Aggregation sheaves for greedy modal decompositions

Michael Robinson
SRC, Inc., and Department of Mathematics and Statistics, American University, Washington, DC 20016, United States of America
E-mail: mrobinson@srcinc.com

Keywords: sheaf, greedy algorithm, modal decomposition

Abstract
This article develops a new theoretical basis for decomposing signals that are formed by the linear superposition of a finite number of modes. Each mode depends linearly on the weights within the superposition and nonlinearly upon several other parameters. The particular focus of this article is upon finding both the weights and the parameters when the number of modes is not known in advance. This article introduces a novel mathematical formalism, aggregation sheaves, and shows how they characterize the behavior of greedy algorithms that attempt to solve modal decomposition problems. It is shown that minimizing the local consistency radius within the aggregation sheaf is guaranteed to solve all modal decomposition problems. Since the modes may or may not be well-separated, a greedy algorithm that identifies the most distinct modes first may not work reliably.

1. Introduction

To explore the fundamental limits on modal imaging, we need an umbrella framework that encompasses all existing ones yet generalizes beyond them to the maximum extent possible. It has been recently argued that sheaves are the appropriate mathematical foundation upon which to build a theory of topological filtering [1]. Because sheaves and topological filters are foundational, they subsume all modal decomposition problems and their algorithmic instantiations as special cases, including those not yet discovered.

Recently, it was demonstrated that improved signal-to-noise performance can be obtained using sheaves when compared to existing approaches if theoretical and implementation hurdles can be surmounted. This article attempts to reconcile these empirical findings with an appropriately general theory, by attempting to lay the groundwork for solving all modal decomposition problems. One cannot overemphasize this fact: regardless of what physical processes or techniques are used for gathering measurements about sources—be they classical, quantum, or otherwise—they are subject to the same mathematical formalism as described in this article.

The plan for the article is as follows. In section 2, we establish a formal specification of a general modal decomposition problem. We also specify two classical modal decomposition problems that we return to in section 5. In section 3, we establish the proper metrical representation for the free parameters in a general modal decomposition problem. To this end, we present a novel definition for a conical space to represent the parameters of a mode (Definition 1), and prove basic properties about such a space and its related constructions. The modal decomposition problem itself is recast into several novel sheaf models in section 4. Which sheaf model is used depends on the desired algorithm, though they all have associated performance guarantees that are proven in that section. In section 5, we demonstrate the usage of the general sheaves on the two specific problems from section 2. Section 6 briefly concludes the article.

1.1. Historical context

Modal decompositions are old problems, so there are many frameworks in use today, such as Fourier bases, wavelets, and their generalizations to frames. The usual solution to such problems involve generalized Fourier decompositions, which rely on the algebraic structure of the modes. For instance, wavelets [2, 3] rely on frames [4, 5], which are overcomplete orthogonal dictionaries of modes.
Compressive sensing methods have become popular for solving decomposition problems (for instance, [6, 7]). They generally rely on the sparsity of the modes, which is to say that a given signal is composed of a few modes only. As a result, compressive sensing generally exploits both algebraic and geometric structure to reduce the necessary search space for a modal decomposition.

Many of these decomposition techniques rely upon a greedy or partially-greedy algorithm. Perhaps the most famous greedy algorithm for source decompositions is the CLEAN algorithm and its generalizations (see [8–10] for instance, among many others). CLEAN relies on the geometric structure of modes, such as a single strongest peak within each mode. Although greedy algorithms can perform extremely well in certain settings, they can also fail spectacularly when their assumptions are violated. This article provides a theoretical basis for ensuring the correct performance of these greedy algorithms.

Finally, recent quantum techniques have raised awareness of the possibility for substantially better performance. For instance, Tsang [11] showed that the quantum modal imaging could separate sources that were separated well below the Rayleigh limit. This idea has already led to improvements in molecular microscopy [12], in which the Wasserstein metric was used to provide statistical robustness.

2. Detailed problem statement

Assume that one obtains a number of samples from a signal

\[ s(x; \mathbf{a}_1, b_1, \mathbf{a}_2, b_2, \ldots, \mathbf{a}_N, b_N) = \sum_{n=1}^{N} a_n \phi(x; b_n), \]

where usually the number of sources \( N \) is fixed but unknown. We will call \( a_n \) and \( b_n \) the parameters for source \( n \). The parameter \( a_n \) will be called the magnitude of the source \( n \), and the parameter \( b_n \) will be called its location. Our goal is to find \( \{ a_n \} \) and \( \{ b_n \} \) from samples of \( s \) at a finite set of values for \( x \), namely \( x_1, x_2, \ldots, x_M \). We will call each \( x_m \) a measurement and \( s(x_m) \) an observation. Equation (1) is meant to be taken fairly generally. We do not require that \( \phi \) be nonnegative, complex valued, or real valued. (For simplicity of dimension counting only, we will assume that \( \phi \) is real valued in sections 3 and 4. The reader may generalize this mutatis mutandis if desired.) That said, we will assume that \( \phi(x; b_n) \) is smooth in the second variable \( b_n \), but we need not assume that it is linear in either variable. If \( \phi(x; b_n) \) is not continuous in the first variable \( x \), this may limit certain strategies for adaptively determining the next measurement. However, even if \( \phi(x; b_n) \) is not continuous in the first variable \( x \), all of the theory and some of the methodology will still work without changes.

What follows are two separate source parameter recovery problems that are special cases of our assumed signal model. When we solve the problem of recovering source parameters in the general case in section 3, our solution automatically applies to each of the special cases described in this section. Although there are algorithmic considerations, the theory provides a surprisingly complete set of solvability conditions (Proposition 1) and prescribes a rather definite algorithmic recipe. Quite specifically, the algorithms that solve each of the problems in this section amount to special cases of solvers for the general optimization problems specified in the statements of corollary 3 and proposition 5. These optimization problems are consistency radius minimization problems whose use has been recently effective in solving difficult signal processing problems [13, 14].

2.1. Fourier decomposition

In the most basic case, equation (1) represents the Fourier series decomposition of a periodic function. For this representation, we merely need to take \( b_n = n \in \mathbb{Z} \) and define

\[ \phi(x; n) = e^{2\pi inx}. \]

The signal model is then simply

\[ s(x) = \sum_{n=-N}^{N} a_n e^{2\pi inx}. \]

Since we are only concerned with finitely many sources, we will not consider what happens if \( N \to \infty \).

2.2. Spectral Estimation

We can generalize the Fourier series decomposition to handle non-integer frequencies by \( b_n = \omega_n \in \mathbb{R} \),

\[ \phi(x; \omega_n) = e^{2\pi i\omega_n x}, \]
resulting in a new signal model

\[ s(x) = \sum_{n=1}^{N} a_n e^{2\pi i n \omega_n x}, \]

so that the complex magnitudes \( \{a_n\} \) and the real frequencies \( \{\omega_n\} \) are parameters.

Of course this signal model really isn’t much different from the Fourier decomposition, but it differs in one important aspect: convergence. With finitely many sources (ie. \( N < \infty \)), this decomposition is a finite sum even though its Fourier decomposition may require infinitely many terms Worse, the Fourier series usually only converges in the \( L^2 \) sense (not pointwise) and may do so slowly.

3. Source parameter analysis

Each source is determined by its magnitude and location. These properties have different units, which suggests that the space of source parameters, jointly, is not Euclidean space. As will be explained in section 3.1, the correct space for sources is a quotient of a product of conical spaces. Interpreting a set of source parameters as a point within such a quotient space, we can connect measurements of the sources (however they are obtained) to the coordinates of this point. Differential topology provides a clear, concrete answer as to how many measurements must be obtained to succeed at this task, which we outline in section 3.2.

3.1. Conical spaces

The problem at hand is determining the parameters of the sources, namely the values of \( a_n \) and \( b_n \) in equation (1) from the values of \( s(x_1), s(x_2), \ldots, s(x_M) \).

To fix ideas, let us first consider each source individually. It happens that this is already an interesting—and apparently underappreciated—problem. Assume that \( a_n \in \mathbb{R}^A \) and \( b_n \in \mathbb{R}^B \) for each \( n \), for some fixed \( A, B \).

Recall that \( a_n \) is the magnitude of the source and \( b_n \) is its location.

It is convenient to encapsulate the source parameters in a single manifold \( I \), whose points consist of pairs \( w = (a, b) \in \mathbb{R}^A \times \mathbb{R}^B \cong \mathbb{R}^{A+B} \) under the identification that \( (0, b) = (0, b') \) for any \( b \) and \( b' \). The identification captures the intuition that if a source has no magnitude, its location does not matter. We therefore call any element \((0, b)\) to be a zero element of \( I \), and call the equivalence class of all zero elements the vertex of \( I \).

The manifold \( I \) should have a metric that reflects the zero elements. We can enforce this requirement by defining the distance

\[
d_i((a_1, b_1), (a_2, b_2)) := \min \{ \|a_1\| + \|a_2\|, \sqrt{\|a_1 - a_2\|^2 + \beta \|b_1 - b_2\|^2} \},
\]

where \( \beta > 0 \) is a constant and the usual norms on \( \mathbb{R}^A \) and \( \mathbb{R}^B \) are assumed for magnitudes and locations, respectively. This metric is called a bottleneck metric, and is widely used in applications of topology. Notice in particular that this reduces to the case of \( d_i((0, b), (0, b')) = 0 \) and \( d_i((a, b), (a', b)) = \|a - a'\| \).

**Lemma 1.** The bottleneck metric is a pseudometric on \( \mathbb{R}^{A+B} \), and is a metric on \( I \).

**Proof.** We simply need to establish the usual properties of a pseudometric:

**Reflexivity:** \( d_i((a, b), (a, b)) = \min \{ \|a\| + \|a\|, \sqrt{\|a - a\|^2 + \beta \|b - b\|^2} \} = \min \{2\|a\|, 0\} \) = 0.

**Symmetry:** \( d_i((a_1, b_1), (a_2, b_2)) = d_i((a_2, b_2), (a_1, b_1)) \) because the defining equation (2) is clearly symmetric.

**Triangle inequality:** The usual trick of adding and subtracting, along with the usual triangle inequality, establishes that

\[
\sqrt{\|a_1 - a_3\|^2 + \beta \|b_1 - b_3\|^2} = \sqrt{\|a_1 - a_3 + a_2 - a_2\|^2 + \beta \|b_1 - b_3 + b_2 - b_2\|^2} \\
\leq \sqrt{\|a_1 - a_2\|^2 + \beta \|b_1 - b_2\|^2} + \sqrt{\|a_2 - a_3\|^2 + \beta \|b_2 - b_3\|^2}
\]

and it is clear that \( \|a_1\| + \|a_3\| \leq \|a_1\| + 2\|a_2\| + \|a_3\| \). Combining these facts,
Finally, if \( d_I((a_1, b_1), (a_2, b_2)) = 0 \), this means that either \( \|a_1\| + \|a_2\| = 0 \) or \( \|a_1 - a_2\| + \beta\|b_1 - b_2\| = 0 \). In the latter case, it is clear that \( a_1 = a_2 \) and \( b_1 = b_2 \), since \( \beta > 0 \) and the usual norms induce metrics on Euclidean space. In the former case, we conclude that \( a_1 = a_2 = 0 \), which means that both points correspond to the vertex of \( I \). These two cases therefore establish that \( d_I \) is a metric on \( I \).}

**Definition 1.** A conical space is a metric space \( I \) that can be written as the quotient of a product \( \mathbb{R}^A \times \mathbb{R}^B \) with a metric \( d_I \) of the form given in equation (2), where the quotient identifies sets of points that are distance zero apart from each other.

As a shorthand, we will use the notation

\[
\| (a, b) \|_I := d_I((0, c), (a, b)) = \min \{ \|a\|, \sqrt{\|a\|^2 + \beta\|c - b\|^2} \} = \|a\|,
\]

which works a bit like a norm on \( I \) because it recovers the norm on \( \mathbb{R}^A \). We will call \( \| (a, b) \|_I \) the magnitude norm of \((a,b)\) in \( I \).

**Corollary 1.** Suppose that \( I \) is a conical space. Then the following properties follow about its magnitude norm:

1. if \( s \) is a real number, \( \| (sa, sb) \|_I = |s|\langle(a, b)\rangle_I \),
2. \( \| (a, b) \|_I = 0 \) if and only if \( a = 0 \), and
3. \( \| (a_1 + a_2, b) \|_I \leq \| (a_1, b) \|_I + \| (a_2, b) \|_I \) for all \( a_1, a_2, \) and \( b \).

**Corollary 2.** A normed space \( V \) is a conical space of the form \( V = V \times \{0\} \), namely the product of \( V \) with the trivial vector space.

### 3.2. Free parameter analysis

Let us start with the situation where the number of sources \( N \) is known, before we turn to the more general setting in section 4.

Let us assume that measurements are given by a function \( s \) taking values \( s(x_m) \in \mathbb{R} \). (These assumptions are not terribly critical in what follows, but it makes explaining the dimension counting easier. The reader can easily adjust the dimensions if desired.)

It is most useful to bundle all of the measurements into a vector \( (s(x_1), s(x_2), \ldots, s(x_M)) \in \mathbb{R}^M \). Under the usual topology on \( \mathbb{R}^M \), this means that

\[
s(x) = \sum_{n=1}^{N} a_n \phi(x_n, b_n)
\]

induces a smooth function \( S: I^N \to \mathbb{R}^M \) from the N-fold product of the conical space \( I \), given by

\[
S(a_1, b_1, a_2, b_2, \ldots, a_N, b_N) := (s(x_1), s(x_2), \ldots, s(x_M)). \tag{3}
\]

Notice that equation (3) is invariant to permutations of the sources. That is,

\[
S(a_1, b_1, a_2, b_2, \ldots, a_j, b_j, \ldots, a_N, b_N) = S(a_1, b_1, a_j, b_j, a_2, b_2, \ldots, a_N, b_N)
\]

for all pairs of integers \( i, j = 1, \ldots, N \). Put somewhat more abstractly, if \( S_N \) is N-fold symmetric group, the group of permutations of \( N \) items, then \( S \) factors through \( I^N / S_N \). This statement is equivalent to the commutative diagram.
In the rest of this article, we will abuse notation by allowing $S$ to refer either to $S : I^N \to \mathbb{R}^M$ or $S' : I^N/S_N \to \mathbb{R}^M$ in the diagram above. Since both $I^N$ and $I^N/S_N$ are manifolds of the same dimension, and both maps are defined by the same formula (equation (3)), this should cause little confusion.

The problem of determining the parameters of the sources can be solved in principle if and only if $S$ is injective. Clearly a necessary condition for this to occur is that $M \geq N(A + B)$.

Under fairly general conditions, $S$ is usually injective if $M$ is large enough.

Proposition 1. [the Whitney embedding theorem [15]; see also [16]]

Suppose that $\epsilon > 0$ is given. If $M > 2N(A + B)$, then there is an injective smooth function $I^N \to \mathbb{R}^M$ from source parameters to measurements that is within a distance of $\epsilon$ to $S$ in the $C^\infty$ topology of functions on $I^N$.

Intuitively, if $M > 2N(A + B)$ then $S$ can fail to be injective only if there is a symmetry present in the source parameters and the samples. This means that in principle, the sources can be completely inferred from the set of samples.

Although the proof of the Whitney embedding theorem is not constructive, it is typically proven using a perturbation-based argument. These perturbations may be rather undesirable within the context of the signal model, for instance they might suggest adding a grating lobe to an antenna pattern. More sensible perturbations can be obtained using a weaker form of the Whitney embedding theorem, called the signal embedding theorem [16], which allows the support of $\hat{\phi}(\cdot ; b, \phi)$ to be compact. Perturbations need not occur outside these supports.

The signal embedding theorem has been applied to invert signal models in the same form as $S$ for source localization [16] and antenna measurement [17]. The perturbation arguments suggest that an adaptive algorithm for inverting $S$ along its image might exist. Nevertheless, the basic problem is that we need an algorithm for this inversion. A canonical algorithm is provided by the sheaf model developed in the next section.

4. Sheaf models of modal decompositions

Since sheaf theory is not widely used in signal processing, we begin with a brief account of their mechanics in section 4.1. In section 4.2, we show how a non-adaptive approach for source decomposition has a straightforward incarnation as a sheaf $\mathcal{F}_P$. In contrast, adaptive approaches require a novel mathematical tool called an aggregation sheaf. Basic mathematical properties of aggregation sheaves are discussed in section 4.3. We apply aggregation sheaves to support general source decompositions in section 4.4, using a larger sheaf $\mathcal{K}_P$.

Finally, we establish performance guarantees for determining the correct number of sources in section 4.5.

4.1. A brief account of sheaves

A sheaf is a mathematical model of how local data are ascribed to portions of a space, and how these data may be compared to one another for consistency. The formalization of 'local' is generally understood to involve the notion of a topological space. For this paper, we only need to consider certain simple kinds of topological spaces, namely those that arise from a partially ordered set.

Definition 2. For a partially ordered set $(P, \leq)$ and any element $x \in P$,

$$U_x = \{y \in P : x \leq y\}$$

is called an upward set. The collection of all unions of upward sets forms a topology on $P$, called the Alexandrov topology. Any union of upward sets is therefore an open set in this topology.

We need a little more structure to capture the values of the set of measurements or parameters on each upward set; this is what defines a sheaf.

Definition 3. [18] A sheaf $\mathcal{S}$ on a partially ordered set $(P, \leq)$ consists of the following data:

- Stalks to each $x$ in $P$, there is associated a pseudometric space $\mathcal{S}(x)$ with pseudometric $d_x$,
- Restrictions to each relation $x \leq y$, there is associated a continuous map $\mathcal{S}_{x \leq y} : \mathcal{S}(x) \to \mathcal{S}(y)$.
such that if $x \leq y$ and $y \leq z$, it follows that $S_{x \leq z} = S_{y \leq z} \circ S_{x \leq y}$. We call the Alexandrov topology for $(P, \leq)$ the base space of the sheaf $S$.

We will often define a sheaf $S$ on $(P, \leq)$ by annotating the Hasse diagram for $(P, \leq)$ with the stalks and restrictions; such a diagram is called a sheaf diagram.

Notice that the symbol $\mathcal{S}$ is polymorphic: we use the same symbol—but with a different argument—to represent a set or a function between sets. This helps keep the notation from becoming overbearing when several sheaves are being used.

Specifying values from some of the stalks should be interpreted as either proposing a possible set of parameters, recording a measurement, or some combination of both. All of these variations are formalized by the concept of an assignment.

**Definition 4.** An assignment $a$ to a sheaf $\mathcal{S}$ on a partially ordered set $(P, \leq)$ supported on $R \subseteq P$ is an element of the Cartesian product of sets

$$a \in \prod_{x \in R} \mathcal{S}(x).$$

Note carefully that each stalk $\mathcal{S}(x)$ is a set. An assignment $a$ supported on $R$ can be thought of as a list of values, one for each element of $R$, in which the permissible values in the list are drawn from the corresponding stalks of $\mathcal{S}$.

It is often useful to distinguish between local and global assignments. A global assignment is one for which the support set $R$ is equal to the entire poset $P$. Any assignment that is not global is a local assignment.

Assignments are not concerned with the restrictions of the underlying sheaf, but when there is consistency between the sheaf and an assignment, this means that

$$\mathcal{S}_{x \leq y}(a(x)) = a(y).$$

The left side of this equation is the output of the restriction function $\mathcal{S}_{x \leq y}$ given the the value of the assignment at $x$, namely $a(x)$, as input. This output is compared with the value already specified by the assignment for $y$, namely $a(y)$.

**Definition 5.** A global assignment for which equation (4) holds for every pair of elements $x, y \in P$ such that $x \leq y$ is called a global section of $\mathcal{S}$.

If a sheaf is used to specify a signal model, global sections are best interpreted as the signals permitted by the model.

Since the stalks of a sheaf $\mathcal{S}$ are pseudometric spaces, we can quantify how far an assignment is from being a global section using the consistency radius.

**Definition 6 (19, Def. 20).** The consistency radius of a global assignment $a$ to a sheaf $\mathcal{S}$ on a partially ordered set $(P, \leq)$ is

$$c_{\mathcal{S}}(a) = \sum_{x \in P} d_{\mathcal{S}}(\mathcal{S}_{x \leq y}(a(x)), a(y))^2. $$

Notice that global sections have zero consistency radius. The most crucial fact about the consistency radius is that it bounds the overall distance from an assignment to the nearest global section.

**Proposition 2 (19, Prop. 23).** Let $a$ be an assignment to a sheaf $\mathcal{S}$ of pseudometric spaces on $(P, \leq)$ in which each restriction map of $\mathcal{S}$ is Lipschitz with constant $K$, then

$$D(a, s) \geq \frac{c_{\mathcal{S}}(a)}{1 + K} $$

for every global section $s$ of $\mathcal{S}$, where

$$D(a, s) = \sum_{x \in P} d_{\mathcal{S}}(a(x), s(x))^2.$$

As a result, we may transform the system of equations required by a global section into an optimization problem that minimizes consistency radius. Solutions to a consistency radius minimization problem are often more robust to noise and uncertainty than solutions to the original system of equations [13].
The consistency radius can be localized to a smaller portion of the underlying partial order. This is important for specifying how greedy algorithms work in the next few sections of the article, though the required definition is quite easy.

**Definition 7 (20, Def. 15, Prop. 8).** Suppose that \( U \) is an open set in the Alexandrov topology of \((P, \leq)\), which means that it is a union of upward sets. The local consistency radius of the assignment \( a \) to a sheaf \( S \) supported on \( U \) is

\[
\epsilon_S(a; U) = \sqrt{\sum_{x \leq y \in U} d_x(S_{x \leq y}(a(x)), a(y))^2}
\]

### 4.2. A sheaf model for non-adaptive decompositions

We are now ready to write modal decomposition problems as sheaves, in which the solutions to these problems are global sections.

If \( M \) real measurements are taken from a signal determined by \( N \) sources, this situation can be modeled by a simple sheaf diagram \( \mathcal{L}_N \)

\[
\begin{align*}
I^N / S_N \\
\downarrow S \\
\mathbb{R}^M
\end{align*}
\]

where the function \( S \) is given by equation (3). The consistency radius of \( \mathcal{L}_N \) on an assignment is simply

\[
\epsilon_{\mathcal{L}_N}((w_1, \ldots, w_N, z)) = \|S(w_1, \ldots, w_N) - z\|,
\]

where \( z \in \mathbb{R}^M \) is the measurement and \( w_i = (a_i, b_i) \) is the parameter of the \( i \)-th source. It bears stating what is perhaps obvious: if the number of sources is known to be \( N \) and the source parameters \( w \) match the measurements \( z \), then the consistency radius of this assignment is zero.

**Corollary 3.** If the \( S \) function is injective, then the consistency radius of an assignment \((w_1, \ldots, w_N, z) \) of \( \mathcal{L}_N \) vanishes precisely when the values of the assignment satisfy the modal decomposition, equation (1).

If we do not know the number of sources, then we must consider various instantiations of the sheaf \( \mathcal{L}_N \). The most obvious way to proceed is simply to combine the diagrams for various values of \( N \), resulting in a new sheaf \( \mathcal{J}_P \).

**Definition 8.** The sheaf \( \mathcal{J}_P \) is defined by the diagram

\[
\begin{align*}
I \quad & \quad I^2 / S_2 \\
\downarrow S(\cdot) \\
\mathbb{R}^M & \quad S(\cdot, \cdot) \quad \quad \downarrow \quad \quad \downarrow \\
& \quad I^3 / S_3 \\
& \quad \quad \vdots \\
& \quad I^P / S_P
\end{align*}
\]

in which each of the restriction maps are given by the signal space map (equation (3)) with different numbers of sources.

The definition of \( \mathcal{J}_P \) explicitly contains an upper limit \( P \) to the maximum number of sources that will be considered, which may not be related to the true number of sources \( N \). Successfully decomposing a signal requires that \( N \leq P \), so that we consider a decomposition with the correct number of sources, though evidently \( N \) is not necessarily known from the outset.

The global consistency radius of an assignment to this sheaf is given by

\[
\epsilon_{\mathcal{J}_P}((w_1, \ldots, z)) = \sqrt{\frac{1}{P} \sum_{i=1}^{P} \|S(w_1, \ldots, w_i) - z\|^2},
\]

where \( z \in \mathbb{R}^M \) is the measurement and the \( w_{ij} \) is the proposed \( j \)-th source parameter when we are considering a representation of the signal with \( i \) sources.

The square of the global consistency radius is simply the sum of squares of consistency radii for each \( \mathcal{L}_i \) that is part of the diagram of \( \mathcal{J}_P \). Each \( \mathcal{L}_i \) can be thought of as a subproblem, in which we propose a number \( k \) of sources (in hopes of finding the correct number). The global consistency radius for \( \mathcal{L}_i \) becomes a local consistency radius for an open set in the base space topology for \( \mathcal{J}_P \). If each of the \( S \) functions is injective, then if \( N > P \), none of these local consistency radii will vanish. Therefore, the minimum value of \( \epsilon_{\mathcal{J}_P} \) will not be zero. If
It is useful to construct a specific kind of sheaf that aggregates observations or parameter values under different assumptions. In this section, we define the concept of an aggregation sheaf that represents the idea of nested sets of parameters.

Suppose that \( I \) is a conical space and \( S_N \) is the group of permutations of \( N \) items. The elements of \( I^N/S_N \) are equivalence classes of elements of \( I^N \) under permutations. We may choose representatives of these equivalence classes so that the components (elements of \( I \)) are sorted according to their magnitude norms.

For instance, a typical representative of an element of \( I^N/S_N \) could be written

\[
[w_1, w_2, \ldots, w_N]
\]

where \( ||w_i|| \geq ||w_j|| \) if \( i \leq j \).

A natural operation is concatenation \( \otimes : (I^M/S_M) \times (I^N/S_N) \rightarrow I^{M+N}/S_{M+N} \) given by

\[
\{u_0, \ldots, u_M \} \otimes \{w_0, \ldots, w_N \} = \{u_0, \ldots, u_M, w_0, \ldots, w_N \}.
\]

We may therefore abuse notation slightly, and think of elements of \( I^N/S_N \) as length \( N \) multisets: unordered lists of length \( N \) in which duplicates are permitted. We will therefore speak of \( w_i \) as being an element of \( \{w_1, \ldots, w_M\} \) without any possible confusion.

**Lemma 2.** If \( I \) is a conical space, then \( I^N/S_N \) is a metric space in which the metric is given by

\[
d(u, w) = \min_{\sigma \in S_N} \sum_{n=1}^{N} d_I(u_n, w_{\sigma(n)}).
\]

**Proof.** We simply need to establish the axioms of a metric:

- Because \( d(u, w) \) is a sum of norms, \( d(u, w) \geq 0 \) is immediate.
- Similarly, \( d(w, w) = 0 \) is immediately apparent by taking \( \sigma = \text{id} \).
- If \( d(u, w) = 0 \), this means that there is a permutation \( \sigma \) such that \( u_n = w_{\sigma(n)} \) for each \( n \) because \( d_I \) is a metric. This implies that \( u \) and \( w \) are in the same equivalence class of \( I^N/S_N \).
- A simple reindexing argument establishes that \( d(u, w) = d(w, u) \).
- Finally, \( d(u, v) \leq d(u, w) + d(w, v) \)

\[
d(u, v) = \min_{\sigma \in S_N} \sum_{n=1}^{N} d_I(u_n, w_{\sigma(n)})
\]

\[
\leq \min_{\tau \in S_N} \min_{\sigma \in S_N} \left( \sum_{n=1}^{N} d_I(u_n, w_{\tau(n)}) + \sum_{n=1}^{N} d_I(w_{\tau(n)}, w_{\sigma(n)}) \right)
\]

\[
\leq \min_{\tau \in S_N} \sum_{n=1}^{N} d_I(u_n, w_{\tau(n)}) + \min_{\tau \in S_N} \min_{\sigma \in S_N} \sum_{n=1}^{N} d_I(w_{\tau(n)}, w_{\sigma(n)})
\]

\[
\leq d(u, w) + d(w, v).
\]

With these basic properties established, we now aggregate copies of \( I \) according to a recipe encoded as a partially ordered set \((P, \leq)\). Intuitively, the elements of \( I \) represent choices of parameters for the sources, and the elements of \( P \) represent different possible ways to configure these sources. The partial order \( \leq \) on \( P \) represents nesting relations between these configurations.

For instance the relation between a configuration with 2 sources can be thought of as a sub-configuration of one with 3 sources in several possible ways. The appropriate representation of the space of parameters along with these relationships is an aggregation sheaf.
**Definition 9.** Suppose that $D_p$ is the subset of $P$ in the partially ordered set $(P, \leq)$ given by

$$D_p := \{ q \in P : q \leq p \},$$

and that $I$ is a conical space. The aggregation sheaf $A$ on $(P, \leq)$ modeled on $I$ is a sheaf of metric spaces constructed according to the following recipe:

**Stalks.** $A(P) := (\bigoplus_{q \in D_p} I) / S_{\# D_p}$, which can be interpreted as an unordered concatenation of conical spaces,

**Restrictions** if $A(q)$ is a multiset of $N$ elements $\{w_1, \ldots, w_N\}$ and $A(p)$ is a multiset of $N + k$ elements, then

$$A_{q \leq p}(\{w_1, \ldots, w_N\}) := \{w_1, \ldots, w_N, (0, b_1), \ldots, (0, b_k)\},$$

where $b_1, \ldots, b_k$ are completely arbitrary, since they all refer to the vertex of $I$. Briefly, $A_{q \leq p}$ copies the elements of $A(q)$, padding with copies of the vertex of $I$ as needed. We call $A_{q \leq p}$ the inclusion of $A(q)$ into $A(p)$.

According to lemma 2, the stalks of an aggregation sheaf are metric spaces. It is immediate that the restrictions are continuous with respect to the metrics on the stalks.

For our purposes here—decomposing signals formed according to equation (1)—we want to understand the most efficient way to solve certain consistency radius minimization problems. Constraining the assignments that must be tested as the optimization problem is solved can yield a performance boost. Greedy algorithms are often preferred in the literature due to their speed of convergence. However, greedy algorithms can fail spectacularly if their assumptions are not met. We seek both the correct optimization problem to decompose equation (1) and a flexible algorithmic framework that becomes greedy when appropriate, but fails more gracefully into a non-greedy algorithm if needed. This graceful degradation is governed by properties of the assignments which minimize consistency radius; a key starting point is a sorting property for assignments of aggregation sheaves.

**Proposition 3.** Sorting property. Suppose that $A$ is an aggregation sheaf for a finite partially ordered set $(P, \leq)$ modeled on the conical space $I$, that $(P, \leq)$ has a unique maximal element $p'$, and that $N = \# P$. Let $a$ be an assignment to $A$ supported on $p'$. Without loss of generality, suppose that $a_p = \{w_1, w_2, \ldots, w_N\}$, where $\|w_i\| \geq \|w_j\|$ if $i \leq j$, recalling that $\|\cdot\|$ is the magnitude norm of $I$. Any extension $b$ of $a$ to all of $A$ which has minimal consistency radius will have the property that if $q \leq p$ as elements in $P$,

$$b_q = b_q \oplus \{w_i, \ldots\},$$

where every element in the set $w_i, \ldots$ has magnitude norm less than or equal to that of every element in $b_q$ and each $w_i$ is an element of $a_p$.

**Proof.** It suffices to establish the conclusion for an arbitrary pair $q \leq p$ in $P$ for which there is no other $r$ in $P$ ‘between’ $q$ and $p$, that is, $q \leq r \leq p$ implies $r = q$ or $r = p$. Under this hypothesis, the lengths of $b_p$ and $b_q$ differ by exactly 1. The sheaf diagram in the vicinity of $p$ and $q$ is therefore of the form

$$I^N / S_N \xrightarrow{A_{q \leq p}} I^{N+1} / S_{N+1}.$$

Without loss of generality, suppose that $b_p = \{w_1, w_2, \ldots, w_{N+1}\}$ in which the elements are sorted in order of descending magnitude norm. The proposition will be established if we can conclude that $b_q = \{w_1, w_2, \ldots, w_N\}$. Suppose that $b_q = \{u_1, u_2, \ldots, u_N\}$, so that

$$d(A_{q \leq p}(b_q), b_p) = \min_{\sigma \in S_{N+1}} \sum_{\mu=1}^{N+1} d_1((A_{q \leq p}(b_q))_{\sigma\mu}, w_{\sigma(1)})$$

$$= \min_{\sigma \in S_{N+1}} \left( \sum_{\mu=1}^{N} d_1(u_{\sigma\mu}, w_{\sigma(1)}) + d_1((0, b), w_{\sigma(N+1)}) \right)$$

$$= \min_{\sigma \in S_{N+1}} \left( \sum_{\mu=1}^{N} d_1(u_{\sigma\mu}, w_{\sigma(1)}) + \|w_{\sigma(N+1)}\| \right).$$

Notice that since $A_{q \leq p}$ is an inclusion, its output contains one extra copy of the vertex of $I$ as padding. This extra vertex copy is paired with $w_{\sigma(N+1)}$, resulting in the magnitude norm in the above calculation. Since we sorted the $w_{\sigma\mu}$, selecting $u_\sigma = w_\mu$ for $n = 1, \ldots, N$ means that the above distance is not more than $\|w_{N+1}\|$. On the other hand, notice that...
Therefore, selecting \( u_n = w_n \) for \( n = 1, \ldots, N \) results in the minimum value of \( d(A_{q \in J}(b_q), b_p) \). Since this is the contribution of \( b_p \) to the consistency radius of the entire assignment \( b \), we conclude that it minimizes consistency radius. \( \square \)

According to proposition 3, the assignments with minimal consistency radius on the aggregation sheaf

\[
I \rightarrow I^2/S_2 \rightarrow \cdots \rightarrow I^N/S_N
\]

are always of the form

\[
\{ w_1 \} \rightarrow \{ w_1, w_2 \} \rightarrow \cdots \rightarrow \{ w_1, \ldots, w_N \}
\]

where \( \| w_i \|_I \geq \| w_j \|_I \) if \( i \leq j \). That is, the elements of \( I \) added as one moves rightward in the diagram are sorted in descending order.

### 4.4. Sheaf model for nested subproblems

To remedy the issues with \( J_p \), and to create a sheaf model suitable for greedy algorithms, let us repackage the top row of equation (6) using an aggregation sheaf. Since we do not know the true number of sources, it makes sense to encapsulate all possibilities in a sheaf \( \mathcal{K} \). This sheaf is an aggregation sheaf

\[
\mathcal{K} \rightarrow \mathcal{K}/S_2 \rightarrow \mathcal{K}/S_3 \rightarrow \cdots \rightarrow \mathcal{K}/S_n \rightarrow \cdots
\]

which represents all possible sets of sources present in the signal along with the nesting structure. At any place in the diagram above, there are a finite number of sources, and these sources are present in all positions to the right.

This sheaf \( \mathcal{K} \) can be connected to the measurements by replicating the diagram in equation (5), resulting in the sheaf \( \mathcal{K} \) defined below.

**Definition 10.** The sheaf \( \mathcal{K} \) is defined by the diagram

\[
\begin{array}{c}
\text{I} \\
S(\cdot) \\
\mathbb{R}^N \rightarrow \text{I}/S_2 \\
\rightarrow \text{I}/S_3 \\

\mathcal{K}/S_p
\end{array}
\]

in which the top row is an aggregation sheaf, and the restrictions from the top to bottom row are given by the signal map \( S \).

It should be immediately apparent to the reader that the consistency radius for this new sheaf \( \mathcal{K} \) is rather formidable to write explicitly, although it is straightforward to construct systematically. Specifically, because of commutativity all restriction maps from the top to bottom row are actually visible in the diagram. Moreover, the top row is an aggregation sheaf, so its consistency radius can be written explicitly using the construction in section 4.3.

Consider an assignment to the above sheaf \( \mathcal{K} \), in which the values in the stalks in the top row are given by

\[
\{ w_{1,1}, \{ w_{2,1}, w_{2,2} \}, \{ w_{3,1}, w_{3,2}, w_{3,3} \}, \ldots
\]

and the value in the bottom row (the \( \mathbb{R}^2 \)) is \( z \). The square of the consistency radius of this assignment is

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} \min_{\sigma \in S_1} \left( \sum_{k=1}^{i} d_i(w_{k,\sigma(k)}, w_{k,\sigma(k)}) + \sum_{m=i+1}^{j} \| w_{k,\sigma(m)} \|_I \right)^2 \sum_{j=1}^{N} \| S(w_{i,1}, \ldots, w_{i,j}) - z \|^2.
\]

Let us compare \( \mathcal{K} \) with the sheaf \( J_p \) defined earlier in equation (6). Although the base spaces of these two sheaves have different topologies, they are written on the same set of elements. The only difference is the partial order on these elements. In particular, every order relation present in the base of \( J_p \) (arrows from the top to the bottom row) is present in the base of \( \mathcal{K} \), though not conversely. This means that the identity map from the base space of \( J_p \) to the base space of \( \mathcal{K} \) is order preserving. Furthermore, notice that the stalks and restrictions in \( J_p \) are present in exactly the same form as in \( \mathcal{K} \), though again not conversely. Therefore, there is a sheaf morphism \( m: \mathcal{K} \rightarrow J_p \) defined along this identity map, which means that every diagram of the form
Proposition 4. The sheaf morphism \( m : \mathcal{K}_p \to \mathcal{J}_p \) induces an isomorphism on the spaces of global sections provided that \( S : I^p / S_p \to \mathbb{R}^M \) is injective.

Proof. The fact that \( m \) takes each section of \( \mathcal{K}_p \) to a section of \( \mathcal{J}_p \) is a general fact for sheaf morphisms and is true regardless of the injectivity of \( S \). The other direction, that sections of \( \mathcal{J}_p \) also correspond to sections of \( \mathcal{K}_p \), requires special consideration of the sheaves involved. First of all, notice that global sections of \( \mathcal{J}_p \) are determined by their value on \( I \), since this is the stalk over the unique minimal element in the base partial order. We merely need to show that a global section of \( \mathcal{K}_p \) is also determined by its value on the corresponding element in its base partial order, even though that element is not the unique minimal element. To this end, suppose that we have a global section of \( \mathcal{J}_p \). Such a global section consists of the following information:

1. A value \( z \in \mathbb{R}^M \), and
2. A set of values \( \{ w_{n,i} \}_{i=1}^n \) in \( I^n / S_n \) for each \( n = 1, \ldots, p \), each of which consists of a pair \( w_{n,i} = (a_{n,i}, b_{n,i}) \) because \( I \) is a conical space,

such that

\[
z = S(w_{n,1}, \ldots, w_{n,n})
\]

for each \( n = 1, \ldots, p \). Because of the injectivity of \( S \), this means that in particular, we have that

\[
z = S(w_{p,1}, \ldots, w_{p,p}).
\]

By the injectivity of \( S \), for this particular value of \( z \), no other set \( \{ w_{p,i} \}_{i=1}^p \) will satisfy the equation. On the other hand, we have that

\[
z = S(w_{1,1}),
\]

and again uniquely so. This means that \( w_{1,1} \) must be present in \( \{ w_{p,i} \}_{i=1}^p \), and furthermore must be the only non-vertex element (recall that the vertex is the equivalence class consisting of all elements with \( a_{i,j} = 0 \)). We merely need to notice that this means that the global section is determined by its value \( w_{1,1} \) on \( I \).

The proof of proposition 4 is easily extended to handle the situation of local sections of \( \mathcal{K}_p \); merely trim off the left side of the diagram!

Corollary 4. Every local section of \( \mathcal{K}_p \) is also a local section of \( \mathcal{J}_p \).

This means that if we use \( \mathcal{K}_p \) instead of \( \mathcal{J}_p \), we essentially have the same information, simply encapsulated in a smaller topology. The smaller topology might help guide our attempts at finding extensions of assignments with minimal consistency radius, and so may be rather convenient even though the consistency radius formula for \( \mathcal{K}_p \) is more complicated than that of \( \mathcal{J}_p \).

4.5. Determining the number of sources adaptively

The general strategy for determining source parameters from measurements is to minimize consistency radius in either the sheaf \( \mathcal{J}_p \) or the sheaf \( \mathcal{K}_p \), subject to the assignment being determined at the observation. In cases where the source magnitudes are widely varying, it is likely that a greedy algorithm is possible (and may be preferable). Minimizing consistency radius on \( \mathcal{K}_p \) will tend to exhibit the sorting property (Proposition 3), and therefore will prioritize sources with large magnitude first. On the other hand, if the source magnitudes are all similar, then a greedy algorithm is likely to fail when source locations are similar. In this situation, minimizing consistency radius in \( \mathcal{K}_p \) may yield misleading results; instead, minimizing consistency radius on \( \mathcal{J}_p \) would be preferable.

Proposition 5. Suppose that the number of sources is given as \( N \) and that \( z \in \mathbb{R}^M \) is given by

\[
z = S(a_1, b_1, \ldots, a_N, b_N),
\]
where $S$ is given by equation (3). Consider an assignment in which $z \in \mathbb{R}^M$ represents the value in stalk in the bottom row of the sheaf $\mathcal{F}_P$ (see equation (6)) or $\mathcal{K}_P$ (see equation (7)). If $P \geq N$, then there is an extension of this assignment to a global assignment in which the local consistency radius vanishes for all sufficiently small open sets.

**Proof.** Establishing the statement for $\mathcal{K}_P$ automatically carries over to $\mathcal{F}_P$, so we merely consider the case of $\mathcal{K}_P$.

That is given by

$$I \rightarrow I^2 / S_2 \rightarrow I^3 / S_3 \rightarrow \cdots \rightarrow I^P / S_P$$

To that end, $z \in \mathbb{R}^M$ is assigned in the bottom row of the diagram. This is merely a consequence of the structure of the global consistency radius, initially shown in equation (8). Suppose that the assignment takes the following values on the top row of the diagram

$$\{(a_{1,1}, b_{1,1}), \{(a_{2,1}, b_{2,1})\}, \{(a_{2,2}, b_{2,2})\}, \{(a_{3,1}, b_{3,1}), (a_{3,2}, b_{3,2}), (a_{3,3}, b_{3,3})\}, \ldots, \{(a_{P,1}, b_{P,1}), \ldots, (a_{P,P}, b_{P,P})\}.$$  

The square of the consistency radius of any assignment supported on the minimal open set (in the sense of inclusion) whose stalk is $I^N / S_N$ is given by

$$\sum_{i=N}^{P} \sum_{j=i+1}^{P} \min_{\sigma \in S_i} \left( \sum_{k=1}^{i} d_i ((a_{i,k}, b_{i,k}), (a_{j,\sigma(k)}, b_{j,\sigma(k)})) + \sum_{m=i+1}^{j} \|a_{j,\sigma(m)}\|^2 \right)^2 + \sum_{i=N}^{P} \|S(a_{i,1}, b_{i,1} \ldots a_{i,j}, b_{i,j}) - z\|^2. \quad (9)$$

(Specifically, in equation (9) rather than starting the $i$ sums at 1 they start at $N$. Additionally, all sums run to $P$ instead of $N$.) The extension to a global assignment places as many of the correct source parameters in each stalk as is possible. Any additional source parameters are unneeded, and may simply be chosen to have zero magnitude.

$$a_{P,k} = \begin{cases} a_k & \text{if } k \leq N, \\ 0 & \text{otherwise}. \end{cases}$$

The other parameters may be set as well, though the $b_k$ parameters for any unneeded source may be chosen arbitrarily, since they are ignored by the conical metrics,

$$b_{P,k} = \begin{cases} b_k & \text{if } k \leq N, \\ \text{arbitrary} & \text{otherwise}. \end{cases}$$

Supposing that we choose $a_{i,k} = a_{j,k}$ and $b_{i,k} = b_{j,k}$ whenever they are both defined, then this definition clearly results in the second sum of equation (9) being zero. On the other hand, the sum

$$\sum_{k=1}^{i} d_i ((a_{i,k}, b_{i,k}), (a_{j,\sigma(k)}, b_{j,\sigma(k)}))$$

will also be made to vanish, simply by letting $\sigma = \text{id}$. What of the remaining sum? It too vanishes because all of the terms in question have index greater than $N$, and therefore each vanishes by construction.  

**Corollary 5.** The assignment for which the local consistency radius of both $\mathcal{F}_P$ and $\mathcal{K}_P$ vanishes occurs precisely at the number of sources $N$, assuming that the function $S$ is injective for $P$ sources.

As a consequence of [20], lemma 3, the local consistency radius is monotonically decreasing as the number of sources being considered increases. With this in mind, corollary 5 implies that it is possible to determine the correct number of sources by sweeping through the local consistency radii to find where the local consistency radius decreases sharply to zero. This is a reasonable methodology even if there is noise in the measurements or if the signal model is only an approximation, because although the local consistency radius will not reach zero, it will still decrease sharply at the correct number of sources.

5. Application to several example problems

This section outlines a few examples of the sheaf constructions described in this article. It also records the results from several of them, along with a description of a general framework for sheaf-based calculations.
5.1. Fourier decomposition

If the samples are equally spaced, then the signal model is given by

\[ S(a_{-N}, -N, a_{-N+1}, -N + 1, \ldots, a_N, N) = \left( \sum_{n=-N}^{N} a_n e^{2\pi i n h}, \sum_{n=-N}^{N} a_n e^{2\pi i (M-1) n} \right) \]

for some fixed \( h \in \mathbb{R} \).

We can express this diagrammatically as

\[
\begin{array}{c}
\mathbb{C}^{2N+1} \\
\downarrow S \\
\mathbb{C}^M
\end{array}
\]

noting that minimizing consistency radius is tantamount to inverting the map \( S \). If the number of sources \( 2N + 1 \) is known, then \( S \) is simply a discrete Fourier transform. Provided \( M = 2N + 1 \), the inverse discrete Fourier transform will recover the source parameters \( a_n \) without any further trouble.

5.2. Nonlinear two-source problems

Consider the following conical space \( I = \mathbb{R} \times \mathbb{R} \) under the metric

\[ d((a_1, b_1), (a_2, b_2)) = \min \{|a_1| + |a_2|, \sqrt{|a_1 - a_2|^2 + \beta |b_1 - b_2|^2} \}. \]

Let us treat \( I \) as the representation of a single source, whose magnitude and position are each given by a scalar. We consider two Examples, 1 and 2, each of which involves the consistency radius of a two-source problem. In Example 1, both magnitudes are known and have the same value, while in Example 2 they can have different values.

In both Examples, the sheaf diagram of the problem has the form

\[
\begin{array}{c}
A_1 \xrightarrow{i} A_2 = I^2 / S_2 \\
\downarrow g \\
B = \mathbb{R}^M
\end{array}
\]

where \( \mathbb{R}^M \) has the usual Euclidean metric. This is an example of the sheaf \( K_2 \) in what follows. To keep the notation clear, we will name the assignment on \( A_1 \) to be \((a_{11}, b_{11})\), the assignment on \( A_2 \) to be \((a_{21}, \ b_{21}), (a_{22}, \ b_{22})\), and the assignment on \( B \) to be \( z \). The restriction map \( f \) is given by the formula

\[ f((a_{21}, b_{21}), (a_{22}, b_{22})) = a_{21} \phi(b_{21}) + a_{22} \phi(b_{22}), \]

where \( \phi: \mathbb{R} \to \mathbb{R}^M \) is a smooth function. Since the top row of the sheaf diagram is an aggregation sheaf, we have that

\[ i((a_{11}, b_{11})) = ((a_{11}, b_{11}), (0, c)), \]

where \( c \) is arbitrary (but not required for well-definedness) since \((0, c)\) is the vertex of \( I \). Commutativity of the sheaf diagram for \( K_2 \) ensures that the restriction \( g = f \circ i \) is given by

\[ g((a_{11}, b_{11})) = a_{11} \phi(b_{11}). \]

**Example 1.** Let us assume that the magnitudes of both sources have the same value. Without loss of generality, we may declare that \( a_1 = a_2 = 1 \). The \( K_2 \) sheaf model for this two-source problem becomes
Given the unit magnitudes, we note that
\[ b = -+ + - + - = + \]
\[ \begin{align*}
1, 1, 1, & \\
11 & \end{align*} \]

This implies that the square of the consistency radius is
\[ c^2(b_1, b_2, b_2) = \beta(b_1 - b_2) + 1 + (f(b_2) - z)^2 \]
\[ = \beta(b_1 - b_2)^2 + 1 + (\phi(b_1) + \phi(b_2) - z)^2 + (\phi(b_1) - z)^2. \]

An expression of corollary 5 is that the third term in the above,
\[ (\phi(b_1) + \phi(b_2) - z)^2, \]
is the local consistency radius that can vanish if the correct source locations are used, since there are precisely two sources. Evidently the global consistency radius cannot vanish (ever)!

**Example 2.** Now let us consider the situation in which the source magnitudes are not known. The sheaf diagram for \( K_2 \) is given by

\[
\begin{align*}
A_1 &= \mathbb{R}^2, & A_2 &= \mathbb{R}^2 \times \mathbb{R}^2, \\
B &= \mathbb{R}^M, & f &= \begin{cases} \\
\end{cases}
\end{align*}
\]

The square of the consistency radius for an assignment to \( K_2 \) is given by
\[
c^2_{K_2}(a_{11}, b_1, a_{21}, a_{22}, b_2) = \|a_{21}\phi(b_1) + a_{22}\phi(b_2) - z\|^2 + \|a_{21}\phi(b_1) - z\|^2.
\]

If there are two sources, namely \( z = a_{21}\phi(b_1) + a_{22}\phi(b_2), \) the minimum consistency radius in \( K_2 \) occurs precisely when the two terms of \( c^2_{K_2} \) are minimized. Since this can be done independently, if \( a_{22} = 0, \) then the second term will never vanish but the first will vanish. On the other hand, if \( a_{22} = 0, \) then the global minimum consistency radius will be zero and will ensure that \( (a_{11}, b_1) = (a_{21}, b_2). \)

Let us now consider what happens with \( K_2 \). The sheaf diagram for \( K_2 \) is

\[
\begin{align*}
A_1 &= \mathbb{R}^2, & A_2 &= \mathbb{R}^2 \times \mathbb{R}^2, \\
B &= \mathbb{R}^M, & f &= \begin{cases} \\
\end{cases}
\end{align*}
\]

The square of the consistency radius of an assignment to \( K_2 \) contains several more terms than \( J_2 \), and is given by
\[
c^2_{K_2}(a_{11}, b_1, a_{21}, a_{22}, b_2, a_{22}) = \|a_{21}\phi(b_1) + a_{22}\phi(b_2) - z\|^2 + \|a_{21}\phi(b_1) - z\|^2.
\]

Let us assume that the single-source problem is a subproblem of the two-source problem. That means that \( (a_{11}, b_1) = (a_{21}, b_2). \)

Under the above assumption, the consistency radius is a function of \( a_{21}, b_1, a_{22}, \) and \( b_2 \) only,
\[
c^2_{K_2}(a_{21}, b_1, a_{22}, b_2) = |a_{22}|^2 + \|a_{21}\phi(b_1) + a_{22}\phi(b_2) - z\|^2 + \|a_{21}\phi(b_1) - z\|^2.
\]

Assume that the measured signal is from two sources, so that \( z = a_{21}\phi(b_1) + a_{22}\phi(b_2). \) The square of the consistency radius becomes
\[
c^2_{K_2} = |a_{22}|^2 + \|a_{21}\phi(b_1) - z\|^2
\]
\[
= |a_{22}|^2 + \|z - a_{22}\phi(b_2)|^2
\]
\[
= |a_{22}|^2(1 + \|\phi(b_2)|^2)
\]

This vanishes if and only if the measurements are from a single source, since in that case \( a_{22} = 0. \) On the other hand, where are the critical points? We begin with a preliminary calculation:
\[
\frac{\partial}{\partial a_{11}} \|a_{11} \phi(b_{11}) - z\|^2 = \frac{\partial}{\partial a_{11}} ((a_{11} \phi(b_{11}) - z) \cdot (a_{11} \phi(b_{11}) - z)) \\
= 2 \phi(b_{11}) \cdot (a_{11} \phi(b_{11}) - z).
\]

That established, the gradient of the square of the consistency radius is given by
\[
d(\epsilon_{\mathbb{K}}^2) = 2 \left( \begin{array}{c} \phi(b_{21}) \cdot (a_{21} \phi(b_{21}) + a_{22} \phi(b_{22}) - 2z) \\
(d\phi(b_{21}))(a_{21} \phi(b_{21}) + a_{22} \phi(b_{22}) - z) \\
(a_{22} + \phi(b_{22}))(a_{21} \phi(b_{21}) + a_{22} \phi(b_{22}) - z) \\
(d\phi(b_{22}))(a_{22} \phi(b_{22}) + a_{22} \phi(b_{22}) - z) \end{array} \right).
\]

Continuing to assume there are two sources, so that
\[
z = a_{21} \phi(b_{21}) + a_{22} \phi(b_{22}),
\]
the gradient of the square of the consistency radius simplifies to
\[
d(\epsilon_{\mathbb{K}}^2) = 2 \left( \begin{array}{c} \phi(b_{21}) \cdot (a_{21} \phi(b_{21}) - z) \\
(d\phi(b_{21}))(a_{21} \phi(b_{21}) - z) \\
0 \\
0 \end{array} \right).
\]

If there is only one source, then each of the coefficients of the gradient do indeed vanish since \(a_{22} = 0\), implying that the minimum of the consistency radius is the correct single source decomposition. On the other hand, if there are two sources, then the gradient cannot vanish in virtue of the third component. This implies that minimizing consistency radius will only recover approximations to the correct source parameters. These approximations improve in the limit when \(|a_{21}| \ll |a_{22}|\), which is effectively a condition required by the greedy algorithm posited by minimizing consistency radius in \(\mathbb{K}_p\).

5.3. Spectral estimation

Let us consider the situation described in section 2.2, where the signal model is a sum of complex sinusoids
\[
s(x) = \sum_{n=1}^{N} a_n e^{2\pi i \omega_n x},
\]
with complex magnitudes \(\{a_n\}\) and real frequencies \(\{\omega_n\}\) being the unknown parameters. For each source, the unknown parameters lie in \((\mathbb{C} \times \mathbb{R})\), which can be given a conical metric. If we make \(M\) measurements, so that \(x \in \mathbb{R}^M\), the \(\mathcal{F}_p\) sheaf therefore has the form
\[
(\mathbb{C} \times \mathbb{R}) \quad (\mathbb{C} \times \mathbb{R})^2 \quad \cdots \quad (\mathbb{C} \times \mathbb{R})^P \\
\mathbb{C}^M
\]

According to proposition 5, this will have zero local consistency radius on sufficiently small open sets provided that \(P \geq N\) and the correct source parameters are found. Therefore, minmization of local consistency radius over all open sets in the Alexandrov topology of the base space of the \(\mathcal{F}_p\) sheaf is guaranteed to recover all source parameters.

The number of open sets in the Alexandrov topology of the base space of \(\mathcal{F}_p\) is \(2^P + 2\). Since this count also contains the empty set and the measurement space \(\mathbb{C}^M\), this means that at most \(2^P\) optimizations need to be performed. However, most of the optimization problems are not relevant to the problem at hand—namely determining the source parameters. In actual fact, only \(P\) optimization problems are needed—one problem for each proposed number of sources. These problems are to be solved independently, since the sheaf structure does not include restriction maps between the elements in the top row of the diagram.

However, a greedy approach is suggested by the \(\mathbb{K}_p\) sheaf, whose diagram is given by
\[
(\mathbb{C} \times \mathbb{R}) \quad (\mathbb{C} \times \mathbb{R})^2 \quad \cdots \quad (\mathbb{C} \times \mathbb{R})^P \\
\mathbb{C}^M
\]

Given the fact that the top row of the \(\mathbb{K}_p\) sheaf is an aggregation sheaf, proposition 3 suggests—but does not establish—that solutions to the minimization of local consistency radius with a small number of sources and minimization of local consistency radius with a larger number of sources ought to be related. One can therefore proceed somewhat greedily, by using the solution for a given number of sources as an initial guess for the solution with either one more or one fewer sources. From a practical standpoint, both incrementally increasing and incrementally decreasing the number of sources are viable approaches.
6. Discussion and conclusions

The topological and sheaf-based framework posited in this article exhibits significant advantages over current representations of solving modal decomposition problems. Foremost is that a sheaf-based representation is essentially the most general possible way to state a modal decomposition problem, yet solvability conditions have been proven, and these conditions prescribe algorithmic strategies for constructing solutions. Regardless of what strategy is posited for solving a modal decomposition problem, it is subject to the mathematical formalism that is described in this article.

We call out three foci where the sheaf theoretic formalism may have a significant impact. First, the formalism applies to many other modal decomposition problems, beyond the simple cases explained in the article, without essential change. Indeed, what needs to be defined are the spaces of source parameters (as conical spaces), the signal map $S$, and the space of possible measurements.

Secondly, because of the generality of the sheaf-based representation, any other modal decomposition approach can be recast as a sheaf. It is a mathematical fact that the data within any well-posed data fusion problem can be recast as a sheaf [19]. Recovering the source parameters from measurements is clearly a kind of data fusion problem, although it is one in which the data are of homogeneous type. Although not stated in this way directly, the discussion in section 4 is simply a manifestation of the general data fusion result in the special case of modal decompositions.

Finally, since our starting point in section 2 was that of the most general modal decomposition, the sheaf-based representation is in the privileged position of acting as an unbiased arbitrator amongst algorithms. Indeed, since any other approach to solving a modal decomposition problem can be recast as a sheaf, several such sheaf models can be compared in an even-handed way.

Acknowledgments

The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00112090125.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Michael Robinson https://orcid.org/0000-0003-0766-3301

References

[1] Robinson M 2014 Topological Signal Processing (Berlin: Springer)
[2] Daubechies I 1992 Ten Lectures on Wavelets (Philadelphia, PA: SIAM)
[3] Rioul O and Vetterli M 1991 Wavelets and signal processing IEEE Signal Process Mag. 8 14–38
[4] Casazza P G, Kutyniok G and Philipp F 2013 Introduction to finite frame theory Finite Frames pp 1–53
[5] Casazza P G et al 2000 The art of frame theory Taiwanese Journal of Mathematics 4 129–201
[6] Fornasier M and Rauhut H 2015 Compressive sensing Handbook of Mathematical Methods in Imaging I 187–229
[7] Baraniuk R G 2007 Compressive sensing [lecture notes] IEEE Signal Process Mag. 24 118–21
[8] Huang J A 1974 Aperture synthesis with a non–regular distribution of interferometer baselines Astronomy and Astrophysics Supplement 15 417–26
[9] McKinnon G C, Burger C and Boesiger P 1990 Spectral baseline correction using CLEAN Magn. Reson. Med. 13 145–9
[10] Stewart I M, Fenech D M and Muxlow T W B 2011 A multiple-beam CLEAN for imaging intra–day variable radio sources Astron. Astrophys. 535
[11] Tsang M 2019 Resolving starlight: a quantum perspective Contemp. Phys. 60 279–98
[12] Mazidia H, Dinda T, Nehoria A and Lewa M D 2020 Quantifying accuracy and heterogeneity in single–molecule super–resolution microscopy Nat. Commun. 11
[13] Robinson M 2019 Hunting for foxes with sheaves Notices of the American Mathematical Society 66
[14] Robinson M, Henrich J, Capraro C and Zulch P 2018 Dynamic sensor fusion using local topology IEEE Aerospace Conference
[15] Lee J M 2013 Smooth Manifolds, Introduction to Smooth Manifolds (Berlin: Springer) pp 1–31
[16] Robinson M and Ghrist R 2012 Topological localization via signals of opportunity IEEE Trans. Signal Process. 60 2362–73
[17] Robinson M 2014 Knowledge–based antenna pattern interpolation IEEE. Trans. Ant. Prop. 72–9
[18] Baclawski K 1975 Whitney numbers of geometric lattices Adv. in Math. 16 125–38
[19] Robinson M 2017 Sheaves are the canonical data structure for sensor integration Information Fusion 36 208–24
[20] Robinson M 2020 Assignments to sheaves of pseudometric spaces Compositio–Tonalita 2