M(f) = \frac{1}{M} \sum_{j=1}^M (\Phi(y_j) - \tilde{b}_{\Phi}(y_j, \mathcal{M})),$$

in which case

$$\lim_{M \rightarrow \infty} \tilde{\Phi}_M(f) = \Phi(f) + \delta$$

almost surely by the law of large numbers. The main difficulty therefore is twofold. On the one hand, one must design a representation  $\Phi$  that satisfies (2.4), (2.5) and (2.8) with a bias  $b$  that can be estimated; on the other hand, the optimization (2.6) must be tractable. For random translation plus additive noise models (i.e., Model 1), the authors of [5,14] describe a representation  $\Phi$  based on Fourier invariants that satisfies the outlined requirements and for which one can solve (2.6) despite the optimization being non-convex. The Fourier invariants include  $\widehat{f}(0)$  (i.e. the integral of  $f$ ), the power spectrum of  $f$  and the bispectrum of  $f$ . Each invariant captures successively more information in  $f$ . While  $\widehat{f}(0)$  carries limited information, the power spectrum recovers the magnitude of the Fourier transform, namely it recovers the non-negative, real-valued function  $\rho(\omega)$  such that  $\widehat{f}(\omega) = \rho(\omega) e^{i\theta(\omega)}$ , but the phase information  $\theta(\omega)$  is lost. Since  $\widehat{T}_t f(\omega) = e^{-i\omega t} \widehat{f}(\omega)$ , the power spectrum is invariant to translations as the Fourier modulus kills the phase factor induced by a translation  $t$  of  $f$ . However, it is in general not possible to recover a signal from its power spectrum, although in certain special cases the phase information can be resolved; results along these lines are in the field of phase retrieval [26,78]. The bispectrum is also translation invariant and invertible so long as  $\widehat{f}(\omega) \neq 0$  [66].

In Section 5 we show that it is impossible to significantly reduce the power spectrum bias for Model 2, which includes translations, dilations and additive noise. We thus propose replacing the power spectrum with the  $\mathbf{L}^2(\mathbb{R})$  norms of the wavelet coefficients of the signal  $f$ . These invariants satisfy

(2.4) and (2.8) for Model 2 and yield a convex formulation of (2.6). They do not satisfy (2.5) for general  $f \in \mathbf{L}^2(\mathbb{R})$ , but Theorem 2.4 in Section 2.2.2 shows that knowing the wavelet invariants of  $f$  is equivalent to knowing the power spectrum of  $f$ , which means that any phase retrieval setting in which recovery is possible will also be possible with the specified wavelet invariants. For example if the signal lives in a spline or shift invariant space in addition to being realvalued, then it can be recovered from its phaseless measurements [26,78].

**2.2.2 Wavelet invariants** We now define the wavelet invariants used in this article. A wavelet  $\psi \in \mathbf{L}^2(\mathbb{R})$  is a waveform that is localized in both space and frequency and has zero average,

$$\int \psi(x) dx = 0.$$

Note throughout this article  $\psi$  will always denote a wavelet in  $\mathbf{L}^1 \cap \mathbf{L}^2(\mathbb{R})$  with zero average, satisfying  $\|\psi\|_2 = 1$  as well as the classic admissability condition  $\int \frac{|\widehat{\psi}(\omega)|^2}{\omega} d\omega < \infty$ . A dilation of the wavelet by a factor  $\lambda \in (0, \infty)$  is denoted,

$$\psi_\lambda(x) = \lambda^{1/2} \psi(\lambda x),$$

where the normalization guarantees that  $\|\psi_\lambda\|_2 = \|\psi\|_2 = 1$ . The continuous wavelet transform  $W$  computes

$$Wf = \{f * \psi_\lambda(x) : \lambda \in (0, \infty), x \in \mathbb{R}\}.$$

The parameter  $\lambda$  corresponds to a frequency variable. Indeed, if  $\xi_0$  is the central frequency of  $\psi$ , the wavelet coefficients  $f * \psi_\lambda$  recover the frequencies of  $f$  in a band of size proportional to  $\lambda$  centered at  $\lambda \xi_0$ . Thus, high frequencies are grouped into larger packets, which we shall use to obtain a stable, invariant representation of  $f$ .

The wavelet transform  $Wf$  is equivariant to translations but not invariant. Integrating the wavelet coefficients over  $x$  yields translation invariant coefficients, but they are trivial since  $\int \psi_\lambda = 0$ . We therefore compute  $\mathbf{L}^2(\mathbb{R})$  norms in the  $x$  variable, yielding the following nonlinear wavelet invariants:

**DEFINITION 2.1** (Wavelet invariants). The  $\mathbf{L}^2$  wavelet invariants of a real-valued signal  $f \in \mathbf{L}^1 \cap \mathbf{L}^2(\mathbb{R})$  are given by

$$(Sf)(\lambda) = \|f * \psi_\lambda\|_2^2, \quad \lambda \in (0, \infty), \tag{2.9}$$

where  $\psi_\lambda(x) = \lambda^{1/2} \psi(\lambda x)$  are dilations of a mother wavelet  $\psi$ .

Throughout this article  $\psi$  can be taken as a Morlet wavelet, in which case  $\psi$  is constructed to have frequency centered at  $\xi$  by  $\psi(x) = C_\xi \pi^{-1/4} e^{-x^2/2} (e^{i\xi x} - e^{-\xi^2/2})$  for  $C_\xi = (1 - e^{-\xi^2} - 2e^{-3\xi^2/4})^{-1/2}$ , but results hold more generally for what we refer to as  $k$ -admissible wavelets, where  $k \geq 0$  is an even integer. See Appendix A for a precise description of this admissibility criteria. The wavelet invariants can be expressed in the frequency domain as

$$(Sf)(\lambda) = \frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 |\widehat{\psi}_\lambda(\omega)|^2 d\omega,$$

which motivates the following definition of ‘wavelet invariant derivatives’.

DEFINITION 2.2 (Wavelet invariant derivatives). The  $n$ -th derivative of  $(Sf)(\lambda)$  is defined as

$$(Sf)^{(n)}(\lambda) := \frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \frac{d^n}{d\lambda^n} |\widehat{\psi}_\lambda(\omega)|^2 d\omega.$$

REMARK 2.2 Definition 2.1 assumes  $f: \mathbb{R} \rightarrow \mathbb{R}$ , which allows the wavelet  $\psi$  to be either real or complex. Our results can easily be extended to complex  $f$ , but a strictly complex wavelet would be needed, with  $Sf(\lambda)$  computed for all  $\lambda \in (-\infty, \infty) \setminus 0$ .

REMARK 2.3 For a discrete signal of length  $n$ , computing the wavelet invariants via a continuous wavelet transform is  $O(n^2)$ , while computing the power spectrum is  $O(n \log n)$ . Thus, one pays a computational cost to achieve greater stability with no loss of information. On the other hand, if wavelet invariants are computed for a dyadic wavelet transform (i.e. only for  $O(\log n)$   $\lambda$ 's), the computational cost is the same and stability is maintained, but more information is lost.

REMARK 2.4 When  $(Pf)(\omega) = |\widehat{f}(\omega)|^2$  is continuous, Definition 2.2 reduces to a normal derivative, i.e. one can check that  $(Sf)^{(n)}(\lambda) = \frac{d^n}{d\lambda^n} (Sf)(\lambda)$ . However, when  $Pf$  is not continuous, in general  $(Sf)^{(n)}(\lambda) \neq \frac{d^n}{d\lambda^n} (Sf)(\lambda)$ , and  $(Sf)^{(n)}(\lambda)$  is more convenient for controlling the error of the estimators proposed in this article. Throughout this article, the notation  $(Sf)^{(n)}(\lambda)$  will thus denote the derivative of Definition 2.2 and  $\frac{d^n}{d\lambda^n} (Sf)(\lambda)$  will denote the standard derivative.

Under mild conditions, one can show that  $S: \mathbf{L}^2(\mathbb{R}) \rightarrow \mathbf{L}^1 \cap \mathbf{C}(0, \infty)$ . The values  $\lambda = 2^j$  for  $j \in \mathbb{Z}$  correspond to rigid versions of first-order  $\mathbf{L}^2(\mathbb{R})$  wavelet scattering invariants [56]. The continuous wavelet transform  $Wf$  is extremely redundant; indeed, for suitably chosen mother wavelets, the dyadic wavelet transform with  $\lambda = 2^j$  for  $j \in \mathbb{Z}$  is a complete representation of  $f$ . However, the corresponding operator  $S$  restricted to  $\lambda = 2^j$  is not invertible. When one utilizes every frequency  $\lambda \in (0, \infty)$ , though, the resulting  $\mathbf{L}^2(\mathbb{R})$  norms  $(Sf)(\lambda) = \|f * \psi_\lambda\|_2^2$  uniquely determine the power spectrum of  $f$ , so long as the wavelet  $\psi$  satisfies a type of independence condition.

CONDITION 2.3 Define

$$|\widehat{\psi}_\lambda^+(\omega)|^2 = \left( |\widehat{\psi}_\lambda(\omega)|^2 + |\widehat{\psi}_\lambda(-\omega)|^2 \right) \cdot \mathbf{1}(\omega \geq 0).$$

If for any finite sequence  $\{\omega_i\}_{i=1}^n$  of distinct positive frequencies, the collection  $\{|\widehat{\psi}_\lambda^+(\omega_i)|^2\}_{i=1}^n$  is linearly independent functions of  $\lambda$ , we say the wavelet  $\psi$  satisfies the linear independence condition.

REMARK 2.5 Condition 2.3 is stated in terms of  $|\widehat{\psi}_\lambda^+(\omega)|^2$  to avoid assumptions on whether  $\psi$  is real or complex. When  $\psi(x) \in \mathbb{R}$ ,  $|\widehat{\psi}_\lambda^+(\omega)|^2 = 2|\widehat{\psi}_\lambda(\omega)|^2$  for  $\omega \geq 0$ . When  $\psi$  is complex analytic,  $|\widehat{\psi}_\lambda^+(\omega)|^2 = |\widehat{\psi}_\lambda(\omega)|^2$ . When  $\psi \in \mathbb{C}$  but not complex analytic,  $|\widehat{\psi}_\lambda^+(\omega)|^2$  simply incorporates a reflection of  $|\widehat{\psi}_\lambda(\omega)|^2$  about the origin. Since we assume  $f(x) \in \mathbb{R}$ ,  $|\widehat{\psi}_\lambda^+(\omega)|^2$  uniquely defines  $(Sf)(\lambda)$ , since  $(Sf)(\lambda) = \frac{1}{2\pi} \langle |\widehat{f}|^2, |\widehat{\psi}_\lambda^+|^2 \rangle$  by the Plancherel and Fourier convolution theorems.

THEOREM 2.4 Let  $f, g \in \mathbf{L}^1 \cap \mathbf{L}^2(\mathbb{R})$  and assume  $\psi$  satisfies Condition 2.3 and  $\widehat{\psi}$  has compact support. Then,

$$Sf = Sg \iff Pf = Pg.$$

*Proof.* First assume  $Pf = Pg$ , which means  $|\widehat{f}(\omega)|^2 = |\widehat{g}(\omega)|^2$  for almost every  $\omega \in \mathbb{R}$ . Using the Plancheral and Fourier convolution theorems,

$$\begin{aligned} (Sf)(\lambda) &= \int |f * \psi_\lambda(x)|^2 dx = \frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 |\widehat{\psi}_\lambda(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \int |\widehat{g}(\omega)|^2 |\widehat{\psi}_\lambda(\omega)|^2 d\omega = (Sg)(\lambda), \quad \forall \lambda \in (0, \infty). \end{aligned}$$

Now suppose  $Sf = Sg$ . Since  $Sf$  and  $Sg$  are continuous in  $\lambda$ , we have

$$0 = (Sf)(\lambda) - (Sg)(\lambda) = \frac{1}{2\pi} \int (|\widehat{f}(\omega)|^2 - |\widehat{g}(\omega)|^2) |\widehat{\psi}_\lambda(\omega)|^2 d\omega, \quad \forall \lambda \in (0, \infty).$$

Since  $f \in \mathbf{L}^1 \cap \mathbf{L}^2(\mathbb{R})$  we have  $\widehat{f} \in \mathbf{L}^2 \cap \mathbf{L}^\infty(\mathbb{R})$  and thus  $Pf \in \mathbf{L}^1 \cap \mathbf{L}^\infty(\mathbb{R})$ . By interpolation we have  $Pf \in \mathbf{L}^2(\mathbb{R})$ , and the same for  $Pg$ . By applying Lemma 2.1 (stated below) with  $p(\omega) = (Pf)(\omega) - (Pg)(\omega)$  (note  $p$  is continuous since  $f, g \in \mathbf{L}^1(\mathbb{R})$ ), we conclude  $Pf = Pg$  for almost every  $\omega$ .  $\square$

LEMMA 2.1 Let  $p \in \mathbf{L}^2(\mathbb{R})$  be continuous and assume  $p(\omega) = p(-\omega)$ ,  $\widehat{\psi}$  has compact support and Condition 2.3. Then,

$$\int p(\omega) |\widehat{\psi}_\lambda(\omega)|^2 d\omega = 0 \quad \forall \lambda > 0 \implies p = 0 \text{ a.e.}$$

The proof of Lemma 2.1 is in Appendix C. We remark that many wavelets satisfy Condition 2.3 and have compactly supported Fourier transform, so Theorem 2.4 is broadly applicable. For example, Proposition 2.1 below proves that any complex analytic wavelet with compactly supported Fourier transform satisfies Condition 2.3. Morlet wavelets satisfy Condition 2.3 (see Lemma C.1 in Appendix C) but do not have compactly supported Fourier transform; however,  $\widehat{\psi}$  does have fast decay for a Morlet wavelet and numerically we observe no issues. We also note, the assumption that  $\widehat{\psi}$  has compact support in Theorem 2.4 can be removed if  $f, g$  are bandlimited. The following Proposition, proved in Appendix C, gives some sufficient conditions guaranteeing Condition 2.3.

PROPOSITION 2.1 The following are sufficient to guarantee Condition 2.3:

- (i)  $|\widehat{\psi}(\omega)|^2$  has a compact support contained in the interval  $[a, b]$ , where  $a$  and  $b$  have the same sign, e.g. complex analytic wavelets with compactly supported Fourier transform.
- (ii)  $|\widehat{\psi}(\omega)|^2 \in \mathbf{C}^\infty(\mathbb{R})$  and there exists an  $N$  such that all derivatives of order at least  $N$  are non-zero at  $\omega = 0$ , e.g. the Morlet wavelet.

REMARK 2.6 In practice,  $Pf, Sf$  are implemented as discrete vectors, and  $Sf$  is obtained from  $Pf$  via matrix multiplication, i.e.  $Sf = F(Pf)$  for some real matrix  $F$  with  $F^T F$  strictly positive definite. Thus,  $\|Pf - Pg\|_2 \leq \sigma_{\min}^{-1} \|Sf - Sg\|_2$ , where  $\sigma_{\min} > 0$  is the smallest singular value of the matrix  $F$ , and the spectral decay of  $F$ , which can be explicitly computed, thus determines the stability of the representation. The smoother the wavelet, the more rapidly the spectrum decays, since when  $P\psi \in C^p$ ,  $F^T F$  is defined by a  $C^p$  kernel and thus has eigenvalues that decay like  $o(1/n^{p+1})$  [20]. There is thus a tradeoff between smoothness and stability. In this article we choose smoothness over stability, since smoothness is required for unbiasing noisy dilation MRA, and in our experiments the Morlet wavelet

yielded the best results. We therefore invert the representation by solving an optimization problem that is initialized to be close to the desired solution (see Section 6.5), and we avoid computing the pseudo-inverse of  $F$ , which is unstable for our smooth wavelet.

### 3. Unbiasing for classic MRA

In this section we consider the classic MRA model (Model 1). We discuss unbiasing results for both the power spectrum and wavelet invariants, as well as simulation results comparing the two methods. In the following proposition we establish unbiasing results for the power spectrum by rederiving some results from [14], extended to the continuum setting. The proposition is proved in Appendix D.

**PROPOSITION 3.1** Assume Model 1. Define the following estimator of  $(Pf)(\omega)$ :

$$(\tilde{P}f)(\omega) := \frac{1}{M} \sum_{j=1}^M (Py_j)(\omega) - \sigma^2.$$

Then with probability at least  $1 - 1/t^2$ ,

$$|(Pf)(\omega) - (\tilde{P}f)(\omega)| \leq \frac{2t\sigma}{\sqrt{M}} (\|f\|_1 + \sigma). \quad (3.1)$$

We obtain an identical result for wavelet invariants (Proposition 3.2) when signals are corrupted by additive noise only. See Appendix D for the proof.

**PROPOSITION 3.2** Assume Model 1. Define the following estimator of  $(Sf)(\lambda)$ :

$$(\tilde{S}f)(\lambda) := \frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda) - \sigma^2.$$

Then with probability at least  $1 - 1/t^2$ ,

$$|(Sf)(\lambda) - (\tilde{S}f)(\lambda)| \leq \frac{2t\sigma}{\sqrt{M}} (\|f\|_1 + \sigma). \quad (3.2)$$

As  $M \rightarrow \infty$ , the error of both the power spectrum and wavelet invariant estimators decays to zero at the same rate, and one can perfectly unbias both representations. As demonstrated in Section 5, this is not possible for noisy dilation MRA (Model 2), as there is a non-vanishing bias term. However, a nonlinear unbiasing procedure on the wavelet invariants can significantly reduce the bias.

We illustrate and compare additive noise unbiasing for power spectrum estimation using  $(\tilde{P}f)$ , the power spectrum method of Proposition 3.1 and  $(\tilde{S}f)$ , the wavelet invariant method of Proposition 3.2. To approximate  $(Pf)$  from the wavelet invariants  $(\tilde{S}f)$ , we apply the convex optimization algorithm described in Section 6.5 to obtain  $(\tilde{P}Sf)$ , the power spectrum approximation that best matches the wavelet invariants  $(\tilde{S}f)$ . Thus, throughout this article,  $(\tilde{P}Sf)$  denotes a power spectrum estimator obtained by first unbiasing wavelet invariants and then running an optimization procedure, while  $(\tilde{P}f)$  denotes an

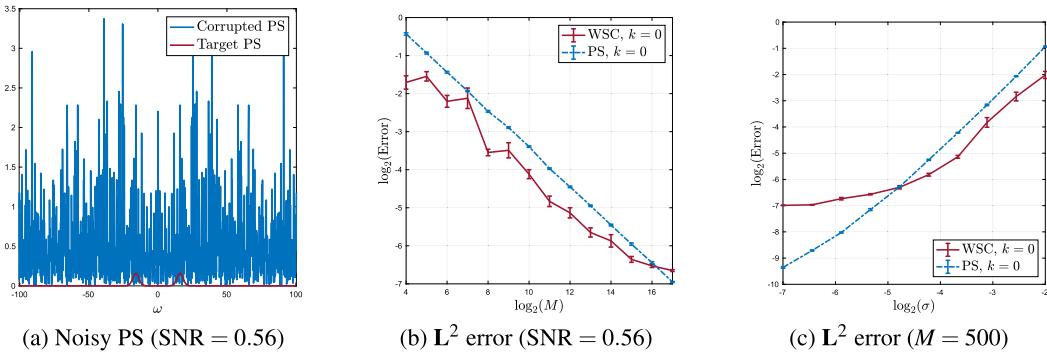


FIG. 2. Simulation results for additive noise model for medium frequency Gabor  $f(x) = e^{-5x^2} \cos(16x)$ .

estimator computed by directly unbiasing the power spectrum. Our simulations compare the  $\mathbf{L}^2$  error of both of these estimators, i.e. we compare  $\|Pf - \tilde{P}f\|_2$  and  $\|Pf - \tilde{P}_{\text{SF}}f\|_2$ .

Figure 2(a) shows the uncorrupted power spectrum (red curve) of a medium frequency Gabor function ( $f(x) = e^{-5x^2} \cos(16x)$ ), and the power spectrum after the signal is corrupted by additive noise with level  $\sigma = 2^{-3}$  (blue curve); the SNR of the experiment is 0.56 (see Section 6.1). Figure 2(b) shows the  $\mathbf{L}^2$  error of the power spectrum estimation for the two methods as a function of  $\log_2(M)$  for a fixed SNR, and Fig. 2(c) shows the  $\mathbf{L}^2$  error as a function of  $\log_2(\sigma)$  for a fixed  $M$ . The  $\mathbf{L}^2$  errors for the two methods are similar; however, estimation via wavelet invariants is advantageous when the sample size  $M$  is small or the additive noise level  $\sigma$  is large. As  $M$  becomes very large or  $\sigma$  very small, the power spectrum method is preferable as the smoothing procedure of the wavelet invariants may numerically erase some extremely small scale features of the original power spectrum.

#### 4. Unbiasing for dilation MRA

In this section we analyze the dilation MRA model (Model 3). We thus assume the signals have been randomly translated and dilated but there is no additive noise.

In fact there is a simple algorithm to recover  $f$  under this model. Since  $\|f_{\tau_j}\|_2^2 = \|f\|_2^2/(1 - \tau_j)$ ,  $\frac{1}{M} \sum_{j=1}^M 1/\|f_{\tau_j}\|_2^2$  is an unbiased estimator of  $1/\|f\|_2^2$ , and so  $\|f\|_2^2$  can be accurately approximated. Once  $\|f\|_2^2$  is recovered, one can take any signal  $y_j$  and dilate it so that  $\|y_j\|_2^2 = \|f\|_2^2$ , and the result will be an accurate approximation of the hidden signal  $f$  for  $M$  large. However, this approach collapses in the presence of even a small amount of additive noise. In the presence of additive noise, an alternative is to attempt a synchronization by centering each signal. The center  $c_f$  of signal  $f$  can be defined in the classical way by

$$c_f = \frac{1}{\|f\|_2^2} \int x |f(x)|^2 \, dx.$$

Since the signals  $y_j(x - (c_f + t_j))$  are perfectly aligned, one can thus attempt an alignment by defining  $\tilde{y}_j(x) = y_j(x - c_{y_j})$ . However  $c_{y_j} - (c_f + t_j) = O(\sigma \vee \sigma^2 + \eta)$ , so significant errors arise in the synchronization that cannot be resolved by averaging. As our goal is ultimately to produce a method

that can be extended to the noisy dilation MRA model, we abandon both the trivial solution (which cannot be extended to noisy dilation MRA) and the synchronization approach (which produces large errors) and explore a method based on empirical averages.

We first observe that random dilations cause  $\frac{1}{M} \sum_{j=1}^M (Py_j)(\omega)$  and  $\frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda)$  to be biased estimators of  $(Pf)(\omega)$  and  $(Sf)(\lambda)$ , and the bias for both is  $O(\eta^2)$ , where  $\eta^2$  is the variance of the dilation distribution. However, if the moments of the dilation distribution are known and  $Pf, Sf$  are sufficiently smooth, one can apply an unbiasing procedure to the above estimators so that the resulting bias is  $O(\eta^{k+2})$ , where  $k \geq 2$  is an even integer.

Throughout this section, we assume  $k \geq 2$  is an even integer, and define the constants  $C_i$  from the first  $k/2$  even moments of  $\tau$  by  $\mathbb{E}[\tau^i] = C_i \eta^i$  for  $i = 2, 4, \dots, k$ . Note since we assume  $\mathbb{E}[\tau^2] = \eta^2$ ,  $C_2 = 1$ . We define the constants  $B_2, B_4, \dots, B_k$  by solving

$$\frac{C_i}{i!} - \frac{B_2 C_{i-2}}{(i-2)!} - \dots - \frac{B_{i-2} C_2}{2!} - B_i = 0 \quad (4.1)$$

for  $i = 2, 4, \dots, k$ ; these constants are deterministic functions of the moments of  $\tau$ . A non-recursive formula related to the Euler numbers can be derived, which defines  $B_i$  explicitly in terms of  $C_2, \dots, C_i$ ; however, the recursive formula (4.1) is easier to implement numerically.

We introduce two additional moment-based constants that are defined by the  $C_i, B_i$  constants:

$$T := \max_{i=0,2,\dots} C_i^{\frac{1}{i}} \quad (4.2)$$

$$E := \max_{i=0,2,\dots,k} \max_{j=0,\dots,k+2-i} \left( \frac{T^j}{j!} |B_i| \right)^{\frac{1}{i+j}}, \quad (4.3)$$

where  $C_0, |B_0| = 1$ , and when  $i = j = 0$  in (4.3),  $\left( \frac{T^j}{j!} |B_i| \right)^{\frac{1}{i+j}}$  is replaced with 1.

**REMARK 4.1** Since the distribution of  $\tau$  is bounded, we are guaranteed that  $T < \infty$ , and in general can consider both  $T$  and  $E$  to be  $O(1)$  constants. For example for the uniform distribution,  $T \leq \sqrt{3}$  and  $|B_i| \leq \frac{|\text{Euler}(i)|}{i!} \leq 1$ , which gives  $E \leq \sqrt{3}$ .

We utilize the following two lemmas, which are proved in Appendix E, to derive results for both the power spectrum and wavelet invariants.

**LEMMA 4.1** Let  $F_\lambda(\tau) = L((1-\tau)\lambda)$  for some function  $L \in \mathbf{C}^{k+2}(0, \infty)$  and a random variable  $\tau$  satisfying the assumptions of Section 2.1, and let  $k \geq 2$  be an even integer. Assume there exist functions  $\Lambda_i : \mathbb{R} \rightarrow \mathbb{R}, R : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$|\lambda^i L^{(i)}(\lambda)| \leq \Lambda_i(\lambda) \text{ for } 0 \leq i \leq k+2, \quad \frac{\Lambda_{k+2}((1-\tau)\lambda)}{\Lambda_{k+2}(\lambda)} \leq R(\lambda),$$

and define the following estimator of  $L(\lambda)$ :

$$G_\lambda(\tau) := F_\lambda(\tau) - B_2 \eta^2 F_\lambda''(\tau) - B_4 \eta^4 F_\lambda^{(4)}(\tau) - \dots - B_k \eta^k F_\lambda^{(k)}(\tau).$$

Then  $G_\lambda(\tau)$  satisfies

$$|\mathbb{E} G_\lambda(\tau) - L(\lambda)| \lesssim kR(\lambda)\Lambda_{k+2}(\lambda)(2E\eta)^{k+2}$$

$$\text{Var } G_\lambda(\tau) \lesssim k^2 R(\lambda)^2 \Lambda(\lambda)^2$$

where

$$\Lambda(\lambda)^2 := \sum_{0 \leq i, j \leq k+2, i+j \geq 2} \Lambda_i(\lambda) \Lambda_j(\lambda) (2E\eta)^{i+j}$$

and  $E$  is the absolute constant defined in (4.3).

LEMMA 4.2 Let the assumptions and notation of Lemma 4.1 hold, and let  $\tau_1, \dots, \tau_M$  be independent. Define

$$\tilde{L}(\lambda) := \frac{1}{M} \sum_{j=1}^M G_\lambda(\tau_j).$$

Then with probability at least  $1 - 1/t^2$

$$|\tilde{L}(\lambda) - L(\lambda)| \lesssim kR(\lambda) \left( \Lambda_{k+2}(\lambda)(2E\eta)^{k+2} + \frac{t\Lambda(\lambda)}{\sqrt{M}} \right).$$

The deviation of the estimator  $\tilde{L}(\lambda)$  from  $L(\lambda)$  thus depends on two things: (1) the bias of the estimator that is  $O(\eta^{k+2})$  and (2) the standard deviation of the estimator that is  $O(\eta M^{-\frac{1}{2}})$ , since  $\Lambda(\lambda) = O(\eta)$ .

#### 4.1 Power spectrum results for dilation MRA

We now show how this unbiasing procedure based on both the moments of  $\tau$  and the even derivatives of  $Py$  can be used to obtain an estimator of  $Pf$ .

PROPOSITION 4.1 Assume Model 3 and  $Pf \in \mathbf{C}^{k+2}(\mathbb{R})$ . Define the following estimator of  $(Pf)(\omega)$ :

$$(\tilde{P}f)(\omega) := \frac{1}{M} \sum_{j=1}^M \left[ (Py_j)(\omega) - B_2 \eta^2 \omega^2 (Py_j)''(\omega) - \dots - B_k \eta^k \lambda^k (Py_j)^{(k)}(\omega) \right]$$

where the constants  $B_i$  satisfy (4.1). Let

$$\Omega_i(\omega) = |\omega^i (Pf)^i(\omega)| \text{ for } 0 \leq i \leq k+2 \quad , \quad R(\omega) = \max_{\tau} \frac{\Omega_{k+2}((1-\tau)\omega)}{\Omega_{k+2}(\omega)}.$$

Then for all  $\omega \neq 0$ , with probability at least  $1 - 1/t^2$ ,

$$|(\tilde{P}f)(\omega) - (Pf)(\omega)| \lesssim kR(\omega) \left( \Omega_{k+2}(\omega)(2E\eta)^{k+2} + \frac{t\Omega(\omega)}{\sqrt{M}} \right), \quad (4.4)$$

where

$$\Omega(\omega) = \sum_{0 \leq i, j \leq k+2, i+j \geq 2} \Omega_i(\omega) \Omega_j(\omega) (2E\eta)^{i+j}.$$

*Proof.* Since  $Pf$  is a translation invariant representation, we can ignore the translation factors  $\{t_k\}_{k=1}^M$  and consider the model  $y_j = L_{\tau_j}f$ . In addition since  $y_j(x) \in \mathbb{R}$ ,  $(Py_j)(\omega) = (Py_j)(-\omega)$  and it is sufficient to consider  $\omega \in (0, \infty)$ . Proposition 4.1 then follows directly from Lemma 4.2 with  $\lambda = \omega$ ,  $L = Pf$  since  $(Py_j)(\omega) = (Pf)((1 - \tau_j)\omega) = F_\omega(\tau_j)$ ,  $\Lambda_i = \Omega_i$ , and  $\Lambda = \Omega$ .  $\square$

We postpone a discussion of the shortcomings of Proposition 4.1 to Section 4.3, where we compare the power spectrum and wavelet invariant results for dilation MRA.

#### 4.2 Wavelet invariant results for dilation MRA

We now apply the same unbiasing procedure to the wavelet invariants. Unlike for the power spectrum, where the error may depend on the frequency  $\omega$  (see (4.4) and Section 4.3), the wavelet invariant error can be uniformly bounded independently of  $\lambda$  with high probability. The following two Lemmas establish bounds on the derivatives of  $(Sf)(\lambda)$  and are needed to prove Proposition 4.2; they are proved in Appendix B.

LEMMA 4.3 (Low-frequency bound). Assume  $P\psi \in \mathbf{C}^m(\mathbb{R})$  and  $f \in \mathbf{L}^1(\mathbb{R})$ . Then the quantity  $|\lambda^m(Sf)^{(m)}(\lambda)|$  can be bounded uniformly over all  $\lambda$ . Specifically:

$$|\lambda^m(Sf)^{(m)}(\lambda)| \leq \Psi_m \|f\|_1^2$$

for  $\Psi_m$  defined in (A1).

LEMMA 4.4 (High-frequency bound for differentiable functions). Assume  $P\psi \in \mathbf{C}^m(\mathbb{R})$ , and  $f' \in \mathbf{L}^1(\mathbb{R})$ . Then the quantity  $|\lambda^m(Sf)^{(m)}(\lambda)|$  can be bounded by

$$|\lambda^m(Sf)^{(m)}(\lambda)| \leq \frac{\Theta_m}{\lambda^2} \|f'\|_1^2$$

for  $\Theta_m$  defined in (A2).

When  $\psi$  is a Morlet wavelet or more generally when  $\psi$  is  $(k+2)$ -admissible as described in Appendix A, these lemmas allow one to bound the error of the order  $k$  wavelet invariant estimator for dilation MRA in terms of the following quantities:

$$\Lambda_i(\lambda) = \Psi_i \|f\|_1^2 \wedge \frac{\Theta_i}{\lambda^2} \|f'\|_1^2, \quad \Lambda(\lambda)^2 = \sum_{0 \leq i, j \leq k+2, i+j \geq 2} \Lambda_i(\lambda) \Lambda_j(\lambda) (2E\eta)^{i+j}, \quad (4.5)$$

where  $\Psi_i, \Theta_i$  are defined in (A1), (A2) and  $E$  is defined in (4.3).

**PROPOSITION 4.2** Assume Model 3, the notation in (4.5), and that  $\psi$  is  $(k+2)$ -admissible. Define the following estimator of  $(Sf)(\lambda)$ :

$$(\tilde{S}f)(\lambda) := \frac{1}{M} \sum_{j=1}^M \left[ (Sy_j)(\lambda) - B_2 \eta^2 \lambda^2 (Sy_j)''(\lambda) - \dots - B_k \eta^k \lambda^k (Sy_j)^{(k)}(\lambda) \right]$$

where the constants  $B_i$  satisfy (4.1). Then with probability at least  $1 - 1/t^2$ ,

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \lesssim k \left( \Lambda_{k+2}(\lambda) (2E\eta)^{k+2} + \frac{t\Lambda(\lambda)}{\sqrt{M}} \right).$$

*Proof.* Since  $Sf$  is a translation invariant representation, we can ignore the translation factors  $\{t_k\}_{k=1}^M$  and consider the model  $y_j = L_{\tau_j} f$ . Since  $\psi$  is  $k+2$ -admissible,  $\widehat{\psi} \in \mathbf{C}^{k+2}(\mathbb{R})$ , which guarantees  $(Sf)(\lambda) \in \mathbf{C}^{k+2}(0, \infty)$ . We note that since  $f \in \mathbf{L}^1(\mathbb{R})$ ,  $Pf$  is continuous, and the Leibniz integral rule guarantees that  $(Sf)^{(n)}(\lambda) = \frac{d^n}{d\lambda^n} (Sf)(\lambda)$  for  $1 \leq n \leq k+2$ . By applying Lemma 4.3, we have  $|\lambda^i (Sf)^{(i)}(\lambda)| \leq \Psi_i \|f\|_1^2$  for all  $0 \leq i \leq k+2$ , so that Lemma 4.2 holds for  $L(\lambda) = (Sf)(\lambda)$ ,  $\Lambda_i(\lambda) = \Psi_i \|f\|_1^2$ , and  $R(\lambda) = 1$ . Now by applying Lemma 4.4, we have  $|\lambda^i (Sf)^{(i)}(\lambda)| \leq \Theta_i \|f'\|_1^2 / \lambda^2$  for all  $0 \leq i \leq k+2$ , so that Lemma 4.2 also holds for  $L(\lambda) = (Sf)(\lambda)$ ,  $\Lambda_i(\lambda) = \Theta_i \|f'\|_1^2 / \lambda^2$ , and  $R(\lambda) = 4$  (note since  $|\tau| \leq \frac{1}{2}$ ,  $\Lambda_{k+2}((1-\tau)\lambda) / \Lambda_{k+2}(\lambda) \leq 4$ ). Thus, Lemma 4.2 in fact holds with  $\Lambda_i(\lambda) = \left( \Psi_i \|f\|_1^2 \wedge \frac{\Theta_i}{\lambda^2} \|f'\|_1^2 \right)$ ; since  $(Sy_j)(\lambda) = (Sf)((1-\tau_j)\lambda) = F_\lambda(\tau_j)$ , we obtain Proposition 4.2.  $\square$

Since  $\Lambda_i(\lambda) \leq \Psi_i \|f\|_1^2$ , Proposition 4.2 guarantees that the error can be uniformly bounded independent of  $\lambda$ . In addition, if the signal is smooth, the error for high-frequency  $\lambda$  will have the favorable scaling  $\lambda^{-2}$ . An important question in practice is how to choose  $k$ , i.e. what order wavelet invariant estimator minimizes the bias. Consider for example when  $f' \notin \mathbf{L}^1(\mathbb{R})$ , and  $\Lambda_{k+2}(\lambda) = \Psi_{k+2} \|f\|_1^2$ . By using a second-order estimator, we can decrease the bias from  $O(\eta^2)$  to  $O(\eta^4)$ , and we can further decrease the bias to  $O(\eta^6)$  by choosing  $k = 4$ . However,  $\Psi_k$  increases very rapidly in  $k$ . Indeed, as can be seen from (A1),  $\Psi_k$  increases like  $k!$ . Thus, one possible heuristic (assuming  $\eta$  is known) is to choose  $k = \tilde{k}$  where  $\tilde{k}$  minimizes the bias upper bound  $k\Psi_{k+2}(2E\eta)^{k+2}$ . Since  $\Psi_k$  increases factorially,  $\Psi_k \sim (Ck)^k$  for some constant  $C$ , and  $\tilde{k} + 2$  will be inversely proportional to  $\eta$ , that is  $(\tilde{k} + 2) \sim \eta^{-1}$ . The following corollary of Proposition 4.2 then holds for any  $k \leq \tilde{k}$ .

**COROLLARY 4.1** Under the assumptions of Proposition 4.2, if  $\Psi_i(2E\eta)^i$  is decreasing for  $i \leq k+2$ , then with probability at least  $1 - 1/t^2$ :

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \lesssim \|f\|_1^2 \left( k\Psi_{k+2}(2E\eta)^{k+2} + \frac{tk^2\eta}{\sqrt{M}} \right). \quad (4.6)$$

Similarly, if  $\Theta_i(2E\eta)^i$  is decreasing for  $i \leq k+2$ , then with probability at least  $1 - 1/t^2$ :

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \lesssim \frac{\|f'\|_1^2}{\lambda^2} \left( k\Theta_{k+2}(2E\eta)^{k+2} + \frac{tk^2\eta}{\sqrt{M}} \right). \quad (4.7)$$

REMARK 4.2 We observe that for a discrete lattice  $I$  of  $\lambda$  values, we can define the discrete 1-norm by  $\|g\|_{\mathbf{L}^1(I)} = \sum_{\lambda \in I} |g(\lambda)| \Delta\lambda$ . Assume the lattice has cardinality  $n$ , and that  $\Psi_i(2E\eta)^i, \Theta_i(2E\eta)^i$  are decreasing for  $i \leq k+2$ . Applying Proposition 4.2 with  $t = \sqrt{ns}$  and a union bound over the lattice gives

$$\|\tilde{S}f - Sf\|_{\mathbf{L}^1(I)} \lesssim k \left( \|f\|_1^2 \Psi_{k+2} + \|f'\|_1^2 \Theta_{k+2} \right) (2E\eta)^{k+2} + \frac{s\sqrt{nk^2\eta}}{\sqrt{M}} \left( \|f\|_1^2 + \|f'\|_1^2 \right)$$

with probability at least  $1 - 1/s^2$ . When  $n \ll M$ , which is the context for MRA, the 1-norm of the error is  $O(\eta^{k+2})$  as  $M \rightarrow \infty$ .

### 4.3 Comparison

Although Propositions 4.2 and 4.1 at first glance appear quite similar, the wavelet invariant method has several important advantages over the power spectrum method, which we enumerate in the following remarks.

REMARK 4.3 Proposition 4.2 (wavelet invariants) applies to any signal satisfying  $f \in \mathbf{L}^1(\mathbb{R})$ , but Proposition 4.1 requires  $Pf \in \mathbf{C}^{k+2}(\mathbb{R})$ . Thus, as  $k$  is increased, the power spectrum results apply to an increasingly restrictive function class. Furthermore, as discussed in Section 5, if the signal contains any additive noise,  $P\psi$  is not even  $\mathbf{C}^1$ , which means the unbiasing procedure of Proposition 4.1 cannot be applied. On the other hand, by choosing  $P\psi \in \mathbf{C}^\infty(\mathbb{R})$ ,  $Sf$  will inherit the smoothness of the wavelet, and the wavelet invariant results will hold for any  $f \in \mathbf{L}^1(\mathbb{R})$  and any  $k$ .

REMARK 4.4 Since  $(Pf_\tau)(\xi) = (Pf)((1 - \tau)\xi)$ , dilation will transport the frequency content at  $\xi$  to  $(1 - \tau)\xi$ , so that the displacement is  $\tau\xi$ . Thus, when  $\xi$  is very large,  $|(Pf)(\xi) - (Pf_\tau)(\xi)|$  can be large even for  $\tau$  small. Because the wavelet invariants bin the frequency content, and these bins become increasingly large in the high frequencies, this does not occur for wavelet invariants. More specifically, there is always a signal  $f$  and frequency  $\xi$  for which  $|(Pf)(\xi) - (\tilde{P}f)(\xi)|$  is large regardless of  $k$ . Consider for example when  $(Pf)(\omega) = e^{-(\omega-\xi)^2}$ . Then  $\Omega_k(\xi) \sim \xi^k$ , and  $|(Pf)(\xi) - (\tilde{P}f)(\xi)| \gtrsim 1$ . However, for  $M$  large enough, the order  $k$  wavelet invariant estimator satisfies  $|(Sf)(\lambda) - (\tilde{S}f)(\lambda)| = O(k\Psi_{k+2}\eta^{k+2})$  for all  $\lambda$ . The wavelet invariants are thus stable for high-frequency signals, where the power spectrum fails.

REMARK 4.5 For the wavelet invariants there will be a unique  $\tilde{k}$  that minimizes  $k\Psi_{k+2}(2E\eta)^{k+2}$ , and  $\tilde{k}$  does not depend on  $\lambda$ . Furthermore,  $\tilde{k}$  can be explicitly computed given the wavelet  $\psi$  and moment constant  $E$ . On the other hand, the minimum of  $k\Omega_{k+2}(\omega)(2E\omega)^{k+2}$  with respect to  $k$  will depend on both the frequency  $\omega$  and the signal  $f$ , so that  $\tilde{k} = \tilde{k}(\omega, f)$ , and it becomes unclear how to choose the unbiasing order.

### 4.4 Simulation results for dilation MRA

We first illustrate the unbiasing procedure of Propositions 4.1 and 4.2 for the high-frequency signal  $f(x) = e^{-5x^2} \cos(32x)$ . Figure 3 shows the power spectrum estimator  $\tilde{P}f$  and the wavelet invariant estimator  $\tilde{P}_{Sf}$  for  $k = 0, 2, 4$  for both small and large dilations, where  $\tilde{P}_{Sf}$  denotes the combined wavelet invariant unbiasing plus optimization procedure (see Section 6.5). Higher order unbiasing is beneficial for both methods for small dilations but fails for the power spectrum for large dilations. Both methods

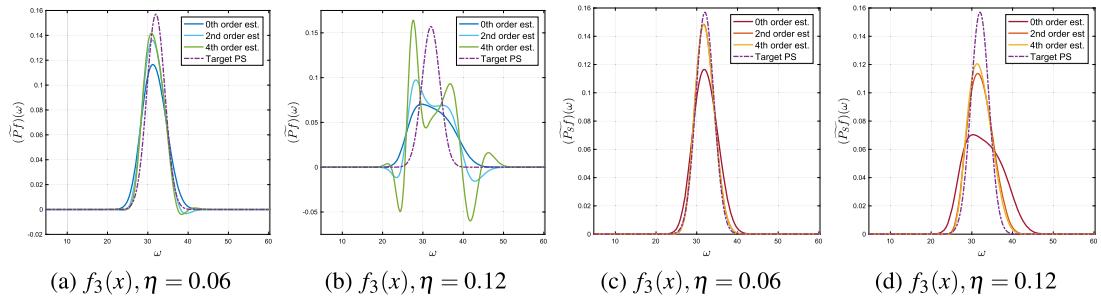


FIG. 3. Order  $k = 0, 2, 4$  power spectrum estimators  $\tilde{P}f$  (first two figures) and wavelet invariant estimators  $\tilde{P}_{Sf}$  (last two figures) for the signal  $f_3(x) = e^{-5x^2} \cos(32x)$ . Figures 3(a) and 3(c) show small dilations and Figs 3(b) and 3(d) show large dilations.

will of course fail for  $\eta$  large enough, but for high-frequency signals the power spectrum fails much sooner.

Next we compare  $\|Pf - \tilde{P}f\|_2$  and  $\|Pf - \tilde{P}_{Sf}\|_2$ , the  $L^2$  error of estimating the power spectrum of the target signal via the power spectrum estimators of Proposition 4.1 and via the wavelet invariant estimators of Proposition 4.2, followed by a convex optimization procedure. We consider order  $k = 0, 2, 4$  estimators for both the power spectrum and wavelet invariants on the following Gabor atoms of increasing frequency:

$$\begin{aligned} f_1(x) &= e^{-5x^2} \cos(8x) \\ f_2(x) &= e^{-5x^2} \cos(16x) \\ f_3(x) &= e^{-5x^2} \cos(32x). \end{aligned}$$

These functions satisfy  $f = \text{Real}(h)$  where  $(Ph)(\omega) = (\pi/5)e^{-(\omega-\xi)^2/10}$  for  $\xi = 8, 16, 32$ , and thus exhibit the behavior described in Remark 4.4.

Simulation results are shown in Fig. 4; the horizontal axis shows  $\log_2(M)$  while the vertical axis shows  $\log_2(\text{Error})$ . For each value of  $M$ , the error was calculated for 10 independent simulations and then averaged. The unbiasing procedure of Propositions 4.1 and 4.2 requires knowledge of the moments of the dilation distribution, but in practice these are unknown. Thus, the first two even moments of the dilation distribution ( $\eta^2, C_4\eta^4$ ) were estimated empirically with the fourth-order estimators described in Section 6.3 (see Definition 6.1). For the low-frequency signal, the fourth-order power spectrum estimator was best for both small and large dilations and is preferable due to the lower computational cost (see Remark 2.3). For the high-frequency signal, the fourth-order wavelet invariant estimator was best for large dilations and WSC  $k = 2$  and  $k = 4$  were best and equivalent for small dilations. For the medium-frequency signal, the higher order power spectrum estimators were best for small dilations while the higher order wavelet invariant estimators were best for large dilations. Thus, the simulation results confirm that the wavelet invariants will have an advantage over Fourier invariants when the signals are either high frequency or corrupted by large dilations. We remark that one obtains nearly identical error plots with oracle knowledge of the dilation moments, indicating that the empirical moment estimation procedure is highly accurate in the absence of additive noise, even for small  $M$  values.

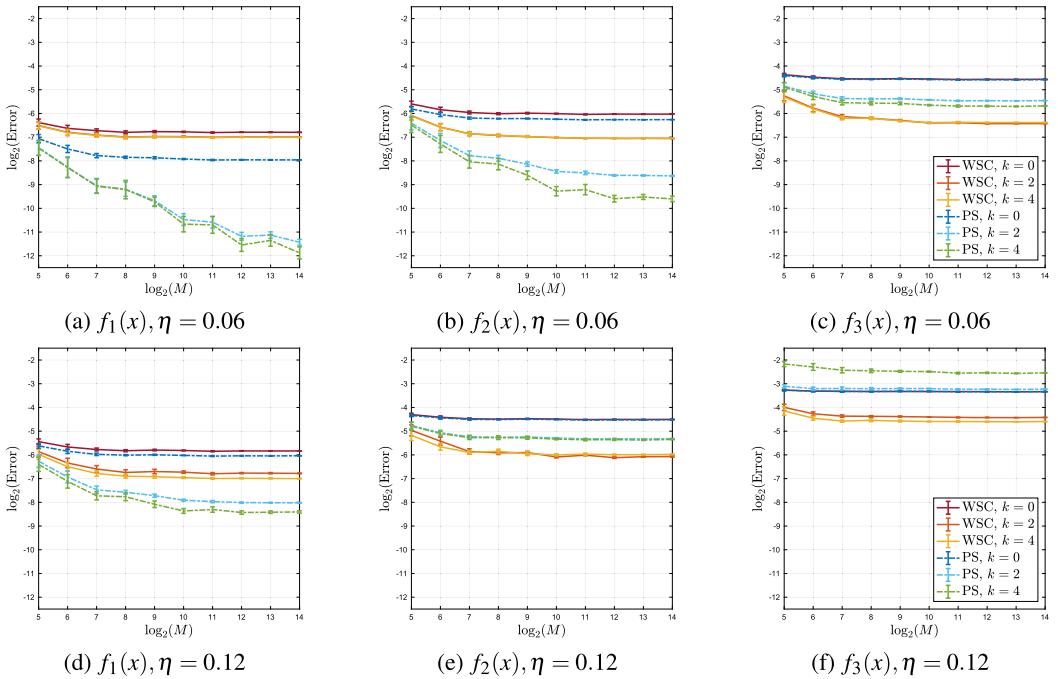


FIG. 4.  $L^2$  error with standard error bars for dilation model (empirical moment estimation). Top row shows results for small dilations ( $\eta = 0.06$ ) and bottom row shows results for large dilations ( $\eta = 0.12$ ). First, second, third column shows results for low, medium, high frequency Gabor signals. All plots have the same axis limits.

## 5. Noisy dilation MRA model

Finally, we consider the noisy dilation MRA model (Model 2) where signals are randomly translated and dilated and corrupted by additive noise. Section 5.1 gives unbiasing results for wavelet invariants and Section 5.2 reports relevant simulations.

### 5.1 Wavelet invariant results for noisy dilation MRA

To state Proposition 5.1 as succinctly as possible, we also define the following quantity

$$\Psi := \sum_{m=0,2,\dots,k} \Psi_m(E\eta)^m, \quad (5.1)$$

where  $E$  is defined in (4.3) and  $\Psi_m$  is defined in (A1).

**PROPOSITION 5.1** Assume Model 2 and that  $\psi$  is  $(k+2)$ -admissible. Define the following estimator of  $(Sf)(\lambda)$ :

$$(\tilde{S}f)(\lambda) := \frac{1}{M} \sum_{j=1}^M \left[ (Sy_j)(\lambda) - B_2 \eta^2 \lambda^2 (Sy_j)''(\lambda) - \dots - B_k \eta^k \lambda^k (Sy_j)^{(k)}(\lambda) \right] - \sigma^2$$

where the constants  $B_i$  satisfy (4.1). Then with probability at least  $1 - 1/t^2$

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \lesssim k\Lambda_{k+2}(\lambda)(2E\eta)^{k+2} + \frac{t}{\sqrt{M}} \left[ k\Lambda(\lambda) + \Psi\sigma^2 + \sqrt{\Psi(\Lambda_0(\lambda) + \Lambda(\lambda))\sigma} \right], \quad (5.2)$$

where  $E, \Lambda(\lambda), \Psi$  are as defined in (4.3), (4.5), (5.1).

The following corollary is an immediate consequence of Proposition 5.1.

**COROLLARY 5.1** Let the assumptions of Proposition 5.1 hold, and in addition assume  $\Psi_i(2E\eta)^i$  is decreasing for  $i \leq k+2$ . Then with probability at least  $1 - 1/t^2$

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \lesssim k\Psi_{k+2}(2E\eta)^{k+2} \|f\|_1^2 + \frac{tk}{\sqrt{M}} \left[ k\eta \|f\|_1^2 + \sigma \|f\|_1 + \sigma^2 \right]. \quad (5.3)$$

We remark that there are two components to the estimation error bounded by the right-hand side of (5.3): the first two terms are the error due to dilation, as in Corollary 4.1 of Proposition 4.2, and the last two terms are the error due to additive noise, as given in Proposition 3.2. Thus, the wavelet invariant representation allows for a decomposition of the error of the noisy dilation MRA model into the sum of the errors of the random dilation model and the additive noise model. This is possible because the representation inherits the differentiability of the wavelet and is not possible when  $P\psi \notin C^k(\mathbb{R})$ , in which case the dilation unbiasing procedure has a more complicated effect on the additive noise. A result equivalent to Proposition 5.1 cannot be made for the power spectrum, because the nonlinear unbiasing procedure of Proposition 4.1 cannot be applied to the power spectra of signals from the noisy dilation MRA corruption model, since they are not differentiable in the presence of additive noise.

*Proof of Proposition 5.1.* Since  $Sf$  is a translation invariant representation, we can ignore the translation factors  $\{t_j\}_{j=1}^M$  and consider the model  $y_j = f_{t_j} + \varepsilon_j$ . For notational convenience, we define the following order  $k$  derivative ‘unbiasing’ operator:

$$A_\lambda g(\lambda) := g(\lambda) - B_2\eta^2\lambda^2 \frac{d}{d\lambda^2}g(\lambda) - \dots - B_k\eta^k\lambda^k \frac{d}{d\lambda^k}g(\lambda), \quad (5.4)$$

which is defined on any function of  $\lambda$ , so that we can express our estimator by

$$\begin{aligned} (\tilde{S}f)(\lambda) &= \frac{1}{M} \sum_{j=1}^M \left[ \frac{1}{2\pi} \int |\hat{y}_j(\omega)|^2 A_\lambda |\hat{\psi}_\lambda(\omega)|^2 d\omega \right] - \sigma^2 \\ &= \frac{1}{M} \sum_{j=1}^M \left[ \frac{1}{2\pi} \int \left( |\hat{f}_{t_j}(\omega)|^2 + \hat{f}_{t_j}(\omega)\bar{\hat{\varepsilon}}_j(\omega) + \bar{\hat{f}}_{t_j}(\omega)\hat{\varepsilon}_j(\omega) + |\hat{\varepsilon}_j(\omega)|^2 \right) A_\lambda |\hat{\psi}_\lambda(\omega)|^2 d\omega \right] - \sigma^2. \end{aligned}$$

We can thus decompose the error as follows:

$$|(\tilde{S}f)(\lambda) - (Sf)(\lambda)| \leq \underbrace{\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int \left( \widehat{f}_{\tau_j}(\omega) \widehat{\varepsilon}_j(\omega) + \widehat{f}_{\tau_j}(\omega) \widehat{\varepsilon}_j(\omega) \right) A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right|}_{\text{Cross Term Error}} \\ + \underbrace{\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int |\widehat{f}_{\tau_j}(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega - (Sf)(\lambda) \right|}_{\text{Dilation Error}} + \underbrace{\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \sigma^2 \right|}_{\text{Additive Noise Error}}.$$

To bound the above terms we utilize the following two Lemmas, which are proved in Appendix F.

**LEMMA 5.1** Let the notation and assumptions of Proposition 5.1 hold, and let  $A_\lambda$  be the operator defined in (5.4). Then with probability at least  $1 - 1/t^2$

$$\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \sigma^2 \right| \leq \frac{2t\sqrt{k}\Psi\sigma^2}{\sqrt{M}}.$$

**LEMMA 5.2** Let the notation and assumptions of Proposition 5.1 hold, and let  $A_\lambda$  be the operator defined in (5.4). Then with probability at least  $1 - 1/t^2$

$$\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int \left( \widehat{f}_{\tau_j}(\omega) \widehat{\varepsilon}_j(\omega) + \widehat{f}_{\tau_j}(\omega) \widehat{\varepsilon}_j(\omega) \right) A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right| \lesssim \frac{t}{\sqrt{M}} \sqrt{\Psi(\Lambda_0(\lambda) + \Lambda(\lambda))} \sigma.$$

Applying Proposition 4.2 to bound the dilation error, Lemma 5.1 to bound the additive noise error, and Lemma 5.2 to bound the cross term error gives (5.2).  $\square$

## 5.2 Simulation results for noisy dilation MRA

We once again consider the Gabor atoms of varying frequency introduced in Section 4.4, and compare the  $L^2$  error of estimating the power spectrum by (1) averaging the power spectra of the noisy signals, and applying additive noise unbiasing; this is the zero-order power spectrum method (PS  $k = 0$ ), defined in Proposition 3.1, and (2) by approximating the wavelet invariants by the estimators given in Proposition 5.1 for  $k = 0, 2, 4$ , and then applying the optimization procedure described in Section 6.5; we refer to these methods as WSC  $k = i$  for  $i = 0, 2, 4$ . We emphasize that for the noisy dilation MRA model, it is impossible to define higher order methods for the power spectrum.

We first consider the errors obtained given oracle knowledge of the noise moments, both additive and dilation. Results are shown in Fig. 5 for all parameter combinations resulting from  $\sigma = 2^{-4}, 2^{-3}$  (giving SNR = 2.2, 0.56) and  $\eta = 0.06, 0.12$ . The horizontal axis shows  $\log_2(M)$  and the vertical axis shows  $\log_2(\text{Error})$ ; for each value of  $M$ , the error was calculated for 10 independent simulations and then averaged. For all simulations  $\tau$  was given a uniform distribution, a challenging regime for dilations, and the sample size ranged over  $16 \leq M \leq 131,072$ . For the medium- and high-frequency signals, for large enough  $M$ , WSC  $k = 2$  and WSC  $k = 4$  have significantly smaller error than the order zero estimators, indicating that the nonlinear unbiasing procedure of Proposition 5.1 contributes a definitive advantage.

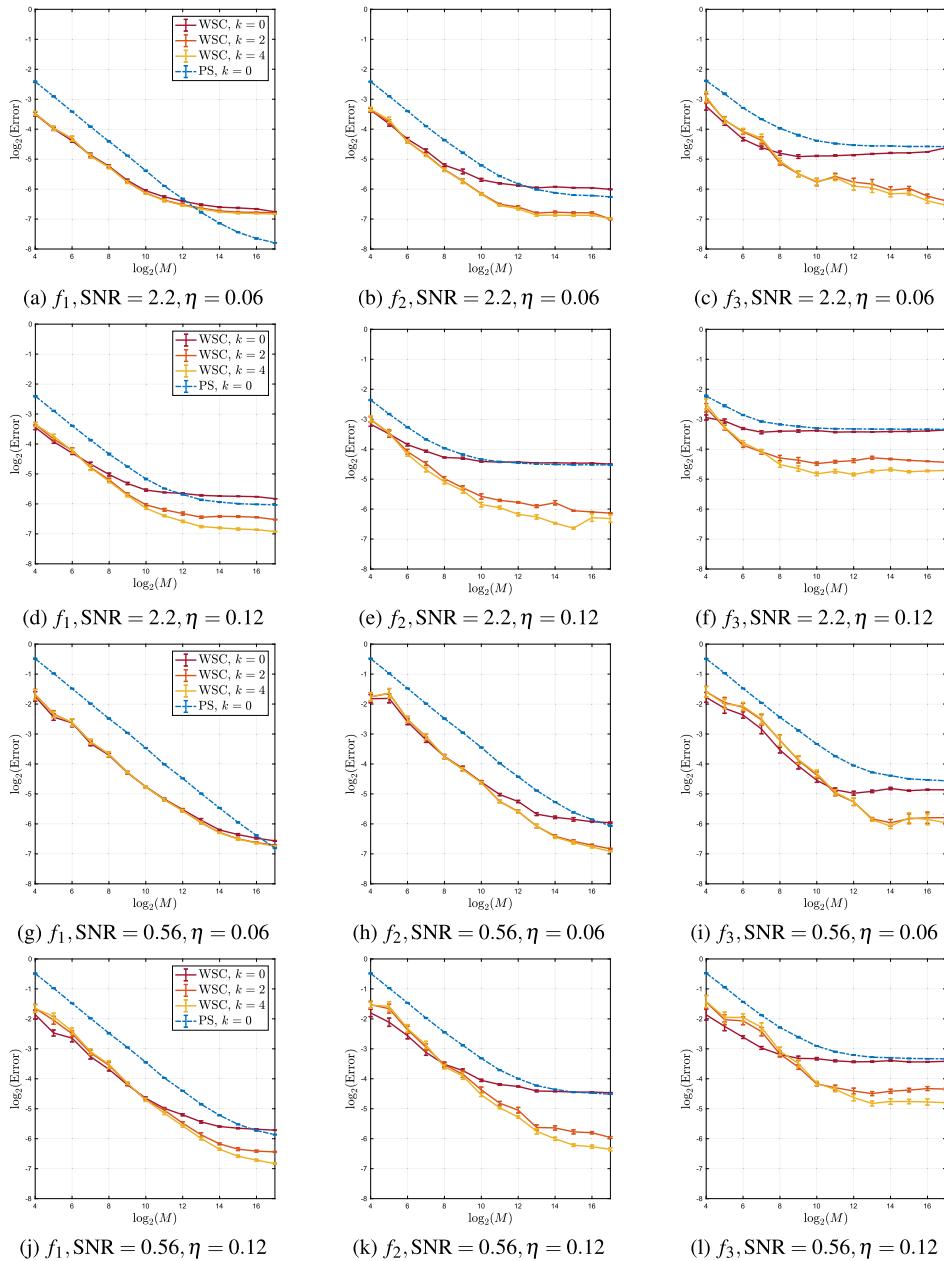


FIG. 5.  $\mathbf{L}^2$  error with standard error bars for noisy dilation MRA model (oracle moment estimation). First, second and third column shows results for low-, medium- and high-frequency Gabor signals. All plots have the same axis limits.

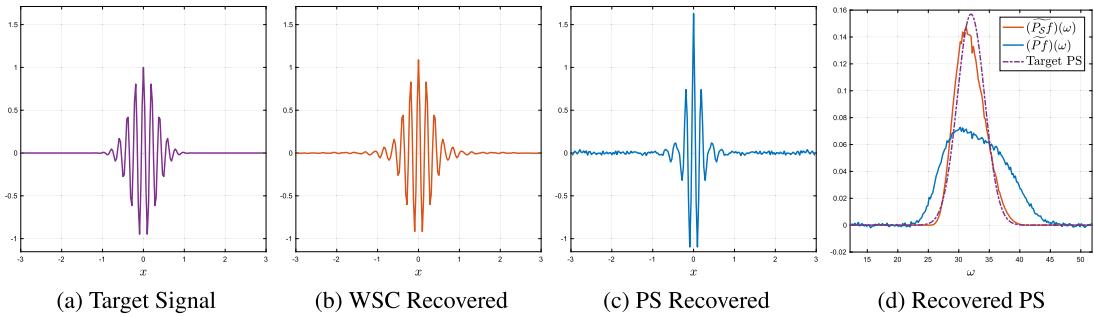


FIG. 6. Signal recovery results for  $f_3(x) = e^{-5x^2} \cos(32x)$  with  $M = 20,000$ ,  $\eta = 0.12$ , SNR = 2.2.

For the high-frequency signal and large  $M$ , the error using WSC  $k = 4$  is decreased by a factor of about 3 from the PS  $k = 0$  error. For small dilations ( $\eta = 0.06$ ), there is not much of a difference in performance between WSC  $k = 2$  and WSC  $k = 4$ , but the gap between these estimators widens for large dilations ( $\eta = 0.12$ ), as the fourth-order correction becomes more important. For the low-frequency signal under small dilations, PS  $k = 0$  achieves the smallest error for large  $M$ . However, when  $M$  is small or the dilations are large, the WSC estimators have the advantage for the low-frequency signal as well, and WSC  $k = 4$  is once again the best estimator for large  $M$ .

We note that although in general recovering the power spectrum is insufficient for recovering the signal, the signal can be recovered when  $\widehat{f}(\omega) \in \mathbb{R}$  and  $\widehat{f}(\omega) \geq 0$  by taking the inverse Fourier transform of the root power spectrum. Figure 6 shows the approximate signals recovered by this procedure from PS  $k = 0$  (Fig. 6(c)) and WSC  $k = 4$  (Fig. 6(b)) for the high-frequency Gabor signal  $f_3(x)$  (Fig. 6(a)). The WSC-recovered signal is a much better approximation of the target signal. The recovered power spectra are shown in Fig. 6(d); PS  $k = 0$  is much flatter than the target power spectrum, while WSC  $k = 4$  is a good approximation of both the shape and height of the target power spectrum.

Appendix G outlines an empirical procedure for estimating the moments of  $\tau$  in the special case when  $t = 0$  in the noisy dilation MRA model (i.e. no random translations). All simulations reported in Fig. 5 are repeated (with minor modifications) with empirical additive and dilation moment estimation, and the results are reported in Fig. G7 of Appendix G.

Appendix H contains additional simulation results for a variety of high-frequency signals.

**REMARK 5.1** One could also solve noisy dilation MRA with an EM algorithm. Appendix I describes how the method proposed in [1] can be extended to solve Model 2. Although EM algorithms provide a flexible tool for accurate parameter estimation in a variety of MRA models, the primary disadvantage is the high computational cost of each iteration. Each iteration costs  $O(Mn^3)$ , while wavelet invariant estimators can be computed in  $O(Mn^2)$ . In addition the statistical priors chosen may bias the signal reconstruction [12], and the algorithm will generally only converge to a local maximum. In this article we thus explore whether it is possible to solve noisy dilation MRA more efficiently and accurately by nonlinear unbiassing procedures.

## 6. Numerical implementation

In this section we describe the numerical implementation of the proposed method used to generate the results reported in Sections 3, 4.4 and 5.2. Section 6.1 describes how signals were generated, and

Sections 6.2 and 6.3 describe empirical procedures for estimating the additive noise level and the moments of the dilation distribution  $\tau$ . Finally, Section 6.4 discusses how the derivatives used for unbiasing were computed, and Section 6.5 describes the convex optimization algorithm used to recover  $Pf$  from  $Sf$ . All simulations used a Morlet wavelet constructed with  $\xi = 3\pi/4$ .

### 6.1 Signal generation and SNR

All signals were defined on  $[-N/4, N/4]$  and then padded with zeros to obtain a signal defined on  $[-N/2, N/2]$ ; the additive noise was also defined on  $[-N/2, N/2]$ . Signals were sampled at a rate of  $1/2^\ell$ , thus resolving frequencies in the interval  $[-2^\ell\pi, 2^\ell\pi]$  with a frequency sampling rate of  $2\pi/N$ . We used  $N = 2^5$  and  $\ell = 5$  in all experiments, keeping the box size and resolution fixed. For each experiment with hidden signal  $f$ , the SNR was calculated by  $\text{SNR} = \left( \frac{1}{N} \int_{-N/2}^{N/2} f(x)^2 dx \right) / \sigma^2$ .

### 6.2 Empirical estimation of additive noise level

The additive noise level  $\sigma^2$  can be estimated from the mean vertical shift of the mean power spectrum  $\frac{1}{M} \sum_{j=1}^M |\hat{y}_j(\omega)|^2$  in the tails of the distribution. Specifically, for  $\Sigma = [-2^\ell\pi, 2^\ell\pi] \setminus [-2^{\ell-1}\pi, 2^{\ell-1}\pi]$ , we define

$$\tilde{\sigma}^2 = \frac{1}{|\Sigma|} \sum_{\omega \in \Sigma} \frac{1}{M} \sum_{j=1}^M |\hat{y}_j(\omega)|^2.$$

If we choose  $\ell$  large enough so that the target signal frequencies are essentially contained in the interval  $[-2^{\ell-1}\pi, 2^{\ell-1}\pi]$ ,  $|\hat{y}_j(\omega)|^2 = |\hat{\varepsilon}_j(\omega)|^2$  for  $\omega \in \Sigma$ , and this is a robust and unbiased estimation procedure since  $\mathbb{E}|\hat{\varepsilon}_j(\omega)|^2 = \sigma^2$  by Lemma D.1.

### 6.3 Empirical moment estimation for dilation MRA

Given the additive noise level, the moments of the dilation distribution  $\tau$  for dilation MRA (Model 3) can be empirically estimated from the mean and variance of the random variables  $\alpha_m(y_j)$  defined by

$$\alpha_m(y_j) = \int_0^{2^\ell\pi} \omega^m |\hat{y}_j(\omega)|^2 d\omega \quad (6.1)$$

for integer  $m \geq 0$ . More specifically, we define the order  $m$  squared coefficient of variation by

$$CV_m := \frac{\text{Var}[\alpha_m(y_j)]}{|\mathbb{E}[\alpha_m(y_j)]|^2}. \quad (6.2)$$

The following proposition guarantees that for  $M$  large the second and fourth moments of the dilation distribution can be recovered from  $CV_0, CV_1$ . In fact one could continue this procedure for higher  $m$  values, i.e.  $\{CV_m\}_{m=0}^{k/2-1}$  will define estimators of the first  $\frac{k}{2}$  even moments of  $\tau$ , accurate up to  $O(\eta^{k+2})$ , but for brevity we omit the general case.

PROPOSITION 6.1 Assume Model 3 and  $CV_0, CV_1$  defined by (6.1) and (6.2). Then

$$\begin{aligned} CV_0 &= \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6) \\ CV_1 &= 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6). \end{aligned}$$

*Proof.* Since  $y_j = L_{\tau_j}f(x - t_j)$ ,

$$\begin{aligned} \alpha_m(y_j) &= \int_0^{2^\ell \pi} \omega^m |\widehat{f}((1 - \tau_j)\omega)|^2 d\omega \\ &= \int_0^{2^\ell \pi(1-\tau_j)} \frac{\xi^m}{(1 - \tau_j)^m} |\widehat{f}(\xi)|^2 \frac{d\xi}{(1 - \tau_j)} \\ &= (1 - \tau_j)^{-(m+1)} \alpha_m(f), \end{aligned}$$

where we assume we have chosen  $\ell$  large enough so that the target signal frequencies are essentially supported in  $[-2^{\ell-1}\pi, 2^{\ell-1}\pi]$ . Thus,

$$CV_m = \frac{\mathbb{E}[\alpha_m(y_j)^2] - (\mathbb{E}[\alpha_m(y_j)])^2}{(\mathbb{E}[\alpha_m(y_j)])^2} = \frac{\mathbb{E}[(1 - \tau_j)^{-2(m+1)}]}{(\mathbb{E}[(1 - \tau_j)^{-(m+1)}])^2} - 1.$$

When  $m = 0$ , we have

$$\begin{aligned} CV_0 &= \frac{\mathbb{E}[(1 - \tau_j)^{-2}]}{(\mathbb{E}[(1 - \tau_j)^{-1}])^2} - 1 \\ &= \frac{\mathbb{E}[1 + 2\tau + 3\tau^2 + 4\tau^3 + 5\tau^4 + O(\tau^5)]}{(\mathbb{E}[1 + \tau + \tau^2 + \tau^3 + \tau^4 + O(\tau^5)])^2} - 1 \\ &= \frac{1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6)}{(1 + \eta^2 + C_4\eta^4 + O(\eta^6))^2} - 1 \\ &= \frac{1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6)}{1 + 2\eta^2 + (2C_4 + 1)\eta^4 + O(\eta^6)} - 1 \\ &= (1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6))(1 - 2\eta^2 + (3 - 2C_4)\eta^4 + O(\eta^6)) - 1 \\ &= \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6). \end{aligned}$$

When  $m = 1$ , we have

$$\begin{aligned}
CV_1 &= \frac{\mathbb{E}[(1 - \tau_j)^{-4}]}{(\mathbb{E}[(1 - \tau_j)^{-2}])^2} - 1 \\
&= \frac{\mathbb{E}[1 + 4\tau + 10\tau^2 + 20\tau^3 + 35\tau^4 + O(\tau^5)]}{(\mathbb{E}[1 + 2\tau + 3\tau^2 + 4\tau^3 + 5\tau^4 + O(\tau^5)])^2} - 1 \\
&= \frac{1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6)}{(1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6))^2} - 1 \\
&= \frac{1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6)}{(1 + 6\eta^2 + (9 + 10C_4)\eta^4 + O(\eta^6))} - 1 \\
&= (1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6))(1 - 6\eta^2 + (27 - 10C_4)\eta^4 + O(\eta^6)) - 1 \\
&= 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6).
\end{aligned}$$

□

We cannot compute  $CV_m$  exactly, but by replacing  $\text{Var}, \mathbb{E}$  with their finite sample estimators, we obtain an approximate  $\widetilde{CV}_m \rightarrow CV_m$  as  $M \rightarrow \infty$ . Motivated by Proposition G.1, we thus use  $\widetilde{CV}_0, \widetilde{CV}_1$  to define estimators of  $\eta^2$  and  $C_4\eta^4$ .

**DEFINITION 6.1** Assume Model 3 and let  $\widetilde{CV}_0, \widetilde{CV}_1$  be the empirical versions of (6.2). Define the second-order estimator of  $\eta^2$  by  $\widetilde{\eta}^2 = \widetilde{CV}_0$ . Define the fourth-order estimators of  $(\eta^2, C_4\eta^4)$  by the unique positive solution  $(\widetilde{\eta}^2, \widetilde{C}_4)$  of

$$\begin{aligned}
\widetilde{CV}_0 &= \eta^2 + (3C_4 - 3)\eta^4 \\
\widetilde{CV}_1 &= 4\eta^2 + (25C_4 - 33)\eta^4.
\end{aligned}$$

For noisy dilation MRA (Model 2), estimating the dilation moments is more difficult. We give a procedure for estimating the moments in the special case  $t = 0$  in Appendix G. Empirical moment estimation procedures that are simultaneously robust to translations, dilations and additive noise are an important area of future research.

#### 6.4 Derivatives

All derivatives were approximated numerically using finite difference calculations. A sixth-order finite difference approximation was used for second derivatives, and a fourth-order finite difference approximation was used for fourth derivatives. This procedure was done on the empirical mean for each representation, not the individual signals. In fact since the wavelet is known,  $\frac{d^n}{d\lambda^n} |\widetilde{\psi}_\lambda(\omega)|^2$  could be computed analytically, and  $(S_{Y_j})^{(n)}(\lambda)$  computed using Definition 2.2. Thus error due to finite difference approximations could be avoided for wavelet invariant derivatives.

#### 6.5 Optimization

In this section we describe the convex optimization algorithm for computing  $(\widetilde{P_{Sf}})$ , the power spectrum approximation that best matches the wavelet invariants  $(\widetilde{Sf})$ . Since the wavelet invariants are only

computed for  $\lambda > 0$ , we also incorporate zero frequency information into the loss function via  $(\tilde{P}f)(0)$ , an approximation of the power spectrum at frequency zero. For all of the examples reported in this article, the quasi-newton algorithm was used to solve an unconstrained optimization problem minimizing the following convex loss function:

$$\text{loss}(\hat{g}) := \sum_{\lambda} \left( \langle \hat{g}^2, |\hat{\psi}_{\lambda}^+|^2 \rangle - \tilde{S}f(\lambda) \right)^2 + \left( \hat{g}(0)^2 - (\tilde{P}f)(0) \right)^2,$$

where

$$|\hat{\psi}_{\lambda}^+(\omega)|^2 = \left( |\hat{\psi}_{\lambda}(\omega)|^2 + |\hat{\psi}_{\lambda}(-\omega)|^2 \right) \cdot \mathbf{1}(\omega \geq 0).$$

Letting  $\hat{g}^*$  denote the minimizer of the above loss function, we then define  $(\tilde{P}_S f) := \hat{g}^*(\omega)^2$ . Theorem 2.4 ensures that when the loss function is defined with the exact wavelet invariants  $Sf$ , it has a unique minimizer corresponding to  $Pf$ . Whenever  $f(x) \in \mathbb{R}$ , the symmetry of  $(Pf)(\omega)$  ensures that  $(Sf)(\lambda) = \langle |\hat{f}|^2, |\hat{\psi}_{\lambda}^+|^2 \rangle$ , and thus it is sufficient to optimize over the non-negative frequencies and then symmetrically extend the solution. Such a procedure ensures the output of the optimization algorithm is symmetric while avoiding adding constraints to the optimization. The algorithm was initialized using the mean power spectrum with additive noise unbiasing only, i.e.  $\text{PS } k = 0$ . The optimization output does depend on various numerical tolerance parameters, which were held fixed for all examples.

**REMARK 6.1** Alternatively, one can invert the representation by applying a pseudo-inverse with Tikhonov regularization. Specifically, if  $F$  is the matrix defining the wavelet invariants, so that  $Sy = F(Py)$ , then one can define  $(\tilde{P}_S f) = (F^T F + \lambda I)^{-1} F^T (\tilde{S}f)$ . This procedure however requires careful selection of the hyper-parameter  $\lambda$  and did not work as well as inverting via optimization in our experiments.

## 7. Conclusion

This article considers a generalization of classic MRA, which incorporates random dilations in addition to random translations and additive noise and proposes solving the problem with a wavelet invariant representation. These wavelet invariants have several desirable properties over Fourier invariants, which allow for the construction of unbiasing procedures that cannot be constructed for Fourier invariants. Unbiasing the representation is critical for high-frequency signals, where even small diffeomorphisms cause a large perturbation. After unbiasing, the power spectrum of the target signal can be recovered from a convex optimization procedure.

Several directions remain for further investigation, including extending results to higher dimensions and considering rigid transformations instead of translations. Such extensions could be especially relevant to image processing, where variations in the size of an object can be modeled as dilations. Incorporating the effect of tomographic projection would also lead to results more directly relevant to problems such as cryo-EM. The tools of the present article, although significantly reducing the bias, do not allow for a completely unbiased estimator for noisy dilation MRA due to the bad scaling of certain intrinsic constants. Thus, an important open question is whether it is possible to define unbiased estimators for noisy dilation MRA using a different approach. The noisy dilation MRA model of this article corresponds to linear diffeomorphisms, and constructing unbiasing procedures that apply to more general diffeomorphisms is also an important future direction. In addition, one can construct wavelet

invariants that characterize higher order auto-correlation functions such as the bispectrum, and future work will investigate full signal recovery with such invariants.

## Funding

Alfred P. Sloan Foundation (Sloan Fellowship FG-2016-6607 to M.H.); Defense Advanced Research Projects Agency (Young Faculty Award D16AP00117 to M.H.); National Science Foundation (grant 1912906 to A.L.; grant 1620216 and CAREER award 1845856 to M.H.).

## Acknowledgements

We would like to thank the reviewers for their detailed comments and insights that greatly improved the manuscript. We would also like to thank Stephanie Hickey for providing useful references on flexible regions of macromolecular structures.

## REFERENCES

1. ABBE, E., BENDORY, T., LEEB, W., PEREIRA, J. M., SHARON, N. & SINGER, A. (2018) Multireference alignment is easier with an aperiodic translation distribution. *IEEE Trans. Inf. Theory*, **65**, 3565–3584.
2. AIZENBUD, Y., LANDA, B. & SHKOLNISKY, Y. (2019) Rank-one multi-reference factor analysis. arXiv preprint arXiv:1905.12442.
3. BAI, X.-c., RAJENDRA, E., YANG, G., SHI, Y. & SCHERES, S. H. W. (2015) Sampling the conformational space of the catalytic subunit of human  $\gamma$ -secretase. *eLife*, **4**, e11182.
4. BANDEIRA, A., CHEN, Y., LEDERMAN, R. R. & SINGER, A. (2020) Non-unique games over compact groups and orientation estimation in cryo-EM. *Inverse Probl.*
5. BANDEIRA, A., RIGOLLET, P. & WEED, J. (2017) Optimal rates of estimation for multi-reference alignment. arXiv preprint at arXiv:1702.08546.
6. BANDEIRA, A. S. (2015) Synchronization problems and alignment. *Topics in Mathematics of Data Science Lecture Notes*. Cambridge, MA: Massachusetts Institute of Technology.
7. BANDEIRA, A. S., BLUM-SMITH, B., KILEEL, J., PERRY, A., WEED, J. & WEIN, A. S. (2017) Estimation under group actions: recovering orbits from invariants. arXiv preprint arXiv:1712.10163.
8. BANDEIRA, A. S., BOUMAL, N. & SINGER, A. (2017) Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Math. Programming*, **163**, 145–167.
9. BANDEIRA, A. S., BOUMAL, N. & VORONINSKI, V. (2016) On the low-rank approach for semidefinite programs arising in synchronization and community detection. *Conf. Learn. Theory*, **49**, 361–382.
10. BANDEIRA, A. S., CHARIKAR, M., SINGER, A. & ZHU, A. (2014) Multireference alignment using semidefinite programming. *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*. ACM, pp. 459–470.
11. BARTESAGHI, A., MERK, A., BANERJEE, S., MATTHIES, D., WU, X., MILNE, J. L. S. & SUBRAMANIAM, S. (2015) 2.2 Åresolution cryo-EM structure of  $\beta$ -galactosidase in complex with a cell-permeant inhibitor. *Science*, **348**, 1147–1151.
12. BENDORY, T., BARTESAGHI, A. & SINGER, A. (March 2020) Single-particle cryo-electron microscopy: mathematical theory, computational challenges, and opportunities. *IEEE Signal Process. Mag.*, **37**, 58–76.
13. BENDORY, T., BOUMAL, N., LEEB, W., LEVIN, E. & SINGER, A. (2019) Multi-target detection with application to cryo-electron microscopy. *Inverse Probl.*
14. BENDORY, T., BOUMAL, N., MA, C., ZHAO, Z. & SINGER, A. (2017) Bispectrum inversion with application to multireference alignment. *IEEE Trans. Signal Process.*, **66**, 1037–1050.
15. BOUMAL, N. (2016) Nonconvex phase synchronization. *SIAM J. Optim.*, **26**, 2355–2377.

16. BOUMAL, N., BENDORY, T., LEDERMAN, R. R. & SINGER, A. (2018) Heterogeneous multireference alignment: a single pass approach. *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*. IEEE, pp. 1–6.

17. BOWMAN, G. D. & POIRIER, M. G. (2015) Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.*, **115**, 2274–2295.

18. BROWN, L. G. (1992) A survey of image registration techniques. *ACM Computing Surv. (CSUR)*, **24**, 325–376.

19. BRUNA, J. & MALLAT, S. (2018) Multiscale sparse microcanonical models. *Math. Stat. Learn.*, **1**, 257–315.

20. BUESCU, J. & PAIXÃO, A. C. (2007) Eigenvalue distribution of positive definite kernels on unbounded domains. *Integral Equ. Oper. Theory*, **57**, 19–41.

21. BUESCU, J., PAIXÃO, A. C., GARCIA, F. & LOURTIE, I. (2004) Positive-definiteness, integral equations and fourier transforms. *J. Integral Equ. Appl.*, **16**, 33–52.

22. CAPODIFERRO, L., CUSANI, R., JACOVITTI, G. & VASCOTTO, M. (1987) A correlation based technique for shift, scale, and rotation independent object identification. *ICASSP'87: IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 12. IEEE, pp. 221–224.

23. CHANDRAN, V. & ELGAR, S. L. (1992) Position, rotation, and scale invariant recognition of images using higher-order spectra. *ICASSP'92: IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, pp. 213–216.

24. CHEN, Y. & CANDÈS, E. J. (2018) The projected power method: an efficient algorithm for joint alignment from pairwise differences. *Comm. Pure Appl. Math.*, **71**, 1648–1714.

25. CHEN, Y., GUIBAS, L. J. & HUANG, Q.-X. (2014) Near-optimal joint object matching via convex relaxation. *Proceedings of the 31st International Conference on Machine Learning*, vol. 32 of Proceedings of Machine Learning Research, pp. 100–108.

26. CHENG, C., JIANG, J. & SUN, Q. (2017) Phaseless sampling and reconstruction of real-valued signals in shift-invariant spaces. *J. Fourier Anal. Appl.*, 1–34.

27. CLERC, M. & MALLAT, S. (2002) The texture gradient equation for recovering shape from texture. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**, 536–549.

28. CLERC, M. & MALLAT, S. (2003) Estimating deformations of stationary processes. *Ann. Stat.*, **31**, 1772–1821.

29. COLLIS, W. B., WHITE, P. R. & HAMMOND, J. K. (1998) Higher-order spectra: the bispectrum and trispectrum. *Mech. Syst. Signal Process.*, **12**, 375–394.

30. DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *J. Royal Stat. Soc. Ser. B (Methodological)*, **39**, 1–22.

31. DESJARLAIS, R. & TUMMINO, P. J. (2016) Role of histone-modifying enzymes and their complexes in regulation of chromatin biology. *Biochemistry*, **55**, 1584–1599.

32. DIAMOND, R. (1992) On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.*, **1**, 1279–1287.

33. DVORNEK, N. C., SIGWORTH, F. J. & TAGARE, H. D. (2015) Subspaceem: a fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction. *J. Struct. Biol.*, **190**, 200–214.

34. EICKENBERG, M., EXARCHAKIS, G., HIRN, M. & MALLAT, S. (2017) Solid harmonic wavelet scattering: predicting quantum molecular energy from invariant descriptors of 3D electronic densities. *Adv. Neural Inf. Proc. Syst. 30 (NIPS 2017)*, 6540–6549.

35. EICKENBERG, M., EXARCHAKIS, G., HIRN, M., MALLAT, S. & THIRY, L. (2018) Solid harmonic wavelet scattering for predictions of molecule properties. *J. Chem. Phys.*, **148**, 241732.

36. EKMAN, D., BJÖRKLUND, A. K., FREY-SKÖTT, J. & ELOFSSON, A. (2005) Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.*, **348**, 231–243.

37. FERNANDEZ-LEIRO, R., CONRAD, J., SCHERES, S. H. W. & LAMERS, M. H. (2015) Cryo-EM structures of the *E. coli* replicative dna polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and t. *eLife*, **4**, e11134.

38. FISCHER, N., NEUMANN, P., KONEVEGA, A. L., BOCK, L. V., FICNER, R., RODNINA, M. V. & STARK, H. (2015) Structure of the *E. coli* ribosome–ef-tu complex at  $< 3$  Åresolution by c s-corrected cryo-EM. *Nature*, **520**, 567–570.

39. FORNERIS, F., WU, J. & GROS, P. (2012) The modular serine proteases of the complement cascade. *Curr. Opin. Struct. Biol.*, **22**, 333–341.
40. FOROOSH, H., ZERUBIA, J. B. & BERTHOD, M. (2002) Extension of phase correlation to subpixel registration. *IEEE Trans. Image Process.*, **11**, 188–200.
41. FRANK, J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford, United Kingdom: Oxford University Press.
42. GAO, F., WOLF, G. & HIRN, M. (2019) Geometric scattering for graph data analysis. *Proceedings of the 36th International Conference on Machine Learning, PMLR*, vol. 97, pp. 2122–2131.
43. GIL-PITA, R., ROSA-ZURERA, M., JARABO-AMORES, P. & LÓPEZ-FERRERAS, F. (2005) Using multilayer perceptrons to align high range resolution radar signals. *International Conference on Artificial Neural Networks*. Springer, pp. 911–916.
44. HANSEN, L. P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.
45. HIRN, M., MALLAT, S. & POILVERT, N. (2017) Wavelet scattering regression of quantum chemical energies. *Multiscale Model. Simul.*, **15**, 827–863. arXiv:1605.04654.
46. HOTTA, K., MISHIMA, T. & KURITA, T. (2001) Scale invariant face detection and classification method using shift invariant features extracted from log-polar image. *IEICE Trans. Inf. Syst.*, **84**, 867–878.
47. HUDSON, S. & PSALTIS, D. (1993) Correlation filters for aircraft identification from radar range profiles. *IEEE Trans. Aerosp. Electron. Syst.*, **29**, 741–748.
48. KAM, Z. (1980) The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, **82**, 15–39.
49. KLEBANER, F. C. (2012) *Introduction to Stochastic Calculus With Applications*. Singapore: World Scientific Publishing Company.
50. LANDA, B. & SHKOLNISKY, Y. (2019) Multi-reference factor analysis: low-rank covariance estimation under unknown translations. arXiv preprint arXiv:1906.00211.
51. LEGGETT, R. M., HEAVENS, D., CACCAMO, M., CLARK, M. D. & DAVEY, R. P. (2015) Nanook: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. *Bioinformatics*, **32**, 142–144.
52. LEVITT, M. (2009) Nature of the protein universe. *Proc. Natl. Acad. Sci.*, **106**, 11079–11084.
53. LIM, W. A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.*, **12**, 61–68.
54. MA, C., BENDORY, T., BOUMAL, N., SIGWORTH, F. & SINGER, A. (2019) Heterogeneous multireference alignment for images with application to 2d classification in single particle reconstruction. *IEEE Trans. Image Process.*, **29**, 1699–1710.
55. MALLAT, S. (2008) *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd edn. Cambridge, MA: Academic Press.
56. MALLAT, S. (2012) Group invariant scattering. *Comm. Pure Appl. Math.*, **65**, 1331–1398.
57. MARTINEC, D. & PAJDLA, T. Robust rotation and translation estimation in multiview reconstruction. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2007. IEEE, pp. 1–8.
58. McGINTY, R. K. & TAN, S. (2016) Recognition of the nucleosome by chromatin factors and enzymes. *Curr. Opin. Struct. Biol.*, **37**, 54–61.
59. MERK, A., BARTESAGHI, A., BANERJEE, S., FALCONIERI, V., RAO, P., DAVIS, M. I., PRAGANI, R., BOXER, M. B., EARL, L. A., MILNE, J. L. S., et al. (2016) Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell*, **165**, 1698–1707.
60. MEYNARD, A. & TORRÉSANI, B. (2018) Spectral analysis for nonstationary audio. *IEEE/ACM Trans. Audio Speech Lang. Process.*, **26**, 2371–2380.
61. OMER, H. & TORRÉSANI, B. (2013) Estimation of frequency modulations on wideband signals; applications to audio signal analysis. *10th International Conference on Sampling Theory and Applications*, Bremen, Germany, 2013, 29–32.
62. OMER, H. & TORRÉSANI, B. (2017) Time-frequency and time-scale analysis of deformed stationary processes, with application to non-stationary sound modeling. *Appl. Comput. Harmon. Anal.*, **43**, 1–22.

63. PALAMINI, M., CANGIANI, A. & FORNERIS, F. (2016) Identifying and visualizing macromolecular flexibility in structural biology. *Front. Mol. Biosci.*, **3**, 47.

64. PARK, W. & CHIRIKJIAN, G. S. (2014) An assembly automation approach to alignment of noncircular projections in electron microscopy. *IEEE Trans. Automat. Sci. Eng.*, **11**, 668–679.

65. PARK, W., MIDGETT, C. R., MADDEN, D. R. & CHIRIKJIAN, G. S. (2011) A stochastic kinematic model of class averaging in single-particle electron microscopy. *Int. J. Robot. Res.*, **30**, 730–754.

66. PERRY, A., WEED, J., BANDEIRA, A., RIGOLLET, P. & SINGER, A. (2017) The sample complexity of multi-reference alignment. *SIAM J. Math. Data Sci.*, **1**, 497–517.

67. PERRY, A., WEIN, A. S., BANDEIRA, A. S. & MOITRA, A. (2018) Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.*, **71**, 2275–2322.

68. PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. & BRUBAKER, M. A. (2017) Cryosparc: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290.

69. ROBINSON, D., FARSIU, S. & MILANFAR, P. (2007) Optimal registration of aliased images using variable projection with applications to super-resolution. *Comput. J.*, **52**, 31–42.

70. SADLER, B. M. & GIANNAKIS, G. B. (1992) Shift-and rotation-invariant object reconstruction using the bispectrum. *JOSA A*, **9**, 57–69.

71. SJORS, H. W., SCHERES, M. V., NUÑEZ, R., SORZANO, C. O. S., MARABINI, R., HERMAN, G. T. & CARAZO, J.-M. (2005) Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.*, **348**, 139–149.

72. SHARON, N., KILEEL, J., KHOO, Y., LANDA, B. & SINGER, A. (2020) Method of moments for 3-D single particle ab initio modeling with non-uniform distribution of viewing angles. *Inverse Probl.*, **36**, 044003.

73. SINGER, A. (2011) Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.*, **30**, 20–36.

74. SINGER, A. (2018) Mathematics for cryo-electron microscopy. *Proceedings of the International Congress of Mathematicians*, vol. **4**. Rio de Janeiro, pp. 4013–4032.

75. SIROHI, D., CHEN, Z., SUN, L., KLOSE, T., PIERSON, T. C., ROSSMANN, M. G. & KUHN, R. J. (2016) The 3.8 Åresolution cryo-EM structure of Zika virus. *Science*, **352**, 467–470.

76. SONDAY, B., SINGER, A. & KEVREKIDIS, I. G. (2013) Noisy dynamic simulations in the presence of symmetry: data alignment and model reduction. *Comput. Math. Appl.*, **65**, 1535–1557.

77. SORZANO, C. O. S., BILBAO-CASTRO, J. R., SHKOLNISKY, Y., ALCORLO, M., MELERO, R., CAFFARENA-FERNÁNDEZ, G., LI, M., XU, G., MARABINI, R. & CARAZO, J. M. (2010) A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.*, **171**, 197–206.

78. SUN, W. (2017) Phaseless sampling and linear reconstruction of functions in spline spaces. arXiv preprint arXiv:1709.04779.

79. THEOBALD, D. L. & STEINDEL, P. A. (2012) Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics*, **28**, 1972–1979.

80. TSATSANIS, M. K. & GIANNAKIS, G. B. (1990) Translation, rotation, and scaling invariant object and texture classification using polyspectra. *Advanced Signal Processing Algorithms, Architectures, and Implementations*, vol. 1348. Bellingham, WA: International Society for Optics and Photonics, pp. 103–115.

81. VILLARREAL, S. A. & STEWART, P. L. (2014) Cryo-em and image sorting for flexible protein/dna complexes. *J. Struct. Biol.*, **187**, 76–83.

82. WEIN, A. S. (2018) Statistical estimation in the presence of group actions. *Ph.D. Thesis*, Massachusetts Institute of Technology.

83. WINKLER, J. & NIRANJAN, M. (2002) *Uncertainty in Geometric Computations*, vol. 704. Berlin, Germany: Springer Science & Business Media.

84. ZHONG, Y. & BOUMAL, N. (2018) Near-optimal bounds for phase synchronization. *SIAM J. Optim.*, **28**, 989–1016.

85. ZWART, J. P., VAN DER HEIDEN, R., GELSEMA, S. & GROEN, F. (2003) Fast translation invariant classification of HRR range profiles in a zero phase representation. *IEE Proc. Radar Sonar Nav.*, **150**, 411–418.

### A. Wavelet admissibility conditions

This appendix describes the wavelet admissibility conditions that are needed for the main results in this article, namely Propositions 4.2 and 5.1. The wavelet  $\psi$  is  $k$ -admissible if  $\widehat{\psi} \in \mathbf{C}^k(\mathbb{R})$  and  $\Psi_k < \infty$ ,  $\Theta_k < \infty$  where

$$\Psi_k := \frac{1}{2\pi} \sum_{i=0}^k \binom{k}{i} \frac{k!}{i!} \|\omega^i (P\psi)^{(i)}(\omega)\|_1, \quad (\text{A.1})$$

$$\Theta_k := \frac{1}{2\pi} \sum_{i=0}^k \binom{k}{i} \frac{k!}{i!} \|\omega^{i-2} (P\psi)^{(i)}(\omega)\|_1. \quad (\text{A.2})$$

For  $\psi$  to be  $k$ -admissible, it is sufficient for  $\widehat{\psi} \in \mathbf{C}^k(\mathbb{R})$ ,  $(P\psi)^{(i)}$  to decay faster than  $\omega^{i+1}$ , and  $\int \frac{|\widehat{\psi}(\omega)|^2}{\omega^2} d\omega < \infty$  (see Lemma B.1 in Appendix B). The condition  $\int \frac{|\widehat{\psi}(\omega)|^2}{\omega^2} d\omega < \infty$  is slightly stronger than the classic admissibility condition  $C_\psi := \int \frac{|\widehat{\psi}(\omega)|^2}{\omega} d\omega < \infty$  [55, Theorem 4.4]. When  $\widehat{\psi}$  is continuously differentiable,  $\widehat{\psi}(0) = 0$  is sufficient to guarantee  $C_\psi < \infty$ ; but here we need  $\widehat{\psi}(\omega) \sim \omega^{\frac{1}{2}+\epsilon}$  for some  $\epsilon > 0$  as  $\omega \rightarrow 0$ . If this condition is removed, we are not guaranteed  $\Theta_k < \infty$ , but all results in fact still hold, with  $\Lambda_k(\lambda) = \Psi_k \|f\|_1^2$  replacing  $\Lambda_k(\lambda) = \Psi_k \|f\|_1^2 \wedge \frac{\Theta_k \|f'\|_1^2}{\lambda^2}$  in Propositions 4.2 and 5.1. Any wavelet with fast decay satisfies this stronger admissibility condition, and it ensures that a smooth signal will enjoy a fast decay of wavelet invariants.

**REMARK A.1** The Morlet wavelet  $\psi(x) = g(x)(e^{i\tilde{\xi}x} - C)$  is  $k$ -admissible for any  $k$ , since  $\widehat{\psi} \in \mathbf{C}^\infty(\mathbb{R})$ ,  $P\psi$  has fast decay, and  $\widehat{\psi}(\omega) \sim \omega$  as  $\omega \rightarrow 0$ . One can also choose  $\widehat{\psi}$  to be an order  $k+1$ -spline of compact support.

### B. Properties of wavelet invariants

This appendix establishes several important properties of wavelet invariants. Lemma B.1 gives sufficient conditions guaranteeing that a wavelet is  $k$ -admissible. Lemmas 4.3 and 4.4 bound wavelet invariant derivatives. Lemma B.2 bounds terms that arise in the dilation unbiasing procedure of Sections 4.2 and 5.

**LEMMA B.1** ( $k$ -admissible). If  $\widehat{\psi} \in \mathbf{C}^k(\mathbb{R})$ ,  $(P\psi)^{(i)}$  decays fast than  $\omega^{i+1}$ , and  $\int \frac{|\widehat{\psi}(\omega)|^2}{\omega^2} d\omega < \infty$ , then  $\psi$  is  $k$ -admissible.

*Proof.* We first note that  $\widehat{\psi} \in \mathbf{C}^k(\mathbb{R})$  guarantees  $P\psi \in \mathbf{C}^k(\mathbb{R})$ . Since  $(P\psi)^{(i)}$  decays faster than  $\omega^{i+1}$  and  $P\psi \in \mathbf{C}^k(\mathbb{R})$ ,  $\omega^i (P\psi)^{(i)}(\omega) \in \mathbf{L}^1(\mathbb{R})$  for  $0 \leq i \leq k$ , so  $\Psi_k < \infty$ . Also  $P\psi \in \mathbf{C}^k(\mathbb{R})$  and  $\omega^i (P\psi)^{(i)} \in \mathbf{L}^1(\mathbb{R})$  implies  $\omega^{i-2} (P\psi)^{(i)} \in \mathbf{L}^1(\mathbb{R})$  for  $2 \leq i \leq k$ . In addition,  $\omega^{-2} (P\psi)(\omega) \in \mathbf{L}^1(\mathbb{R})$  by assumption. Thus, to conclude  $\Theta_k < \infty$ , it only remains to show  $\omega^{-1} (P\psi)'(\omega) \in \mathbf{L}^1(\mathbb{R})$ . Since  $(P\psi)'$  is continuous and decays faster than  $\omega^2$ , only the integrability around the origin needs to be verified. We note that  $\int \frac{|\widehat{\psi}(\omega)|^2}{\omega^2} d\omega < \infty$  and  $P\psi$  continuous implies  $P\psi \sim \omega^{1+\epsilon}$  for some  $\epsilon > 0$  as  $\omega \rightarrow 0$ . Thus,  $(P\psi)' \sim \omega^\epsilon$  as  $\epsilon \rightarrow 0$ , so that  $\omega^{-1} (P\psi)' \sim \omega^{\epsilon-1}$ ; the function is thus integrable around the origin since  $\epsilon-1 > -1$ .  $\square$

LEMMA 4.3 (Low frequency bound). Assume  $P\psi \in \mathbf{C}^m(\mathbb{R})$  and  $f \in \mathbf{L}^1(\mathbb{R})$ . Then the quantity  $|\lambda^m(Sf)^{(m)}(\lambda)|$  can be bounded uniformly over all  $\lambda$ . Specifically,

$$|\lambda^m(Sf)^{(m)}(\lambda)| \leq \Psi_m \|f\|_1^2$$

for  $\Psi_m$  defined in (A.1).

*Proof.* Let  $g(\omega) = (P\psi)(\omega) = |\widehat{\psi}(\omega)|^2$ , and let

$$g_\lambda(\omega) := \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) = |\widehat{\psi}_\lambda(\omega)|^2.$$

Utilizing Definition 2.2 we obtain

$$\lambda^m(Sf)^{(m)}(\lambda) = \frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \left[ \lambda^m \frac{d^m}{d\lambda^m} g_\lambda(\omega) \right] d\omega.$$

Expanding the derivative gives

$$\begin{aligned} \lambda^m \frac{d^m}{d\lambda^m} g_\lambda(\omega) &= C_{m,0} g_\lambda(\omega) + C_{m,1} \omega g'_\lambda(\omega) + C_{m,2} \omega^2 g''_\lambda(\omega) + \dots + C_{m,m} \omega^m g_\lambda^{(m)}(\omega), \\ C_{m,i} &= (-1)^m \binom{m}{i} \frac{m!}{i!}. \end{aligned}$$

Utilizing  $\|\widehat{f}\|_\infty \leq \|f\|_1$  and  $g_\lambda^{(i)}(\omega) = \frac{1}{\lambda^{i+1}} g^{(i)}\left(\frac{\omega}{\lambda}\right)$ , one obtains

$$\begin{aligned} |\lambda^m(Sf)^{(m)}(\lambda)| &\leq \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\widehat{f}(\omega)|^2 |\omega^i g_\lambda^{(i)}(\omega)| d\omega \\ &\leq \|f\|_1^2 \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega^i g_\lambda^{(i)}(\omega)| d\omega \\ &= \|f\|_1^2 \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega^i g^{(i)}(\omega)| d\omega \\ &= \|f\|_1^2 \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \cdot \|\omega^i g^{(i)}(\omega)\|_1 \\ &= \Psi_m \|f\|_1^2. \end{aligned}$$

□

LEMMA 4.4 (High frequency bound for differentiable functions). Assume  $P\psi \in \mathbf{C}^m(\mathbb{R})$ , and  $f' \in \mathbf{L}^1(\mathbb{R})$ . Then the quantity  $|\lambda^m(Sf)^{(m)}(\lambda)|$  can be bounded by

$$|\lambda^m(Sf)^{(m)}(\lambda)| \leq \frac{\Theta_m}{\lambda^2} \|f'\|_1^2$$

for  $\Theta_m$  defined in (A.2).

*Proof.* Recall from the proof of Lemma 4.3 that

$$|\lambda^m (Sf)^{(m)}(\lambda)| \leq \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\widehat{f}(\omega)|^2 |\omega^i g_\lambda^{(i)}(\omega)| d\omega$$

where  $g_\lambda(\omega) = \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) = |\widehat{\psi}_\lambda(\omega)|^2$  and  $C_{m,i} = (-1)^m \binom{m}{i} \frac{m!}{i!}$ . Since  $\|\omega \widehat{f}(\omega)\|_\infty \leq \|f'\|_1$  and  $g_\lambda^{(i)}(\omega) = \frac{1}{\lambda^{i+1}} g^{(i)}\left(\frac{\omega}{\lambda}\right)$ , we obtain

$$\begin{aligned} |\lambda^m (Sf)^{(m)}(\lambda)| &\leq \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega \widehat{f}(\omega)|^2 |\omega^{i-2} g_\lambda^{(i)}(\omega)| d\omega \\ &\leq \|f'\|_1^2 \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega^{i-2} g_\lambda^{(i)}(\omega)| d\omega \\ &= \frac{\|f'\|_1^2}{\lambda^2} \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega^{i-2} g^{(i)}(\omega)| d\omega \\ &= \frac{\|f'\|_1^2}{\lambda^2} \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \cdot \|\omega^{i-2} g^{(i)}(\omega)\|_1 \\ &= \frac{\Theta_m}{\lambda^2} \|f'\|_1^2. \end{aligned}$$

□

LEMMA B.2 Assume  $Pf \in C^0(\mathbb{R})$  and  $\psi$  is  $m$ -admissible, and let  $B_m, E, \Psi_m, \Theta_m$  be as defined in (4.1), (4.3), (A.1) (A.2). Then,

$$\frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \cdot \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \leq (E\eta)^m \Lambda_m(\lambda),$$

where

$$\Lambda_m(\lambda) = \left( \|f\|_1^2 \Psi_m \wedge \frac{\|f'\|_1^2 \Theta_m}{\lambda^2} \right).$$

*Proof.* From the proof of Lemma 4.3:

$$\frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \cdot \left| \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \leq \Psi_m \|f\|_1^2.$$

From the proof of Lemma 4.4:

$$\frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \cdot \left| \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \leq \Theta_m \frac{\|f'\|_1^2}{\lambda^2}.$$

Utilizing  $|B_m| \leq E^m$  gives

$$\frac{1}{2\pi} \int |\widehat{f}(\omega)|^2 \cdot \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \leq (E\eta)^m \left( \|f\|_1^2 \Psi_m \wedge \frac{\|f'\|_1^2 \Theta_m}{\lambda^2} \right).$$

□

The following Corollary is obtained from Lemma B.2 when  $f$  is a dirac-delta function.

COROLLARY B.1 Assume  $\psi$  is  $m$ -admissible, and let  $B_m, E, \Psi_m$  be as defined in (4.1), (4.3), (A1). Then,

$$\frac{1}{2\pi} \int \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \leq (E\eta)^m \Psi_m.$$

### C. Power spectrum and wavelet invariant equivalence

This appendix contains supporting results for demonstrating the equivalence of the power spectrum and wavelet invariants. Lemma 2.1 establishes that wavelet invariants uniquely determine any bandlimited  $L^2$  function, as long as the wavelet satisfies the linear independence Condition 2.3 and a mild integrability condition. Proposition 2.1 gives two criteria that are sufficient to guarantee Condition 2.3. Finally, Lemma C.1 establishes that the Morlet wavelet satisfies Condition 2.3.

LEMMA 2.1 Let  $p \in L^2(\mathbb{R})$  be continuous and assume  $p(w) = p(-w)$ ,  $\widehat{\psi}$  has compact support and Condition 2.3. Then

*Proof.* Since  $p$  is continuous, there exists an  $\epsilon > 0$  such that on  $(0, \epsilon)$  one either has  $p = 0$ ,  $p > 0$ , or  $p < 0$ . Claim: one must have  $p = 0$ . Suppose not, and without loss of generality assume  $p > 0$  on  $(0, \epsilon)$  and that the support of  $|\widehat{\psi}^+(\omega)|^2$  is contained in the interval  $[1, 2]$ . Now choose  $\lambda_0$  small enough so that  $|\widehat{\psi}_{\lambda_0}^+(\omega)|^2$  is supported on  $[\epsilon/2, \epsilon]$ , i.e.  $\lambda_0 = \epsilon/2$ . Clearly, there must exist a subset  $\mathcal{M} \subseteq [\epsilon/2, \epsilon]$  of positive measure such that  $|\widehat{\psi}_{\lambda_0}^+(\omega)|^2 > 0$  on  $\mathcal{M}$ . Then,

$$0 = \int_0^\infty p(\omega) |\widehat{\psi}_{\lambda_0}^+(\omega)|^2 d\omega = \int_{\epsilon/2}^\epsilon p(\omega) |\widehat{\psi}_{\lambda_0}^+(\omega)|^2 d\omega \geq \int_{\mathcal{M}} p(\omega) |\widehat{\psi}_{\lambda_0}^+(\omega)|^2 d\omega \geq 0.$$

We conclude

$$\int_{\mathcal{M}} p(\omega) |\widehat{\psi}_{\lambda_0}^+(\omega)|^2 d\omega = 0,$$

but this is impossible since the integrand is strictly positive on  $\mathcal{M}$ . We thus conclude that  $p = 0$  on  $(0, \epsilon)$ . Thus, it is sufficient to only consider frequencies  $[\epsilon, \infty)$ .

Assume  $\int p(\omega) |\widehat{\psi}_\lambda(\omega)|^2 d\omega = 0$  for all  $\lambda$ . Since  $p(\omega) = p(-\omega)$ ,

$$\int p(\omega) |\widehat{\psi}_\lambda(\omega)|^2 d\omega = \int_0^\infty p(\omega) |\widehat{\psi}_\lambda^+(\omega)|^2 d\omega = \int_\epsilon^\infty p(\omega) |\widehat{\psi}_\lambda^+(\omega)|^2 d\omega = \langle p, |\widehat{\psi}_\lambda^+|^2 \rangle_I = 0 \quad \forall \lambda,$$

where  $I = [\epsilon, \infty)$ . We now define  $|\widehat{\phi}_\lambda^+(\omega)|^2 := \lambda^{-\beta} |\widehat{\psi}_\lambda^+(\omega)|^2$  for some  $\beta > 0$ , and observe that

$$\int_0^\infty p(\omega) |\widehat{\phi}_\lambda^+(\omega)|^2 d\omega = 0 \quad \forall \lambda \quad \implies \quad \int_0^\infty |\langle p, |\widehat{\phi}_\lambda^+|^2 \rangle_{\mathbb{R}^+}|^2 d\lambda = \int_0^\infty |\langle p, |\widehat{\phi}_\lambda^+|^2 \rangle_I|^2 d\lambda = 0.$$

Note

$$\begin{aligned} \int_0^\infty |\langle p, |\widehat{\phi}_\lambda^+|^2 \rangle_I|^2 d\lambda &= \int_0^\infty \langle p, |\widehat{\phi}_\lambda^+|^2 \rangle_I \langle \overline{p}, |\widehat{\phi}_\lambda^+|^2 \rangle_I d\lambda \\ &= \int_0^\infty \left( \int_I p(\omega_1) |\widehat{\phi}_\lambda^+(\omega_1)|^2 d\omega_1 \right) \left( \int_I \overline{p(\omega_2)} |\widehat{\phi}_\lambda^+(\omega_2)|^2 d\omega_2 \right) d\lambda \\ &= \int_I \overline{p(\omega_2)} \left( \int_I p(\omega_1) \left( \int_0^\infty |\widehat{\phi}_\lambda^+(\omega_1)|^2 |\widehat{\phi}_\lambda^+(\omega_2)|^2 d\lambda \right) d\omega_1 \right) d\omega_2. \end{aligned}$$

We now apply the change of variable  $\omega_i = 1/\xi_i$ , and let  $g(\xi_i) = p(1/\xi_i)$ . We obtain

$$0 = \int_0^{1/\epsilon} \overline{g(\xi_2)} \left( \int_0^{1/\epsilon} g(\xi_1) \left( \int_0^\infty \frac{1}{\xi_1^2 \xi_2^2} \left| \widehat{\phi}_\lambda^+ \left( \frac{1}{\xi_1} \right) \right|^2 \left| \widehat{\phi}_\lambda^+ \left( \frac{1}{\xi_2} \right) \right|^2 d\lambda \right) d\xi_1 \right) d\xi_2. \quad (\text{C.1})$$

Now consider the kernel

$$k(\xi_1, \xi_2) = \int_0^\infty \frac{1}{\xi_1^2 \xi_2^2} \left| \widehat{\phi}_\lambda^+ \left( \frac{1}{\xi_1} \right) \right|^2 \left| \widehat{\phi}_\lambda^+ \left( \frac{1}{\xi_2} \right) \right|^2 d\lambda.$$

Note that  $k$  is a strictly positive definite kernel function if for any finite sequence  $\{\xi_i\}_{i=1}^n$  in  $[0, 1/\epsilon]$ , the  $n$  by  $n$  matrix  $A$  defined by

$$A_{ij} = k(\xi_i, \xi_j)$$

is strictly positive definite [83]. Viewing  $\tilde{\xi}_i(\lambda) = \xi_i^{-2} |\widehat{\phi}_\lambda^+(1/\xi_i)|^2$  as functions of  $\lambda$ , we see that

$$A_{ij} = \langle \tilde{\xi}_i(\lambda), \tilde{\xi}_j(\lambda) \rangle_{\mathbb{R}^+}$$

and  $A$  is thus a Gram matrix. Since the  $\tilde{\xi}_i(\lambda)$  are linearly independent if and only if the  $|\widehat{\psi}_\lambda^+(\omega_i)|^2$  are linearly independent, and the  $|\widehat{\psi}_\lambda^+(\omega_i)|^2$  are linearly independent by assumption, we can conclude that  $A$  and thus  $k$  are strictly positive definite. Now consider the corresponding integral operator on  $[0, 1/\epsilon]$ :

$$Kg(\xi_2) = \int_0^{1/\epsilon} g(\xi_1) k(\xi_1, \xi_2) d\xi_1.$$

Since  $\psi \in \mathbf{L}^1(\mathbb{R})$ ,  $|\widehat{\psi}_\lambda^+|^2$  and thus  $|\widehat{\phi}_\lambda^+|^2$  are continuous, and  $k$  will thus be continuous as long as it remains bounded. To check boundedness we observe that  $k(\xi_1, \xi_2)^2 \leq k(\xi_1, \xi_1)k(\xi_2, \xi_2)$  [21], and

$$\begin{aligned} k(\xi, \xi) &= \int_0^\infty \frac{1}{\xi^4} \left| \widehat{\phi}_\lambda^+ \left( \frac{1}{\xi} \right) \right|^4 d\lambda \\ &= \int_0^\infty \frac{1}{\xi^4} \frac{1}{\lambda^{2+2\beta}} \left| \widehat{\psi}^+ \left( \frac{1}{\lambda\xi} \right) \right|^4 d\lambda \\ &= \int_0^\infty \frac{1}{\xi^4} (\omega\xi)^{2+2\beta} |\widehat{\psi}^+(\omega)|^4 \frac{d\omega}{\xi\omega^2} \\ &= \xi^{2\beta-3} \int_0^\infty \omega^{2\beta} |\widehat{\psi}^+(\omega)|^4 d\omega \\ &\leq 3\xi^{2\beta-3} \int_0^\infty \omega^{2\beta} |\widehat{\psi}(\omega)|^4 d\omega \\ &\leq 3\xi^{2\beta-3} \|\omega^\beta P\psi\|_2^2. \end{aligned}$$

Since  $\widehat{\psi}$  has a compact support, clearly  $\|\omega^\beta P\psi\|_2^2 < \infty$ , and  $k$  is thus bounded on the compact interval  $[0, 1/\epsilon]$  as long as  $\beta \geq 3/2$ . Since  $k$  is continuous and  $[0, 1/\epsilon]$  is compact,  $K : \mathbf{L}^2[0, 1/\epsilon] \rightarrow \mathbf{L}^2[0, 1/\epsilon]$  is a compact, self-adjoint operator and by Mercer's Theorem  $K$  is also strictly positive definite [83]. Since  $\langle Kg, g \rangle_{[0, 1/\epsilon]} = 0$  by (C.1), we conclude  $g = 0$  in  $\mathbf{L}^2[0, 1/\epsilon]$ . Thus,  $p(1/\xi) = 0$  for almost every  $\xi \in (0, 1/\epsilon]$ , which implies  $p(\omega) = 0$  for almost every  $\omega \in [\epsilon, \infty)$ . Since  $p(\omega) = p(-\omega)$  and  $p = 0$  on  $(0, \epsilon)$ ,  $p = 0$  for almost every  $\omega \in \mathbb{R}$ .  $\square$

PROPOSITION 2.1 The following are sufficient to guarantee Condition 2.3:

- (i)  $|\widehat{\psi}(\omega)|^2$  has a compact support contained in the interval  $[a, b]$ , where  $a$  and  $b$  have the same sign, e.g. complex analytic wavelets with compactly supported Fourier transform.
- (ii)  $|\widehat{\psi}(\omega)|^2 \in \mathbf{C}^\infty(\mathbb{R})$  and there exists an  $N$  such that all derivatives of order at least  $N$  are non-zero at  $\omega = 0$ , e.g. the Morlet wavelet.

*Proof.* Let  $\{\omega_i\}_{i=1}^n$  be a finite sequence of distinct positive frequencies, and let  $\tilde{\omega}_i(\lambda) = \frac{1}{|\lambda|} |\widehat{\psi}^+(\frac{\omega_i}{\lambda})|^2$  denote the corresponding functions of  $\lambda$ .

First assume (i). Without loss of generality we assume that  $[a, b]$  is a positive interval and that  $|\widehat{\psi}(\omega)|^2 > 0$  on  $(a, a + \epsilon)$  for some  $\epsilon > 0$ . Clearly,  $|\widehat{\psi}^+(\omega)|^2 = |\widehat{\psi}(\omega)|^2$ . A simple calculation shows that the support of  $\tilde{\omega}_i(\lambda)$  is contained in the interval  $[\frac{\omega_i}{b}, \frac{\omega_i}{a}]$ , and  $\tilde{\omega}_i(\lambda) > 0$  in a neighborhood of  $\frac{\omega_i}{a}$ . Assume we have ordered the  $\omega_i$  so that  $\omega_1 > \dots > \omega_n > 0$ . Now suppose

$$c_1 \tilde{\omega}_1(\lambda) + \dots + c_n \tilde{\omega}_n(\lambda) = 0.$$

Note  $\tilde{\omega}_1(\lambda)$  is the only function in the above collection with support in a neighborhood of  $\frac{\omega_1}{a}$ ; thus, we must have  $c_1 = 0$ , so that

$$c_2 \tilde{\omega}_2(\lambda) + \dots + c_n \tilde{\omega}_n(\lambda) = 0.$$

But now  $\tilde{\omega}_2(\lambda)$  is the only function in the above collection with support in a neighborhood of  $\frac{\omega_2}{a}$ , so we must have  $c_2 = 0$ , and proceeding iteratively we conclude that  $c_1 = \dots = c_n = 0$ . Thus,  $\{\tilde{\omega}_i(\lambda)\}_{i=1}^n$  is a linearly independent set, and Condition 2.3 holds.

Now assume (ii). Since  $\frac{d^n}{d\omega^n} (|\widehat{\psi}^+(\omega)|^2) |_{\omega=0} = 2 \frac{d^n}{d\omega^n} (|\widehat{\psi}(\omega)|^2) |_{\omega=0}$ ,  $|\widehat{\psi}^+(\omega)|^2$  is  $\mathbf{C}^\infty(\mathbb{R})$  and all derivatives of order at least  $N$  are non-zero at  $\omega = 0$ . Note  $\{\tilde{\omega}_i(\lambda)\}_{i=1}^n = \{|\lambda|^{-1} |\widehat{\psi}^+(\omega_i/\lambda)|^2\}_{i=1}^n$  are linearly independent if and only if  $\{|\widehat{\psi}^+(\omega_i/\lambda)|^2\}_{i=1}^n$  are linearly independent. Defining  $\tilde{\lambda} = 1/\lambda$ , this holds if and only if  $\{|\widehat{\psi}^+(\omega_i \tilde{\lambda})|^2\}_{i=1}^n = \{g(\omega_i \tilde{\lambda})\}_{i=1}^n$  are linearly independent as functions of  $\tilde{\lambda}$ , where we define  $g(\omega) = |\widehat{\psi}^+(\omega)|^2$ . Assume

$$c_1 g(\omega_1 \tilde{\lambda}) + c_2 g(\omega_2 \tilde{\lambda}) + \dots + c_n g(\omega_n \tilde{\lambda}) = 0.$$

Differentiating  $m$  times for  $N \leq m \leq N+n-1$ , we obtain

$$\begin{aligned} c_1 \omega_1^N g^{(N)}(\omega_1 \tilde{\lambda}) + \dots + c_n \omega_n^N g^{(N)}(\omega_n \tilde{\lambda}) &= 0 \\ &\vdots \\ c_1 \omega_1^{N+n-1} g^{(N+n-1)}(\omega_1 \tilde{\lambda}) + \dots + c_n \omega_n^{N+n-1} g^{(N+n-1)}(\omega_n \tilde{\lambda}) &= 0. \end{aligned}$$

The above holds for all  $\tilde{\lambda}$ . We now take the limit as  $\tilde{\lambda} \rightarrow 0$  to obtain

$$\begin{aligned} g^{(N)}(0)(\omega_1^N c_1 + \omega_2^N c_2 + \dots + \omega_n^N c_n) &= 0 \\ g^{(N+1)}(0)(\omega_1^{N+1} c_1 + \omega_2^{N+1} c_2 + \dots + \omega_n^{N+1} c_n) &= 0 \\ &\vdots \\ g^{(N+n-1)}(0)(\omega_1^{N+n-1} c_1 + \omega_2^{N+n-1} c_2 + \dots + \omega_n^{N+n-1} c_n) &= 0. \end{aligned}$$

Since  $g^{(m)}(0) \neq 0$ , we obtain

$$\underbrace{\begin{bmatrix} \omega_1^N & \dots & \omega_n^N \\ \omega_1^{N+1} & \dots & \omega_n^{N+1} \\ \vdots & & \vdots \\ \omega_1^{N+n-1} & \dots & \omega_n^{N+n-1} \end{bmatrix}}_{:=A} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}}_{:=B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ \omega_1 & \dots & \omega_n \\ \vdots & & \vdots \\ \omega_1^{(n-1)} & \dots & \omega_n^{(n-1)} \end{bmatrix}}_{:=A} \underbrace{\begin{bmatrix} \omega_1^N & 0 & \dots & 0 \\ 0 & \omega_2^N & \dots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \dots & \omega_n^N \end{bmatrix}}_{:=B} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since  $A$  is a Vandermonde matrix constructed from distinct  $\omega_i$ ,  $\det(A) \neq 0$ . Since the  $\omega_i$  are non-zero,  $\det(B) \neq 0$ . Thus,  $\det(AB) = \det(A)\det(B) \neq 0$ . We conclude  $AB$  is invertible and so all  $c_i = 0$ , which gives Condition 2.3.  $\square$

LEMMA C.1 Suppose we construct a Morlet wavelet with parameter  $\xi$ , that is  $\psi(x) = C_\xi \pi^{-1/4} e^{-x^2/2} (e^{i\xi x} - e^{-\xi^2/2})$  for  $C_\xi = (1 - e^{-\xi^2} - 2e^{-3\xi^2/4})^{-1/2}$ . Then, for almost all  $\xi \in \mathbb{R}^+$ , the wavelet satisfies Condition 2.3.

*Proof.* The Fourier transform  $\hat{\psi}$  has form

$$\hat{\psi}(\omega) = \tilde{C}_\xi e^{-\omega^2/2} (e^{\xi\omega} - 1)$$

for some constant  $\tilde{C}_\xi$  depending on  $\xi$ , so that

$$g(\omega) := \tilde{C}_\xi^{-2} |\hat{\psi}(\omega)|^2 = e^{-\omega^2} (e^{\xi\omega} - 1)^2.$$

From direct calculation or a computer algebra system, one obtains

$$g^{(n)}(0) = \begin{cases} H_n(\xi) - 2H_n(\xi/2) & n \text{ odd} \\ H_n(\xi) - 2H_n(\xi/2) + \frac{(-1)^{\frac{n}{2}} n!}{(\frac{n}{2})!} & n \text{ even} \end{cases}$$

where  $H_n(\xi)$  is the  $n^{\text{th}}$  degree physicist's Hermite polynomial. We have  $g'(0) = 0$ , but for  $n > 1$ ,  $g^{(n)}(0) = 0$  only when  $\xi$  is a root of the above polynomial. Since the set of roots of the polynomials  $\{g^{(n)}(0)\}_{n=1}^\infty$  is countable, if  $\xi$  is selected at random from  $\mathbb{R}$ , it is not a root of any of these polynomials with probability 1, and  $g^{(n)}(0) \neq 0$  for all  $n$ . Thus, the wavelet satisfies criterion (ii) of Proposition 2.1, and thus the linear independence Condition 2.3.  $\square$

## D. Supporting results: classic MRA

This appendix contains supporting results for Section 3. The first two lemmas (Lemmas D.1 and Lemma D.2) establish additive noise bounds for the power spectrum and are needed to prove Proposition 3.1. The next two lemmas (Lemmas D.3 and Lemma D.4) establish additive noise bounds for wavelet invariants and are needed to prove Proposition 3.2.

LEMMA D.1 Let  $\varepsilon(x)$  be a white noise processes on  $[-\frac{1}{2}, \frac{1}{2}]$  with variance  $\sigma^2$ . Then, for all frequencies  $\omega, \xi$ ,

$$\mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^2 \right] = \sigma^2 \quad (\text{D.1})$$

$$\mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^4 \right] \leq 3\sigma^4 \quad (\text{D.2})$$

$$\mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^2 |\widehat{\varepsilon}(\xi)|^2 \right] \leq 3\sigma^4. \quad (\text{D.3})$$

*Proof.* By Proposition J.1,

$$\begin{aligned} \mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^2 \right] &= \mathbb{E} \left[ \widehat{\varepsilon}(\omega) \overline{\widehat{\varepsilon}(\omega)} \right] \\ &= \mathbb{E} \left[ \left( \int_{-1/2}^{1/2} e^{-i\omega x} dB_x \right) \left( \int_{-1/2}^{1/2} e^{i\omega x} dB_x \right) \right] \\ &= \sigma^2 \int_{-1/2}^{1/2} dx \\ &= \sigma^2, \end{aligned}$$

which shows (D.1). By Proposition J.2,

$$\begin{aligned} \mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^4 \right] &= \mathbb{E} \left[ \widehat{\varepsilon}(\omega)^2 \left( \overline{\widehat{\varepsilon}(\omega)} \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \int_{-1/2}^{1/2} e^{-i\omega x} dB_x \right)^2 \left( \int_{-1/2}^{1/2} e^{i\omega x} dB_x \right)^2 \right] \\ &= 2\sigma^4 \left( \int_{-1/2}^{1/2} dx \right)^2 + \sigma^4 \left( \int_{-1/2}^{1/2} e^{-2i\omega x} dx \right) \left( \int_{-1/2}^{1/2} e^{2i\omega x} dx \right) \\ &\leq 2\sigma^4 + \sigma^4 \left( \int_{-1/2}^{1/2} |e^{-2i\omega x}| dx \right) \left( \int_{-1/2}^{1/2} |e^{2i\omega x}| dx \right) \\ &= 3\sigma^4, \end{aligned}$$

which shows (D.2). Finally, by Proposition J.3, we have

$$\begin{aligned} \mathbb{E} \left[ |\widehat{\varepsilon}(\omega)|^2 |\widehat{\varepsilon}(\xi)|^2 \right] &= \mathbb{E} \left[ \left( \int_{-1/2}^{1/2} e^{-i\omega x} dB_x \right) \left( \int_{-1/2}^{1/2} e^{i\omega x} dB_x \right) \left( \int_{-1/2}^{1/2} e^{-i\xi x} dB_x \right) \left( \int_{-1/2}^{1/2} e^{i\xi x} dB_x \right) \right] \\ &= \sigma^4 \left[ \left( \int_{-1/2}^{1/2} e^{-i(\omega+\xi)x} dx \right) \left( \int_{-1/2}^{1/2} e^{i(\omega+\xi)x} dx \right) \right] \\ &\quad + \sigma^4 \left[ \left( \int_{-1/2}^{1/2} e^{i(\xi-\omega)x} dx \right) \left( \int_{-1/2}^{1/2} e^{i(\omega-\xi)x} dx \right) + \left( \int_{-1/2}^{1/2} dx \right) \left( \int_{-1/2}^{1/2} dx \right) \right] \\ &\leq \sigma^4 \left[ 3 \left( \int_{-1/2}^{1/2} dx \right) \left( \int_{-1/2}^{1/2} dx \right) \right] \\ &= 3\sigma^4, \end{aligned}$$

which gives (D.3).  $\square$

LEMMA D.2 Let  $\varepsilon(x)$  be a white noise processes on  $[-\frac{1}{2}, \frac{1}{2}]$  with variance  $\sigma^2$ . Then, for any signal  $f \in \mathbf{L}^1(\mathbb{R})$ ,

$$\begin{aligned}\mathbb{E}[(P(f + \varepsilon))(\omega)] &= (Pf)(\omega) + \sigma^2 \\ \text{Var}[(P(f + \varepsilon))(\omega)] &\leq 4\sigma^2(Pf)(\omega) + 2\sigma^4.\end{aligned}$$

*Proof.* Since  $\mathbb{E}[\widehat{\varepsilon}(\omega)] = \mathbb{E}[\overline{\widehat{\varepsilon}(\omega)}] = 0$  and  $\mathbb{E}[|\widehat{\varepsilon}(\omega)|^2] = \sigma^2$  by Lemma D.1,

$$\begin{aligned}\mathbb{E}[(P(f + \varepsilon))(\omega)] &= \mathbb{E}\left[\left(\widehat{f}(\omega) + \widehat{\varepsilon}(\omega)\right)\left(\overline{\widehat{f}(\omega)} + \overline{\widehat{\varepsilon}(\omega)}\right)\right] \\ &= \mathbb{E}\left[|\widehat{f}(\omega)|^2 + \widehat{f}(\omega)\overline{\widehat{\varepsilon}(\omega)} + \widehat{\varepsilon}(\omega)\overline{\widehat{f}(\omega)} + |\widehat{\varepsilon}(\omega)|^2\right] \\ &= (Pf)(\omega) + \sigma^2.\end{aligned}$$

We now control  $\text{Var}[(P(f + \varepsilon))(\omega)]$ . Note that:

$$[(P(f + \varepsilon))(\omega)]^2 = \left(|\widehat{f}(\omega)|^2 + \widehat{f}(\omega)\overline{\widehat{\varepsilon}(\omega)} + \widehat{\varepsilon}(\omega)\overline{\widehat{f}(\omega)} + |\widehat{\varepsilon}(\omega)|^2\right)^2$$

and that

$$\begin{aligned}\mathbb{E}\left[|\widehat{\varepsilon}(\omega)|^2\widehat{\varepsilon}(\omega)\right] &= \mathbb{E}\left[\left(\int_{-1/2}^{1/2} e^{-i\omega x} dB_x\right)\left(\int_{-1/2}^{1/2} e^{i\omega s} dB_s\right)\left(\int_{-1/2}^{1/2} e^{-i\omega p} dB_p\right)\right] \\ &= 0,\end{aligned}$$

since even when  $x = s = p$ ,  $\mathbb{E}[(\Delta B_x)^3] = 0$ . Ignoring the terms with zero expectation, we thus get

$$\begin{aligned}\mathbb{E}[(P(f + \varepsilon))(\omega)^2] &= \mathbb{E}\left(|\widehat{f}(\omega)|^4 + 4|\widehat{f}(\omega)|^2|\widehat{\varepsilon}(\omega)|^2 + |\widehat{\varepsilon}(\omega)|^4 + \widehat{f}(\omega)^2\overline{\widehat{\varepsilon}(\omega)}^2 + \widehat{\varepsilon}(\omega)^2\overline{\widehat{f}(\omega)}^2\right) \\ &\leq \mathbb{E}\left(|\widehat{f}(\omega)|^4 + 6|\widehat{f}(\omega)|^2|\widehat{\varepsilon}(\omega)|^2 + |\widehat{\varepsilon}(\omega)|^4\right) \\ &= [(Pf)(\omega)]^2 + 6\sigma^2(Pf)(\omega) + 3\sigma^4\end{aligned}$$

where the last line follows from Lemma D.1. Thus,

$$\begin{aligned}\text{Var}[(P(f + \varepsilon))(\omega)] &= \mathbb{E}[(P(f + \varepsilon))(\omega)^2] - (\mathbb{E}[(P(f + \varepsilon))(\omega)])^2 \\ &\leq [(Pf)(\omega)]^2 + 6\sigma^2(Pf)(\omega) + 3\sigma^4 - ((Pf)(\omega) + \sigma^2)^2 \\ &= 4\sigma^2(Pf)(\omega) + 2\sigma^4.\end{aligned}$$

□

PROPOSITION 3.1 Assume Model 1. Define the following estimator of  $(Pf)(\omega)$ :

$$(\widetilde{P}f)(\omega) := \frac{1}{M} \sum_{j=1}^M (Py_j)(\omega) - \sigma^2.$$

Then, with probability at least  $1 - 1/t^2$ ,

$$|(Pf)(\omega) - (\tilde{P}f)(\omega)| \leq \frac{2t\sigma}{\sqrt{M}} (\|f\|_1 + \sigma). \quad (3.1)$$

*Proof.* Let  $f^{t_j}(x) = f(x - t_j)$  so that  $y_j = f^{t_j} + \varepsilon_j$ . We first note since  $\widehat{f^{t_j}}(\omega) = e^{-i\omega t_j} \widehat{f}(\omega)$ , the power spectrum is translation invariant, that is  $(Pf^{t_j})(\omega) = (Pf)(\omega)$  for all  $\omega, t_j$ . Thus, by Lemma D.2,

$$\mathbb{E}[(Py_j)(\omega)] = \mathbb{E}[(P(f^{t_j} + \varepsilon_j))(\omega)] = (Pf^{t_j})(\omega) + \sigma^2 = (Pf)(\omega) + \sigma^2$$

and

$$\text{Var}[(Py_j)(\omega)] = \text{Var}[(P(f^{t_j} + \varepsilon_j))(\omega)] \leq 4\sigma^2(Pf^{t_j})(\omega) + 2\sigma^4 = 4\sigma^2(Pf)(\omega) + 2\sigma^4.$$

Since the  $y_j$  are independent,

$$\text{Var}\left(\frac{1}{M} \sum_{j=1}^M (Py_j)(\omega)\right) \leq \frac{1}{M} (4\sigma^2(Pf)(\omega) + 2\sigma^4).$$

Applying Chebyshev's inequality to the random variable  $X = \frac{1}{M} \sum_{j=1}^M (Py_j)(\omega)$ , we obtain

$$\mathbb{P}\left(\left|\frac{1}{M} \sum_{j=1}^M (Py_j)(\omega) - ((Pf)(\omega) + \sigma^2)\right| \geq \frac{t(2\sigma\sqrt{(Pf)(\omega)} + \sqrt{2}\sigma^2)}{\sqrt{M}}\right) \leq \frac{1}{t^2}.$$

Observing that  $\sqrt{(Pf)(\omega)} = |\widehat{f}(\omega)| \leq \|f\|_1$  gives (3.1). □

LEMMA D.3 Let  $\varepsilon(x)$  be a white noise processes on  $[-\frac{1}{2}, \frac{1}{2}]$  with variance  $\sigma^2$ . Then,

$$\mathbb{E}[(S\varepsilon)(\lambda)] = \sigma^2$$

$$\mathbb{E}[(S\varepsilon)(\lambda)^2] \leq 3\sigma^4.$$

*Proof.* Since  $\mathbb{E}[|\widehat{\varepsilon}(\omega)|^2] = \sigma^2$  by Lemma D.1, we have

$$\begin{aligned} \mathbb{E}[(S\varepsilon)(\lambda)] &= \mathbb{E}\left[\|\varepsilon * \psi_\lambda\|_2^2\right] \\ &= \mathbb{E}\left[\frac{1}{2\pi} \|\widehat{\varepsilon} \cdot \widehat{\psi}_\lambda\|_2^2\right] \\ &= \mathbb{E}\left[\frac{1}{2\pi} \int |\widehat{\varepsilon}(\omega)|^2 |\widehat{\psi}_\lambda(\omega)|^2 d\omega\right] \\ &= \frac{\sigma^2}{2\pi} \int |\widehat{\psi}_\lambda(\omega)|^2 d\omega \\ &= \sigma^2 \|\psi_\lambda\|_2^2 \\ &= \sigma^2. \end{aligned}$$

Since by Lemma D.1,  $\mathbb{E} [ |\widehat{\varepsilon}(\omega)|^2 |\widehat{\varepsilon}(\xi)|^2 ] \leq 3\sigma^4$ , we also have

$$\begin{aligned}
\mathbb{E} [(S\varepsilon)(\lambda)^2] &= \mathbb{E} \left[ \|\varepsilon * \psi_\lambda\|_2^4 \right] \\
&= \mathbb{E} \left[ \frac{1}{(2\pi)^2} \|\widehat{\varepsilon} \cdot \widehat{\psi}_\lambda\|_2^2 \|\widehat{\varepsilon} \cdot \widehat{\psi}_\lambda\|_2^2 \right] \\
&= \mathbb{E} \left[ \frac{1}{(2\pi)^2} \int \int |\widehat{\varepsilon}(\omega)|^2 |\widehat{\varepsilon}(\xi)|^2 |\widehat{\psi}_\lambda(\omega)|^2 |\widehat{\psi}_\lambda(\xi)|^2 d\omega d\xi \right] \\
&\leq \frac{3\sigma^4}{(2\pi)^2} \int \int |\widehat{\psi}_\lambda(\omega)|^2 |\widehat{\psi}_\lambda(\xi)|^2 d\omega d\xi \\
&= 3\sigma^4 \left( \|\psi_\lambda\|_2^2 \right)^2 \\
&= 3\sigma^4.
\end{aligned}$$

□

LEMMA D.4 Let  $\varepsilon(x)$  be a white noise processes on  $[-\frac{1}{2}, \frac{1}{2}]$  with variance  $\sigma^2$ . Then, for any signal  $f \in \mathbf{L}^1(\mathbb{R})$ ,

$$\begin{aligned}
\mathbb{E}[(S(f + \varepsilon))(\lambda)] &= (Sf)(\lambda) + \sigma^2 \\
\text{Var}[(S(f + \varepsilon))(\lambda)] &\leq 4\sigma^2(Sf)(\lambda) + 2\sigma^4.
\end{aligned}$$

*Proof.* Utilizing  $\mathbb{E}[\varepsilon] = \mathbb{E}[\bar{\varepsilon}] = 0$  and Lemma D.3, we have

$$\begin{aligned}
\mathbb{E}[(S(f + \varepsilon))(\lambda)] &= \mathbb{E} \left[ \int |(f + \varepsilon) * \psi_\lambda(u)|^2 du \right] \\
&= \int |f * \psi_\lambda(u)|^2 + \mathbb{E} \left[ \int |\varepsilon * \psi_\lambda(u)|^2 du \right] \\
&= (Sf)(\lambda) + \mathbb{E}[(S\varepsilon)(\lambda)] \\
&= (Sf)(\lambda) + \sigma^2.
\end{aligned}$$

To bound  $\mathbb{E}[(S(f + \varepsilon))(\lambda)^2]$ , note that

$$\begin{aligned}
&[(S(f + \varepsilon))(\lambda)]^2 \\
&= \left( \int |f * \psi_\lambda(u_1)|^2 + (\varepsilon * \psi_\lambda(u_1))(\bar{f} * \overline{\psi_\lambda(u_1)}) + (f * \psi_\lambda(u_1))(\bar{\varepsilon} * \overline{\psi_\lambda(u_1)}) + |\varepsilon * \psi_\lambda(u_1)|^2 du_1 \right) \\
&\quad \cdot \left( \int |f * \psi_\lambda(u_2)|^2 + (\varepsilon * \psi_\lambda(u_2))(\bar{f} * \overline{\psi_\lambda(u_2)}) + (f * \psi_\lambda(u_2))(\bar{\varepsilon} * \overline{\psi_\lambda(u_2)}) + |\varepsilon * \psi_\lambda(u_2)|^2 du_2 \right).
\end{aligned}$$

When we take expectation, any term involving one or three  $\varepsilon$  terms disappears, so that

$$\begin{aligned}
\mathbb{E}[(S(f + \varepsilon))(\lambda)^2] &= \mathbb{E} \left[ \int \int |f * \psi_\lambda(u_1)|^2 |f * \psi_\lambda(u_2)|^2 du_1 du_2 \right. \\
&\quad + \int \int |f * \psi_\lambda(u_1)|^2 |\varepsilon * \psi_\lambda(u_2)|^2 du_1 du_2 \\
&\quad + \int \int (\varepsilon * \psi_\lambda(u_1))(\bar{f} * \overline{\psi_\lambda(u_1)})(\varepsilon * \psi_\lambda(u_2))(\bar{f} * \overline{\psi_\lambda(u_2)}) du_1 du_2 \\
&\quad + \int \int (\varepsilon * \psi_\lambda(u_1))(\bar{f} * \overline{\psi_\lambda(u_1)})(f * \psi_\lambda(u_2))(\bar{\varepsilon} * \overline{\psi_\lambda(u_2)}) du_1 du_2 \\
&\quad + \int \int (f * \psi_\lambda(u_1))(\bar{\varepsilon} * \overline{\psi_\lambda(u_1)})(\varepsilon * \psi_\lambda(u_2))(\bar{f} * \overline{\psi_\lambda(u_2)}) du_1 du_2 \\
&\quad + \int \int (f * \psi_\lambda(u_1))(\bar{\varepsilon} * \overline{\psi_\lambda(u_1)})(f * \psi_\lambda(u_2))(\bar{\varepsilon} * \overline{\psi_\lambda(u_2)}) du_1 du_2 \\
&\quad + \int \int |\varepsilon * \psi_\lambda(u_1)|^2 |f * \psi_\lambda(u_2)|^2 du_1 du_2 \\
&\quad \left. + \int \int |\varepsilon * \psi_\lambda(u_1)|^2 |\varepsilon * \psi_\lambda(u_2)|^2 du_1 du_2 \right] \\
&\leq \mathbb{E} \left[ \int \int |f * \psi_\lambda(u_1)|^2 |f * \psi_\lambda(u_2)|^2 du_1 du_2 \right. \\
&\quad + 6 \int \int |f * \psi_\lambda(u_1)|^2 |\varepsilon * \psi_\lambda(u_2)|^2 du_1 du_2 \\
&\quad \left. + \int \int |\varepsilon * \psi_\lambda(u_1)|^2 |\varepsilon * \psi_\lambda(u_2)|^2 du_1 du_2 \right] \\
&= \mathbb{E} \left[ [(Sf)(\lambda)]^2 + 6(Sf)(\lambda)(S\varepsilon)(\lambda) + [(\varepsilon S)(\lambda)]^2 \right] \\
&= [(Sf)(\lambda)]^2 + 6\sigma^2(Sf)(\lambda) + 3\sigma^4,
\end{aligned}$$

where the last line follows from Lemma D.3. Thus,

$$\begin{aligned}
\text{Var}[(S(f + \varepsilon))(\lambda)] &= \mathbb{E}[(S(f + \varepsilon))(\lambda)^2] - (\mathbb{E}[(S(f + \varepsilon))(\lambda)])^2 \\
&\leq [(Sf)(\lambda)]^2 + 6\sigma^2(Sf)(\lambda) + 3\sigma^4 - [(Sf)(\lambda) + \sigma^2]^2 \\
&= 4\sigma^2(Sf)(\lambda) + 2\sigma^4.
\end{aligned}$$

□

PROPOSITION 3.2 Assume Model 1. Define the following estimator of  $(Sf)(\lambda)$ :

$$(\tilde{S}f)(\lambda) := \frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda) - \sigma^2.$$

Then, with probability at least  $1 - 1/t^2$ ,

$$|(Sf)(\lambda) - (\tilde{S}f)(\lambda)| \leq \frac{2t\sigma}{\sqrt{M}}(\|f\|_1 + \sigma). \quad (\text{D.4})$$

*Proof.* Let  $f^{t_j}(x) = f(x - t_j)$  so that  $y_j = f^{t_j} + \varepsilon_j$ . We first note that the wavelet invariants are translation invariant, that is  $Sf^{t_j} = Sf$  for all  $t_j$ . We now compute the mean and variance of the coefficients  $(Sy_j)(\lambda)$ . By Lemma D.4,

$$\mathbb{E}[(Sy_j)(\lambda)] = \mathbb{E}[(S(f^{t_j} + \varepsilon_j))(\lambda)] = (Sf^{t_j})(\lambda) + \sigma^2 = (Sf)(\lambda) + \sigma^2$$

and

$$\text{Var}[(Sy_j)(\lambda)] = \text{Var}[(S(f^{t_j} + \varepsilon_j))(\lambda)] \leq 4\sigma^2(Sf^{t_j})(\lambda) + 2\sigma^4 = 4\sigma^2(Sf)(\lambda) + 2\sigma^4.$$

Since the  $y_j$  are independent,

$$\text{Var}\left[\frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda)\right] \leq \frac{1}{M} [4\sigma^2(Sf)(\lambda) + 2\sigma^4].$$

Applying Chebyshev's inequality to the random variable  $X = \frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda)$  gives

$$\mathbb{P}\left(\left|\frac{1}{M} \sum_{j=1}^M (Sy_j)(\lambda) - [(Sf)(\lambda) + \sigma^2]\right| \geq \frac{t(2\sigma\sqrt{(Sf)(\lambda)} + \sqrt{2}\sigma^2)}{\sqrt{M}}\right) \leq \frac{1}{t^2}.$$

By Young's convolution inequality,  $(Sf)(\lambda) = \|f * \psi_\lambda\|_2^2 \leq \|f\|_1^2 \|\psi_\lambda\|_2^2 = \|f\|_1^2$ , which gives (D.4).  $\square$

## E. Supporting results: dilation MRA

This appendix contains the technical details of the dilation unbiasing procedure that is central to Propositions 4.1, 4.2 and 5.1. Lemma 4.1 bounds the bias and variance of the estimator, and Lemma 4.2 bounds the error of the estimator given  $M$  independent samples.

**LEMMA 4.1** Let  $F_\lambda(\tau) = L((1 - \tau)\lambda)$  for some function  $L \in \mathbf{C}^{k+2}(0, \infty)$  and a random variable  $\tau$  satisfying the assumptions of Section 2.1, and let  $k \geq 2$  be an even integer. Assume there exist functions  $\Lambda_i : \mathbb{R} \rightarrow \mathbb{R}$ ,  $R : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$|\lambda^i L^{(i)}(\lambda)| \leq \Lambda_i(\lambda) \text{ for } 0 \leq i \leq k+2, \quad \frac{\Lambda_{k+2}((1 - \tau)\lambda)}{\Lambda_{k+2}(\lambda)} \leq R(\lambda),$$

and define the following estimator of  $L(\lambda)$ :

$$G_\lambda(\tau) := F_\lambda(\tau) - B_2\eta^2 F_\lambda''(\tau) - B_4\eta^4 F_\lambda^{(4)}(\tau) - \dots - B_k\eta^k F_\lambda^{(k)}(\tau).$$

Then  $G_\lambda(\tau)$  satisfies

$$|\mathbb{E} G_\lambda(\tau) - L(\lambda)| \lesssim kR(\lambda)\Lambda_{k+2}(\lambda)(2E\eta)^{k+2}$$

$$\text{Var } G_\lambda(\tau) \lesssim k^2 R(\lambda)^2 \Lambda(\lambda)^2$$

where

$$\Lambda(\lambda)^2 := \sum_{0 \leq i, j \leq k+2, i+j \geq 2} \Lambda_i(\lambda) \Lambda_j(\lambda) (2E\eta)^{i+j}$$

and  $E$  is the absolute constant defined in (4.3).

*Proof.* We Taylor expand  $F_\lambda(\tau)$  about  $\tau = 0$

$$F_\lambda(\tau) = F_\lambda(0) + F'_\lambda(0)\tau + \frac{F''_\lambda(0)}{2}\tau^2 + \dots + \frac{F^{(k+1)}_\lambda(0)}{(k+1)!}\tau^{k+1} + \underbrace{\int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k+1)!}(\tau-t)^{k+1} dt}_{:=R_0(\tau, \lambda)}.$$

We note

$$\mathbb{E}[F_\lambda(\tau)] = F_\lambda(0) + \frac{F''_\lambda(0)}{2}\eta^2 + \dots + \frac{F^k_\lambda(0)}{k!}C_k\eta^k + \mathbb{E}[R_0(\tau, \lambda)],$$

which motivates an unbiasing with the first  $k/2$  even derivatives, and thus a Taylor expansion of these derivatives

$$\begin{aligned} F_\lambda(\tau) &= F_\lambda(0) + F'_\lambda(0)\tau + \dots + \frac{F^{(k+1)}_\lambda(0)}{(k+1)!}\tau^{k+1} + \underbrace{\int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k+1)!}(\tau-t)^{k+1} dt}_{:=R_0(\tau, \lambda)} \\ F''_\lambda(\tau) &= F''_\lambda(0) + F^{(3)}_\lambda(0)\tau + \dots + \frac{F^{(k+1)}_\lambda(0)}{(k-1)!}\tau^{k-1} + \underbrace{\int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k-1)!}(\tau-t)^{k-1} dt}_{:=R_2(\tau, \lambda)} \\ F^{(4)}_\lambda(\tau) &= F^{(4)}_\lambda(0) + F^{(5)}_\lambda(0)\tau + \dots + \frac{F^{(k+1)}_\lambda(0)}{(k-3)!}\tau^{k-3} + \underbrace{\int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k-3)!}(\tau-t)^{k-3} dt}_{:=R_4(\tau, \lambda)} \\ &\vdots \\ F^{(k)}_\lambda(\tau) &= F^{(k)}_\lambda(0) + F^{(k+1)}_\lambda(0)\tau + \underbrace{\int_0^\tau F^{(k+2)}_\lambda(t)(\tau-t) dt}_{:=R_k(\tau, \lambda)}. \end{aligned}$$

Multiplication of the  $i^{\text{th}}$  even derivative by  $B_i \eta^i$  gives

$$\begin{aligned} F_\lambda(\tau) &= F_\lambda(0) + F'_\lambda(0)\tau + \dots + \frac{F_\lambda^{(k+1)}(0)}{(k+1)!}\tau^{k+1} + R_0(\tau, \lambda) \\ B_2 \eta^2 F_\lambda''(\tau) &= B_2 \eta^2 F_\lambda''(0) + B_2 \eta^2 F_\lambda^{(3)}(0)\tau + \dots + B_2 \eta^2 \frac{F_\lambda^{(k+1)}(0)}{(k-1)!}\tau^{k-1} + B_2 \eta^2 R_2(\tau, \lambda) \\ B_4 \eta^4 F_\lambda^{(4)}(\tau) &= B_4 \eta^4 F_\lambda^{(4)}(0) + B_4 \eta^4 F_\lambda^{(5)}(0)\tau + \dots + B_4 \eta^4 \frac{F_\lambda^{(k+1)}(0)}{(k-3)!}\tau^{k-3} + B_4 \eta^4 R_4(\tau, \lambda) \\ &\vdots \\ B_k \eta^k F_\lambda^{(k)}(\tau) &= B_k \eta^k F_\lambda^{(k)}(0) + B_k \eta^k F_\lambda^{(k+1)}(0)\tau + B_k \eta^k R_k(\tau, \lambda). \end{aligned}$$

We want an estimator that targets  $F_\lambda(0) = L(\lambda)$ . We thus consider the following variable as an estimator:

$$G_\lambda(\tau) := F_\lambda(\tau) - B_2 \eta^2 F_\lambda''(\tau) - B_4 \eta^4 F_\lambda^{(4)}(\tau) - \dots - B_k \eta^k F_\lambda^{(k)}(\tau)$$

and show that  $\mathbb{E}[G_\lambda(\tau)] = F_\lambda(0) + O(\eta^{k+2})$  for constants  $B_i$  chosen according to (4.1). We have

$$\begin{aligned} \mathbb{E}[F_\lambda(\tau)] &= F_\lambda(0) + F_\lambda''(0) \frac{C_2}{2} \eta^2 + \dots + F_\lambda^{(k)}(0) \frac{C_k}{k!} \eta^k + \mathbb{E}[R_0(\tau, \lambda)] \\ \mathbb{E}[B_2 \eta^2 F_\lambda''(\tau)] &= F_\lambda''(0) B_2 \eta^2 + F_\lambda^{(4)}(0) \frac{B_2 C_2}{2} \eta^4 + \dots + F_\lambda^{(k)}(0) \frac{B_2 C_{k-2}}{(k-2)!} \eta^k + \mathbb{E}[B_2 \eta^2 R_2(\tau, \lambda)] \\ \mathbb{E}[B_4 \eta^4 F_\lambda^{(4)}(\tau)] &= F_\lambda^{(4)}(0) B_4 \eta^4 + F_\lambda^{(6)}(0) \frac{B_4 C_2}{2} \eta^6 + \dots + F_\lambda^{(k)}(0) \frac{B_4 C_{k-4}}{(k-4)!} \eta^k + \mathbb{E}[B_4 \eta^4 R_4(\tau, \lambda)] \\ &\vdots \\ \mathbb{E}[B_{k-2} \eta^{k-2} F_\lambda^{(k-2)}(\tau)] &= F_\lambda^{(k-2)}(0) B_{k-2} \eta^{k-2} + F_\lambda^{(k)}(0) \frac{B_{k-2} C_2}{2} \eta^k + \mathbb{E}[B_{k-2} \eta^{k-2} R_{k-2}(\tau, \lambda)] \\ \mathbb{E}[B_k \eta^k F_\lambda^{(k)}(\tau)] &= F_\lambda^{(k)}(0) B_k \eta^k + \mathbb{E}[B_k \eta^k R_k(\tau, \lambda)]. \end{aligned}$$

That is,

$$\begin{aligned} \mathbb{E}[G_\lambda(\tau)] &= F_\lambda(0) + F_\lambda''(0) \left( \frac{C_2}{2!} - B_2 \right) \eta^2 + F_\lambda^{(4)}(0) \left( \frac{C_4}{4!} - \frac{B_2 C_2}{2!} - B_4 \right) \eta^4 \\ &\quad + F_\lambda^{(6)}(0) \left( \frac{C_6}{6!} - \frac{B_2 C_4}{4!} - \frac{B_4 C_2}{2!} - B_6 \right) \eta^6 \dots \\ &\quad + F_\lambda^{(k)}(0) \left( \frac{C_k}{k!} - \frac{B_2 C_{k-2}}{(k-2)!} - \dots - \frac{B_{k-2} C_2}{2!} - B_k \right) \eta^k + H_1(\lambda) \end{aligned}$$

where

$$H_1(\lambda) = \mathbb{E}[R_0(\lambda, \tau) - B_2 \eta^2 R_2(\tau, \lambda) - \dots - B_k \eta^k R_k(\lambda, \tau)].$$

Since (4.1) guarantees that

$$\begin{aligned}
 B_2 &= \frac{C_2}{2!} \\
 B_4 &= \frac{C_4}{4!} - \left( \frac{C_2}{2!} \right)^2 \\
 B_6 &= \frac{C_6}{6!} - \frac{C_2 C_4}{2! 4!} - \left( \frac{C_4}{4!} - \left( \frac{C_2}{2!} \right)^2 \right) \frac{C_2}{2!} \\
 &\vdots \\
 B_k &= \frac{C_k}{k!} - \frac{B_2 C_{k-2}}{(k-2)!} - \dots - \frac{B_{k-2} C_2}{2!},
 \end{aligned}$$

the coefficients of  $\eta^2, \eta^4, \dots, \eta^k$  vanish, and we obtain

$$\mathbb{E}[G_\lambda(\tau)] = F_\lambda(0) + H_1(\lambda).$$

First we bound the bias  $H_1(\lambda)$ . In the remainder of the proof we let  $B_0 = -1$  to simplify notation, so that

$$H_1(\lambda) = \sum_{i=0,2,\dots,k} -B_i R_i(\lambda, \tau) \eta^i.$$

We first obtain a bound for  $|B_i R_i(\lambda, \tau) \eta^i|$ . Note

$$\begin{aligned}
 (k+1-i)! \eta^i R_i(\lambda, \tau) &= \eta^i \int_0^\tau F_\lambda^{(k+2)}(t) (\tau-t)^{k+1-i} dt \\
 &= \eta^i \int_0^\tau \lambda^{k+2} L^{(k+2)}((1-t)\lambda) (\tau-t)^{k+1-i} dt.
 \end{aligned}$$

We observe that

$$\begin{aligned}
 \left| ((1-t)\lambda)^{k+2} L^{(k+2)}((1-t)\lambda) \right| &\leq \Lambda_{k+2}((1-t)\lambda) \\
 \left| \lambda^{k+2} L^{(k+2)}((1-t)\lambda) \right| &\leq \frac{1}{(1-t)^{k+2}} \frac{\Lambda_{k+2}((1-t)\lambda)}{\Lambda_{k+2}(\lambda)} \Lambda_{k+2}(\lambda) \\
 \left| \lambda^{k+2} L^{(k+2)}((1-t)\lambda) \right| &\leq \frac{R(\lambda) \Lambda_{k+2}(\lambda)}{(1-t)^{k+2}}
 \end{aligned}$$

so that

$$-\frac{R(\lambda) \Lambda_{k+2}(\lambda)}{(1-t)^{k+2}} \leq \lambda^{k+2} L^{(k+2)}((1-t)\lambda) \leq \frac{R(\lambda) \Lambda_{k+2}(\lambda)}{(1-t)^{k+2}}.$$

Now assume first of all that  $\tau$  is positive. We have

$$\begin{aligned}
|(k+1-i)! \eta^i R_i(\lambda, \tau)| &\leq \eta^i R(\lambda) \Lambda_{k+2}(\lambda) \int_0^\tau \frac{(\tau-t)^{k+1-i}}{(1-t)^{k+2}} dt \\
&\leq \eta^i R(\lambda) \Lambda_{k+2}(\lambda) \int_0^\tau \frac{\tau^{k+1-i}}{(1-t)^{k+2}} dt \\
&= \eta^i \tau^{k+1-i} R(\lambda) \Lambda_{k+2}(\lambda) \frac{1}{(k+1)} \left( \frac{1}{(1-\tau)^{k+1}} - 1 \right) \\
&\leq \frac{2^{k+2} R(\lambda)}{k+1} \eta^i \tau^{k+2-i} \Lambda_{k+2}(\lambda)
\end{aligned}$$

where the last line follows since  $\frac{1}{(1-\tau)^{k+1}} \leq 2 \cdot 2^{k+1} \tau$  for  $\tau \in [0, \frac{1}{2}]$ . A similar argument can be applied when  $\tau$  is negative, and we can conclude

$$|B_i \eta^i R_i(\lambda, \tau)| \leq \frac{2^{k+2} R(\lambda)}{(k+1)(k+1-i)!} \Lambda_{k+2}(\lambda) |B_i| \eta^i |\tau|^{k+2-i}, \quad (\text{E.1})$$

which gives

$$\begin{aligned}
\mathbb{E} |B_i \eta^i R_i(\lambda, \tau)| &\leq \frac{2^{k+2} R(\lambda)}{(k+1)(k+1-i)!} \Lambda_{k+2}(\lambda) T^{k+2-i} |B_i| \eta^{k+2} \\
&= \frac{2^{k+2} (k+2-i) R(\lambda)}{k+1} \Lambda_{k+2}(\lambda) \frac{T^{k+2-i}}{(k+2-i)!} |B_i| \eta^{k+2}.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
|\mathbb{E}[G_\lambda(\tau)] - L(\lambda)| &= |H_1(\lambda)| \leq \frac{R(\lambda) \Lambda_{k+2}(\lambda)}{k+1} (2E\eta)^{k+2} \sum_{i=0,2,\dots,k} (k+2-i) \\
&\lesssim R(\lambda) k \Lambda_{k+2}(\lambda) (2E\eta)^{k+2},
\end{aligned}$$

which establishes the bound on the bias. We now bound the variance. We note

$$G_\lambda(\tau) = \underbrace{\sum_{i=0,2,\dots,k} \sum_{j=0,1,\dots,k+1-i} \frac{-B_i}{j!} F_\lambda^{(i+j)}(0) \eta^i \tau^j}_{:=(\mathbf{I})} + \underbrace{\sum_{i=0,2,\dots,k} -B_i R_i(\lambda, \tau) \eta^i}_{:=(\mathbf{II})}.$$

Thus,

$$\begin{aligned}
\text{Var}[G_\lambda(\tau)] &= \mathbb{E}[G_\lambda(\tau)^2] - \mathbb{E}[G_\lambda(\tau)]^2 \\
&= \mathbb{E}[(\mathbf{I})(\mathbf{I})] + 2\mathbb{E}[(\mathbf{I})(\mathbf{II})] + \mathbb{E}[(\mathbf{II})(\mathbf{II})] - F_\lambda(0)^2 - 2F_\lambda(0)H_1(\lambda) - H_1(\lambda)^2 \\
&\leq \underbrace{(\mathbb{E}[(\mathbf{I})(\mathbf{I})] - F_\lambda(0)^2)}_{:=(\mathbf{A})} + \underbrace{(2\mathbb{E}[(\mathbf{I})(\mathbf{II})] - 2F_\lambda(0)H_1(\lambda))}_{:=(\mathbf{B})} + \underbrace{\mathbb{E}[(\mathbf{II})(\mathbf{II})]}_{:=(\mathbf{C})}
\end{aligned}$$

and we proceed to bound each term.

$$(\mathbf{I})(\mathbf{I}) - F_\lambda(0)^2 = \sum_{i=0,2,\dots,k} \sum_{\ell=0,2,\dots,k} \sum_{j=0}^{k+1-i} \sum_{s=0}^{k+1-\ell} \frac{B_i B_\ell}{j! \ell!} F_\lambda^{(i+j)}(0) F_\lambda^{(\ell+s)}(0) \eta^{i+\ell} \tau^{j+s} \mathbf{1}_E$$

where  $\mathbf{1}_E$  is an indicator function indicating that  $i, j, \ell, s$  are not all zero. We have

$$\begin{aligned} \mathbb{E} \left| \frac{B_i B_\ell}{j! \ell!} F_\lambda^{(i+j)}(0) F_\lambda^{(\ell+s)}(0) \eta^{i+\ell} \tau^{j+s} \right| &\leq \frac{|B_i B_\ell|}{j! \ell!} C_{j+s} \Lambda_{i+j}(\lambda) \Lambda_{\ell+s}(\lambda) \eta^{i+\ell+j+s} \\ &\leq \frac{|B_i B_\ell|}{j! \ell!} T^j T^s \Lambda_{i+j}(\lambda) \Lambda_{\ell+s}(\lambda) \eta^{i+\ell+j+s} \\ &\leq E^{i+j} E^{\ell+s} \Lambda_{i+j}(\lambda) \Lambda_{\ell+s}(\lambda) \eta^{i+\ell+j+s} \\ &= (\Lambda_{i+j}(\lambda) (E\eta)^{i+j}) (\Lambda_{\ell+s}(\lambda) (E\eta)^{\ell+s}). \end{aligned}$$

Noting that only terms where  $j + s$  is even survive expectation, and letting  $\tilde{i} = i + j$  and  $\tilde{\ell} = \ell + s$ , we obtain

$$\begin{aligned} \mathbb{E}[(\mathbf{I})(\mathbf{I})] - F_\lambda(0)^2 &\leq \sum_{i=0,2,\dots,k} \sum_{\ell=0,2,\dots,k} \sum_{j=0}^{k+1-i} \sum_{s=0}^{k+1-\ell} \Lambda_{i+j}(\lambda) (4T\eta)^{i+j} \Lambda_{\ell+s}(\lambda) (4T\eta)^{\ell+s} \mathbf{1}_E \mathbf{1}(j+s \text{ even}) \\ &= \sum_{\tilde{i}=0}^{k+1} \sum_{\tilde{\ell}=0}^{k+1} C_{\tilde{i}, \tilde{\ell}} \Lambda_{\tilde{i}}(\lambda) (E\eta)^{\tilde{i}} \Lambda_{\tilde{\ell}}(\lambda) (E\eta)^{\tilde{\ell}} \end{aligned}$$

for coefficients  $C_{\tilde{i}, \tilde{\ell}}$  such that  $C_{0,0} = 0$ ,  $C_{\tilde{i}, \tilde{\ell}} = 0$  if  $\tilde{i} + \tilde{\ell}$  is odd, and  $C_{\tilde{i}, \tilde{\ell}} \leq k^2$ . Thus,

$$\mathbb{E}[(\mathbf{I})(\mathbf{I})] - F_\lambda(0)^2 \leq k^2 \sum_{\substack{2 \leq \tilde{i} + \tilde{\ell} \leq 2k+2 \\ \tilde{i} + \tilde{\ell} \text{ even}}} \Lambda_{\tilde{i}}(\lambda) \Lambda_{\tilde{\ell}}(\lambda) (E\eta)^{\tilde{i} + \tilde{\ell}} \leq k^2 \Lambda(\lambda)^2.$$

Next we bound  $\mathbb{E}[(\mathbf{II})(\mathbf{II})]$ .

$$(\mathbf{II})(\mathbf{II}) = \sum_{i=0,2,\dots,k} \sum_{\ell=0,2,\dots,k} B_i B_\ell R_i(\lambda, \tau) R_\ell(\lambda, \tau) \eta^{i+\ell}.$$

Utilizing Equation (E.1), we have

$$\left| B_i B_\ell R_i(\lambda, \tau) R_\ell(\lambda, \tau) \eta^{i+\ell} \right| \leq \frac{2^{2k+4} R(\lambda)^2 |B_i B_\ell|}{(k+1)^2 (k+1-i)! (k+1-\ell)!} \Lambda_{k+2}(\lambda)^2 \eta^{i+\ell} |\tau|^{2k+4-i-\ell},$$

which gives

$$\begin{aligned}
& \mathbb{E} \left| B_i B_\ell R_i(\lambda, \tau) R_\ell(\lambda, \tau) \eta^{i+\ell} \right| \\
& \leq \frac{2^{2k+4} R(\lambda)^2 T^{2k+4-i-\ell} |B_i B_\ell|}{(k+1)^2 (k+1-i)! (k+1-\ell)!} \Lambda_{k+2}(\lambda)^2 \eta^{2k+4} \\
& \leq \frac{R(\lambda)^2 (k+2-i)(k+2-\ell)}{(k+1)^2} \left( \frac{T^{k+2-i} |B_i|}{(k+2-i)!} \right) \left( \frac{T^{k+2-\ell} |B_\ell|}{(k+2-\ell)!} \right) \Lambda_{k+2}(\lambda)^2 (2\eta)^{2k+4} \\
& \leq \frac{R(\lambda)^2 (k+2-i)(k+2-\ell)}{(k+1)^2} \Lambda_{k+2}(\lambda)^2 (2E\eta)^{2k+4}
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E} [(\mathbf{II})(\mathbf{III})] & \leq \frac{R(\lambda)^2}{(k+1)^2} \Lambda_{k+2}(\lambda)^2 (2E\eta)^{2k+4} \sum_{i=0,2,\dots,k} \sum_{\ell=0,2,\dots,k} (k+1-i)(k+2-\ell) \\
& \lesssim k^2 R(\lambda)^2 \Lambda_{k+2}(\lambda)^2 (2E\eta)^{2k+4} \\
& \leq k^2 R(\lambda)^2 \Lambda(\lambda)^2.
\end{aligned}$$

Finally we bound the cross term  $2\mathbb{E}[(\mathbf{I})(\mathbf{II})] - 2F_\lambda(0)H_1(\lambda)$ .

$$(\mathbf{I})(\mathbf{II}) = \sum_{i=0,2,\dots,k} \sum_{j=0}^{k+1-i} \sum_{\ell=0,2,\dots,k} \frac{B_i}{j!} F_\lambda^{(i+j)}(0) \eta^i \tau^j B_\ell R_\ell(\lambda, \tau) \eta^\ell \quad (\text{E.2})$$

Since  $|F_\lambda^{(i+j)}(0)| \leq \Lambda_{i+j}(\lambda)$  and  $|B_\ell R_\ell(\lambda, \tau) \eta^\ell| \leq \frac{2^{k+2} R(\lambda) |B_\ell|}{(k+1)(k+1-\ell)!} \Lambda_{k+2}(\lambda) \eta^\ell \tau^{k+2-\ell}$  from (E.1), we have

$$\left| \frac{B_i}{j!} F_\lambda^{(i+j)}(0) \eta^i \tau^j B_\ell R_\ell(\lambda, \tau) \eta^\ell \right| \leq \frac{2^{k+2} R(\lambda) |B_i B_\ell|}{(k+1)j! (k+1-\ell)!} \Lambda_{i+j}(\lambda) \Lambda_{k+2}(\lambda) \eta^{i+\ell} \tau^{k+2+j-\ell}$$

so that

$$\begin{aligned}
& \mathbb{E} \left| \frac{B_i}{j!} F_\lambda^{(i+j)}(0) \eta^i \tau^j B_\ell R_\ell(\lambda, \tau) \eta^\ell \right| \\
& \leq \frac{2^{k+2} R(\lambda) T^{k+2+j-\ell} |B_i B_\ell|}{(k+1)j! (k+1-\ell)!} \Lambda_{i+j}(\lambda) \Lambda_{k+2}(\lambda) \eta^{i+j+k+2} \\
& = \frac{2^{k+2} R(\lambda) (k+2-\ell)}{(k+1)} \left( \frac{T^j |B_i|}{j!} \right) \left( \frac{T^{k+2-\ell} |B_\ell|}{(k+2-\ell)!} \right) \Lambda_{i+j}(\lambda) \Lambda_{k+2}(\lambda) \eta^{i+j+k+2} \\
& = \frac{R(\lambda) (k+2-\ell)}{(k+1)} \left[ (E\eta)^{i+j} \Lambda_{i+j}(\lambda) \right] \cdot \left[ (2E\eta)^{k+2} \Lambda_{k+2}(\lambda) \right].
\end{aligned}$$

The same bound holds for the terms of  $F_\lambda(0)H_1(\lambda)$ , which arise from  $i = 0, j = 0$  in (E.2), so that

$$\begin{aligned}
& 2\mathbb{E}[(\mathbf{I})(\mathbf{II})] - 2F_\lambda(0)H_1(\lambda) \\
& \lesssim \left( \sum_{i=0,2,\dots,k} \sum_{j=0}^{k+1-i} (E\eta)^{i+j} \Lambda_{i+j}(\lambda) \right) \left( \sum_{\ell=0,2,\dots,k} \frac{R(\lambda)(k+2-\ell)}{(k+1)} (2E\eta)^{k+2} \Lambda_{k+2}(\lambda) \right) \\
& \lesssim \left( k \sum_{\tilde{i}=0}^{k+1} \Lambda_{\tilde{i}}(\lambda) (E\eta)^{\tilde{i}} \right) \left( kR(\lambda) (2E\eta)^{k+2} \Lambda_{k+2}(\lambda) \right) \\
& \leq k^2 R(\lambda) \sum_{\tilde{i}=0}^{k+1} \Lambda_{\tilde{i}}(\lambda) \Lambda_{k+2}(\lambda) (2E\eta)^{\tilde{i}+k+2} \\
& \leq k^2 R(\lambda) \Lambda(\lambda)^2.
\end{aligned}$$

Thus,  $\text{Var}[G_\lambda(\tau)] \lesssim k^2 R(\lambda)^2 \Lambda(\lambda)^2$  and the lemma is proved.  $\square$

LEMMA 4.2 Let the assumptions and notation of Lemma 4.1 hold, and let  $\tau_1, \dots, \tau_M$  be independent. Define

$$\tilde{L}(\lambda) := \frac{1}{M} \sum_{j=1}^M G_\lambda(\tau_j).$$

Then, with probability at least  $1 - 1/t^2$ ,

$$|\tilde{L}(\lambda) - L(\lambda)| \lesssim kR(\lambda) \left( \Lambda_{k+2}(\lambda) (2E\eta)^{k+2} + \frac{t\Lambda(\lambda)}{\sqrt{M}} \right).$$

*Proof.* By Lemma 4.1 and the independence of the  $\tau_j$ , we have

$$\begin{aligned}
|L(\lambda) - \mathbb{E} \tilde{L}(\lambda)| & \lesssim kR(\lambda) \Lambda_{k+2}(\lambda) (2E\eta)^{k+2} \\
\text{Var } \tilde{L}(\lambda) & \lesssim \frac{1}{M} k^2 \Lambda(\lambda)^2
\end{aligned}$$

so by Chebyshev's inequality we can conclude that with probability at least  $1 - 1/t^2$ , we have

$$|\tilde{L}(\lambda) - \mathbb{E}[\tilde{L}(\lambda)]| \leq \frac{tkR(\lambda)\Lambda(\lambda)}{\sqrt{M}},$$

which gives

$$\begin{aligned}
|L(\lambda) - \tilde{L}(\lambda)| & \leq |L(\lambda) - \mathbb{E}[\tilde{L}(\lambda)]| + |\mathbb{E}[\tilde{L}(\lambda)] - \tilde{L}(\lambda)| \\
& \lesssim kR(\lambda) \Lambda_{k+2}(\lambda) (2E\eta)^{k+2} + \frac{tkR(\lambda)\Lambda(\lambda)}{\sqrt{M}}.
\end{aligned}$$

$\square$

## F. Supporting results: noisy dilation MRA

This appendix contains supporting results needed to prove Proposition 5.1, which defines a wavelet invariant estimator for noisy dilation MRA. Lemma 5.1 controls the additive noise error and Lemma 5.2 controls the cross-term error. Lemma F.1 guarantees that the dilation unbiasing procedure applied to the additive noise still has mean  $\sigma^2$ , which is needed to prove Lemma 5.1.

LEMMA 5.1 Let the notation and assumptions of Proposition 5.1 hold, and let  $A_\lambda$  be the operator defined in (5.4). Then, with probability at least  $1 - 1/t^2$ ,

$$\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \sigma^2 \right| \leq \frac{2t\sqrt{k}\Psi\sigma^2}{\sqrt{M}}.$$

*Proof.* Let

$$D(\varepsilon_j, \lambda) := \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega.$$

By Lemma D.1,  $\mathbb{E}_\varepsilon [|\widehat{\varepsilon}_j(\omega)|^2] = \sigma^2$ , and we thus obtain

$$\begin{aligned} \mathbb{E}_\varepsilon [D(\varepsilon_j, \lambda)] &= \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right] \\ &= \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 B_2 \eta^2 \lambda^2 \frac{d}{d\lambda^2} |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \dots \right. \\ &\quad \left. - \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 B_k \eta^k \lambda^k \frac{d}{d\lambda^k} |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right] \\ &= \sigma^2 \left( \frac{1}{2\pi} \int |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \frac{B_2 \eta^2}{2\pi} \int \lambda^2 \frac{d}{d\lambda^2} |\widehat{\psi}_\lambda(\omega)|^2 d\omega - \dots \right. \\ &\quad \left. - \frac{B_k \eta^k}{2\pi} \int \lambda^k \frac{d}{d\lambda^k} |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right) \\ &= \sigma^2 (1 - 0 - \dots - 0) \\ &= \sigma^2, \end{aligned}$$

where we have used Lemma F.1 to conclude  $\int \lambda^m \left( \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right) d\omega = 0$  for  $m = 2, \dots, k$ . Also since  $(a_1 + \dots + a_n)^2 \leq n(a_1^2 + \dots + a_n^2)$  by the Cauchy–Schwarz inequality, we obtain

$$\mathbb{E}_\varepsilon [D(\varepsilon_j, \lambda)^2] \leq \mathbb{E}_\varepsilon \left[ k \sum_{m=0,2,\dots,k} \left( \frac{B_m \eta^m}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right)^2 \right]$$

where we let  $\frac{d}{d\lambda^0} |\widehat{\psi}_\lambda(\omega)|^2$  denote  $|\widehat{\psi}_\lambda(\omega)|^2$  and  $B_0 = 1$ . By Lemma D.1, we have  $\mathbb{E}_\varepsilon \left[ |\varepsilon_j(\omega)|^2 |\varepsilon_j(\xi)|^2 \right] \leq 3\sigma^4$  for all frequencies  $\omega, \xi$ , so that

$$\begin{aligned} & \mathbb{E}_\varepsilon \left[ \left( \frac{B_m \eta^m}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right)^2 \right] \\ & \leq \mathbb{E}_\varepsilon \left[ \frac{B_m^2 \eta^{2m}}{4\pi^2} \int \int |\widehat{\varepsilon}_j(\omega)|^2 |\widehat{\varepsilon}_j(\xi)|^2 \left| \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| \cdot \left| \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\xi)|^2 \right| d\omega d\xi \right] \\ & \leq 3\sigma^4 \left( \frac{1}{2\pi} \int \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right)^2 \\ & \leq 3\sigma^4 \Psi_m^2(E\eta)^{2m}, \end{aligned}$$

where the last line follows from Corollary B.1 in Appendix B. We thus obtain

$$\begin{aligned} \mathbb{E}_\varepsilon \left[ D(\varepsilon_j, \lambda)^2 \right] & \leq k \sum_{m=0,2,\dots,k} \mathbb{E}_\varepsilon \left[ \left( \frac{B_m \eta^m}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right)^2 \right] \\ & \leq 3k\sigma^4 \sum_{m=0,2,\dots,k} \Psi_m^2(E\eta)^{2m} := (\mathbf{I}) \end{aligned}$$

so that

$$\begin{aligned} \mathbb{E}_\varepsilon \left[ D(\varepsilon_j, \lambda) - \sigma^2 \right] & = 0 \\ \text{Var}_\varepsilon \left[ D(\varepsilon_j, \lambda) - \sigma^2 \right] & = \text{Var}_\varepsilon \left[ D(\varepsilon_j, \lambda) \right] \leq \mathbb{E}_\varepsilon \left[ (D(\varepsilon_j, \lambda))^2 \right] \leq (\mathbf{I}). \end{aligned}$$

Thus,

$$\text{Var}_\varepsilon \left( \frac{1}{M} \sum_{j=1}^M D(\varepsilon_j, \lambda) - \sigma^2 \right) \leq \frac{(\mathbf{I})}{M}$$

so that by Chebyshev's inequality with probability at least  $1 - 1/t^2$

$$\left| \frac{1}{M} \sum_{j=1}^M D(\varepsilon_j, \lambda) - \sigma^2 \right| \leq \frac{t\sqrt{(\mathbf{I})}}{\sqrt{M}} \leq t\sqrt{3k} \left( \sum_{m=0,2,\dots,k} \Psi_m(E\eta)^m \right) \frac{\sigma^2}{\sqrt{M}} = 2t\sqrt{k}\Psi \frac{\sigma^2}{\sqrt{M}}.$$

□

LEMMA 5.2 Let the notation and assumptions of Proposition 5.1 hold, and let  $A_\lambda$  be the operator defined in (5.4). Then, with probability at least  $1 - 1/t^2$ ,

$$\left| \frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int \left( \widehat{f}_{\tau_j}(\omega) \overline{\widehat{\varepsilon}_j(\omega)} + \overline{\widehat{f}_{\tau_j}(\omega)} \widehat{\varepsilon}_j(\omega) \right) A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega \right| \lesssim \frac{t}{\sqrt{M}} \sqrt{\Psi(\Lambda_0(\lambda) + \Lambda(\lambda))} \sigma.$$

*Proof.* We have

$$\frac{1}{M} \sum_{j=1}^M \frac{1}{2\pi} \int \left( \widehat{f}_{\tau_j}(\omega) \overline{\widehat{\varepsilon}_j}(\omega) + \overline{\widehat{f}_{\tau_j}}(\omega) \widehat{\varepsilon}_j(\omega) \right) A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega = \frac{1}{M} \sum_{j=1}^M Y_j + \overline{Y_j}$$

where

$$Y_j := \frac{1}{2\pi} \int \left( \overline{\widehat{f}_{\tau_j}}(\omega) \widehat{\varepsilon}_j(\omega) \right) A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 d\omega.$$

The random variable  $Y_j$  has randomness depending on both  $\varepsilon_j$  and  $\tau_j$ . Note that

$$\mathbb{E}_{\varepsilon, \tau}[Y_j] = \mathbb{E}_{\varepsilon, \tau} \left[ \mathbb{E}_{\varepsilon, \tau}[Y_j | \tau_j] \right]$$

since  $Y_j$  is integrable. Thus, since  $\mathbb{E}_{\varepsilon, \tau}[\widehat{\varepsilon}_j(\omega)] = 0$ , we obtain  $\mathbb{E}_{\varepsilon, \tau}[Y_j | \tau_j] = 0$ , which yields  $\mathbb{E}_{\varepsilon, \tau}[Y_j] = 0$ . We also have

$$\begin{aligned} \text{Var}_{\varepsilon, \tau}[Y_j] &= \mathbb{E}_{\varepsilon, \tau}[Y_j^2] \\ &\leq \mathbb{E}_{\varepsilon, \tau} \left[ \left( \frac{1}{2\pi} \int |\overline{\widehat{f}_{\tau_j}}(\omega)| \cdot |\widehat{\varepsilon}_j(\omega)| \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right)^2 \right] \\ &\leq \mathbb{E}_{\varepsilon, \tau} \left[ \left( \frac{1}{2\pi} \int |\overline{\widehat{f}_{\tau_j}}(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right) \left( \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right) \right] \\ &= \mathbb{E}_\tau \left[ \frac{1}{2\pi} \int |\overline{\widehat{f}_{\tau_j}}(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right] \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right]. \end{aligned}$$

Letting  $B_0 = 1$  and applying Lemma B.2, we have

$$\begin{aligned} \mathbb{E}_\tau \left[ \frac{1}{2\pi} \int |\overline{\widehat{f}_{\tau_j}}(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right] &\leq \mathbb{E}_\tau \left[ \sum_{m=0,2,\dots,k} \frac{1}{2\pi} \int |\widehat{f}_{\tau_j}(\omega)|^2 \cdot \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right] \\ &\leq \mathbb{E}_\tau \left[ \sum_{m=0,2,\dots,k} (E\eta)^m \left( \|f_{\tau_j}\|_1^2 \Psi_m \wedge \frac{\|f'_{\tau_j}\|_1^2 \Theta_m}{\lambda^2} \right) \right] \\ &\leq \sum_{m=0,2,\dots,k} (E\eta)^m \left( \|f\|_1^2 \Psi_m \wedge \frac{4\|f'\|_1^2 \Theta_m}{\lambda^2} \right) \\ &\leq 4 \sum_{m=0,2,\dots,k} (E\eta)^m \Lambda_m(\lambda) \\ &\lesssim \Lambda_0(\lambda) + \Lambda(\lambda) \end{aligned}$$

since  $\|\tau_j\|_\infty \leq \frac{1}{2}$  guarantees  $\|f'_{\tau_j}\|_1 = \frac{1}{1-\tau_j} \|f'\|_1 \leq 2 \|f'\|_1$ . Also,

$$\begin{aligned} \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \cdot \left| A_\lambda |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right] &\leq \mathbb{E}_\varepsilon \left[ \sum_{m=0,2,\dots,k} \frac{1}{2\pi} \int |\widehat{\varepsilon}_j(\omega)|^2 \cdot \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right] \\ &= \sigma^2 \left( \sum_{m=0,2,\dots,k} \frac{1}{2\pi} \int \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right| d\omega \right) \\ &\leq \sigma^2 \sum_{m=0,2,\dots,k} (E\eta)^m \Psi_m \\ &= \sigma^2 \Psi \end{aligned}$$

where the second line follows from Lemma D.1 in Appendix D and the next to last line from Corollary B.1 in Appendix B. We thus have

$$\begin{aligned} \mathbb{E}_{\varepsilon,\tau}[Y_j] &= 0 \\ \text{Var}_{\varepsilon,\tau}[Y_j] &\lesssim \sigma^2 \Psi (\Lambda_0(\lambda) + \Lambda(\lambda)) \end{aligned}$$

and an identical argument can be applied to the  $\bar{Y}_j$  so that by Chebyshev's inequality with probability at least  $1 - 1/t^2$

$$\left| \frac{1}{M} \sum_{j=1}^M Y_j + \bar{Y}_j \right| \leq \left| \frac{1}{M} \sum_{j=1}^M Y_j \right| + \left| \frac{1}{M} \sum_{j=1}^M \bar{Y}_j \right| \lesssim t \sqrt{\Psi} \sqrt{\Lambda_0(\lambda) + \Lambda(\lambda)} \frac{\sigma}{\sqrt{M}}. \quad \square$$

LEMMA F.1 Assume  $\psi$  is  $k$ -admissible. Then,

$$\int \lambda^m \left( \frac{d^m}{d\lambda^m} |\widehat{\psi}_\lambda(\omega)|^2 \right) d\omega = 0 \quad (\text{F.1})$$

for all  $1 \leq m \leq k$ .

*Proof.* We recall that since  $\psi$  is  $k$ -admissible,  $|\widehat{\psi}_\lambda(\omega)|^2 \in \mathbf{C}^k(\mathbb{R})$ , and to simplify notation we let  $g = |\widehat{\psi}|^2$  and

$$g_\lambda(\omega) = \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) = |\widehat{\psi}_\lambda(\omega)|^2.$$

We first establish that

$$\lambda^k \left( \frac{d}{d\lambda^k} g_\lambda(\omega) \right) = \frac{d}{d\omega} \left( -\omega \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) - (k-1) \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega). \quad (\text{F.2})$$

The proof is by induction. When  $k = 1$ , we obtain

$$\text{LHS of Eqn. (F.2)} = \lambda \frac{d}{d\lambda} \left( \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) \right) = -\frac{\omega}{\lambda^2} g'\left(\frac{\omega}{\lambda}\right) - \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) = -\omega g'_\lambda(\omega) - g_\lambda(\omega)$$

and

$$\text{RHS of Eqn. (F.2)} = \frac{d}{d\omega} (-\omega g_\lambda(\omega)) = -\omega g'_\lambda(\omega) - g_\lambda(\omega),$$

so the base case is established. We now assume that Equation (F.2) holds and show it also holds for  $k+1$  replacing  $k$ . By the inductive hypothesis

$$\begin{aligned} \frac{d}{d\lambda^k} g_\lambda(\omega) &= \frac{d}{d\omega} \left( -\omega \lambda^{-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) - (k-1) \lambda^{-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \\ \frac{d}{d\lambda^{k+1}} g_\lambda(\omega) &= \frac{d}{d\omega} \left( -\omega \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) + \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \omega \lambda^{-2} \right) \\ &\quad - (k-1) \left( \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) + \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) (-\lambda^{-2}) \right) \\ &= \frac{d}{d\omega} \left( -\omega \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) \right) - (k-1) \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) \\ &\quad + \underbrace{\frac{d}{d\omega} \left( \omega \lambda^{-2} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) + (k-1) \lambda^{-2} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega)}_{= -\lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) \text{ by inductive hypothesis}} \\ &= \frac{d}{d\omega} \left( -\omega \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) \right) - k \lambda^{-1} \frac{d}{d\lambda^k} g_\lambda(\omega) \end{aligned}$$

so that

$$\lambda^{k+1} \frac{d}{d\lambda^{k+1}} g_\lambda(\omega) = \frac{d}{d\omega} \left( -\omega \lambda^k \frac{d}{d\lambda^k} g_\lambda(\omega) \right) - k \lambda^k \frac{d}{d\lambda^k} g_\lambda(\omega).$$

Thus, (F.2) is established. We now use integration by parts to show (F.2) implies (F.1) in the Lemma. The proof of (F.1) is once again by induction. When  $k=1$ , we have already shown

$$\lambda \left( \frac{d}{d\lambda} g_\lambda(\omega) \right) = -\omega g'_\lambda(\omega) - g_\lambda(\omega). \quad (\text{F.3})$$

Integration by parts gives

$$\int \omega g'_\lambda(\omega) d\omega = (\omega g_\lambda(\omega)) \Big|_{-\infty}^{\infty} - \int g_\lambda(\omega) d\omega = \int g_\lambda(\omega) d\omega.$$

Note  $\omega g_\lambda(\omega)$  vanishes at  $\pm\infty$  since  $g \in \mathbf{L}^1(\mathbb{R})$  guarantees  $g_\lambda \in \mathbf{L}^1(\mathbb{R})$ , and thus  $g_\lambda$  must decay faster than  $\omega^{-1}$ . Utilizing (F.3),

$$\int \omega g'_\lambda(\omega) - g_\lambda(\omega) d\omega = 0 \implies \int \lambda \left( \frac{d}{d\lambda} g_\lambda(\omega) \right) d\omega = 0$$

and the base case is established. We now assume

$$\int \lambda^{k-1} \left( \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) d\omega = 0.$$

By integrating Equation (F.2), we obtain

$$\begin{aligned}
& \int \lambda^k \left( \frac{d}{d\lambda^k} g_\lambda(\omega) \right) d\omega \\
&= \int \frac{d}{d\omega} \left( -\omega \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) d\omega - (k-1) \underbrace{\int \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) d\omega}_{=0 \text{ by induc. hypo.}} \\
&= \int -\omega \frac{d}{d\omega} \left( \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \right) d\omega - \underbrace{\int \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) d\omega}_{=0 \text{ by induc. hypo.}} \\
&= -\omega \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) \Big|_{-\infty}^{\infty} + \underbrace{\int \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) d\omega}_{=0 \text{ by induc. hypo.}} \\
&= 0.
\end{aligned}$$

We are guaranteed  $-\omega \lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega)$  vanishes at  $\pm\infty$  since in the proof of Lemma 4.3 we showed  $\lambda^{k-1} \frac{d}{d\lambda^{k-1}} g_\lambda(\omega) = \sum_{j=0}^{k-1} C_j \omega^j g_\lambda^{(j)}(\omega)$ , and  $\omega^j g_\lambda^{(j)} \in L^1(\mathbb{R})$  implies  $\omega^{j+1} g_\lambda^{(j)}$  vanishes at  $\pm\infty$ .  $\square$

## G. Moment estimation for noisy dilation MRA

In this appendix we outline a moment estimation procedure for noisy dilation MRA (Model 2) in the special case  $t = 0$ , i.e. signals are randomly dilated and subjected to additive noise but are not translated. This procedure is a generalization of the method presented in Section 6.3.

Given the additive noise level, the moments of the dilation distribution  $\tau$  can be empirically estimated from the mean and variance of the random variables  $\beta_m(y_j)$  defined by

$$\beta_m(y_j) = \int_0^{2^\ell \pi} \omega^m \widehat{y}_j(\omega) d\omega \quad (\text{G.1})$$

for integer  $m \geq 0$ . To account for the effect of additive noise on the above random variables, we define

$$g_m(\ell, \sigma) = \int_0^{2^\ell \pi} \int_0^{2^\ell \pi} \frac{2\sigma^2 \xi^m \omega^m \sin(\frac{1}{2}(\xi - \omega))}{(\xi - \omega)} d\omega d\xi \quad (\text{G.2})$$

and an order  $m$  additive noise adjusted squared coefficient of variation by

$$CV_m := \frac{\text{Var}[\beta_m(y_j)] - g_m(\ell, \sigma)}{|\mathbb{E}[\beta_m(y_j)]|^2}. \quad (\text{G.3})$$

REMARK G.1 If the noisy signals are supported in  $[-\frac{N}{2}, \frac{N}{2}]$  instead of  $[-\frac{1}{2}, \frac{1}{2}]$ , (G.2) is replaced with

$$g_m(N, \ell, \sigma) = \int_0^{2^\ell \pi} \int_0^{2^\ell \pi} \frac{2\sigma^2 \xi^m \omega^m \sin(\frac{N}{2}(\xi - \omega))}{(\xi - \omega)} d\omega d\xi.$$

The following proposition mirrors Proposition 6.1 for dilation MRA; its proof appears at the end of Appendix G.

PROPOSITION G.1 Assume Model 2 with  $t = 0$  and  $CV_0, CV_1$  defined by (G.1), (G.2) and (G.3). Then,

$$\begin{aligned} CV_0 &= \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6) \\ CV_1 &= 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6). \end{aligned}$$

Once again we cannot compute  $CV_m$  exactly, but by replacing  $\text{Var}, \mathbb{E}$  with their finite sample estimators, we obtain approximations  $\widetilde{CV}_m$  that can be used to define estimators of the dilation moments.

DEFINITION G.1 Assume Model 3 with  $t = 0$  and  $\widetilde{CV}_0, \widetilde{CV}_1$  the empirical counterparts of (G.3). Define the second-order estimator of  $\eta^2$  by  $\widetilde{\eta}^2 = \widetilde{CV}_0$ . Define the fourth-order estimators of  $(\eta^2, C_4\eta^4)$  by the unique positive solution  $(\widetilde{\eta}^2, \widetilde{C}_4)$  of

$$\begin{aligned} \widetilde{CV}_0 &= \eta^2 + (3C_4 - 3)\eta^4 \\ \widetilde{CV}_1 &= 4\eta^2 + (25C_4 - 33)\eta^4. \end{aligned}$$

As  $M \rightarrow \infty$ , the second-order moment estimator is accurate up to  $O(\eta^4)$  and the fourth-order moment estimators are accurate up to  $O(\eta^6)$ . However, in the finite sample regime, the  $g_m(\ell, \sigma)$  appearing in (G.3) will be replaced with  $g_m(\ell, \sigma) \pm O(\sigma^2/\sqrt{M})$ , so that the estimators given in Definition G.1 are subject to an error of order  $O(\sigma^2/\sqrt{M})$ . More generally, the additive noise fluctuations imply that to estimate the first  $k/2$  even moments of  $\tau$  up to an  $O(\eta^{k+1})$  error will require  $\sigma^2/\sqrt{M} \leq \eta^{k+1}$ , or  $M \geq \sigma^4 \eta^{-2(k+1)}$ .

Having established an empirical moment estimation procedure for noisy dilation MRA when  $t = 0$ , we repeat the simulations of Section 5.2 on the restricted model, but estimate the additive and dilation moments empirically. Since accurately estimating the moments of  $\tau$  is difficult for  $\sigma$  large, we make three modifications to the oracle set-up. First, we lower the additive noise level by a factor of 2 from the oracle simulations, and consider all parameter combinations resulting from  $\sigma = 2^{-5}, 2^{-4}$  (giving SNR = 9.0, 2.2) and  $\eta = 0.06, 0.12$ . Secondly, we take  $M$  substantially larger than for the oracle simulations, with  $16,384 \leq M \leq 370,727$ . Thirdly, we compute WSC  $k = 4$  only for large dilations. For large dilations  $(\eta^2, C_4\eta^4)$  are approximated with fourth-order estimators, while for small dilations  $\eta^2$  is approximated with a second-order estimator (see Definition G.1).

Results are shown in Fig. G7, and the same overall behavior observed in the oracle simulations for large  $M$  holds. The additive noise level was estimated empirically as described in Section 6.2. For the medium- and high-frequency signal, WSC  $k = 2$  has substantially smaller error than both PS  $k = 0$  and WSC  $k = 0$ ; for the large-frequency signal, the error is decreased by at least a factor of 2 for large dilations and a factor of 4 for small dilations relative to both zero order estimators. When WSC  $k = 4$  is defined, it has a smaller error than WSC  $k = 2$  for the high-frequency signal, while WSC  $k = 2$  is preferable for the low- and medium-frequency signal. We observe that for the oracle simulations WSC  $k = 4$  is preferable for all frequencies, so this is most likely due to error in the moment estimation degrading the WSC  $k = 4$  estimator. For the low-frequency signal, PS  $k = 0$  once again achieves the smallest error for small dilations, while for large dilations the higher order wavelet methods appear to surpass PS  $k = 0$  for  $M$  large enough.

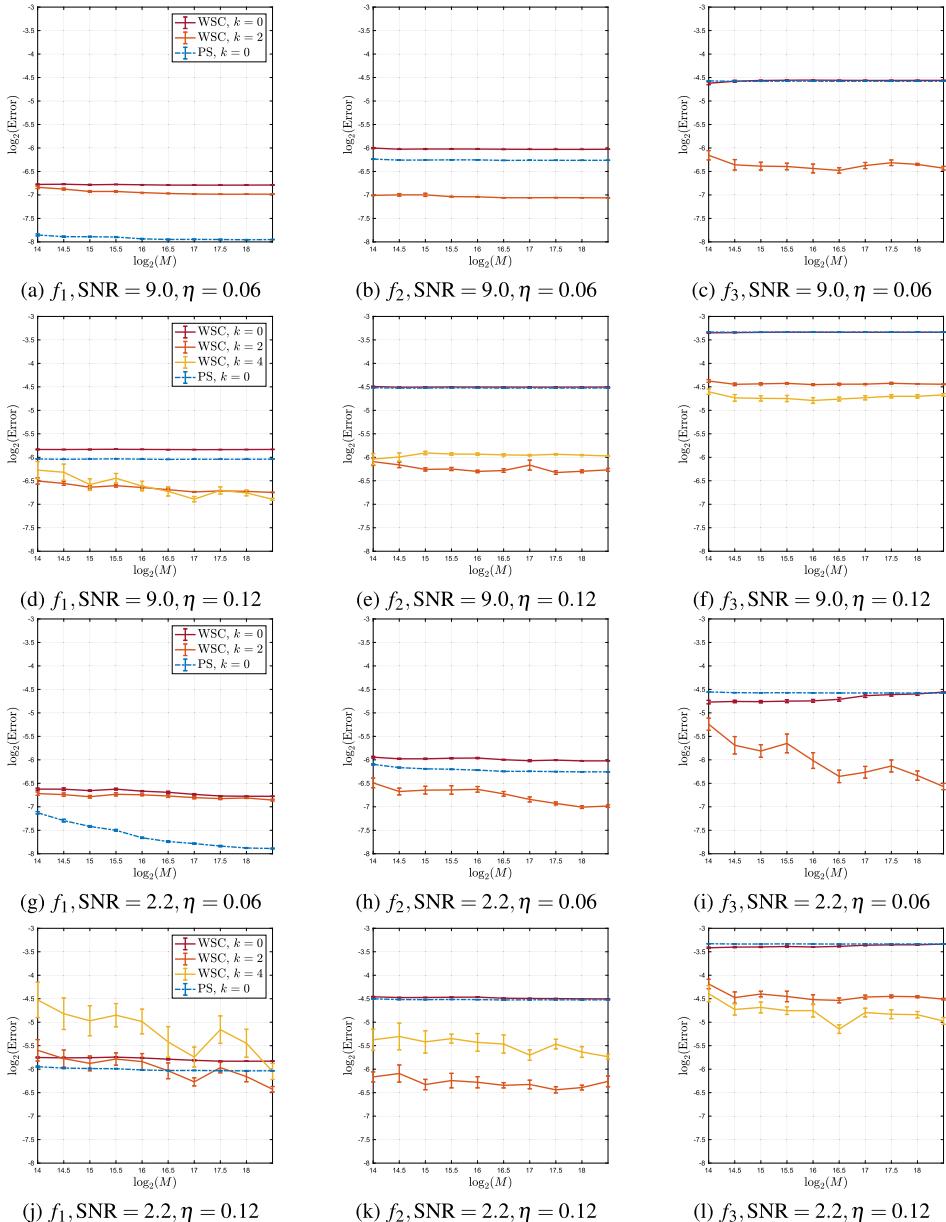


FIG. G7.  $L^2$  error with standard error bars for noisy dilation MRA model ( $t = 0$ , empirical moment estimation). First, second and third column shows results for low-, medium- and high-frequency signals. All plots have the same axis limits.

*Proof of Proposition G.1.* Since  $y_j = L_{\tau_j} f + \varepsilon_j$ , we have

$$\begin{aligned}
\mathbb{E}[\beta_m(y_j)] &= \mathbb{E} \left[ \int_0^{2^\ell \pi} \omega^m (\widehat{f}_{\tau_j}(\omega) + \widehat{\varepsilon}_j(\omega)) \, d\omega \right] \\
&= \mathbb{E} \left[ \int_0^{2^\ell \pi} \omega^m \widehat{f}_{\tau_j}(\omega) \, d\omega \right] \\
&= \mathbb{E} \left[ \int_0^{2^\ell \pi} \omega^m \widehat{f}((1 - \tau_j)\omega) \, d\omega \right] \\
&= \mathbb{E} \left[ \int_0^{2^\ell \pi(1-\tau_j)} \frac{\xi^m}{(1 - \tau_j)^m} \widehat{f}(\xi) \frac{d\xi}{(1 - \tau_j)} \right] \\
&= \beta_m(f) \mathbb{E} \left[ (1 - \tau_j)^{-(m+1)} \right].
\end{aligned}$$

We now compute the variance. We first establish that

$$g_m(\ell, \sigma) = \mathbb{E} \left[ \left( \int_0^{2^\ell \pi} \omega^m \widehat{\varepsilon}_j(\omega) \, d\omega \right) \left( \int_0^{2^\ell \pi} \omega^m \overline{\widehat{\varepsilon}_j(\omega)} \, d\omega \right) \right].$$

By Thm 4.5 of [49]

$$\begin{aligned}
\mathbb{E} \left[ \widehat{\varepsilon}_j(\omega) \overline{\widehat{\varepsilon}_j(\xi)} \right] &= \mathbb{E} \left[ \left( \int_{-1/2}^{1/2} e^{-i\omega t} \, dB_t \right) \left( \int_{-1/2}^{1/2} e^{i\xi t} \, dB_t \right) \right] \\
&= \sigma^2 \int_{-1/2}^{1/2} e^{i(\xi - \omega)t} \, dt \\
&= \frac{2\sigma^2 \sin(\frac{1}{2}(\xi - \omega))}{(\xi - \omega)}
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E} \left[ \left( \int_0^{2^\ell \pi} \widehat{\varepsilon}_j(\omega) \, d\omega \right) \left( \int_0^{2^\ell \pi} \overline{\widehat{\varepsilon}_j(\omega)} \, d\omega \right) \right] &= \int_0^{2^\ell \pi} \int_0^{2^\ell \pi} \omega^m \xi^m \mathbb{E} \left[ \widehat{\varepsilon}_j(\omega) \overline{\widehat{\varepsilon}_j(\xi)} \right] \, d\omega \, d\xi \\
&= \int_0^{2^\ell \pi} \int_0^{2^\ell \pi} \omega^m \xi^m \frac{2\sigma^2 \sin(\frac{1}{2}(\xi - \omega))}{(\xi - \omega)} \, d\omega \, d\xi \\
&= g_m(\ell, \sigma).
\end{aligned}$$

We thus obtain

$$\begin{aligned}
\left[ |\beta_m(y_j)|^2 \right] &= \mathbb{E} \left[ \left( \int_0^{2^\ell \pi} \omega^m (\widehat{f}_{\tau_j}(\omega) + \widehat{\varepsilon}_j(\omega)) \, d\omega \right) \left( \int_0^{2^\ell \pi} \omega^m (\overline{\widehat{f}_{\tau_j}(\omega)} + \overline{\widehat{\varepsilon}_j(\omega)}) \, d\omega \right) \right] \\
&= \mathbb{E} \left[ \left( \int_0^{2^\ell \pi} \omega^m \widehat{f}((1 - \tau_j)\omega) \, d\omega \right) \left( \int_0^{2^\ell \pi} \omega^m \overline{\widehat{f}((1 - \tau_j)\omega)} \, d\omega \right) \right. \\
&\quad \left. + \left( \int_0^{2^\ell \pi} \omega^m \widehat{\varepsilon}_j(\omega) \, d\omega \right) \left( \int_0^{2^\ell \pi} \omega^m \overline{\widehat{\varepsilon}_j(\omega)} \, d\omega \right) \right] \\
&= \mathbb{E} \left[ (1 - \tau_j)^{-2(m+1)} \beta_m(f) \overline{\beta_m(f)} \right] + g_m(\ell, \sigma) \\
&= |\beta_m(f)|^2 \mathbb{E} \left[ (1 - \tau_j)^{-2(m+1)} \right] + g_m(\ell, \sigma).
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Var}[\beta_m(y_j)] - g_m(\ell, \sigma) &= \mathbb{E} \left[ |\beta_m(y_j)|^2 \right] - g_m(\ell, \sigma) - |\mathbb{E} \left[ \beta_m(y_j) \right]|^2 \\
&= |\beta_m(f)|^2 \mathbb{E} \left[ (1 - \tau_j)^{-2(m+1)} \right] - |\beta_m(f)|^2 \left( \mathbb{E} \left[ (1 - \tau_j)^{-(m+1)} \right] \right)^2.
\end{aligned}$$

Dividing by  $|\mathbb{E} \left[ \beta_m(y_j) \right]|^2$  gives

$$CV_m = \frac{\mathbb{E}[(1 - \tau_j)^{-2(m+1)}]}{(\mathbb{E}[(1 - \tau_j)^{-(m+1)}])^2} - 1,$$

and the remainder of the proof is identical to the proof of Proposition 6.1.  $\square$

## H. Additional simulations for noisy dilation MRA

We investigate the  $\mathbf{L}^2$  error of estimating the power spectrum using PS ( $k = 0$ ) and WSC ( $k = 0, 2, 4$ ) for three additional high-frequency functions:

$$\begin{aligned}
f_4(x) &= 1.175 \cos(32x) \cdot \mathbf{1}(x \in [-0.2, 0.2]) \\
f_5(x) &= 0.299 \exp^{-0.04x^2} \cos(30x + 1.5x^2) \\
f_6(x) &= (2.304/\pi) \cos(35x) \text{sinc}(3x).
\end{aligned}$$

The multiplicative constants were chosen so that the  $\mathbf{L}^2$  norms of  $f_4, f_5, f_6$  are comparable with the  $\mathbf{L}^2$  norms of the Gabor signals  $f_1, f_2, f_3$  defined in Section 4.4. The signal  $f_4$  is not continuous and has compact support, with a slowly decaying, oscillating Fourier transform given by  $\widehat{f}_4(\omega)/0.47 = \text{sinc}(0.2(\omega - 32)) + \text{sinc}(0.2(-\omega - 32))$ . The signal  $f_5$  is a linear chirp with a constantly varying instantaneous frequency. The signal  $f_6$  is slowly decaying in space, with a discontinuous Fourier transform of compact support given by  $\widehat{f}_6(\omega)/0.384 = \mathbf{1}(\omega \in [-38, -32]) + \mathbf{1}(\omega \in [32, 38])$ .

Implementation details were as described in Section 6, and simulations were run with oracle moment estimation on the full model (parameter values as described in Section 5.2). Figure H8 shows the  $\mathbf{L}^2$  error. As for the high-frequency Gabor in Section 5.2, WSC ( $k = 2$ ) and WSC ( $k = 4$ ) significantly

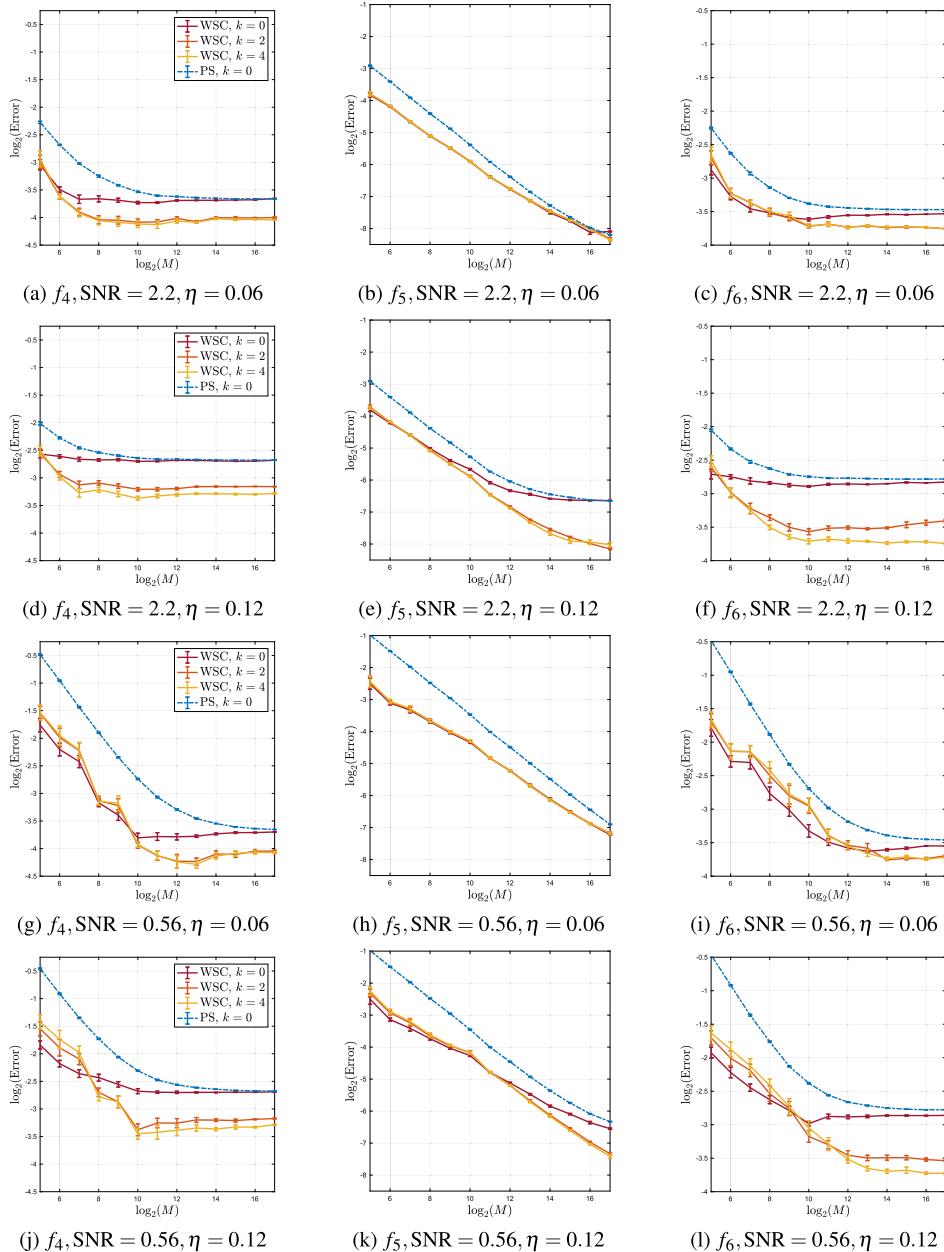


FIG. H8.  $L^2$  error with standard error bars for noisy dilation MRA model (oracle moment estimation). First, second and third columns show results for  $f_4, f_5$  and  $f_6$ . All plots for the same signal have the same axis limits.

outperformed the zero order estimators. In addition for large dilations, the WSC ( $k = 4$ ) outperformed WSC ( $k = 2$ ) on  $f_4$  and  $f_6$ .

## I. Expectation maximization algorithm for noisy dilation MRA

In this appendix we discuss how the expectation-maximization (EM) algorithm proposed in [1] can be extended to solve noisy dilation MRA. We first summarize the EM framework, which differentiates between observed data  $y = \{y_j\}_{j=1}^M$ , latent variables  $s = \{s_j\}_{j=1}^M$  and model parameters  $x$ . The goal is to produce the  $x$  that maximizes the marginalized likelihood function

$$p(y|x) = \int p(y, s|x) ds.$$

Maximizing  $p(y|x)$  directly is generally not tenable because enumerating the various values for  $s$  is too costly. However, EM algorithms can be used to find local maxima of the above function, by iterating between estimating the conditional distribution of latent variables given the current estimate of parameters (E-step) and estimating parameters given the current estimate of the conditional distribution of latent variables (M-step). Specifically, the iterative procedure updates  $x^k$ , the current estimate of  $x$ , by

$$Q(x|x^k) = \mathbb{E}_{s|y,x^k} [\log p(y, s|x)] \quad \text{E-step} \quad (I.1)$$

$$x^{k+1} = \arg \max_x Q(x|x^k) \quad \text{M-step.} \quad (I.2)$$

Since (under certain conditions)  $\log p(y|x)$  improves at least as much as  $Q$  at each iteration [30], the algorithm converges to a local maximum of  $p(y|x)$ . This framework can be applied to noisy dilation MRA, and explicit formulas for both the E-step and M-step can be derived. Assume for simplicity that signals have been discretized to have length  $n$  and that the translation distribution  $\rho_t$  and dilation distribution  $\rho_\tau$  are unknown and also discrete with  $n$  possible values  $\{t^\ell\}_{\ell=1}^n$  and  $\{\tau^q\}_{q=1}^n$ , respectively. Letting  $x = (f, \rho_t, \rho_\tau)$  denote the parameters,  $s_j = (t_j, \tau_j)$  denote the latent/nuisance variables, and  $p_x$  denote conditioning on  $x$ , the likelihood function has form

$$p(y, s|x) = p_x(y|s)p_x(s) = \prod_{j=1}^M \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|L_{\tau_j} T_{t_j} f - y_j\|_2^2\right) \rho_t(t_j) \rho_\tau(\tau_j).$$

Thus (up to a constant), the log likelihood has form

$$\log p(y, s|x) = \sum_{j=1}^M -\frac{1}{2\sigma^2} \|L_{\tau_j} T_{t_j} f - y_j\|_2^2 + \sum_{j=1}^M \log \rho_t(t_j) + \sum_{j=1}^M \log \rho_\tau(\tau_j). \quad (I.3)$$

Given the current estimate  $x^k = (f^k, \rho_t^k, \rho_\tau^k)$  of parameters, the E-step is performed by first computing the conditional distribution of the latent variables

$$w_k^{\ell,q,j} = \mathbb{P}(t_j = t^\ell, \tau_j = \tau^q | x^k) = C_k^j \exp\left(-\frac{1}{2\sigma^2} \|L_{\tau_j} T_{t_j} f^k - y_j\|_2^2\right) \rho_t^k(t^\ell) \rho_\tau^k(\tau^q), \quad (I.4)$$

where  $C_k^j$  is a normalizing constant so that  $\sum_{\ell,q} w_k^{\ell,q,j} = 1$ . These weights are then used to compute  $Q$ , that is, by combining (I1), (I3) and (I4):

$$Q(f, \rho_t, \rho_\tau | f^k, \rho_t^k, \rho_\tau^k) = \sum_{j=1}^M \sum_{\ell=1}^n \sum_{q=1}^n w_k^{\ell,q,j} \left( -\frac{1}{2\sigma^2} \|L_{\tau_j} T_{t_j} f - y_j\|_2^2 + \log \rho_t(t^\ell) + \log \rho_\tau(\tau^q) \right), \quad (\text{I.5})$$

up to a constant. The M-step is then computed by

$$(f^{k+1}, \rho_t^{k+1}, \rho_\tau^{k+1}) = \arg \max_{f, \rho_t, \rho_\tau} Q(f, \rho_t, \rho_\tau | f^k, \rho_t^k, \rho_\tau^k). \quad (\text{I.6})$$

Since  $f, \rho_t, \rho_\tau$  all appear in distinct sums in (I5), performing the maximization in (I6) is straightforward. Since  $\|L_{\tau_j} T_{t_j} f - y_j\|_2^2 = \frac{1}{1-\tau_q} f - T_\ell^{-1} L_\tau^{-1} y_j \|_2^2$ , it is easy to check that

$$f^{k+1} = \frac{1}{C} \sum_{j=1}^M \sum_{\ell=1}^n \sum_{q=1}^n \frac{w_k^{\ell,q,j}}{(1-\tau_q)} T_\ell^{-1} L_\tau^{-1} y_j, \quad C = \sum_{j=1}^M \sum_{\ell=1}^n \sum_{q=1}^n \frac{w_k^{\ell,q,j}}{(1-\tau_q)}. \quad (\text{I.7})$$

Using Lemma 15 in [1], one can also obtain closed form expressions for the updates to  $\rho_t^k, \rho_\tau^k$ :

$$\rho_t^{k+1}(t^\ell) = \frac{\tilde{w}_k^\ell}{\sum_{\ell'} \tilde{w}_k^{\ell'}} \text{ for } \tilde{w}_k^\ell = \sum_j \sum_q w_k^{\ell,q,j}, \quad \rho_\tau^{k+1}(\tau^q) = \frac{\tilde{v}_k^q}{\sum_{q'} \tilde{v}_k^{q'}} \text{ for } \tilde{v}_k^q = \sum_j \sum_\ell w_k^{\ell,q,j}.$$

Note when a discrete signal defined on some fixed grid is dilated, its dilation is defined on a different grid. Thus, computing (I.4) and (I.7) will involve off-grid interpolation, a subtlety not arising in classic MRA, and this interpolation may contribute additional error. We also note that one can always force the translation distribution to be uniform by retranslating the signals uniformly, and in this case all sums over  $\ell$  in this section could be eliminated. This would improve the computational complexity of the algorithm but may be disadvantageous in terms of sample complexity, as in classic MRA a uniform translation distribution requires a larger sample size for accurate estimation than an aperiodic translation distribution [1].

## J. Supporting results: stochastic calculus

This appendix contains several stochastic calculus results that are used to control the statistics of the additive noise. Proposition J.1 is a simple generalization of Thm 4.5 of [49]. Proposition J.2 controls the second moment of the stochastic quantity in Proposition J.1 and is in fact a special case of Proposition J.3. Both Propositions J.2 and J.3 are proved with standard techniques from stochastic calculus, and for brevity we omit the proofs.

**PROPOSITION J.1** Assume  $\int_0^T f(t)^2 dt < \infty$ ,  $\int_0^T \overline{f(t)^2} dt < \infty$ , and let  $B_t$  be a Brownian motion with variance  $\sigma^2$ . Then,

$$\mathbb{E} \left[ \left( \int_0^T f(t) dB_t \right) \left( \int_0^T \overline{f(t)^2} dB_t \right) \right] = \sigma^2 \int_0^T f(t) \overline{f(t)^2} dt.$$

PROPOSITION J.2 Let  $f(t)$  be a bounded and continuous complex deterministic function on  $[0, T]$ , and let  $B_t$  be a Brownian motion with variance  $\sigma^2$ . Then, for a fixed non-random time  $T$ , we have

$$\mathbb{E} \left[ \left( \int_0^T f(t) dB_t \right)^2 \left( \int_0^T \overline{f(t)} dB_t \right)^2 \right] = 2\sigma^4 \left( \int_0^T |f(t)|^2 dt \right)^2 + \sigma^4 \left( \int_0^T f(t)^2 dt \right) \left( \int_0^T \overline{f(t)}^2 dt \right).$$

COROLLARY J.1 When  $f(t)$  is real, the above reduces to

$$\mathbb{E} \left[ \left( \int_0^T f(t) dB_t \right)^4 \right] = 3\sigma^4 \left( \int_0^T f(t)^2 dt \right)^2.$$

PROPOSITION J.3 Let  $f(t), g(t)$  be bounded and continuous complex deterministic functions on  $[0, T]$ , and let  $B_t$  be a Brownian motion with variance  $\sigma^2$ . Then, for a fixed non-random time  $T$ , we have

$$\begin{aligned} & \mathbb{E} \left[ \left( \int_0^T f(t) dB_t \right) \left( \int_0^T \overline{f(t)} dB_t \right) \left( \int_0^T g(t) dB_t \right) \left( \int_0^T \overline{g(t)} dB_t \right) \right] \\ &= \sigma^4 \left[ \left( \int_0^T f(t)g(t) dt \right) \left( \int_0^T \overline{f(t)g(t)} dt \right) + \left( \int_0^T f(t)\overline{g(t)} dt \right) \left( \int_0^T \overline{f(t)}g(t) dt \right) \right. \\ & \quad \left. + \left( \int_0^T |f(t)|^2 dt \right) \left( \int_0^T |g(t)|^2 dt \right) \right]. \end{aligned}$$