De-noising by thresholding operator adapted wavelets

Gene Ryan Yoo1 · Houman Owhadi2

Published online: 21 September 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Donoho and Johnstone (Ann Stat 26(3):879–921, 1998) proposed a method from reconstructing an unknown smooth function $u$ from noisy data $u + \zeta$ by translating the empirical wavelet coefficients of $u + \zeta$ towards zero. We consider the situation where the prior information on the unknown function $u$ may not be the regularity of $u$ but that of $Lu$ where $L$ is a linear operator (such as a PDE or a graph Laplacian). We show that the approximation of $u$ obtained by thresholding the gamblet (operator adapted wavelet) coefficients of $u + \zeta$ is near minimax optimal (up to a multiplicative constant), and with high probability, its energy norm (defined by the operator) is bounded by that of $u$ up to a constant depending on the amplitude of the noise. Since gamblets can be computed in $O(N \text{polylog } N)$ complexity and are localized both in space and eigenspace, the proposed method is of near-linear complexity and generalizable to nonhomogeneous noise.

Keywords Probabilistic numerics · Denoising · Thresholding · Wavelets · Gamblet transform

1 Introduction
Donoho et al. (1990), Donoho (1995) and Donoho et al. (1998) addressed the problem of recovering of a smooth signal from noisy observations by soft-thresholding empirical wavelet coefficients (Donoho 1995). More recently, Ding and Mathé (2017) considered the recovery of $x \in X$ based on the observation of $Tx + \zeta$, where $\zeta_i$ is i.i.d. $\mathcal{N}(0, \sigma^2)$ and $T$ is a compact linear operator between Hilbert spaces $X$ and $Y$ with the prior that $x$ lies in an ellipsoid defined by the eigenvectors of $T^*T$. Ding and Mathé (2017) showed that thresholding the coefficients of the corrupted signal $Tx + \zeta$ in the basis formed by the SVD of $T$ (which can be computed in $O(N^3)$ complexity) approached the minimax recovery to a fixed multiplicative constant.

In this paper we are interested in the fast recovery of a signal $u$ based on noisy observations $u + \zeta$ and a bound the regularity of $Lu$ where $L$ is a linear operator. Our main motivation is to approximate solutions of PDEs or Graph Laplacians based on their noisy observations (which is why we use $u$ for the signal rather than $x$).

Let $s \in \mathbb{N}^*$ and $\Omega$ be a regular bounded domain of $\mathbb{R}^d$ ($d \in \mathbb{N}$). Let $\mathcal{H}^s_0(\Omega)$ be the Sobolev space (Gazzola et al. 2010, Sec. 2.2.1) defined as the closure of the set of smooth functions with compact support in $\Omega$.

$$\mathcal{H}^s_0(\Omega) = \text{closure of the set of smooth functions with compact support in } \Omega$$

Our first setting will be that of a symmetric positive linear bijection mapping $\mathcal{H}^s_0(\Omega)$ to $\mathcal{H}^{-s}(\Omega)$, i.e.

$$\mathcal{L} : \mathcal{H}^s_0(\Omega) \rightarrow \mathcal{H}^{-s}(\Omega)$$

We also assume $\mathcal{L}$ to be local, i.e. $\int_{\Omega} u \mathcal{L} v = 0$ for $u, v \in \mathcal{H}^s_0(\Omega)$ with disjoint supports (this assumption is used to achieve near-linear complexity in the recovery). Let $\| \cdot \|$ be the energy-norm defined by

$$\|u\|^2 := \int_{\Omega} u \mathcal{L} u, \quad (1.2)$$

and write

$$\langle u, v \rangle := \int_{\Omega} u \mathcal{L} v, \quad (1.3)$$

by

$$\|u\|^2 := \int_{\Omega} u \mathcal{L} u, \quad (1.2)$$

and write

$$\langle u, v \rangle := \int_{\Omega} u \mathcal{L} v, \quad (1.3)$$
for the associated scalar product.

Let

\[ \zeta \sim \mathcal{N}(0, \sigma^2 \delta(x - y)), \]

be white noise defined as a centered Gaussian process on \( \Omega \) with covariance function \( \sigma^2 \delta(x - y) \). Consider the following problem.

**Problem 1** Let \( u \) be an unknown element of \( \mathcal{H}_0^1(\Omega) \). Given the noisy observation \( \eta = u + \zeta \) and a prior bound on \( \| \mathcal{L} u \|_{L^2} \), find an approximation of \( u \) that is as accurate as possible in the energy norm \( \| \cdot \| \).

**Example 1** As a running illustrative example we will consider the case where \( s = 1 \) and \( \mathcal{L} \) is the differential operator \(- \text{div} (a(x) \nabla \cdot )\) where the conductivity \( a \) is a uniformly elliptic symmetric \( d \times d \) matrix with entries in \( L^\infty(\Omega) \). This example is of practical importance in groundwater flow modeling (where \( a \) is the porosity of the medium) and in electrostatics (where \( a \) is the dielectric constant), and in both applications \( a \) may be rough (non-smooth).

We define \( \lambda_{\text{min}}(a) \) as the largest constant and \( \lambda_{\text{max}}(a) \) as the smallest constant such that for all \( x \in \Omega \) and \( l \in \mathbb{R}^d \),

\[ \lambda_{\text{min}}(a)||l||^2 \leq l^T a(x) l \leq \lambda_{\text{max}}(a)||l||^2, \]

where \( ||l|| \) represents the Euclidean norm of \( l \). Problem 1 then corresponds to the problem of recovering the solution of the PDE

\[
\begin{cases}
    - \text{div} (a(x) \nabla u(x)) = f(x) & \text{in } \Omega; \\
    u = 0 & \text{on } \partial \Omega,
\end{cases}
\]

from its noisy observation \( \eta = u + \zeta \) and a prior bound on \( \| f \|_{L^2(\Omega)} \).

To solve Problem 1 we will decompose \( \eta \) over wavelets adapted to the operator \( \mathcal{L} \) and filter its wavelet coefficients. In Sect. 2, we will first summarize (see Owhadi 2017; Owhadi and Scovel 2017, 2020; Owhadi and Zhang 2017; Schäfer et al. 2017 for details) the main properties of these operator adapted wavelets, named gamblets in reference to their game theoretic interpretation (Owhadi 2017; Owhadi and Scovel 2017, 2020). Given that gamblets can also be interpreted in the frameworks of Information Based Complexity (Woźniakowski 1986), Bayesian numerical analysis (Diaconis 1988), optimal recovery (Michelli and Rivlin 1977), and probabilistic numerics (Hennig et al. 2015), the results of this paper suggest that probabilistic numerical methods (Hennig et al. 2015; Chkrebtii et al. 2016; Owhadi 2015; Briol et al. 2015; Cockayne et al. 2016, 2017; Raissi et al. 2017) may not only lead to efficient quadrature rules (Schober et al. 2014), seamless integration of model uncertainty with numerical errors (Oates et al. 2017), fast solvers (Schäfer et al. 2017), and optimal pipelines of computation (Cockayne et al. 2017), they may also lead to near-optimal methods for the de-noising of solutions of differential equations (see Remark 5).

## 2 Gamblets

This section will give a brief introduction to gamblets, more detailed exposition can be found in Owhadi (2017), Owhadi and Scovel (2017, 2020), Owhadi and Zhang (2017) and Schäfer et al. (2017).

### 2.1 Hierarchy of measurement functions

Let \( q \in \mathbb{N}^s \) (used to represent a number of scales). Let \( (\mathcal{T}^{(k)})_{1 \leq k \leq q} \) be a hierarchy of labels defined as follows. \( \mathcal{T}^{(k)} \) is a set of \( q \)-tuples consisting of elements \( i = (i_1, \ldots, i_q) \).

For \( 1 \leq k \leq q \) and \( i \in \mathcal{T}^{(q)}, i^{(k)} := (i_1, \ldots, i_k) \) and \( \mathcal{T}^{(k)} \) is the set of \( k \)-tuples \( \mathcal{T}^{(k)} = \{ i^{(k)} | i \in \mathcal{T}^{(q)} \} \). For \( 1 \leq r \leq k \leq q \) and \( j = (j_1, \ldots, j_r) \in \mathcal{T}^{(k)} \), we write \( j^{(r)} = (j_1, \ldots, j_r) \). We say that \( \mathcal{M} \) is a \( \mathcal{T}^{(k)} \times \mathcal{T}^{(l)} \) matrix if its rows and columns are indexed by elements of \( \mathcal{T}^{(k)} \) and \( \mathcal{T}^{(l)} \) respectively.

Let \( \phi_i^{(k)} | k \in \{1, \ldots, q\}, i \in \mathcal{T}^{(k)} \) be a nested hierarchy of elements of \( \mathcal{H}^{-s}(\Omega) \) such that \( \phi_i^{(q)} \) are linearly independent and

\[ \phi_i^{(k)} = \sum_{j \in \mathcal{T}^{(k+1)}} \pi_{i,j}^{(k+1)} \phi_j^{(k+1)}, \]

for \( i \in \mathcal{T}^{(k)}, k \in \{1, \ldots, q - 1\} \), where \( \pi^{(k+1)} \) is an \( \mathcal{T}^{(k)} \times \mathcal{T}^{(k+1)} \) matrix and

\[ \pi^{(k+1)} \pi^{(k+1),T} = I^{(k)}. \]

In (2.2), \( \pi^{(k+1),T} \) is the transpose of \( \pi^{(k+1)} \) and \( I^{(k)} \) is the \( \mathcal{T}^{(k)} \times \mathcal{T}^{(k)} \) identity matrix.

### 2.2 Hierarchy of operator adapted pre-wavelets

Let \( \psi_i^{(k)} | i \in \mathcal{T}^{(k)} \) be the hierarchy of optimal recovery splines associated with \( \phi_i^{(k)} | i \in \mathcal{T}^{(k)} \), i.e. for \( k \in \{1, \ldots, q\} \) and \( i \in \mathcal{T}^{(k)} \),

\[ \psi_i^{(k)} = \sum_{j \in \mathcal{T}^{(k)}} A_{i,j}^{(k)} L^{-1} \phi_j^{(k)}, \]

where

\[ A^{(k)} := (\Theta^{(k)})^{-1}, \]
and $\Theta^{(k)}$ is the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k)}$ symmetric positive definite Gramian matrix with entries (writing $[\phi, u]$ for the duality pairing between $\phi \in \mathcal{H}^{-3}(\Omega)$ and $v \in \mathcal{H}_0^1(\Omega)$)

\begin{equation}
\Theta_{i,j}^{(k)} = [\phi_i^{(k)}, L^{-1}\phi_j^{(k)}].
\end{equation}

Note that $A^{(k)}$ is the stiffness matrix of the elements $(\psi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ in the sense that

\begin{equation}
A_{i,j}^{(k)} = [\psi_i^{(k)}, \psi_j^{(k)}].
\end{equation}

Writing $\Phi^{(k)} := \text{span}\{\phi_i^{(k)} | i \in \mathcal{I}^{(k)}\}$ and $\Psi^{(k)} := \text{span}\{\psi_i^{(k)} | i \in \mathcal{I}^{(k)}\}$, $\Phi^{(k)} \subset \Phi^{(k+1)}$ and $\Psi^{(k)} = L^{-1}\Phi^{(k)}$ imply $\Psi^{(k)} \subset \Psi^{(k+1)}$. The $(\phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ and $(\psi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ form a bi-orthogonal system in the sense that

\begin{equation}
[\phi_i^{(k)}, \psi_j^{(k)}] = \delta_{i,j} \text{ for } i, j \in \mathcal{I}^{(k)},
\end{equation}

and the $\langle \cdot, \cdot \rangle$-orthogonal projection of $u \in \mathcal{H}_0^1(\Omega)$ on $\psi^{(k)}$ is

\begin{equation}
u^{(k)} := \sum_{i \in \mathcal{I}^{(k)}} [\phi_i^{(k)}, u] \psi_i^{(k)}.
\end{equation}

### 2.3 Operator adapted wavelets

Let $(\mathcal{J}^{(k)})_{2 \leq k \leq q}$ be a hierarchy of labels such that (writing $|\mathcal{J}^{(k)}|$ for the cardinal of $\mathcal{J}^{(k)}$)

\begin{equation}
|\mathcal{J}^{(k)}| = |\mathcal{I}^{(k)}| - |\mathcal{I}^{(k-1)}|.
\end{equation}

For $k \in \{2, \ldots, q\}$, let $W^{(k)}$ be a $\mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ matrix such that

\begin{equation}
\text{Ker}(\pi^{(k-1,k)}) = \text{Im}(W^{(k)}, T).
\end{equation}

For $k \in \{2, \ldots, q\}$ and $i \in \mathcal{J}^{(k)}$ define

\begin{equation}
\chi_i^{(k)} := \sum_{j \in \mathcal{I}^{(k)}} W_{i,j} \psi_j^{(k)},
\end{equation}

and write $\mathcal{X}^{(k)} := \text{span}\{\chi_i^{(k)} | i \in \mathcal{J}^{(k)}\}$. Then $\mathcal{X}^{(k)}$ is the $\langle \cdot, \cdot \rangle$-orthogonal complement of $\psi^{(k-1)}$ in $\psi^{(k)}$, i.e., $\psi^{(k)} = \psi^{(k-1)} \oplus \mathcal{X}^{(k)}$, and

\begin{equation}
\psi^{(q)} = \psi^{(1)} \oplus \mathcal{X}^{(2)} \oplus \cdots \oplus \mathcal{X}^{(q)}.
\end{equation}

For $k \in \{2, \ldots, q\}$ write

\begin{equation}
B^{(k)} := W^{(k)} A^{(k)} W^{(k)}, T.
\end{equation}

Note that $B^{(k)}$ is the stiffness matrix of the elements $(\chi_i^{(k)})_{j \in \mathcal{J}^{(k)}}$, i.e.

\begin{equation}
B_{i,j}^{(k)} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle.
\end{equation}

Further, for $k \in \{2, \ldots, q\}$, define

\begin{equation}
N^{(k)} := A^{(k)} W^{(k)}, T B^{(k)}, -1
\end{equation}

and for $i \in \mathcal{J}^{(k)}$,

\begin{equation}
\phi_i^{(k), \mathcal{X}} := \sum_{j \in \mathcal{J}^{(k)}} N_{i,j}^{(k)} T \phi_j^{(k)}.
\end{equation}

Then defining $u^{(k)}$ as in (2.8), it holds true that for $k \in \{2, \ldots, q\}$, $u^{(k)} - u^{(k-1)}$ is the $\langle \cdot, \cdot \rangle$-orthogonal projection of $u$ on $\mathcal{X}^{(k)}$ and

\begin{equation}
u^{(k)} = \sum_{i \in \mathcal{J}^{(k)}} [\phi_i^{(k), \mathcal{X}}, u] \chi_i^{(k)}.
\end{equation}

To simplify notations, write $\mathcal{J}^{(1)} := \mathcal{I}^{(1)}$, $B^{(1)} := A^{(1)}$, $N^{(1)} := I^{(1)}$, $\phi_i^{(1), \mathcal{X}} := \phi_i^{(1)}$ for $i \in \mathcal{J}^{(1)}$, $\mathcal{J} := \mathcal{J}^{(1)} \cup \cdots \cup \mathcal{J}^{(q)}$, $\chi_i := \chi_i^{(k)}$ and $\phi_i^{\mathcal{X}} := \phi_i^{(k), \mathcal{X}}$ for $i \in \mathcal{J}^{(k)}$ and $1 \leq k \leq q$. Then the $\phi_i^{\mathcal{X}}$ and $\chi_i$ form a bi-orthogonal system, i.e.

\begin{equation}
[\phi_i^{\mathcal{X}}, \chi_j] = \delta_{i,j} \text{ for } i, j \in \mathcal{J},
\end{equation}

and

\begin{equation}
u^{(q)} = \sum_{i \in \mathcal{J}} [\phi_i^{\mathcal{X}}, u] \chi_i.
\end{equation}

Simplifying notations further, we will write $[\phi_i^{\mathcal{X}}, u]$ for the $\mathcal{J}$ vector with entries $[\phi_i^{\mathcal{X}}, u]$ and $\chi$ for the $\mathcal{J}$ vector with entries $\chi_i$ so that (2.19) can be written

\begin{equation}
u^{(q)} = [\phi_i^{\mathcal{X}}, u] \cdot \chi.
\end{equation}

Further, define the $\mathcal{J}$ by $\mathcal{J}$ block-diagonal matrix $B$ defined as $B_{i,j} = B_{i,j}^{(k)}$ if $i, j \in \mathcal{J}^{(k)}$ and $B_{i,j} = 0$ otherwise. Note that it holds that $B_{i,j} = \langle \chi_i, \chi_j \rangle$. When $q = \infty$, and $\cup_{k=1}^\infty \mathcal{X}^{(k)}$ is dense in $\mathcal{H}^{-3}(\Omega)$ then (writing $\mathcal{X}^{(1)} := \psi^{(1)}$)

\begin{equation}
\mathcal{H}_0^1(\Omega) = \oplus_{k=1}^\infty \mathcal{X}^{(k)},
\end{equation}

$u^{(q)} = u$ and (2.19) is the corresponding multi-resolution decomposition of $u$. When $q < \infty$, $u^{(q)}$ is the projection of $u$ on $\oplus_{k=1}^q \mathcal{X}^{(k)}$ and (2.20) is the corresponding multi-resolution decomposition.

---

1. We write $M^{(k), T}$ and $M^{(k), -1}$ for the transpose and inverse of a matrix $M^{(k)}$.

2. The dependence on $k$ is left implicit to simplify notation, for $i \in \mathcal{J}$ there exists a unique $k$ such that $i \in \mathcal{J}^{(k)}$. 

---

© Springer
of these sets satisfy (2.9).

\[ \phi_{ij}^{(k)} = \frac{1}{|\tau_i^{(k)}|} \]

\[ \phi_{ij}^{(2)} = \frac{1}{|\tau_j^{(2)}|} \]

\[ \phi_{ij}^{(3)} = \frac{1}{|\tau_j^{(3)}|} \]

Fig. 1 Pre-Haar wavelets with $\Omega = [0, 1]^2$ and $[\tau_i^{(k)}]_{i \in \mathcal{I}^{(k)}}$ forms a uniform partition of $\Omega$ into $2^k \times 2^k$ squares. The top row shows the supports of sample pre-Haar measurement functions. The bottom row shows examples of rows of $\pi^{(1,2)}$ and $\pi^{(2,3)}$. Used from forthcoming book (Owhadi and Scovel 2020) with permission from Cambridge University Press

### 2.4 Pre-Haar wavelet measurement functions

The gamblers used in the subsequent developments will use pre-Haar wavelets (defined below and whose supports are illustrated in a particular instantiation in Fig. 1) as measurement functions $\phi_{ij}^{(k)}$ and our main near optimal de-noising estimates will be derived from their properties (summarized in Theorem 2).

Let $\delta, h \in (0, 1)$. Let $(\tau_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ be uniformly Lipschitz convex sets forming a nested partition of $\Omega$, i.e. such that $\Omega = \bigcup_{i \in \mathcal{I}^{(k)}} \tau_i^{(k)}$, $k \in \{1, \ldots, q\}$ is a disjoint union except for the boundaries, and $\tau_i^{(k)} = \bigcup_{j \in \mathcal{I}^{(k+1)}; j(k) = i} \tau_j^{(k+1)}$, $k \in \{1, \ldots, q-1\}$. Assume that each $\tau_i^{(k)}$ contains a ball of radius $\delta h^k$, and is contained in the ball of radius $\delta^{-1} h^k$. Writing $|\tau_i^{(k)}|$ for the volume of $\tau_i^{(k)}$, take

\[ \phi_{ij}^{(k)} := \frac{1}{|\tau_i^{(k)}|} |\tau_j^{(k)}|^{\frac{1}{2}}. \]  

The nesting relation (2.1) is then satisfied with $\pi_{ij}^{(k+1)} := |\tau_j^{(k+1)}|^{\frac{1}{2}} |\tau_i^{(k)}| - \frac{1}{2}$ for $j(k) = i$ and $\pi_{ij}^{(k+1)} := 0$ otherwise.

For $k \in \{2, \ldots, q\}$ let $\mathcal{J}^{(k)}$ be a finite set of $k$-tuples of the form $j = (j_1, \ldots, j_k)$ such that $\{j(k-1) \mid j \in \mathcal{J}^{(k)}\} = \mathcal{I}^{(k-1)}$ and for $i \in \mathcal{I}^{(k-1)}$, $\text{Card} \{j \in \mathcal{J}^{(k)} \mid j(k-1) = i\} = \text{Card} \{j \in \mathcal{I}^{(k-1)} \mid j(k-1) = i\} - 1$. Note that the cardinalities of these sets satisfy (2.9).

Write $J^{(k)}$ for the $\mathcal{I}^{(k)} \times \mathcal{J}^{(k)}$ identity matrix. For $k = 2, \ldots, q$ let $W^{(k)}$ be a $\mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ matrix such that $\text{Im}(W^{(k)T}) = \text{Ker}(\sigma^{(k-1, k)})$, $W^{(k)}(W^{(k)})^T = J^{(k)}$ and $W_{i,j} = 0$ for $i(k-1) \neq j(k-1)$.

**Theorem 2** With pre-Haar wavelet measurement functions, it holds true that

1. For $k \in \{1, \ldots, q\}$ and $u \in \mathcal{L}^{-1}L^2(\Omega)$,

\[ \|u - u^{(k)}\| \leq C h^{k_2} \|Lu\|_{L^2(\Omega)}. \]  

2. Writing $\text{Cond}(M)$ for the condition number of a matrix $M$ we have for $k \in \{1, \ldots, q\}$,

\[ C^{-1} h^{-2(k-1)s} f(k) \leq B^{(k)} \leq C h^{-2k_2} f(k) \]

and $\text{Cond}(B^{(k)}) \leq C h^{-2s}$.

3. For $i \in \mathcal{I}^{(k)}$ and $x_i^{(k)} \in \tau_i^{(k)}$,

\[ \|\psi_i\|_{\mathcal{H}^{s}(\Omega \setminus B(x_i^{(k)}, nh))} \leq C h^{-s} e^{-n/C} \]  

4. The wavelets $\psi_{i}^{(k)}, x_i^{(k)}$ and stiffness matrices $A^{(k)}, B^{(k)}$ can be computed to precision $\epsilon$ (in $\|\cdot\|_{-\infty}$-norm for elements of $\mathcal{H}^{s}(\Omega)$ and in Frobenius norm for matrices) in $O(N \log^{2d+1} N \epsilon)$ complexity.

**Furthermore the constant $C$ depends only on $\delta, \Omega, d, s$,**

\[ \|\mathcal{L}\| := \sup_{u \in \mathcal{H}^{s}(\Omega)} \frac{\|Lu\|_{\mathcal{H}^{-s}(\Omega)}}{\|u\|_{\mathcal{H}^{s}(\Omega)}} \] and

\[ \|\mathcal{L}^{-1}\| := \sup_{u \in \mathcal{H}^{s}(\Omega)} \frac{\|Lu\|_{\mathcal{H}^{-s}(\Omega)}}{\|u\|_{\mathcal{H}^{s}(\Omega)}}. \]  

**Proof** 1 and 2 follows from an application of Prop. 4.17 and Theorems 4.14 and 3.19 from Owhadi and Scovel (2017). 3 follows from Thm. 2.23 of Owhadi and Scovel (2017). 4 follows from the complexity analysis of Alg. 6 of Owhadi and Scovel (2017). See Owhadi and Scovel (2020) for detailed proofs.

**Remark 3** The wavelets $\psi_{i}^{(k)}, x_i^{(k)}$ and stiffness matrices $A^{(k)}, B^{(k)}$ can also be computed in $O(N \log^{2d} N \epsilon)$ complexity using the incomplete Cholesky factorization approach of Schäfer et al. (2017).

### 3 On $\mathbb{R}^N$

Gamblers have a natural generalization in which $\mathcal{H}^{s}(\Omega)$ and $\mathcal{H}^{-s}(\Omega)$ are replaced by $\mathbb{R}^N$ (or a finite-dimensional space) and $\mathcal{L}$ is replaced by an $N \times N$ symmetric positive definite matrix $A$ (Owhadi and Scovel 2017, 2020). This generalization is relevant to practical applications requiring the prior numerical discretization of the continuous operator $\mathcal{L}$. In these applications $1$ $\mathcal{H}^{s}(\Omega)$ is replaced by the linear space spanned by the finite-elements $\psi_i$ (e.g. piecewise linear or...
bi-linear tent functions on a fine mesh/grid in the setting of Example 1) used to discretize the operator \( \mathcal{L}(2) \mathcal{T}^{(q)} \) is used to label the elements \( \tilde{\psi}_i \) for \( i \in \mathcal{I}^{(q)} \). Algorithms 1 and 2 summarize the discrete gamblet-transform (Owhadi 2017; Owhadi and Scovel 2017, 2020) and the discrete Gamblet solve of the linear system \( \mathcal{L}u = f \). Note that Algorithm 1 computes gamblets at level \( k - 1 \) from those at level \( k \) with the interpolation matrices \( \mathcal{R}^{(k-1,k)} \) defined using the matrices \( \pi^{(k-1,k)} \) and \( W^{(k)} \), introduced at and around Eqs. (2.2) and (2.10). The near-linear complexities mentioned in Theorem 2 are based on the near-sparsity of the interpolation matrices \( \mathcal{R}^{(k-1,k)} \) and the well-conditioning and near-sparsity of the \( \mathcal{B}^{(k)} \), and achieved by localizing the computation of gamblets, the inversion of \( \mathcal{B}^{(k)} \) and truncating the \( \mathcal{A}^{(k)} \) (see Owhadi 2017; Owhadi and Scovel 2017, 2020 for details).

### Algorithm 1 The Gamblet transform

1: \( \psi_j^{(q)} = \tilde{\psi}_i \)
2: \( x_j^{(q)} = (\tilde{\psi}_i, \psi_j^{(q)}) \)
3: for \( k = q \) to 2 do
4: \( B^{(k)} = W^{(k)} \mathcal{A}^{(k)} W^{(k),T} \)
5: \( \tilde{x}_j^{(k)} = \sum_{i \in \mathcal{I}^{(k)}} W_{i,j} \tilde{\psi}_i \)
6: \( \tilde{x}_j^{(1-k)}(\mathcal{I}^{(k)} - \mathcal{A}^{(k)} W^{(k),T} B^{(k-1)} W^{(k)}) \)
7: \( \tilde{x}_j^{(1-k)} = \sum_{i \in \mathcal{I}^{(k)}} \mathcal{A}^{(k-1)} \tilde{x}_j^{(k)} \psi_j \)
9: end for

### Algorithm 2 The Gamblet solve

1: \( x_j^{(q)} = \int_{\mathcal{D}} f \tilde{\psi}_i \)
2: for \( k = q \) to 2 do
3: \( \tilde{x}_j^{(k)} = B^{(k-1)} W^{(k),T} f^{(k)} \)
4: \( v^{(k)} = \sum_{i \in \mathcal{I}^{(k)}} w_{i,j}^{(k)} \tilde{x}_j^{(k)} \)
5: \( f^{(k-1)} = \mathcal{R}^{(k-1,k)} f^{(k)} \)
6: end for
7: \( v^{(1)} = A^{(1-1)} f^{(1)} \)
8: \( v^{(l)} = \sum_{i \in \mathcal{I}^{(l)}} w_{l,j}^{(1)} \tilde{\psi}_i^{(1)} \)
9: \( u = v^{(1)} + v^{(2)} + \ldots + v^{(q)} \)

### 4 De-noising by filtering gamblet coefficients

In this section we will show that filtering the gamblet coefficients of \( \eta \) in Problem 1 produces an approximation of \( u \) that is minimax optimal up to a multiplicative constant, i.e., “near minimax”.

#### 4.1 Near minimax recovery

This work will consider the following discrete variant of Problem 1 within the finite dimensional space spanned by Gamblet wavelets (using pre-Haar measurement functions defined in Sect. 2.4) taken to the \( q \)th level (note there are no mathematical constraints to the number of levels taken in the decomposition). In addition, the discrete noise used in this problem, \( \zeta \in \Psi^{(q)} \), is the projection of the noise (1.4) onto \( \Psi^{(q)} \) (due to (2.8)).

**Problem 2** Let \( u \) be an unknown element of \( \Psi^{(q)} \subset \mathcal{H}_0^2(\Omega) \) for \( q < \infty \). Let \( \zeta \) be a centered Gaussian vector in \( \Psi^{(q)} \) such that

\[
\mathbb{E}[\langle \psi_j^{(q)}, \zeta], \langle \phi_j^{(q)}, \zeta \rangle] = \sigma^2 \delta_{i,j}.
\]

Given the noisy observation \( \eta = u + \zeta \) and a prior bound \( M > 0 \) on \( \| Lu \|_{L^2(\Omega)} \), find an approximation of \( u \) in \( \Psi^{(q)} \) that is as accurate as possible in the energy norm \( \| \cdot \| \).

To justify this discrete approximation, recall that by Theorem 2, we have \( \| u - u^{(q)} \| \leq Ch^{\theta} \| f \|_{L^2(\Omega)} \). Hence with the prior bound on \( \| Lu \|_2 \), this approximation is arbitrarily accurate with \( q \) large enough. Conversely, it will be shown that if \( q \) is taken to be small enough relative to the level of noise, the identity recovery, \( \eta^{(q)} = \eta \), is nearly optimal (as in Theorem 4). We will elaborate on this point after Theorem 4. Let \( \eta \) be as in Problem 2 and let Gamblets be defined as in Sect. 2.4 (with pre-Haar measurement functions). For \( l \in \{1, \ldots, q\} \) let

\[
\eta^{(l)} = \sum_{k=1}^{l} [\langle \phi_k^{(l)}, \chi \rangle, \eta] \cdot \chi^{(k)}
\]

and \( \eta^{(0)} = 0 \in \Psi^{(q)} \). Let \( M > 0 \) and write

\[
\mathcal{V}_M^{(q)} = \{ u \in \Psi^{(q)} \mid \| Lu \|_{L^2(\Omega)} \leq M \}.
\]

Assume that \( \sigma > 0 \) and write

\[
l^* = \arg\min_{l \in \{0, \ldots, q\}} \beta_l,
\]

for

\[
\beta_l = \begin{cases} 
  h^{2s} M^2 & \text{if } l = 0 \\
  \sigma^2 h^{-(2s+d)+l} + h^{2(l+1)} M^2 & \text{if } 1 \leq l \leq q - 1 \\
  h^{-(2s+d)q} \sigma^2 & \text{if } l = q.
\end{cases}
\]
The following theorem shows that $\eta^{(l)}$ is a near minimax recovery of $u$, by which we mean that the $\| \cdot \|^2$ recovery error is minimax optimal up to a multiplicative constant [depending only on $\|L\|$, $\|L^{-1}\|$, $\Omega$, $d$, $\delta$ and whose value can be made explicit using the estimates of Owhadi and Scovel (2020)]. We will also refer to $\eta^{(l)}$ as the smooth recovery of $u$ because, with probability close to 1, it is nearly as regular (in energy norm) as $u$. Note that in the context of example 1, this implies that the $H^l(\Omega)$ Sobolev norm of the recovery is bounded by the corresponding norm of $u$.

**Theorem 4** Suppose $v^\dagger(\eta) = \eta^{(l)}$, then there exists a constant $C$ depending only on $h$, $s$, $\|L\|$, $\|L^{-1}\|$, $\Omega$, $d$, and $\delta$ such that

$$
\sup_{u \in V_M^{(q)}} \mathbb{E}[\|u - v^\dagger(\eta)\|^2] < C \inf_{v(\eta) \in V_M^{(q)}} \mathbb{E}[\|u - v(\eta)\|^2],
$$

(4.6)

where the infimum is taken over all measurable functions $v : \psi^{(q)} \to \psi^{(q)}$. Furthermore, if $l^\dagger \neq 0$, then with probability at least $1 - \varepsilon$,

$$
\|\eta^{(l)}\| \leq \|u\| + C \sqrt{\frac{1}{\varepsilon} \log \frac{1}{\delta} \frac{2^{4-s}M}{4^{4-s}}}.
$$

(4.7)

**Proof** See Sect. 7. \hfill \Box

Note that $l^\dagger = q$ occurs (approximately) when $q$ is such that $h^q > (\frac{\sigma}{M})^{\frac{2}{4-s}}$, i.e. when

$$
\frac{\sigma}{M} < h^{\frac{4-s}{2}},
$$

(4.8)

and in this case $\eta^{(q)}$ is a near minimax optimal recovery of $u^{(q)}$. On the other extreme $l^\dagger = 0$ occurs (approximately) when $(\frac{\sigma}{M})^{\frac{2}{4-s}} > h$, i.e. when

$$
\frac{\sigma}{M} > h^{\frac{4-s}{2}}
$$

(4.9)

and in this case, the zero signal is a near optimal recovery.

**Remark 5** Write $\xi \sim \mathcal{N}(0, L^{-1})$ for the centered Gaussian field on $H^0(\Omega)$ with covariance operator $L^{-1}$ (i.e. $[\phi, \xi] \sim \mathcal{N}(0, [\phi, L^{-1}\phi])$ for $\phi \in H^{-1}(\Omega)$). According to Owhadi (2017) and Owhadi and Scovel (2017, 2020), $\xi$ emerges as an optimal mixed strategy in an adversarial game opposing 2 players (in this game Player I selects $u \in H^0(\Omega)$ and player II must approximate $u$ with $v$ based on a finite number of linear measurements on $u$ and receives the loss $\|u - v\|/\|u\|$). Furthermore for $k \in \{1, \ldots, q\}$, $\eta^{(k)}$ also admits the representation

$$
\eta^{(k)} = \mathbb{E}[\xi | [\phi^{(k)}, \xi] = [\phi^{(k)}, \eta]].
$$

(4.10)

Therefore the near-optimal recovery of $u$ can be obtained by conditioning the optimal mixed strategy $\xi$ with respect to linear measurements on $\eta$ made at level $k = l^\dagger$.

### 4.2 Numerical illustrations

#### 4.2.1 Example 1 with $d = 1$

Consider Example 1 with $d = 1$. Take $\Omega = [0, 1] \subset \mathbb{R}$, $q = 10$ and $\phi^{(k)}_i = 1_{[\frac{i-1}{2^{k-1}}, \frac{i}{2^k}]}$ for $1 \leq i \leq 2^k$. Let $W^{(k)}$ be the $2^{k-1} \times 2^k$ matrix with non-zero entries defined by $W_{i,2i-1} = \frac{1}{\sqrt{2}}$ and $W_{i,2i} = -\frac{1}{\sqrt{2}}$. Let $L := -\text{div}(a \nabla \cdot)$ with $a(x) := \prod_{k=1}^{10} (1 + 0.25 \cos(2^k x)).$

(4.11)

In Fig. 2 we select $f(x)$ at random uniformly over the unit $L^2(\Omega)$-sphere of $\Phi^{(q)}$ and let $\zeta$ be white noise [as in (1.4)] with $\sigma = 0.001$ and $\eta = u + \zeta$.

We next consider a case where $f$ is smooth, i.e. $f(x) = \frac{\sin(\pi x)}{\pi}$ on $x \in (0, 1)$ and $f(0) = \pi$. Let $\zeta$ be white noise with standard deviation $\sigma = 0.01$. See Fig. 3 for the corresponding numerical illustrations. Both figures show that (1) $v(\eta)$ and $\nabla v(\eta)$ are accurate approximations of $u$ and $\nabla u$ (2) the accuracy of these approximations increases with the regularity of $f$.

#### 4.2.2 Example 1 with $d = 2$

Consider Example 1 with $d = 2$. Take $\Omega = [0, 1]^2 \subset \mathbb{R}$ and $q = 7$. Use the pre-Haar wavelets defined as $\phi^{(k)}_{i,j} = 1_{[\frac{i-1}{2^k}, \frac{i}{2^k}] \times [\frac{j-1}{2^k}, \frac{j}{2^k}]}$ for $1 \leq i, j \leq 2^k$. Let $W^{(k)}$ be defined be the $2^k(4^k-1)$ by $4^k$ matrix defined as in construction 4.13 of Owhadi (2017).

In Fig. 4 we select $f(x)$ at random uniformly over the unit $L^2(\Omega)$-sphere of $\Phi^{(q)}$ and let $\zeta$ be white noise [as in (1.4)] with $\sigma = 0.001$ and $\eta = u + \zeta$.

Let $L := -\text{div}(a \nabla \cdot)$ with $a(x, y) := \prod_{k=1}^{7} \left[ (1 + \frac{1}{4} \cos(2^k \pi x + y)) \right.$

$$
\left. (1 + \frac{1}{4} \cos(2^k \pi x - 3y)) \right].
$$

(4.12)

Next consider a case where $f$ is smooth, i.e. $f(x, y) = \cos(3x+y)+\sin(3y)+\sin(7x-5y)$. Let $\zeta$ be white noise with standard deviation $\sigma = 0.01$. See Fig. 5 for the corresponding numerical illustrations. As with the $d = 1$ plots, the $d = 2$ plots show the accuracy of the recovery of $u$ and $\nabla u$ and the positive impact of the regularity of $f$ that accuracy.
Fig. 2 The plots of $a, f, u, \eta$, the near minimax recovery $v(\eta) = \eta^{(l)}$, its error from $u$, and the derivatives of $u$ and $v(\eta)$

Fig. 3 The plots of $a$, smooth $f, u, \eta$, $v(\eta) = \eta^{(l)}$, its error from $u$, and the derivatives of $u$ and $v(\eta)$

5 Comparisons

Since hard- and soft-thresholding have been used in Donoho and Johnstone (1994); Donoho (1995); Donoho et al. (1998) for the near minimax recovery of regular signals we will compare the accuracy of (4.2) with that of hard- and soft-thresholding the Gamblet transform of the noisy signal (and regularization).

5.1 Hard- and soft-thresholding

We call hard-thresholding the recovery of $u$ with

$$v(\eta) = \sum_{k=1}^{q} \sum_{i \in J^{(k)}} H^{(l)}_{i}([\phi_{i}^{(k),x}, \eta]) \chi_{i}^{(k)}$$

and

$$H^{\beta}(x) = \begin{cases} x & |x| > \beta \\ 0 & |x| \leq \beta. \end{cases}$$

(5.1)

(5.2)

We call soft-thresholding the recovery of $u$ with

$$v(\eta) = \sum_{k=1}^{q} \sum_{i \in J^{(k)}} S^{(l)}_{i}([\phi_{i}^{(k),x}, \eta]) \chi_{i}^{(k)}$$

and

$$S^{\beta}(x) = \begin{cases} x - \beta \text{sgn}(x) & |x| > \beta \\ 0 & |x| \leq \beta. \end{cases}$$

(5.3)

(5.4)
The parameters \((t_1, \ldots, t_q)\) are adjusted to achieve minimal average errors. Since the mass matrix of \(\phi^x\) is comparable to identity (see Theorem 10), decomposing \(f\) over \(\phi^x\) and the bi-orthogonality identities \([\phi^x, \chi_j] = \delta_{i,j}\) imply that \([f, \chi]\) is approximately uniformly sampled on the unit sphere of \(\mathbb{R}^J\) and the variance of \([f, \chi_{(k)}]\) can be approximated by \(1/|J|\). Therefore \([\phi^x, u] = B^{(k)}_{-1}[f, \chi_{(k)}]\) and (2.24) imply that the standard deviation of \([\phi^{(k)}x, u]\) can be approximated by \(h^{-2k}/\sqrt{|J|}\). Therefore optimal choices follow the power law \(t_k = h^{-2k} t_0\) for some parameter \(t_0\).

### 5.2 Regularization

We call regularized\ the recovery of \(u\) with \(v(\eta)\) defined as the minimizer of

\[
\|v(\eta) - \eta\|_{L^2(\Omega)}^2 + \alpha\|v(\eta)\|^2.
\]

(5.5)

For practical implementation we consider \(A_{i,j} = \langle \tilde{\psi}_i, \tilde{\psi}_j \rangle\), the \(N \times N\) stiffness matrix obtained by discretizing \(\mathcal{L}\) with finite elements \(\tilde{\psi}_1, \ldots, \tilde{\psi}_N\), and write \(\eta = \sum_{i=1}^N v_i \tilde{\psi}_i\) and
basis \( \zeta = \sum_{i=1}^{N} z_i \hat{\psi}_i \) for the representation of \( \eta \) and \( \zeta \) over this basis \( (\eta = u + \zeta \) and \( z \sim \mathcal{N}(0, \sigma^2 I_d) \), writing \( I_d \) for the identity matrix). In that discrete setting we have

\[
v(\eta) = \sum_{i=1}^{N} x_i \hat{\psi}_i
\]

(5.6)

where \( x \) is the minimizer of

\[
|x - y|^2 + \alpha x^T A x
\]

(5.7)

Theorem 6 and Corollary 7 show that this recovery corresponds to minimizing the energy norm \( \|v\|^2 = x^T A x \) subject to \( |x - y| \leq \gamma \) with

\[
y = |(I - (\alpha A + I)^{-1}) y|
\]

(5.8)

In practice \( y \) would correspond to a level of confidence (e.g. chosen so that \( \mathbb{P}[|z| > \gamma] = 0.05 \) with \( z \sim \mathcal{N}(0, \sigma^2 I_d) \)).

**Theorem 6** Let \( x \) be the minimizer of

\[
\begin{align*}
\text{Minimize} & \quad x^T A x \\
\text{subject to} & \quad |x - y| \leq \gamma.
\end{align*}
\]

(5.9)

If \( |y| \leq \gamma \), then \( x = 0 \). Otherwise (if \( |y| > \gamma \)), then \( x = (\alpha A + I)^{-1} y \) where \( \alpha \) is defined as the solution of (5.8).

**Proof** Supposing \( |y| \leq \gamma \), then if \( x = 0 \), then \( |x - y| \leq \gamma \). Further, \( x = 0 \) is the global minimum of \( x^T A x \). Therefore in this case, \( x = 0 \).

If \( |y| > \gamma \) then at minimum \( x \) the hyperplane tangent to the ellipsoid of center zero must also be tangent to the sphere of center \( y \) which implies that \( A x = \alpha^{-1} (y - x) \) for some parameter \( \alpha \). We therefore have \( x = (\alpha A + I)^{-1} y \) and \( \alpha \) is determined by the equation \( |x - y| = \gamma \) which leads to

\[
|(I - (\alpha A + I)^{-1}) y| = \gamma.
\]

(5.10)

\[\square\]

**Corollary 7** If \( |y| > \gamma \), then the minimizers of (5.9) and (5.7) are identical with \( \alpha \) identified as in (5.10).

**Proof** \( \nabla_x (|x - y|^2 + \alpha x^T A x) = 0 \) is equivalent to \( x - y + \alpha A x = 0 \) which leads to \( x = (\alpha A + I)^{-1} y \). \[\square\]

### 5.3 Numerical experiments

**5.3.1 Example 1 with \( d = 1 \)**

Consider the same example as in Sect. 4.2.1. Table 1 shows a comparison of errors measured in \( L^2 \) and energy norms averaged over 3000 independent random realizations of \( f \) and \( \zeta \) (\( f \) is uniformly distributed over the unit sphere of \( L^2(\Omega) \) and \( \zeta \) is white noise with \( \sigma = 0.001 \)). The hard variable thresholding recovery is as defined in Sect. 5.1, regularization recovery is as defined in Sect. 5.2, and the near minimax recovery refers to \( v^\dagger(\eta) = \eta^\dagger \) in Theorem 4. The best performing algorithm in each category is in bold. In this experiment, the proposed near minimax recovery outperforms the other methods in terms of average error and error variance.

For reference, the average and standard deviation of the (discrete) energy norm of \( \zeta \) used in this trial were 1.68 and 0.06 respectively.

**5.3.2 Example 1 with \( d = 2 \)**

Consider the same example as in Sect. 4.2.2. Table 2 shows errors measured in \( L^2 \) and energy norms averaged over 100 independent random realizations of \( f \) and \( \zeta \) (\( f \) is uniformly distributed over the unit sphere of \( L^2(\Omega) \) and \( \zeta \) is white noise with \( \sigma = 0.001 \)). In this experiment, the proposed

---

**Table 1** Comparison of the performance of de-noising algorithms for \( d = 1 \)

| Algorithm              | \( L \) Error AVG | \( L \) Error STDEV | \( L^2 \) Error AVG | \( L^2 \) Error STDEV |
|------------------------|-------------------|----------------------|---------------------|-----------------------|
| Hard variable threshold| \( 4.78 \times 10^{-3} \) | \( 9.64 \times 10^{-4} \) | \( 2.25 \times 10^{-4} \) | \( 1.07 \times 10^{-4} \) |
| Soft variable threshold| \( 4.27 \times 10^{-3} \) | \( 7.70 \times 10^{-4} \) | \( 1.65 \times 10^{-4} \) | \( 5.63 \times 10^{-5} \) |
| Regularization recovery| \( 4.37 \times 10^{-3} \) | \( 7.93 \times 10^{-4} \) | \( 2.82 \times 10^{-4} \) | \( 7.83 \times 10^{-5} \) |
| Near minimax recovery  | \( 3.90 \times 10^{-3} \) | \( 5.30 \times 10^{-4} \) | \( 1.24 \times 10^{-4} \) | \( 2.50 \times 10^{-5} \) |

**Table 2** Comparison of the performance of de-noising algorithms for \( d = 2 \)

| Algorithm              | \( L \) Error AVG | \( L \) Error STDEV | \( L^2 \) Error AVG | \( L^2 \) Error STDEV |
|------------------------|-------------------|----------------------|---------------------|-----------------------|
| Hard variable threshold| \( 6.95 \times 10^{-3} \) | \( 9.78 \times 10^{-5} \) | \( 1.42 \times 10^{-4} \) | \( 7.76 \times 10^{-6} \) |
| Soft variable threshold| \( 7.18 \times 10^{-3} \) | \( 1.57 \times 10^{-4} \) | \( 1.90 \times 10^{-4} \) | \( 2.35 \times 10^{-5} \) |
| Regularization recovery| \( 6.90 \times 10^{-3} \) | \( 1.03 \times 10^{-4} \) | \( 1.86 \times 10^{-4} \) | \( 1.88 \times 10^{-5} \) |
| Near minimax recovery  | \( 6.94 \times 10^{-3} \) | \( 9.58 \times 10^{-5} \) | \( 1.40 \times 10^{-4} \) | \( 7.29 \times 10^{-6} \) |
near minimax recovery is the best or near the best in every error metric (it is slightly outperformed by regularization in average $L$ error).

For reference, the average and standard deviation of the (discrete) $L$ norm of this trial’s $\zeta$ were 0.250 and 0.06 respectively.

6 De-noising graph Laplacians

Since gamblets can also be constructed for graph Laplacians (Owhadi and Scovel 2017, 2020) the proposed method can also be used for de-noising such operators. This section will provide a succinct description of this extension with numerical illustrations.

6.1 The problem

Let $G = (V, E)$ be a simple graph (with vertex set $V$ and edge set $E$) and let $L$ be its graph Laplacian, which is defined as follows:

$$L_{i,j} = \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

Fix $i_0 \in V$ and let $\Psi := \{u \in \mathbb{R}^V \mid u_{i_0} = 0\}$. Note that $L$ is a symmetric positive linear bijection from $\Psi$ to $\Psi$ with respect to the inner product $(u, v)_L = u^T L v$. Consider the following problem.

Problem 3 Recover the solution $u \in \Psi$ of

$$Lu = f \quad (6.2)$$

as accurately as possible in the energy norm, defined as $\|u\|_L^2 = (u, u)_L$, given the information that $f \in \Psi$ with $\|f\|_2 \leq M$ and the observation $\eta = u + \zeta$ where $\zeta$ is a centered Gaussian vector on $\Psi$ with covariance matrix $\sigma^2$.

6.2 Near minimax recovery

Similarly to Sect. 4.1, with $M > 0$, define

$$V_M = \{u \in \Psi \mid \|Lu\|_2 \leq M\}. \quad (6.3)$$

With gamblets as defined in Sect. 3 (Owhadi and Scovel 2017, 2020), we provide sufficient conditions under which the matrices $B^{(k)}$ and $Z$ are uniformly well conditioned (as in Theorem 2), in particular under these conditions there exists $H \in (0, 1)$ such that $\text{Cond}(B^{(k)}) \leq C H^{-2}$ and

$$C^{-1}H^{-2(k-1)} f^{(k)} \leq B^{(k)} \leq C H^{-2k} f^{(k)}. \quad (6.4)$$

In this brief outlook we will simply numerically estimate $H$ and define $d$ to be such that $H^{-d} = \Theta(\mathcal{J}^{(k)})$. Assuming $\sigma > 0$, set

$$l^\dagger = \arg\min_{l \in [0, \ldots, q]} \beta_l, \quad (6.5)$$

for

$$\beta_l = \begin{cases} H^2 M^2 & \text{if } l = 0 \\ H^{-(2+d)l} \sigma^2 + H^{2(l+1)} M^2 & \text{if } 1 \leq l \leq q - 1 \\ H^{-(2+d)q} \sigma^2 & \text{if } l = q. \end{cases} \quad (6.6)$$

Finally, we use hard thresholding (5.1) and

$$v^\dagger(\eta) := \eta(l^\dagger), \quad (6.7)$$

to recover $u$ from $\eta$.

6.3 Numerical illustration

For our numerical illustration, we choose the edges and vertices of the graph are the streets and intersections of Pasadena, CA (obtained using the python OSMNX package, Boeing 2017). The latitude and longitude of each vertex is known through the package, and we (arbitrarily) select $f$ as $f(v_i) = \cos(3x(v_i) + y(v_i)) + \sin(3y(v_i)) + \sin(7x(v_i) - 5y(v_i))$, where $x(v_i)$ and $y(v_i)$ are the latitudes and longitudes of vertex $v_i$ normalized to $[0, 1]$. The noise, $\zeta$, is a centered Gaussian vector on $\Psi$ with $\sigma = 400$. Figure 6 shows (6.7), and the recovery error $v^\dagger(\eta) - u$.

The measurement functions used for the gamblet decomposition are defined as

$$\phi_{i,j}^{(k)}(u) = \frac{1}{\sqrt{|S_{i,j}^{(k)}|}} \sum_{l \in S_{i,j}^{(k)}} u_l \quad (6.8)$$

where $S_{i,j}^{(k)}$ contain all vertices with relative latitudes in $[\frac{i-1}{2^k}, \frac{i}{2^k}]$ and relative longitudes in $[\frac{j-1}{2^k}, \frac{j}{2^k}]$. The $W^{(k)}$ matrix are as in Sect. 2.4. Further, in this graph, we estimate $H \approx 0.5996$ and $d \approx 2.327$.

The figures show $f$, $u$, $\eta$, $v^\dagger(\eta)$ (using Average errors were computed using 100 independent noise samples. The de-noising errors of the hard gamblet thresholding with fixed thresholds ($t^{(k)} = t$ for all $k$) is compared to the near minimax recovery analog, $v^\dagger$, is given in Table 3.

For reference, the average and standard deviation of the $L$-energy norm of $\zeta$ used in this trial were 49,265 and 576 respectively.
Fig. 6 The plot of \(f, u, \eta, v(\eta), \) and \(v(\eta) - u\) on the Pasadena graph

Table 3 Comparisons between de-noising algorithms

| Algorithm               | L Error AVG | L Error STDEV | L\(^2\) Error AVG | L\(^2\) Error STDEV |
|-------------------------|-------------|---------------|--------------------|----------------------|
| Estimator (6.7)         | 578.3       | 11.6          | 6216               | 234                  |
| Hard variable threshold | 541.6       | 40.7          | 5644               | 446                  |

7 Proof of near minimax recovery in energy norm

This section will provide a proof of Theorem 4. Throughout this section we will use the pre-Haar wavelets of Sect. 2.4 as measurement functions \(\phi_i^{(k)}\).

7.1 Bounds on the covariance matrix of the noise in the gamblet basis

For \(s < k\), write

\[\pi^{(s,k)} := \pi^{(s,s+1)}\pi^{(s+1,s+2)}\ldots\pi^{(k-1,k)}\]  

(7.1)

and let \(\pi^{(k,s)} := (\pi^{(s,k)})^T\). Let \(\pi^{(k,k)} := I^{(k)}\). Let \(Z\) be the \(J \times J\) matrix (recall that \(J\) is the union of \(J^{(k)}\), the index set for \(k\)th level gamblet wavelets, for \(1 \leq k \leq q\)) defined by

\[Z_{i,j} = (N^{(s)}_i, T^{(s,k)} \pi^{(s,k)} N^{(k)}_{j})_{i,j}\]  

(7.2)

for \(i \in J^{(s)}\) and \(j \in J^{(k)}\).

Observe that by linearity

\[[\phi^k, \eta] = [\phi^k, u] + [\phi^k, \xi].\]  

(7.3)

where \([\phi^k, \xi]\) is a centered Gaussian vector whose covariance matrix is \(\sigma^2 Z\) as shown in the following lemma.

Lemma 8 It holds true that

\[\mathbb{E}[\phi^k, \xi][\phi^k, \xi]^T] = \sigma^2 Z.\]  

(7.4)
Furthermore, for $x \in \mathbb{R}^J$,
\[
x^T Z x = \| x \cdot \phi^k \|^2_{L^2(\Omega)},
\]
(7.5)

Proof. $E[\phi_i(q), \zeta \phi_j(q), \zeta] = \sigma^2(\phi_i(q), \phi_j(q))_{L^2(\Omega)} = \sigma^2 \delta_{i,j}$ implies
\[
E\left[ \phi_i^{(k)}(x), \zeta \phi_j^{(k)}(x), \zeta \right] = (N^{(k)}, T N^{(k)})_{3,5} \sigma^2
\]
(7.6)

Therefore, $\phi_i^{(k)} = \pi^{(s, k)} \phi_i(q)$ for $1 \leq s < k \leq q$ implies
\[
E\left[ \phi_i^{(k)}(x), \zeta \phi_j^{(k)}(x), \zeta \right] = (N^{(s)}, T \pi^{(s, k)} N^{(k)})_{3,5} \sigma^2
\]
(7.7)
for $i \in \mathcal{J}^{(s)}$ and $j \in \mathcal{J}^{(k)}$. The proof of (7.5) is identical. \hfill \Box

There exists a large literature comparing thresholding and component filtering with minimax signal recovery. Most rigorous results make the assumption that the noise in the decomposition [i.e. $\epsilon \xi$ in $y_j = \theta_j + \epsilon \xi$ as in Donoho (1995) and Donoho et al. (1996)] is i.i.d. $\mathcal{N}(0, \sigma^2)$. Although the situation in the gamblet decomposition is slightly different (since the entries of $[\phi^{(k)}]_j \zeta$ may be correlated and non-identically distributed), the uniform bound $I_d \leq Z \leq (1 + C)I_d$ obtained in Theorem 10 will be sufficient to prove near minimax recovery in energy norm. To obtain Theorem 10 we will first need the following lemma.

Lemma 9 It holds true that for $k \in \{2, \ldots, q\}$, $z^{(k)} \in \mathbb{R}^{T^{(k)}}$ and $y^{(k-1)} \in \mathbb{R}^{T^{(k-1)}}$
\[
|z^{(k)}|^2 + |y^{(k-1)}|^2 \leq |\pi^{(k, k-1)} y^{(k-1)} + N^{(k)} z^{(k)}|^2
\]
\[
\leq (1 + C)|z^{(k)}|^2 + |y^{(k-1)}|^2
\]
(7.8)
where $C$ depends only on $\|L\|, \|L^{-1}\|, \Omega, s, d$ and $\delta$.

Proof. We start with the argument of the proof of Owhadi and Scovel (2020, Lem. 13.27).

Since $\text{Im}(W^{(k), T})$ and $\text{Im}(\pi^{(k, k-1)})$ are orthogonal and
\[
\dim(\mathbb{R}^{T^{(k)}}) = \dim \left( \text{Im}(W^{(k), T}) \right) + \dim \left( \text{Im}(\pi^{(k, k-1)}) \right),
\]
(7.9)
for $x \in \mathbb{R}^{T^{(k)}}$ there exists a unique $z \in \mathbb{R}^{T^{(k)}}$ and $y \in \mathbb{R}^{T^{(k-1)}}$ such that
\[
x = W^{(k), T} z + \pi^{(k, k-1)} y.
\]
(7.10)

$W^{(k), \pi^{(k, k-1)}, T} = 0$ and $W^{(k), W^{(k), T}} = J^{(k)}$ implies $W^{(k), T} x = z$, $R^{(k-1), \pi^{(k, k-1)}, T} = I^{(k-1)}$ implies $R^{(k-1), k} x = R^{(k-1), k} y$. Writing
\[
P^{(k)} = \pi^{(k, k-1)} R^{(k-1), k},
\]
(7.11)
observes that $P^{(k)}$ is a projection (since $(P^{(k)})^2 = P^{(k)}$) and
\[
x = W^{(k), T} W^{(k), T} x + P^{(k)} (I^{(k)} - W^{(k), T} W^{(k)}) x.
\]
(7.12)

Observe that $N^{(k)} = A^{(k), W^{(k), T} B^{(k-1), T}}$ implies that $W^{(k)} N^{(k)}$ $N^{(k)} = J^{(k)}$ and $P^{(k)} N^{(k)} = \pi^{(k, k-1)} A^{(k-1), k} \pi^{(k-1, k)} Q^{(k)} N^{(k)} = 0$. Therefore, taking $x = N^{(k)} z^{(k)}$ in (7.12) implies
\[
N^{(k)} z^{(k)} = W^{(k), T} z^{(k)} - P^{(k)} W^{(k), T} z^{(k)}.
\]
(7.13)

Using $P^{(k), \pi^{(k, k-1)}} = \pi^{(k, k-1)}$ we deduce that
\[
\pi^{(k, k-1)} y^{(k-1)} + N^{(k)} z^{(k)} = W^{(k), T} z^{(k)} + P^{(k)} (\pi^{(k, k-1)} y^{(k-1)} - W^{(k), T} z^{(k)}).
\]
(7.14)

Using $W^{(k), P^{(k)}} = 0$ we deduce that
\[
|\pi^{(k, k-1)} y^{(k-1)} + N^{(k)} z^{(k)}|^2
\]
\[
= |z^{(k)}|^2 + |P^{(k)} (\pi^{(k, k-1)} y^{(k-1)} - W^{(k), T} z^{(k)})|^2.
\]
(7.15)

Using $P^{(k), \pi^{(k, k-1)}} = \pi^{(k, k-1)}$, $W^{(k), P^{(k)}} = 0$ and $W^{(k), \pi^{(k, k-1)}} = 0$ we have
\[
|P^{(k)} (\pi^{(k, k-1)} y^{(k-1)} - W^{(k), T} z^{(k)})|^2
\]
\[
= |y^{(k-1)}|^2 + |P^{(k)} W^{(k), T} z^{(k)}|^2.
\]
(7.16)

Summarizing we have obtained that
\[
|\pi^{(k, k-1)} y^{(k-1)} + N^{(k)} z^{(k)}|^2
\]
\[
= |z^{(k)}|^2 + |y^{(k-1)}|^2 + |P^{(k)} W^{(k), T} z^{(k)}|^2.
\]
(7.17)

(Owhadi and Scovel 2020, Lem. 14.7) and (Owhadi and Scovel 2020, Lem. 14.53) imply that $\|P^{(k)}\|_2 \leq C$ which concludes the proof. \hfill \Box

Theorem 10. It holds true that for $z \in \mathbb{R}^J$,
\[
|z|^2 \leq \|z \cdot \phi^k\|^2_{L^2(\Omega)} \leq (1 + C)|z|^2
\]
(7.18)
where $C$ depends only on $\|L\|, \|L^{-1}\|, \Omega, s, d$ and $\delta$. In particular,
\[
I_d \leq Z \leq (1 + C) I_d.
\]
(7.19)

Proof. (7.19) follows from (7.5) and (7.18). To prove (7.18) write $z = (z^{(1)}, \ldots, z^{(q)})$ with $z^{(k)} \in \mathbb{R}^{T^{(k)}}$. Observe that
\[
\|z \cdot \phi^k\|^2_{L^2(\Omega)} = |\pi^{(q, 1)} z^{(1)}|^2 + \sum_{k=2}^{q} |\pi^{(q, k)} N^{(k)} z^{(k)}|^2
\]
(7.20)
Therefore
\[ \|z \cdot \phi^T\|_{L^2(\Omega)}^2 = \|\pi^{(q,q-1)} y^{(q-1)} + N^{(q)} z^{(q)}\|^2 \] (7.21)
with
\[ y^{(q-1)} = \pi^{(q-1,1)} z^{(1)} + \sum_{k=2}^{q-1} \pi^{(q-1,k)} N^{(k)} z^{(k)} \]
(7.22)

Using Lemma 9 with \( k = q \) implies that
\[ |z^{(q)}|^2 + |y^{(q-1)}|^2 \leq |\pi^{(q,q-1)} y^{(q-1)} + N^{(q)} z^{(q)}|^2 \]
\[ \leq (1 + C) |z^{(q)}|^2 + |y^{(q-1)}|^2 \] (7.23)

We conclude the proof of (7.18) by a simple induction using Lemma 9 iteratively. For the sake of clarity we will write the next step of this iteration. We have
\[ y^{(q-1)} = \pi^{(q-1,q-2)} y^{(q-2)} + N^{(q-1)} z^{(q-1)} \] (7.24)
with
\[ y^{(q-2)} = \pi^{(q-2,1)} z^{(1)} + \sum_{k=2}^{q-2} \pi^{(q-2,k)} N^{(k)} z^{(k)} \]
(7.25)
and Lemma 9 with \( k = q - 1 \) implies
\[ |z^{(q-1)}|^2 + |y^{(q-2)}|^2 \leq |y^{(q-1)}|^2 \]
\[ \leq (1 + C) |z^{(q-1)}|^2 + |y^{(q-2)}|^2. \] (7.26)

\[ \square \]

7.2 Near minimax recovery results

For \( T \in \{0, 1\}^J \), write
\[ v_T(\eta) := \sum_{i \in J} T_i [\phi_i^T, \eta] \chi_i. \] (7.27)

Let \( M > 0 \) and write \( V_M^{(q)} := \{ u \in \Psi^{(q)} | \|Lu\|_{L^2(\Omega)} \leq M \} \). Define \( T_i^\dagger \in \{0, 1\}^J \) with \( T_i^\dagger = 1 \) if and only if \( i \in J^{(k)} \) for \( k \leq l^\dagger \) with
\[ l^\dagger = \arg \min_{l \in \{0, \ldots, q\}} \beta_l, \] (7.28)
for
\[ \beta_l = \begin{cases} h^{2s} M^2 & \text{if } l = 0 \\ \sigma^2 h^{-(2s+d)+l} + h^{2s(1+l)} M^2 & \text{if } 1 \leq l \leq q - 1 \\ h^{-(2s+d)+q} \sigma^2 & \text{if } l = q. \end{cases} \] (7.29)

We will first prove the following theorem showing the near optimality of \( v_T^\dagger \) in the class of estimators of the form (7.27).

**Theorem 11** It holds true that
\[ \sup_{u \in V_M^{(q)}} \mathbb{E} \left[ \|u - v_{T^\dagger}(\eta)\|^2 \right] < C \inf_{T \in \{0, 1\}^J} \sup_{u \in V_M^{(q)}} \mathbb{E} \left[ \|u - v_T(\eta)\|^2 \right], \] (7.30)
where \( C \) depends only on \( h, s, \|\mathcal{L}\|, \|\mathcal{L}^{-1}\|, \Omega, d, \) and \( \delta \).

**Proof** To simplify notations, we will write \( C \) for any constant depending only on \( h, s, \|\mathcal{L}\|, \|\mathcal{L}^{-1}\|, \Omega, d, \) and \( \delta \) (therefore \( Ch^{-2s}(1 + C) \) will still be written \( C \)). Begin by writing
\[ [\phi^T, u - v_T(\eta)] = (1 - T_i)[\phi_i^T, \eta] u^2 + T_i [\phi_i^T, \eta] \]
(7.31)

Abusing notations and writing \( T \) for the \( \mathcal{J} \times \mathcal{J} \) diagonal matrix with entries \( T_{i,j} = T_i \), we have for \( u \in \Psi^{(q)} \),
\[ \|u - v_T(\eta)\|^2 = ((1 - T) [\phi^T, u] - T [\phi^T, \xi])^T B ((1 - T) [\phi^T, u] - T [\phi^T, \xi]). \] (7.32)

Therefore the bounds (2.24) on the (diagonal) blocks of \( B \) imply that (where \( C \) absorbs a factor of \( h^{-2s} \))
\[ \inf_T \mathbb{E} \left[ \|u - v_T(\eta)\|^2 \right] \geq C^{-1} \inf_T \sup_u \sum_{k=1}^{q} h^{-2sk} \sum_{i \in J^{(k)}} \mathbb{E}[(7.31)]. \] (7.33)

Theorem 10 implies
\[ \sigma^2 \leq \mathbb{E} \left[ [\phi_i^{(k)}]^{\chi_i} \right] \leq (1 + C) \sigma^2, \] (7.34)
and therefore
\[ \mathbb{E}[(7.31)] \geq (1 - T_i) [\phi_i^{(k)}]^{\chi_i} u^2 + T_i \sigma^2, \] (7.35)
which implies
\[ \inf_T \mathbb{E} \left[ \|u - v_T(\eta)\|^2 \right] \geq C^{-1} \inf_T \sup_u \sum_{k=1}^{q} h^{-2sk} \sum_{i \in J^{(k)}} (7.35). \] (7.36)

Next, if \( |\cdot|^2 \) is the \( l^2 \) norm on \( \mathbb{R}^J \), \( Lu = [\phi^T, u] \cdot B \phi^T \) and Theorem 10 imply
\[ \|\phi^T, u\|^T B\|_2^2 \leq \|Lu\|^2 \leq (1 + C)[\phi^T, u] B\|_2^2. \] (7.37)
We deduce that $cV^{(q)}_M \subset V^{(q)}_M \subset \tilde{V}^{(q)}_M$ with

$$V^{(q)}_M := \left\{ u \in \Psi^{(q)} | \sum_{k=1}^{q} h^{-4k} \| \phi^{(k)} \cdot u \|_2^2 \leq \frac{M^2}{C} \right\}$$

(7.38)

and

$$\tilde{V}^{(q)}_M := \left\{ u \in \Psi^{(q)} | \sum_{k=1}^{q} h^{-4k} \| \phi^{(k)} \cdot u \|_2^2 < C M^2 \right\}.$$  

(7.39)

We deduce that

$$\inf_{T} \sup_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2] \geq C^{-1} \inf_{T} \sup_{u \in Y^{(q)}_M} \sum_{k=1}^{q} h^{-2sk} \sum_{i \in \mathcal{J}^{(k)}} (7.35).$$

(7.40)

Let us now show that for some $0 \leq l \leq q$, the minimizer $T^{(l),*}$ of the right-hand side of (7.40) satisfies $T^{(l),*} = 1$ for all $i \in \mathcal{J}^{(k)}$ with $k \leq l$ and $T^{(l),*} = 0$ otherwise. For $l \in \{0, \ldots, q-1\}$, write

$$F^{(l)} := \{ T \in \{0,1\}^{\mathcal{J}} | T_i = 1 \text{ for } i \in \mathcal{J}^{(k)} \text{ with } k \leq l, \text{ and } \exists i \in \mathcal{J}^{(l+1)} \text{ s.t. } T_i = 0 \},$$

(7.41)

and write $F^{(q)} := \{1\}^{\mathcal{J}}$. Let us now identify the minimizer in $T$ of

$$\sup_{u \in V^{(q)}_M} \sum_{k=1}^{q} h^{-2sk} \sum_{i \in \mathcal{J}^{(k)}} ((1 - T_i) \| \phi^{(k)} \|_2^2 + T_i \sigma^2).$$

(7.42)

Notice that the $u$ which maximizes (7.42) can be found by maximizing

$$\sum_{k=1}^{q} h^{-2sk} \sum_{i \in \mathcal{J}^{(k)}} (1 - T_i) \| \phi^{(k)} \|_2^2$$

over $u \in \Psi^{(q)}$ such that

$$\sum_{k=1}^{q} h^{-4ks} \sum_{i \in \mathcal{J}^{(k)}} \| \phi^{(k)} \|_2^2 < \frac{M^2}{C}.$$ 

(7.43)

Writing $y_i := h^{-2sk} \| \phi^{(k)} \|_2^2$ over $y \in \mathbb{R}^{\mathcal{J}}$ such that $\|y\|^2 \leq \frac{M^2}{C}$, for $T \in F^{(l)}$ with $1 \leq l \leq q-1$ the maximum is achieved by taking $y_i = M/\sqrt{C}$ for a single $i \in \mathcal{J}^{(l+1)}$ such that $T_i = 0$ and is equal to $h^{2s(l+1)} M^2/C$. Therefore, using $C \sum_{k=1}^{l} h^{-2sk} \| \mathcal{J}^{(k)} \| \geq h^{-2d+1/l}$ we deduce that

$$\inf_{T \in F^{(l)}} \sup_{u \in V^{(q)}_M} \sum_{k=1}^{q} h^{-2sk} \sum_{i \in \mathcal{J}^{(k)}} (7.35) \geq \frac{h^{2s(l+1)} M^2 + h^{-2d+l} \sigma^2}{C}.$$ 

(7.45)

For notational convenience, define the sequence $\beta_i$ to be the left-hand side of (7.45) for $0 \leq l \leq q$. Further, if $l = 0$, then $\beta_0 \geq h^{2s} M^2/C$ with optimum filter $T = \{0\}^{\mathcal{J}}$ and if $l = q$, by definition, $T = \{1\}^{\mathcal{J}}$. Hence $\beta_q \geq h^{2s+q} \sigma^2/C$. Therefore, (7.40) implies

$$\inf_{T} \sup_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2] \geq \min \beta_i.$$ 

(7.46)

Similarly using (7.32), (7.34) and the bounds (2.24) on the (diagonal) blocks of $B$ and (7.39) imply that

$$\sup_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2] \leq C \sup_{u \in V^{(q)}_M} \sum_{k=1}^{q} h^{-2sk} \sum_{i \in \mathcal{J}^{(k)}} (7.35).$$

(7.47)

Let $\tilde{\beta}_l$ be the value of the right-hand side of (7.47) for $T = T^{(l),*}$. By the same arguments as those following (7.43), we obtain that, $\tilde{\beta}_l \leq C(h^{2s(l+1)} M^2 + h^{-2d+l} \sigma^2)$. Hence, for some fixed constant, $\tilde{\beta}_l \leq C \beta_l$ for all $l$. Therefore taking $T = T^{*}$ and bounding the right-hand side of (7.47) as in (7.42) leads to

$$\sup_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2] \leq \min \beta_l \leq C \min \beta_l \leq C \inf_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2].$$

(7.48)

which concludes the proof. \[\square\]

The following theorem shows that gambit filtering yields a recovery that is near optimal (in energy norm and in the class of all estimators) up to a constant.

**Theorem 12** Suppose that $T \in \{0,1\}^{\mathcal{J}}$. Then the following holds for a constant $C$ dependent on $h$, $s$, $\|\mathcal{C}\|$, $\|\mathcal{C}^{-1}\|$, $\Omega$, $d$, and $\delta$:

$$\inf_{T \in \{0,1\}^{\mathcal{J}}} \sup_{u \in V^{(q)}_M} \mathbb{E}[\|u - v_T(\eta)\|^2] \leq C \inf_{v(\eta) \in V^{(q)}_M} \mathbb{E}[\|u - v(\eta)\|^2].$$

(7.49)
where the infimum on the right-hand side is taken over all measurable functions \( v : \Psi^{(q)} \rightarrow \Psi^{(q)} \).

**Proof** Recalling that

\[
\mathbb{E}[\|u - v_{T}(n)\|^2] = \mathbb{E}[(T[\phi^X, \xi] - (I - T)[\phi^X, u])^T B(T[\phi^X, \xi] - (I - T)[\phi^X, u])],
\]

the bounds (2.24) on the (diagonal) blocks of \( B \) imply that

\[
(7.50) \leq CE \left[ \sum_{k} \sum_{i} \left( T_{i} h^{-2sk}(1 + C)\sigma^2 + (1 - T_{i}) (h^{-2sk} [\phi_{i}^{(k),X}, u])^2 \right) \right].
\]

(7.51)

We will now introduce preliminary results associated with the problem of recovering \( \theta \in \mathbb{R}^{\mathcal{J}} \) from the observation of \( y = \theta + z \) where \( z \) is a centered Gaussian vector with independent entries \( z_{i} \sim \mathcal{N}(0, \sigma_{i}^{2}) \). Estimators in this \( \mathbb{R}^{\mathcal{J}} \) space will be written \( \hat{\theta} \) (we use \( v \) when the problem is formulated in \( \Psi^{(q)} \)). For \( T \in (0,1)^{\mathcal{J}} \) define the estimator \( \hat{\theta}_{T} \) via

\[
(7.52)
\]

Then, writing \( | \cdot |_{2} \) for the \( l^{2} \) (Euclidean) norm in \( \mathbb{R}^{\mathcal{J}} \), the expected error of the recovery is:

\[
\mathbb{E}[|\theta - \hat{\theta}_{T}(y)|_{2}^2] = \mathbb{E} \left[ \sum_{i} \left( T_{i} z_{i}^2 + (1 - T_{i}) \theta_{i}^2 \right) \right].
\]

(7.53)

Let \( \hat{\theta}(u) \) be the \( \mathbb{R}^{\mathcal{J}} \) vector defined by \( \hat{\theta}_{i}(u) = h^{-sk}[\phi_{i}^{(k),X}, u] \) for \( i \in \mathcal{J} \). Let \( V^{(q)}_{M} \subset V^{(q)}_{M} \) be defined as in (7.39) and write \( \hat{\theta}(V^{(q)}_{M}) \) for the image of \( V^{(q)}_{M} \) under the map \( \hat{\theta} \). Consider the problem of recovering \( \theta \in \hat{\theta}(V^{(q)}_{M}) \) from the observations \( y'_{i} = \theta_{i} + z'_{i} \) where the \( z'_{i} \) are independent centered Gaussian random variables with variance \( \text{Var}(z'_{i}) = (1 + C) h^{-2sk} \) for \( i \in \mathcal{J}^{(k)} \). We will need the following lemma which is directly implied by Ding and Mathé (2017, Lem. 2).

**Lemma 13** Let \( \mathcal{J} \) be an index set and for \( \theta \in \mathbb{R}^{\mathcal{J}} \) let \( y \) be the noisy observation of \( \theta \) defined by \( y = \theta + z \) where \( z \) is a Gaussian vector with independent entries \( z_{i} \sim \mathcal{N}(0, \sigma_{i}^{2}) \). For \( a \in \mathbb{R}^{\mathcal{J}} \setminus \{0\} \) and \( M > 0 \) write

\[
\Theta_{a}(M) := \left\{ \theta \in \mathbb{R}^{\mathcal{J}} \mid \sum_{i} \theta_{i}^2 \leq M^2 \right\}.
\]

(7.54)

Then for \( \hat{\theta}_{T} \) defined as in (7.52), it holds true that

\[
\inf_{T \in (0,1)^{\mathcal{J}}} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}_{T}(y) - \theta|_{2}^2]
\]

\[
\leq 4.44 \inf_{\hat{\theta}(y) \in \hat{\theta}(V^{(q)}_{M})} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}(y) - \theta|_{2}^2]
\]

(7.55)

where the infimum in the right-hand side of (7.55) is taken over all functions \( \hat{\theta} \) of \( y \).

Lemma 13 implies that

\[
\inf_{T} \sup_{\hat{\theta}(y) \in \hat{\theta}(V^{(q)}_{M})} \mathbb{E}[|\hat{\theta}(y) - \theta|_{2}^2] \leq 4.44 \inf_{\hat{\theta}(y) \in \hat{\theta}(V^{(q)}_{M})} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}(y) - \theta|_{2}^2].
\]

(7.56)

Let \( C_{0} > 1 \) be a constant larger than \((1 + C)h^{-1} \) [with \( C \) being the constant in (2.24)] and also larger than \( \sqrt{1 + C} \) [with \( C \) being the constant in (7.18)]. Then, it is true that

\[
\hat{\theta}(V^{(q)}_{M}) \subset \hat{\theta}(V^{(q)}_{M}) \subset \hat{\theta}(V^{(q)}_{M})
\]

(7.57)

Further, for \( \theta \in \mathbb{R}^{\mathcal{J}} \) let \( y'' = \theta + z'' \) be the noisy observation of \( \theta \) where \( z'' \) is a centred Gaussian vector with independent entries of variance \( \text{Var}(z''_{i}) = C_{0}^{-2} \text{Var}(z'_{i}) \). We deduce that

\[
\inf_{\hat{\theta}(y'') \in \hat{\theta}(V^{(q)}_{M})} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}(y'') - \theta|_{2}^2] = C_{0}^{2} \inf_{\hat{\theta}(y'') \in \hat{\theta}(V^{(q)}_{M})} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}(y') - \theta|_{2}^2].
\]

(7.58)

Define the set of affine recoveries to be of the form \( \hat{\theta}_{A}(y) = Ty + y_{0} \) with \( T \) linear and \( y_{0} \in \mathbb{R}^{\mathcal{J}} \). Since this is a subset of all recoveries \( \hat{\theta} \) (and replacing \( C_{0} \) with \( C \)), it holds true that

\[
(7.59)
\]

If we define \( \theta_{i} := h^{-sk}[\phi_{i}^{(k),X}, u], z_{i} := h^{-sk}[\phi_{i}^{(k),X}, \xi], \) and \( y_{i} = \theta_{i} + z_{i}, \) it is true that \( \text{Cov}(z) > \text{Cov}(z'') \). Since, in the space of affine recoveries, the recovery error increases with the covariance matrix, and since \( \hat{\theta}(V^{(q)}_{M}) \subset \hat{\theta}(V^{(q)}_{M}) \), we have

\[
(7.59) \leq C \inf_{\hat{\theta}_{A}(y) \in \hat{\theta}(V^{(q)}_{M})} \sup_{\theta \in \Theta_{a}(M)} \mathbb{E}[|\hat{\theta}(y) - \theta|_{2}^2]
\]

(7.60)
The following lemma is a direct application of Donoho (1994, Cor. 1) (with $K = I$ and $L = I$, see also the remarks in Donoho (1994, sec. 13.1)). It will be used to compare affine and general minimax recovery errors.

**Lemma 14** Let $\mathcal{J}$ be an index set and $z$ be a centered Gaussian vector of $\mathbb{R}^{\mathcal{J}}$ with Cov($z$) = $Z$ where $Z$ is invertible. Let $V$ be a closed, bounded, and convex subset and for $\theta \in V$ let $y = \theta + z$ be the noisy observation of $\theta$. It holds true that

$$
\inf_{\hat{\theta} \in V} \sup_{\theta \in V} \mathbb{E}[|\hat{\theta}(y) - \theta|^2] \leq 1.25 \inf_{\theta \in V} \mathbb{E}[|\hat{\theta}(y) - \theta|^2].
$$

(7.61)

Since ellipsoids are closed, bounded, and convex, we can apply the lemma and bounds on the (diagonal) blocks of $B$ in (2.24) to obtain the desired result,

$$
(7.60) \leq C \inf_{\theta \in V} \sup_{\hat{\theta} \in V} \mathbb{E}[|\hat{\theta}(y) - \theta|^2] 
\leq C h^{-2s} \inf_{\hat{\theta} \in V} \mathbb{E}[|\hat{\theta}(y) - \theta|^2].
$$

(7.62)

\[\square\]

### 7.3 Proof of smooth recovery

The final statement of Theorem 4 will be proven in this section. First, the following lemma will be proven.

**Lemma 15** Let $T^\dagger$ be as in Theorem 4 such that $T^\dagger \neq 0$, then it holds true that

$$
||v_{T^\dagger}(\eta)|| \leq ||u|| + C \sqrt{x} \sigma \frac{2 \epsilon}{\delta} M \frac{2 + d}{-\epsilon} + d
$$

(7.63)

with probability at least $1 - (z e^{1-\epsilon})|\mathcal{J}(T^\dagger)|/2$ for $C$ a constant dependent on $h$, $s$, $||\mathcal{L}||$, $||\mathcal{L}^{-1}||$, $\Omega$, $d$, and $\delta$.

**Proof** Start by observing that since $v_{T^\dagger}(\eta) = u(T^\dagger) + \xi(T^\dagger)$,

$$
||v_{T^\dagger}(\eta)|| \leq ||u(T^\dagger)|| + ||\xi(T^\dagger)|| \leq ||u|| + ||\xi(T^\dagger)||.
$$

(7.64)

Next, the latter part of the sum on the right-hand side will be bounded.

$$
||\xi(T^\dagger)||^2 \leq \sum_{k=1}^{l} (1 + C) h^{-2s} k \sum_{k=1}^{l} \phi_{i}^{(k)} \xi^2 
\leq (1 + C) h^{-2s} \sum_{k=1}^{l} \sum_{i=1}^{l} \phi_{i}^{(k)} \xi^2. 
$$

(7.65)

Defining the Gaussian vector $X$ such that $X_i = \phi^{(k)} \xi$ for $i \in \mathcal{J}(k)$ for $k = \{0, \ldots, l\}$, it holds that Cov($X$) < $\sigma^2 I$ by Theorem 10. Defining Gaussian vector $\tilde{X}$ such that Cov($\tilde{X}$) = Cov($X$), it holds that $M := \sigma \sqrt{1 + C} Cov(X)^{-1/2} > 1$. Note that since $P[|MX|^2 \geq |X|^2] = 1$ and $MX$ has an identical distribution to $\tilde{X}$, it is true that

$$
P[|X|^2 < x] \leq P[|\tilde{X}|^2 < x] = P[|\tilde{X}|^2 < x].
$$

(7.66)

Hence, to get a tail bound of (7.65), we apply a Chernoff bound on $\chi^2$ distributions, which states for $z > 1$ the CDF of $Q_k$, a $\chi^2(k)$-distribution, is bounded by

$$
P(Q_k < zk) \geq 1 - (ze^{1-\epsilon})^{k/2}.
$$

(7.67)

This bound can be deduced from Lemma 2.2 of Dasgupta and Gupta (2003). Hence, with probability at least $1 - (ze^{1-\epsilon})|\mathcal{J}(T^\dagger)|/2$ ($C$ absorbs a constant dependent on $h$, $s$, and $d$)

$$
(7.65) \leq C \sigma^2 h^{-2s} M \frac{2 + d}{-\epsilon} + d
\leq C \frac{\sigma \log(1 + C)}{\delta} M \frac{2 + d}{-\epsilon} + d
$$

(7.68)

\[\square\]

**Theorem 16** Assuming $T^\dagger \neq 0$, the following inequality holds with probability at least $1 - \epsilon$,

$$
||\eta(T^\dagger)|| \leq ||u|| + C \sqrt{\log(1 + C)} \sigma \frac{2 \epsilon}{\delta} M \frac{2 + d}{-\epsilon} + d,
$$

(7.69)

where $C$ depends only on $h$, $d$, $\delta$ and $s$.

**Proof** Given Lemma 15, for $z > 1$, we begin by defining

$$
\epsilon = (z e^{1-\epsilon})|\mathcal{J}(T^\dagger)|/2.
$$

(7.70)

Note that for any value of $\epsilon \in (0, 1)$, there exists a unique solution $z > 1$. Therefore, we have $ze^{1-\epsilon} \leq 2e^{-\epsilon/2}$. Moreover,

$$
\frac{4}{\log(2)} \log(1 + C) \sigma \frac{2 \epsilon}{\delta} M \frac{2 + d}{-\epsilon} + d
$$

(7.71)

Hence, for $C$ depending only on $h$, $d$, $\delta$ and $s$,

$$
C \log(1 + C) \sigma \frac{2 \epsilon}{\delta} M \frac{2 + d}{-\epsilon} + d
$$

(7.72)

This yields an upper bound on $z$ and we apply Lemma 15 to obtain the result:

$$
(7.68) \leq C \sigma \log(1 + C) \sigma \frac{2 \epsilon}{\delta} M \frac{2 + d}{-\epsilon} + d.
$$

(7.73)

\[\square\]
Acknowledgements

The authors gratefully acknowledges this work supported by the Air Force Office of Scientific Research and the DARPA EQUiPS Program under Award Number FA9550-16-1-0054 (Computational Information Games) and the Air Force Office of Scientific Research under Award Number FA9550-18-1-0271 (Games for Computation and Learning). We also thank two anonymous referees for detailed reviews and helpful comments.

References

Boeing, G.: OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–135 (2017)

Briol, F.X., Oates, C.J., Girolami, M., Osborne, M.A., Sejdinovic, D.: Probabilistic integration: a role for statisticians in numerical analysis? (2015). arXiv:1512.00933

Chkrebtii, O.A., Campbell, D.A., Calderhead, B., Girolami, M.A.: Bayesian solution uncertainty quantification for differential equations. Bayesian Anal. 11(4), 1239–1267 (2016)

Cockayne, J., Oates, C.J., Sullivan, T., Girolami, M.: Probabilistic meshless methods for Bayesian inverse problems (2016). arXiv:1605.07811

Cockayne, J., Oates, C., Sullivan, T., Girolami, M.: Bayesian probabilistic numerical methods (2017). arXiv:1702.03673

Dasgupta, S., Gupta, A.: An elementary proof of a theorem of Johnson and Lindenstrauss. Random Struct. Algorithms 22, 60–65 (2003)

Diaconis, P.: Bayesian numerical analysis. In: Berger, J., Gupta, S. (eds.) Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986), pp. 163–175. Springer, New York (1988)

Ding, L., Mathé, P.: Minimax rates for statistical inverse problems under general source conditions (2017). arxiv:1707.01706v2

Donoho, D.L.: Statistical estimation and optimal recovery. Ann. Stat. 22, 238–270 (1994)

Donoho, D.L.: De-noising by soft-thresholding. IEEE Trans. Inf. Theory 41(3), 613–627 (1995)

Donoho, D., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3), 425–455 (1994)

Donoho, D., Liu, K., MacGibbon, B.: Minimax risk over hyperrectangles and implications. Ann. Stat. 18(3), 1416–1437 (1990)

Donoho, D.L., Johnstone, I.M., et al.: Minimax estimation via wavelet shrinkage. Ann. Stat. 26(3), 879–921 (1998)

Gazzola, F., Grunau, H.C., Sweers, G.: Polyharmonic Boundary Value Problems: Positivity Preserving and Nonlinear Higher Order Elliptic Equations in Bounded Domains. Springer, Berlin (2010)

Hennig, P., Osborne, M.A., Girolami, M.: Probabilistic numerics and uncertainty in computations. Proc. R. Soc. A. 471(2179), 20150142 (2015). https://doi.org/10.1098/rspa.2015.0142

Michelli, C.A., Rivlin, T.J.: A survey of optimal recovery. In: Michelli, C.A., Rivlin, T.J. (eds.) Optimal Estimation in Approximation Theory, pp. 1–54. Springer (1977)

Oates, C., Cockayne, J., Aykroyd, R.G.: Bayesian probabilistic numerical methods for industrial process monitoring (2017). arXiv:1707.06107

Owhadi, H.: Bayesian numerical homogenization. Multiscale Model. Simul. 13(3), 812–828 (2015). https://doi.org/10.1137/140974596

Owhadi, H.: Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. SIAM Rev. 59(1), 99–149 (2017)

Owhadi, H., Scovel, C.: Universal scalable robust solvers from computational information games and fast eigenspace adapted multiresolution analysis (2017). arXiv:1703.10761

Owhadi, H., Scovel, C.: Operator Adapted Wavelets, Fast Solvers, and Numerical Homogenization from a Game Theoretic Approach to Numerical Approximation and Algorithm Design. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge (2020)

Owhadi, H., Zhang, L.: Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients. J. Comput. Phys. 347, 99–128 (2017)

Raisis, M., Perdikaris, P., Karniadakis, G.E.: Inferring solutions of differential equations using noisy multi-fidelity data. J. Comput. Phys. 335, 736–746 (2017)

Schäfer, F., Sullivan, T.J., Owhadi, H.: Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity (2017). arXiv:1706.02205

Schober, M., Duvenaud, D.K., Hennig, P.: Probabilistic ODE solvers with Runge–Kutta means. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems 27, pp. 739–747. Curran Associates, Inc., Red Hook (2014)

Woźniakowski, H.: Probabilistic setting of information-based complexity. J. Complex. 2(3), 255–269 (1986). https://doi.org/10.1016/0885-064X(86)90005-1

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.