Wavelet invariants for statistically robust multi-reference alignment

MATTHEW HIRN
Department of Computational Mathematics, Science, and Engineering,
Department of Mathematics,
Michigan State University, East Lansing, MI
mhirn@msu.edu

AND

ANNA LITTLE
Department of Computational Mathematics, Science, and Engineering,
Michigan State University, East Lansing, MI
littl119@msu.edu

October 31, 2019

Abstract

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 obtain the target signal up to an unknown phase. Extensive numerical experiments demonstrate the statistical robustness of this approximation procedure. 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 signal $f : \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}) \rightarrow \mathbb{R}$:

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

where:

(i) $\text{supp}(y_j) \subseteq [-\frac{1}{2}, \frac{1}{2}]$ 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 $\sigma^2$.

1
The signal is thus subjected to both random translation and additive noise. The MRA problem arises in numerous applications, including structural biology [1–6], single cell genomic sequencing [7], radar [8, 9], crystalline simulations [10], image registration [11–13], and signal processing [8]. It is a simplified model relevant for Cryo-Electron Microscopy (Cryo-EM), an imaging technique for molecules which achieves near atomic resolution [14–16]. In this application one seeks to recover a 3d reconstruction of the molecule from many noisy 2d images/projections [17]. Although MRA ignores the tomographic projection of Cryo-EM, investigation of the simplified model provides important insights. For example, [18, 19] 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 [20, 21]. Recent work has established an improved sample complexity of \( M = \Theta(\sigma^4) \) for MRA when the translation distribution is aperiodic [22], and this rate has been shown to also hold in the more complicated setting of Cryo-EM, if the viewing angles are nonuniformly distributed [23]. 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 \( f_1, \ldots, f_k \) [19, 24–26], 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 [27].

Approaches for solving MRA generally fall into two categories: synchronization methods and the method of invariants. Synchronization methods [28–30] attempt to recover the signal by aligning the translations and then averaging, but these methods fail in the low signal-to-noise regime. The method of invariants 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 [31, 32]. Recent work [16, 18] utilizes such Fourier invariants (mean, power spectrum, and bispectrum), and recovers \( \hat{f} \) by solving a nonconvex optimization problem on the manifold of phases.

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

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

where \( L_{\tau} \) is a dilation operator which 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). The generalized MRA model captures small stretches of the molecules in Cryo-EM, and is thus less restrictive than the classic MRA model. 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, [33–35] consider a stationary random signal \( f(x) \) which is time-warped, i.e. \( D_\tau f(x) = \sqrt{\zeta'(x)}f(\zeta(x)) \), and uses a maximum likelihood approach to estimate \( \zeta \). In [36,37], a similar stochastic time warping model is analyzed using wavelet based techniques. The generalized MRA model considered here corresponds to the simplest case of time-warping, when \( \zeta \) is an affine function.

Fourier invariants will fail for the generalized MRA problem, since they are 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 [38], 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 of the wavelet coefficients, thus providing rigid invariants which can be statistically unbiased. Similar invariant coefficients have been utilized in a number of applica-
tions including predicting molecular properties [39, 40] and quantum chemical energies [41], and in microcanonical ensemble models for texture synthesis [42]. Recent work [43] has also generalized such coefficients to graphs.

1.1 Notation

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

\[
\hat{f}(\omega) = \int f(x) e^{-ix\omega} dx.
\]

The power spectrum is the nonlinear transform \( P : L^2(\mathbb{R}) \to L^1(\mathbb{R}) \) that maps \( f \) to

\[
(Pf)(\omega) = |\hat{f}(\omega)|^2, \quad \omega \in \mathbb{R}.
\]

We denote \( f(x) \leq C g(x) \) for some absolute constant \( C \) by \( f(x) \lesssim g(x) \). We also write \( f(x) = O(g(x)) \) if \( |f(x)| \leq C g(x) \) for all \( x \geq x_0 \) for some constants \( x_0, C > 0 \); \( f(x) = \Theta(g(x)) \) denotes \( C_1 g(x) \leq |f(x)| \leq C_2 g(x) \) for all \( x \geq x_0 \) for some constants \( x_0, C_1, C_2 > 0 \). The minimum of \( a \) and \( b \) is denoted \( a \wedge b \).

2 MRA models and the method of invariants

Standard multi-reference alignment (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 \( L^2(\mathbb{R}) \) wavelet coefficient invariants. Theorem 2.4 proves wavelet coefficient invariants computed with a continuous wavelet transform and a suitable mother wavelet have the same capacity as 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 multi-reference alignment (MRA) scenario considers the problem of recovering a signal \( f \in L^2(\mathbb{R}) \) 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 is low, as registration methods become intractable. In [16,18,25,26,44,45] 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 [46] consider 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:

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

where \( \xi_j \in C^1(\mathbb{R}) \) 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 \( L_{\xi} \) is the action of the diffeomorphism \( \xi \) on \( f \),

\[
L_{\xi} f(x) = f(\xi(x)).
\]

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

One of the keys to the Fourier invariant approach of [16,18,25,26,44,45] 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) 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 generalized MRA model.

Model 2 (Generalized MRA data model) The generalized MRA data model consists of \( M \) independent observations of a signal \( f: L^1 \cap L^2(\mathbb{R}) \to \mathbb{R} \):

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

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

\[
L_{\tau} f(x) = (1 - \tau)^{-1} f((1 - \tau)^{-1} x).
\]

In addition, we assume:

(i) \( \text{supp}(y_j) \subseteq [-\frac{1}{2}, \frac{1}{2}] \) 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) \( \{\epsilon_j(x)\}_{j=1}^M \) are independent white noise processes on \([-\frac{1}{2}, \frac{1}{2}]\) with variance \( \sigma^2 \).

Remark 2.1 The interval \([-\frac{1}{2}, \frac{1}{2}]\) is arbitrary, i.e. it can be replaced with any interval of length 1. In addition, the spatial box size is arbitrary, i.e. \([-\frac{1}{2}, \frac{1}{2}]\) can be replaced with \([-\frac{N}{2}, \frac{N}{2}]\). 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 \( \epsilon(x) \) is a white noise process if \( \epsilon(x) = dB_x \), i.e. it is the derivative of a Brownian motion with variance \( \sigma^2 \).

While the generalized 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 \( L_{\tau} \hat{f}(\omega) \) relative to \( \hat{f}(\omega) \) 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 [18]. 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 signal $f : L^1 \cap L^2(\mathbb{R}) \to \mathbb{R}$:

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

(4)

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

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

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 translation by $t$ operator acting on a signal $f$. Invariant measurement models seek a representation $\Phi(f) \in \mathcal{B}$ in a Banach space $\mathcal{B}$ of the signal $f$ such that

$$\Phi(T_t f) = \Phi(f), \quad \forall t \in \mathbb{R}. \quad (5)$$

In MRA problems, one additionally requires that

$$\Phi(f) = \Phi(g) \iff g = T_t f \text{ for some } t \in \mathbb{R}. \quad (6)$$

The first condition (5) removes the need to align random translations of the signal $f$, whereas the second condition (6) 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

$$f^* = \arg \inf_{g \in L^1 \cap L^2(\mathbb{R})} \|\Phi(g) - \Phi(f)\|_{\mathcal{B}}, \quad (7)$$

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

$$\overline{\Phi}_M(f) = \frac{1}{M} \sum_{j=1}^M \Phi(y_j), \quad (8)$$

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

$$\mathbb{E} \Phi(y_j) = \Phi(f) + \tilde{b}_\Phi(f, \mathcal{M}), \quad (9)$$

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 (9) 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 (8) to reduce the bias:

$$\overline{\Phi}_M(f) = \frac{1}{M} \sum_{j=1}^M (\Phi(y_j) - \tilde{b}_\Phi(y_j, \mathcal{M})). \quad (5)$$
in which case
\[
\lim_\limits_{M \to \infty} \tilde{\Phi}_M(f) = \Phi(f) + \delta
\]
by the law of large numbers. The main difficulty therefore is twofold. On the one hand, one must design a representation \( \Phi \) that satisfies (5), (6), and (9) with a bias \( b \) that can be estimated; on the other hand, the optimization (7) must be tractable. For random translation plus additive noise models (i.e., Model 1), the authors of [16, 18] describe a representation \( \Phi \) based on Fourier invariants that satisfies the outlined requirements and for which one can solve (7) despite the optimization being non-convex. The Fourier invariants include \( \hat{f}(0) \) (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 \( \hat{f}(0) \) carries limited information, the power spectrum recovers \( f \) up to its phase, which in some cases can be resolved; results along these lines are in the field of phase retrieval [47, 48]. The power spectrum is invariant to translations since the Fourier modulus kills the phase factor induced by a translation \( t \) of \( f \). The bispectrum is also translation invariant and invertible so long as \( \hat{f}(\omega) \neq 0 \) [19].

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 \( L^2(\mathbb{R}) \) norms of the wavelet coefficients of the signal \( f \). These invariants satisfy (5) and (9) for Model 2, and yield a convex formulation of (7). They do not satisfy (6) for general \( f \in L^2(\mathbb{R}) \), 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.

2.2.2 Wavelet invariants

We now define the wavelet invariants used in this article. A wavelet \( \psi \in L^2(\mathbb{R}) \) is a waveform that is localized in both space and frequency and has zero average,
\[
\int \psi(x) \, dx = 0.
\]
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 \( \| \psi_\lambda \|_2 = \| \psi \|_2 \); we assume \( \| \psi \|_2 = 1 \). The continuous wavelet transform \( W \) computes
\[
W f = \{ f * \psi_\lambda(x) : \lambda \in (0, \infty), x \in \mathbb{R} \}.
\]
The parameter \( \lambda = 2^\alpha \) for \( \alpha \in \mathbb{R} \) corresponds to a \( \log_2 \)-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 \( \lambda \xi_0 \). 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 \( L^2(\mathbb{R}) \) norms in the \( x \) variable, yielding the following nonlinear wavelet invariants:

**Definition 2.1** (Wavelet invariants) The \( L^2 \) wavelet invariants of a signal \( f : L^1 \cap L^2(\mathbb{R}) \to \mathbb{R} \) are given by
\[
(S f)(\lambda) = \| f * \psi_\lambda \|_2^2, \quad \lambda \in (0, \infty),
\]
where $\psi_\lambda(x) = \lambda^{1/2} \psi(\lambda x)$ are dilations of a mother wavelet $\psi$. The wavelet $\psi$ is $k$-admissable if $\hat{\psi} \in \mathbb{C}^k$ and $\Psi_k < \infty, \Theta_k < \infty$ where

$$
\Psi_k := \frac{1}{2\pi} \sum_{i=0}^{k} \frac{k!}{i!} \| \omega^i (P\psi)^{\langle i \rangle}(\omega) \|_1, \tag{11}
$$

$$
\Theta_k := \frac{1}{2\pi} \sum_{i=0}^{k} \frac{k!}{i!} \| \omega^{i-2} (P\psi)^{\langle i \rangle}(\omega) \|_1. \tag{12}
$$

The wavelet invariants can be expressed in the frequency domain as

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

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:

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

For $\psi$ to be $k$-admissable, it is sufficient for $\hat{\psi} \in \mathbb{C}^k$, $(P\psi)^{\langle i \rangle}$ to decay faster than $\omega^{i+1}$, and $\int \frac{|\hat{\psi}(\omega)|^2}{\omega^2} \, d\omega < \infty$ (see Lemma A.1 in Appendix A). The condition $\int \frac{|\hat{\psi}(\omega)|^2}{\omega} \, d\omega < \infty$ is slightly stronger than the classic admissability condition $C_\psi := \int \frac{|\hat{\psi}(\omega)|^2}{\omega} \, d\omega < \infty$ [49, Theorem 4.4]. When $\hat{\psi}$ is continuously differentiable, $\hat{\psi}(0) = 0$ is sufficient to guarantee $C_\psi < \infty$; but here we need $\hat{\psi} \sim \omega^{\frac{1}{2}+\epsilon}$ for some $\epsilon > 0$ as $\omega \to 0$. 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 $\Lambda_k(\lambda) = \Psi_k \| f \|_1^2 \wedge \Theta_k \| f \|_1^2$ in Propositions 4.2 and 5.1.

**Remark 2.2** The Morlet wavelet $\psi(x) = g(x)(e^{ix} - 1)$ is $k$-admissable for any $k$, since $\hat{\psi} \in \mathbb{C}^\infty$, $P\psi$ has fast decay, and $\hat{\psi}(\omega) \sim \omega^{\frac{1}{2}+\epsilon}$ for some $\epsilon > 0$.

**Remark 2.3** Definition 2.1 assumes $f \in \mathbb{R}$, 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 $\lambda \in (-\infty, \infty)$.

**Remark 2.4** 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 $O(\log n)$’s), the computational cost is the same and stability is maintained, but more information is lost.

**Remark 2.5** When $Pf$ 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.

When $C_\psi < \infty$, one can show that $S : L^2(\mathbb{R}) \to L^1 \cap C(0, \infty)$. The values $\lambda = 2^j$ for $j \in \mathbb{Z}$ correspond to rigid versions of first order $L^2(\mathbb{R})$ wavelet scattering invariants [38]. 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 $\lambda = 2^j$ loses a considerable amount of information from $f$. When one utilizes every $\log_2$-frequency $\lambda$, though, the resulting first order scattering coefficients uniquely determine the power spectrum of $f$, so long as the wavelet $\psi$ satisfies a type of independence condition.
Condition 2.3 Define
\[ |\hat{\psi}_\lambda^+(\omega)|^2 = (|\hat{\psi}_\lambda(\omega)|^2 + |\hat{\psi}_\lambda(-\omega)|^2) \cdot 1(\omega \geq 0). \]
If for any finite sequence \( \{\omega_i\}_{i=1}^n \) of distinct positive frequencies, the collection \( \{|\hat{\psi}_\lambda^+(\omega_i)|^2\}_{i=1}^n \) are linearly independent functions of \( \lambda \), we say \( \psi \) satisfies the linear independence condition.

**Remark 2.6** Condition 2.3 is stated in terms of \( |\hat{\psi}_\lambda^+(\omega)|^2 \) to avoid assumptions on whether \( \psi \) is real or complex. When \( \psi \in \mathbb{R} \), \( |\hat{\psi}_\lambda^+(\omega)|^2 = 2|\hat{\psi}_\lambda(\omega)|^2 \) for \( \omega \geq 0 \). When \( \psi \) is complex analytic, \( |\hat{\psi}_\lambda^+(\omega)|^2 = |\hat{\psi}_\lambda(\omega)|^2 \). When \( \psi \in \mathbb{C} \) but not complex analytic, \( |\hat{\psi}_\lambda^+(\omega)|^2 \) simply incorporates a reflection of \( |\hat{\psi}_\lambda(\omega)|^2 \) about the origin. Since we assume \( f \in \mathbb{R} \), \( |\hat{\psi}_\lambda^+(\omega)|^2 \) uniquely defines \((Sf)(\lambda)\), since \((Sf)(\lambda) = \frac{1}{2\pi} \langle |\hat{f}|^2, |\hat{\psi}_\lambda^+|^2 \rangle \) by the Plancherel and Fourier convolution theorems.

**Theorem 2.4** Let \( f, g \in L^1 \cap L^2(\mathbb{R}) \) and assume that \( \psi \) satisfies Condition 2.3. Then:
\[ Sf = Sg \iff Pf = Pg. \]

**Proof.** First assume \( Pf = Pg \), which means \( |\hat{f}(\omega)|^2 = |\hat{g}(\omega)|^2 \) for almost every \( \omega \in \mathbb{R} \). Using the Plancherel and Fourier convolution theorems,
\[ (Sf)(\lambda) = \int |f * \psi_\lambda(x)|^2 \, dx = \frac{1}{2\pi} \int |\hat{f}(\omega)|^2 |\hat{\psi}_\lambda(\omega)|^2 \, d\omega \]
\[ = \frac{1}{2\pi} \int |\hat{g}(\omega)|^2 |\hat{\psi}_\lambda(\omega)|^2 \, d\omega = (Sg)(\lambda), \quad \forall \lambda \in (0, \infty). \]

Now suppose \( Sf = Sg \). Since \( Sf \) and \( Sg \) are continuous in \( \lambda \), we have:
\[ 0 = (Sf)(\lambda) - (Sg)(\lambda) = \frac{1}{2\pi} \int (|\hat{f}(\omega)|^2 - |\hat{g}(\omega)|^2) |\hat{\psi}_\lambda(\omega)|^2 \, d\omega, \quad \forall \lambda \in (0, \infty). \]
Since \( f \in L^1 \cap L^2(\mathbb{R}) \) we have \( \hat{f} \in L^2 \cap L^\infty(\mathbb{R}) \) and thus \( Pf \in L^1 \cap L^\infty(\mathbb{R}) \). By interpolation we have \( Pf \in L^2(\mathbb{R}) \), and the same for \( Pg \). By applying Lemma 2.1 (stated below) with \( p(\omega) = (Pf)(\omega) - (Pg)(\omega) \), we conclude \( Pf = Pg \) for almost every \( \omega \). \( \square \)

**Lemma 2.1** Let \( p \in L^2 \) and assume \( p(\omega) = p(-\omega) \) and Condition 2.3. Then
\[ \int p(\omega)|\hat{\psi}_\lambda(\omega)|^2 \, d\omega = 0 \quad \forall \lambda > 0 \implies p = 0 \text{ a.e.} \]

The proof of Lemma 2.1 is in Appendix B. We remark that most wavelets satisfy Condition 2.3 (for example Morlet wavelets; see Lemma B.1 in Appendix B), so Theorem 2.4 is broadly applicable. The following Proposition, proved in Appendix B, gives some sufficient conditions guaranteeing Condition 2.3.

**Proposition 2.5** The following are sufficient to guarantee Condition 2.3:

(i) \( |\hat{\psi}(\omega)|^2 \) 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) \( |\hat{\psi}(\omega)|^2 \in C^\infty \) and there exists an \( N \) such that all derivatives of order at least \( N \) are nonzero at \( \omega = 0 \), e.g., the Morlet wavelet.

3 Unbiasing for classic MRA

We first consider the classic MRA model (Model 1). Section 3.1 discusses unbiasing results for the power spectrum, Section 3.2 discusses unbiasing results for wavelet invariants, and Section 3.3 gives simulation results comparing the two methods.
3.1 Power spectrum results for classic MRA

In this section we rederive some results from [16], extended to the continuum setting. These results utilize the following lemma, which is proved in Appendix C.

**Lemma 3.1** Let \( \epsilon(x) \) be a white noise processes on \([-\frac{1}{2}, \frac{1}{2}] \) with variance \( \sigma^2 \). Then for any signal \( f \in L^1(\mathbb{R}) \):

\[
\mathbb{E}[(P(f + \epsilon))(\omega)] = (Pf)(\omega) + \sigma^2 \\
\text{Var}[(P(f + \epsilon))(\omega)] \leq 4\sigma^2(Pf)(\omega) + 2\sigma^4.
\]

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

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

Then with probability at least \( 1 - 1/t^2 \),

\[
|(Pf)(\omega) - (\hat{P}f)(\omega)| \leq \frac{2t\sigma}{\sqrt{M}} (\|f\|_1 + \sigma).
\] (13)

**Proof.** Let \( f^{t_j}(x) = f(x - t_j) \) so that \( y_j = f^{t_j} + \epsilon_j \). We first note since \( \hat{f}^{t_j}(\omega) = e^{-i\omega t_j}\hat{f}(\omega) \), the power spectrum is translation invariant, that is \((Pf^{t_j})(\omega) = (Pf)(\omega)\) for all \( \omega, t_j \). Thus by Lemma 3.1,

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

and

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

Since the \( y_j \) are independent,

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

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 \( \sqrt{(Pf)(\omega)} = |\hat{f}(\omega)| \leq \|f\|_1 \) gives (13). \( \square \)

3.2 Wavelet invariant results for classic MRA

We obtain an identical result for wavelet invariants (Proposition 3.2) when signals are corrupted by additive noise only. The proof utilizes the following lemma, which is proved in Appendix C.

**Lemma 3.2** Let \( \epsilon(x) \) be a white noise processes on \([-\frac{1}{2}, \frac{1}{2}] \) with variance \( \sigma^2 \). Then for any signal \( f \in L^1(\mathbb{R}) \):

\[
\mathbb{E}[(S(f + \epsilon))(\lambda)] = (Sf)(\lambda) + \sigma^2 \\
\text{Var}[(S(f + \epsilon))(\lambda)] \leq 4\sigma^2(Sf)(\lambda) + 2\sigma^4.
\]
**PROPOSITION 3.2** Assume Model 1. Define the following estimator of \((Sf) (\lambda)\):

\[
(\widetilde{Sf}) (\lambda) := \frac{1}{M} \sum_{j=1}^{M} (Sy_j)(\lambda) - \sigma^2.
\]

Then with probability at least \(1 - 1/t^2\),

\[
| (Sf)(\lambda) - (\widetilde{Sf})(\lambda) | \leq \frac{2t\sigma}{\sqrt{M}} (\|f\|_1 + \sigma).
\]

*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 3.2:

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

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,

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

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

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

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 (14). □

As \(M \to \infty\), 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 generalized MRA (Model 2), as there is a nonvanishing bias term. However a nonlinear unbiasing procedure on the wavelet invariants can significantly reduce the bias.

### 3.3 Simulation results for classic MRA

We illustrate and compare additive noise unbiasing for power spectrum estimation using the power spectrum method of Proposition 3.1 and the wavelet invariant method of Proposition 3.2. Figure 1a 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). Figure 1b shows the \(L^2\) error of the power spectrum estimation for the two methods as a function of \(\log_2(M)\) for a fixed \(\sigma\), and Figure 1c shows the \(L^2\) error as a function of \(\log_2(\sigma)\) for a fixed \(M\). The \(L^2\) 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.
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. Random dilations cause $\frac{1}{M}\sum_{j=1}^{M}(P y_j)(\omega)$ and $\frac{1}{M}\sum_{j=1}^{M}(S y_j)(\lambda)$ to be biased estimators of $(P f)(\omega)$ and $(S f)(\lambda)$, 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 $(P f), (S f)$ 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 $i = 2, 4, \ldots, k$. Note since we assume $\mathbb{E}[\tau^2] = \eta^2, C_2 = 1$. We define the constants $B_2, B_4, \ldots, B_k$ by solving

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

for $i = 2, 4, \ldots, k$; these constants are deterministic functions of the moments of $\tau$. A nonrecursive formula related to the Euler numbers can be derived which defines $B_i$ explicitly in terms of $C_2, \ldots, C_i$; however the recursive formula (15) is easier to implement numerically.

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

$$T := \max_{i=0, 2, \ldots} \frac{C_i^i}{i!}$$

and

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

where $C_0, B_0 = 1$, and when $i = j = 0$ in (17), $\left( \frac{T^j}{j!} \left| B_i \right| \right)_{j=0}^{k+2-i}$ 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 D, 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 $L \in \mathbb{C}^{k+2}(0, \infty)$ and a random variable $\tau$ satisfying...
Define \( |\lambda^i L^{(i)}(\lambda)| \leq \Lambda_i(\lambda) \) for \( 0 \leq i \leq k+2 \), \( \frac{\Lambda_{k+2}((1-\tau)\lambda)}{\Lambda_{k+2}(\lambda)} \leq R \), for some absolute constant \( R \) and define the following estimator of \( L(\lambda) \):

\[
G_{\lambda}(\tau) := F_{\lambda}(\tau) - B_2\eta^2 F''_{\lambda}(\tau) - B_4\eta^4 F^{(4)}_{\lambda}(\tau) - \ldots - B_k\eta^k F^{(k)}_{\lambda}(\tau).
\]

Then \( G_{\lambda}(\tau) \) satisfies

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

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

where

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

and \( E \) is an absolute constant defined in (17).

**Lemma 4.2** Let the assumptions and notation of Lemma 4.1 hold, and let \( \tau_1, \ldots, \tau_M \) be independent. Define:

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

Then with probability at least \( 1 - 1/t^2 \)

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

The deviation of the estimator \( \tilde{L}(\lambda) \) from \( L(\lambda) \) thus depends on two things: (1) the bias of the estimator which is \( O(\eta^{k+2}) \) and (2) the standard deviation of the estimator which 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 \( Pf \in C^{k+2} \). Define the following estimator of \( (Pf)(\omega) \):

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

where the constants \( B_i \) satisfy (15). Let:

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

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

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

where

\[
\Omega(\omega) = \sum_{0 \leq i, j \leq k+2, i+j \geq 2} \Omega_i(\omega)\Omega_j(\omega)(2\eta)^{i+j}.
\]
where the constants \( \Theta \) for MRA in terms of the following quantities: \( \Psi \) for \( \omega \) where the error may depend on the frequency \( \omega \) (see (18) 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 A.

**Lemma 4.3** [Low Frequency Bound] Assume \( P\psi \in C^m \) and \( f \in L^1 \). Then the quantity \( |\lambda^m(Sf)(\lambda)| \) can be bounded uniformly over all \( \lambda \). Specifically:

\[
|\lambda^m(Sf)(\lambda)| \leq \Psi_m \|f\|_1^2
\]

for \( \Psi_m \) defined in (11).

**Lemma 4.4** [High Frequency Bound for Differentiable Functions] Assume \( P\psi \in C^m \), and \( f' \in L^1 \). Then the quantity \( |\lambda^m(Sf)(\lambda)| \) can be bounded by:

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

for \( \Theta_m \) defined in (12).

These lemmas allow one to bound the error of the order \( k \) wavelet invariant estimator for dilation MRA in terms of the following quantities:

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

where \( E \) is defined in (17).

**Proposition 4.2** Assume Model 3, the notation in (19), and that \( \psi \) is \((k+2)\)-admissible. Define the following estimator of \((Sf)(\lambda)\):

\[
(Sf')(\lambda) := \frac{1}{M} \sum_{j=1}^{M} \left( (Sy_j)(\lambda) - B_j \eta^2 \lambda^2 (Sy_j)''(\lambda) - \ldots - B_k \eta^k \lambda^k (Sy_j)^{(k)}(\lambda) \right)
\]

where the constants \( B_i \) satisfy (15). Then with probability at least \( 1 - 1/r^2 \),

\[
|\lambda^i(Sf)(\lambda) - (Sf)(\lambda)| \lesssim k \left( \Lambda_{k+2}(\lambda)(2E\eta)^{k+2} + \frac{t\Lambda(\lambda)}{\sqrt{M}} \right).
\]

**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, \( \tilde{\psi} \in C^{k+2} \) which guarantees \((Sf)(\lambda) \in C^{k+2}\). We note that since \( f \in L^1 \), \( Pf \) is continuous, and the Leibniz integral rule guarantees that \( (Sf)(\lambda) = \frac{d^n}{d\lambda^n}(Sf)(\lambda) \) for \( 1 \leq n \leq k+2 \). By applying Lemma 4.3, we have \( |\lambda^i(Sf)(\lambda)| \leq \Psi_i \|f\|_1^2 \) for all \( 0 \leq i \leq k + 2 \).
so that Lemma 4.2 holds for \( L(\lambda) = (Sf)(\lambda) \), \( \Lambda(\lambda) = \Psi_1 f \| f \|_1^2 \), and \( R = 1 \). Now by applying Lemma 4.4, we have \( |\lambda^2(Sf)(i)(\lambda)| \leq \Theta_i f' \| f' \|_1^2 / \lambda^2 \) for all \( 0 \leq i \leq k + 2 \), so that Lemma 4.2 also holds for \( L(\lambda) = (Sf)(\lambda) \), \( \Lambda(\lambda) = \Theta_i f' \| f' \|_1^2 / \lambda^2 \), and \( R = 4 \) (note since \( |\tau| \leq 1/2 \), \( \Lambda_{k+2}((1-\tau)\lambda)/\Lambda_{k+2}(\lambda) \leq 4 \). Thus Lemma 4.2 in fact holds with \( \Lambda(\lambda) = \left( \Psi_i f' \| f' \|_1^2 \wedge \Theta_i f' \| f' \|_1^2 \right) \); since \( (Sy_j)(\lambda) = (Sf)((1-\tau_j)\lambda) = F_\lambda(\tau_j) \), we obtain Proposition 4.2.

Since \( \Lambda(\lambda) \leq \Psi_i f \| f \|_1^2 \), 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 \( f' \notin L^1 \), and \( \Lambda_{k+2}(\lambda) = \Psi_{k+2} f \| f \|_1^2 \). 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 (11), \( \Psi_k \) 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 \( k \Psi_{k+2} (2\eta^6)^{k+2} \). 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 \( \Psi_i(2\eta^6)^j \) is decreasing for \( i \leq k + 2 \), then with probability at least \( 1 - 1/t^2 \):

\[
|\tilde{Sf}(\lambda) - (Sf)(\lambda)| \lesssim f \| f \|_1^2 \left( k \Psi_{k+2}(2\eta^6)^{k+2} + \frac{tk^2 \eta}{\sqrt{M}} \right).
\]

Similarly, if \( \Theta_i(2\eta^6)^j \) is decreasing for \( i \leq k + 2 \), then with probability at least \( 1 - 1/t^2 \):

\[
|\tilde{Sf}(\lambda) - (Sf)(\lambda)| \lesssim \frac{f \| f \|_1^2}{\lambda^2} \left( k \Theta_{k+2}(2\eta^6)^{k+2} + \frac{tk^2 \eta}{\sqrt{M}} \right).
\]

**Remark 4.2** We observe that for a discrete lattice \( L \) of \( \lambda \) values, we can define the discrete 1-norm by \( \| g \|_{L^1(L)} = \sum_{\lambda \in L} |g(\lambda)| \Delta \lambda \). Assume the lattice has cardinality \( n \), and that \( \Psi_i(2\eta^6)^j, \Theta_i(2\eta^6)^j \) are decreasing for \( i \leq k + 2 \). Applying Proposition 4.2 with \( t = \sqrt{n}s \) and a union bound over the lattice gives

\[
\| \tilde{Sf} - Sf \|_{L^1(L)} \lesssim k \left( \| f \|_1^2 \Psi_{k+2} + \| f' \|_1^2 \Theta_{k+2} \right) (2\eta^6)^{k+2} + \frac{s \sqrt{n}k^2 \eta^2}{\sqrt{M}} \left( \| f \|_1^2 + \| f' \|_1^2 \right)
\]

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 \( M \to \infty \).

**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 \( f \in L^1(\mathbb{R}) \) but Proposition 4.1 requires \( Pf \in C^{k+2} \). 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, \( Py_j \) is not even \( C^1 \), which means the unbiasing procedure of Proposition 4.1 cannot be applied. On the other hand, by choosing \( \Psi \in C^\infty \), \( Sf \) will inherit the smoothness of the wavelet, and the wavelet invariant results will hold for any \( f \in L^1(\mathbb{R}) \) and any \( k \).
We first illustrate the unbiasing procedure of Propositions 4.1 and 4.2 for the high frequency signal for the signal $f_3(x) = e^{-5x^2} \cos(32x)$. Figures 2a and 2c show small dilations and Figures 2b and 2d show large dilations.

**Remark 4.4** There is always a signal $f$ and frequency $\xi$ for which $|P(f)(\xi) - \langle P \rangle(\xi)|$ is large regardless of $k$. Consider for example when $P(f)(\omega) = e^{-(\omega-\xi)^2}$. Then $\Omega_k(\xi) \sim \xi^k$, and $|P(f)(\xi) - \langle P \rangle(\xi)| \geq 1$. However for $M$ large enough, the order $k$ wavelet invariant estimator satisfies $|S(f)(\lambda) - \langle S \rangle(\lambda)| = O(k\Psi_{k+2}^2)$ 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 $\tilde{k}$ which 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 2 shows the $k = 0, 2, 4$ estimators of the power spectrum and wavelet invariants for both small and large dilations. Higher order unbiasing is beneficial for both methods for small dilations, but fails for the power spectrum for large dilations. Both methods will of course fail for $\eta$ large enough, but for high frequency signals the power spectrum fails much sooner.

Next we compare 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:

\[
\begin{align*}
  f_1(x) &= e^{-5x^2} \cos(8x) \\
  f_2(x) &= e^{-5x^2} \cos(16x) \\
  f_3(x) &= e^{-5x^2} \cos(32x).
\end{align*}
\]

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

Simulation results are shown in Figure 3; 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
5 Generalized MRA model

Finally, we consider the generalized MRA model (Model 2) where signals are randomly translated and dilated and corrupted by additive noise. To state Proposition 5.1 as succinctly as possible, we also define the following quantity

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

where $E$ is defined in (17) and $\Psi_m$ is defined in (11).

Figure 3: $L^2$ 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.
**Proposition 5.1** Assume Model 2 and that $\psi$ is $(k + 2)$-admissible. Define the following estimator of $(Sf)(\lambda)$:

$$
(Sf)(\lambda) := \frac{1}{M} \sum_{j=1}^{M} \left( (Sy_j)(\lambda) - B_2 \frac{\eta^2}{2} (Sy_j)''(\lambda) - \ldots - B_k \eta^k \lambda^k (Sy_j)^{(k)}(\lambda) \right) - \sigma^2
$$

where the constants $B_i$ satisfy (15). Then with probability at least $1 - 1/t^2$

$$
\left| \left( \tilde{S}f \right)(\lambda) - (Sf)(\lambda) \right| \leq k \Lambda^2 (2En)^k \frac{t}{\sqrt{M}} \left[ k\Lambda(\lambda) + \Psi \sigma^2 + \sqrt{\Psi(\Lambda_0(\lambda) + \Lambda(\lambda))} \sigma \right],
$$

where $E, \Lambda(\lambda), \Psi$ are as defined in (17), (19), (22).

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 $\Psi_i(2En)^i$ is decreasing for $i \leq k + 2$. Then with probability at least $1 - 1/t^2$

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

We remark that there are two components to the estimation error bounded by the right-hand side of (24): 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 generalized 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 $P\psi \notin C^k$, 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 generalized 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 $t_j^{M-1}$ and consider the model $y_j = f_{t_j} + \epsilon_j$. For notational convenience, we define the following order $k$ derivative “unbiasing” operator:

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

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

$$
(Sf)(\lambda) = \frac{1}{M} \sum_{j=1}^{M} \left( \frac{1}{2\pi} \int |\tilde{y}_j(\omega)|^2 A_\lambda |\tilde{\psi}_\lambda(\omega)|^2 d\omega \right) - \sigma^2
$$

$$
= \frac{1}{M} \sum_{j=1}^{M} \left( \frac{1}{2\pi} \int \left( |\tilde{f}_{t_j}(\omega)|^2 + |\tilde{\epsilon}_j(\omega)|^2 \right) A_\lambda |\tilde{\psi}_\lambda(\omega)|^2 d\omega \right) - \sigma^2
$$

We can thus decompose the error as follows:

$$
| (Sf)(\lambda) | \leq \left[ \frac{1}{M} \sum_{j=1}^{M} \left( \frac{1}{2\pi} \int (|\tilde{f}_{t_j}(\omega)|^2 + |\tilde{\epsilon}_j(\omega)|^2) A_\lambda |\tilde{\psi}_\lambda(\omega)|^2 d\omega \right) \right] + \left( \frac{1}{M} \sum_{j=1}^{M} \left( \frac{1}{2\pi} \int |\tilde{\epsilon}_j(\omega)|^2 A_\lambda |\tilde{\psi}_\lambda(\omega)|^2 d\omega \right) \right) + \sigma^2
$$

To bound the above terms we utilize the following two Lemmas, which are proved in Appendix E.
Lemma 5.1 Let the notation and assumptions of Proposition 5.1 hold, and let $A_\lambda$ be the operator defined in (25). Then with probability at least $1 - 1/t^2$

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

Lemma 5.2 Let the notation and assumptions of Proposition 5.1 hold, and let $A_\lambda$ be the operator defined in (25). Then with probability at least $1 - 1/t^2$

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

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 (23).

$\square$

5.1 Simulation results for generalized MRA model

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 $k = 0, 2, 4$; we refer to these methods as WSC $k = i$ for $i = 0, 2, 4$. We emphasize that for the generalized 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 Figure 4 for all parameter combinations resulting from $\sigma = 2^{-4}, 2^{-3}$ and $\eta = 0.06, 0.12$. The horizontal axis shows $\log_2(M)$ and the vertical axis shows $\log_2$(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. 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 $\hat{f}(\omega) \geq 0$ by taking the inverse Fourier transform of the root power spectrum. Figure 5 shows the approximate signals recovered by this procedure from PS $k = 0$ (Figure 5c) and WSC $k = 4$ (Figure 5b) for the high frequency Gabor signal $f_3(x)$ (Figure 5a). The WSC recovered signal is a much better approximation of the target signal. The recovered power spectra are shown in Figure 5d; 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 F outlines an empirical procedure for estimating the moments of $\tau$ in the special case when $t = 0$ in the generalized MRA model (i.e., no random translations). All simulations reported in Figure 4
Figure 4: $L^2$ error with standard error bars for generalized MRA model (oracle moment estimation). First, second, third column shows results for low, medium, high frequency Gabor signals. All plots have the same axis limits.
Figure 5: Signal recovery results for $f_3(x) = e^{-5x^2} \cos(32x)$ with $M = 20,000, \eta = 0.12, \sigma = 2^{-4}$.

are repeated (with minor modifications) with empirical additive and dilation moment estimation, and the results are reported in Figure 6 of Appendix F.

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

6 Numerical implementation

In this section we describe the numerical implementation of the proposed method used to generate the results reported in Sections 3.3, 4.4, and 5.1. 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 describes the convex optimization algorithm used to recover $Pf$ from $Sf$. All simulations used a Morlet wavelet constructed with $\xi_0 = 3\pi/4$.

6.1 Signal synthesis

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 $2^\ell [-\pi, \pi]$ 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.

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 $\Omega = 2^\ell [-\pi, \pi] \setminus 2^{\ell-1} [-\pi, \pi]$, we define

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

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

\[
\alpha_m(y_j) = \int_0^{2\pi} \omega^m |\hat{f}((1-\tau_j)\omega)|^2 \, d\omega
\]  

(26)

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

\[
CV_m := \frac{\text{Var}[\alpha_m(y_j)]}{\text{E}[\alpha_m(y_j)]^2}.
\]  

(27)

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. \( (CV_m)_{m=1}^{k/2-1} \) will define estimators of the first \( k \) 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 (26) and (27). Then

\[
CV_0 = \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6)
\]

\[
CV_1 = 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6).
\]

**Proof.** Since \( y_j = L_\tau f(x - t_j) \),

\[
\alpha_m(y_j) = \int_0^{2\pi} \omega^m |\hat{f}((1-\tau_j)\omega)|^2 \, d\omega
\]

\[
= \int_0^{2\pi(1-\tau_j)} \frac{\xi^m}{(1-\tau_j)^m} |\hat{f}(\xi)|^2 \, d\xi
\]

\[
= (1-\tau_j)^{-(m+1)} \alpha_m(f),
\]

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

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

When \( m = 0 \), we have

\[
CV_0 = \frac{\text{E}[(1-\tau_j)^{-2}]}{(\text{E}[(1-\tau_j)^{-1}])^2} - 1
\]

\[
= \frac{\text{E}[1 + 2\tau + 3\tau^2 + 4\tau^3 + 5\tau^4 + O(\tau^5)]}{(\text{E}[1 + \tau + \tau^2 + \tau^3 + \tau^4 + O(\tau^5)])^2} - 1
\]

\[
= \frac{1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6)}{1 + 2\eta^2 + (2C_4 + 1)\eta^4 + O(\eta^6)} - 1
\]

\[
= (1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6))(1 - 2\eta^2 + (3 - 2C_4)\eta^4 + O(\eta^6)) - 1
\]

\[
= \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6).
\]
When \( m = 1 \), we have

\[
CV_1 = \frac{\mathbb{E}[(1 - \tau_j)^{-4}]}{(\mathbb{E}[(1 - \tau_j)^{-2}])^2} - 1
\]

\[
= \frac{\mathbb{E}[1 + 4\tau + 10\tau^2 + 20\tau^3 + 35\tau^4 + O(\tau^5)]}{(\mathbb{E}[1 + 2\tau + 3\tau^2 + 4\tau^3 + 5\tau^4 + O(\tau^5)])^2} - 1
\]

\[
= 1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6)
\]

\[
(1 + 3\eta^2 + 5C_4\eta^4 + O(\eta^6))^2 - 1
\]

\[
= 1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6)
\]

\[
(1 + 6\eta^2 + (9 + 10C_4)\eta^4 + O(\eta^6)) - 1
\]

\[
= (1 + 10\eta^2 + 35C_4\eta^4 + O(\eta^6))(1 - 6\eta^2 + (27 - 10C_4)\eta^4 + O(\eta^6)) - 1
\]

\[
= 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6).
\]

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

**Definition 6.2** Assume Model 3 and let \( \widehat{CV}_0, \widehat{CV}_1 \) be the empirical versions of (27). Define the second order estimator of \( \eta^2 \) by \( \widehat{\eta}^2 = \widehat{CV}_0 \). Define the fourth order estimators of \( (\eta^2, C_4 \eta^4) \) by the unique positive solution \( (\widehat{\eta}^2, \widehat{C}_4) \) of

\[
\widehat{CV}_0 = \eta^2 + (3C_4 - 3)\eta^4
\]

\[
\widehat{CV}_1 = 4\eta^2 + (25C_4 - 33)\eta^4.
\]

For generalized 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 F. Empirical moment estimation procedures which are simultaneously robust to translations, dilations, and additive noise is an important area of future research.

### 6.4 Optimization

To approximate \( \hat{P}f \) from the wavelet invariants \( \hat{S}f \), we apply a convex optimization algorithm which returns the power spectrum which best matches the given wavelet invariants. We denote the output of the wavelet invariant plus optimization procedure by \( \hat{P}_S f \), to distinguish such estimators from those computed by directly unbiasing the power spectrum, which we continue to denote by \( \hat{P}f \). Since the wavelet invariants are only computed for \( \lambda > 0 \), we also incorporate zero frequency information into the loss function via \( \hat{P}f(0) \), 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 loss function:

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

where

\[
|\hat{\psi}_{\lambda}^+(\omega)|^2 = (|\hat{\psi}_{\lambda}(\omega)|^2 + |\hat{\psi}_{\lambda}(-\omega)|^2) \cdot 1(\omega \geq 0).
\]

Letting \( \hat{g}^* \) denote the minimizer of the above loss function, we then define \( \hat{P}_S f := \hat{g}^*(\omega)^2 \). 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 $f \in \mathbb{R}$, the symmetry of $(Pf)(\omega)$ ensures that $(Sf)(\lambda) = \langle |\hat{f}|^2, |\hat{\psi}|^2 \rangle$, and thus it is sufficient to optimize over the nonnegative 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 optimization output does depend on various numerical tolerance parameters which were held fixed for all examples.

7 Conclusion

This article considers a generalization of classic MRA which incorporates random dilation in addition to random translation 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 which 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 generalized 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 generalized MRA using a different approach. The generalized MRA model of this article corresponds to linear diffeomorphisms, and constructing unbiasing procedures which apply to more general diffeomorphisms is also an important future direction. In addition, one can construct wavelet invariants which characterize higher order auto-correlation functions such as the bispectrum, and future work will investigate full signal recovery with such invariants.

Funding

This work was supported by: the Alfred P. Sloan Foundation [Sloan Fellowship FG-2016-6607 to M.H.]; the Defense Advanced Research Projects Agency [Young Faculty Award D16AP00117 to M.H.]; and the National Science Foundation [grant 1620216 and CAREER award 1845856 to M.H.].

A Properties of wavelet invariants

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

Lemma A.1 ($k$-admissible) If $\hat{\psi} \in C^k$, $(P\psi)^{(i)}$ decays fast than $\omega^{i+1}$, and $\int \frac{|\hat{\psi}(\omega)|^2}{\omega^2} d\omega < \infty$, then $\psi$ is $k$-admissible.

Proof. We first note that $\hat{\psi} \in C^k$ guarantees $P\psi \in C^k$. Since $(P\psi)^{(i)}$ decays faster than $\omega^{i+1}$ and $P\psi \in C^k$, $\omega^i (P\psi)^{(i)}(\omega) \in L^1$ for $0 \leq i \leq k$, so $\forall k < \infty$. Also $P\psi \in C^k$ and $\omega^i (P\psi)^{(i)} \in L^1$ implies $\omega^{-2} (P\psi)^{(i)} \in L^1$ for $2 \leq i \leq k$. In addition, $\omega^{-2} (P\psi)(\omega) \in L^1$ by assumption. Thus to conclude $\Theta_k < \infty$, it only remains to show $\omega^{-1} (P\psi)(\omega) \in L^1$. Since $(P\psi)$ is continuous and decays faster than $\omega^2$, only the integrability around the
Recall from the proof of Lemma 4.3 that:

**Proof.**

for \( \Theta \)

Expanding the derivative gives:

\[ \epsilon \]

the origin since \( \epsilon > 0 \) as \( \omega \to 0 \). Thus \( (P\psi)' \sim \omega^\epsilon \) as \( \epsilon \to 0 \), so that \( \omega^{-1}(P\psi)' \sim \omega^{-1} \); the function is thus integrable around the origin since \( \epsilon - 1 > -1 \).

**Lemma 4.3** [Low Frequency Bound] Assume \( P\psi \in \mathbb{C}^m \) and \( f \in L^1 \). Then the quantity \( |\lambda^m(Sf)^{(m)}(\lambda)| \) can be bounded uniformly over all \( \lambda \). Specifically:

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

for \( \Psi_m \) defined in (11).

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

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

Utilizing Definition 2.2 we obtain

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

Expanding the derivative gives:

\[ \lambda^m \frac{d^m}{d\lambda^m} g_\lambda(\omega) = C_{m,0} g_\lambda(\omega) + C_{m,1} \omega g_\lambda'(\omega) + C_{m,2} \omega^2 g_\lambda''(\omega) + \ldots + C_{m,m} \omega^m g_\lambda^{(m)}(\omega), \]

\[ C_{m,i} = (-1)^m \binom{m}{i} \frac{m!}{i!}. \]

Utilizing \( \|\hat{f}\|_\infty \leq \|f\|_1 \) and \( g^{(i)}_\lambda(\omega) = \frac{1}{\lambda^{i+1}} g^{(i)}(\omega) \), one obtains:

\[ |\lambda^m(Sf)^{(m)}(\lambda)| \leq \sum_{i=0}^{m} \frac{|C_{m,i}|}{2\pi} \int |\hat{f}(\omega)||^2 |\omega^i g^{(i)}_\lambda(\omega)| d\omega \]

\[ \leq \|f\|_1^2 \sum_{i=0}^{m} \frac{|C_{m,i}|}{2\pi} \int |\omega^i g^{(i)}_\lambda(\omega)| d\omega \]

\[ = \|f\|_1^2 \sum_{i=0}^{m} \frac{|C_{m,i}|}{2\pi} \|\omega^i g^{(i)}_\lambda(\omega)\|_1 \]

\[ = \Psi_m \|f\|^2_1. \]

**Lemma 4.4** [High Frequency Bound for Differentiable Functions] Assume \( P\psi \in \mathbb{C}^m \), and \( f' \in L^1 \). Then the quantity \( |\lambda^m(Sf)^{(m)}(\lambda)| \) can be bounded by:

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

for \( \Theta_m \) defined in (12).

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

\[ |\lambda^m(Sf)^{(m)}(\lambda)| \leq \sum_{i=0}^{m} \frac{|C_{m,i}|}{2\pi} \int |\hat{f}(\omega)||^2 |\omega^i g^{(i)}_\lambda(\omega)| d\omega \]

24
where \( g_A(\omega) = \frac{1}{\lambda} g\left(\frac{\omega}{\lambda}\right) = |\hat{\psi}_A(\omega)|^2 \) and \( C_{m,i} = (-1)^m \frac{(m)}{i!} \cdot \frac{m!}{\pi} \). Since \( \|\omega \hat{f}(\omega)\|_\infty \leq \|f'\|_1 \) and \( g_A^{(i)}(\omega) = \frac{1}{\lambda^{i+1}} g_A^{(i)}\left(\frac{\omega}{\lambda}\right) \), we obtain:

\[
|\lambda^m(Sf)^{(m)}(\lambda)| \leq \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega \hat{f}(\omega)|^2 |\omega^{i-2} g_A^{(i)}(\omega)| \, d\omega
\]

\[
\leq \|f'\|_1^2 \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \int |\omega^{i-2} g_A^{(i)}(\omega)| \, d\omega
\]

\[
= \frac{\|f'\|_1^2}{\lambda^2} \sum_{i=0}^m \frac{|C_{m,i}|}{2\pi} \cdot \|\omega^{i-2} g_A^{(i)}(\omega)\|_1
\]

\[
= \Theta_m \frac{\|f'\|_1^2}{\lambda^2}.
\]

**Lemma A.2** Assume \( Pf \in C^0 \) and \( \psi \) is \( m \)-admissible, and let \( B_m, E, \Psi_m, \Theta_m \) be as defined in (15), (17), (11) (12). Then:

\[
\frac{1}{2\pi} \int |\hat{f}(\omega)|^2 \cdot |B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\hat{\psi}_A(\omega)|^2 \, d\omega \leq (\eta)^m \Lambda_m(\lambda),
\]

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:

\[
\frac{1}{2\pi} \int |\hat{f}(\omega)|^2 \cdot \left| \lambda^m \frac{d^m}{d\lambda^m} |\hat{\psi}_A(\omega)|^2 \right| \, d\omega \leq \Psi_m \|f\|_1^2.
\]

From the proof of Lemma 4.4:

\[
\frac{1}{2\pi} \int |\hat{f}(\omega)|^2 \cdot \left| \lambda^m \frac{d^m}{d\lambda^m} |\hat{\psi}_A(\omega)|^2 \right| \, d\omega \leq \Theta_m \frac{\|f'\|_1^2}{\lambda^2}.
\]

Utilizing \( |B_m| \leq E^m \) gives

\[
\frac{1}{2\pi} \int |\hat{f}(\omega)|^2 \cdot \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\hat{\psi}_A(\omega)|^2 \right| \, d\omega \leq (\eta)^m \left( \|f\|_1^2 \Psi_m \wedge \frac{\|f'\|_1^2 \Theta_m}{\lambda^2} \right).
\]

The following Corollary is obtained from Lemma A.2 when \( f \) is a dirac-delta function.

**Corollary A.1** Assume \( \psi \) is \( m \)-admissible, and let \( B_m, E, \Psi_m \) be as defined in (15), (17), (11). Then:

\[
\frac{1}{2\pi} \int |B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\hat{\psi}_A(\omega)|^2 \, d\omega \leq (\eta)^m \Psi_m.
\]
B PS 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 \( L^2 \) function, as long as the wavelet satisfies the linear independence Condition 2.3. Proposition 2.5 gives two criteria which are sufficient to guarantee Condition 2.3. Finally, Lemma B.1 establishes that the Morlet wavelet satisfies Condition 2.3.

**Lemma 2.1** Let \( p \in L^2 \) and assume \( p(\omega) = p(-\omega) \) and Condition 2.3. Then

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

**Proof.**

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

\[
\int p(\omega)|\hat{\psi}_\lambda(\omega)|^2 \, d\omega = \int_0^\infty p(\omega)|\hat{\psi}_\lambda^+(\omega)|^2 \, d\omega = 0 \quad \forall \lambda.
\]

We will show

\[
\int_0^\infty |\langle p, |\hat{\psi}_\lambda^+|^2 \rangle_{\mathbb{R}^1}|^2 \, d\lambda \geq A\|p\|^2_{L^2(\mathbb{R}^1)}
\]

for some positive constant \( A > 0 \). Note:

\[
\int_0^\infty |\langle p, |\hat{\psi}_\lambda^+|^2 \rangle_{\mathbb{R}^1}|^2 \, d\lambda = \int_0^\infty \langle p, |\hat{\psi}_\lambda^+|^2 \rangle_{\mathbb{R}^1} \langle \overline{p}, |\hat{\psi}_\lambda^+|^2 \rangle_{\mathbb{R}^1} \, d\lambda
\]

\[
= \int_0^\infty \left( \int_0^\infty p(\omega_1)|\hat{\psi}_\lambda^+(\omega_1)|^2 \, d\omega_1 \right) \left( \int_0^\infty \overline{p(\omega_2)}|\hat{\psi}_\lambda^+(\omega_2)|^2 \, d\omega_2 \right) \, d\lambda
\]

\[
= \int_0^\infty p(\omega_2) \left( \int_0^\infty p(\omega_1) \left( \int_0^\infty |\hat{\psi}_\lambda^+(\omega_1)|^2 |\hat{\psi}_\lambda^+(\omega_2)|^2 \, d\lambda \right) \, d\omega_1 \right) \, d\omega_2.
\]

Now consider the kernel

\[
k(\omega_1, \omega_2) = \int_0^\infty |\hat{\psi}_\lambda^+(\omega_1)|^2 |\hat{\psi}_\lambda^+(\omega_2)|^2 \, d\lambda
\]

and the corresponding integral operator defined on \( \mathbb{R}^+ \):

\[
Kp(\omega_2) = \int_0^\infty p(\omega_1)k(\omega_1, \omega_2) \, d\omega_1.
\]

Note that

\[
\int_0^\infty |k(\omega_1, \omega_2)| \, d\omega_1 = \int_0^\infty \int_0^\infty |\hat{\psi}_\lambda^+(\omega_1)|^2 |\hat{\psi}_\lambda^+(\omega_2)|^2 \, d\lambda \, d\omega_1
\]

\[
= \int_0^\infty |\hat{\psi}_\lambda^+(\omega_2)|^2 \left( \int_0^\infty |\hat{\psi}_\lambda^+(\omega_1)|^2 \, d\omega_1 \right) \, d\lambda
\]

\[
\leq \|\hat{\psi}_\lambda^+\|_2^2 \int_0^\infty |\hat{\psi}_\lambda^+(\omega_2)|^2 \, d\lambda
\]

\[
= \|\hat{\psi}\|_2^2 \int_0^\infty |\hat{\psi}(\xi)|^2 \, d\xi
\]

\[
= C_\psi
\]

\[26\]
where we have assumed that $\|\hat{\psi}\|_2^2 = 1$, and an identical argument can be applied to show $\int_0^\infty |k(\omega_1, \omega_2)|\ d\omega_2 \leq C_{\psi}$. Thus by Schur’s lemma, $\|K\| \leq C_{\psi}$, and so $\|Kp\|_{L^2(\mathbb{R}^+)^2} \leq C_{\psi}\|p\|_{L^2(\mathbb{R}^+)}$ and $K : L^2(\mathbb{R}^+) \to L^2(\mathbb{R}^+)$. Thus

$$\int_0^\infty |\langle p, |\hat{\psi}(\lambda)|^2 \rangle_{\mathbb{R}^+}^2| d\lambda = \int_0^\infty \frac{p(\omega_2)}{p(\omega_1)} \left( \int_0^\infty p(\omega_1)k(\omega_1, \omega_2) d\omega_1 \right) d\omega_2$$

$$= \int_0^\infty \frac{p(\omega_2)}{p(\omega_1)}Kp(\omega_2) d\omega_2$$

$$= \langle Kp, p \rangle_{\mathbb{R}^+}.$$ 

We now show that $K$ is a positive definite operator. Note $K$ is positive definite if for any finite sequence $\{\omega_i\}_{i=1}^n$ with all $\omega_i > 0$, the $n$ by $n$ matrix $A$ defined by

$$A_{ij} = k(\omega_i, \omega_j)$$

is positive definite [50]. Viewing $\tilde{\omega}_i(\lambda) = |\hat{\psi}(\omega_i)|^2$ as functions of $\lambda$, we see that

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

and $A$ is thus a Gram matrix. Since the $\tilde{\omega}_i(\lambda)$ are linearly independent by assumption, we can conclude that $A$ and thus $K$ is positive definite. Letting $\lambda_{\min} > 0$ denote the smallest eigenvalue of $K$, we then have $\langle Kp, p \rangle_{\mathbb{R}^+} \geq \lambda_{\min} \|p\|_{L^2(\mathbb{R}^+)}^2$. Thus $\langle \hat{\psi}(\lambda)|^2 \rangle_{\mathbb{R}^+} = 0$ for all $\lambda$ implies $0 \geq \lambda_{\min} \|p\|_{L^2(\mathbb{R}^+)}^2$. We conclude that $\|p\|_{L^2(\mathbb{R}^+)} = 0$ and thus $p = 0$ for almost every $\omega > 0$. Since $p(\omega) = p(\omega)$, $p = 0$ for almost every $\omega \in \mathbb{R}$.

\[ \square \]

**Proposition 2.5** The following are sufficient to guarantee Condition 2.3:

1. $|\hat{\psi}(\omega)|^2$ has a compact support contained in the interval $[a, \alpha]$, where $a$ and $\alpha$ have the same sign, e.g., complex analytic wavelets with compactly supported Fourier transform.

2. $|\hat{\psi}(\omega)|^2 \in C^\infty$ and there exists an $N$ such that all derivatives of order at least $N$ are nonzero 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 $\tilde{\omega}_i(\lambda) = \frac{1}{|\lambda|} |\hat{\psi}(\frac{\omega_i}{\lambda})|^2$ denote the corresponding functions of $\lambda$.

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

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

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

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

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

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

Assume

\[ c_1 g(\omega_1\lambda) + c_2 g(\omega_2\lambda) + \cdots + c_n g(\omega_n\lambda) = 0. \]

Differentiating \( m \) times for \( N \leq m \leq N + n - 1 \), we obtain:

\[ c_1 \omega_1^N g^{(N)}(\omega_1\lambda) + \cdots + c_n \omega_n^N g^{(N)}(\omega_n\lambda) = 0 \]

\[ : \]

\[ c_1 \omega_1^{N+n-1} g^{(N+n-1)}(\omega_1\lambda) + \cdots + c_n \omega_n^{N+n-1} g^{(N+n-1)}(\omega_n\lambda) = 0 \]

The above holds for all \( \lambda \). We now take the limit as \( \lambda \to 0 \) to obtain:

\[ g^{(N)}(0)(\omega_1^N c_1 + \omega_2^N c_2 + \cdots + \omega_n^N c_n) = 0 \]

\[ g^{(N+1)}(0)(\omega_1^{N+1} c_1 + \omega_2^{N+1} c_2 + \cdots + \omega_n^{N+1} c_n) = 0 \]

\[ : \]

\[ g^{(N+n-1)}(0)(\omega_1^{N+n-1} c_1 + \omega_2^{N+n-1} c_2 + \cdots + \omega_n^{N+n-1} c_n) = 0 \]

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

\[
\begin{bmatrix}
\omega_1^N & \cdots & \omega_n^N \\
\omega_1^{N+1} & \cdots & \omega_n^{N+1} \\
\vdots & \ddots & \vdots \\
\omega_1^{N+n-1} & \cdots & \omega_n^{N+n-1}
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
c_n
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & \cdots & 1 \\
\omega_1 & \cdots & \omega_n \\
\vdots & \ddots & \vdots \\
\omega_1^{(N-1)} & \cdots & \omega_n^{(N-1)}
\end{bmatrix}
\begin{bmatrix}
\omega_1^N & 0 & \cdots & 0 \\
0 & \omega_2^N & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \omega_n^N
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
c_n
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

Since \( A \) is a Vandermonde matrix constructed from distinct \( \omega_i \), \( \det(A) \neq 0 \). Since the \( \omega_i \) are nonzero, \( \det(B) \neq 0 \). 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 B.1** Suppose we construct a Morlet wavelet with parameter \( \xi \), that is \( \psi(x) = C_\xi \pi^{-1/4} e^{-x^2/2} (e^{i\xi^2} - e^{-\xi^2}) \) 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

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

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

\[ g(\omega) := \tilde{C}_\xi^{-2} |\hat{\psi}(\omega)|^2 = e^{-\omega^2}(e^{i\omega} - 1)^2. \]
From direct calculation or a computer algebra system (CAS), one obtains:

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

where \(H_n(\xi)\) is the \(n\)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 2.5, and thus the linear independence Condition 2.3. □

C Supporting results: classic MRA

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

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

\[
\begin{align*}
\mathbb{E} \left[ |\hat{\varepsilon}(\omega)|^2 \right] &= \sigma^2 \quad (28) \\
\mathbb{E} \left[ |\hat{\varepsilon}(\omega)|^4 \right] &\leq 3\sigma^4 \quad (29) \\
\mathbb{E} \left[ |\hat{\varepsilon}(\omega)|^2 |\hat{\varepsilon}(\xi)|^2 \right] &\leq 3\sigma^4. \quad (30)
\end{align*}
\]

**Proof.** By Proposition H.1,

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

which shows (28). By Proposition H.2,

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

29
which shows (29). Finally, by Proposition H.3, we have

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

which gives (30).

\[\square\]

**Lemma 3.1** Let \(\varepsilon(x)\) be a white noise process on \([-1/2, 1/2]\) with variance \(\sigma^2\). Then for any signal \(f \in L^1(\mathbb{R})\):

\[
\mathbb{E} \left[ (P(f + \varepsilon))(\omega) \right] = (Pf)(\omega) + \sigma^2 \\
\text{Var} \left[ (P(f + \varepsilon))(\omega) \right] \leq 4\sigma^2 (Pf)(\omega) + 2\sigma^4.
\]

**Proof.** Since \(\mathbb{E} [\hat{\varepsilon}(\omega)] = \mathbb{E} [\hat{\xi}(\xi)] = 0\) and \(\mathbb{E} [|\hat{\varepsilon}(\omega)|^2] = \sigma^2\) by Lemma C.1,

\[
\mathbb{E} \left[ (P(f + \varepsilon))(\omega) \right] = \mathbb{E} \left[ \hat{f}(\omega) + \hat{\varepsilon}(\omega) \left( \hat{f}(\omega) + \hat{\varepsilon}(\omega) \right) \right] \\
= \mathbb{E} \left[ |\hat{\varepsilon}(\omega)|^2 + \hat{f}(\omega) \hat{\varepsilon}(\omega) + \hat{\varepsilon}(\omega) \hat{f}(\omega) + |\hat{\varepsilon}(\omega)|^2 \right] \\
= (Pf)(\omega) + \sigma^2.
\]

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

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

and that

\[
\mathbb{E} \left[ |\hat{\varepsilon}(\omega)|^2 \hat{\varepsilon}(\omega) \right] = \mathbb{E} \left[ \left( \int_{-1/2}^{1/2} e^{-i\omega x} \, dB_x \right) \left( \int_{-1/2}^{1/2} e^{i\omega x} \, dB_x \right) \left( \int_{-1/2}^{1/2} e^{-i\omega p} \, dB_p \right) \right] \\
= 0,
\]

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

\[
\mathbb{E}[(P(f + \varepsilon))(\omega)^2] = \mathbb{E} \left[ |\hat{f}(\omega)|^4 + 4|\hat{f}(\omega)|^2 |\hat{\varepsilon}(\omega)|^2 + |\hat{\varepsilon}(\omega)|^4 + \hat{f}(\omega)^2 \hat{\varepsilon}(\omega)^2 + \hat{\varepsilon}(\omega)^2 \hat{f}(\omega)^2 \right] \\
\leq \mathbb{E} \left[ |\hat{f}(\omega)|^4 + 6|\hat{f}(\omega)|^2 |\hat{\varepsilon}(\omega)|^2 + |\hat{\varepsilon}(\omega)|^4 \right] \\
= [(Pf)(\omega)]^2 + 6\sigma^2 (Pf)(\omega) + 3\sigma^4
\]

where the last line follows from Lemma C.1. Thus

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

\[\square\]
LEMMA C.2 Let $\epsilon(x)$ be a white noise processes on $[-\frac{1}{2}, \frac{1}{2}]$ with variance $\sigma^2$. Then:
\[
E[(S\epsilon)(\lambda)] = \sigma^2
\]
\[
E[(S\epsilon)(\lambda)^2] \leq 3\sigma^4.
\]

Proof. Since $E[|\hat{\epsilon}(\omega)|^2] = \sigma^2$ by Lemma C.1, we have:
\[
E[(S\epsilon)(\lambda)] = E\left[ \int |\epsilon \ast \psi_\lambda|^2 \, d\omega \right] = \int \left( \frac{1}{2\pi} \right) \hat{\epsilon}(\lambda) \hat{\psi}_\lambda(\omega) d\omega = \sigma^2 \int |\hat{\psi}_\lambda(\omega)|^2 d\omega = \sigma^2 \int \hat{\psi}_\lambda(\omega)^2 d\omega
\]

Since by Lemma C.1, $E[|\hat{\epsilon}(\omega)|^2|\hat{\epsilon}(\xi)|^2] \leq 3\sigma^4$, we also have:
\[
E[(S\epsilon)(\lambda)^2] = E\left[ \int |\epsilon \ast \psi_\lambda|^2 \, d\omega \right] = \int \left( \frac{1}{2\pi} \right) \hat{\epsilon}(\lambda) \hat{\psi}_\lambda(\omega)^2 d\omega = \int \left( \frac{1}{2\pi} \right) \hat{\epsilon}(\lambda) \hat{\psi}_\lambda(\omega)^2 d\omega
\]

\[
\leq 3\sigma^4 \int \hat{\psi}_\lambda(\omega)^2 d\omega = 3\sigma^4
\]

\[
\square
\]

LEMMA 3.2 Let $\epsilon(x)$ be a white noise processes on $[-\frac{1}{2}, \frac{1}{2}]$ with variance $\sigma^2$. Then for any signal $f \in L^1(\mathbb{R})$:
\[
E[(S(f + \epsilon))(\lambda)] = (Sf)(\lambda) + \sigma^2
\]
\[
\text{Var}[(S(f + \epsilon))(\lambda)] \leq 4\sigma^2(Sf)(\lambda) + 2\sigma^4.
\]

Proof. Utilizing $E[\epsilon] = E[\hat{\epsilon}] = 0$ and Lemma C.2, we have:
\[
E[(S(f + \epsilon))(\lambda)] = E\left[ \int |(f + \epsilon) \ast \psi_\lambda(u)|^2 \, du \right] = \int |f \ast \psi_\lambda(u)|^2 + E\left[ \int |\epsilon \ast \psi_\lambda(u)|^2 \, du \right] = (Sf)(\lambda) + E[(S\epsilon)(\lambda)] = (Sf)(\lambda) + \sigma^2.
\]

To bound $E[(S(f + \epsilon))(\lambda)^2]$, note that:
\[
(S(f + \epsilon))(\lambda)^2 = \left( \int |f \ast \psi_\lambda(u_1)|^2 + (\epsilon \ast \psi_\lambda(u_1))(\overline{\hat{f}} \ast \hat{\psi}_\lambda(u_1)) + (f \ast \psi_\lambda(u_1))(\overline{\hat{\epsilon}} \ast \hat{\psi}_\lambda(u_1)) + |\epsilon \ast \psi_\lambda(u_1)|^2 \, du_1 \right)
\]
\[
\cdot \left( \int |f \ast \psi_\lambda(u_2)|^2 + (\epsilon \ast \psi_\lambda(u_2))(\overline{\hat{f}} \ast \hat{\psi}_\lambda(u_2)) + (f \ast \psi_\lambda(u_2))(\overline{\hat{\epsilon}} \ast \hat{\psi}_\lambda(u_2)) + |\epsilon \ast \psi_\lambda(u_2)|^2 \, du_2 \right)
\]

31
When we take expectation, any term involving one or three \( \epsilon \) terms disappear, so that:

\[
\mathbb{E}[(S(f + \epsilon))(\lambda)^2] = \mathbb{E} \left[ \int \int |f * \psi_{\lambda}(u_1)|^2 |f * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int |f * \psi_{\lambda}(u_1)|^2 |\epsilon * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int (|\epsilon * \psi_{\lambda}(u_1)|^2 f * \psi_{\lambda}(u_2))^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int (\epsilon * \psi_{\lambda}(u_1))^2 (f * \psi_{\lambda}(u_2))^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int (f * \psi_{\lambda}(u_1))^2 (\epsilon * \psi_{\lambda}(u_2))^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int (\epsilon * \psi_{\lambda}(u_1))^2 (f * \psi_{\lambda}(u_2))^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int |\epsilon * \psi_{\lambda}(u_1)|^2 |f * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int |\epsilon * \psi_{\lambda}(u_1)|^2 |\epsilon * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 \right]
\]

\[
\leq \mathbb{E} \left[ \int \int |f * \psi_{\lambda}(u_1)|^2 |f * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 
\right. \\
+ \left. 6 \int \int |f * \psi_{\lambda}(u_1)|^2 |\epsilon * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 
\right. \\
+ \left. \int \int |\epsilon * \psi_{\lambda}(u_1)|^2 |\epsilon * \psi_{\lambda}(u_2)|^2 \ du_1 \ du_2 \right]
\]

\[
= \mathbb{E} \left[ [(Sf)(\lambda)]^2 + 6(Sf)(\lambda)(Se)(\lambda) + [(Se)(\lambda)]^2 \right]
\]

\[
= [(Sf)(\lambda)]^2 + 6\sigma^2(Sf)(\lambda) + 3\sigma^4.
\]

where the last line follows from Lemma C.2. Thus

\[
\text{Var}[(S(f + \epsilon))(\lambda)] = \mathbb{E}[(S(f + \epsilon))(\lambda)^2] - \left[ \mathbb{E}[(S(f + \epsilon))(\lambda)] \right]^2 
\]

\[
\leq [(Sf)(\lambda)]^2 + 6\sigma^2(Sf)(\lambda) + 3\sigma^4 - [(Sf)(\lambda) + \sigma^2]^2
\]

\[
= 4\sigma^2(Sf)(\lambda) + 2\sigma^4.
\]

\[
\square
\]

\section{D Supporting results: dilation MRA}

This appendix contains the technical details of the dilation unbiasing procedure which 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.

\textbf{Lemma 4.1} Let \( F_{\lambda}(\tau) = L((1 - \tau)\lambda) \) for some function \( L \in C^{k+2}(0, \infty) \) and a random variable \( \tau \) satisfying the assumptions of Section 2.1, and let \( k \geq 2 \) be an even integer. Assume

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

for some absolute constant \( R \) and define the following estimator of \( L(\lambda) \):

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

32
Then $G_\lambda(\tau)$ satisfies

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

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

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 an absolute constant defined in (17).

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

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

We note:

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

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

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

$$F''_\lambda(\tau) = F''_\lambda(0) + F^{(3)}_\lambda(0)\tau + \ldots + \frac{F^{(k+1)}_\lambda(0)}{(k-1)!} \tau^{k-1} + \int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k-1)!} (\tau - t)^{k-1} dt. \quad := R_2(\tau, \lambda)$$

$$F^{(4)}(\tau) = F^{(4)}_\lambda(0) + F^{(5)}_\lambda(0)\tau + \ldots + \frac{F^{(k+1)}_\lambda(0)}{(k-3)!} \tau^{k-3} + \int_0^\tau \frac{F^{(k+2)}_\lambda(t)}{(k-3)!} (\tau - t)^{k-3} dt. \quad := R_4(\tau, \lambda)$$

$$\vdots$$

$$F^{(k)}(\tau) = F^{(k)}_\lambda(0) + F^{(k+1)}_\lambda(0)\tau + \int_0^\tau \frac{F^{(k+2)}_\lambda(t)(\tau - t)}{(k)!} dt. \quad := R_k(\tau, \lambda)$$
Multiplication of the $i^{th}$ even derivative by $B_i \eta^i$ gives:

$$F_\lambda(\tau) = F_\lambda(0) + F''_\lambda(0) \tau + \ldots + \frac{F^{(k+1)}(0)}{(k+1)!} \tau^{k+1} + R_0(\tau, \lambda)$$

$$F''_\lambda(\tau) = \int F''_\lambda(0) + \frac{B_2 \eta^2 F''_\lambda(0) + B_2 \eta^2 F^{(3)}(0) + \ldots + B_2 \eta^2 \frac{F^{(k+1)}(0)}{(k-1)!}}{2} \tau^{k-1} + B_2 \eta^2 R_2(\tau, \lambda)$$

$$F^{(4)}_\lambda(\tau) = \int B_4 \eta^4 F^{(4)}_\lambda(0) + B_4 \eta^4 F^{(5)}(0) + \ldots + B_4 \eta^4 \frac{F^{(k+1)}(0)}{(k-3)!} \tau^{k-3} + B_4 \eta^4 R_4(\tau, \lambda)$$

$$\vdots$$

$$F^{(k)}_\lambda(\tau) = B_k \eta^k F^{(k)}(0) + B_k \eta^k F^{(k-1)}(0) \tau + B_k \eta^k R_k(\tau, \lambda).$$

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

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

and show that $E[G_\lambda(\tau)] = F_\lambda(0) + O(\eta^{k+2})$ for constants $B_i$ chosen according to (15). We have:

$$E[F_\lambda(\tau)] = F_\lambda(0) + F''_\lambda(0) \frac{C_2}{2} \eta^2 + \ldots + \frac{F^{(4)}(0)}{4!} \eta^4 + \ldots + \frac{F^{(6)}(0)}{6!} \eta^6 + \ldots + \frac{F^{(k)}(0)}{k!} \eta^k + E[R_0(\tau, \lambda)]$$

$$E[B_2 \eta^2 F''(\tau)] = \int B_2 \eta^2 F^{(2)}(0) + B_2 \eta^2 F^{(3)}(0) + \ldots + B_2 \eta^2 \frac{F^{(k-2)}(0)}{(k-2)!} \tau^{(k-2)} + E\left[B_2 \eta^2 R_2(\tau, \lambda)\right]$$

$$E[B_4 \eta^4 F^{(4)}(\tau)] = \int B_4 \eta^4 F^{(4)}(0) + B_4 \eta^4 F^{(5)}(0) + \ldots + B_4 \eta^4 \frac{F^{(k)}(0)}{(k-4)!} \eta^k + E\left[B_4 \eta^4 R_4(\tau, \lambda)\right]$$

$$\vdots$$

$$E[B_k \eta^k F^{(k)}(\tau)] = \int B_k \eta^k F^{(k)}(0) B_k \eta^k + E\left[B_k \eta^k R_k(\tau, \lambda)\right]$$

That is:

$$E[G_\lambda(\tau)] = F_\lambda(0) + F''_\lambda(0) \left(\frac{C_2}{2!} - B_2\right) \eta^2 + \frac{F^{(4)}(0)}{4!} \left(\frac{C_4}{4!} - \frac{B_2 C_2}{2!} - B_4\right) \eta^4$$

$$+ \frac{F^{(6)}(0)}{6!} \left(\frac{C_6}{6!} - \frac{B_2 C_4}{4!} - \frac{B_4 C_2}{2!} - B_6\right) \eta^6$$

$$\ldots + \frac{F^{(k)}(0)}{k!} \left(\frac{C_k}{k!} - \frac{B_2 C_{k-2}}{2!(k-2)!} - \ldots - \frac{B_{k-2} C_2}{2!} - B_k\right) \eta^k + H_1(\lambda)$$

where

$$H_1(\lambda) = E\left[R_0(\lambda, \tau) - B_2 \eta^2 R_2(\tau, \lambda) - \ldots - B_k \eta^k R_k(\lambda, \tau)\right].$$

Since (15) guarantees that

$$B_2 = \frac{C_2}{2!}$$

$$B_4 = \frac{C_4}{4!} - \left(\frac{C_2}{2!}\right)^2$$

$$B_6 = \frac{C_6}{6!} - \frac{C_2 C_4}{2!4!} - \left(\frac{C_4}{4!} - \left(\frac{C_2}{2!}\right)^2\right) \frac{C_2}{2!}$$

$$\vdots$$

$$B_k = \frac{C_k}{k!} - \frac{B_2 C_{k-2}}{2!(k-2)!} - \ldots - \frac{B_{k-2} C_2}{2!}.$$
the coefficients of $\eta^2, \eta^4, \ldots, \eta^k$ vanish, and we obtain:

$$E[G_A(\tau)] = F_A(0) + H_1(\lambda).$$

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

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

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

$$(k + 1 - i)! \eta^i R_i(\lambda, \tau) = \eta^i \int_0^\tau P^{(k+2)}(t)(\tau - t)^{k+1-i} \, dt$$

$$= \eta^i \int_0^\tau \lambda^{k+2} L^{(k+2)}((1-t)\lambda)(\tau - t)^{k+1-i} \, dt.$$  

We observe that:

$$\left| \left( (1-t)\lambda \right)^{k+2} L^{(k+2)}((1-t)\lambda) \right| \leq \Lambda_{k+2}((1-t)\lambda)$$

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

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

so that

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

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

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

$$\leq \eta^i R\Lambda_{k+2}(\lambda) \int_0^\tau \frac{t^{k+1-i}}{(1-t)^{k+2}} \, dt$$

$$= \eta^i \tau^{k+1-i} \Lambda_{k+2}(\lambda) \frac{1}{(k+1)} \left( \frac{1}{(1-\tau)^{k+1}} - 1 \right)$$

$$\leq \frac{2^{k+2} R}{k+1} \eta^i \tau^{k+2-i} \Lambda_{k+2}(\lambda)$$

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

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

(31)

which gives

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

$$= \frac{2^{k+2} (k+2-i) R}{k+1} \Lambda_{k+2}(\lambda) T^{k+2-i} |B_i| \eta^{k+2}.$$  

35
We thus obtain
\begin{align*}
|E[G_\lambda(\tau) - L(\lambda)| = |H_1(\lambda)| & \leq \frac{R\Lambda k+2(\lambda)}{k+1} (2En)^{k+2} \sum_{i=0,2,\ldots,k} (k+2-i) \\
& \lesssim Rk\Lambda k+2(\lambda)(2En)^{k+2},
\end{align*}
which establishes the bound on the bias. We now bound the variance. We note:
\begin{align*}
G_\lambda(\tau) = \sum_{i=0,2,\ldots,k} \sum_{j=0,1,\ldots,k+1-i} \frac{-B_i}{j!} F_\lambda^{(j+1)}(0)\eta^j \tau^j + \sum_{i=0,2,\ldots,k} -B_i R_i(\lambda,\tau)\eta^i.
\end{align*}
Thus:
\begin{align*}
Var[G_\lambda(\tau)] &= E[G_\lambda(\tau)^2] - E[G_\lambda(\tau)]^2 \\
&= E[(I(I) + 2E[(I(II)] + E[(II)(II)]) - F_\lambda(0)^2 - 2F_\lambda(0)H_1(\lambda) - H_1(\lambda)^2 \\
& \leq (E[(I(I)]) - F_\lambda(0)^2) + (2E[(I(II))] - 2F_\lambda(0)H_1(\lambda) + E[(II)(II)])
\end{align*}
and we proceed to bound each term.
\begin{align*}
(I(I) - F_\lambda(0))^2 &= \sum_{i=0,2,\ldots,k} \sum_{j=0,1,\ldots,k+1-i} \sum_{\ell=0,2,\ldots,k} \sum_{s=0}^{k+1-i-k+1-\ell} \frac{B_i B_\ell}{j! \ell!} F_\lambda^{(j+1)}(0)F_\lambda^{(\ell+s)}(0)\eta^j \tau^j \eta^\ell \tau^s 1_E
\end{align*}
where $1_E$ is an indicator function indicating that $i, j, \ell, s$ are not all zero. We have
\begin{align*}
E \left| \frac{B_i B_\ell}{j! \ell!} F_\lambda^{(j+1)}(0)F_\lambda^{(\ell+s)}(0)\eta^j \tau^j \eta^\ell \tau^s \right| & \leq \left| \frac{B_i B_\ell}{j! \ell!} \right| C_{j+s}\Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)\eta^j \tau^j \eta^\ell \tau^s \\
& \leq \left| \frac{B_i B_\ell}{j! \ell!} \right| T^j T^\ell \Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)\eta^j \tau^j \eta^\ell \tau^s \\
& \leq E^j E_{\ell+s} \Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)\eta^j \tau^j \eta^\ell \tau^s \\
& = \left(\Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)\eta^j \tau^j \eta^\ell \tau^s \right).
\end{align*}
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{align*}
E [(I(I)] - F_\lambda(0)^2 & \leq \sum_{i=0,2,\ldots,k} \sum_{j=0,1,\ldots,k+1-i} \sum_{\ell=0,2,\ldots,k} \sum_{s=0}^{k+1-i-k+1-\ell} \Lambda_i j + j(\lambda)(4T\eta)^j \Lambda_{\ell+s}(\lambda)(4T\eta)^{\ell+s} 1_E(j + s even) \\
& = \sum_{i=0,2,\ldots,k} \sum_{\ell=0,2,\ldots,k} \Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)(4T\eta)^j \Lambda_{\ell+s}(\lambda)(4T\eta)^\ell \\
& \leq k^2 \Lambda(\lambda)^2.
\end{align*}
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:
\begin{align*}
E [(I(I)] - F_\lambda(0)^2 & \leq k^2 \sum_{2\leq j + \ell \leq 2k+2} \sum_{I + \ell \text{ even}} \Lambda_i j + j(\lambda)\Lambda_{\ell+s}(\lambda)(4T\eta)^j \Lambda_{\ell+s}(\lambda)(4T\eta)^\ell \\
& \leq k^2 \Lambda(\lambda)^2.
\end{align*}
Next we bound $E[(II)(II)]$.
\begin{align*}
(II)(II) = \sum_{i=0,2,\ldots,k} \sum_{\ell=0,2,\ldots,k} B_i B_\ell R_i(\lambda,\tau)R_\ell(\lambda,\tau)\eta^{i+\ell}
\end{align*}
Utilizing Equation (31), we have:

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

which gives

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

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

so that

$$\mathbb{E} [\mathbb{E} (\mathbb{E})] \leq \frac{R^2}{(k+1)^2} \Lambda_{k+2}(\lambda)^2 (2\eta)^{2k+4} \sum_{i=0,2,\ldots,k} \sum_{\ell=0,2,\ldots,k} (k+1-i)(k+2-\ell)$$

$$\leq k^2 R^2 \Lambda_{k+2}(\lambda)^2 (2\eta)^{2k+4} \leq k^2 R^2 \Lambda(\lambda)^2.$$

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

$$\begin{align*}
(\mathbb{E} (\mathbb{E}) (\mathbb{E})) &= \sum_{i=0,2,\ldots,k} \sum_{j=0,2,\ldots,k} \sum_{\ell=0,2,\ldots,k} \frac{B_i}{j!} F^{(i+j)}(\lambda)(0) \eta^i \tau^j B_\ell R_\ell(\lambda, \tau) \eta^{\ell} \\
(\mathbb{E} (\mathbb{E}) (\mathbb{E})) &= \sum_{i=0,2,\ldots,k} \sum_{j=0,2,\ldots,k} \sum_{\ell=0,2,\ldots,k} \frac{B_i}{j!} F^{(i+j)}(\lambda)(0) \eta^i \tau^j B_\ell R_\ell(\lambda, \tau) \eta^{\ell} \\
&\leq \frac{2^{k+2} R^2 T^{k+2-j-\ell} |B_i B_\ell|}{(k+1) j!(k+1-\ell)!} \Lambda_{i+j}(\lambda) \Lambda_{k+2}(\lambda)^2 \eta^{i+j} \eta^{k+2-j+\ell}
\end{align*}$$

so that

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

37
The same bound holds for the terms of $F_\lambda(0)H_1(\lambda)$, which arise from $i = 0, j = 0$ in (32), so that

$$2E[(\text{I}) (\text{II})] - 2F_\lambda(0)H_1(\lambda) \lesssim \sum_{i=0}^{k} \frac{R(k + 2 - \ell)}{(k + 1)} \left( \sum_{j=0}^{k} (E\eta)^{i+j} \Lambda_i \Lambda_j(\lambda) \right)$$

$$\lesssim \left( \sum_{i=0}^{k+1} \Lambda_i(\lambda)(E\eta)^{i} \right) \left( kR(2E\eta)^{k+2} \Lambda_{k+2}(\lambda) \right)$$

$$\leq k^2 R \left( \sum_{i=0}^{k+1} \Lambda_i(\lambda) \Lambda_{k+2}(\lambda)(2E\eta)^{i+k+2} \right)$$

$$\leq k^2 R \Lambda(\lambda)^2$$

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

\[\square\]

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

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

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

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

**Proof.**

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

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

$$\text{Var}[\bar{L}(\lambda)] \lesssim \frac{1}{M} k^2 \Lambda(\lambda)^2$$

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

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

which gives

$$|L(\lambda) - \bar{L}(\lambda)| \leq |L(\lambda) - \mathbb{E}[\bar{L}(\lambda)]| + |\mathbb{E}[\bar{L}(\lambda)] - \bar{L}(\lambda)|$$

$$\lesssim kR \Lambda_{k+2}(\lambda)(2E\eta)^{k+2} + \frac{tkR\Lambda(\lambda)}{\sqrt{M}}.$$

\[\square\]

**E Supporting results: generalized MRA**

This appendix contains supporting results needed to prove Proposition 5.1, which defines a wavelet invariant estimator for generalized MRA. Lemma 5.1 controls the additive noise error and Lemma 5.2 controls the cross-term error. Lemma E.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 (25). Then with probability at least $1 - 1/t^2$

$$|\sum_{j=1}^{M} \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 A_\lambda |\bar{\psi}_\lambda(\omega)|^2 \, d\omega - \sigma^2| \leq \frac{2t\sqrt{k}\sigma^2}{\sqrt{M}}.$$ 

Proof. Let

$$D(\epsilon, \lambda) := \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 A_\lambda |\bar{\psi}_\lambda(\omega)|^2 \, d\omega.$$ 

By Lemma C.1, $E_\epsilon [|\tilde{e}_j(\omega)|^2] = \sigma^2$, and we thus obtain:

$$E_\epsilon [D(\epsilon, \lambda)] = E_\epsilon \left[ \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 A_\lambda |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right]$$

$$= E_\epsilon \left[ \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 |\bar{\psi}_\lambda(\omega)|^2 \, d\omega - \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 B_2 \eta^2 \lambda^2 \frac{d}{d\lambda^2} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega - \frac{1}{2\pi} \int |\tilde{e}_j(\omega)|^2 B_k \eta^2 \lambda^k \frac{d}{d\lambda^k} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right]$$

$$= \sigma^2 \left[ \frac{1}{2\pi} \int |\bar{\psi}_\lambda(\omega)|^2 \, d\omega - \frac{B_2 \eta^2}{2\pi} \int \lambda^2 \frac{d}{d\lambda^2} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega - \frac{B_k \eta^k}{2\pi} \int \lambda^k \frac{d}{d\lambda^k} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right]$$

$$= \sigma^2 (1 - 0 - \ldots - 0)$$

$$= \sigma^2,$$

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

$$E_\epsilon [D(\epsilon, \lambda)^2] \leq E_\epsilon \left[ k \sum_{m=0,2,\ldots,k} \left( \frac{B_m \eta^m}{2\pi} \int |\tilde{e}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right)^2 \right]$$

where we let $\frac{d}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2$ denote $|\bar{\psi}_\lambda(\omega)|^2$ and $B_0 = 1$. By Lemma C.1, we have $E_\epsilon [|e_j(\omega)|^2 | e_j(\xi)|^2] \leq 3\sigma^4$ for all frequencies $\omega, \xi$, so that

$$E_\epsilon \left[ \left( \frac{B_m \eta^m}{2\pi} \int |\tilde{e}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right)^2 \right]$$

$$\leq E_\epsilon \left[ \frac{B^2_m \eta^{2m}}{4\pi^2} \int |\tilde{e}_j(\omega)|^2 |\bar{\psi}_\lambda(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\xi)|^2 \, d\omega \, d\xi \right]$$

$$\leq 3\sigma^4 \left( \frac{1}{2\pi} \int \left| B_m \eta^m \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2 \right|^2 \, d\omega \right)^2$$

$$\leq 3\sigma^8 \Psi^2_m (E\eta)^{2m},$$

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

$$E_\epsilon [D(\epsilon, \lambda)^2] \leq k \sum_{m=0,2,\ldots,k} E_\epsilon \left[ \left( \frac{B_m \eta^m}{2\pi} \int |\tilde{e}_j(\omega)|^2 \lambda^m \frac{d^m}{d\lambda^m} |\bar{\psi}_\lambda(\omega)|^2 \, d\omega \right)^2 \right]$$

$$\leq 3k\sigma^4 \sum_{m=0,2,\ldots,k} \Psi^2_m (E\eta)^{2m} := (I)$$
so that
\[ \mathbb{E}_\varepsilon \left[ D(\varepsilon_j, \lambda) - \sigma^2 \right] = 0 \]
\[ \text{Var}_\varepsilon \left[ D(\varepsilon_j, \lambda) - \sigma^2 \right] = \text{Var}_\varepsilon \left[ D(\varepsilon_j, \lambda) \right] \leq \mathbb{E}_\varepsilon \left[ (D(\varepsilon_j, \lambda))^2 \right] \leq (I). \]

Thus
\[ \text{Var}_\varepsilon \left( \frac{1}{M} \sum_{j=1}^{M} D(\varepsilon_j, \lambda) - \sigma^2 \right) \leq \frac{(I)}{M} \]
so that by Chebyshev's Inequality with probability at least \(1 - 1/t^2\)
\[ \left| \frac{1}{M} \sum_{j=1}^{M} D(\varepsilon_j, \lambda) - \sigma^2 \right| \leq t \frac{\sqrt{(I)}}{\sqrt{M}} \leq t \sqrt{3k} \left( \sum_{m=0,2,...,k} \Psi_m(En)^m \right) \frac{\sigma^2}{\sqrt{M}} = 2t \sqrt{k} \Psi \frac{\sigma^2}{\sqrt{M}}. \]

**Lemma 5.2** Let the notation and assumptions of Proposition 5.1 hold, and let \( A_\lambda \) be the operator defined in (25). Then with probability at least \(1 - 1/t^2\)
\[ \left| \frac{1}{M} \sum_{j=1}^{M} \frac{1}{2\pi} \int \left( \hat{f}_{\varepsilon_j}(\omega) \hat{e}_j(\omega) + \hat{f}_{\varepsilon_j}(\omega) \hat{e}_j(\omega) \right) A_\lambda |\hat{\psi}_\lambda(\omega)|^2 d\omega \right| \leq \frac{t}{\sqrt{M}} \sqrt{\Psi(\Lambda_0(\lambda) + \Lambda(\lambda))}. \]

**Proof.** We have
\[ \frac{1}{M} \sum_{j=1}^{M} \frac{1}{2\pi} \int \left( \hat{f}_{\varepsilon_j}(\omega) \hat{e}_j(\omega) + \hat{f}_{\varepsilon_j}(\omega) \hat{e}_j(\omega) \right) A_\lambda |\hat{\psi}_\lambda(\omega)|^2 d\omega = \frac{1}{M} \sum_{j=1}^{M} Y_j + \overline{Y_j} \]
where
\[ Y_j := \frac{1}{2\pi} \int \left( \hat{f}_{\varepsilon_j}(\omega) \hat{e}_j(\omega) \right) A_\lambda |\hat{\psi}_\lambda(\omega)|^2 d\omega. \]
The random variable \( Y_j \) has randomness depending on both \( \varepsilon_j \) and \( \tau_j \). Note that
\[ \mathbb{E}_{\varepsilon, \tau}[Y_j] = \mathbb{E}_{\varepsilon, \tau}[\mathbb{E}_{\varepsilon, \tau}[Y_j | \tau_j]] \]
since \( Y_j \) is integrable. Thus since \( \mathbb{E}_{\varepsilon, \tau}[\hat{e}_j(\omega)] = 0 \), we obtain \( \mathbb{E}_{\varepsilon, \tau}[Y_j | \tau_j] = 0 \), which yields \( \mathbb{E}_{\varepsilon, \tau}[Y_j] = 0 \). We also have:
\[ \text{Var}_{\varepsilon, \tau}[Y_j] = \mathbb{E}_{\varepsilon, \tau}[Y_j^2] \]
\[ \leq \mathbb{E}_{\varepsilon, \tau} \left[ \left( \frac{1}{2\pi} \int |\hat{f}_{\varepsilon_j}(\omega)| \cdot |\hat{e}_j(\omega)| \cdot |A_\lambda |\hat{\psi}_\lambda(\omega)|^2| \ d\omega \right)^2 \right] \]
\[ \leq \mathbb{E}_{\varepsilon, \tau} \left[ \left( \frac{1}{2\pi} \int |\hat{f}_{\varepsilon_j}(\omega)|^2 \cdot |A_\lambda |\hat{\psi}_\lambda(\omega)|^2| \ d\omega \right) \left( \frac{1}{2\pi} \int |\hat{e}_j(\omega)|^2 \cdot |A_\lambda |\hat{\psi}_\lambda(\omega)|^2| \ d\omega \right) \right] \]
\[ = \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\hat{f}_{\varepsilon_j}(\omega)|^2 \cdot |A_\lambda |\hat{\psi}_\lambda(\omega)|^2| \ d\omega \right] \mathbb{E}_\varepsilon \left[ \frac{1}{2\pi} \int |\hat{e}_j(\omega)|^2 \cdot |A_\lambda |\hat{\psi}_\lambda(\omega)|^2| \ d\omega \right]. \]
Letting \( B_0 = 1 \) and applying Lemma A.2, we have:

\[
\mathbb{E}_\tau \left[ \frac{1}{2\pi} \int |\hat{f}_{r_j}(\omega)|^2 \cdot |A_\lambda | \hat{\psi}_\lambda(\omega)|^2 \right. \] 
\[
\left. \cdot d\omega \right] \leq \mathbb{E}_\tau \left[ \sum_{m=0,2,\ldots} \frac{1}{2\pi} \int |\hat{f}_{r_j}(\omega)|^2 \cdot \left| B_m \eta^m \frac{d^m}{d\lambda^m} |\hat{\psi}_\lambda(\omega)|^2 \right| \right. \] 
\[
\left. \cdot d\omega \right] 
\leq \mathbb{E}_\tau \left[ \sum_{m=0,2,\ldots} (\eta^m) \left( \int |\hat{f}_{r_j}(\omega)|^2 \cdot \left( \frac{\|f'\|_1^2 \Psi_m}{\lambda^2} \right) \right) \right. \] 
\[
\left. \cdot d\omega \right] 
\leq \sum_{m=0,2,\ldots} (\eta^m) \left( \frac{4 \|f'\|_1^2 \Psi_m}{\lambda^2} \right) 
\leq 4 \sum_{m=0,2,\ldots} (\eta^m) \Lambda_m(\lambda) 
\lesssim \Lambda_0(\lambda) + \Lambda(\lambda)
\]
since \( \|r_j\|_\infty \leq \frac{1}{2} \) guarantees \( \|f'\|_1 = \frac{1}{1-r_j} \|f'\|_1 \leq 2 \|f'\|_1 \). Also:

\[
\mathbb{E}_\tau \left[ \frac{1}{2\pi} \int |\tilde{\psi}_j(\omega)|^2 \cdot |A_\lambda | \hat{\psi}_\lambda(\omega)|^2 \right. \] 
\[
\left. \cdot d\omega \right] \leq \mathbb{E}_\tau \left[ \sum_{m=0,2,\ldots} \frac{1}{2\pi} \int |\tilde{\psi}_j(\omega)|^2 \cdot \left| B_m \eta^m \frac{d^m}{d\lambda^m} |\hat{\psi}_\lambda(\omega)|^2 \right| \right. \] 
\[
\left. \cdot d\omega \right] 
= \sigma^2 \left( \sum_{m=0,2,\ldots} \frac{1}{2\pi} \int \left| B_m \eta^m \frac{d^m}{d\lambda^m} |\hat{\psi}_\lambda(\omega)|^2 \right| \right. \] 
\[
\left. \cdot d\omega \right) 
\leq \sigma^2 \sum_{m=0,2,\ldots} (\eta^m) \Psi_m 
= \sigma^2 \Psi
\]

where the second line follows from Lemma C.1 in Appendix C and the next to last line from Corollary A.1 in Appendix A. We thus have:

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

and an identical argument can be applied to the \( \bar{Y}_j \) so that by Chebyshev’s Inequality with probability at least \( 1 - 1/\tau^2 \):

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

\[\square\]

**Lemma E.1** Assume \( \psi \) is \( k \)-admissible. Then:

\[
\int \lambda^m \left( \frac{d^m}{d\lambda^m} |\hat{\psi}_\lambda(\omega)|^2 \right) d\omega = 0
\]

for all \( 1 \leq m \leq k \).

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

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

We first establish that:

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

(34)
The proof is by induction. When $k = 1$, we obtain

$$\text{LHS of Eqn. (34)} = \lambda \frac{d}{d\lambda} \left( \frac{1}{\lambda} \frac{\omega}{\lambda^2} \right) = -\frac{\omega}{\lambda^2} g'(\frac{\omega}{\lambda}) - \frac{1}{\lambda} g'(\frac{\omega}{\lambda}) = -\omega g'_{\lambda}(\omega) - g_{\lambda}(\omega)$$

and

$$\text{RHS of Eqn. (34)} = \frac{d}{d\omega} (-\omega g_{\lambda}(\omega)) = -\omega g'_{\lambda}(\omega) - g_{\lambda}(\omega),$$

so the base case is established. We now assume that Equation (34) holds and show it also holds for $k + 1$ replacing $k$. By the inductive hypothesis:

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

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

$$+ \frac{d}{d\omega} \left( \omega \lambda^{-2} \frac{d}{d\lambda} g_{\lambda}(\omega) \right) + (k - 1) \lambda^{-2} \frac{d}{d\lambda} g_{\lambda}(\omega)$$

$$= -\lambda^{-1} \frac{d}{d\lambda} g_{\lambda}(\omega) \quad \text{by inductive hypothesis}$$

so that

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

Thus (34) is established. We now use integration by parts to show (34) implies (33) in the Lemma. The proof of (33) is once again by induction. When $k = 1$, we have already shown

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

Integration by parts gives

$$\int \omega g'_{\lambda}(\omega) \, d\omega = \left[ \omega g_{\lambda}(\omega) \right]_{-\infty}^{\infty} - \int g_{\lambda}(\omega) \, d\omega = \int g_{\lambda}(\omega) \, d\omega.$$ 

Note $\omega g_{\lambda}(\omega)$ vanishes at $\pm\infty$ since $g \in L^1$ guarantees $g_{\lambda} \in L^1$, and thus $g_{\lambda}$ must decay faster that $\omega^{-1}$. Utilizing (35),

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

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 (34), we obtain:

\[
\int \lambda^k \left( \frac{d}{d\lambda} g_\lambda(\omega) \right) d\omega = \int \frac{d}{d\omega} \left( -\omega \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) \right) d\omega - (k-1) \int \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) d\omega
\]

\[=0 \text{ by induc. hypo.}\]

\[
\int \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) d\omega = \int \frac{d}{d\omega} \lambda^{k-1} g_\lambda(\omega) d\omega - \int \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) d\omega
\]

\[=0 \text{ by induc. hypo.}\]

\[
-\omega \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) \bigg|_{-\infty}^{\infty} + \int \lambda^{k-1} \frac{d}{d\lambda} g_\lambda(\omega) d\omega = 0 \text{ by induc. hypo.}
\]

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

### F Moment estimation for generalized MRA

In this appendix we outline a moment estimation procedure for generalized 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

\[
\beta_m(y_j) = \int_0^{2\pi} \omega^m \tilde{y}_j(\omega) d\omega
\]

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

\[
g_m(\ell, \sigma) = \int_0^{2\pi} \int_0^{2\pi} 2\alpha^2 \xi^m \omega^m \sin\left(\frac{1}{2}(\xi - \omega)\right) \frac{d\omega d\xi}{(\xi - \omega)}
\]

and an order \(m\) additive noise adjusted squared coefficient of variation by:

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

**Remark F.1** If the noisy signals are supported in \([-N/2, N/2]\) instead of \([-1/2, 1/2]\), (37) is replaced with:

\[
g_m(N, \ell, \sigma) = \int_0^{2\pi} \int_0^{2\pi} 2\alpha^2 \xi^m \omega^m \sin\left(\frac{N}{2}(\xi - \omega)\right) \frac{d\omega d\xi}{(\xi - \omega)}.
\]

The following proposition mirrors Proposition 6.1 for dilation MRA; its proof appears at the end of Appendix F.
**Proposition F.1** Assume Model 2 with \( t = 0 \) and \( CV_0, CV_1 \) defined by (36), (37), and (38). Then

\[
CV_0 = \eta^2 + (3C_4 - 3)\eta^4 + O(\eta^6)
\]
\[
CV_1 = 4\eta^2 + (25C_4 - 33)\eta^4 + O(\eta^6).
\]

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

**Definition F.2** Assume Model 3 with \( t = 0 \) and \( \tilde{CV}_0, \tilde{CV}_1 \) the empirical counterparts of (38). Define the second order estimator of \( \eta^2 \) by \( \tilde{\eta}^2 = \tilde{CV}_0 \). Define the fourth order estimators of \( (\eta^2, C_4\eta^4) \) by the unique positive solution \( (\tilde{\eta}^2, \tilde{C}_4) \) of

\[
\tilde{CV}_0 = \eta^2 + (3C_4 - 3)\eta^4
\]
\[
\tilde{CV}_1 = 4\eta^2 + (25C_4 - 33)\eta^4.
\]

As \( M \to \infty \), 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 (38) will be replaced with \( g_m(\ell, \sigma) \pm O(\sigma^2/\sqrt{M}) \), so that the estimators given in Definition F.2 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 generalized MRA when \( t = 0 \), we repeat the simulations of Section 5.1 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 \( \sigma = 2^{-5}, 2^{-4} \) 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 \( \tilde{\eta}^2 \) is approximated with a second order estimator (see Definition F.2).

Results are shown in Figure 6, 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 low 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.
Figure 6: $L^2$ error with standard error bars for generalized MRA model ($t = 0$, empirical moment estimation). First, second, third column shows results for low, medium, high frequency signals. All plots have the same axis limits.
Proof of Proposition F1. Since $y_j = L_t, f + \varepsilon_j$, we have

$$
\mathbb{E}[\alpha_m(y_j)] = \mathbb{E} \left[ \int_0^{2\pi} \omega^m (\hat{f}_j(\omega) + \hat{\varepsilon}_j(\omega)) \, d\omega \right] \\
= \mathbb{E} \left[ \int_0^{2\pi} \omega^m \hat{f}_j(\omega) \, d\omega \right] \\
= \mathbb{E} \left[ \int_0^{2\pi} \omega^m \hat{f}((1 - \tau_j) \omega) \, d\omega \right] \\
= \mathbb{E} \left[ \int_0^{2\pi} \omega^m \frac{\xi^m}{(1 - \tau_j) \hat{f}(\xi)} \frac{d\xi}{(1 - \tau_j)} \right] \\
= \alpha_m(f) \mathbb{E} \left[ (1 - \tau_j)^{-(m+1)} \right].
$$

We now compute the variance. We first establish that

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

By Thm 4.5 of [51]

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

so that

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

We thus obtain:

$$
\left| \alpha_m(y_j) \right|^2 = \mathbb{E} \left[ \left( \int_0^{2\pi} \omega^m (\hat{f}_j(\omega) + \hat{\varepsilon}_j(\omega)) \, d\omega \right) \left( \int_0^{2\pi} \omega^m (\overline{\hat{f}_j(\omega) + \hat{\varepsilon}_j(\omega)}) \, d\omega \right) \right] \\
= \mathbb{E} \left[ \left( \int_0^{2\pi} \omega^m \hat{f}(\omega) \, d\omega \right) \left( \int_0^{2\pi} \omega^m \overline{\hat{f}(\omega)} \, d\omega \right) \right] \\
+ \mathbb{E} \left[ \left( \int_0^{2\pi} \omega^m \hat{\varepsilon}_j(\omega) \, d\omega \right) \left( \int_0^{2\pi} \omega^m \overline{\hat{\varepsilon}_j(\omega)} \, d\omega \right) \right] \\
= \mathbb{E} \left[ (1 - \tau_j)^{-2(m+1)} \alpha_m(f) \overline{\alpha_m(f)} \right] + g_m(\ell, \sigma) \\
= |\alpha_m(f)|^2 \mathbb{E} \left[ (1 - \tau_j)^{-2(m+1)} \right] + g_m(\ell, \sigma).
$$
Thus:
\[
\text{Var}[\alpha_m(y_j)] - g_m(\ell, \sigma) = \mathbb{E} \left[ |\alpha_m(y_j)|^2 \right] - g_m(\ell, \sigma) - |\mathbb{E}[\alpha_m(y_j)]|^2 \\
= |\alpha_m(f)|^2 \mathbb{E} \left[ (1 - r_j)^{-2(m+1)} \right] - |\alpha_m(f)|^2 \left( \mathbb{E} \left[ (1 - r_j)^{-2(m+1)} \right] \right)^2.
\]
Dividing by $|\mathbb{E}[\alpha_m(y_j)]|^2$ gives:
\[
CV_m = \frac{\mathbb{E} \left[ (1 - r_j)^{-2(m+1)} \right]}{\left( \mathbb{E} \left[ (1 - r_j)^{-2(m+1)} \right] \right)^2} - 1,
\]
and the remainder of the proof is identical to the proof of Proposition 6.1. □

G Additional simulations for generalized MRA

We investigate the $L^2$ error of estimating the power spectrum using PS ($k = 0$) and WSC ($k = 0, 2, 4$) for three additional high frequency functions:
\[
\begin{align*}
    f_4(x) &= 2.5 \cos(32x) \cdot 1(x \in [-0.2, 0.2]) \\
    f_5(x) &= \exp^{-0.04x^2} \cos(30x + 1.5x^2) \\
    f_6(x) &= (6/\pi) \cos(35x) \text{sinc}(3x).
\end{align*}
\]
The signal $f_4$ is not continuous and has compact support, with a slowly decaying, oscillating Fourier transform given by $\hat{f}_4(\omega) = \text{sinc}(0.2(\omega - 32)) + \text{sinc}(0.2(-\omega - 32))$. 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 $\hat{f}_6(\omega) = 1(\omega \in [-38, -32]) + 1(\omega \in [32, 38])$.

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 5.1) and with empirical moment estimation on the $t = 0$ restricted model (parameter values and moment estimation as described in Section 6.2 and Appendix F). Figures 7 and 8 show the $L^2$ error for oracle and empirical moment estimation respectively. As for the high frequency Gabor in Section 5.1, WSC ($k = 2$) and WSC ($k = 4$) significantly outperformed the zero order estimators. In addition for large dilations, the WSC ($k = 4$) outperformed WSC ($k = 2$).

H Supporting results: stochastic calculus

This appendix contains several stochastic calculus results which are used to control the statistics of the additive noise. Proposition H.1 is a simple generalization of Thm 4.5 of [51]. Proposition H.2 controls the second moment of the stochastic quantity in Proposition H.1, and is in fact a special case of Proposition H.3. Both Propositions H.2 and H.3 are proved with standard techniques from stochastic calculus, and for brevity we omit the proofs.

**Proposition H.1** Assume $\int_0^T f(t)^2 \, dt < \infty$, $\int_0^T f(t)^{-2} \, dt < \infty$, and let $B_t$ be a Brownian motion with variance $\sigma^2$. Then:
\[
\mathbb{E} \left[ \left( \int_0^T f(t) \, dB_t \right) \left( \int_0^T f(t) \, dB_t \right) \right] = \sigma^2 \int_0^T f(t) \, dt.
\]

**Proposition H.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 nonrandom time $T$, we have:
\[
\mathbb{E} \left[ \left( \int_0^T f(t) \, dB_t \right)^2 \left( \int_0^T f(t)^{-2} \, dt \right) \right] = 2\sigma^4 \left( \int_0^T f(t)^2 \, dt \right)^2 + \sigma^4 \left( \int_0^T f(t)^{-2} \, dt \right) \left( \int_0^T f(t)^{-2} \, dt \right).
\]
Figure 7: $L^2$ error with standard error bars for generalized MRA model (oracle moment estimation). First, second, third column shows results for $f_4$, $f_5$, $f_6$. All plots for the same signal have the same axis limits.
Figure 8: $L^2$ error with standard error bars for generalized MRA model ($t = 0$, empirical moment estimation). First, second, third column shows results for $f_1, f_5, f_6$. All plots for the same signal have the same axis limits.
Corollary H.1 When \( f(t) \) is real, the above reduces to:

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

We now derive a slightly more general version of Proposition H.2.

Proposition H.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 nonrandom time \( T \), we have:

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

References


