DISJOINT SPARSITY FOR SIGNAL SEPARATION AND APPLICATIONS TO HYBRID INVERSE PROBLEMS IN MEDICAL IMAGING

GIOVANNI S. ALBERTI AND HABIB AMMARI

Abstract. The main focus of this work is the reconstruction of the signals $f$ and $g_i$, $i = 1, \ldots, N$, from the knowledge of their sums $h_i = f + g_i$, under the assumption that $f$ and the $g_i$'s can be sparsely represented with respect to two different dictionaries $A_f$ and $A_g$. This generalizes the well-known “morphological component analysis” to a multi-measurement setting. The main result of the paper states that $f$ and the $g_i$'s can be uniquely and stably reconstructed by finding sparse representations of $h_i$ for every $i$ with respect to the concatenated dictionary $[A_f, A_g]$, provided that enough incoherent measurements $g_i$ are available. The incoherence is measured in terms of their mutual disjoint sparsity.

This method finds applications in the reconstruction procedures of several hybrid imaging inverse problems, where internal data are measured. These measurements usually consist of the main unknown multiplied by other unknown quantities, and so the disjoint sparsity approach can be directly applied. As an example, we show how to apply the method to the reconstruction in quantitative photoacoustic tomography, also in the case when the Grüneisen parameter, the optical absorption and the diffusion coefficient are all unknown.

1. Introduction

Hybrid, or coupled-physics, inverse problems have been extensively studied over the last years, both from the mathematical and the experimental points of view. A hybrid imaging modality consists in the combination of two types of techniques, one exhibiting the high contrast of tissues and a second one providing high resolution. Thus, the main drawbacks of the standard imaging modalities can be overcome, at least theoretically. Many combinations have been considered, such as optical and acoustic waves (photoacoustic tomography [40]), electric currents and ultrasounds (ultrasound modulated EIT [8]) or microwaves and ultrasounds (thermoacoustic tomography [40]). The reader is referred to [7, 39, 16, 5, 6, 47] for a review on the mathematical aspects related to hybrid imaging problems.

In general, the inversion for such problems involves two steps. In the first step, an inverse problem related to the high resolution wave provides certain internal measurements. Such internal data are usually functionals of the unknown parameters and of certain solutions of partial differential equations (the unknowns are...
normally the coefficients of the PDE). In the second step, the unknown parameter has to be reconstructed from the knowledge of the internal measurements. This is sometimes referred to as the quantitative step, since the information on the tissue properties contained in the internal data is only qualitative. In this paper we suppose that the first inversion has been performed, and focus only on the second step.

The quantitative step is normally solved with PDE-based methods, by combining the internal data with the PDE modeling the problem. Such approach is sometimes very powerful in the reconstruction [21, 13, 16, 2, 3]. However, there may be difficulties in using these methods. First, the PDE model may be accurate only in some circumstances but not in others [16]. Second, even if the PDE model is accurate, there may be too many unknowns to have unique reconstruction [19]. Third, even in cases when the reconstruction is unique, this may require the differentiation of the data [13, 14, 4], which is known to be an unstable process, or may require additional assumptions to be satisfied [13, 1, 17, 14, 4].

The main focus of this paper is an alternative approach to such problem based on the use of sparse representations, as it was first done by Rosenthal et al. in quantitative photoacoustic tomography (QPAT) [45]. The internal data in a domain \( \Omega \) can often be expressed as the product of the unknown(s) and an expression involving the solutions of the PDE. (For example, in QPAT the internal data have the form \( H = \Gamma \mu u \), where \( \Gamma \) is the Grüneisen parameter, \( \mu \) is the optical absorption and \( u \) is the light intensity.) Taking the logarithm, the inversion corresponds to recovering two functions \( f \) and \( g \) from the knowledge of their sum

\[
h(x) = f(x) + g(x), \quad x \in \Omega.
\]

This problem is, in general, clearly unsolvable. However, it is possible to exploit the different levels of smoothness of \( f \) and \( g \). Indeed, since \( f \) represents a property of the medium, such as the log conductivity, it is typically highly discontinuous. On the other hand, \( g \) is an expression involving the solutions of a PDE, and as such enjoys higher regularity properties. As a consequence, \( f \) and \( g \) have different features, and this can be used to separate them by using a sparse representation approach.

Two signals \( f, g \in \mathbb{R}^n \) can be reconstructed from the knowledge of their sum \( h = f + g \) provided that they have different characteristics. More precisely, they need to be sparsely represented, i.e., with few atoms, with respect to two incoherent dictionaries \( A_f \) and \( A_g \). This method is usually called “morphological component analysis” (MCA), and was introduced by Starck et al. in [49] (see [23, 37, 22] for several generalizations).

In this work, motivated by hybrid imaging techniques, where multiple measurements with the parameters fixed can be taken, we extend this method to a multi-measurement setting. In general terms, this corresponds to the reconstruction of \( f \) (and \( g_i \)) from the knowledge of their sums

\[
h_i = f + g_i, \quad i = 1, \ldots, N.
\]

We prove that the MCA approach gives unique and stable reconstruction, provided that enough incoherent measurements \( g_i \)'s are available. The incoherence is measured in terms of their mutual disjoint sparsity. In vague terms, the atoms from \( A_g \) used to represent \( g_i \) should change for different measurements (see Definition 2 for
DISJOINT SPARSITY FOR HYBRID IMAGING

Numerical simulations show that taking several solutions to the relevant PDE yields the necessary incoherence.

As an example, we discuss the inversion for QPAT, both in the simpler case when \( \Gamma = 1 \) and in the case with non-constant \( \Gamma \) in the diffusive regime for light propagation. For the \( \Gamma = 1 \) case, this method has the advantages of being very robust to noise and of not requiring a particular model for light propagation, if compared to the PDE-based approaches [21, 18]. In the case when \( \Gamma \neq 1 \) there is no uniqueness in the reconstruction, even with multiple measurements [19]; if the parameters are piecewise constant, uniqueness can be guaranteed, but the inversion may be very sensible to noise [12]. We propose a combination of the disjoint sparsity signal separation method and of the PDE method, which provides a satisfactory reconstruction, without requiring piecewise constant parameters. Numerical simulations are provided.

This work is structured as follows. In Section 2 we recall the basic notions related to sparse representations and present the method of morphological component analysis. In Section 3 the signal separation method based on multiple measurements and disjoint sparsity is described in detail and the main reconstruction result is proved. The numerical implementation of the method, together with the possible choices for the dictionaries \( A_f \) and \( A_g \) in hybrid imaging, are discussed in Section 4. Next, this method is applied to hybrid imaging in Section 5 and several numerical simulations are provided. Finally, some concluding remarks are contained in Section 6.

2. SPARSE REPRESENTATIONS AND MORPHOLOGICAL COMPONENT ANALYSIS

In this section, we introduce the basic notions related to sparse representations and morphological components analysis.

2.1. Introduction to sparse representations. This presentation follows [24]. Let \( n \) be the dimension of the signal space, namely we shall consider signals \( f \in \mathbb{R}^n \). Since in this paper we shall consider only images, we should think of \( n \) as being the resolution of the image, namely \( n = d \times d \), where \( d \) is the number of pixels in each row and column. However, in this section we shall use the more general notation \( f \in \mathbb{R}^n \), and think of a signal as a column vector of length \( n \). We now discuss how a signal can be represented as a superposition of given atoms in a fixed dictionary. More precisely, let \( A \in \mathbb{R}^{n \times m} \) be a dictionary, namely a collection of \( m \) atoms, that are the column vectors of \( A \). We assume \( m \geq n \) and that \( A \) has full rank. Thus, it is always possible to express \( f \) as a linear combination of these atoms, i.e. to write

\[
 f = Ay
\]

for some vector of coefficients \( y \in \mathbb{R}^m \). The most common situation is when \( m = n \) and \( A \) is an orthonormal basis: in this case the coefficient \( y \) is uniquely determined. However, the situation we are interested in is when \( m > n \). In this case the dictionary \( A \) is redundant, since \( f \) can be represented in many different ways as a combination of the atoms in \( A \). In other words, the system (1) is underdetermined and has many solutions \( y \in \mathbb{R}^m \).

The key observation is that this non-uniqueness can be exploited by selecting the best representation \( y \). One way to measure the quality of a representation \( y \) is given by its sparsity, which can be quantified by the number of non-zero coefficients
of \( y \)
\[
\|y\|_0 := \#\{\alpha \in \{1, \ldots, m\} : y(\alpha) \neq 0\},
\]
where the symbol \# denotes the cardinality of a set. The representation (\ref{eq:1}) is called sparse if \( \|y\|_0 \ll m \). From the theoretical point of view, the sparsest representation can be found by minimizing the following problem
\[
\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } Ay = f.
\]
The practical search for the minimum poses highly non trivial difficulties, and the description of the main issues goes beyond the scope of this work. Algorithms such as Matching Pursuit [41] or Basis Pursuit [29] can often be successfully used to find the sparsest solution. More details will be given in Section 4.

The choice of the dictionary \( A \) clearly plays a fundamental role in this context. Indeed, a signal \( f \) admits a sparse representation with respect to a dictionary \( A \) if \( f \) can be written as combination of few atoms in \( A \). Therefore, the dictionary \( A \) has to be chosen to capture the main features of the signals we consider. Many choices of dictionaries for images are available, and a detailed discussion is presented in Section 4.

In the presence of noise, it is helpful to consider a relaxation of (2) and to allow a small error between the signal \( f \) and its representation in terms of the atoms of \( A \). Thus, the minimization problem becomes
\[
\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|Ay - f\|_2 \leq \varepsilon,
\]
for some \( \varepsilon > 0 \), or equivalently
\[
\min_{y \in \mathbb{R}^m} \|y\|_0 + \lambda \|Ay - f\|_2,
\]
for a suitable Lagrange multiplier \( \lambda > 0 \).

2.2. Introduction to morphological component analysis (MCA). One of the relevant applications of sparse representations is related to the separation of a signal into its constitutive components, provided they have different features. We shall describe the method discussed in [49]. Suppose that a signal \( h \in \mathbb{R}^n \) can be written as a sum
\[
h = \tilde{f} + \tilde{g},
\]
with \( \tilde{f}, \tilde{g} \in \mathbb{R}^n \). The problem under consideration is the reconstruction of \( \tilde{f} \) and \( \tilde{g} \) from the knowledge of \( h \), under the assumption that \( \tilde{f} \) and \( \tilde{g} \) have distinctive characteristics. This assumption can be expressed in terms of sparse representations. Namely, suppose that there exist two dictionaries \( A_f \in \mathbb{R}^{n \times m_f} \) and \( A_g \in \mathbb{R}^{n \times m_g} \) such that:

1. \( \tilde{f} \) can be sparsely represented with respect to \( A_f \) but not with respect to \( A_g \);
2. \( \tilde{g} \) can be sparsely represented with respect to \( A_g \) but not with respect to \( A_f \).

Under these conditions, a strategy to find \( \tilde{f} \) and \( \tilde{g} \) may be to find a sparse representation \( y = [y_f \ y_g] \) of \( h \) with respect to the concatenated dictionary \( A = [A_f \ A_g] \) and then to write \( f = A_f y_f \) and \( g = A_g y_g \). As we have seen before, the sparse representation \( y \) is the minimum of the minimization problem
\[
\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 \quad \text{subject to } [A_f \ A_g] [y_f \ y_g] = h,
\]
or, in the presence of noise, of
\[
\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 + \lambda \| [A_f, A_g] \begin{bmatrix} y_f \\ y_g \end{bmatrix} - h \|_2.
\]

Even though this procedure is successful in many practical cases [49, 35, 45], a proof of the correct reconstruction, i.e., \( f = \tilde{f} \) and \( g = \tilde{g} \), is only valid in an ideal situation, which we now describe. In the case when \( A_f \) and \( A_g \) are both orthonormal bases, the proof is based on the following uncertainty principle.

**Proposition 1** ([34, Theorem 1]). Take a vector \( h \in \mathbb{R}^n \setminus \{0\} \) and suppose it has the following representations
\[
h = A_y A = B y_B
\]
with respect to two orthonormal bases \( A = [a_1, \ldots, a_n] \) and \( B = [b_1, \ldots, b_n] \). Then
\[
\|A_y\|_0 + \|B y_B\|_0 \geq 2/M,
\]
where \( M = \max_{i,j} |(a_i, b_j)_2| \) is the mutual coherence.

As a simple consequence of this result [34, Theorem 2], we have that if \( y^1 \in \mathbb{R}^{2n} \) and \( y^2 \in \mathbb{R}^{2n} \) are two representations of \( h \) with respect to the concatenated dictionary \( A = [A_f, A_g] \), then
\[
\|y^1\|_0 + \|y^2\|_0 \geq 2/M.
\]
Therefore, if \( \tilde{f} \) and \( \tilde{g} \) have representations \( \tilde{y}_f \) and \( \tilde{y}_g \) satisfying \( \|\tilde{y}_f\|_0 + \|\tilde{y}_g\|_0 < 1/M \), then the above method provides the correct reconstruction.

In practice, the assumption \( \|\tilde{y}_f\|_0 + \|\tilde{y}_g\|_0 < 1/M \) is almost never satisfied, and so the above argument remains only a theoretical speculation. However, when the multi-measurement case is considered, correct and stable reconstruction can be rigorously proved. This theoretical result is also validated by several numerical simulations. These aspects are discussed in the following sections.

3. **DISJOINT SPARSITY FOR MORPHOLOGICAL COMPONENT ANALYSIS**

Motivated by several hybrid imaging modalities (see Section 5), we generalize the MCA problem to a multi-measurement setting. The reader is referred to [23, 37, 22] for other similar variations.

Let \( h_1, \ldots, h_N \in \mathbb{R}^n \) be \( N \) signals that can be decomposed as
\[
h_i = \tilde{f} + \tilde{g}_i, \quad i = 1, \ldots, N,
\]
with \( \tilde{f}, \tilde{g}_i \in \mathbb{R}^n \). We want to study the problem of finding \( \tilde{f} \) and \( \tilde{g}_i \) from the knowledge of \( h_i \) for \( i = 1, \ldots, N \). The case \( N = 1 \) was discussed in the previous section. We shall show that as \( N \) becomes bigger, the above problem becomes much more treatable, and that the sparsity approach introduced before always provides the correct reconstruction, also in the presence of noise. As before, let \( A_f \) and \( A_g \) be the dictionaries with respect to which \( \tilde{f} \) and \( \tilde{g}_i \) have sparse representations, respectively. Note that all the \( \tilde{g}_i \)'s can be sparsely represented with the same dictionary \( A_g \). Assume that the atoms of \( A_f \) and \( A_g \) are linearly independent. As a consequence, \( A_f \) and \( A_g \) admit left inverses, which with an abuse of notation we shall denote by \( A_f^{-1} \) and \( A_g^{-1} \), respectively.
The reconstruction method applied to this case consists in the minimization of
\[
\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 + \lambda \sum_{i=1}^{N} \left\| \begin{bmatrix} A_f & y_i \end{bmatrix} \begin{bmatrix} y_f \\ y_i \end{bmatrix} - h_i \right\|_2,
\]
where we have used the notation \( y = \begin{bmatrix} y_f, & y_1^g, & \ldots, & y_N^g \end{bmatrix} \). Here, the superscript \( t \) denotes the transpose. To model the case with added noise, we write
\[
h_i = \tilde{f} + \tilde{g}_i + n_i, \quad i = 1, \ldots, N,
\]
where \( \tilde{f} \) and the \( \tilde{g}_i \)'s represent the true signals, the \( h_i \)'s are the measured signals and \( n_i \) is such that
\[
\|n_i\|_2 \leq \varepsilon, \quad i = 1, \ldots, N
\]
for some small \( \varepsilon > 0 \). We shall prove that under suitable assumptions, a minimizer of (3) provides the correct reconstruction, up to a factor that is small in \( \varepsilon \).

In the applications we have in mind (Section 5), the signal \( \tilde{f} \) represents (the logarithm of) a physical constitutive parameter, while the \( \tilde{g}_i \)'s usually quantify the injected fields, e.g., the electric field or the light intensity. As such, \( \tilde{f} \) is given and fixed, and we have no control on it. On the other hand, the \( \tilde{g}_i \)'s come from the measurements, and can be indirectly controlled. More precisely, the \( \tilde{g}_i \)'s depend on the solutions of a certain PDE, whose coefficients are unknown, but whose boundary values can be chosen: in this sense the \( \tilde{g}_i \)'s can be controlled. It is therefore natural to give some assumptions on the \( \tilde{g}_i \)'s.

The main requirements are that the \( \tilde{g}_i \)'s should be sufficiently many and incoherent. This will be mainly expressed by means of their disjoint sparsity with respect to \( A_g \). (Disjoint sparsity was used in [22], while joint sparsity has been extensively used in compressive sensing [30, 28].) We shall therefore write
\[
\tilde{f} \approx A_f \tilde{y}_f, \quad \tilde{g}_i \approx A_g \tilde{y}_i^g, \quad i = 1, \ldots, N.
\]
The approximation allows a small error between the true signals and their sparse representations. In terms of the coefficient vectors \( \{\tilde{y}_1^g, \ldots, \tilde{y}_N^g\} \), the required assumptions can be written as follows.

Definition 2. Take \( \beta > 0 \), \( N \in \mathbb{N}^* \), \( \tilde{y}_f \in \mathbb{R}^{m_f} \) and \( \tilde{y}_1^g, \ldots, \tilde{y}_N^g \in \mathbb{R}^{m_g} \). We say that \( \{\tilde{y}_1^g, \ldots, \tilde{y}_N^g\} \) is a \( \beta \)-complete set of measurements if the following two conditions hold true:

**CS1:** if \( |\tilde{y}_i^g(\alpha) - \tilde{y}_j^g(\alpha)| \leq \beta \) for some \( \alpha \in \{1, \ldots, m_g\} \) then \( i = j \);

**CS2:** for every \( \eta > 0 \) and \( p \in \mathbb{R}^n \) such that \( \{\alpha : |(A_g^{-1} p)(\alpha)| \geq \eta\} \neq \emptyset \) there holds
\[
\#\text{supp} (A_f^{-1} p) \setminus \text{supp} \tilde{y}_f + \sum_{i=1}^{N} \#\{\alpha : |(A_g^{-1} p)(\alpha)| \geq \eta\} \setminus \text{supp} \tilde{y}_i^g > 6 + \# \bigcup_{i=1}^{N} \text{supp} \tilde{y}_i^g + \|\tilde{y}_f\|_0.
\]

1 Similar assumptions of enough independent measurements are required also when using PDE methods for hybrid inverse problems (see § 5.1).
Lemma 4. \( \text{Proofs are immediate.} \)

In the proof of this theorem we shall make use of the following properties, whose proofs are immediate.

Theorem 3. \( \text{Take } \beta > 0, N \in \mathbb{N}^*, \delta \in (0, 1), \hat{x}_f \in \mathbb{R}^{m_f} \) and \( \tilde{y}_f^1, \ldots, \tilde{y}_f^N \in \mathbb{R}^{m_g} \).

Assume that \( \{\tilde{y}_f^1, \ldots, \tilde{y}_f^N\} \) is \( \beta \)-complete. \( \text{There exists } \varepsilon > 0 \text{ depending only on } N, \delta, A_g, \|\tilde{y}_f\|_0 \text{ and } \beta \text{ such that for every } \varepsilon \leq \tilde{\varepsilon} \text{ the following is true.} \)

Assume that \( \hat{f}, \hat{g}, n_i \in \mathbb{R}^n \text{ satisfy } (3) \) and

\[
\|A_f \hat{y}_f - \hat{f}\|_2 \leq \varepsilon, \quad \|A_g \hat{y}_g - \hat{g}_i\|_2 \leq \varepsilon, \quad i = 1, \ldots, N, \tag{7}
\]

and let \( y_f \in \mathbb{R}^{m_f} \) and \( y_g^i \in \mathbb{R}^{m_g} \) realize the minimum of

\[
\min_{y \in \mathbb{R}^{m_f} + N m_g} \|y\|_0 + \frac{1}{\varepsilon N} \sum_{i=1}^N \|A_f y_f + A_g y_g^i - h_i\|_2, \tag{8}
\]

where \( h_i \) is given by \( (4) \). Then

\[
\|A_f y_f - \hat{f}\|_2 \leq C \varepsilon^{1-\delta}, \quad \|A_g y_g^i - \hat{g}_i\|_2 \leq C \varepsilon^{1-\delta}, \quad i = 1, \ldots, N
\]

for some \( C > 0 \) depending only on \( A_g \).

Thanks to this result, the reconstruction can be performed by minimizing \( (5) \) and then taking \( \hat{f} \approx A_f y_f \) and \( \hat{g}_i \approx A_g y_g^i \). The practical details of such minimization will be discussed in Section 4.

In the proof of this theorem we shall make use of the following properties, whose proofs are immediate.

Lemma 4. \( \text{The following properties hold true.} \)
(1) For every \(a, b \in \mathbb{R}^m\) there holds
\[
\|a + b\|_0 = \|a\|_0 + \#(\mathrm{supp} \ b \setminus \mathrm{supp} \ a) - \#\{\alpha : a(\alpha) = -b(\alpha) \neq 0\}.
\]

(2) Let \(A \in \mathbb{R}^{n \times m}\) be a left-invertible matrix. There exist \(C_1, C_2 > 0\) such that
\[
C_1 \|y\|_2 \leq \|Ay\|_2 \leq C_2 \|y\|_2, \quad y \in \mathbb{R}^m.
\]

We are now ready to prove Theorem [3].

Proof of Theorem [3]. Take \(\varepsilon > 0\) such that \(2\varepsilon^{1-\delta} \leq \beta\), assume that \(\hat{f}, \hat{g}, n_i \in \mathbb{R}^n\) satisfy (5) and (7) and let \(y_f \in \mathbb{R}^{m_f}\) and \(y_g' \in \mathbb{R}^{m_g}\) realize the minimum of (8).

Write \(f := A_f y_f, g_i := A_g y_i, e_f := y_f - \hat{y}_f, e_g := y_i - \hat{y}_g, p := A_f e_f, e_g := -A_g^{-1} p\) and \(r := e_i - e_g\).

Since \(y_f\) and \(y_g'\) realize the minimum of (8) there holds
\[
\|y\|_0 + \frac{1}{\varepsilon N} \sum_{i=1}^N \|f + g_i - h_i\|_2 \leq \|\hat{y}\|_0 + \frac{1}{\varepsilon N} \sum_{i=1}^N \|A_f \hat{y}_f + A_g \hat{y}_g' - h_i\|_2,
\]
where we have set \(\hat{y} = [\hat{y}_f, \hat{y}_g, \ldots, \hat{y}_g^N]\). By (4), (5) and (7) we have
\[
\|A_f \hat{y}_f + A_g \hat{y}_g' - h_i\|_2 \leq \|A_f \hat{y}_f + A_g \hat{y}_g' - (\hat{f} + \hat{g}_i + n_i)\|_2 \leq 3\varepsilon.
\]

Similarly
\[
\|f + g_i - h_i\|_2 = \|(f - \hat{f}) + (g_i - \hat{g}_i) - n_i\|_2
\]
\[
= \|p + (A_f \hat{y}_f - \hat{f}) + A_g e_g^i + (A_g \hat{y}_g' - \hat{g}_i) - n_i\|_2,
\]
whence
\[
\|f + g_i - h_i\|_2 \geq \|A_g r^i\|_2 - 3\varepsilon
\]
Inserting (10) and (11) into (9) we obtain
\[
\|y\|_0 + \frac{1}{\varepsilon N} \sum_{i=1}^N \|A_g r^i\|_2 \leq \|\hat{y}\|_0 + 6, \text{ or, equivalently,}
\]
\[
((\|y\|_0 - \|\hat{y}\|_0) + \sum_{i=1}^N (\|y_g\|_0 - \|\hat{y}_g\|_0)) + \frac{1}{\varepsilon N} \sum_{i=1}^N \|A_g r^i\|_2 \leq 6.
\]

Thus, since \(y_f = \hat{y}_f + e_f\) and \(y_g^i = \hat{y}_g^i + e_g^i\), Lemma [4] yields
\[
\|(\hat{y}_f\|_0 - \|\hat{y}\|_0) + \sum_{i=1}^N (\|y_g\|_0 - \|\hat{y}_g\|_0)) + \frac{1}{\varepsilon N} \sum_{i=1}^N \|A_g r^i\|_2 \leq 6.
\]

Observe now that
\[
\#\{\alpha : \hat{y}_f(\alpha) = -e_f(\alpha) \neq 0\} \leq \|\hat{y}_f\|_0,
\]
\[
\mathrm{supp} e_g^i = \mathrm{supp} (e_g + r^i) \supseteq \{\alpha : |e_g(\alpha)| \geq 2\varepsilon^{1-\delta}, |r^i(\alpha)| \leq \varepsilon^{1-\delta}\},
\]
\[
\#\{\alpha : \hat{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0\} = \#\{\alpha : \hat{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| > \varepsilon^{1-\delta}\}
\]
\[
+ \#\{\alpha : \hat{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| \leq \varepsilon^{1-\delta}\}.
\]
Since \(2\varepsilon^{1-\delta} \leq \beta\) and \(\{\tilde{y}_g^1, \ldots, \tilde{y}_g^N\}\) is \(\beta\)-complete, by Definition 2 (condition CS1), we have

\[
\sum_{i=1}^{N} \#\{\alpha : \tilde{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| \leq \varepsilon^{1-\delta}\} = \# \bigcup_{i=1}^{N} \{\alpha : \tilde{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| \leq \varepsilon^{1-\delta}\} \leq \# \bigcup_{i=1}^{N} \text{supp} \tilde{y}_g^i.
\]

Inserting (13) and (14) into (12) gives

\[
\#(\text{supp} \ y_f \setminus \text{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\{\alpha : |e_g(\alpha)| \geq 2\varepsilon^{1-\delta}, |r^i(\alpha)| \leq \varepsilon^{1-\delta}\} \setminus \text{supp} \tilde{y}_g^i
\]

\[
- \sum_{i=1}^{N} \#\{\alpha : \tilde{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| > \varepsilon^{1-\delta}\} + \frac{C(A_g)}{\varepsilon N} \sum_{i=1}^{N} \|r^i\|_2
\]

\[
\leq 6 + \|\tilde{y}_f\|_0 + \# \bigcup_{i=1}^{N} \text{supp} \tilde{y}_g^i.
\]

For brevity of notation, write \(z(i, \varepsilon) = \#\{\alpha : |r^i(\alpha)| > \varepsilon^{1-\delta}\}\). In view of the inequalities

\[
\#\{\alpha : \tilde{y}_g^i(\alpha) = -e_g^i(\alpha) \neq 0, |r^i(\alpha)| > \varepsilon^{1-\delta}\} \leq z(i, \varepsilon),
\]

\[
\|r^i\|_2^2 = \sum_{\alpha} |r^i(\alpha)|^2 \geq \sum_{\alpha : |r^i(\alpha)| > \varepsilon^{1-\delta}} |r^i(\alpha)|^2 \geq \varepsilon^{2(1-\delta)} z(i, \varepsilon),
\]

we obtain

\[
\#(\text{supp} \ y_f \setminus \text{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\{\alpha : |e_g(\alpha)| \geq 2\varepsilon^{1-\delta}, |r^i| \leq \varepsilon^{1-\delta}\} \setminus \text{supp} \tilde{y}_g^i
\]

\[
+ \frac{C(A_g)}{\varepsilon^{\delta} N} \sum_{i=1}^{N} z(i, \varepsilon) \frac{1}{2} - \sum_{i=1}^{N} z(i, \varepsilon) \leq 6 + \|\tilde{y}_f\|_0 + \# \bigcup_{i=1}^{N} \text{supp} \tilde{y}_g^i.
\]

We now claim that \(z(i, \varepsilon) = 0\) if \(\varepsilon\) is small enough. The previous inequality implies

\[
\frac{C(A_g)}{\varepsilon^{\delta} N} \sum_{i=1}^{N} z(i, \varepsilon) \frac{1}{2} - \sum_{i=1}^{N} z(i, \varepsilon) \leq 6 + \|\tilde{y}_f\|_0 + \# \bigcup_{i=1}^{N} \text{supp} \tilde{y}_g^i.
\]

If \(z(i, \varepsilon) = 0\) for all \(i\), the claim is proved. Assume that \(z(i, \varepsilon) > 0\) for some \(i\). Rearranging the terms and observing that \(z(i, \varepsilon) \leq m_g\) and \(\# \bigcup \text{supp} \tilde{y}_g^i \leq m_g\) we obtain

\[
\frac{C(A_g)}{\varepsilon^{\delta} N} \leq 6 + \|\tilde{y}_f\|_0 + (N + 1)m_g,
\]

whence

\[
\varepsilon^{\delta} \geq C(A_g)N^{-1}(6 + \|\tilde{y}_f\|_0 + (N + 1)m_g)^{-1}.
\]

This leads us to define

\[
\bar{\varepsilon} = \min \left(\frac{C(A_g)N^{-1}(6 + \|\tilde{y}_f\|_0 + (N + 1)m_g)^{-1}}{2}, (\beta/2)^{1-\tau}\right).
\]

As a consequence, \(2\varepsilon^{1-\delta} \leq \beta\) and for every \(\varepsilon \leq \bar{\varepsilon}\) and every \(i = 1, \ldots, N\) there holds \(z(i, \varepsilon) = 0\).
Suppose now that \( \varepsilon \leq \bar{\varepsilon} \). Note that \( z(i, \varepsilon) = 0 \) implies that \( |r^i(\alpha)| \leq \varepsilon^{1-\delta} \) for all \( \alpha \). As a result, by (15) we have

\[
\#(\text{supp} (A_f^{-1}p) \setminus \text{supp} \tilde{y}_f) + \sum_{i=1}^N \#(\{\alpha : |(A_g^{-1}p)(\alpha)| \geq 2\varepsilon^{1-\delta}\} \setminus \text{supp} \tilde{y}_g^i) \leq 6 + ||\tilde{y}_f||_0 + \# \bigcup_{i=1}^N \text{supp} \tilde{y}_g^i.
\]

Therefore, since \( \{\tilde{y}_g^1, \ldots, \tilde{y}_g^N\} \) is \( \beta \)-complete, by Definition 2 (condition CS2) we obtain that \( \{\alpha : |(A_g^{-1}p)(\alpha)| \geq 2\varepsilon^{1-\delta}\} = \emptyset \). Thus \( |e_g(\alpha)| \leq 2\varepsilon^{1-\delta} \) for all \( \alpha \). Recall that \( |r^i(\alpha)| \leq \varepsilon^{1-\delta} \) for all \( \alpha \). Hence

\[
(e_g)\|_2 \leq C(m_g)\varepsilon^{1-\delta}, \quad |r^i|\|_2 \leq C(m_g)\varepsilon^{1-\delta}, \quad i = 1, \ldots, N.
\]

This implies \( \|y_g^i - \tilde{y}_g^i\|_2 \leq C(m_g)\varepsilon^{1-\delta} \) for every \( i \). As a consequence, by Lemma 4 \( \|A_gy_g^i - A_g\tilde{y}_g^i\|_2 \leq C(A_g)\varepsilon^{1-\delta} \). In view of (7) (since \( \varepsilon \leq 1 \) and \( g_i = A_gy_g^i \), this gives

\[
\|g_i - \tilde{g}_i\|_2 \leq C(A_g)\varepsilon^{1-\delta}, \quad i = 1, \ldots, N.
\]

This proves the desired bound for \( g_i \). It remains to prove the corresponding estimate for \( f = A_fy_f \). In order to do this, observe that by (7), (11) and (16) there holds

\[
\|f + g_i - (\tilde{f} + \tilde{g}_i)\|_2 = \|(A_fy_f - \tilde{f}) + A_gr^i + (A_gy_g^i - \tilde{y}_g^i)\|_2 \leq C(A_g)\varepsilon^{1-\delta},
\]

whence by (17)

\[
\|f - \tilde{f}\|_2 \leq \|f + g_i - (\tilde{f} + \tilde{g}_i) + (\tilde{g}_i - g_i)\|_2 \leq C(A_g)\varepsilon^{1-\delta}.
\]

This concludes the proof. \( \square \)

4. Numerical implementation

4.1. Orthogonal matching pursuit. The simplest available algorithm for the minimization of problems of the type

\[
\min_{y \in \mathbb{R}^n} \|y\|_0 \quad \text{subject to} \quad \|Ay - f\|_2 \leq \varepsilon,
\]

for \( f \in \mathbb{R}^n \) and \( A \in \mathbb{R}^{n \times m} \), is the orthogonal matching pursuit (OMP) [41, 24]. This algorithm belongs to a wider class of methods called greedy algorithms, in which the set of the used atoms of \( A \) is increased step by step. In OMP, the best coefficients for the atoms are recomputed at each iteration, which makes it more efficient compared to the standard matching pursuit. Greedy algorithms have been proved to perform well in the minimization of the above problem [51]. The reader is referred to [24] for more details on this topic.

The adaptation of OMP to the problem of our interest is quite straightforward. By Theorem 3 we need to minimize (8) or, equivalently,

\[
\min_{y \in \mathbb{R}^{n \times m_y}} \|y\|_0 \quad \text{subject to} \quad \|A_fy_f + A_gy_g^i - h_i\|_2 \leq \varepsilon, \quad i = 1, \ldots, N;
\]

\[
\min_{y \in \mathbb{R}^{n \times m_y}} \|y\|_0 \quad \text{subject to} \quad \|A_fy_f + A_gy_g^i - h_i\|_2 \leq \varepsilon, \quad i = 1, \ldots, N;
\]

\[
\min_{y \in \mathbb{R}^{n \times m_y}} \|y\|_0 \quad \text{subject to} \quad \|A_fy_f + A_gy_g^i - h_i\|_2 \leq \varepsilon, \quad i = 1, \ldots, N;
\]
given $N$ signals $h_i \in \mathbb{R}^n$ and two dictionaries $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$. Setting

$$A = \begin{bmatrix} A_f & A_g & 0 & \cdots & 0 \\ A_f & 0 & A_g & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ A_f & 0 & \cdots & 0 & A_g \end{bmatrix}, \quad y = \begin{bmatrix} y_f \\ y_1^g \\ y_2^g \\ \vdots \\ y_N^g \end{bmatrix} \quad \text{and} \quad h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix},$$

the above problem is equivalent to

$$\min_{y \in \mathbb{R}^{m_f + N_m_g}} \|y\|_0 \quad \text{subject to} \quad \|Ay - h\|_2 \leq \sqrt{N} \varepsilon,$$

for which the OMP can be applied directly, as discussed above.

4.2. Dictionaries for image content. Let us now discuss what choices may be done for the dictionaries $A_f$ and $A_g$ in the context we are interested in, namely hybrid imaging inverse problems. As we shall see in Section 5, in such problems the signal $\tilde{f}$ will typically represent (the logarithm of) a constitutive parameter of the tissue under investigation, like for instance the electric permittivity, electric conductivity or the optical absorption. As such, the image $\tilde{f}$ can be supposed to be piecewise smooth: the discontinuities are usually the inclusions we would like to determine. On the other hand, the signals $\tilde{g}_i$’s usually represent the injected fields, such as the electric field or the light intensity, and are the solutions to certain PDE. As such, they enjoy higher regularity properties, and the images $\tilde{g}_i$’s can be supposed to be smooth.

These different features represent the foundation of the signal separation method discussed in the previous section. In order to exploit this diversity it is necessary to choose suitable dictionaries $A_f$ and $A_g$, with respect to which $\tilde{f}$ and the $\tilde{g}_i$’s have sparse representations, respectively.

As far as $A_f$ is concerned, wavelets have been widely used to sparsely represent piecewise smooth images [50]. This is our choice for this paper, and the details of the types of wavelet used will be given in the following section, when the numerical results are presented. It is worth noting that in recent years a large number of new multi-dimensional transforms have been introduced in order to capture the directional features of two-dimensional images [46]. Among the most known, there are the curvelets [26], the ridgelets [25] and the shearlets [38]. The use of these transforms may give better results, but a deep investigation of the best choice for the dictionaries goes beyond the scope of this paper. Thus, we have decided to make the simplest choice of the wavelets, which is sufficient to properly illustrate the disjoint sparsity signal separation method.

The situation is simpler for the choice of $A_g$. Indeed, a good representation of smooth functions may be obtained by choosing low frequency trigonometric polynomials. This is a simple consequence of the correspondence between the smoothness of a function and the decay of its Fourier transform. This represents our choice for $A_g$ in this paper.

These dictionaries purely represent a general guideline for the choices of $A_f$ and $A_g$. Additional information on the particular physical model may be used to select dictionaries more adapted to the features of the images under consideration.
5. Applications to Hybrid Inverse Problems

5.1. Introduction. We have seen in the introduction that a hybrid problem usually involves two steps. First, internal functionals are measured inside the domain and, second, from their knowledge the unknown coefficients of the PDE have to be reconstructed. These internal data are linear or quadratic functionals of the unknowns and of the solutions of the direct problem. Let us mention some examples.

- In photoacoustic tomography \cite{11, 12, 21, 19} the internal data take the form
  \[ H(x) = \Gamma(x)\mu(x)u(x), \quad x \in \Omega, \]
  where \( \Gamma \) is the Grüneisen parameter, \( \mu \) is the optical absorption and \( u \) is the light intensity. The main unknown of the problem is \( \mu \). The photoacoustic image \( H \) gives only qualitative information on the medium, as \( \mu \) is multiplied by \( \Gamma \) and \( u \). The problem of quantitative photoacoustics is the reconstruction of \( \mu \) from \( H \). Under the diffusion approximation, the light intensity \( u \) solves
  \[ -\text{div}(D\nabla u) + \mu u = 0 \quad \text{in} \; \Omega, \]
  where \( D \) is the diffusion parameter. This equation should be augmented with suitable boundary conditions, such as of Dirichlet or Robin type.

- In thermoacoustic tomography \cite{14} the internal data take the form
  \[ H(x) = \sigma(x)|u(x)|^2, \quad x \in \Omega, \]
  where \( \sigma \) is the unknown conductivity and \( u \) is the electric field. The problem of quantitative thermoacoustics is the reconstruction of \( \sigma \) from \( H \). In the scalar approximation, \( u \) is the solution of the Helmholtz equation
  \[ \Delta u + (\omega^2 + i\omega\sigma)u = 0 \quad \text{in} \; \Omega, \]
  where \( \omega \) is the angular frequency. As before, this equation should be augmented with suitable boundary conditions.

- In microwave imaging by ultrasound deformation \cite{13, 11} the internal data take the form
  \[ H(x) = \varepsilon(x)u(x)^2, \quad x \in \Omega, \]
  where \( \varepsilon \) is the unknown electric permittivity and \( u \) is the electric field. The problem is now to reconstruct \( \varepsilon \) from \( H \). In the scalar approximation, \( u \) is the solution of the Helmholtz equation
  \[ \Delta u + \omega^2\varepsilon u = 0 \quad \text{in} \; \Omega, \]
  augmented with suitable boundary conditions.

- In ultrasound modulated diffuse optical tomography \cite{10, 9, 15} the internal data are \( \text{div}(u^2\nabla \mu) \), where \( u \) solves \eqref{18} and \( \mu \) is the optical absorption.

In all the above examples, the measurement \( H \) is the product of the desired unknown and other quantities. Thus, the problem is extracting the desired unknown from \( H \). PDE techniques are usually used to solve the problem, but, as discussed in the introduction, have several drawbacks.

All the above problems consist in the determination of two functions from the knowledge of their product. Taking logarithms, in a multi-measurement setting this is equivalent to finding \( f(x) \) and \( g_i(x) \) from the knowledge of their sums
  \[ h(x) = f(x) + g_i(x), \quad x \in \Omega. \]
The disjoint sparsity signal separation method can be applied since \( f \) and the \( g_i \)'s have different level of smoothness, and so can be sparsely represented with respect to different dictionaries (see § 4.2).

In particular, Theorem 3 guarantees unique and stable reconstruction of the coefficients, provided that we can construct many incoherent \( u_i \) of the above problems (by changing the boundary values), so that the corresponding \( g_i \)'s give a complete set of measurements, according to Definition 2. As we shall see below in the numerical simulations, this can be easily achieved in practice. However, a formal proof of this behavior is still missing, which represents a very interesting open problem regarding the interplay of the PDE theory with the disjoint sparsity approach. It is worth mentioning that similar assumptions of incoherence, usually in terms of linear independence of gradients of solutions, are often necessary for the PDE-based reconstruction methods (see, e.g., [27, 21, 13, 14, 1, 17, 2, 3, 42] and references therein).

As an example, in the rest of this section we shall apply the method to the reconstruction in quantitative photoacoustic tomography. All the other modalities mentioned above can be treated with minor modifications.

5.2. **Quantitative photoacoustic tomography, the case \( \Gamma = 1 \).** In photoacoustic tomography, the object under investigation is illuminated with light radiation, whose absorption causes local heating of the medium. The temperature rise creates an expansion of the tissue, thereby producing an acoustic wave, that can be measured on the boundary of the domain. The first step of this hybrid modality consists in the recovery of the amount of absorbed radiation by inverting the wave equation, from the knowledge of the boundary values. This problem has attracted much attention in the last years: the reader is referred to [40] and to the references therein for more details. In this paper, we shall suppose that the first step has been performed, and that we have access to the amount of absorbed radiation

\[
H(x) = \Gamma(x) \mu(x) u(x), \quad x \in \Omega,
\]

where \( \Gamma \) is the Grüneisen parameter, \( \mu \) is the optical absorption and \( u \) is the light intensity. The problem of quantitative photoacoustic tomography consists in the reconstruction of \( \mu \) from the knowledge of \( H \). Much research has been done on this over the last years, see e.g. [32, 45, 33, 11, 31, 48, 36, 44, 43] and references therein. Sparse representations were first used in [45].

Light propagation may be modeled by a radiative transport equation or, when the scattering coefficient is large and the absorption is small, by its diffusion approximation [18]. We consider here the simplest case when \( \Gamma = 1 \), the general case is discussed later in § 5.3. By using multiple measurements, \( \mu \) can be recovered both in the transport regime [18] and in the diffusive regime [21], under suitable assumptions on the solutions.

We now describe how to apply the disjoint sparsity approach to this problem. The separation of \( \mu \) and \( u \) does not require the use of a PDE, and so no specific model has to be assumed for the inverse problem. For simplicity, we shall construct the solutions \( u_i \) via [18] with \( D = 1 \) and Dirichlet boundary values, namely

\[
\begin{align*}
-\Delta u_i + \mu u_i &= 0 \quad \text{in } \Omega, \\
u_i &= \varphi_i \quad \text{on } \partial \Omega.
\end{align*}
\]

However, this equation will not be used for the inversion.
The joint sparsity method will be applied as follows. Let \( \tilde{\mu} \) denote the true absorption. After constructing \( N \) solutions \( \tilde{u}_1, \ldots, \tilde{u}_N \) to the above equation, the quantities

\[
H_i(x) = \tilde{\mu}(x)\tilde{u}_i(x), \quad x \in \Omega,
\]

are measured. Taking logarithms and adding white Gaussian noise \( n_i \), we obtain

\[
(20) \quad h_i = \log \tilde{\mu} + \log \tilde{u}_i + n_i, \quad i = 1, \ldots, N.
\]

The reconstruction of \( \tilde{\mu} \) from the knowledge of the \( h_i \)'s exactly corresponds to the problem discussed in Section 3. The method based on Theorem 3 and whose numerical implementation is described in Section 4 will be used for the reconstruction.

In all the examples, we shall consider the two-dimensional domain \([0, 1]^2\) with \(128 \times 128\) pixels. As far as the choice for the dictionaries is concerned, we let \( A_f \) consist of the orthonormal basis of Haar wavelets and let \( A_g \) consist of 961 low frequency trigonometric polynomials, periodic over \([0, 1]^2\). Note that \( A_f \) and \( A_g \) are left invertible, as required.

The choice for \( A_g \) forces to choose periodic boundary conditions, and so we set

\[
(21) \quad \begin{cases}
\varphi_1(x) = 1, \\
\varphi_2(x) = 1 - \sin(2\pi x_1)/4, \\
\varphi_3(x) = 1 - \sin(2\pi x_2)/4, \\
\varphi_4(x) = 1 - \cos(2\pi x_2)/4, \\
\varphi_5(x) = 1 - \cos(2\pi x_1)/4.
\end{cases}
\]

(For physical constraints, all the quantities have to be positive.) Non-periodic choices for the boundary conditions would be allowed with no added difficulties for the reconstruction, provided that the dictionary \( A_g \) is properly chosen. The above boundary values are expected to generate incoherent solutions to (19) in such a way to satisfy the conditions of complete sets as in Definition 2. In this case, the assumptions of Theorem 3 are verified and the joint sparsity separation method is guaranteed to provide the right reconstruction, even in presence of noise.

5.2.1. Example 1 - convex inclusions. We consider three convex constant inclusions in a homogeneous background, as shown in Figure 1a. We choose to stop the iteration procedure of OMP after 1500 iterations, which gives satisfactory results. More accurate stopping criteria may be considered [45], but this is not among the aims of this work.

In a first experiment we consider one noise-free measurement, namely we set \( N = 1 \) and \( n_1 = 0 \) in (20). The results are shown in Figure 1. The solution to (19) with boundary value \( \varphi_1 = 1 \) is shown in Figure 1b, and the corresponding measurement \( H_1 = \tilde{\mu}\tilde{u}_1 \) in Figure 1c. As it is evident from the images, the inclusions are still clearly recognizable in \( H_1 \), but the corresponding values of the absorption are corrupted by the multiplication by \( \tilde{u}_1 \). The sparse separation approach is thus applied as discussed above, and the reconstructed \( \mu \) is shown in Figure 1d. The relative reconstruction error is

\[
\frac{\|\log \mu - \log \tilde{\mu}\|_2}{\|\log \tilde{\mu}\|_2} \approx 1.5\%.
\]

This shows that, in absence of noise, the reconstruction is almost exact, even with only one measurement.
In a second experiment (Figure 2) we add white Gaussian noise \( n_i \) in (20). The noise level is so that 

\[
\frac{\|n_i\|_2}{\|\log(\tilde{\mu}\tilde{u}_i)\|_2} \approx 17.6\%.
\]

We tested the reconstruction procedure for \( N = 1, \ldots, 5 \) and the boundary values \( \varphi_i \) as in (21). The data \( H_i, i = 1, 3, 5 \), are shown in Figures 2a, 2c and 2e respectively. The reconstructed \( \mu' \)'s for \( N = 1, N = 3 \) and \( N = 5 \) are shown in Figures 2b, 2d and 2f respectively. The reconstruction errors for \( N = 1, \ldots, 5 \) are given in Table 1.

| \( N \) | 1 | 2 | 3 | 4 | 5 |
|--------|---|---|---|---|---|
| \( \|\log \mu - \log \hat{\mu}\|_2 / \|\log \hat{\mu}\|_2 \) | 74.4% | 32.2% | 28.4% | 15.8% | 7.5% |

It is evident that the larger \( N \) is, the more accurate the reconstruction becomes. More precisely, when \( N = 1 \) the reconstructed values of the absorption in the inclusions are completely wrong. This is due to the fact that the inclusions are roughly approximated by smooth atoms in \( A_g \) and then corrected with fewer atoms in \( A_f \), and so the sparsest approximation does not separate the two components as desired. However, the problem is solved when more measurements are added: this phenomenon is a simple verification of the general theory discussed in Section 3. In particular, the conditions in Definition 2 are easily satisfied when \( N \) becomes bigger.
The reconstruction with $N = 5$ is very satisfactory if measurement and reconstruction errors are compared. Indeed, the noise from the data has almost disappeared in the reconstruction, without a separate regularization. This is due to the implicit regularizing effect that sparse representations provide. An a priori total variation (TV) regularization of the measurements may give even better results; this has not been investigated in this work.

5.2.2. Example 2 - The Shepp–Logan phantom. Here, we let $\tilde{\mu}$ be the well-known Shepp-Logan phantom (shown in Figure 3a). We choose to stop the iterative procedure of OMP after 2000 iterations. As above, we consider the boundary conditions

Figure 2. Example 1. Noisy case with multiple measurements.
ϕ_i as in (21) and the corresponding solutions ũ_i to (19), for i = 1, ..., 5, and measure the quantities h_i as in (20).

In a first experiment we consider the case without noise (Figure 3). The reconstruction errors for N = 1, ..., 5 are shown in Table 2. We see that the reconstruction quality improves as N increases, as it is expected from the general theory discussed in Section 3. From a comparison with the previous case without noise, we notice that more measurements are needed to have a satisfactory reconstruction. This is due to the more complicated structure of the phantom, which has a less sparse representation in terms of Haar wavelets than the absorption considered in the previous case. Thus, a higher N is needed to satisfy the conditions in Definition 2.

Table 2. Example 2. Reconstruction errors for the noise-free case.

| N   | \( \frac{\| \log \mu - \log \tilde{\mu} \|_2}{\| \log \tilde{\mu} \|_2} \) |
|-----|------------------------------------------|
| 1   | 68.6%                                    |
| 2   | 24.8%                                    |
| 3   | 18.6%                                    |
| 4   | 11.4%                                    |
| 5   | 5.4%                                     |

In a second experiment we add white Gaussian noise n_i to the data in (20). The noise level is such that \( \frac{\| H_i - \tilde{\mu} \tilde{u}_i \|_2}{\| \tilde{\mu} \tilde{u}_i \|_2} \approx 17.8\% \).

Motivated by the noisy-free case, we perform the reconstruction with N = 5 measurements. The reconstruction error is \( \frac{\| \log \mu - \log \tilde{\mu} \|_2}{\| \log \tilde{\mu} \|_2} = 17.0\% \), that is comparable to the measurement error. The results are shown in Figure 4. It is expected that adding more measurements would improve the quality of the reconstruction.
5.3. **Quantitative photo-acoustic tomography in the diffusive regime with variable $\Gamma$.**

5.3.1. **Introduction.** We consider here the problem of quantitative photoacoustic tomography, as introduced above, without the further assumption $\Gamma = 1$. Thus, the unknown absorption $\mu$ has to be reconstructed from

$$H(x) = \Gamma(x)\mu(x)u(x), \quad x \in \Omega.$$  

We consider the diffusion approximation [18] of light propagation:

$$-\text{div}(D\nabla u) + \mu u = 0 \quad \text{in} \quad \Omega.$$  

For simplicity, here we shall augment this equation with Dirichlet boundary conditions, that are supposed to be measurable. Note that $D$, $\Gamma$ and $\mu$ are unknowns of the problem. In contrast with the case $\Gamma = 1$, where (22) was merely used to compute the data but not in the inverse problem, we shall make full use of this PDE. Let us briefly review the main known results on this inverse problem. Bal and Ren [19] showed that when $\Gamma$, $D$ and $\mu$ are all unknown, then there is no uniqueness for the inverse problem even with all the measurements $H$ for all solutions $u$ to (22). If any of these parameters is known, then the others may be reconstructed by using the PDE. The same authors have proved that all the coefficients may be uniquely reconstructed by using multi-frequency measurements, under certain assumptions on the dependency of the parameters on the frequency [20]. Scherzer and Naetar [42] studied the case of piecewise constant parameters: all unknowns can be uniquely determined, but the method may be very sensitive to noise.

We propose here for the single-frequency case a mixed approach combining the following aspects.

- As in [19], the PDE [22] can be used in the reconstruction. However, one degree of freedom for the parameters remain.
- As in [42], the PDE method gives unique reconstruction under the finite dimensionality assumption of the coefficient spaces.
- The disjoint sparsity signal separation method may be applied to this case as in § 5.2.

The combination of such approaches consists in substituting the piecewise constant assumption with the sparsity assumption, and then in the use of [22] to reconstruct all the parameters. More precisely, the reconstruction algorithm proposed here is substantially divided into the following three main steps.
(1) By using the disjoint sparsity signal separation method applied to
\[ h_i = \log H_i = \log(\Gamma \mu) + \log u_i, \quad i = 1, \ldots, N, \]
the solutions \( u_i \) are reconstructed.

(2) Following [19], by using the PDE
\[ -\text{div}(Du_i \nabla u_j) = 0 \quad \text{in} \quad \Omega, \]
with three suitable measurements, the diffusion \( D \) can be uniquely determined.

(3) Finally, the absorption can be directly reconstructed via
\[ \mu = \text{div}(D \nabla u_i) \quad \text{in} \quad \Omega, \]
and possibly averaging over \( i \).

5.3.2. The reconstruction algorithm. Even though theoretically satisfactory, the algorithm summarized above is not applicable in practice as it stands. Indeed, the reconstruction of \( D \) in (2) is not too sensitive to errors in \( u_j \), but that of \( \mu \) in (3) is. To understand this, we compare the solutions \( u_i \) to (22) with \( D = 1 \) and \( \mu \) as in Figure 1a and the solutions \( u_0^i \) to (22) with \( D = 1 \) and \( \mu = \mu^0 = 1 \), with boundary conditions given by (21):
\[
\begin{aligned}
\begin{cases}
-\text{div}(Du_i \nabla u_i) + \mu u_i = 0 & \text{in} \quad \Omega, \\
-\text{div}(Du_0^i \nabla u_0^i) + \mu^0 u_0^i = 0 & \text{in} \quad \Omega, \\
u_i = u_0^i = \varphi_i & \text{on} \quad \partial \Omega.
\end{cases}
\end{aligned}
\]

The solutions are shown in Figure 5. Looking at the first two columns, it is clear that the variations between \( u_i \) and \( u_0^i \) are minimal. This is due to the fact that the two problems have the same diffusion coefficients and small variations in the absorption coefficients. As we saw in §5.2, the reconstruction at step (1) cannot be at this level of precision, and therefore \( \mu \) cannot be reconstructed in this simple way. In order to overcome this difficulty, we make the following observation.

Remark 5. The ratios \( u_i/u_0^i \) are almost independent of \( i \), provided that \( \mu \) is a small variation around a known background \( \mu^0 \). This is evident from the third column of Figure 5 and can be proved by arguing as follows. A direct calculation gives that \( v_i = u_i/u_0^i \) satisfies
\[
\begin{aligned}
\begin{cases}
-\text{div}(D \nabla v_i) + (\mu - \mu^0)v_i = 2 \frac{\nabla u_0^i}{u_0^i} \cdot \left( \frac{\nabla u_i}{u_i} - \frac{\nabla u_0^i}{u_0^i} \right) & \text{in} \quad \Omega, \\
v_i = 1 & \text{on} \quad \partial \Omega.
\end{cases}
\end{aligned}
\]
When \( \mu \) is close to \( \mu^0 \), the right-hand side of this equation becomes negligible with respect to the other terms. Thus, \( v_i \) is substantially independent of \( i \).

Let us now describe the precise reconstruction algorithm based on these observations. It consists of two initial steps and an iterative procedure consisting of three more substeps. For simplicity, we shall discuss only the noise-free case. We suppose that \( \mu \) is a small perturbation around a known coefficient \( \mu^0 \) and that \( D \) is known at one point of the domain.

(1) By using the disjoint sparsity signal separation method applied to
\[ h_i = \log(\Gamma \mu) + \log u_i, \quad i = 1, \ldots, N, \]
Figure 5. The solutions to \([23]\) and their ratios.

a first approximation \(u_i(0)\) of the solutions \(u_i\) is reconstructed. As discussed in Section 4, this can be done by minimizing

\[
\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 + \lambda \sum_{i=1}^{N} \|A_f y_f + A_g y_g - h_i\|_2
\]

with OMP, and then writing \(u_i(0) = \exp(A_g y_g)\).

(2) By using the computed \(u_i(0)\) and the PDE

\[
- \text{div}(D u_i \nabla \frac{u_j}{u_i}) = 0 \quad \text{in} \ \Omega
\]

with three suitable measurements, a first approximation \(D(0)\) of the diffusion can be obtained. Indeed, choose three boundary values \(\varphi_1, \varphi_2\) and \(\varphi_3\) such that

\[
\det(\nabla \frac{u_2}{u_1}, \nabla \frac{u_3}{u_1}) > 0, \quad \text{in} \ \Omega.
\]
(This can be easily done in two dimensions [19].) Then the above PDE may be rewritten as
\[ t(\nabla \log D) = - \left[ \text{div}(u_1 \nabla \frac{u_2}{u_1}) / u_2 \right. - \left. \text{div}(u_1 \nabla \frac{u_3}{u_1}) / u_3 \right] \left[ \nabla \frac{u_2}{u_1} \cdot \nabla \frac{u_3}{u_1} \right]^{-1} \text{ in } \Omega, \]
which can be integrated over \( \Omega \) and gives a unique solution for the diffusion coefficient, since \( D \) is known at one point of the domain. Since the solutions \( u_i \) are very sensitive to changes in \( D \), we expect this reconstruction to be satisfactory. From the numerical point of view, an optimal control approach may be applied to [24] to find \( D \).

(3) We now start the main iterative procedure. Initialize \( \mu(0) = \mu^0 \) and let \( u_i(0) \) and \( D(0) \) be as reconstructed in points (1) and (2). From \( \mu(k) \), \( u_i(k) \) and \( D(k) \), the following iteration is computed as follows.
(a) Given \( D(k) \) and \( \mu(k) \), let \( u_i^0(k) \) be the unique solution to
\[
\begin{cases}
-\text{div}(D(k) \nabla u_i^0(k)) + \mu(k) u_i^0(k) = 0 & \text{in } \Omega, \\
u_i^0(k) = \varphi_i & \text{on } \partial \Omega.
\end{cases}
\]
Since \( \varphi_i \) is known, \( u_i^0(k) \) is a known datum. Therefore we can measure
\[
h_i^0 = \log(H_i/u_i^0(k)) = \log(\Gamma u) + \log \left| \frac{u_i}{u_i^0(k)} \right|, \quad i = 1, \ldots, N.
\]
In view of Remark 5, the quantities \( u_i/u_i^0(k) \) are almost independent of \( i \). This leads to the minimization of
\[
\min_{y \in \mathbb{R}^{m_f + (N+1)m_g}} \| y \|_0 + \lambda_1 \sum_{i=1}^N \left\| A_f y_f + A_g y_g^f - h_i \right\|_2^2 + \lambda_2 \sum_{i=1}^N \left\| A_f y_f + A_g y_g^{N+1} - h_i^0 \right\|_2^2
\]
with \( \lambda_1 \ll \lambda_2 \). The second term maintains the incoherence among the \( y_g^f \)'s, on which this disjoint sparsity approach is based. The third term forces the quantities \( u_i/u_i^0(k) \) to be independent of \( i \), and numerical evidence shows that this gives a much better reconstruction than the one performed at point (1). The multipliers \( \lambda_1 \) and \( \lambda_2 \) may be taken dependent of \( k \). Set \( u_i(k+1) = u_i^0(k) \exp(A_g y_g^{N+1}) \).
(b) Given \( u_i(k+1) \), find a better approximation \( D(k+1) \) of the diffusion coefficient by proceeding as in (2).
(c) Reconstruct \( \mu(k+1) \) via
\[
\mu(k+1) = \frac{1}{N} \sum_{i=1}^N \frac{\text{div}(D(k+1) \nabla u_i(k+1))}{u_i(k+1)} \text{ in } \Omega.
\]
From the numerical point of view, it may be useful to regularize \( u_i(k+1) \) and \( D(k+1) \) before taking the derivatives. Finally, a TV-regularization of \( \mu(k+1) \) may reduce the accumulated noise.

There is no obvious stopping criterion for this iterative procedure. However, in the numerical simulations less than five iterations were sufficient.

In the above algorithm we have assumed for simplicity that the boundary values \( \varphi_i \) are measurable. However, this is probably not a necessary conditions, since they may be obtained from point (1) as \( \varphi_i = u_i(0)|_{\partial \Omega} \).
5.3.3. Numerical simulations. We have tested the algorithm discussed above with the absorption map $\tilde{\mu}$ considered in § 5.2.1 (see Figure 6c). The same dictionaries considered in § 5.2 are chosen. The light intensities $\tilde{u}_i$ are the solutions of

$$
\begin{cases}
-\text{div}(\tilde{D}\nabla \tilde{u}_i) + \tilde{\mu}\tilde{u}_i = 0 & \text{in } \Omega, \\
\tilde{u}_i = \varphi_i & \text{on } \partial\Omega,
\end{cases}
$$

where the diffusion coefficient $\tilde{D}$ is shown in Figure 6a and five boundary values are chosen as follows:

$$
\varphi_1(x) = 1, \quad \varphi_2(x) = 1 - \sin(2\pi x_1)/8, \quad \varphi_3(x) = 1 - \sin(2\pi x_2)/8, \\
\varphi_4(x) = x_1/4 + 7/8, \quad \varphi_5(x) = x_2/4 + 7/8.
$$

The internal data take the form

$$
H_i = \tilde{\Gamma}\tilde{D}\tilde{u}_i, \quad i = 1, \ldots, 5,
$$

where the Grüneisen parameter is shown in Figure 6b. The measurements corresponding to the first three boundary conditions will be used for the disjoint sparsity signal separation method (with $N = 3$), namely for the steps (1) and (3a) of the above algorithm; those corresponding to $\varphi_1$, $\varphi_4$ and $\varphi_5$ will be used in the steps (2) and (3b), in order to satisfy (25). All measurements are used in the last step (3c).

The OMP iterative procedures are stopped after 2000 iterations. If the absorption $\mu$ were recovered via (26) immediately after steps (1) and (2), the reconstruction would not be satisfactory, as it can be seen in Figure 6e. This makes steps (3a) and (3b) necessary: after repeating step (3) twice, the quality of the reconstruction is sensibly improved (see Figure 6f). The corresponding reconstruction of $D$ is shown in Figure 6d. As anticipated before, the reconstruction of $D$ from $u_i$ is much more stable than that of $\mu$. Note that the absorption was supposed to be known...
near the boundary of the domain, to avoid problems with the second derivatives in \(\partial_{\varepsilon}u\). The case with noise was not studied in this paper, since the robustness to noise of the disjoint sparsity signal separation algorithm has already been tested in §5.2. The robustness to noise of the other steps of the reconstruction method discussed above is well-known, and standard regularization method can be employed. It is worth noticing that since the absorption \(\mu\) is found in step (3c) from the reconstructed values of the light intensities \(u_i\), the signal to noise ratio of \(u_i\) has to be sufficiently large. Unfortunately, as it can be seen from the third column of Figure 5, the amplitude of the information about \(\mu\) captured in \(u_i\) is very small, and so the noise has to be comparable to this.

6. Conclusions

In this work we have studied a method for signal separation based on the disjoint sparsity of multiple measurements. A theorem giving unique and stable reconstruction was proved. The result is based on the incoherence of the measurements. Then, the method was applied to hybrid imaging problems, and in particular to quantitative photoacoustic tomography. This technique has been successfully tested on several numerical simulations, and results to be very robust to noise.

It would be interesting to see whether the assumption requiring \(A_f\) and \(A_g\) to be left-invertible can be removed. This would allow to consider over-redundant dictionaries, such as the curvelets or the shearlets, which are known to provide sparser representations of images if compared to the wavelets.

The incoherence between the measurements \(g_i\) is the main foundation of the method, and the numerical simulations have shown that such property holds true with different solutions to the same PDE, which is the relevant case for hybrid imaging. It would be very interesting to prove this result in general. Randomly chosen boundary conditions may give the necessary incoherence.

Finally, it would be very interesting to investigate whether the main ideas behind this method can be applied to other inverse problems with multiple measurements consisting of two components, of which only one remains fixed.

References

[1] G. S. Alberti. On multiple frequency power density measurements. *Inverse Problems*, 29(11):115007, 25, 2013.
[2] G. S. Alberti. Enforcing local non-zero constraints in PDEs and applications to hybrid imaging problems. *ArXiv e-prints: 1406.3248*, June 2014.
[3] G. S. Alberti. On multiple frequency power density measurements II. The full Maxwell’s equations. *Journal of Differential Equations*, 258(8):2767 – 2793, 2015.
[4] G. S. Alberti, H. Ammari, and K. Ruan. Multi-frequency acousto-electromagnetic tomography. *Contemp. Math. (to appear).*
[5] G. S. Alberti and Y. Capdebroos. A propos de certains problèmes inverses hybrides. In *Séminaire: Equations aux Dérivées Partielles. 2013–2014*, Sémin. Équ. Dériv. Partielles, page Exp. No. II. École Polytech., Palaiseau.
[6] G. S. Alberti and Y. Capdebroos. *Lectures on elliptic methods for hybrid inverse problems*. In preparation.
[7] H. Ammari. *An Introduction to Mathematics of Emerging Biomedical Imaging*, volume 62 of *Math. Appl.* Springer-Verlag, Berlin, 2008.
[8] H. Ammari, E. Bonnetier, Y. Capdebroosq, M. Tanter, and M. Fink. Electrical impedance tomography by elastic deformation. *SIAM J. Appl. Math.*, 68(6):1557–1573, 2008.
[9] H. Ammari, E. Bossy, J. Garnier, L. H. Nguyen, and L. Seppecher. Reconstruction of a piecewise smooth absorption coefficient by an acousto-optic process. *Comm. Part. Diff. Equat.*, 38(10):1737–1762, 2013.

[10] H. Ammari, E. Bossy, J. Garnier, L. H. Nguyen, and L. Seppecher. A reconstruction algorithm for ultrasound-modulated diffuse optical tomography. *Proc. Amer. Math. Soc.*, 142(9):3221–3236, 2014.

[11] H. Ammari, E. Bossy, V. Jugnon, and H. Kang. Mathematical modeling in photoacoustic imaging of small absorbers. *SIAM Rev.*, 52(4):677–695, 2010.

[12] H. Ammari, E. Bossy, V. Jugnon, and H. Kang. Reconstruction of the optical absorption coefficient of a small absorber from the absorbed energy density. *SIAM J. Appl. Math.*, 71(3):676–693, 2011.

[13] H. Ammari, Y. Capdeboscq, F. de Gournay, A. Rozanova-Pierrat, and F. Triki. Microwave imaging by elastic deformation. *SIAM J. Appl. Math.*, 71(6):2112–2130, 2011.

[14] H. Ammari, J. Garnier, W. Jing, and L. H. Nguyen. Quantitative thermo-acoustic imaging: An exact reconstruction formula. *J. Differential Equations*, 254(3):1375–1395, 2013.

[15] H. Ammari, L. H. Nguyen, and L. Seppecher. Reconstruction and stability in acousto-optic imaging for absorption maps with bounded variation. *J. Funct. Anal.*, 267(11):4361–4398, 2014.

[16] G. Bal. Hybrid inverse problems and internal functionals. In G. Uhlmann, editor, *Inverse problems and applications: inside out. II*, volume 60 of *Math. Sci. Res. Inst. Publ.*, pages 325–368. Cambridge Univ. Press, Cambridge, 2013.

[17] G. Bal, E. Bonnetier, F. Monard, and F. Triki. Inverse diffusion from knowledge of power densities. *Inverse Probl. Imaging*, 7(2):353–375, 2013.

[18] G. Bal, A. Jollivet, and V. Jugnon. Inverse transport theory of photoacoustics. *Inverse Problems*, 26(2):025011, 35, 2010.

[19] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 20, 2011.

[20] G. Bal and K. Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 13, 2012.

[21] G. Bal and G. Uhlmann. Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8):085010, 20, 2010.

[22] J. Bobin, Y. Moudden, and J.-L. Starck. Enhanced source separation by morphological component analysis. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V, May 2006.

[23] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353 – 2362, 2001.

[24] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):43–101, 2009.

[25] E. J. Candès and D. L. Donoho. Ridgelets: a key to higher-dimensional intermittency? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 357(1760):2495–2509, 1999.

[26] E. J. Candés and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise $C^2$ singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[27] Y. Capdeboscq, J. Fehrenbach, F. de Gournay, and O. Kavian. Imaging by modification: numerical reconstruction of local conductivities from corresponding power density measurements. *SIAM J. Imaging Sci.*, 2(4):1003–1030, 2009.

[28] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *Signal Processing, IEEE Transactions on*, 54(12):4634–4643, Dec 2006.

[29] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001. Reprinted from SIAM J. Sci. Comput. 20 (1998), no. 1, 33–61 (electronic) [ MR1639094 (99h:94013)].

[30] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *Signal Processing, IEEE Transactions on*, 53(7):2477–2488, July 2005.

[31] B. Cox, T. Tarvainen, and S. Arridge. Multiple illumination quantitative photoacoustic tomography using transport and diffusion models. In *Tomography and inverse transport theory*, volume 559 of *Contemp. Math.*, pages 1–12. Amer. Math. Soc., Providence, RI, 2011.
[32] B. T. Cox, S. R. Arridge, K. P. K"ostli, and P. C. Beard. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Appl. Opt.*, 45(8):1866–1875, Mar 2006.

[33] B. T. Cox, J. G. Laufer, and P. C. Beard. The challenges for quantitative photoacoustic imaging, 2009.

[34] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, 2002.

[35] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Appl. Comput. Harmon. Anal.*, 19(3):340–358, 2005.

[36] H. Gao, S. Osher, and H. Zhao. Quantitative photoacoustic tomography. In *Mathematical modeling in biomedical imaging. II*, volume 2035 of *Lecture Notes in Math.*, pages 131–158. Springer, Heidelberg, 2012.

[37] P. Georgiev, F. Theis, and A. Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *Neural Networks, IEEE Transactions on*, 16(4):992–996, July 2005.

[38] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, 39(1):298–318, 2007.

[39] P. Kuchment. Mathematics of hybrid imaging: A brief review. In I. Sabadini and D. C. Struppa, editors, *The Mathematical Legacy of Leon Ehrenpreis*, volume 16 of *Springer Proceedings in Mathematics*, pages 183–208. Springer Milan, 2012.

[40] P. Kuchment and L. Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 817–865. Springer New York, 2011.

[41] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, Dec 1993.

[42] W. Naetar and O. Scherzer. Quantitative photoacoustic tomography with piecewise constant material parameters. *SIAM J. Imaging Sci.*, 7(3):1755–1774, 2014.

[43] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. A Bayesian approach to spectral quantitative photoacoustic tomography. *Inverse Problems*, 30(6):065012, 18, 2014.

[44] K. Ren, H. Gao, and H. Zhao. A hybrid reconstruction method for quantitative PAT. *SIAM J. Imaging Sci.*, 6(1):32–55, 2013.

[45] A. Rosenthal, D. Razansky, and V. Ntziachristos. Quantitative optoacoustic signal extraction using sparse signal representation. *Medical Imaging, IEEE Transactions on*, 28(12):1997–2006, Dec 2009.

[46] R. Rubinstein, A. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, June 2010.

[47] J. K. Seo and E. J. Woo. *Nonlinear Inverse Problems in Imaging*. Wiley, 2013.

[48] J.-L. Starck, M. Elad, and D. Donoho. Redundant multiscale transforms and their application for morphological component separation. In *Advances in Imaging and Electron Physics*, volume 132 of *Advances in Imaging and Electron Physics*, pages 287 – 348. Elsevier, 2004.

[49] J.-L. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *Image Processing, IEEE Transactions on*, 14(10):1570–1582, Oct 2005.

[50] J. Tropp. Greed is good: algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50(10):2231–2242, Oct 2004.