Efficient regularization with wavelet sparsity constraints in photoacoustic tomography

To cite this article: Jürgen Frikel and Markus Haltmeier 2018 Inverse Problems 34 024006

View the article online for updates and enhancements.
Efficient regularization with wavelet sparsity constraints in photoacoustic tomography

Jürgen Frikel\textsuperscript{1} and Markus Haltmeier\textsuperscript{2}\textsuperscript{©}

\textsuperscript{1} Department of Computer Science and Mathematics, Universitätsstraße 31, D-93053 Regensburg, Germany
\textsuperscript{2} Department of Mathematics, University of Innsbruck, Technikstraße 13, A-6020 Innsbruck, Austria

E-mail: juergen.frikel@oth-regensburg.de and markus.haltmeier@uibk.ac.at

Received 28 March 2017, revised 28 September 2017
Accepted for publication 11 December 2017
Published 17 January 2018

Abstract
In this paper, we consider the reconstruction problem of photoacoustic tomography (PAT) with a flat observation surface. We develop a direct reconstruction method that employs regularization with wavelet sparsity constraints. To that end, we derive a wavelet-vaguelette decomposition (WVD) for the PAT forward operator and a corresponding explicit reconstruction formula in the case of exact data. In the case of noisy data, we combine the WVD reconstruction formula with soft-thresholding, which yields a spatially adaptive estimation method. We demonstrate that our method is statistically optimal for white random noise if the unknown function is assumed to lie in any Besov-ball. We present generalizations of this approach and, in particular, we discuss the combination of PAT-vaguelette soft-thresholding with a total variation (TV) prior. We also provide an efficient implementation of the PAT-vaguelette transform that leads to fast image reconstruction algorithms supported by numerical results.

Keywords: photoacoustic tomography, wavelet-vaguelette decomposition, variational regularization, sparsity constraints, image reconstruction

(Some figures may appear in colour only in the online journal)
1. Introduction

Photoacoustic tomography (PAT) is a novel coupled-physics (hybrid) modality for non-invasive biomedical imaging that combines the high contrast of optical tomography with the high spatial resolution of acoustic imaging \cite{7, 68, 102, 105, 107}. Its principle consists in illuminating a sample by an electromagnetic pulse that, due to the photoacoustic effect, generates pressure waves inside of the object; see figure 1. The generated pressure waves (the acoustic signals) then propagate through the sample and beyond, and the pressure is recorded outside of the sample. Finally, mathematical algorithms are used to reconstruct an image of the interior (see, for example \cite{69, 86, 107}).

In this paper, we work with the standard model of PAT, where the acoustic pressure $u: \mathbb{R}^d \times (0, \infty) \to \mathbb{R}$ solves the standard wave equation (see (3)). The goal of PAT is to recover the initial pressure distribution (at time $t = 0$), given by a function $h: \mathbb{R}^d \to \mathbb{R}$, from measurements $U_h: = u|_{\partial \mathbb{H}_+ \times (0, \infty)}$ of the pressure that is recorded on the hyperplane $\partial \mathbb{H}_+$. Here, $\mathbb{H}_+$ denotes the half space $\mathbb{R}^d - 1 \times (0, \infty)$. That is, we assume that we know $u$ on $\partial \mathbb{H}_+ = \mathbb{R}^{d-1} \times \{0\}$ (the acquisition surface), and our goal is to reconstruct $h$ from (possibly approximate) knowledge of $U_h$. This is an inverse problem which amounts to an (approximate) inversion of the operator $U$. We will refer to that problem as the inverse problem of PAT with a planar acquisition geometry. Note that similar approaches can be considered for other acquisition geometries as well.

In the case of exact data $U_h$, several approaches have been derived for solving the inverse problem of PAT considering different acquisition geometries. This includes time reversal (see \cite{16, 41, 62, 81, 99}), Fourier domain algorithms (see \cite{2, 4, 13, 38, 61, 63, 67, 71, 82, 108}), analytic reconstruction formulas of back-projection type (see \cite{40, 41, 53, 55, 60, 70, 77, 84, 106}), as well as iterative approaches \cite{32, 59, 85, 87, 93, 95, 103, 104, 109}. For the case of noisy data, it is well known that iterative methods (including variational methods and sparse reconstructions) tend to be more accurate than analytic methods.

Among those iterative techniques, sparse regularization approaches have gained a lot of attention during the last years as they have proven to perform well for noisy data as well as for incomplete data problems. One of their main advantages consists in their ability to combine efficient regularization with good feature preservation and (to some extent) to compensate for the missing data \cite{10, 46, 56, 90}. However, these advantages of sparse regularization methods come with the cost of typically significantly longer reconstruction times than FBP-type approaches. This is because the forward and adjoint operators have to be evaluated repeatedly. Due to that reason, FBP type methods (or other direct approaches) are often preferred over the more elaborate sparse regularization techniques \cite{3, 5, 16, 88, 89}.

In this paper, we develop a numerically efficient reconstruction method for PAT with planar geometry that effectively deals with noisy data $g = U_h + z$, where regularization is achieved by enforcing sparsity constraints in the reconstruction with respect to wavelet coefficients of $h$. More precisely, we derive a direct method for calculating a minimizer of the $\ell^1$-Tikhonov functional

$$
\Phi_{g,w}(f) := \frac{1}{2} \|A f - g\|_2^2 + \sum_{\lambda \in \Lambda} w_\lambda |\langle \psi_\lambda, f \rangle|,
$$

(1)

where $A f := (r^{-1/2} \circ U \circ y^{1/2}) f$ (see (4) for a precise definition), $(\psi_\lambda)_{\lambda \in \Lambda}$ denotes an orthonormal wavelet basis and $w_\lambda > 0$ are weights. To achieve that, as one of our main results, we construct a wavelet-vaguelette decomposition (WVD) for the operator $A$ (and also for $U$; see theorem 6). That is, given a wavelet basis $(\psi_\lambda)_{\lambda \in \Lambda}$, we construct a family $(\nu_\lambda)_{\lambda \in \Lambda}$ in ran $(A)$ that,
in the case of exact data, gives rise to an inversion formula of the form \( f = \sum_{\lambda} \kappa_{\lambda} \langle A f, v_{\lambda} \rangle \psi_{\lambda} \).

Given noisy data \( g = Af + z \), we show that a combination of this formula with soft-thresholding \( s_{\omega} \) (see equation (11)), namely

\[
f^* = \sum_{\lambda} \kappa_{\lambda} s_{\omega} \langle \langle g, v_{\lambda} \rangle \rangle \psi_{\lambda},
\]

provides a minimizer of the functional (1). Additionally, we derive an efficient algorithm for the PAT-vaguelette transform \( g \mapsto \langle \langle g, v_{\lambda} \rangle \rangle_{\lambda} \) and provide an implementation for the WVD reconstruction (2). Moreover, we show order optimality of our method in the case of deterministic noise as well as the case of Gaussian random noise. We also consider generalizations of our method and, in particular, we show how WVD reconstruction can be combined with an additional TV-prior.

Sparse regularization has been widely used as a reconstruction method for general inverse problems and there is a vast literature on that topic (see, for example, [14, 31, 42, 50, 51, 54, 72, 91, 94, 101]). In the statistical setting, \( \ell_1 \)-Tikhonov regularization is known as LASSO [11, 83, 98, 110] or basis pursuit [22]. In most cases, the reconstructions are computed by employing iterative algorithms (such as iterative soft-thresholding) to minimize the \( \ell_1 \)-Tikhonov functional [8, 12, 27, 28, 31, 39]. As mentioned above, those methods have the disadvantage to be slower than FBP type methods, as the forward and adjoint problem have to be solved repeatedly.

WVD and generalizations like biorthogonal curvelet decompositions and shearlet decompositions have been derived for the classical Radon transform in 2D, see [1, 17, 26, 35, 66]. Wavelet methods for inverse problems and tomographic applications have also been studied, for example, in [24, 33, 65, 73, 92]. To the best of our knowledge, this paper is the first to provide a WVD for PAT as well as an efficient direct implementation of sparse regularization using wavelets for that case.

1.1. Organization of the paper

In section 2 we review the mathematical principles of PAT with a flat observation surface and collect results required for our further analysis. In section 3 we derive the WVD for the forward operator. The WVD is then used to define the soft-thresholding estimator in section 4. In that section we also discuss the equivalence to variational estimation such as \( \ell_1 \)-Tikhonov regularization and introduce hybrid estimators combining the WVD with general regularization terms like the TV-prior. The efficient implementation of the WVD based estimators
requires an efficient algorithm for evaluating the PAT-vaguelette transform. Such an algorithm is derived in section 5, where we also present results of our numerical simulation.

2. PAT with a flat observation surface

Let \( C_0^\infty(\mathbb{H}_+^d) \) denote the space of compactly supported functions \( h: \mathbb{R}^d \to \mathbb{R} \) that are supported in the half space \( \mathbb{H}_+ := \mathbb{R}^{d-1} \times (0, \infty) \), where \( d \geq 2 \). We write \((x, y) \in \mathbb{R}^{d-1} \times \mathbb{R} = \mathbb{R}^d\) and consider the initial value problem

\[
\begin{align*}
(\partial_t^2 - \Delta_{(x,y)}) u(x, y, t) &= 0, & (x, y, t) &\in \mathbb{R}^d \times (0, \infty), \\
 u(x, y, 0) &= h(x, y), & (x, y) &\in \mathbb{R}^d, \\
 \partial_t u(x, y, 0) &= 0, & (x, y) &\in \mathbb{R}^d,
\end{align*}
\]

(3)

with \( h \in C_0^\infty(\mathbb{H}_+^d) \). We assume that the pressure data is observed on the hyperplane \( \partial \mathbb{H}_+ = \mathbb{R}^{d-1} \times \{0\} \) and define the corresponding PAT forward operator as

\[
U: C_0^\infty(\mathbb{H}_+^d) \to C^\infty(\mathbb{R}^d)
\]

\[
(Uh)(x, t) := \begin{cases} u(x, 0, t), & \text{for } t > 0, \\
0, & \text{else},
\end{cases}
\]

where \( u \) denotes the unique solution of (3). Our aim is to recover \( h \) from exact or approximate knowledge of \( Uh \). We are particularly interested in the cases \( d = 2 \) and \( d = 3 \), as they are of practical relevance in PAT (see [15, 69]). Nevertheless, in what follows, we consider the case of general dimension since this does not introduce additional difficulties.

2.1. Isometry property

The following isometry property for the wave equation is central in the analysis we derive below. For odd dimensions, it has been obtained in [13], and for even dimensions in [76]; see also [9].

**Lemma 1 (Isometry property for the operator \( U \)).** For any \( h \in C_0^\infty(\mathbb{H}_+^d) \) we have

\[
\int_{\mathbb{R}^{d-1}} \int_0^\infty \frac{|Uh(x, t)|^2}{t} \, dx \, dt = \int_{\mathbb{R}^{d-1}} \int_0^\infty \frac{|h(x, y)|^2}{y} \, dy \, dx.
\]

**Proof.** See [13] for \( d \) odd and [76] for \( d \) even. \( \square \)

From lemma 1 it follows that \( U \) extends to an isometry on the space \( \text{cl}(C_0^\infty(\mathbb{H}_+^d)) \) with respect to the scalar product \( \int_{\mathbb{R}^{d-1}} \int_0^\infty h_1(x, y)h_2(x, y) y^{-1} \, dy \, dx \). In view of the isometry property and the desired WVD, instead of the operator \( U \) (that represents the actual measured data), it is more convenient to work with the modified operator

\[
A: C_0^\infty(\mathbb{H}_+^d) \to C^\infty(\mathbb{R}^d)
\]

\[
(Af)(x, t) := \begin{cases} (r^{-1/2} \circ U \circ y^{1/2})(f)(x, t) & \text{for } t > 0, \\
0, & \text{else}.
\end{cases}
\]

(4)

It is not hard to see that \( A \) is an isometry with respect to the inner product

\[
\langle f_1, f_2 \rangle_{L^2(\mathbb{H}_+^d)} := \langle f_1, f_2 \rangle_{L^2(\mathbb{H}_+^d)} = \int_{\mathbb{R}^{d-1}} \int_0^\infty f_1(x, y) \overline{f_2(x, y)} \, dy \, dx.
\]
We have the following result.

**Lemma 2 (Isometry property for the operator A).** Let $A$ be defined as in (4). Then, the following assertions hold:

(a) For all $f \in C_0^\infty(\mathbb{H}_+)$ we have $\|Af\|_{L^2(\mathbb{H}_+)} = \|f\|_{L^2(\mathbb{H}_+)}$.

(b) The operator $A$ uniquely extends to an isometry $A : L^2(\mathbb{H}_+) \to L^2(\mathbb{H}_+)$. 

**Proof.**

(a) According to lemma 1 for every $f \in C_0^\infty(\mathbb{H}_+)$ we have

$$
\|Af\|_{L^2(\mathbb{H}_+)}^2 = \int_{\mathbb{R}^d-1} \int_0^\infty |Af(x,t)|^2 \, dt \, dx
$$

$$
= \int_{\mathbb{R}^d-1} \int_0^\infty \left| (t^{-1/2} U y^{1/2})f(x,t) \right|^2 \, dt \, dx
$$

$$
= \int_{\mathbb{R}^d-1} \int_0^\infty \left| U(y^{1/2}f)(x,t) \right|^2 \frac{dt}{t} \, dx
$$

$$
= \int_{\mathbb{R}^d-1} \int_0^\infty \left| y^{1/2}f(x,y) \right|^2 \frac{dy}{y} \, dx
$$

$$
= \int_{\mathbb{R}^d-1} \int_0^\infty |f(x,y)|^2 \, dy \, dx
$$

$$
= \|f\|_{L^2(\mathbb{H}_+)}^2.
$$

(b) The set $C_0^\infty(\mathbb{H}_+)$ is dense in $L^2(\mathbb{H}_+)$ and, according to (a), the operator $A$ is an isometry on $C_0^\infty(\mathbb{H}_+)$. Consequently, it follows from the general Hilbert space theory that the operator $A$ can be extended in a unique manner to an isometry $A : L^2(\mathbb{H}_+) \to L^2(\mathbb{H}_+)$. Finally, from the construction of $A$ it is clear that $A : L^2(\mathbb{H}_+) \subseteq L^2(\mathbb{H}_+)$. 

2.2. Isometric extension to $L^2(\mathbb{R}^d)$

For the following considerations it will be convenient to apply the operator $A$ to functions that are defined on $\mathbb{R}^d$ rather than on the half space $\mathbb{H}_+$. For that purpose, we need to extend the operator $A : L^2(\mathbb{H}_+) \to L^2(\mathbb{H}_+)$ in a meaningful way to an operator $A : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$. One possibility for doing this, is to consider the wave equation (3) with initial data $f \in C_0^\infty(\mathbb{R}^d)$ and then to proceed as above. However, any function that is odd in the last variable would be in the kernel of the resulting operator. Therefore, we use a different extension that leads to an isometric operator on $L^2(\mathbb{R}^d)$.

To that end, we define the operator $S : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ by $(Sf)(x,y) := (Sf)(x,-y)$. Then $S$ is an isometric isomorphism with $S^{-1} = S$ and $S(L^2(\mathbb{H}_+)) = L^2(\mathbb{H}_-)$, where $\mathbb{H}_-$ is defined in an obvious way. We are now able to define the announced extension of $A$.

**Definition 3.** We define the operator $A : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ by

$$
A(f) := A \left( P_{L^2(\mathbb{H}_+)} f \right) + (S \circ A \circ S) \left( P_{L^2(\mathbb{H}_-)} f \right).
$$

Here and below $P_{L^2(\mathbb{H}_-)} f$ denotes the orthogonal projection onto a closed subspace $V$ of $L^2(\mathbb{R}^d)$.

**Theorem 4.** The operator $A : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is an isometry.
Proof. Any function $f \in L^2(\mathbb{R}^d)$ can be written in the form $f = P_{L^2(\mathbb{H}_+)}f + P_{L^2(\mathbb{H}_-)}f$ and satisfies $\|f\|_{L^2}^2 = \|P_{L^2(\mathbb{H}_+)f}\|_{L^2}^2 + \|P_{L^2(\mathbb{H}_-)f}\|_{L^2}^2$. From definition 3 and lemma 2 we then conclude that $\|Af\|_{L^2} = \|f\|_{L^2}$ for every $f \in L^2(\mathbb{R}^d)$.

In what follows, we will also consider the operator

$$U: \text{dom}(U) \subseteq L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d): h \mapsto \left(|t|^{1/2} \circ A \circ |y|^{-1/2}\right)h$$

as a densely defined operator on $L^2(\mathbb{R}^d)$. Here, $\text{dom}(U)$ is the set of all (equivalence classes of) functions $h: \mathbb{R}^d \to \mathbb{R}$ such that $|y|^{-1/2}h \in L^2(\mathbb{R}^d)$.

2.3. Explicit expressions for $A$ and its dual

In this section, we will state explicit expressions for the operator $A$ and its dual. For that purpose, we consider the spherical Radon transform $M$, which is defined as follows:

$$\forall (x, t) \in \partial \mathbb{H}_+ \times (0, \infty): \quad (Mf)(x, r) := \frac{1}{|S^{n-1}|} \int_{S^{n-1}} f(x + r\omega) dS(\omega),$$

where $S^{n-1} := \{x \in \mathbb{R}^n \mid |x| = 1\}$ denotes the unit sphere in $\mathbb{R}^n$ and $|S^{n-1}|$ is its surface measure. A simple calculation (application of Fubini’s theorem) shows that the dual $M^*$ of the operator $M$ is given by

$$(M^*g)(x, y) = \frac{1}{|S^{n-1}|} \int_{S^{n-1}} g\left(z, \sqrt{\|z - y\|^2 + x^2}\right) dS(z).$$

The operator $M^*$ is called spherical backprojection operator, because $(M^*g)(x, z)$ integrates the function $g$ over all spheres $(z, r)$ that pass through the point $(x, z)$.

We will also consider the (fractional) differentiation operators

$$D^\mu_t := \begin{cases} ((2t)^{-1}\partial_t)^\mu, & \text{for } \mu \in \mathbb{N}, \\ ((2t)^{-1}\partial_t)^{\mu+1/2}A, & \text{for } \mu \in \mathbb{N} - 1/2. \end{cases}$$

The formal $L^2$ adjoints of those operators are given by $(D^\mu_t)^* = (-1)^{\mu}tD^\mu_t^{-1}$ for $\mu \in \mathbb{N}$ and $(D^\mu_t)^* = (-1)^{\mu+1/2}A^*D^{\mu+1/2}_t^{-1}$ for $\mu \in \mathbb{N} - 1/2$, where

$$(A^*g)(t) = \frac{2t}{\sqrt{\pi}} \int_t^\infty \frac{g(s)}{\sqrt{s^2 - t^2}} ds, \quad \text{for } t \in (0, \infty).$$

We are now able to provide explicit expressions for the operator $A$ and its dual $A^*$.

Lemma 5. We have

$$\forall f \in C^\infty_0(\mathbb{H}_+) : \quad Af = \frac{\sqrt{\pi}}{2\Gamma(n/2)} \left(t^{-1/2}D^{(n-3)/2}\rho^{n/2}M_1f\right), \quad \text{(5)}$$

$$\forall g \in C^\infty_0(\mathbb{H}_+) : \quad A^*g = \frac{\sqrt{\pi}}{2\Gamma(n/2)} \left(\rho^{1/2}M^*_1\rho^{n/2}D^{(n-3)/2}_t\rho^{-1/2}g\right). \quad \text{(6)}$$
Proof. The identity (5) follows from the well known explicit expression for the solution of the wave equation (see [29, p 682] and [37, p 80]). The identity (6) follows from (5) by applying calculation rules for the adjoint. □

3. Wavelet vaguelette decomposition (WVD)

In this section, for a given wavelet basis \((\psi_\lambda)_{\lambda \in \Lambda}\) of \(\mathbb{R}^d\), we construct the WVD of the operators \(U\) and \(A\) that we defined in the previous section and prove inversion formulae for the case of exact data. To that end, we particularly will make use of the isometry relation that we proved in section 2.

The basic idea of the WVD is to start with an orthogonal wavelet basis and to construct a possibly non-orthogonal basis of the image space in such a way that the operator and the prior information are simultaneously (nearly) diagonalized [35]. For readers convenience, we summarized some basic facts about wavelets in appendix.

3.1. The idea of the WVD

Let \(K : \text{dom}(K) \subseteq L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)\) be a linear, not necessarily bounded, operator and let \((\psi_\lambda)_{\lambda \in \Lambda}\) be an orthonormal wavelet basis of \(L^2(\mathbb{R}^d)\). The construction of a WVD for the operator \(K\) amounts to finding families \((u_\lambda)_{\lambda}, (v_\lambda)_{\lambda}\) in \(L^2(\mathbb{R}^d)\) and a family \((\kappa_j)\) of positive numbers satisfying the following properties:

(WVD1) Quasi-singular relations (with \(\lambda = (j, k, \varepsilon)\)):

\[
K \psi_\lambda = \kappa_j \nu_\lambda,
\]

\[
K^* u_\lambda = \kappa_j \psi_\lambda.
\]

(WVD2) Biorthogonal relations:

\[
\langle u_\lambda, v_\lambda' \rangle_{L^2} = \delta_{\lambda,\lambda'}.
\]

(WVD3) Near-orthogonality relations:

\[
\forall c \in L^2(\Lambda) : \| \sum_{\lambda \in \Lambda} c_\lambda u_\lambda \|_{L^2} \approx \| c \|_{\ell^2},
\]

\[
\forall c \in L^2(\Lambda) : \| \sum_{\lambda \in \Lambda} c_\lambda v_\lambda \|_{L^2} \approx \| c \|_{\ell^2}.
\]

Here, \(f \approx g\) means that there are constants \(A, B > 0\) such that \(A g \leq f \leq B g\).

Such a decomposition \((\psi_\lambda, u_\lambda, v_\lambda, \kappa_j)_{\lambda \in \Lambda}\) (if it exists) is called a WVD for the operator \(K\). Given such WVD for an operator \(K\), one can always obtain an explicit inversion formula for the operator \(K\) of the form

\[
\forall f \in \text{dom}(K) : \quad f = \sum_{\lambda \in \Lambda} \kappa_j^{-1} \langle Kf, u_\lambda \rangle_{L^2} \psi_\lambda.
\]

(7)

Note the analogy between (7) and inverting an operator via its singular value decomposition (SVD). The numbers \(\kappa_j\) depend here only on the scale parameter \(j\) and have the same meaning as singular values in the SVD. Thus, \(\kappa_j\) are referred to as quasi singular values. Similarly to the SVD, the decay of the quasi singular values \(\kappa_j\) reflects the ill-posedness of the inverse problem \(g = Kf\).
A WVD decomposition has been constructed for the classical computed tomography modeled by the two dimensional Radon transform, where \( \kappa_j = 2^{j/2} \); see [35]. In the case of the two dimensional Radon transform, a generalization of the WVD, a so-called biorthogonal curvelet decomposition was constructed in [17]. In [26], the authors derived biorthogonal shearlet decompositions for two and three dimensional Radon transforms.

In the case of PAT, there are no such decompositions available so far. In the next subsection, we establish a WVD for the operators \( \mathbf{U} \) and \( \mathbf{A} \), which serve as forward operators for PAT with a planar observation surface, and so automatically obtain an inversion formula for exact data.

### 3.2. Construction of the WVD for PAT

Let \((\psi_\lambda)_{\lambda \in \Lambda}\) be a wavelet basis of \( L^2(\mathbb{R}^d) \). The desired WVD of \( \mathbf{U} \) and \( \mathbf{A} \) is in fact a direct consequence of the isometry relation of theorem 4.

**Theorem 6 (The WVD for PAT).** For every \( \lambda \in \Lambda \) define \( u_\lambda := A \psi_\lambda \) and let \( v_\lambda = u_\lambda \). Then, the family \((v_\lambda,u_\lambda,\psi_\lambda,1)_{\lambda \in \Lambda}\) is a WVD for the operator \( \mathbf{A} \). Moreover, \((u_\lambda)_{\lambda \in \Lambda}\) is an orthonormal basis of \( \text{ran} \ (\mathbf{A}) \subseteq L^2(\mathbb{R}^d) \) and we have the inversion formula

\[
\forall f \in L^2(\mathbb{R}^d) : \ f = \sum_{\lambda \in \Lambda} \langle A f, u_\lambda \rangle_{L^2} \psi_\lambda, \tag{8}
\]

\[
\forall h \in \text{dom} \ (\mathbf{U}) : \ h = \sum_{\lambda \in \Lambda} \langle \mathbf{U} h, |t|^{-1/2} u_\lambda \rangle_{L^2} |y|^{1/2} \psi_\lambda. \tag{9}
\]

**Proof.** We start with an arbitrary function \( f \in L^2(\mathbb{R}^d) \) and express this function in terms of a wavelet expansion \( f = \sum_{\lambda \in \Lambda} \langle f, \psi_\lambda \rangle_{L^2} \psi_\lambda \) with respect to \((\psi_\lambda)_{\lambda \in \Lambda}\). Then according to the isometry property \((u_\lambda)_{\lambda \in \Lambda}\) is an orthonormal basis of \( \text{ran} \ (\mathbf{A}) \) and further \( A^* u_\lambda = \psi_\lambda \). In particular, this amounts to a WVD decomposition with \( v_\lambda = u_\lambda \) and \( \kappa_j = 1 \). Further, (8) is a consequence of the WVD. Next, let \( f = |t|^{-1/2} h \) be an element in \( \text{dom} \ (\mathbf{U}) \). Then \( \langle A f, u_\lambda \rangle_{L^2} = \langle |t|^{-1/2} \mathbf{U} h, u_\lambda \rangle_{L^2} = \langle \mathbf{U} h, |t|^{-1/2} u_\lambda \rangle_{L^2} \). Consequently, applying (8) yields (9).

Note that the identity (9) in theorem 6 is an explicit inversion formula for the operator \( \mathbf{U} \) in the spirit of a WVD. Instead of an orthonormal wavelet basis it uses the family \((|t|^{1/2} \psi_\lambda)_{\lambda \in \Lambda}\), which is non-orthogonal with respect to the \( L^2 \) inner product. Restricted to functions in \( L^2(\mathbb{R}^d) \) that vanish outside \( K := K_+ \cup S(K_+) \), where \( K_+ \subseteq H_+ \) is any compact subset, the operator \( h \mapsto |t|^{1/2} h \) is an isomorphism. Then, \((|t|^{1/2} \psi_\lambda)_{\lambda \in \Lambda}\) allows to characterize the Besov norm \( \| \cdot \|_{\mathcal{B}_{2,1}} \) of any function that is supported in \( K \).

Figure 2 shows a vertical, a horizontal and a diagonal Daubechies 6 wavelet (scaled and displaced versions of \( \psi^{(1)}(x,y) \), \( \psi^{(2)}(x,y) \) and \( \psi^{(3)}(x,y) \); see appendix A.2) and the corresponding PAT-vaguelettes obtained by application of the operator \( \mathbf{A} \).

### 4. Inversion from noisy data

In what follows, we assume \((\psi_\lambda)_{\lambda \in \Lambda}\) to be an orthonormal compactly supported wavelet basis of \( L^2(\mathbb{R}^d) \). Further, by \((u_\lambda)_{\lambda \in \Lambda}\) we denote the corresponding PAT-vaguelette basis of \( \text{ran} \ (\mathbf{A}) \subseteq L^2(\mathbb{R}^d) \) that satisfies \( \mathbf{A} \psi_\lambda = u_\lambda \).
If exact data are available, then the WVD decomposition (8) provides an exact reconstruction formula for the unknown $f$. However, in practical applications the data $A f$ (or $U h$) are only known up to some errors (e.g. noise). We therefore assume that we are given erroneous (noisy) data

$$g = Af + z,$$

where $z$ is the error term and $f$ the exact unknown. We consider both the deterministic and the stochastic case, in which different models for $z$ are assumed:

- In the deterministic case, we assume that a bound $\|z\|_{L^2} \leq \delta$ is available.
- In the stochastic case, we assume that $z = \delta Z$, where $Z$ is a white noise process.

In the deterministic situation the constant $\delta > 0$ is referred to as the noise level; in the stochastic case it is the noise standard deviation. The goal in both situations is to estimate the unknown $f \in L^2(\mathbb{R}^d)$ from data in (10).

### 4.1. PAT-vaguelette-thresholding estimator

In section 3 we have shown that the reproducing formula $f = A^* Af = \sum_{\lambda \in \Lambda} \langle Af, u_{\lambda} \rangle_{L^2} \psi_{\lambda}$ holds in the case of exact data. If the data is corrupted by noise, i.e. $g = Af + z$, the inner products $\langle Af, u_{\lambda} \rangle_{L^2}$ cannot be computed exactly. Instead, they are estimated by first evaluating $\langle g, u_{\lambda} \rangle_{L^2}$ and then applying the soft-thresholding operation

$$s_q: \mathbb{R} \rightarrow \mathbb{R}: y \mapsto \begin{cases} 
y + q & \text{if } y < -q \\
0 & \text{if } y \in [-q, q] \\
y - q & \text{if } y > q \end{cases}$$

with appropriate (level dependent) thresholds $q = w_j$.

**Definition 7 (WVD soft thresholding estimator).** For any nonnegative sequence $w = (w_j)$, the WVD soft thresholding-estimator for the solution of (10) is defined by
\[ S_w : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d) : g \mapsto \sum_{\lambda} s_w((g, u_\lambda)_{L^2}) \psi_\lambda, \]  
with the non-linear soft-thresholding function \( s_w \) defined in (11).

In the deterministic case we assume that the error term \( z \) is known to satisfy \( \| z \| \leq \delta \). In this case, due the isometry property of the operator \( A \), one obtains a stable reconstruction by applying the adjoint \( A^* \),

\[ \forall f \in L^2(\mathbb{R}^d) : \sup_{\| z \| \leq \delta} \| f - A^*(Af + z) \|_{L^2} = \delta. \]

As the adjoint operator satisfies \( A^* = S_0 \), i.e. \( A^* \) corresponds to the WVD-thresholding estimator with \( w = 0 \), it follows that also the class of WVD-thresholding estimators yields the optimal error estimate. Without additional knowledge about the noise there is no need to apply the WVD soft-thresholding estimator with \( w \neq 0 \), even if the function \( f \) is known to belong to some smoothness class, such as a Sobolev or Besov ball. However, if a portion of the noise energy is known to correspond to high frequency components, then using non-zero thresholds may significantly outperform \( A^* \) given that \( f \) belongs to some smoothness class. For example, this happens in the case of stochastic white noise, where the energy-spectrum is uniformly distributed.

4.2. Optimality of PAT-vaguelette-thresholding

Suppose \( z = \delta Z \), where \( Z \) is a white noise process. In the case of random noise, it is common to measure the performance of an estimator \( R : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d) \) in terms of the worst-case risk \([34, 35, 64, 100]\) of a subset \( M \subseteq L^2(\mathbb{R}^d) \),

\[ \Delta(R, \delta, M) = \sup_{f \in M} \mathbb{E} \left( \| f - R(Af + \delta Z) \|_{L^2}^2 \right). \]

Further, define the nonlinear minimax risk, the minimax risk using the WVD estimator (12) and the linear minimax risk, respectively by

\[ \Delta_N(\delta, M) := \inf_{R} \sup_{f \in M} \mathbb{E} \left( \| f - R(Af + \delta Z) \|_{L^2}^2 \right), \]
\[ \Delta_w(\delta, M) := \inf_{w \in [0, \infty)^p} \sup_{f \in M} \mathbb{E} \left( \| f - S_w(Af + \delta Z) \|_{L^2}^2 \right), \]
\[ \Delta_L(\delta, M) := \inf_{R \text{ linear}} \sup_{f \in M} \mathbb{E} \left( \| f - R(Af + \delta Z) \|_{L^2}^2 \right). \]

From the definition it is clear that no reconstruction method \( R : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d) \) can have worst-case risk \( \Delta(R, \delta, M) \) smaller than the non-linear minimax risk \( \Delta_N(\delta, M) \). We are in particular interested in asymptotic behavior for the case \( \delta \to 0 \) and \( M \) is a ball in a Besov space.

**Theorem 8.** Suppose that \( r > d(1/p - 1/2) \) and let \( M \) be a ball in the Besov-norm having the form

\[ M = \left\{ f \in L^2(\mathbb{R}^d) \mid \| f \|_{B_{r,q}} < \varrho \right\} \text{ for some } \varrho > 0. \]

Then, as \( \delta \to 0 \), the following hold
(a) $\Delta_N(\delta, M) \asymp \delta^{\frac{2}{d+1}}$
(b) $\Delta_w(\delta, M) \leq c_w \Delta_N(\delta, M)$ for some constant $c_w > 0$.
(c) $\Delta_1(\delta, M) \asymp \delta^{\frac{d+1}{2d+1}}$ with $p_- := \min \{2, p\}$.

**Proof.** Follows from [35, theorem 4] for the special case $\kappa_j = 1$. □

Theorem 8 implies that, despite its simplicity, the WVD-thresholding estimator is order optimal on any Besov-ball and the rate cannot be improved by any other estimator (up to some constant factors). On the other hand, if $p < 2$, then the exponent in the linear minimax rate is strictly smaller than the exponent in the non-linear minimax rate, $\frac{r+d(1/2-1/p_-)}{r+d(1-1/p_-)} < \frac{r}{r+d}$. Therefore, no linear estimator can give the optimal convergence order. In particular, this implies that the WVD-thresholding estimator outperforms filter based regularization methods including the truncated SVD or quadratic Tikhonov regularization.

### 4.3. Variational characterizations and extensions

The WVD-based soft thresholding estimation can be characterized via various variational minimization schemes that, as we shall discuss later, in turn offer several extension of the WVD-estimators.

**Theorem 9 (Variational characterizations of vaguelette thresholding).** Let $(w_j)$ be a sequence of thresholds and let $g, f \in L^2(\mathbb{R}^d)$. Then the following assertions are equivalent:

1. $\hat{f} = S_w(g)$;
2. $\hat{f} = \arg \min \left\{ \frac{1}{2} \|A\hat{f} - g\|_2^2 + \sum_{\lambda \in \Lambda} w_{j} |\langle \psi_{\lambda}, f \rangle| : f \in L^2(\mathbb{R}^d) \right\}$;
3. $\hat{f}$ is the unique solution of the constraint optimization problem

\[
\begin{cases}
\min_{f \in L^2(\mathbb{R}^d)} \frac{1}{2} \|f\|_2^2 \\
\text{such that } \max_{\lambda \in \Lambda} \frac{|\langle \alpha, g - A\hat{f} \rangle|}{w_j} \leq 1.
\end{cases}
\] (13)

**Proof.** (1) $\Leftrightarrow$ (2): Let $\hat{f}$ denote the minimizer of $\frac{1}{2} \|A\hat{f} - g\|_2^2 + \sum_{\lambda \in \Lambda} w_{j} |\langle \psi_{\lambda}, f \rangle|$. Because $(u_{\lambda})_{\lambda \in \Lambda}$ is an orthonormal basis of $\text{ran}(A)$ we have $\|A\hat{f} - g\|_2^2 = \|P_{\text{ran}(A)}(g)\|_2^2 + \sum_{\lambda \in \Lambda} |\langle A\hat{f} - g, u_{\lambda} \rangle|^2$. Further, $\sum_{\lambda \in \Lambda} w_{j} |\langle \psi_{\lambda}, f \rangle| = \sum_{\lambda \in \Lambda} w_{j} |\langle u_{\lambda}, A\hat{f} \rangle|$, which shows that $\hat{f}$ is the unique minimizer of

\[
\sum_{\lambda \in \Lambda} \frac{1}{2} |\langle A\hat{f} - g, u_{\lambda} \rangle|^2 + w_{j} |\langle u_{\lambda}, A\hat{f} \rangle|.
\]

The latter functional is minimized by minimizing every summand

\[
\Phi(\langle A\hat{f}, u_{\lambda} \rangle ; w_j) = \frac{1}{2} |\langle A\hat{f}, u_{\lambda} \rangle - \langle g, u_{\lambda} \rangle|^2 + w_{j} |\langle u_{\lambda}, A\hat{f} \rangle|
\]

independently in the first argument. The minimizer is given by one-dimensional soft thresholding which gives $\langle \hat{f}, \psi_{\lambda} \rangle = \langle A\hat{f}, u_{\lambda} \rangle = s_{w_j}(\langle g, u_{\lambda} \rangle)$ and therefore $\hat{f} = S_w(g)$. 


(1) ⇔ (3): Let \( \hat{f} \) denote the solution of (13). Using that \( (\psi_\lambda)_{\lambda \in \Lambda} \) is an orthogonal basis and \( \psi_\lambda = A^* u_\lambda \), shows that \( \hat{f} \) can be equivalently characterized as the minimizer of

\[
\begin{align*}
\min_{f \in L^2(\mathbb{R}^d)} & \quad \frac{1}{2} \sum_{\lambda \in \Lambda} |\langle A f, u_\lambda \rangle|^2 \\
\text{such that} & \quad \max_{\lambda \in \Lambda} \frac{|\langle u_\lambda, g - A f \rangle|}{w_j} \leq 1.
\end{align*}
\]

(14)

The minimization problem (14) can again be solved separately for every component \( \langle A f, u_\lambda \rangle \) which is again given by the one-dimensional soft thresholding \( \langle \hat{A} f, u_\lambda \rangle = s_{w_j} (\langle g, u_\lambda \rangle) \). As above this implies \( \hat{f} = S_{w}(g) \).

The variational characterizations of theorem 9 have several important implications. First, they provide an explicit minimizer for the \( \ell_1 \)-Tikhonov functional

\[
\Phi_{g, u}(f) := \frac{1}{2} \|A f - g\|_{L^2}^2 + \sum_{\lambda \in \Lambda} w_j |\langle \psi_\lambda, f \rangle|,
\]

which in general has to be minimized by an iterative algorithm, such as the iterative soft thresholding algorithm and its variants [8, 27, 28, 31]. For the analysis of \( \ell_1 \)-Tikhonov regularization for inverse problems see [31, 50, 51, 72]. Another important consequence is that the WVD-soft thresholding estimator can be generalized in various directions. In particular, one can get a generalization of (13) by replacing the \( L^2 \)-norm in (13) by an appropriate regularization functional \( \mathcal{J} \) (for example, \( \mathcal{J} \) can be chosen as the total variation norm, see section 5.3). That is, instead of solving the problem (13), one aims at solving the problem

\[
\begin{align*}
\min_{f \in L^2(\mathbb{R}^d)} & \quad \mathcal{J}(f) \\
\text{such that} & \quad \max_{\lambda \in \Lambda} \frac{|\langle u_\lambda, g - A f \rangle|}{w_j} \leq 1.
\end{align*}
\]

(15)

This generalization constitutes a hybrid version that combines the WVD estimator with more general regularization functionals \( \mathcal{J} \). Such hybrid approaches have been introduced independently in [18, 36, 74, 96] (see also [21, 25]). It is also related to the Dantzig and multiscale estimators of [19, 43, 45, 52, 58, 78].

5. Numerical implementation

Throughout the following, \( (\psi_\lambda)_{\lambda \in \Lambda} \) denotes an orthonormal wavelet basis of \( L^2(\mathbb{R}^d) \) and \( (u_\lambda)_{\lambda \in \Lambda} \) is the corresponding orthogonal PAT-vaguelette basis of \( \text{ran}(A) \subseteq L^2(\mathbb{R}^d) \) defined by \( u_\lambda := A \psi_\lambda \) (see theorem 6). In this section, we provide algorithms for the calculation of the PAT-vaguelette transform \( g \mapsto (\langle g, u_\lambda \rangle)_{\lambda \in \Lambda} \) and the corresponding WVD estimator defined by (12). Moreover, we consider a hybrid version of the WVD estimator that combines wavelets and TV regularization and discuss its implementation. We also present results of our numerical studies.

5.1 Practical aspects

The considered noisy data model (10) assumes that the data can be collected on the whole hyperplane \( \partial \mathcal{H}^+ \). However, this is not feasible in practice since the data can be collected only on a finite subset. In this subsection, we discuss the effects of partial (or limited view) data and discretization.
First, we address the limited data issue. In the considered imaging setup, the data is collected on a subset \( Q \times [0, T] \) where \( Q \subseteq \mathbb{R}^{d-1} \) is the finite measurement aperture (see figure 3) and \( T \in (0, \infty) \) the maximal measurement time. Such partial or limited view data can be modeled by

\[
g^\delta(x, t) = (\chi_{Q \times [0, T]} A f)(x, t) + \delta Z, \tag{16}\]

where \( \chi_{Q \times [0, T]} \) denotes the characteristic function of \( Q \times [0, T] \), and \( \delta Z \) is noise with variance \( \delta^2 \). Using partial data, only certain features of \( f \) can be reconstructed in a stable way, see \([6, 47, 80, 97]\) and 3. Consequently, the practical problem of reconstructing the unknown function \( f \) from partial data, is a limited data problem and therefore severely ill-posed. It is therefore common to incorporate a priori information into the reconstruction and so to regularize the reconstruction. In this work, we are doing it in two ways: first, we incorporate wavelet sparsity assumptions. This is what the WVD estimator does, which is implemented by evaluating \( f^0 = \sum_{\lambda} \langle \chi_{Q \times [0, T]} A f, u_\lambda \rangle_{L^2} \psi_\lambda \) (see definition 7). Here and below, for simplicity, we take the thresholding sequence \( w = (w_j) \) being constant, and denote its value by \( q = w_j \). In the noise free case, taking \( q = 0 \), noting \( u_\lambda = A \psi_\lambda \), and using the linearity of \( A \) and \( A^* \), we obtain

\[
f^0 = \sum_{\lambda} \langle \chi_{Q \times [0, T]} A f, u_\lambda \rangle_{L^2} \psi_\lambda = \sum_{\lambda} \langle A^*(\chi_{Q \times [0, T]} A f), u_\lambda \rangle_{L^2} \psi_\lambda = A^*(\chi_{Q \times [0, T]} A f). \tag{17}\]

The partial reconstruction \( f^0 = A^*(\chi_{Q \times [0, T]} A f) \) is known to be a ‘good’ approximate reconstruction for \( f \), since it recovers all visible boundaries of \( f \) correctly (as follows from \([6, 47, 80, 97]\)) and can be evaluated stably. Second, we combine wavelet sparsity with TV regularization by using the hybrid estimator (15) to impose even more regularization.

Another practical restriction is that only a finite number of samples of the pressure can be measured. Assuming equidistant sampling and limited view data \( g \) as in (16) with \( Q = [0, T]^{d-1} \), the actual sampled data are given by

\[
g[n, m] := g(n \Delta_n, m \Delta_m) + z[n, m], \quad (n, m) \in \{1, \ldots, N\}^{d-1} \times \{1, \ldots, M\}. \tag{18}\]
Here \( N \) and \( M \) are natural numbers, \( \Delta_N := X/N \) and \( \Delta_M := T/M \) are the sampling step sizes and \( z[n,m] \) describes the noise in the data. Using Shannon sampling theory it can be shown that (18) is correctly sampled for \( \Delta_N = \Delta_M \leq \pi/\Omega \), where \( \Omega \) is the essential bandwidth of \( f \); see [61]. The precise analysis of discretization effects on the considered vaguelette estimators is beyond the scope of this paper.

5.2. Implementation of the PAT-vaguelette transform

Analogously to the wavelet transform, we can define the vaguelette transform of \( g \in L^2(\mathbb{R}^d) \) corresponding to the operator \( A \) by \( Vg := (\langle u_\lambda, g \rangle_{L^2})_{\lambda \in \Lambda} \). Using the representation \( u_\lambda = A^* \psi_\lambda \) (compare (WVD1) and theorem 6) we see that the PAT-vaguelette transform is given by

\[
\forall g \in L^2(\mathbb{R}^d) \quad \forall \lambda \in \Lambda: \quad (Vg)(\lambda) = (A^* \psi_\lambda, g)_{L^2} = (\psi_\lambda, A^* g)_{L^2}.
\]

Hence, the PAT-vaguelette transform can be computed by first applying the adjoint \( A^* \) to the data \( g \) and then calculating the wavelet transform of \( A^* g \).

**Algorithm 1.** Discrete PAT-vaguelette transform.

1: Compute \( Bg \in \mathbb{R}^{N_d-1 \times M} \) with \( Bg[n,m] \approx A^* g(n\Delta_N, \Delta_M) \)
2: Compute the discrete wavelet transform \( WBg \)

Assuming discrete data \( g \in \mathbb{R}^{N_d-1 \times M} \) of the form \( g[n,m] = g(n\Delta_N, m\Delta_M) \), the discrete PAT-vaguelette transform can be efficiently computed by algorithm 1. Both the steps in this algorithm are well known and numerically efficient (compare with remark 10 below). For the first step, we use a numerical approximation of \( A^* \) by numerically implementing (6) with a filtered backprojection (FBP) algorithm (see [15, 16]). For evaluating the second step we use the implementation of the fast discrete wavelet transform provided by the Matlab function `wavedec2`.

5.3. Implementation of the reconstruction algorithms

For the numerical experiments, that will be presented in the next section, we implement the WVD soft-thresholding estimator defined by (12) as well as a hybrid vaguelette-TV approach defined by (15) with \( J(f) = \|f\|_{TV} \), being the TV norm of \( f \). In both cases, we take constant weights \( w_j = q \), where \( q > 0 \) is some threshold. In the case of random noise, a typical choice for the threshold is Donoho’s so called universal threshold \( \delta \sqrt{2 \log (N^d-1M)} \), that can be derived from extreme value theory [34, 57]. The generalization to non-constant weights is straightforward.

The implementation of the WVD soft-thresholding estimator is summarized in algorithm 2. It is based on the discrete PAT-vaguelette transform that has been presented in algorithm 1.

**Algorithm 2.** WVD soft-thresholding estimator.

1: Compute PAT-vaguelette coefficients \( c \approx Vg \) by applying algorithm 1
2: Apply soft thresholding to \( c \) using threshold \( q \)
3: Apply the inverse wavelet transform

**Remark 10 (Numerical complexity of the WVD soft-thresholding algorithm).** In two spatial dimensions, and assuming \( N \sim M \), the FBP algorithm requires \( O(N^3) \) operations (see [16]) and the discrete wavelet transform requires \( O(N^2 \log N) \) operations. Therefore, the
discrete PAT-vaguelette transform (algorithm 1) as well as the proposed WVD soft thresholding estimator given by algorithm 2 have a complexity of $O(N^3)$, which is the same as for the standard FBP algorithm. Using Fourier algorithms for evaluating $A^*$ (see [61, 63, 67, 82, 108]) the numerical effort even reduces to $O(N^2 \log N)$.

Given discrete data, the hybrid vaguelette-TV approach can be written in the form

$$
\min_f \| f \|_{TV} \quad \text{such that} \quad \| W(Bg - f) \|_{\infty} \leq q,
$$

(19)

where $B$ denotes the discrete FBP operator. One recognizes that (19) can be implemented by applying a hybrid wavelet-TV denoising algorithm to $Bg$. Following [44], we rewrite the optimization problem (19) in the form

$$
\min_{f,v} \| f \|_{TV} + i\{\| v \|_{\infty} \leq w\} \quad \text{such that} \quad f + v = Bg,
$$

(20)

and introduce the associated augmented Lagrangian operator

$$
\mathcal{L}_\epsilon(f,v,\mu) := \| f \|_{TV} + i\{\| v \|_{\infty} \leq w\} + (\mu, f + v - Bg) + \frac{1}{2\epsilon} \| f + v - Bg \|_2^2.
$$

Then, (20) can be solved by the alternating direction method of multipliers (ADMM), introduced in [48, 49], which alternatingly performs minimization steps with respect to $f$ and $v$ and maximization steps with respect to $\mu$. The resulting ADMM algorithm for solving (19) is summarized in algorithm 3.
Algorithm 3. Hybrid vaguelette-TV approach.

1: $f_0 = v_0 = \mu_0 = 0$
2: for $k = 0, 1, \ldots, N_{\text{iter}}$ do
3:  $f_{k+1} \in \arg \min \frac{1}{2} \| f - (Bg - v_k - c\mu_k) \|_2^2 + c \| f \|_{TV}$
4:  $v_{k+1} = q(Bg - f_{k+1} - c\mu_k) / \max \{ q, \|Bg - f_{k+1} - c\mu_k\|_2 \}$
5:  $\mu_{k+1} := \mu_k + c^{-1}(f_{k+1} + v_{k+1} - Bg)$
6: end for

For performing the $f$-update in algorithm 3 we have to solve the unconstraint total variation minimization problem $\frac{1}{2} \| f - (Bg - v_k - c\mu_k) \|_2^2 + c \| f \|_{TV}$. For this purpose we use Chambolle’s dual projection algorithm [20]. In two spatial dimensions, assuming $M \sim N$ and using the FBP algorithm for evaluating $Bg$ (which is required only once), the numerical complexity of algorithm 3 is $O(N^3 + N^3 N_{\text{iter}} N_{\text{inner}})$. Here $N_{\text{iter}}$ is the number of outer iterations and $N_{\text{inner}}$ the number of inner iterations with the Chambolle algorithm. In the numerical example presented in the following subsection, we use $N = 512$, $N_{\text{iter}} = 100$ and $N_{\text{inner}} = 50$. In such a situation, the numerical complexity is dominated by $N^2 N_{\text{inner}} N_{\text{iter}}$, which is confirmed by the runtimes.

5.4. Numerical example 1: Disc phantom

In this section, we present a numerical example testing the PAT-vaguelette soft-thresholding and the hybrid approach. For that purpose we first consider a simple phantom that consists
of the sum of four non-overlapping uniformly absorbing discs as illustrated in the top image in figure 4. The simple phantom has been chosen in order to clearly identify the reconstruction artefacts due to noise and the limited detection view. Note the amount of visible boundaries depends of the location of the disc relative to the observation surface (see figure 3). More boundaries are visible for the inner discs than for the outer discs. The initial phantom $f$ is discretized using $128 \times 512$ equidistant samples in $[0, 1] \times [-2, 2]$. The data $g^0 = Af$ are computed numerically using an implementation according to (5) and evaluated at $512 \times 512$ equidistant samples in $[0, 4] \times [-2, 2]$.

To simulate measurement errors, we added i.i.d. Gaussian noise with standard deviation $\delta = 0.25$. The relative $\ell^2$-error in the noisy data $g^\delta \in \mathbb{R}^{512 \times 512}$ is $\|g^\delta - g^0\|_2 / \|g^0\|_2 = 1.05$.

The simulated data with noise added is shown in middle row in figure 4. Note that the data are shown for times restricted to $[0, 2]$ but simulated until the maximal measurement time $T = 4$, and have been multiplied by $|t|^{1/2}$ to represent actual PAT measurement data. The last row in figure 4 shows a reconstruction from exact data using the FBP algorithm (denoted as B and implemented following [15]). Reconstructions from noisy data are presented below. We point out here, that the FBP algorithm equals our WVD soft-thresholding estimator when the threshold parameters are chosen to be constantly zero (see algorithms 1 and 2).

Reconstruction results for noisy data are shown figure 5. The top image shows the non-regularized reconstruction $f^\delta_{\text{FBP}} = B g^\delta$, the middle image the WVD soft-thresholding reconstruction $f^\delta_{\text{WVD}}$, and the bottom image the reconstruction with the hybrid vaguelette-TV method $f^\delta_{\text{hybrid}}$ using 100 steps of the ADMM with 50 inner iterations for Chambolle’s TV minimization algorithm. For the WVD and the hybrid reconstruction, the weights $w_j = q = 0.63$ are chosen to be constant which value equals half of Donoho’s universal threshold $\delta \sqrt{2 \log(512 \times 512)}$. 

![Figure 6. Differences between the disc phantom and the reconstructions shown in figure 4 with $\delta = 0.25$. Top row: $f^\delta_{\text{FBP}} - f$. Middle row: $f^\delta_{\text{WVD}} - f$. Bottom row: $f^\delta_{\text{TV}} - f$.](image-url)
Figure 6 shows corresponding difference images between the reconstructions and the true phantom \( f \). The reconstruction times have been 0.46 seconds for the FBP, 0.51 seconds for the PAT-vaguelette thresholding estimator and 35.34 seconds for the hybrid estimator. The results have been obtained with a Matlab implementation using Intel Xeon CPU E3-1225 (3.20 Ghz) on a standard desktop PC.

Figure 7. Shepp–Logan phantom and reconstructions with FBP and WVD from noisy data with \( \delta = 0.15 \). Top row: head phantom (left) and FBP reconstruction (right). Rows 2, 3, 4: WVD soft-thresholding reconstruction \( f^{\delta}_{\text{WVD}} \) using various threshold values. (All images use the same colormap.)
One clearly observes that the vaguelette estimators significantly reduce the error in the reconstruction. This can be quantified by the relative $\ell^2$-errors, which are equal to $\|f_{\delta \text{FBP}} - f\|_2 / \|f\|_2 = 0.66$ for the unregularized (FBP) reconstruction, $\|f_{\delta \text{WVD}} - f\|_2 / \|f\|_2 = 0.38$ for the PAT-vaguelette thresholding estimator and $\|f_{\delta \text{hybrid}} - f\|_2 / \|f\|_2 = 0.42$ for the hybrid estimator. The relative $\ell^2$-error for the hybrid reconstruction is slightly larger than for the vaguelette-thresholding estimator. This is reasonable since the PAT-vaguelette-thresholding estimator minimizes the $\ell^2$-norm (see (13)) among all potential reconstructions $f$ that are compatible with the data, whereas the hybrid estimator minimizes the total variation. However, in some more appropriate error measure the hybrid reconstruction may outperform the thresholding estimator. Further note that all reconstruction methods contain some limited data artifacts, which cannot be removed completely by wavelet methods [6, 47, 61, 79, 97].

**Figure 8.** Hybrid reconstructions $f_{\delta \text{hybrid}}$ of the Shepp–Logan phantom from noisy data with $\delta = 0.15$ using various thresholds. (Same colormap as in figure 7.)
5.5. Numerical example 2: Shepp–Logan phantom

Next we present reconstruction results using a more complex phantom. We also discuss the effects of different noise levels and different thresholds. The Shepp–Logan head phantom \( \mathbf{f} \) is discretized using \( 256 \times 512 \) equidistant samples in \([0, 1] \times [-1, 1]\) and the (limited view) data \( \mathbf{g}^0 = A \mathbf{f} \) using \( 512 \times 512 \) equidistant samples in \([0, 2] \times [-1, 1]\). To the simulated data \( \mathbf{g}^0 \) we added i.i.d. Gaussian noise with variable noise levels (different standard deviations in the interval \([0, 0.2]\)).

We first consider the case of noise standard deviation equals \( \delta = 0.15 \), in which the relative noise level is about \( \| \mathbf{g}^\delta - \mathbf{g}^0 \|_2 / \| \mathbf{g}^0 \|_2 = 0.8 \). The phantom and the reconstructions from the noisy data using various thresholds \( q = 0, 0.15, 0.3, 0.45, 0.65, 0.8, 0.95 \) are shown in figure 7. Note the result with the threshold \( q = 0 \) corresponds to the result with the FBP algorithm. The relative reconstruction errors for the above values of threshold are \( \| \mathbf{f}^\delta_{\text{FBP}} - \mathbf{f}^\delta \|_2 / \| \mathbf{f}^\delta \|_2 = 0.78 \) for FBP and \( \| \mathbf{f}^\delta_{\text{WVD}} - \mathbf{f}^\delta \|_2 / \| \mathbf{f}^\delta \|_2 = 0.65, 0.60, 0.59, 0.58, 0.59, 0.59 \) for the WVD reconstructions. One notices that the errors first decrease with increasing thresholds and then start to increase again. Figure 8 shows results using the hybrid WVD-TV approach using...
the same thresholds \( q = 0.15, 0.3, 0.45, 0.65, 0.8, 0.95 \). The relative reconstruction errors are 
\[
\frac{\|f_{\text{hybrid}}^\delta - f\|_2}{\|f\|_2} = 0.72, 0.65, 0.62, 0.60, 0.59, 0.58.
\]
1.09 seconds for the PAT-vaguelette thresholding estimator and 39.26 s for the hybrid estimator. As in the disc example, the reconstruction errors (that are measured in the \( L^2 \)-norm) are slightly larger for the hybrid method, while the visual results might be slightly more appealing. This again indicates that the \( L^2 \)-norm might not be the optimal error measure (at least) for the hybrid method. As a consequence, the following detailed error study is only performed for the WVD soft thresholding method.

To investigate the dependence of the reconstructions on the noise variance and the selected thresholds in more detail, we applied the WVD soft thresholding estimator for noisy data with 21 equidistant values of \( \delta \) in \([0, 0.2]\) and 21 equidistant values of the threshold \( q \) in \([0, 1.5]\).

The relative reconstruction errors \( E_1 := \frac{\|f_{\text{WVD}}^\delta - f\|_2}{\|f\|_2} \) are shown in the two images in the left column of figure 9. For any value of the standard deviation, the reconstruction errors first decreased with increasing threshold and then start to increase again. The relatively large values of the reconstruction errors are due to the limited view artifacts that are present in any reconstruction. The limited view artifacts are also partially removed by the soft thresholding estimator which affects that similar thresholds is optimal for a large range of noise levels (in the sense that they minimize \( E_1 \), note the solid white line with added white dots). In the right image in figure 9, we therefore also depict the relative distance \( E_2 := \frac{\|f_{\text{WVD}}^\delta - f^0\|_2}{\|f^0\|_2} \), where \( f^0 = Bg^0 \) is the noise free partial reconstruction. The threshold that minimize \( E_2 \) now shows a clear (linear) dependence on the variance. The dashed lines in the images in the top row show Donoho’s universal threshold \( \delta \sqrt{2 \log(512 \times 512)} \) that is known to completely remove noise with high probability but also is known to yield over-smoothing [34, 57].

6. Conclusion

In this paper, we developed a regularization framework using wavelet sparsity in photoacoustic tomography (PAT). For that purpose we derived wavelet-vaguelette decompositions (WVDs) (see theorem 6) and efficient implementation of the corresponding PAT-vaguelette transform (see algorithm 1). Using the WVD we derived an explicit formula for minimizing the sparse Tikhonov functional that can be implemented without any iterative reconstruction procedure (see algorithm 2). The considered regularization approach has been shown to provide optimal error estimates in the deterministic as well as in the stochastic setting (see theorem 8). The computational complexity of the given implementation is \( O(N^3) \) for recovering a phantom at \( N \times N \) reconstruction points, which is the same as for the standard filtered backprojection (FBP) algorithm. In order to account for wavelet artifacts we also developed hybrid regularization methods combining wavelet sparsity with total variation (see algorithm 3). Numerical results demonstrate the feasibility of our reconstruction approaches. Future work will be done to extend our approach to more general measurements geometries in PAT and also to different Radon type inverse problems.

Appendix. Orthogonal wavelets

We recall some basic facts about orthogonal wavelets as we need them for our purpose (in particular in section 3). For the task of function estimation, wavelets are known to sparsely represent many signals and, hence, they can be used to effectively encode prior information. Another useful property of wavelets consists in the ability to characterize several classical
smoothness measures (e.g. Sobolev and Besov norms) in terms of the decay properties of wavelet coefficients. We will also use that property in section 4. For a detailed introduction to wavelets we refer to [23, 30, 75].

A.1. One-dimensional wavelets

We first briefly recall the basic definitions and notations of orthonormal wavelet bases (wavelet ONB) in one spatial dimension, which will be then extended to higher dimensions.

The construction of a wavelet ONB is based on the concept of a multiresolution analysis (MSA), which is given by a sequence subspaces \((V_j)_{j \in \mathbb{Z}}\) in \(L^2(\mathbb{R})\) that satisfy the following requirements (see [23, 30, 75]):

- For all \(j \in \mathbb{Z}\) it holds that \(V_j \subseteq V_{j+1}\).
- The union \(\bigcup_{j \in \mathbb{Z}} V_j\) is dense in \(L^2(\mathbb{R})\).
- \(\bigcap_{j \in \mathbb{Z}} V_j = \{0\}\).
- For every \(j \in \mathbb{N}\), we have \(f \in V_j \iff f(\cdot - k) \in V_{j+1}\).
- There is a function \(\varphi \in L^2(\mathbb{R})\) such that the translates \(\langle \varphi(\cdot - k) \rangle_{k \in \mathbb{Z}}\) constitute an ONB of the scaling space \(V_0\).

The function \(\varphi\) is called scaling function and the spaces \(V_j\) are called scaling (or approximation) spaces at scale \(j\). To each scaling space \(V_j\), one can associate a wavelet (or detail) space \(W_j\), that are defined to be the orthogonal complements of \(V_j\) in \(V_{j+1}\), i.e. \(V_{j+1} = V_j \oplus W_j\). Because of the above properties it holds that, for each \(j \in \mathbb{Z}\), the functions \(t \mapsto \varphi_{j,k}(t) := 2^{j/2} \varphi(2^j t - k)\) constitute an ONB for the scaling space \(V_j\). One can also show that from the existence of the scaling function it follows that there exists a so-called generating wavelet (or mother wavelet) \(\psi \in L^2(\mathbb{R})\) such that, for each \(j \in \mathbb{Z}\), the functions \(t \mapsto \psi_{j,k}(t) := 2^{j/2} \varphi(2^j t - k)\) constitute an ONB for the spaces \(W_j\). Hence, we have that, for every \(j \in \mathbb{Z}\), the following mappings are bijections:

\[
V_j \to \ell^2(\mathbb{Z}): f \mapsto (\langle \varphi_{j,k}, f \rangle \mid k \in \mathbb{Z}),
\]

\[
W_j \to \ell^2(\mathbb{Z}): f \mapsto (\langle \psi_{j,k}, f \rangle \mid k \in \mathbb{Z}).
\]

The above constructions provide the following decompositions of the signal space \(L^2(\mathbb{R})\) into the sum of the scaling spaces \(V_j\) and the wavelet spaces \(W_j\), or into a sum of only wavelet spaces:

\[
L^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j \geq 0} W_j = \bigoplus_{j \in \mathbb{Z}} W_j.
\]

From this decomposition we immediately get the following decompositions of signals \(f \in L^2(\mathbb{R})\):

\[
f = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{0,k} \rangle \varphi_{0,k} + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}.
\]

The coefficients of the above decomposition are called the wavelet and scaling coefficients of \(f\), respectively, and the corresponding mapping that maps \(f\) to those coefficients is called the wavelet transform of \(f\). A detailed construction of orthogonal (and biorthogonal) wavelet basis together with many interesting details may be found, for example, in [23, 30, 75].
A.2. Wavelets in higher dimension

Wavelet bases in higher dimensions can be defined by taking tensor products of the one-dimensional wavelet and scaling functions. As a concrete example consider \( L^2(\mathbb{R}^2) \cong L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \) and let \( \psi, \varphi : \mathbb{R} \rightarrow \mathbb{R} \) be the generating wavelet and scaling functions, respectively. Then, an orthogonal wavelet basis of \( L^2(\mathbb{R}^2) \) is defined by translates and scaled versions of the (tensor product) functions

\[
\begin{align*}
\psi^{(1)} : (x, y) &\mapsto \psi(x)\varphi(y) \\
\psi^{(2)} : (x, y) &\mapsto \varphi(x)\psi(y) \\
\psi^{(3)} : (x, y) &\mapsto \psi(x)\psi(y).
\end{align*}
\]

The corresponding scaling functions at scale \( j \in \mathbb{Z} \) of two variables are defined as translated and scaled versions of

\[
\varphi^{(0)}(x, y) := \varphi(x)\varphi(y).
\]

Analogously to the above construction in two dimensions, one can define a wavelet ONB in \( \mathbb{R}^d \) by a tensor product construction. In this case, we get \( 2^d - 1 \) different generating wavelets \( \psi^{(c)} \) and one generating scaling function \( \varphi^{(0)} \). Therefore, a wavelet ONB in \( \mathbb{R}^d \) is a family \( (\psi_\lambda)_{\lambda \in \Lambda} \) in \( L^2(\mathbb{R}^d) \), where the index \( \lambda = (j, k, \varepsilon) \) consists of three parameters, the scale index \( j \in \mathbb{Z} \), the location index \( k \in \mathbb{Z}^d \), and the orientation index \( \varepsilon \in \{1, 2, \ldots, 2^d - 1\} \). The wavelet basis elements are again defined as dilates and translates of the generating wavelets \( \psi^{(c)} \):

\[
\psi_{j,k,\varepsilon}(x) = 2^{-jd/2} \psi^{(c)}(2^j x - k).
\]

If we also add the index set \( \Lambda_{-1} := \{0\} \times \mathbb{Z}^d \) and consider the scaling functions defined by \( \varphi_{j,k}(x) = 2^{-jd/2} \varphi^{(0)}(2^j x - k) \), then we get similar decompositions to the one-dimensional case:

\[
f = \sum_{k \in \mathbb{Z}^d} \langle f, \varphi_{0,k} \rangle \varphi_{0,k} + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}^d} \sum_{\varepsilon=1}^{2^d-1} \langle f, \psi_{j,k,\varepsilon} \rangle \psi_{j,k,\varepsilon} = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^d} \sum_{\varepsilon=1}^{2^d-1} \langle f, \psi_{j,k,\varepsilon} \rangle \psi_{j,k,\varepsilon}.
\]

For \( f \in L^2(\mathbb{R}^d) \) we define the wavelet transform as

\[
(Wf) : \Lambda \rightarrow \mathbb{R} : \lambda \mapsto (Wf)(\lambda) := \langle \psi_\lambda, f \rangle.
\]

Then the wavelet transform \( W : L^2(\mathbb{R}^d) \rightarrow \ell^2(\Lambda) : f \mapsto Wf \) is an isometric isomorphism with the reproducing property \( f = W^* Wf \).

A.3. Besov spaces

Orthonormal wavelet bases can be used to characterize smoothness of functions in terms of the decay properties of wavelet coefficients. In particular, wavelets provide a convenient and a numerically efficient way to characterize Besov spaces and calculate Besov smoothness of functions, see [23]. Skipping the details, we here only mention that the Besov spaces \( B^{s}_{p,q}(\mathbb{R}^d) \) can be defined by the condition that the Besov norm

\[
\|f\|_{B^{s}_{p,q}(\mathbb{R}^d)} = \left( \sum_{j \in \mathbb{N}} 2^{jsq} \|Wf(j, \cdot)\|_p^q \right)^{1/q} \quad \text{with} \quad s = r + \frac{d}{2} - \frac{d}{p},
\]

where \( Wf(j, x) = \sum_{k \in \mathbb{Z}^d} \langle f, \psi_{j,k} \rangle \psi_{j,k}(x) \) are the wavelet coefficients and \( (\psi_{j,k})_k \) is the wavelet basis.
is well defined and finite. We note that, the given definition of $\| \cdot \|_{B^r_{p,q}}$ is an equivalent norm to the classical definition of Besov norms. The above characterization holds as long as the generating wavelet has $m > r$ vanishing moments and is $m$-times continuously differentiable.

In section 4, we use the above characterization of Besov norms in terms of wavelet coefficients in order to evaluate the performance of our method for functions that lie in balls of Besov spaces $B^r_{p,q}(\mathbb{R}^d)$ for some given norm parameters $p, q \geq 1$ and a smoothness parameter $r \geq 0$.

**ORCID iDs**

Markus Haltmeier https://orcid.org/0000-0001-5715-0331

**References**


[7] Beard P 2011 Biomedical photoacoustic imaging Interface Focus 1 602–31


[61] Hristova Y, Kuchment P and Nguyen L 2008 Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media Inverse Problems 24 055006


[63] Johnstone I M 2015 Gaussian estimation: sequence and wavelet models Draft of a Monograph Version October 6


[70] Kunyansky L A 2007 A series solution and a fast algorithm for the inversion of the spherical mean Radon transform Inverse Problems 23 S11–20

[71] Lorenz D 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints J. Inverse Ill-Posed Problems 16 463–78


[73] Narayanan E K and Rakesh 2010 Spherical means with centers on a hyperplane in even dimensions Inverse Problems 26 035014
SSR Teckhn. Kibern. 3 50–60
[79] Nguyen L V 2015 How strong are streak artifacts in limited angle computed tomography? Inverse
Problems Imaging 31 055003
[80] Nguyen L V 2015 On artifacts in limited data spherical Radon transform: flat observation surfaces
[81] Nguyen L V and Kunyansky L A 2016 A dissipative time reversal technique for photoacoustic
tomography in a cavity SIAM J. Imaging Sci. 9 748–69
28 202–20
9 319–37
[84] Palamodov V P 2012 A uniform reconstruction formula in integral geometry Inverse Problems
28 065014
algorithms for limited view photoacoustic tomography with line detectors Inverse Problems
23 881–94
integrating area and line detectors Photoacoustic Imaging and Spectroscopy (Boca Raton, FL:
CRC Press) ch 20, pp 251–63
optoacoustic imaging J. Opt. Soc. Am. 112 1536–44
[88] Pan X, Sidky E Y and Vannier M 2009 Why do commercial ct scanners still employ traditional,
filtered back-projection for image reconstruction? Inverse Problems 25 123009
[89] Poudel J, Matthews T P, Li L, Anastasio M A and Wang L V 2017 Mitigation of artifacts due to
isolated acoustic heterogeneities in photoacoustic computed tomography using a variable data
[90] Provost J and Lesage F 2009 The application of compressed sensing for photo-acoustic tomography
problems with sparsity constraints Numer. Math. 104 177–203
[92] Rieder A 1997 A wavelet multilevel method for ill-posed problems stabilized by Tikhonov
regularization Numer. Math. 75 501–22
tomography: a review Curr. Med. Imaging Rev. 9 318
Imaging (Applied Mathematical Sciences vol 167) (New York: Springer)
[95] Schwab J, Pereverzyev S Jr and Haltmeier M 2016 A Galerkin least squares approach for
wavelets and curvelets Proc. SPIE 4478 9
[97] Stefanov P and Uhlmann G 2013 Is a curved flight path in SAR better than a straight one? SIAM
J. Appl. Math. 73 1596–612
[99] Treby B E and Cox B T 2010 k-wave: Matlab toolbox for the simulation and reconstruction of
photoacoustic wave-fields J. Biomed. Opt. 15 021314
(New York: Springer)
Inf. Theory 59 2001–16
principles Handbook of Mathematical Methods in Imaging (Berlin: Springer) pp 781–815
for three-dimensional optoacoustic tomography using radially symmetric expansion functions