Journal of Fourier Analysis and Applications (2023) 29:43
https://doi.org/10.1007/s00041-023-10023-5

Power Spectrum Unbiasing for Dilation-Invariant Multi-reference Alignment

Matthew Hirn[1] · **Anna Little**[2]

Received: 25 January 2022 / Revised: 27 February 2023 / Accepted: 11 April 2023 /
Published online: 8 July 2023

Abstract
This article discusses a generalization of the 1-dimensional multi-reference alignment problem. The goal is to recover a hidden signal from many noisy observations, where each noisy observation includes a random translation and random dilation of the hidden signal, as well as high additive noise. We propose a method that recovers the power spectrum of the hidden signal by applying a data-driven, nonlinear unbiasing procedure, and thus the hidden signal is obtained up to an unknown phase. An unbiased estimator of the power spectrum is defined, whose error depends on the sample size and noise levels, and we precisely quantify the convergence rate of the proposed estimator. The unbiasing procedure relies on knowledge of the dilation distribution, and we implement an optimization procedure to learn the dilation variance when this parameter is unknown. Our theoretical work is supported by extensive numerical experiments on a wide range of signals.

Keywords Multi-reference alignment · Method of invariants · Dilations · Signal processing

1 Introduction

In classic multi-reference alignment (MRA), one attempts to recover a hidden signal $f : \mathbb{R} \to \mathbb{R}$ from many noisy observations, where each noisy observation has been

Communicated by Afonso Bandeira.

✉ Anna Little
little@math.utah.edu

[1] Department of Computational Mathematics, Science and Engineering, Department of Mathematics, and the Center for Quantum Computing, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

[2] Department of Mathematics, Utah Center for Data Science, University of Utah, Salt Lake City, UT 84112, USA

randomly translated and corrupted by additive noise, as described in the following model.

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

$$y_j(x) = f(x - t_j) + \varepsilon_j(x), \quad 1 \le j \le M, \tag{1}$$

where:

(i) $supp(y_j) \subseteq [-\frac{1}{2}, \frac{1}{2}]$ *for* $1 \le j \le M$.
(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* σ^2.

This toy model is a first step towards more realistic models arising in cryo-electron microscopy (cryo-EM), and is relevant in many other applications including structural biology [25, 44, 45, 48, 49, 54]; radar [30, 58]; single cell genomic sequencing [35]; image registration [12, 28, 47]; and signal processing [58]. Some methods solve Model 1 via *synchronization* [2–5, 10, 16, 17, 46, 52, 57], i.e., the translation factors $\{t_j\}_{j=1}^M$ are explicitly recovered and the signals aligned. Synchronization approaches will fail in the high noise regime when the signal-to-noise ratio (SNR) is low, but the hidden signal can still be recovered by methods which avoid alignment; these include the *method of moments* [31, 34, 50], which contain the *method of invariants* [6, 8, 21] as a special case, and *expectation-maximization type algorithms* [1, 23]. The method of invariants leverages translation invariant Fourier features such as the power spectrum and bispectrum, i.e., *Fourier invariants*, as they are especially useful for solving Model 1. Recall the Fourier transform (FT) of a signal $f \in \mathbf{L}^1(\mathbb{R})$ is defined as

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

and its power spectrum is then defined by $(Pf)(\omega) = |\widehat{f}(\omega)|^2$ and its bispectrum by $(Bf)(\omega_1, \omega_2) = \widehat{f}(\omega_1)\widehat{f}(\omega_2)^* \widehat{f}(\omega_2 - \omega_1)$. Furthermore, continuous time white noise is defined as the derivative of a Brownian motion, i.e., $\epsilon(x) = dB_x$ so that $\widehat{\epsilon}(\omega) = \int_{-1/2}^{1/2} e^{-ix\omega} dB_x$.

Although foundational, Model 1 fails to capture important sources of randomness which appear in many applications. For example in cryo-EM, 3D molecules are randomly rotated and only the 2D tomographic projection is observed. Furthermore, macromolecular structures contain flexible regions whose dynamics have important affects on molecular interactions [26, 27, 36, 37] and the regulation of histone tails [11, 24, 40], but these dynamics make imaging highly challenging. Mathematically, one can model the situation as a *diffeomorphism*, i.e., there is an underlying structure $f(x)$ but one observes $f(\zeta(x))$ for some random diffeomorphism $\zeta(x)$. Translations, i.e., $\zeta(x) = x - t$, are the simplest diffeomorphism and give rise to classic MRA. In this article we analyze the following generalization of classic MRA, where signals are also corrupted by a random scale change (i.e., dilation) in addition to random

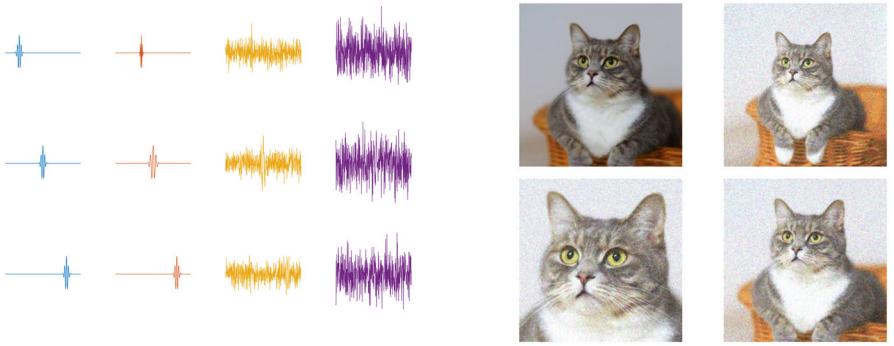


Fig. 1 Left: Illustration of Model 2. A hidden signal is randomly translated (column 1), randomly dilated (column 2), and then corrupted by additive noise (columns 3 and 4). Column 3 shows SNR $= \frac{1}{2}$ and column 4 shows SNR $= 2$, which is the noise level considered in the simulations reported in Sect. 6. Right: Illustration of Noisy dilation MRA in two dimensions. From a collection of noisy images (bottom and top right), one seeks to recover a high-resolution image (top left)

translation and additive noise; see Fig. 1. This corresponds to the next simplest class of diffeomorphisms, i.e., affine functions $\zeta(x) = C(x - t)$.

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

$$y_j(x) = f\left((1-\tau_j)^{-1}(x - t_j)\right) + \varepsilon_j(x)\,, \quad 1 \le j \le M\,. \tag{2}$$

In addition, we assume:

(i) $supp(y_j) \subseteq [-\frac{1}{2}, \frac{1}{2}]$ *for* $1 \le j \le M$.
(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 from a uniformly distributed random variable* τ *with mean zero and variance* $\eta^2 \le 1/12$*, so that:*

$$\tau \in [-\sqrt{3}\eta, \sqrt{3}\eta] \quad , \quad \mathbb{E}(\tau) = 0 \quad , \quad Var(\tau) = \eta^2.$$

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

Dilations are highly relevant in many applications, for example in the analysis of time-warped audio signals [41–43] and in imaging [13, 15, 33, 39, 47, 55], where distance to an object and camera zoom determine the scale of an object. Note the box size in Models 1 and 2 is arbitrary; more generally, the signals may be supported on any finite interval $[-\frac{N}{2}, \frac{N}{2}]$. All results still hold with $\sigma\sqrt{N}$ replacing σ. The assumption $\eta^2 \le \frac{1}{12}$ is also arbitrary but arises from requiring that the dilation factors satisfy $(1 - \tau) \in [\frac{1}{2}, 2]$; more generally, one can replace this requirement with $(1 - \tau) \in [r_{\min}, r_{\max}]$ for some $r_{\min}$ lower bounded away from 0; this will affect constants but all key results still hold.

Remark 1 We consider $\mathbf{L}^\infty(\mathbb{R})$ normalized dilations in Model 2, i.e., we consider a dilation operator $D_C f(x) = f(C^{-1}x)$ which preserves the infinity norm, since $\|D_C f\|_\infty = \|f\|_\infty$. However the method is easily modified to accommodate $\mathbf{L}^p(\mathbb{R})$ normalized dilation operators where $D_C f(x) = C^{-1/p} f(C^{-1}x)$, so that $\|D_C f\|_p = \|f\|_p$. We consider $p = \infty$ because this is natural for imaging applications, but other normalizations are useful in other contexts. For example it is natural to take $p = 1$ in the statistical context, where one may observe samples from a family of distributions which are shifts and rescalings of an underlying distribution.

Solving Model 2 is highly challenging. Dilations cause instabilities in the high frequencies of a signal, where even a small dilation can lead to a large perturbation of the frequency values. In this paper we generalize the *method of invariants* to the setting of dilation corruptions, i.e., we seek to fully solve Model 2 using Fourier invariants. Note kth order Fourier invariants are obtained by taking the Fourier transform of the kth order auto-correlation function, a process which always results in translation-invariant features; for example power spectra are second order Fourier invariants while bispectra are third order. Fully recovering the hidden signal in Model 2 via Fourier invariants involves two key challenges (1) remove the dilation bias from the Fourier invariants and (2) invert the Fourier invariants to obtain the hidden signal. In this paper we focus on the first key challenge, and for simplicity we start with a method for recovering the *power spectrum* of the hidden signal, i.e., $k = 2$. However we expect that the methodology developed here can in fact be generalized to Fourier invariants of any order. For a 1d signal to be uniquely determined by its Fourier invariants, $k = 3$ is generally required, since the bispectrum is invertible under very general conditions [56].

In this paper we do not focus on key challenge (2), but we note that there are already many works in the literature regarding this topic. Although the hidden signal is in general not uniquely determined by its power spectrum, there are some exceptions, such as when the signal lives in a spline or shift invariant space [20, 53], or is sufficiently sparse [7]; see [9, 51] for an overview of phase retrieval results. Inverting the bispectrum is more tenable, and methods include non-convex optimization over the manifold of phases [8], iterative phase synchronization [8], semi-definite programming, phase unwrapping, frequency marching [29, 48], and the spectral method proposed in [18]. Thus the main contribution of the present work is a novel technique for key challenge (1), i.e., dilation unbiasing of Fourier invariants; future work will investigate full signal recovery by extending the method from the power spectrum to the bispectrum, and then combining with existing bispectrum inversion algorithms.

We note in [32] an alternate method for dilation unbiasing of Model 2 is proposed, assuming $\mathbf{L}^1(\mathbb{R})$ normalized dilations. The authors define wavelet-based, translation invariant features and unbias for dilations by utilizing the first few moments of the dilation distribution (note similar unbiasing procedures have been applied in the deconvolution context [14, 22]). The method has two main short-comings: although it can reduce the bias due to dilations, it cannot remove it entirely, i.e., the method of [32] does not define an *unbiased estimator* of the true features. In addition, inverting the wavelet-based features to recover the power spectrum of the hidden signal is numerically unstable, as it is driven by the condition number of a low rank matrix. This article

proposes a method which overcomes both of these challenges: by working directly on the power spectrum, we avoid a numerically unstable inversion, and we develop a new unbiasing procedure which yields an *unbiased estimator* of the power spectrum of the hidden signal; we refer to this unbiasing procedure as *inversion unbiasing*. To achieve this we assume explicit knowledge of the dilation distribution instead of knowledge of the first few moments. To illustrate inversion unbiasing, it is helpful to define the following model in which signals are randomly translated and dilated, but not corrupted by additive noise.

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

$$y_j(x) = f\left((1-\tau_j)^{-1}(x-t_j)\right), \quad 1 \le j \le M. \tag{3}$$

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

Since Model 3 lacks additive noise, it can in fact be trivially solved by first estimating $\|f\|_2$, and then dilating any observed signal to have the right norm (for further details see [32]). We use Model 3 to build a theory to solve Model 2, but note it is not of independent interest.

Remark 2 This article generalizes the *method of invariants* for Model 2. Another possible approach is to generalize the *expectation maximization* algorithm proposed in [1] (see Appendix I in [32]). Although this may work well in many cases, some disadvantages are: (1) the high computational cost of each iteration, (2) convergence to local (not global) minima, and (3) error accumulation due to off grid interpolation of dilated function values.

The remainder of the article is organized as follows. Section 2 motivates inversion unbiasing by first considering the infinite sample size case. Section 3 presents our main results for solving Models 2 and 3 in the finite sample regime. Section 4 discusses how inversion unbiasing is implemented via an optimization algorithm. Section 5 discusses the discretization error incurred when the continuous signals of Model 2 are discretely sampled. Section 6 reports simulation results testing the performance of inversion unbiasing. Section 7 concludes the article and summarizes future research directions.

1.1 Notation

Let $f_j(x) = f\left((1-\tau_j)^{-1}(x-t_j)\right)$ denote the jth signal which is dilated by $1-\tau_j$. We note that $\widehat{f_j}(\omega) = e^{-it_j\omega}(1-\tau_j)\widehat{f}((1-\tau_j)\omega)$, so that

$$(Pf_j)(\omega) = (1-\tau_j)^2(Pf)((1-\tau_j)\omega).$$

We let $g = Pf$, and for Models 2 and 3 we define

$$g_\eta(\omega) := \mathbb{E}_\tau\left[(Pf_j)(\omega)\right]. \tag{4}$$

Thus $g(\omega)$ is the power spectrum of the hidden signal, while $g_\eta(\omega)$ is the expected value of the power spectrum under dilation corruption. Note for Model 2, it is easy to show that

$$g_\eta(\omega) = \mathbb{E}_{\tau,\epsilon}[(Py_j)(\omega) - \sigma^2]\,,$$

since $y_j = f_j + \epsilon_j$ and $\mathbb{E}_{\tau,\epsilon}[(P\epsilon_j)(\omega)] = \sigma^2$; see for example Proposition 3.1 in [32]. We let $g'_\eta(\omega) = \frac{d}{d\omega} g_\eta(\omega)$ denote the normal derivative. We also define the following constants which depend on η:

$$B_0 = \frac{(1-\sqrt{3}\eta)}{(1+\sqrt{3}\eta)}\,,\ B_1 = 2\sqrt{3}\eta\,,\ B_2 = \frac{1}{1+\sqrt{3}\eta}\,, \tag{5}$$

and we let $(L_C g)(\omega) = C^3 g(C\omega)$ be a dilation operator. We use a^* to denote the complex conjugate of a, and $a \wedge b$ to denote $\min\{a, b\}$. When $a \le Cb$ for an absolute constant C, we say $a = O(b)$ and write $a \lesssim b$. Finally, we let $\mathbf{C}^0(\mathbb{R})$ denote the space of continuous functions on $\mathbb{R}$, and $\mathbf{C}^k(\mathbb{R})$ functions on $\mathbb{R}$ with k continuous derivatives.

2 Infinite Sample Estimate

To motivate our finite sample procedure, we first consider how to define an unbiased estimator in the infinite sample limit. We can recover Pf from g_η, as stated in the following Proposition.

Proposition 1 *Assume $Pf \in \mathbf{C}^0(\mathbb{R})$ and g_η as defined in (4). Then for $\omega \neq 0$:*

$$(Pf)(\omega) = (I - L_{B_0})^{-1} B_1 L_{B_2}(3g_\eta(\omega) + \omega g'_\eta(\omega))\,,$$

where B_0, B_1, B_2 are as defined in (5).

Proof of Proposition 1 Since τ has a uniform distribution with variance η^2, the probability density function (pdf) of τ has form $p_\tau(\omega) = \frac{1}{2\sqrt{3}\eta} \mathbb{1}_{[-\sqrt{3}\eta, \sqrt{3}\eta]}(\omega)$. Thus:

$$\begin{aligned} g_\eta(\omega) &:= \mathbb{E}_\tau[(1-\tau)^2 g((1-\tau)\omega)] \\ &= \int (1-\tau)^2 g((1-\tau)\omega) p_\tau(\omega)\, d\tau \\ &= \frac{1}{2\sqrt{3}\eta} \int_{-\sqrt{3}\eta}^{\sqrt{3}\eta} (1-\tau)^2 g((1-\tau)\omega)\, d\tau \\ &= \frac{1}{2\sqrt{3}\eta} \int_{(1-\sqrt{3}\eta)\omega}^{(1+\sqrt{3}\eta)\omega} \frac{\widetilde{\tau}^2}{\omega^2} g(\widetilde{\tau}) \frac{1}{\omega}\, d\widetilde{\tau}\,, \end{aligned}$$

where we have applied the change of variable $\widetilde{\tau} = (1-\tau)\omega$,

$d\tau = -\frac{1}{\omega}\, d\widetilde{\tau}$. Letting $h(x) = x^2\, g(x)$ and $H(x)$ an antiderivative of h, by the Fundamental Theorem of Calculus we thus obtain:

$$\begin{aligned} 2\sqrt{3}\eta\omega^3 g_\eta(\omega) &= \int_{(1-\sqrt{3}\eta)\omega}^{(1+\sqrt{3}\eta)\omega} \widetilde{\tau}^2 g(\widetilde{\tau})\, d\widetilde{\tau} \\ &= \int_{(1-\sqrt{3}\eta)\omega}^{(1+\sqrt{3}\eta)\omega} h(\widetilde{\tau})\, d\widetilde{\tau} \\ &= H((1+\sqrt{3}\eta)\omega) - H((1-\sqrt{3}\eta)\omega)\,. \end{aligned}$$

Differentiating with respect to ω yields:

$$\begin{aligned} &2\sqrt{3}\eta\left(3\omega^2 g_\eta(\omega) + \omega^3 g'_\eta(\omega)\right) \\ &\quad = (1+\sqrt{3}\eta)h((1+\sqrt{3}\eta)\omega) - (1-\sqrt{3}\eta)h((1-\sqrt{3}\eta)\omega)\,, \end{aligned}$$

and dividing by ω^2 gives:

$$\begin{aligned} &2\sqrt{3}\eta\left(3g_\eta(\omega) + \omega g'_\eta(\omega)\right) \\ &= (1+\sqrt{3}\eta)^3 g((1+\sqrt{3}\eta)\omega) - (1-\sqrt{3}\eta)^3 g((1-\sqrt{3}\eta)\omega)\,. \end{aligned}$$

Applying the dilation operator L_{B_2} then gives:

$$\begin{aligned} &B_1 L_{B_2}(3g_\eta + \omega g'_\eta(\omega)) \\ &= g\,(\omega) - \left(\frac{1-\sqrt{3}\eta}{1+\sqrt{3}\eta}\right)^3 g\left(\left(\frac{1-\sqrt{3}\eta}{1+\sqrt{3}\eta}\right)\omega\right) \\ &= (I - L_{B_0})g(\omega)\,. \end{aligned}$$

Since $B_0 < 1$, the series $I + L_{B_0} + L_{B_0}^2 + L_{B_0}^3 + \ldots$ converges, and $I - L_{B_0}$ is invertible. We thus obtain

$$g(\omega) = (I - L_{B_0})^{-1} B_1 L_{B_2}(3g_\eta(\omega) + \omega g'_\eta(\omega))\,,$$

which proves the proposition. □

Proposition 1 indicates that when τ is uniformly distributed, Pf can be recovered from g_η, which is known in the infinite sample limit. It thus provides the key insight on how to unbias for dilations.

3 Finite Sample Estimates

Since we are only given a finite sample, we do not have access to g_η, but for large M, g_η is well approximated by:

$$\widetilde{g}_\eta(\omega) := \frac{1}{M}\sum_{j=1}^{M}(Pf_j)(\omega)\,. \tag{6}$$

For dilation MRA, $\widetilde{g}_\eta$ can be computed exactly, and we describe the resulting estimator in Sect. 3.1. For noisy dilation MRA, $\widetilde{g}_\eta$ cannot be computed exactly due to the additive noise, but an unbiased estimator can still be defined as described in Sect. 3.2.

3.1 Results for Dilation MRA

Motivated by Proposition 1, we define the following estimator for dilation MRA:

$$(\widetilde{Pf})(\omega) := (I - L_{B_0})^{-1}B_1L_{B_2}(3\widetilde{g}_\eta(\omega) + \omega\widetilde{g}'_\eta(\omega))\,, \tag{7}$$

where $\widetilde{g}_\eta$ is as defined in (6) and B_0, B_1, B_2 are as defined in (5). We note that in practice one does not have a closed form formula for applying $(I - L_{B_0})^{-1}$, but $(\widetilde{Pf})(\omega)$ can be obtained by solving the following convex optimization problem:

$$\operatorname{argmin}_{\mathring{g}\geq 0}\,\|(I - L_{B_0})\mathring{g} - B_1L_{B_2}(3\widetilde{g}_\eta(\omega) + \omega\widetilde{g}'_\eta(\omega))\|_2^2\,.$$

We describe this optimization procedure in detail in Sect. 4, but first we analyze the statistical properties of the estimator $(\widetilde{Pf})(\omega)$. The key quantity we bound is the mean squared error (MSE) $\mathbb{E}\left[\|Pf - \widetilde{Pf}\|_2^2\right]$. The following lemma establishes that when $\widetilde{g}_\eta, \widetilde{g}'_\eta$ are good approximations of g_η, g'_η, $\widetilde{Pf}$ is a good approximation of Pf, so we can reduce the problem to controlling $\widetilde{g}_\eta, \widetilde{g}'_\eta$.

Lemma 1 *Assume Model 3, $Pf \in \mathbf{C}^1(\mathbb{R})$, and the estimator $(\widetilde{Pf})(\omega)$ defined in (7). Then:*

$$\|Pf - \widetilde{Pf}\|_2^2 \lesssim \|g_\eta - \widetilde{g}_\eta\|_2^2 + \|\omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2\,.$$

Proof The proof of Lemma 1 is given in Appendix A. □

Lemma 1 thus establishes that to bound $\mathbb{E}\left[\|Pf - \widetilde{Pf}\|_2^2\right]$, it is sufficient to bound $\mathbb{E}\left[\|g_\eta - \widetilde{g}_\eta\|_2^2\right]$ and $\mathbb{E}\left[\|\omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2\right]$. Utilizing Lemma 1 yields the following Theorem, which bounds the MSE of (7) for dilation MRA. To control higher order terms we define:

$$(\overline{Pf})^k(\omega) := \max_{\xi\in[\omega/2,2\omega]}|(Pf)^k(\xi)|\,.$$

In general for well behaved functions, $(\overline{g})^k$ and g^k have the same decay rate; for example, if g^k is monotonic, $(\overline{g})^k(\omega) = g^k(2\omega)$.

Theorem 1 *Assume Model 3, the estimator* $(\widetilde{Pf})(\omega)$ *defined in* (7)*,* $Pf \in \mathbf{C}^3(\mathbb{R})$*, and that* $\omega^k(\overline{Pf})^{(k)}(\omega) \in \mathbf{L}^2(\mathbb{R})$ *for* $k = 2, 3$*. Then:*

$$\mathbb{E}\left[\|Pf - \widetilde{Pf}\|_2^2\right] \lesssim \frac{\eta^2}{M}\left(\|(Pf)(\omega)\|_2^2 + \|\omega(Pf)'(\omega)\|_2^2 + \|\omega^2(Pf)''(\omega)\|_2^2\right) + r\,,$$

where r *is a higher-order term satisfying*

$$r \leq \frac{\eta^4}{M}\left(\|\omega^2(\overline{Pf})''(\omega)\|_2^2 + \|\omega^3(\overline{Pf})'''(\omega)\|_2^2\right).$$

Proof By Lemma 1, it is sufficient to bound $\mathbb{E}\left[\|[\|]g_\eta - \widetilde{g}_\eta\|_2^2\right]$ and $\mathbb{E}\left[\|[\|]\omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2\right]$. Since $\widetilde{g}_\eta(\omega) = \frac{1}{M}\sum_{j=1}^M Pf_j(\omega)$, we have

$$(\widetilde{g}_\eta(\omega) - g_\eta(\omega))^2 \leq \left(\frac{1}{M}\sum_{j=1}^M (Pf_j)(\omega) - g_\eta(\omega)\right)^2.$$

Let $X_j = (Pf_j)(\omega) - g_\eta(\omega) = (Pf_j)(\omega) - \mathbb{E}\left[(Pf_j)(\omega)\right]$. Thus because $\frac{1}{M}\sum_{j=1}^M X_j$ is a centered random variable, we have

$$\mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^M X_j\right)^2\right] = \mathrm{var}\left[\frac{1}{M}\sum_{j=1}^M X_j\right] = \frac{\mathrm{var}(X_j)}{M}. \tag{8}$$

Note that we can write:

$$\begin{aligned} X_j &= (Pf_j)(\omega) - (Pf)(\omega) + (Pf)(\omega) - \mathbb{E}\left[(Pf_j)(\omega)\right] \\ X_j^2 &\leq 2\left((Pf_j)(\omega) - (Pf)(\omega)\right)^2 + 2\left((Pf)(\omega) - \mathbb{E}\left[(Pf_j)(\omega)\right]\right)^2. \end{aligned}$$

Since it is easy to check that

$$\mathbb{E}\left[\left((Pf)(\omega) - \mathbb{E}\left[(Pf_j)(\omega)\right]\right)^2\right] \leq \mathbb{E}\left[\left((Pf_j)(\omega) - (Pf)(\omega)\right)^2\right],$$

we obtain

$$\mathbb{E}\left[X_j^2\right] \leq 4\mathbb{E}\left[\left((Pf_j)(\omega) - (Pf)(\omega)\right)^2\right].$$

Taylor expanding $(Pf)((1-\tau_j)\omega)$ gives:

$$(Pf)((1-\tau_j)\omega) = (Pf)(\omega) + (Pf)'(\omega)\cdot\omega\tau_j \pm \frac{1}{2}(\overline{Pf})''(\omega)\cdot\omega^2\tau_j^2.$$

Multiplying by $(1-\tau_j)^2$ and rearranging:

$$\begin{aligned}&(1-\tau_j)^2(Pf)((1-\tau_j)\omega)-(Pf)(\omega)\\&\quad=(-2\tau_j+\tau_j^2)(Pf)(\omega)+(1-\tau_j)^2(Pf)'(\omega)\cdot\omega\tau_j\pm\frac{(1-\tau_j)^2}{2}(\overline{Pf})''(\omega)\cdot\omega^2\tau_j^2\,.\end{aligned}$$

Utilizing $a+b-c\le d\le a+b+c\implies d^2\lesssim a^2+b^2+c^2$, we square and take expectation to obtain

$$\mathbb{E}\left[\left((Pf_j)(\omega)-(Pf)(\omega)\right)^2\right]\lesssim[(Pf)(\omega)]^2\,\eta^2+\left[\omega(Pf)'(\omega)\right]^2\eta^2+\left[\omega^2(\overline{Pf})''(\omega)\right]^2\eta^4\,.$$

Thus

$$\mathrm{var}[X_j]=\mathbb{E}[X_j^2]\lesssim\left([(Pf)(\omega)]^2+\left[\omega(Pf)'(\omega)\right]^2\right)\eta^2+\left[\omega^2(\overline{Pf})''(\omega)\right]^2\eta^4\,.$$

Utilizing (8), we obtain

$$\mathbb{E}\left[(\widetilde{g}_\eta(\omega)-g_\eta(\omega))^2\right]\lesssim\frac{\eta^2}{M}\left([(Pf)(\omega)]^2+\left[\omega(Pf)'(\omega)\right]^2+\left[\omega^2(\overline{Pf})''(\omega)\right]^2\eta^2\right)$$

so that

$$\begin{aligned}\mathbb{E}\left[\|g_\eta-\widetilde{g}_\eta\|_2^2\right]&=\int\mathbb{E}\left[(\widetilde{g}_\eta(\omega)-g_\eta(\omega))^2\right]\,d\omega\\&\lesssim\frac{\eta^2}{M}\left(\|(Pf)(\omega)\|_2^2+\|\omega(Pf)'(\omega)\|_2^2+\|\omega^2(\overline{Pf})''(\omega)\|_2^2\,\eta^2\right)\,.\end{aligned}$$

We now bound $\mathbb{E}\left[\|[\|]\omega(g_\eta'(\omega)-\widetilde{g}_\eta'(\omega))_2^2\right]$. Letting $g_j=Pf_j$, we have

$$\omega\widetilde{g}_\eta'(\omega)-\omega g_\eta'(\omega)=\frac{1}{M}\sum_{j=1}^M\omega g_j'(\omega)-\omega g_\eta'(\omega)=\frac{1}{M}\sum_{j=1}^M Z_j$$

where

$$Z_j=\omega g_j'(\omega)-\omega g_\eta'(\omega)\,.$$

We note $\mathbb{E}[Z_j]=0$, and a similar argument as the one applied to X_j gives

$$\begin{aligned}Z_j^2&\le 2\left(\omega g_j'(\omega)-\omega g'(\omega)\right)^2+2\left(\omega g'(\omega)-\omega g_\eta'(\omega)\right)^2\\\mathbb{E}\left[Z_j^2\right]&\le 4\mathbb{E}\left[\left(\omega g_j'(\omega)-\omega g'(\omega)\right)^2\right]\,.\end{aligned}$$

Taylor expanding $(Pf)'((1-\tau_j)\omega)$ gives

$$(Pf)'((1-\tau_j)\omega) = (Pf)'(\omega) + (Pf)''(\omega)\cdot\omega\tau_j \pm \frac{1}{2}(\overline{Pf})'''(\omega)\cdot\omega^2\tau_j^2\,.$$

Since $\omega g_j'(\omega) = \omega(Pf_j)'(\omega) = (1-\tau_j)^3\omega(Pf)'((1-\tau_j)\omega)$, we multiply by $(1-\tau_j)^3\omega$ to obtain:

$$\begin{aligned}\omega(Pf_j)'(\omega) = (1-\tau_j)^3\omega(Pf)'(\omega) + \tau_j(1-\tau_j)^3\omega^2(Pf)''(\omega)\\ \pm\frac{1}{2}\tau_j^2(1-\tau_j)^3\omega^3(\overline{Pf})'''(\omega)\end{aligned}$$

Rearranging:

$$\begin{aligned}\omega(Pf_j)'(\omega) - \omega(Pf)'(\omega) = (-3\tau_j + 3\tau_j^2 - \tau_j^3)\omega(Pf)'(\omega)\\ + \tau_j(1-\tau_j)^3\omega^2(Pf)''(\omega) \pm \frac{1}{2}\tau_j^2(1-\tau_j)^3\omega^3(\overline{Pf})'''(\omega)\,.\end{aligned}$$

Squaring and taking expectation:

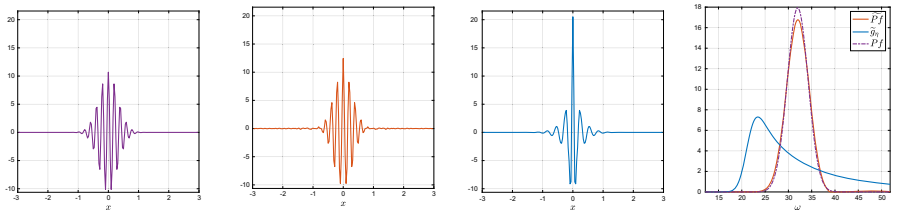
$$\begin{aligned}\mathbb{E}\left[\left(\omega g_j'(\omega) - \omega g'(\omega)\right)^2\right] \lesssim \left[\omega(Pf)'(\omega)\right]^2\eta^2\\ + \left[\omega^2(Pf)''(\omega)\right]^2\eta^2 + \left[\omega^3(\overline{Pf})'''(\omega)\right]^2\eta^4\,.\end{aligned}$$

Having bounded $\mathrm{var}\left[Z_j\right]$, an identical argument as the one used to control $\mathbb{E}\left[\|[\|]g_\eta - \widetilde{g}_\eta\|_2^2\right]$ gives

$$\begin{aligned}&\mathbb{E}\left[\|[\|]\omega g_\eta'(\omega) - \omega\widetilde{g}_\eta'(\omega)\|_2^2\right]\\ &\lesssim \frac{\eta^2}{M}\left(\|[\|]\omega(Pf)'(\omega)\|_2^2 + \|[\|]\omega^2(Pf)''(\omega)\|_2^2 + \|[\|]\omega^3(\overline{Pf})'''(\omega)\|_2^2\eta^2\right),\end{aligned}$$

which proves the Theorem. □

Figure 2d illustrates how much is gained from inversion unbiasing for a specific high frequency signal; the mean power spectrum under Model 3 is greatly perturbed due to large dilations, but $\widetilde{Pf}$ is still an accurate approximation of Pf. Although in general a signal is not uniquely defined by its power spectrum, if $\widehat{f}$ is real and positive as in Fig. 2, f can be approximated by taking the inverse Fourier Transform of $(\widetilde{Pf})^{\frac{1}{2}}$. Figure 2a–c illustrate how in this case inversion unbiasing yields a signal which accurately approximates the target.



(a) Target signal (b) Signal recovered via $\widetilde{P}f$ (c) Signal recovered via $\widetilde{g}_\eta$ (d) Power spectra

Fig. 2 Power spectrum estimation and signal recovery for high frequency Gabor signal $f_3(x) = C_3 \exp^{-5x^2} \cos(32x)$ under Model 3 with $\eta = 12^{-1/2}$ and $M = 100,000$. The mean power spectrum $\widetilde{g}_\eta$ is greatly perturbed from the target power spectrum Pf, but applying inversion unbiasing to $\widetilde{g}_\eta$ yields an approximation $\widetilde{P}f$ which is quite close to Pf (see Fig. 2d). Figure 2a shows the target signal, and Figs. 2b, c show the target signal approximations obtained by inverting $\widetilde{g}_\eta$, $\widetilde{P}f$; in 2b the relative $\mathbf{L}^2$ error is 9.5%, while in 2c it is 89%

3.2 Results for Noisy Dilation MRA

Solving noisy dilation MRA presents several additional challenges which are lacking in dilation MRA. First of all, since white noise has a flat power spectrum, i.e., $\mathbb{E}[(P\epsilon)(\omega)] = \sigma^2$ for all frequencies ω, the MSE can only be controled on a finite frequency interval. We thus restrict to a finite frequency interval Ω, and consider the MSE of an estimator $\widetilde{P}f$ over the finite interval, i.e., $\mathbb{E}\left[\|Pf - \widetilde{P}f\|^2_{\mathbf{L}^2(\Omega)}\right]$. We note the residual error from working on Ω decays to zero as $|\Omega| \to \infty$. In addition, in any numerical implementation one is always restricted to a finite frequency interval.

Another challenge is that one does not have direct access to $\widetilde{g}_\eta$; rather one only has access to

$$\frac{1}{M}\sum_{j=1}^{M} Py_j - \sigma^2 = \widetilde{g}_\eta + \widetilde{g}_\sigma \tag{9}$$

where

$$\widetilde{g}_\sigma := \frac{1}{M}\sum_{j=1}^{M} \widehat{f}_j\widehat{\epsilon}_j^* + \widehat{f}_j^*\widehat{\epsilon}_j + P\epsilon_j - \sigma^2 .$$

Although the compact support of the hidden signal guarantees the smoothness of $\widetilde{g}_\eta, \widetilde{g}_\sigma$ is not smooth due to the additive noise. To extend the unbiasing procedure of Sect. 3.1 to the additive noise context, it is thus necessary to smooth the noisy power spectra. We thus compute $(\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L$ where $\phi_L(\omega) = (2\pi L^2)^{-\frac{1}{2}} e^{-\frac{\omega^2}{2L^2}}$ is a Gaussian filter with width L, and then define the following estimator:

$$(\widetilde{Pf})(\omega) := (I - L_{B_0})^{-1} B_1 L_{B_2} \left[3(\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L(\omega) + \omega \left((\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L \right)'(\omega) \right], \tag{10}$$

where $\widetilde{g}_\eta + \widetilde{g}_\sigma$ are as in (9) and B_0, B_1, B_2 as in (5). As $M \to \infty$ and $L \to 0$, (10) is an unbiased estimator of Pf. To quantify how the error of the estimator depends on L, we need the following two lemmas.

Lemma 2 *Let $h \in \mathbf{L}^2(\mathbb{R})$ and assume $|\widehat{h}(\omega)|$ decays like $|\omega|^{-\alpha}$ for some integer $\alpha \geq 1$. Then for L small enough:*

$$\|h - h * \phi_L\|_2^2 \lesssim \|h\|_2^2 L^4 + L^{4 \wedge (2\alpha - 1)}.$$

Proof The proof of Lemma 2 is given in Appendix B. □

Lemma 3 *Let $xh(x) \in \mathbf{L}^2(\mathbb{R})$ and assume $|\widehat{(\cdot)h(\cdot)}(\omega)|$ decays at least like $|\omega|^{-\alpha}$ for some integer $\alpha \geq 1$. Then for L small enough:*

$$\|x(h - h * \phi_L)\|_2^2 \lesssim (L^3 \|h\|_2^2) \wedge (L^4 \|h'\|_2^2) + \|xh\|_2^2 L^4 + L^{4 \wedge (2\alpha - 1)}.$$

Proof The proof of Lemma 3 is given in Appendix C. □

We now state the main result of the article.

Theorem 2 *Assume Model 2, the estimator $(\widetilde{Pf})(\omega)$ defined in (10), $Pf \in \mathbf{C}^3(\mathbb{R})$, and that $\omega^k (\overline{Pf})^{(k)}(\omega) \in \mathbf{L}^2(\mathbb{R})$ for $k = 2, 3$. Then*

$$\mathbb{E}\left[\|Pf - \widetilde{Pf}\|_{\mathbf{L}^2(\Omega)}^2 \right] \lesssim C_{f,\Omega} \left(\frac{\eta^2}{M} + L^4 + \frac{\sigma^2 \vee \sigma^4}{L^2 M} \right).$$

Proof From Proposition 1 and a proof similar to Lemma 1

$$\|Pf - \widetilde{Pf}\|_{\mathbf{L}^2(\Omega)}^2 \lesssim \|g_\eta + \omega g_\eta'(\omega) - (\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L - \omega((\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L)'(\omega)\|_{\mathbf{L}^2(\Omega)}^2.$$

By the triangle inequality

$$\begin{aligned} \|Pf - \widetilde{Pf}\|_{\mathbf{L}^2(\Omega)}^2 &\lesssim \|g_\eta + \omega g_\eta'(\omega) - \widetilde{g}_\eta - \omega \widetilde{g}_\eta'(\omega)\|_2^2 \\ &\quad + \|\widetilde{g}_\eta + \omega \widetilde{g}_\eta'(\omega) - (\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L - \omega((\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L)'(\omega)\|_{\mathbf{L}^2(\Omega)}^2 \\ &:= (A) + (B). \end{aligned}$$

From the proof of Theorem 1,

$$\mathbb{E}[(A)] \lesssim \frac{\eta^2}{M} \left(\|(Pf)(\omega)\|_2^2 + \|\omega(Pf)'(\omega)\|_2^2 + \|\omega^2(Pf)''(\omega)\|_2^2 \right) + r,$$

where $r = C_f \eta^4 / M$ for a constant C_f depending on f. It remains to control (B). We have

$$\begin{aligned}(B) &\lesssim \|\widetilde{g}_\eta - \widetilde{g}_\eta * \phi_L\|_2^2 + \|\omega \widetilde{g}'_\eta - \omega(\widetilde{g}_\eta * \phi_L)'\|_2^2 \\ &\quad + \|\widetilde{g}_\sigma * \phi_L\|_{\mathbf{L}^2(\Omega)}^2 + \|\omega(\widetilde{g}_\sigma * \phi_L)'\|_{\mathbf{L}^2(\Omega)}^2 \\ &:= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)}\,.\end{aligned}$$

We control (I) with Lemma 2 and (II) with Lemma 3; we note in both cases α can be chosen arbitrarily large since the signals have compact support. By Lemma 2,

$$\text{(I)} = \|\widetilde{g}_\eta - \widetilde{g}_\eta * \phi_L\|_2^2 \lesssim L^4 \|\widetilde{g}_\eta\|_2^2 \lesssim L^4 \|Pf\|_2^2\,,$$

since $\|Pf_j\|_2 = (1-\tau_j)^{\frac{3}{2}} \|Pf\|_2 \leq (\frac{3}{2})^{\frac{3}{2}} \|Pf\|_2$. By Lemma 3,

$$\begin{aligned}\text{(II)} &= \|\omega \widetilde{g}'_\eta - \omega(\widetilde{g}_\eta * \phi_L)'\|_2^2 \\ &\lesssim L^4 \|\widetilde{g}''_\eta\|_2^2 + L^4 \|\omega \widetilde{g}'_\eta(\omega)\|_2^2 + L^4 \\ &\lesssim L^4 \left(\|(Pf)''\|_2^2 + \|\omega (Pf)'(\omega)\|_2^2 + 1 \right)\,.\end{aligned}$$

For (III), note that by Young's Convolution Inequality

$$\begin{aligned}\|\widetilde{g}_\sigma * \phi_L\|_{\mathbf{L}^2(\Omega)}^2 &\leq \|\phi_L\|_1^2 \cdot \|\widetilde{g}_\sigma\|_{\mathbf{L}^2(\Omega)}^2 \\ &= \|\widetilde{g}_\sigma\|_{\mathbf{L}^2(\Omega)}^2 \\ &\lesssim \|[\|\Big] \frac{1}{M} \sum_{j=1}^M \widehat{f_j} \widehat{\epsilon}_j^* \Big\|_2^2 + \|[\|\Big] \frac{1}{M} \sum_{j=1}^M P\epsilon_j - \sigma^2 \Big\|_{\mathbf{L}^2(\Omega)}^2\,.\end{aligned}$$

We have

$$\begin{aligned}\mathbb{E}\left[\|\frac{1}{M} \sum_{j=1}^M \widehat{f_j} \widehat{\epsilon}_j^* \|_2^2 \right] &= \int \mathbb{E}\left(\frac{1}{M} \sum_{j=1}^M \widehat{f_j}(\omega) \widehat{\epsilon}_j^*(\omega) \right)^2 d\omega \\ &\leq \int \frac{1}{M^2} \sum_{j=1}^M \widehat{f_j}(\omega)^2 \sigma^2 \, d\omega \\ &\lesssim \frac{\sigma^2}{M} \|f\|_2^2\,.\end{aligned}$$

Since $\mathbb{E}[P\epsilon_j] = \sigma^2$, $\mathbb{E}[(P\epsilon_j)^2] \leq 3\sigma^4$ (see Lemma D.1 in [32]), one has

$$\mathbb{E}\left(\frac{1}{M} \sum P\epsilon_j - \sigma^2 \right)^2 = \frac{\mathrm{var}(P\epsilon_j)}{M} \leq \frac{3\sigma^4}{M}\,,$$

which implies

$$\mathbb{E}\left[\|\frac{1}{M}\sum P\epsilon_j - \sigma^2\|^2_{\mathbf{L}^2(\Omega)}\right] \lesssim |\Omega|\frac{\sigma^4}{M}\,.$$

Thus

$$\mathbb{E}\,[(\mathrm{III})] \lesssim \frac{\sigma^2}{M}\left(\|f\|_2^2 + |\Omega|\sigma^2\right)\,.$$

For (IV), note that since $\|\phi_L'\|_1^2 \sim L^{-2}$,

$$\begin{aligned}\|\omega(\widetilde{g}_\sigma * \phi_L)'\|^2_{\mathbf{L}^2(\Omega)} &\le |\Omega|^2\,\|\widetilde{g}_\sigma * \phi_L'\|^2_{\mathbf{L}^2(\Omega)}\\ &\le |\Omega|^2\|\phi_L'\|_1^2\|\widetilde{g}_\sigma\|^2_{\mathbf{L}^2(\Omega)}\\ &\lesssim \frac{|\Omega|^2}{L^2}\|\widetilde{g}_\sigma\|^2_{\mathbf{L}^2(\Omega)}\,,\end{aligned}$$

so that utilizing our previous bound for $\mathbb{E}\left[\|\widetilde{g}_\sigma\|^2_{\mathbf{L}^2(\Omega)}\right]$ one obtains

$$\mathbb{E}\,[(\mathrm{IV})] \lesssim \frac{|\Omega|^2\sigma^2}{L^2 M}\left(\|f\|_2^2 + |\Omega|\sigma^2\right)\,.$$

Adding up the error terms:

$$\begin{aligned}\|Pf - \widetilde{Pf}\|^2_{\mathbf{L}^2(\Omega)} &\lesssim \frac{\eta^2}{M}\left(\|(Pf)(\omega)\|_2^2 + \|\omega(Pf)'(\omega)\|_2^2 + \|\omega^2(Pf)''(\omega)\|_2^2\right) + r\\ &\quad + L^4\left(\|Pf\|_2^2 + \|(Pf)''\|_2^2 + \|\omega(Pf)'(\omega)\|_2^2 + 1\right) + \frac{|\Omega|^2\sigma^2}{L^2M}\left(\|f\|_2^2 + |\Omega|\sigma^2\right)\\ &\lesssim C_{f,\Omega}\left(\frac{\eta^2}{M} + L^4 + \frac{\sigma^2\vee\sigma^4}{L^2M}\right)\,,\end{aligned}$$

which proves the theorem. □

To minimize the error upper bound in Theorem 2, we balance the last two terms, i.e., we choose L such that $L^4 \sim \frac{\sigma^2\vee\sigma^4}{L^2M}$. In the high noise regime where $\sigma \ge 1$, this gives $L \sim \left(\frac{\sigma^4}{M}\right)^{\frac{1}{6}}$, which yields the following important corollary.

Corollary 1 *Let the assumptions of Theorem 2 hold and in addition let* $\sigma \ge 1$ *and* $L = \left(\frac{\sigma^4}{M}\right)^{\frac{1}{6}}$. *Then:*

$$\mathbb{E}\left[\|Pf - \widetilde{Pf}\|^2_{\mathbf{L}^2(\Omega)}\right] \lesssim C_{f,\Omega}\left[\frac{\eta^2}{M} + \left(\frac{\sigma^4}{M}\right)^{\frac{2}{3}}\right]\,.$$

Thus to achieve an MSE bounded by δ requires $M = O\left(\frac{\eta^2}{\delta} + \frac{\sigma^4}{\delta^{3/2}}\right)$ samples.

Remark 3 In the classic MRA context (Model 1), one can estimate Pf simply by $\frac{1}{M}\sum_{i=1}^{M} Py_i - \sigma^2$, and the MSE for $\sigma \geq 1$ is $O(\frac{\sigma^4}{M})$, so that $M = O\left(\frac{\sigma^4}{\delta}\right)$ samples are required for MSE $\leq \delta$. The more restrictive sample size requirement in Corollary 1 is due to the fact that to unbias for dilations one must estimate not only g_η but also g'_η. Although g_η can be estimated with MSE $= O\left(\frac{\eta^2+\sigma^4}{M}\right)$, estimating g'_η requires first smoothing the noisy data via convolution, and this increases the MSE to $O\left(\frac{\eta^2}{M} + \left(\frac{\sigma^4}{M}\right)^{\frac{2}{3}}\right)$.

Remark 4 The inversion unbiasing procedure can also be directly applied to the wavelet-based features $(Sy)(\lambda) = \|y * \psi_\lambda\|_2^2$, where $\psi_\lambda(x) = \sqrt{\lambda}\psi(\lambda x)$ is a wavelet with frequency λ, proposed in [32] for Model 2. Because these features are smooth by design, no additional smoothing is necessary, and when $\sigma \geq 1$ this will yield an estimator $\widetilde{Sf}$ with error

$$\mathbb{E}\left[\|Sf - \widetilde{Sf}\|^2_{\mathbf{L}^2(\Omega)}\right] \lesssim C_{f,\Omega}\left[\frac{\eta^2}{M} + \frac{\sigma^4}{M}\right].$$

The additive noise convergence rate for the wavelet-based features is slightly better than the convergence rate for the power spectrum given in Corollary 1. A power spectrum estimator $\widetilde{Pf}$ can then be obtained from $\widetilde{Sf}$, since the wavelet-based features are defined by an invertible operator on the power spectrum. However, this inversion process is highly unstable numerically, as its accuracy is governed by the smallest eigenvalue of a low rank matrix. In practice, applying inversion unbiasing directly to the power spectrum yielded a lower error in our numerical experiments.

Remark 5 If the signals are only corrupted by dilations and additive noise, i.e., $y_j(x) = f((1-\tau_j)x) + \epsilon_j(x)$ for uniform τ_j, then inversion unbiasing can be applied directly in the spatial domain and Fourier invariants are unnecessary. Defining

$$\widetilde{f}(x) := (I - L_{B_0})^{-1} B_1 L_{B_2}\left[\widetilde{y} * \phi_L(x) + x\,(\widetilde{y} * \phi_L)'(x)\right]$$

where $\widetilde{y} = \frac{1}{M}\sum_{i=1}^{M} y_i$ and $L_{B_2} f(x) = B_2 f(B_2 x)$, a nearly identical analysis gives that for $\sigma \geq 1$ and optimal width L,

$$\mathbb{E}\left[\|f - \widetilde{f}\|^2_{\mathbf{L}^2(\mathbb{R})}\right] \lesssim C_f\left[\frac{\eta^2}{M} + \left(\frac{\sigma^2}{M}\right)^{\frac{2}{3}}\right].$$

Thus the advantages over Model 2 are two-fold: (1) there is an improved sample complexity with respect to the additive noise, since $M = O\left(\frac{\eta^2}{\delta} + \frac{\sigma^2}{\delta^{3/2}}\right)$ samples are required for MSE $\leq \delta$ and (2) the signal is fully recovered from inversion unbiasing and inverting Fourier invariants can be avoided.

4 Optimization

In practice, two subtleties arise in the computation of the estimator (10). First of all, there is no simple formula for applying the inverse operator $(I - L_{B_0})^{-1}$, but the estimator can be computed by solving an optimization problem as described in this section. Secondly, although our theoretical results apply to continuous signals, in practice one only obtains discrete vectors sampled from these continuous signals; Sect. 5 discusses the various errors arising from discretization.

To avoid inversion of the operator $I - L_{B_0}$, we compute (10) by solving a convex optimization problem. In the infinite sample limit, one has access to the perfect data term

$$d(\omega) = 3g_\eta(\omega) + \omega g'_\eta(\omega)\,,$$

and Proposition 1 guarantees that $g = Pf$ can be recovered from d by

$$g^\star = \operatorname{argmin}_{\mathring{g}\geq 0}\ \mathcal{L}(\mathring{g}) \quad , \quad \mathcal{L}(\mathring{g}) = \|[\|](I - L_{B_0})\mathring{g} - B_1 L_{B_2} d\|_2^2\,, \tag{11}$$

where the constants B_i depend on η. If η is known, the optimization (11) is a linearly constrained quadratic program, which is convex. It has a unique global minimum given by $g^\star = Pf$.

To compute the gradient of this loss function, $\nabla_{\mathring{g}}\mathcal{L}(\mathring{g})$, we calculate the Frechet derivative of $\mathcal{L}(\mathring{g})$. Let $A = I - L_{B_0}$; note

$$\mathcal{L}(\mathring{g}) = \|A\mathring{g} - B_1 L_{B_2} d\|_2^2 = N(A\mathring{g})\,,$$

where $Nf = \|f - B_1 L_{B_2} d\|_2^2$. Thus by the chain rule, the functional derivative at $\mathring{g}$ applied to a test function h is

$$(D\mathcal{L})(\mathring{g})h = (DN)(A\mathring{g}) \circ D(A\mathring{g})h = (DN)(A\mathring{g}) \circ Ah\,,$$

since A is a linear operator. To compute DN, note that

$$\frac{|N(f+h) - Nf - 2\langle f - B_1 L_{B_2} d, h\rangle|}{\|h\|_2} = \frac{\|h\|_2^2}{\|h\|_2} \to 0$$

as $\|h\|_2 \to 0$, so $(DN)(f)h = 2\langle f - B_1 L_{B_2} d, h\rangle$. Thus

$$\begin{aligned} (D\mathcal{L})(\mathring{g})h &= 2\langle A\mathring{g} - B_1 L_{B_2} d, Ah\rangle \\ &= \langle 2A^*(A\mathring{g} - B_1 L_{B_2} d), h\rangle \\ \implies \nabla\mathcal{L}(\mathring{g}) &= 2A^*(A\mathring{g} - B_1 L_{B_2} d)\,, \end{aligned}$$

where A^* is the adjoint of A. A straightforward calculation shows $A^*h(\omega) = h(\omega) - {B_0}^2\, h\left(\frac{\omega}{B_0}\right)$.

Remark 6 In order for $g^\star$ to be a valid power spectrum of a real-valued signal, it must also be symmetric, i.e., we must have $g^\star(\omega) = g^\star(-\omega)$. However since the operator $I - L_{B_0}$ acts pointwise and d is an even function, $g^\star$ is always an even function, and imposing an explicit symmetry constrain in the optimization is unnecessary.

Remark 7 Another approach to solving (11) is to define $\mathring{p} = \sqrt{\mathring{g}}$, optimize over $\mathring{p}$ to obtain the optimal $p^\star$, and then define $g^\star = (p^\star)^2$; such a procedure ensures $g^\star$ is nonnegative without constraining $\mathring{g}$ in the optimization. However, the resulting optimization over $\mathring{p}$ is nonconvex.

In practice the variation parameter η may be unknown and one only has access to the finite sample data term, so the relevant loss function is

$$\widetilde{\mathcal{L}}(\mathring{g}, \mathring{\eta}) := \|[\|\Big(I - L_{B_0(\mathring{\eta})}\Big)\, \mathring{g} - B_1(\mathring{\eta}) L_{B_2(\mathring{\eta})} \widetilde{d}\|_2^2 ,$$

for finite sample data term

$$\widetilde{d}(\omega) := 3(\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L(\omega) + \omega \left[(\widetilde{g}_\eta + \widetilde{g}_\sigma) * \phi_L'\right](\omega) .$$

Optimization of this loss function can be delicate as it is no longer convex. Fluctuations induced by the noise in the finite sample regime mean that $g^\star$ is not guaranteed to be non-negative without imposing the constraint $\mathring{g} \geq 0$. Furthermore, since $\widetilde{\mathcal{L}}(\mathring{g}, 0) = 0$ for any $\mathring{g}$, there is a large plateau defined by $\eta = 0$ where loss values are small even for $\mathring{g}$ very far from Pf. It thus becomes necessary to constrain η to be bounded away from 0, i.e., in practice we compute:

$$(g^\star, \eta^\star) = \operatorname{argmin}_{\mathring{g} \geq 0,\, \mathring{\eta} > \delta} \widetilde{\mathcal{L}}(\mathring{g}, \mathring{\eta}), \tag{12}$$

for $\delta > 0$. Section 6 describes specific implementation details.

5 Discretization

Although our theoretical results apply to continuous signals, in practice one only obtains discrete vectors sampled from these continuous signals. More specifically, if the noisy signals are supported on the compact interval $[-\frac{N}{2}, \frac{N}{2}]$ and are sampled at a rate $\Delta x = 1/2^\ell$, we obtain a discrete set of frequencies ω_i on the interval $\Omega = [-2^\ell \pi, 2^\ell \pi]$, sampled at rate $\Delta\omega = \frac{2\pi}{N}$. The optimization of Sect. 4 is thus carried out not on a continuous g, but on a vector $\mathbf{g}$ with entries $\mathbf{g}_i = g(\omega_i)$. The discretization of the signals introduces discretization errors which are not reflected in Theorem 2. Although we do not attempt a precise analysis of all discretization errors, in this section we outline what these sources are and how they impact the solution we recover. In particular, we analyze the error incurred when using a fourth order finite difference method to approximate derivatives and a cubic spline interpolation algorithm with not-a-knot boundary conditions to approximate dilation operators, as this reflects the procedure used in the simulations reported in Sect. 6.

To simplify the discussion, we consider the error incurred in the infinite sample limit, i.e., as $M \to \infty$, under a Fourier decay assumption. In particular, we assume there exist constants $C, \beta > 0$ such that $|\widehat{f}(\omega)| \leq C/(1+|\omega|^{\beta})$. Note if β is an integer, this implies $f \in \mathbf{C}^{\beta-2}(\mathbb{R})$, and if β is non-integer, that $f \in \mathbf{C}^{\lfloor \beta-1 \rfloor}(\mathbb{R})$ [38]. According to the continuum theory (Proposition 1), we know:

$$g = (I - L_{B_0})^{-1} B_1 L_{B_2}(3g_\eta(\omega) + \omega g'_\eta(\omega)) = (I - L_{B_0})^{-1} B_1 L_{B_2} d_c \,, \tag{13}$$

where $d_c = 3g_\eta(\omega) + \omega g'_\eta(\omega)$ is the continuum data term. Let $f_\eta(x) = \mathbb{E}_\tau[f((1-\tau)^{-1}x)]$ and $\mathbf{f}_\eta$ denote the discretization of f_η; note $\widehat{f_\eta}$ also decays like $|\omega|^{-\beta}$, since $(1-\tau)$ is contained in $[\frac{1}{2}, 2]$. In the infinite sample (but discrete resolution) limit, one has access to the power spectrum of $\mathbf{f}_\eta$ computed via a discrete Fourier transform (DFT), but not to $\mathbf{g}_\eta$, where $(\mathbf{g}_\eta)_i = (Pf_\eta)(\omega_i)$ is the discretization of the continuous power spectrum. However, if β is large, the DFT is a good approximation of the Fourier transform, and in particular:

$$\left\| \mathbf{g} - \frac{1}{2^{2\ell}} |\mathrm{DFT}(\mathbf{f}_\eta)|^2 \right\|_\infty \leq C_f \left(\frac{1}{2^\ell} \right)^\beta$$

(see Exercise 3.21 in [38]). The discretization errors impacting $g'_\eta(\omega)$ also involve finite difference approximations of derivatives. Since we use a fourth order finite difference quotient approximation (FDQ) for g'_η, if we had access to $\mathbf{g}_\eta$, this would incur an error bounded by $O(\frac{1}{N^4})$. However if these function values are observed with error ϵ, the derivative approximation will have error $O(N\epsilon + \frac{1}{N^4})$. Since we observe $\mathbf{g}_\eta$ with DFT error, our error in approximating $\mathbf{g}'_\eta$ becomes

$$\left\| \mathbf{g}'_\eta - \mathrm{FDQ}\left(\frac{1}{2^{2\ell}} |\mathrm{DFT}(\mathbf{f}_\eta)|^2 \right) \right\|_\infty \leq C_f \left(N \left(\frac{1}{2^\ell} \right)^\beta + \frac{1}{N^4} \right) .$$

Letting $\mathbf{d}_c$ denote the discretization of the continuum data term $3g_\eta(\omega) + \omega g'_\eta(\omega)$, and $\mathbf{d}_d$ the vector which is actually computed via DFT and FDQ, we thus have:

$$\|\mathbf{d}_c - \mathbf{d}_d\|_\infty \leq C_f 2^\ell \left(N \left(\frac{1}{2^\ell} \right)^\beta + \frac{1}{N^4} \right) ,$$

since $|\omega_i| \leq 2^\ell \pi$. Let $\mathbf{g}$ denote the discretization of $g = Pf = (I - L_{B_0})^{-1} B_1 L_{B_2} d_c$. We are not able to compute $\mathbf{g}$ exactly, but instead we compute

$$\mathbf{g}_d = (I - \mathbf{L}_{B_0})^{-1} B_1 \mathbf{L}_{B_2} \mathbf{d}_d \,, \tag{14}$$

where $\mathbf{L}_C$ is a grid-based approximation of the dilation operator L_C. Since we use a cubic spline interpolation algorithm with not-a-knot boundary conditions, the error

scales like a fourth power of the step size, i.e., as $O(\frac{1}{N^4})$ [19], so that the interpolation error on d_c is bounded by:

$$\|(\mathbf{L_C d_c}) - \mathbf{L}_C \mathbf{d}_c\|_\infty \leq C_f \frac{1}{N^4},$$

where $(\mathbf{L_C d_c})$ denotes the discretization of the continuous function $L_C d_c$. We can now bound $\|\mathbf{g} - \mathbf{g}_d\|_2$; note by (14):

$$\begin{aligned}\|\mathbf{g} - \mathbf{g}_d\|_2 &\leq \|(I - \mathbf{L}_{B_0})^{-1}\|_2 \cdot \|(I - \mathbf{L}_{B_0})(\mathbf{g} - \mathbf{g}_d)\|_2 \\ &= \|(I - \mathbf{L}_{B_0})^{-1}\|_2 \cdot \|(I - \mathbf{L}_{B_0})\mathbf{g} - B_1 \mathbf{L}_{B_2} \mathbf{d}_d)\|_2 .\end{aligned}$$

We thus have:

$$\begin{aligned}(I - \mathbf{L}_{B_0})\mathbf{g} &= \mathbf{g} - \mathbf{L}_{B_0}\mathbf{g} \\ &= \mathbf{g} - (\mathbf{L_{B_0} g}) + O(N^{-4}) && \text{(interpolation error)} \\ &= B_1(\mathbf{L_{B_2} d_c}) + O(N^{-4}) && \text{(by (13))} \\ &= B_1 \mathbf{L}_{B_2} \mathbf{d_c} + O(N^{-4}) && \text{(interpolation error)}\end{aligned}$$

so that $\|(I - \mathbf{L}_{B_0})\mathbf{g} - B_1 \mathbf{L}_{B_2} \mathbf{d_c}\|_\infty = O(N^{-4})$. We thus obtain:

$$\begin{aligned}&\|\mathbf{g} - \mathbf{g}_d\|_2 \\ &\leq \|(I - \mathbf{L}_{B_0})^{-1}\|_2 \cdot \left(\|(I - \mathbf{L}_{B_0})\mathbf{g} - B_1 \mathbf{L}_{B_2} \mathbf{d_c}\|_2 + \|B_1 \mathbf{L}_{B_2} \mathbf{d_c} - B_1 \mathbf{L}_{B_2} \mathbf{d}_d)\|_2\right) \\ &\leq \|(I - \mathbf{L}_{B_0})^{-1}\|_2 \cdot \left((N2^\ell)^{\frac{1}{2}} \|(I - \mathbf{L}_{B_0})\mathbf{g} - B_1 \mathbf{L}_{B_2} \mathbf{d_c}\|_\infty + B_1 \|\mathbf{L}_{B_2}\|_2 \cdot \|\mathbf{d_c} - \mathbf{d}_d)\|_2\right) \\ &\leq (N2^\ell)^{\frac{1}{2}} \|(I - \mathbf{L}_{B_0})^{-1}\|_2 \cdot \left(\|(I - \mathbf{L}_{B_0})\mathbf{g} - B_1 \mathbf{L}_{B_2} \mathbf{d_c}\|_\infty + B_1 \|\mathbf{L}_{B_2}\|_2 \cdot \|\mathbf{d_c} - \mathbf{d}_d)\|_\infty\right) .\end{aligned}$$

Since $\|\mathbf{L}_{B_2}\|_2$, $\|(I - \mathbf{L}_{B_0})^{-1}\|_2$ can be bounded independently of the grid size (but dependent on η), we get:

$$\|\mathbf{g} - \mathbf{g}_d\|_2 \lesssim (N2^\ell)^{\frac{1}{2}} \left(\underbrace{\frac{1}{N^4}}_{\text{Interpolation}} + \underbrace{N2^\ell \left(\frac{1}{2^\ell}\right)^\beta}_{\text{DFT}} + \underbrace{\frac{2^\ell}{N^4}}_{\text{FDQ}} \right) = O(h^{\beta-3} + h^2),$$

where we assume for simplicity that the spatial and frequency grid have equal resolution, i.e., $\frac{2\pi}{N} = \frac{1}{2^\ell} = h$. To compare with the error in Theorem 2, we want to report how well we can approximate the continuous g. Let $g_d = \text{interp}(\mathbf{g}_d)$ be a cubic Hermite spline of the approximated grid values $\mathbf{g}_d$, where derivatives at the grid points have first been approximated with a fourth order method. Decomposing $g = g_\Omega + g_{\Omega^c}$, where $g_\Omega(\omega) = g(\omega)\mathbf{1}(\omega \in \Omega)$,

$$\|g - g_d\|_2^2 \lesssim \|g_{\Omega^c}\|_2^2 + \|g_\Omega - \text{interp}(\mathbf{g})\|_2^2 + \|\text{interp}(\mathbf{g}) - \text{interp}(\mathbf{g}_d)\|_2^2 .$$

The fourier decay assumption guarantees $\|g_{\Omega^C}\|_2^2 \lesssim h^{2\beta-1}$. Furthermore, $|g_\Omega - \text{interp}(\mathbf{g})| \leq O(h^4)$, giving $\|g_\Omega - \text{interp}(\mathbf{g})\|_2^2 \leq O(h^7)$. Finally, a straightforward calculation gives

$$\|\text{interp}(\mathbf{g}) - \text{interp}(\mathbf{g}_d)\|_2^2 \lesssim \|\mathbf{g} - \mathbf{g}_d\|_2^2 \cdot h = O(h^{2\beta-5} + h^5)\,.$$

We thus obtain:

$$\|g - g_d\|_2^2 = O(h^{2\beta-5} + h^5)\,.$$

Again, assuming that $\beta \geq 5$, we obtain an approximation g_d (from solving the discrete problem) of the continuum g with squared error $\|g - g_d\|_2^2 = O(h^5)$. As the grid is taken finer both in space and frequency, the solution recovered via discretization thus converges to the continuum solution. Although we omit a detailed analysis for the finite M case, we expect that the squared error will be the sum of this $O(h^5)$ discretization error and the sampling error reported in Theorem 2.

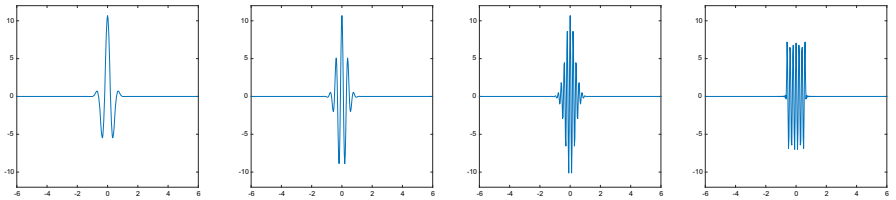
Remark 8 If the signal is sufficiently smooth (more specifically, if $\beta \geq 5$), then $\|\mathbf{g} - \mathbf{g}_d\|_2 = O(h^2)$ which implies that on the grid points one obtains an entry-wise root mean squared discretization error of $O(h^3)$.

Remark 9 In our simulations, we actually compute the error based on the DFT of the discretized true sample, as computing the continuum power spectrum for some of the signals is cumbersome. Thus the error decay in Figs. 4 and 6 includes discretization errors due to finite difference and interpolation, but not the DFT error. However for most of our signals we expect β is large and the discretization error is dominated by the interpolation and finite difference error.

6 Simulation Results

In this section we investigate the proposed inversion unbiasing procedure on the following collection of synthetic signals which capture a variety of features:

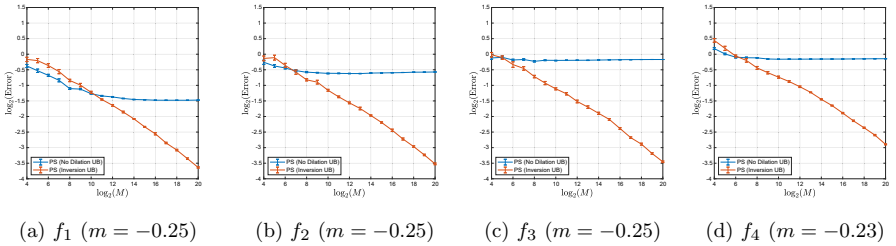
$$\begin{aligned}
f_1(x) &= C_1 \exp^{-5x^2} \cos(8x)\\
f_2(x) &= C_2 \exp^{-5x^2} \cos(16x)\\
f_3(x) &= C_3 \exp^{-5x^2} \cos(32x)\\
\widehat{f_4}(\omega) &= C_4 \left[\text{sinc}(0.2(\omega - 32)) + \text{sinc}(0.2(-\omega - 32))\right]\\
f_5(x) &= C_5 \exp^{-0.04x^2} \cos(30x + 1.5x^2)\\
\widehat{f_6}(\omega) &= C_6 \left[\mathbf{1}(\omega \in [-38, -32]) + \mathbf{1}(\omega \in [32, 38])\right]\\
\widehat{f_7}(\omega) &= C_7 \left[\text{zigzag}\,(0.2(\omega + 40)) + \text{zigzag}\,(0.2(\omega + 40))\right]^{1/2}\\
f_8(x) &= 0.
\end{aligned}$$



(a) f_1 (b) f_2 (c) f_3 (d) f_4

(e) f_5 (f) f_6 (g) f_7 (h) f_8

Fig. 3 The signals discussed in Sect. 6

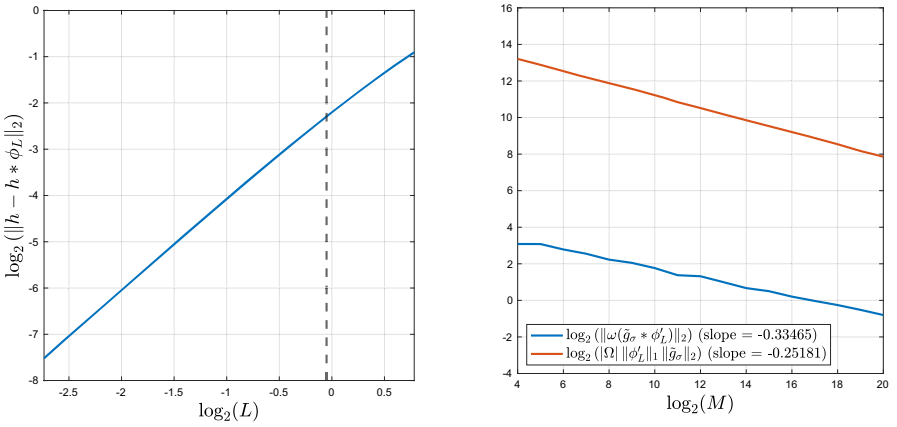


(a) f_1 $(m = -0.25)$ (b) f_2 $(m = -0.25)$ (c) f_3 $(m = -0.25)$ (d) f_4 $(m = -0.23)$

(e) f_5 $(m = -0.22)$ (f) f_6 $(m = -0.11)$ (g) f_7 $(m = -0.24)$ (h) f_8 $(m = -0.23)$

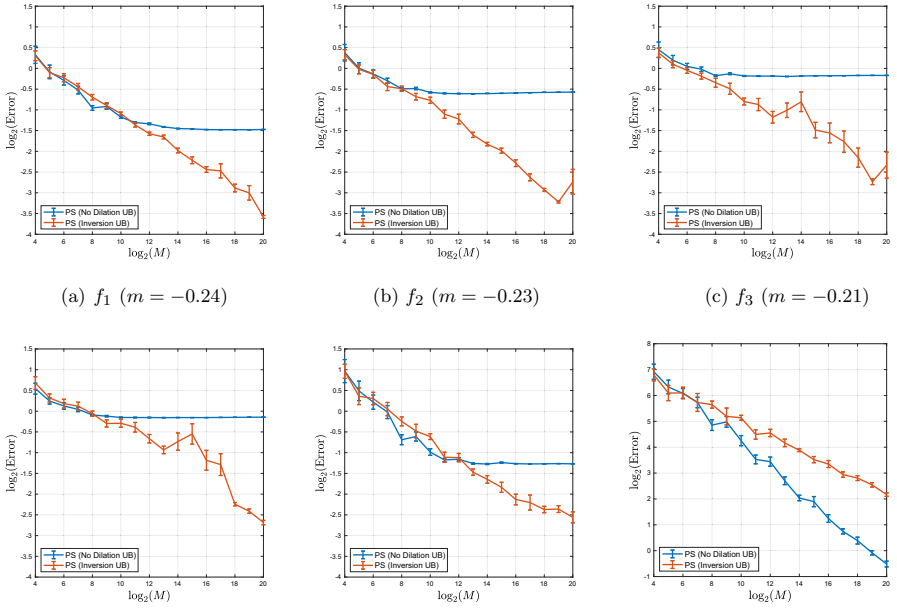
Fig. 4 Error decay with standard error bars for Model 2 (oracle moment estimation). All plots show relative $\mathbf{L}^2$ error and have the same axis limits, except Fig. 4h, which shows absolute error; m reports the slope of linear regression on the right half of the plot, i.e., for $12 \le \log_2(M) \le 20$

The hidden signals were defined on $[-\frac{N}{4}, \frac{N}{4}]$ and the corresponding noisy signals on $[-\frac{N}{2}, \frac{N}{2}]$. The signals were sampled at rate $1/2^\ell$, resolving frequencies in the interval $[-2^\ell\pi, 2^\ell\pi]$; $N = 2^5$ and $\ell = 5$ were used for all simulations. As indicated above, f_4, f_6, f_7 were sampled directly in the frequency domain, while the rest were sampled in the spatial domain. The normalization constants C_i were chosen so that all signals would have the same SNR for a fixed additive noise level, specifically $(\mathrm{SNR})^{-1} = \sigma^2$, where $\mathrm{SNR} = \left(\frac{1}{N}\int_{-N/2}^{N/2} f(x)^2\,dx\right)/\sigma^2$.



(a) Smoothing decay rate (b) Young's Inequality

Fig. 5 Plots explaining the small discrepancy between theoretical and empirical convergence rates. **a** The right side of the dashed line shows the L values corresponding to $12 \leq \log_2(M) \leq 20$, i.e., the upper range of values used in our simulations. In the simulation regime, the slope in the log-log plot is 1.65; however for small L (left side of dashed line), the slope is 1.96, which closely matches the L^2 rate given in Lemma 3. **b** the additive noise term exhibits a decay rate of -0.25 in the range of M values used for our simulations, while the upper bound due to Young's Inequality decays at the faster rate of -0.33



(a) f_1 $(m = -0.24)$ (b) f_2 $(m = -0.23)$ (c) f_3 $(m = -0.21)$

(d) f_4 $(m = -0.27)$ (e) f_5 $(m = -0.17)$ (f) f_8 $(m = -0.29)$

Fig. 6 Error decay with standard error bars for Model 2 (empirical moment estimation). All plots show relative $\mathbf{L}^2$ error and have the same axis limits, except Fig. 6h, which shows absolute error; m reports the slope of linear regression on the right half of the plot, i.e., for $12 \leq \log_2(M) \leq 20$

See Fig. 3 for plots of f_1 through f_8. The Gabors f_1, f_2, f_3 are smooth with a fast decay in both space and frequency; f_4 is discontinuous in space, with a smooth but slowly decaying FT; f_5 is a linear chirp with a non-constant instantaneous frequency; f_6 is discontinuous in frequency; f_7 is continuous but not smooth in frequency. The zero signal was included to investigate the effect of the inversion unbiasing procedure when applied directly to additive noise, i.e., in the absence of any signal. We investigate the ability of inversion unbiasing to solve Model 2 in the challenging regime of both low SNR and large dilations. Specifically we choose SNR $= \frac{1}{2}$ and τ uniform on $[-\frac{1}{2}, \frac{1}{2}]$ (thus $\sigma = \sqrt{2}$ and $\eta = 12^{-1/2} \approx 0.2887$). For comparison, the simulations in [32] were restricted to $\eta \leq 0.12$.

We first assume oracle knowledge of the additive noise and dilation variances σ^2, η^2. We let M increase exponentially from 16 to 1, 048, 576, and for each value of M we run 10 simulations of Model 2 and let $(\mathbf{Pf})_d$ be the discrete implementation of the estimator $\widetilde{Pf}$ given in (10), i.e., $(\mathbf{Pf})_d$ is computed via optimization on a discrete grid as described in Sects. 4 and 5. The width of the Gaussian filter L is chosen as in Corollary 1, and the optimization was implemented using Matlab's fminunc function with a Quasi–Newton algorithm. For each simulation, the relative error of the resulting power spectrum estimator is computed as

$$\text{Error} := \frac{\|\mathbf{Pf} - (\mathbf{Pf})_d\|_2}{\|\mathbf{Pf}\|_2},$$

and the mean error is then computed across simulations. Figure 4 shows the decay of the mean error as the sample size M increases. All signals exhibit a linear error decay in the log–log plots; as the error decay does not plateau, the simulations support that $\widetilde{Pf}$ is an unbiased estimator of Pf as shown in Theorem 2 and Corollary 1.

More specifically, for signals with a smooth power spectrum ($f_1, \ldots, f_5, f_8$), when the discretization error is negligible, Corollary 1 predicts that the error should decay like $M^{-1/3}$, i.e., we would expect to observe a slope of $-1/3$ in the log-log plots. In practice the error decay is slightly slower, with a slope of about $-1/4$ for the smooth signals. There are a few possible reasons for the small mismatch between the theory and simulations. First of all, Lemmas 2 and 3 are based on Taylor expansions about $L = 0$, and so the decay rates in terms of L are only sharp for L small enough; the decay rate is slightly worse in the range of L values used in our simulations; see Fig. 5a. In practice when the continuous theory is implemented on a computer, one can never take L smaller than the discrete frequency resolution. Secondly, the proof of Theorem 2 applies Young's Convolution Inequality to control the additive noise terms, but simulations indicate that the actual decay rate of the additive noise terms is smaller than this upper bound for the simulation range of M values. See Fig. 5b; as $M \to \infty$, the decay rates do converge. Thirdly, there are discretization errors impacting the simulation results which are not accounted for in Corollary 1.

For the non-smooth signals, recall that f_7 has a power spectrum which is continuous but not differentiable while f_6 has a discontinuous power spectrum. The decay rate of f_7 matches that of the smooth signals, but $Pf_7 \notin \mathbf{C}^1(\mathbb{R})$, indicating that perhaps $Pf \in \mathbf{C}^3(\mathbb{R})$ is not required to achieve the rate in Theorem 2 but an artifact of the proof technique. We note the infinite sample result (Proposition 1) holds under the

much milder assumption $Pf \in \mathbf{C}^0(\mathbb{R})$. In practice, the decay rate seems to be driven by the α appearing in Lemma 2; for f_6, Lemma 2 would apply with $\alpha = 1$ to give an error decay like $\sqrt{L}$ and a predicted slope of $-1/12 = -0.083$; we observe -0.1071 in Fig. 4f.

We next investigate the ability of inversion unbiasing to solve Model 2 without oracle knowledge of the variances σ^2, η^2. The additive noise level can be reliably estimated from the mean vertical shift in the tail of the mean power spectrum. In particular, for $\Sigma = [-2^\ell\pi, 2^\ell\pi]\backslash[-2^{\ell-1}\pi, 2^{\ell-1}\pi]$, we define

$$\widetilde{\sigma}^2 := \frac{1}{|\{\omega_i \in \Sigma\}|} \sum_{\omega_i \in \Sigma} |\widehat{y}(\omega_i)|^2 .$$

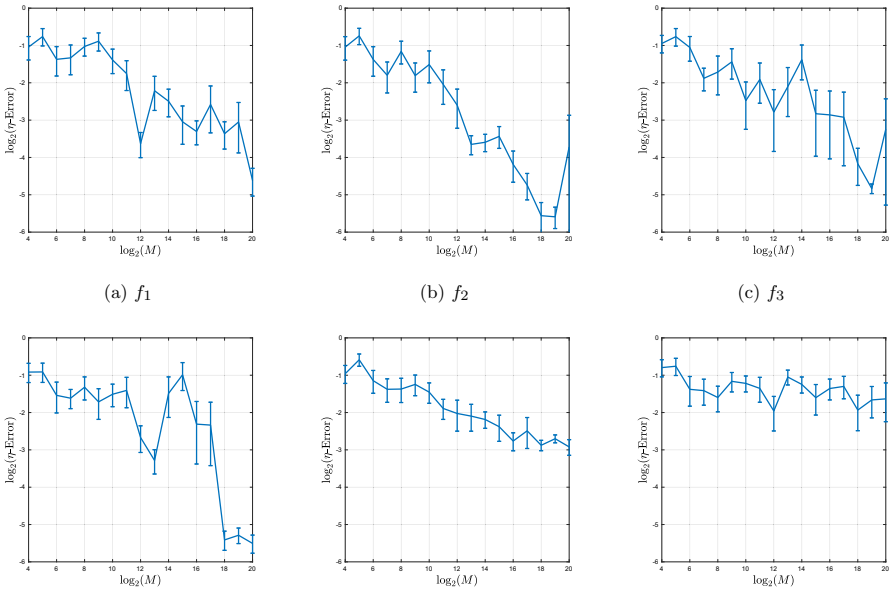
Assuming the frequency box is chosen large enough so that the support of $\widehat{f}$ is contained in $[-2^{\ell-1}\pi, 2^{\ell-1}\pi]$, this an unbiased estimator of σ^2 (see Lemma D.1 in [32]), with error scaling like $\frac{\sigma^2}{\sqrt{M}}$.

Estimating η is more complex and we implement a joint optimization procedure to simultaneously learn η and Pf. The optimization to learn η must be constrained since $\eta = 0$ minimizes the loss function; η is thus constrained to lie in the interval $[0.05, 0.40]$ and we initialize η on a coarse grid ranging from 0.10 to 0.35. For each initialization, the learned η value is recorded; a set of candidate η values is obtained by discarding learned η values which are close to the boundary, and η is then selected as the candidate value with the smallest loss. Note the constrained optimization was implemented using Matlab's fmincon function with an Interior-point algorithm. Figure 7 shows the mean relative error decay for the empirical η estimation for the signals with smooth power spectra. Figure 6 shows the relative error decay for the mean power spectrum; error decay is similar to the oracle case but more variable. Note η cannot be reliably learned with this gradient descent procedure when the power spectrum is not smooth, so f_6 and f_7 are omitted from Figs. 6 and 7.

Remark 10 For the zero signal f_8, the power spectrum estimator with inversion unbiasing is worse than the power spectrum estimator without dilation unbiasing (see Figs. 4h, 6h). Indeed, in this case smoothing the empirical power spectrum is unnecessary and reduces the theoretical convergence rate from $M^{-1/2}$ to $M^{-1/3}$. Note for f_8 recovery of η is impossible; see Fig. 7f.

7 Conclusion

This article considers a generalization of MRA which includes random dilations in addition to random translations and additive noise. The proposed method has several desirable properties compared with previous work. The bias due to dilations is eliminated (not just reduced as in [32]) as the sample size increases. In addition, the method is numerically stable, as the unbiasing procedure operates directly on the power spectrum, rather than features derived from the power spectrum.



(a) f_1 (b) f_2 (c) f_3

(d) f_4 (e) f_5 (f) f_8

Fig. 7 Error decay with standard error bars for empirical estimation of η. All plots show relative error and have the same axis limits

Many compelling directions remain for future research, as the present work only recovers the power spectrum of the hidden signal. There are two possible approaches for full signal recovery building on the current work: (1) apply inversion unbiasing on the power spectrum and then learn the phase via phase retrieval methods or (2) extend inversion unbiasing to operate directly on the bispectrum and then invert the bispectrum. The second approach is the most promising and most general, as there is a well-established history of solving MRA problems via the bispectrum [6, 8, 21], and phase retrieval is only possible in limited circumstances.

Another natural direction is to consider nonuniform dilation distributions. Indeed, preliminary work suggests that inversion unbiasing can be extended to a broad class of dilation distributions as long as their underlying density functions are known; the general procedure is to invert a density-dependent but computable operator. Thus innovative methods for robustly learning the dilation distribution are critical for these methods to become competitive for real world applications. Finally, extensions to 2-dimensional signals and other additive noise models are also of interest.

Acknowledgements AL thanks NSF DMS 2309570 and NSF DMS 2136198.

Appendix A Proof of Lemma 1

Proof From Proposition 1 and (7)

$$Pf - \widetilde{Pf} = (I - L_{B_0})^{-1} B_1 L_{B_2} \left[3(g_\eta - \widetilde{g}_\eta) + \omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega)) \right],$$

for constants B_0, B_1, B_2 defined in (5). Letting $\|\cdot\|$ denote the spectral norm, we thus obtain:

$$\begin{aligned}\|Pf - \widetilde{Pf}\|_2^2 &\le {B_1}^2 \|(I - L_{B_0})^{-1}\|^2 \|L_{B_2}\|^2 \times \|3(g_\eta - \widetilde{g}_\eta) + \omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2 \\ &\le 2{B_1}^2 \|(I - L_{B_0})^{-1}\|^2 \|L_{B_2}\|^2 \times \left(9\|g_\eta - \widetilde{g}_\eta\|_2^2 + \|\omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2 \right).\end{aligned}$$

We first observe that $\|L_C^i\| = C^{\frac{5i}{2}}$ since

$$\begin{aligned}\|L_C^i g\|_2^2 &= \int (C^{3i} g(C^i \omega))^2 \, d\omega \\ &= \int C^{6i} g(\tilde{\omega})^2 \, \frac{d\tilde{\omega}}{C^i} \quad \text{for} \quad \tilde{\omega} = C^i \omega \\ &= C^{5i} \|g\|_2^2 .\end{aligned}$$

Thus

$$\begin{aligned}\|(I - L_{B_0})^{-1}\| = \|[\|] \sum_{i=0}^{\infty} L_{B_0}^i \le \sum_{i=0}^{\infty} {B_0}^{\frac{5i}{2}} = \frac{1}{1 - {B_0}^{\frac{5}{2}}} = O(\eta^{-1}) \\ \|L_{B_2}\| = {B_2}^{\frac{5}{2}} = O(1) \\ B_1 = O(\eta)\end{aligned}$$

so that

$$2{B_1}^2 \|(I - L_{B_0})^{-1}\|^2 \|L_{B_2}\|^2 = O(1) O(\eta^2) O(\eta^{-2}) = O(1)$$

and we obtain

$$\|Pf - \widetilde{Pf}\|_2^2 \lesssim \|g_\eta - \widetilde{g}_\eta\|_2^2 + \|\omega(g'_\eta(\omega) - \widetilde{g}'_\eta(\omega))\|_2^2 .$$

□

Appendix B Proof of Lemma 2

Proof Note by assumption there exist constants $C > 0$, $\omega_0 \ge 1$ such that $|\widehat{h}(\omega)| \le C|\omega|^{-\alpha}$ for $|\omega| \ge \omega_0$. Also note that $\widehat{\phi}_L(\omega) = e^{-L^2\omega^2/2}$, so that $1 - \widehat{\phi}_L(\omega) = \frac{L^2\omega^2}{2} + O(L^3)$ for small L. We have:

$$\begin{aligned}\|h - h * \phi_L\|_2^2 &= (2\pi)^{-1} \|\widehat{h}(1 - \widehat{\phi}_L)\|_2^2 \\ &= \frac{1}{2\pi} \int_{|\omega| < \omega_0} |\widehat{h}(\omega)|^2 |1 - \widehat{\phi}_L(\omega)|^2 \, d\omega\end{aligned}$$

$$+\frac{1}{2\pi}\int_{|\omega|\geq\omega_0} C^2|\omega|^{-2\alpha}|1-\widehat{\phi}_L(\omega)|^2\,d\omega$$
$$:= (I)+(II)\,.$$

Note:

$$\begin{aligned}(I) &\leq \int_{|\omega|<\omega_0} |\widehat{h}(\omega)|^2|1-\widehat{\phi}_L(\omega)|^2\,d\omega\\ &\leq 2\int_0^{\omega_0}|\widehat{h}(\omega)|^2\left(\frac{L^2\omega^2}{2}+O(L^3)\right)^2 d\omega\\ &\leq 2\left(\frac{L^4\omega_0^4}{4}+O(L^5)\right)\int_0^{\omega_0}|\widehat{h}(\omega)|^2\,d\omega\\ &\leq \frac{\omega_0^4}{2}\|h\|_2^2L^4+O(L^5)\,.\end{aligned}$$

To control the second term, note

$$\begin{aligned}(II) &\leq 2C^2\int_1^\infty \omega^{-2\alpha}\left(1-e^{-\frac{L^2\omega^2}{2}}\right)^2 d\omega\\ &= 2C^2\int_L^\infty\left(\frac{L}{\tilde{\omega}}\right)^{2\alpha}\left(1-e^{-\frac{\tilde{\omega}^2}{2}}\right)^2\frac{d\tilde{\omega}}{L}\\ &= 2C^2L^{2\alpha-1}\int_L^\infty\omega^{-2\alpha}\left(1-e^{-\frac{\omega^2}{2}}\right)^2 d\omega\,.\end{aligned}$$

Explicit evaluation of the upper bound with a computer algebra system gives:

$$\begin{aligned}\alpha=1:&\quad C_1L+O(L^4)\\ \alpha=2:&\quad C_2L^3+O(L^4)\\ \alpha=3:&\quad C_3L^4+O(L^5)\end{aligned}$$

Also since

$$\begin{aligned}&\frac{d}{d\alpha}\int_1^\infty\omega^{-2\alpha}\left(1-e^{-\frac{L^2\omega^2}{2}}\right)^2 d\omega\\ &\quad=\int_1^\infty -2\ln(\omega)\omega^{-2\alpha}\left(1-e^{-\frac{L^2\omega^2}{2}}\right)^2 d\omega<0\,,\end{aligned}$$

the upper bound is decreasing in α, and we can conclude $(II)\lesssim L^{4\wedge(2\alpha-1)}$ and the lemma is proved. □

Appendix C Proof of Lemma 3

Proof First observe:

$$\begin{aligned}\|x(h - h * \phi_L)\|_2^2 &= (2\pi)^{-1}\|[\|\Big]\frac{d}{d\omega}\left(\widehat{h} - \widehat{h}\widehat{\phi}_L\right)_2^2 \\ &= (2\pi)^{-1}\|\widehat{h'} - \widehat{h'}\widehat{\phi}_L - \widehat{h}\widehat{\phi}_L'\|_2^2 \\ &\lesssim \|\widehat{h'} - \widehat{h'}\widehat{\phi}_L\|_2^2 + \|\widehat{h}\widehat{\phi}_L'\|_2^2 .\end{aligned}$$

To bound the first term, we apply Lemma 2 to the function xh to obtain,

$$\begin{aligned}\|\widehat{h'} - \widehat{h'}\widehat{\phi}_L\|_2^2 &= 2\pi\|xh - (xh) * \phi_L\|_2^2 \\ &\lesssim \|xh\|_2^2 L^4 + L^{4\wedge(2\alpha-1)} .\end{aligned}$$

To bound the second term, note $\widehat{\phi}_L'(\omega) = -L^2\omega e^{-L^2\omega^2/2}$, and that $\|\omega^2 e^{-L^2\omega^2}\|_\infty = (eL)^{-1}$. Thus

$$\|\widehat{h}\widehat{\phi}_L'\|_2^2 = L^4\int |\widehat{h}(\omega)|^2\omega^2 e^{-L^2\omega^2}\, d\omega \le L^3\|h\|_2^2 .$$

Note we could get a higher power for L by

$$\|\widehat{h}\widehat{\phi}_L'\|_2^2 \le L^4\int \omega^2|\widehat{h}(\omega)|^2\, d\omega = L^4\|\omega\widehat{h}\|_2^2 \lesssim L^4\|h'\|_2^2 ,$$

which proves the lemma. □

References

1. Abbe, E., Bendory, T., Leeb, W., Pereira, J.M., Sharon, N., Singer, A.: Multireference alignment is easier with an aperiodic translation distribution. IEEE Trans. Inf. Theory **65**(6), 3565–3584 (2018)
2. Bandeira, A.S., Boumal, N., Voroninski, V.: On the low-rank approach for semidefinite programs arising in synchronization and community detection. In: Proceedings of the Conference on Learning Theory, pp. 361–382 (2016)
3. Bandeira, A.S., Charikar, M., Singer, A., Zhu, A.: Multireference alignment using semidefinite programming. In: Proceedings of the 5th Conference on Innovations in Theoretical Computer Science, pp. 459–470 (2014)
4. Bandeira, A.S., Boumal, N., Singer, A.: Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. Math. Program. **163**(1–2), 145–167 (2017)
5. Bandeira, A.S., Chen, Y., Lederman, R.R., Singer, A.: Non-unique games over compact groups and orientation estimation in cryo-em. Inverse Probl. **36**(6), 064002 (2020)
6. Bandeira, A.S., Niles-Weed, J., Rigollet, P.: Optimal rates of estimation for multi-reference alignment. Math. Stat. Learn. **2**(1), 25–75 (2020)
7. Bendory, T., Edidin, D.: Toward a mathematical theory of the crystallographic phase retrieval problem. SIAM J. Math. Data Sci. **2**(3), 809–839 (2020)
8. Bendory, T., Boumal, N., Ma, C., Zhao, Z., Singer, A.: Bispectrum inversion with application to multireference alignment. IEEE Trans. Signal Process. **66**(4), 1037–1050 (2017)
9. Bendory, T., Beinert, R., Eldar, Y.C.: Fourier phase retrieval: uniqueness and algorithms. In: Boche, H., Calderbank, R., Kutyniok, G., Vybiral, J. (eds.) Compressed Sensing and Its Applications, pp. 55–91. Birkhäuser, Cham (2017)
10. Boumal, N.: Nonconvex phase synchronization. SIAM J. Optim. **26**(4), 2355–2377 (2016)

11. Bowman, G.D., Poirier, M.G.: Post-translational modifications of histones that influence nucleosome dynamics. Chem. Rev. **115**(6), 2274–2295 (2015)
12. Brown, L.G.: A survey of image registration techniques. ACM Comput. Surv. (CSUR) **24**(4), 325–376 (1992)
13. Capodiferro, L., Cusani, R., Jacovitti, G., Vascotto, M.: A correlation based technique for shift, scale, and rotation independent object identification. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 12, pp. 221–224 (1987)
14. Carroll, R.J., Hall, P.: Low order approximations in deconvolution and regression with errors in variables. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **66**(1), 31–46 (2004)
15. Chandran, V., Elgar, S.L.: Position, rotation, and scale invariant recognition of images using higher-order spectra. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 213–216 (1992)
16. Chen, Y., Guibas, L.J., Huang, Q.X.: Near-optimal joint object matching via convex relaxation. In: Proceedings of the 31st International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 32, pp. 100–108 (2014)
17. Chen, Y., Candès, E.J.: The projected power method: an efficient algorithm for joint alignment from pairwise differences. Commun. Pure Appl. Math. **71**(8), 1648–1714 (2018)
18. Chen, H., Zehni, M., Zhao, Z.: A spectral method for stable bispectrum inversion with application to multireference alignment. IEEE Signal Process. Lett. **25**(7), 911–915 (2018)
19. Cheney, E.W., Kincaid, D.R.: Numerical Mathematics and Computing. Brooks/Cole Cengage Learning, Boston (2012)
20. Cheng, C., Jiang, J., Sun, Q.: Phaseless sampling and reconstruction of real-valued signals in shift-invariant spaces. J. Fourier Anal. Appl. **25**, 1361–1394 (2019)
21. Collis, W., White, P., Hammond, J.: Higher-order spectra: the bispectrum and trispectrum. Mech. Syst. Signal Process. **12**(3), 375–394 (1998)
22. Delaigle, A.: An alternative view of the deconvolution problem. Stat. Sin. **18**(3), 1025–1045 (2008)
23. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc.: Ser. B (Methodol.) **39**(1), 1–22 (1977)
24. DesJarlais, R., Tummino, P.J.: Role of histone-modifying enzymes and their complexes in regulation of chromatin biology. Biochemistry **55**(11), 1584–1599 (2016)
25. Diamond, R.: On the multiple simultaneous superposition of molecular structures by rigid body transformations. Protein Sci. **1**(10), 1279–1287 (1992)
26. Ekman, D., Björklund, Å.K., Frey-Skött, J., Elofsson, A.: Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. J. Mol. Biol. **348**(1), 231–243 (2005)
27. Forneris, F., Wu, J., Gros, P.: The modular serine proteases of the complement cascade. Curr. Opin. Struct. Biol. **22**(3), 333–341 (2012)
28. Foroosh, H., Zerubia, J.B., Berthod, M.: Extension of phase correlation to subpixel registration. IEEE Trans. Image Process. **11**(3), 188–200 (2002)
29. Giannakis, G.B.: Signal reconstruction from multiple correlations: frequency-and time-domain approaches. J. Opt. Soc. Am. A **6**(5), 682–697 (1989)
30. Gil-Pita, R., Rosa-Zurera, M., Jarabo-Amores, P., López-Ferreras, F.: Using multilayer perceptrons to align high range resolution radar signals. In: Proceedings of 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications, pp. 911–916 (2005)
31. Hansen, L.P.: Large sample properties of generalized method of moments estimators. Econometrica **50**(4), 1029–1054 (1982)
32. Hirn, M., Little, A.: Wavelet invariants for statistically robust multi-reference alignment. Inf. Inference J. IMA **10**(4), 1287–1351 (2021)
33. Hotta, K., Mishima, T., Kurita, T.: Scale invariant face detection and classification method using shift invariant features extracted from log-polar image. IEICE Trans. Inf. Syst. **84**(7), 867–878 (2001)
34. Kam, Z.: The reconstruction of structure from electron micrographs of randomly oriented particles. J. Theor. Biol. **82**(1), 15–39 (1980)
35. Leggett, R.M., Heavens, D., Caccamo, M., Clark, M.D., Davey, R.P.: Nanook: multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. Bioinformatics **32**(1), 142–144 (2015)
36. Levitt, M.: Nature of the protein universe. Proc. Natl. Acad. Sci. **106**(27), 11079–11084 (2009)
37. Lim, W.A.: The modular logic of signaling proteins: building allosteric switches from simple binding domains. Curr. Opin. Struct. Biol. **12**(1), 61–68 (2002)

38. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way, 3rd edn. Academic Press, Boston (2008)
39. Martinec, D., Pajdla, T.: Robust rotation and translation estimation in multiview reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
40. McGinty, R.K., Tan, S.: Recognition of the nucleosome by chromatin factors and enzymes. Curr. Opin. Struct. Biol. **37**, 54–61 (2016)
41. Meynard, A., Torrésani, B.: Spectral analysis for nonstationary audio. IEEE/ACM Trans. Audio Speech Lang. Process. **26**(12), 2371–2380 (2018)
42. Omer, H., Torrésani, B.: Estimation of frequency modulations on wideband signals; applications to audio signal analysis. http://arxiv.org/abs/1305.3095 (2013)
43. Omer, H., Torrésani, B.: Time-frequency and time-scale analysis of deformed stationary processes, with application to non-stationary sound modeling. Appl. Comput. Harmon. Anal. **43**(1), 1–22 (2017)
44. Park, W., Chirikjian, G.S.: An assembly automation approach to alignment of noncircular projections in electron microscopy. IEEE Trans. Autom. Sci. Eng. **11**(3), 668–679 (2014)
45. Park, W., Midgett, C.R., Madden, D.R., Chirikjian, G.S.: A stochastic kinematic model of class averaging in single-particle electron microscopy. Int. J. Robot. Res. **30**(6), 730–754 (2011)
46. Perry, A., Wein, A.S., Bandeira, A.S., Moitra, A.: Message-passing algorithms for synchronization problems over compact groups. Commun. Pure Appl. Math. **71**(11), 2275–2322 (2018)
47. Robinson, D., Farsiu, S., Milanfar, P.: Optimal registration of aliased images using variable projection with applications to super-resolution. Comput. J. **52**(1), 31–42 (2007)
48. Sadler, B.M., Giannakis, G.B.: Shift- and rotation-invariant object reconstruction using the bispectrum. J. Opt. Soc. Am. A **9**(1), 57–69 (1992)
49. Scheres, S.H., Valle, M., Nuñez, R., Sorzano, C.O., Marabini, R., Herman, G.T., Carazo, J.M.: Maximum-likelihood multi-reference refinement for electron microscopy images. J. Mol. Biol. **348**(1), 139–149 (2005)
50. Sharon, N., Kileel, J., Khoo, Y., Landa, B., Singer, A.: Method of moments for 3-D single particle ab initio modeling with non-uniform distribution of viewing angles. Inverse Probl. **36**(4), 044003 (2020)
51. Shechtman, Y., Eldar, Y.C., Cohen, O., Chapman, H.N., Miao, J., Segev, M.: Phase retrieval with application to optical imaging: a contemporary overview. IEEE Signal Process. Mag. **32**(3), 87–109 (2015)
52. Singer, A.: Angular synchronization by eigenvectors and semidefinite programming. Appl. Comput. Harmon. Anal. **30**(1), 20–36 (2011)
53. Sun, W.: Phaseless sampling and linear reconstruction of functions in spline spaces. http://arxiv.org/abs/1709.04779 (2017)
54. Theobald, D.L., Steindel, P.A.: Optimal simultaneous superpositioning of multiple structures with missing data. Bioinformatics **28**(15), 1972–1979 (2012)
55. Tsatsanis, M.K., Giannakis, G.B.: Translation, rotation, and scaling invariant object and texture classification using polyspectra. In: SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations, vol. 1348, pp. 103–115 (1990)
56. Yellott, J.I., Iverson, G.J.: Uniqueness properties of higher-order autocorrelation functions. J. Opt. Soc. Am. A **9**(3), 388–404 (1992)
57. Zhong, Y., Boumal, N.: Near-optimal bounds for phase synchronization. SIAM J. Optim. **28**(2), 989–1016 (2018)
58. Zwart, J.P., van der Heiden, R., Gelsema, S., Groen, F.: Fast translation invariant classification of HRR range profiles in a zero phase representation. IEE Proc. **150**(6), 411–418 (2003)