Sparse Representation in Structured Dictionaries with Application to Synthetic Aperture Radar

Kush R. Varshney, *Graduate Student Member, IEEE,* Müjdat Çetin, *Member, IEEE,* John W. Fisher, III, *Member, IEEE,* and Alan S. Willsky, *Fellow, IEEE*

Abstract

Sparse signal representations and approximations from overcomplete dictionaries have become an invaluable tool recently. In this paper, we develop a new, heuristic, graph-structured, sparse signal representation algorithm for overcomplete dictionaries that can be decomposed into subdictionaries and whose dictionary elements can be arranged in a hierarchy. Around this algorithm, we construct a methodology for advanced image formation in wide-angle synthetic aperture radar (SAR), defining an approach for joint anisotropy characterization and image formation. Additionally, we develop a coordinate descent method for jointly optimizing a parameterized dictionary and recovering a sparse representation using that dictionary. The motivation is to characterize a phenomenon in wide-angle SAR that has not been given much attention before: migratory scattering centers, i.e. scatterers whose apparent spatial location depends on aspect angle. Finally, we address the topic of recovering solutions that are sparse in more than one objective domain by introducing a suitable sparsifying cost function. We encode geometric

This work was supported in part by the Air Force Research Laboratory under Grant FA8650-04-1-1719, and Grant FA8650-04-C-1703 (through subcontract 04079-6918 from BAE Systems Advanced Information Technologies), and in part by a National Science Foundation Graduate Research Fellowship.

K. R. Varshney and A. S. Willsky are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: krv@mit.edu; willsky@mit.edu).

M. Çetin is with the Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı, Tuzla 34956 İstanbul, Turkey (e-mail: mcetin@sabanciuniv.edu).

J. W. Fisher, III is with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: fisher@csail.mit.edu).

objectives into SAR image formation through sparsity in two domains, including the normal parameter space of the Hough transform.

Index Terms

sparse signal representations, overcomplete dictionaries, optimization methods, tree searching, inverse problems, synthetic aperture radar, Hough transforms

I. INTRODUCTION

Whether for filtering, compression, or higher level tasks such as content understanding, the transformation of signals to domains and representations with desirable properties forms the heart of signal processing. The last decades have seen overcomplete dictionaries and sparse representations take a place in the processing of signals such as those that are multiscale in nature or can be traced to physical phenomena. By sparse, it is explicitly meant that a signal can be adequately represented using a small number of dictionary elements. Sparse signal representation and approximation has proven successful in solving inverse problems arising in a variety of application areas such as array processing [1], time-delay estimation [2], coherent imaging [3], electroencephalography [4], astronomical image restoration [5], and others. Inverse problems may be cast as sparse signal representation or approximation problems in conjunction with dictionaries whose elements have a physical interpretation, having been constructed based on the observation model of a particular application.

Representing a signal $\mathbf{g} \in \mathbb{C}^N$ using an overcomplete dictionary $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_M\}$, $M > N$ involves finding coefficients a_m such that $\mathbf{g} = \sum_{m=1}^{M} a_m \boldsymbol{\phi}_m$. Since the dictionary is overcomplete, there is no unique solution for the coefficients; additional constraints or objectives, e.g. sparsity, are needed to specify a unique solution. Among other properties, sparsity and overcomplete dictionaries have been known to deal well with undersampled data, and provide superresolution, parsimony, and robustness to noise. Traditionally, sparsity is measured using the ℓ_0 criterion, which counts the number of non-zero values. The problem of finding the optimally sparse representation, i.e. with minimum $\|\mathbf{a}\|_0^0$ where $\mathbf{a}$ is the set of coefficients taken as a vector in $\mathbb{C}^M$, is a combinatorial optimization problem in general. Due to the difficulty in solving large combinatorial problems, greedy algorithms such as matching pursuit [6] and relaxed formulations such as basis pursuit [7] that are computationally tractable have been developed for general overcomplete dictionaries. Methodologies such as these have been proven to produce optimally sparse solutions under certain conditions on the dictionary [8]–[10]. A sparse signal approximation is a set of coefficients subject to a sparse penalty such that $\|\mathbf{g} - \sum_{m=1}^{M} a_m \boldsymbol{\phi}_m\|_2^2$ is less than a small positive

constant.

Oftentimes, the dictionary elements ϕ_m, termed *atoms*, are chosen to have a physical interpretation. Atoms may correspond to different scales, translations, frequencies, and rotations or the dictionary may comprise subdictionaries, often given the name *molecules* [11]. Many popular sparse signal representation methods and algorithms are general and do not exploit natural decompositions of the dictionary into molecules or hierarchical structure that may be present in the collection of atoms. Some approaches do exist in the literature that take advantage of structured dictionaries, e.g. [11]–[16]. A main contribution of this paper is an approximate algorithm for sparse signal representation, related to heuristic search, that uses graphs, one per molecule, constructed with atoms as nodes connected according to hierarchical structure.

In the context of solving inverse problems using sparse signal representation techniques, the design of atoms based on the observation model is predicated on complete knowledge of the observation process. However, it may be the case that the functional form of the observation process is known, but there is dependence on some parameter or parameters that is not known a priori. In this case, it is of interest to both optimize the dictionary over the unknown parameters and to find sparse solution coefficients. In overcomplete representation contexts other than inverse problems, this can be viewed as signal-dependent dictionary refinement. A second contribution of this work is a coordinate descent approach that simultaneously refines the dictionary and determines a sparse representation.

Notationally, we take $\mathbf{\Phi}$ to be a matrix whose columns are atoms from the overcomplete dictionary, and $\mathbf{\Phi}(\boldsymbol{\eta})$ to reflect parametric dependence on the set of parameters $\boldsymbol{\eta}$. The matrix for a dictionary with L molecules is the concatenation of L blocks: $[\mathbf{\Phi}_1 \cdots \mathbf{\Phi}_L]$ or $[\mathbf{\Phi}_1(\boldsymbol{\eta}_1) \cdots \mathbf{\Phi}_L(\boldsymbol{\eta}_L)]$.

A fundamental premise of sparse signal representation is of underlying sparsity in some domain, but signals may be sparse in more than one complementary, or loosely speaking ‘orthogonal,’ domain. Accounting for and imposing simultaneous sparsity in multiple domains is important for recovering parsimonious representations. Representational redundancy that may not be apparent in one domain, but apparent in some other domain, can be appropriately reduced through sparsity in that other domain. We consider this problem of sparsity in more than one domain and, as a third contribution, develop a formulation whose objective function includes a carefully composed sparsity term for each domain.

Here we develop a general approach for sparse signal representation or approximation in which we exploit both molecular structure in dictionaries and hierarchical structure within molecules. Additionally, we incorporate dictionary optimization and simultaneously sparsity in multiple domains. While the methods have wider applicability, we focus on modeling wide-angle spotlight-mode synthetic aperture

radar (SAR) as an illustrative application. As a consequence, we advance the state of the art in radar imaging as well.

SAR is a technology for producing high quality imagery of the ground using a radar mounted on a moving aircraft. Radar pulses are transmitted and received from many points along the flight path. The full collection of measurements is used to form images; conventional image formation techniques are based on the inverse Fourier transform. In principle, very long flight paths—wide-angle synthetic apertures—which have become possible due to advances in sensor technologies, should allow for the reconstruction of images with high resolution. However, phenomena such as anisotropy and migratory scattering, described in the sequel, which arise in wide-angle imaging scenarios are not accounted for by conventional image formation techniques and cause inaccuracies in reconstructed images. As we proceed in the development of novel sparse signal representation methods for structured dictionaries, we use the methods described herein in a way that does account for such phenomenology.

In Section II we describe a heuristic graph-structured algorithm for producing sparse representations in hierarchical overcomplete dictionaries. Section III expands the scope of the algorithm to dictionaries composed of molecules. The motivating application in Section II and Section III is the characterization of anisotropy in wide-angle SAR measurements, a hurdle that once cleared, not only relieves inaccuracies in image reconstruction, but also provides a wealth of information for understanding and inference tasks such as automatic target recognition. Section IV discusses parameterized dictionaries and the joint optimization of the expansion coefficients and the atoms themselves. The SAR problem investigated in this section is of extracting object-level information as part of the image formation process from migratory scatterers. Section V introduces the objective of sparsity in multiple domains, focusing primarily on the two domain case, specifically with the Hough transform domain and the SAR measurement domain. The applications in Section IV and Section V take steps towards bridging low-level radar signal processing and higher-level object-based processing in ways not seen in the SAR literature before. Section VI provides a summary of our contributions.

II. Graph-Structured Algorithm for Hierarchical Dictionaries

At the outset, we consider a dictionary that does not decompose into molecules and is known and fixed. We look at a particular type of dictionary with a hierarchical arrangement of atoms that permits the construction of a graph with the atoms as nodes. Then, we describe an algorithm based on hill-climbing search, a heuristic search method also known as guided depth-first search. The final part of the section applies the algorithm to the characterization of anisotropy of a point-scattering center from

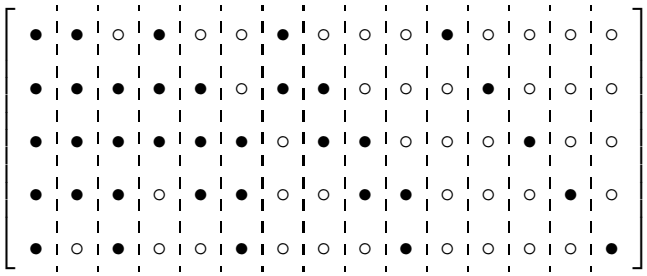


Fig. 1. Illustration of matrix $\mathbf{\Phi}$ for $N = 5$. The solid dots (•) indicate a non-zero value and the empty dots (○) indicate a zero value.

wide-angle SAR measurements.

A. Graph Structure

Oftentimes in overcomplete dictionaries, including for example wavelet packet dictionaries [17], B-spline dictionaries [18], and discrete complex Gabor dictionaries [6], the atoms have a notion of scale and consequently a coarse-scale to fine-scale hierarchy. Translations or rotations are applied at finer scales to create sets of atoms that have a common size but are differentiated in the placement of their region of support; the regions of support may or may not overlap. Some dictionaries are constructed dyadically such that the support of a coarser atom is twice the size of the next finer atom or atoms.

In this work, we consider dictionaries in which the size of the support changes arithmetically rather than geometrically between scales. The matrix $\mathbf{\Phi}$ of such a dictionary for one-dimensional signals of length N is illustrated in Fig. 1; the coarsest atom is the first column and the finest atoms are the N right-most columns. A full set of such atoms with all widths and all shifts has large cardinality ($M = \frac{1}{2}N^2 + \frac{1}{2}N$ atoms), but is appealing for inverse problems because of the possibility that a superposition of very few atoms, perhaps just one, corresponds to a physical phenomenon of interest. As discussed in Section II-C, for SAR anisotropy characterization, the signal $\mathbf{g}$ and atoms $\boldsymbol{\phi}_m$ are such that $\mathbf{g}$ is non-zero for contiguous intervals and zero for other parts of the domain, and is well-represented by few atoms $\boldsymbol{\phi}_m$.

Due to the regular structure of this type of dictionary, we can take the atoms as nodes and arrange them in a graph. As shown in Fig. 2, the coarsest atom is the root node, the finest atoms are leaves, and the graph has N levels. Each node has two children (except for those at the finest level). It is a weakly connected directed acyclic graph, with a topological sort that is exactly the ordering from left to right of the columns in $\mathbf{\Phi}$ illustrated in Fig. 1. As we proceed, we make use of the graph structure, which we term the *molecular graph*, treating the sparse signal representation problem as a graph search.

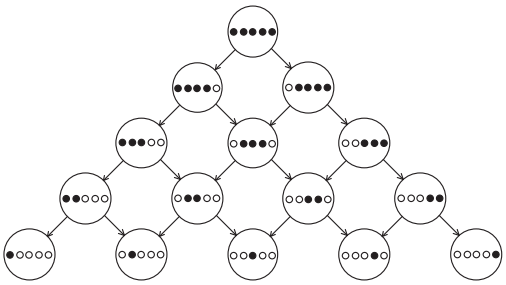


Fig. 2. Illustration of graph structure for overcomplete dictionary, $N = 5$. Coarse-scale atoms are at the top and fine-scale atoms are at the bottom. Different translations are in order from left to right.

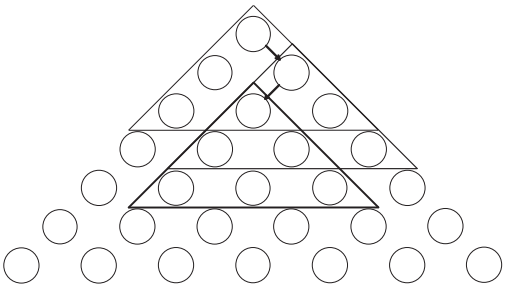


Fig. 3. Illustration of search-based algorithm for $N = 7$, $G = 3$. The guiding graph, a subgraph of the full molecular graph indicated by triangular outline, is moved iteratively to find a sparse representation. The initialization and first two iterations are shown. Molecular graph edges and node labels are omitted.

B. Algorithm Based on Hill-Climbing

As mentioned in Section I, many general methods for obtaining sparse representations give provably optimal solutions (under certain conditions), but require the same computation and memory regardless of whether the dictionary has structure. As an alternative approach for structured dictionaries, we propose a heuristically-based technique with reduced complexity. The idea to have in mind during the exposition of the algorithm is of a small subgraph, given the name *guiding graph*, iteratively moving through an N-level molecular graph, searching for a parsimonious representation. The specifics of the guiding graph, the search strategy, and search steps are presented below. Fig. 3 illustrates the central idea of the algorithm for a small dictionary; in practice, the dictionary and therefore molecular graph are of much larger cardinality.

We assume that $\mathbf{g}$, the signal to be represented or approximated, can be composed using a few atoms whose nodes are close together in the molecular graph under a common parent node. This assumption

is not as restrictive as it may seem: that the signal has a representation with a few atoms is basic for sparsity. Contributing nodes are close together in the graph when the signal is localized in the domain. Prior knowledge can guide the choice of atom shape and standard families of atoms may be used. The assumptions are reasonable for SAR and other applications that lend themselves to such hierarchical structures.

The problem of finding coefficients $\mathbf{a}$ such that $\mathbf{\Phi a}$ equals or well-approximates $\mathbf{g}$ with few non-zero a_m may be reformulated as a search for a node or a few nodes in the molecular graph. In addition to finding nodes, i.e. atoms $\boldsymbol{\phi}_m$ that contribute to the expansion, the corresponding coefficient values a_m must also be determined. Numerous search algorithms exist to find nodes in a graph. Blind search algorithms incorporate no prior information to guide the search. In contrast, heuristic search algorithms have some notion of proximity to the goal available during the search process, allowing the search to proceed along paths that are likely to lead to the goal and reduce average-case running time.

Hill-climbing search is an algorithm similar to depth-first search that makes use of a heuristic. In depth-first search, one path is followed from root to leaf in a predetermined way, such as: "always proceed to the left-most unvisited child." In contrast, hill-climbing search will "proceed to the most promising unvisited child based on a heuristic." In both algorithms, if the goal is not found on the way down and the bottom is reached, there is back-tracking. The approach presented here has hill-climbing search as its foundation.

In standard graph search problems, nodes are labeled and the goal of the search is fixed and specified with a label, e.g. "find node K." Thus the stopping criterion for the search is simply whether the label of the current node matches the goal of the search. Also, there is often a notion of intrinsic distance between nodes that leads to simple search heuristics.

When the sparse signal representation problem is reformulated as a search on an N-level molecular graph, stopping criteria and heuristics are not obvious. One clear desideratum is that calculation of both should require less memory and computation than solving the full problem. The guiding graph, chosen to be a G-level molecular graph, $G \ll N$, with its root at the current node of the search, guides the search by providing search heuristics and stopping conditions.

Intuition about the problem suggests that if the atom or atoms that would contribute in an optimally sparse solution are not included in the guiding graph when solving for coefficients in a sparsity enforcing manner, then the resulting solution will have a non-zero coefficient for the atom most 'similar' to the signal $\mathbf{g}$. In terms of the N-level molecular graph, this suggests that if the optimal sparse representation is far down in the molecular graph, but the problem is solved with a small dictionary containing atoms

from a guiding graph near the top of the molecular graph, then coefficients in the first $G-1$ levels will be zero and one or more coefficients in level G non-zero. In the same vein, if the guiding graph is rooted below the optimal representation, then the root coefficient may be non-zero and the coefficients in levels two through G will be zero. If the guiding graph is such that it contains the optimal atoms, then the corresponding coefficients will be non-zero and the rest of the coefficients zero. This intuition is demonstrated empirically; details are in the appendix.

A simple heuristic for the search based on the coefficient values of the G nodes in level G is apparent from the intuition and experimental validation. Due to the structure of the molecular graph, each node has two children, so the heuristic is used to determine whether to proceed to the left child or the right child. We find the center of mass of the bottom level coefficient magnitudes—the search is guided towards the side that contains the center of mass. A stopping criterion is also apparent: stopping when all of the nodes in level G are zero during the search.

Hill-climbing search finds a single node—a single atom. However, the algorithm that we propose is able to find a small subset of atoms due to the guiding graph. When the stopping criterion is met, i.e. when the finest-scale coefficients are all zero in the sparse solution of the representation problem with atoms from the current guiding graph, then that sparse solution is taken as the solution to the full problem. Consequently, the guiding graph allows a subset of atoms rather than a single atom to be used in the representation.

In summary, the algorithm based on the molecular graph and hill-climbing search is as follows.

(1) Initialization: Let $i \leftarrow 1$ and $\mathbf{\Phi}^{(i)} \leftarrow$ atoms from the top G levels of the molecular graph.

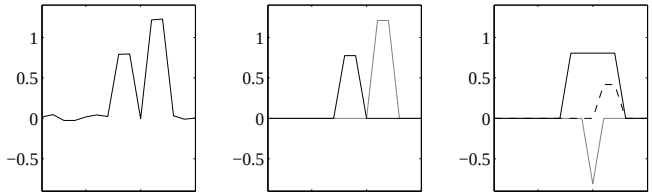
(2) Find a sparse $\mathbf{a}^{(i)}$ such that $\mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$ approximates $\mathbf{g}$.

(3) Calculate weighted sum of bottom row coefficient magnitudes: $\mu \leftarrow \sum_{m=1}^{G} m|a^{(i)}_{\frac{1}{2}G^2-\frac{1}{2}G+m}|$.

(4) If $\mu = 0$ then stop. Otherwise, $i \leftarrow i+1$. If bottom row nodes are leaves of the molecular graph or both children of the guiding graph have been visited before, then $\mathbf{\Phi}^{(i)} \leftarrow$ atoms from the highest unvisited guiding graph.

Else, $\mathbf{\Phi}^{(i)} \leftarrow$ ($\mu < \frac{G+1}{2}\sum_{m=1}^{G}|a^{(i)}_{\frac{1}{2}G^2-\frac{1}{2}G+m}|$ and left child unvisited ? atoms from the left child guiding graph : atoms from the right child guiding graph). Iterate to step (2).

The graph-structured algorithm that we propose is able to produce representations in which there are contributions from atoms that lie within the span of a guiding graph. The approximate nature of the approach is controlled by G; by increasing the size of the guiding graph we may, at the expense of



(a) (b) (c)

Fig. 4. Comparison of graph-structured algorithm and matching pursuit: (a) the signal $\mathbf{g}$; (b) atoms scaled by coefficients in solution obtained with graph-structured algorithm; (c) atoms scaled by coefficients in solution obtained with matching pursuit.

increased complexity, draw from a larger subset of atoms in the solution. The smaller problem with $\mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$ is more tractable than the large problem with $\mathbf{\Phi a}$.

While any of a number of formulations and techniques may be used to solve the smaller problem, here we use a non-convex, ℓ_p, $p < 1$, relaxation, minimizing the cost function:

$$J\left(\mathbf{a}^{(i)}\right) = \left\|\mathbf{g} - \mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}\right\|_2^2 + \alpha \left\|\mathbf{a}^{(i)}\right\|_p^p, \quad p < 1, \tag{1}$$

by a quasi-Newton technique detailed in [19] to obtain a sparse vector of coefficients $\mathbf{a}^{(i)}$. Each step of the quasi-Newton minimization involves solving a set of M_G linear equations, where M_G is the number of atoms in the guiding graph. Direct solution requires $\mathcal{O}(M_G^3)$ computations. However, the particular matrix involved is Hermitian, positive semidefinite, and usually sparse, so the equations may be solved efficiently via iterative algorithms. We use the conjugate gradient method and terminate it when the residual becomes smaller than a threshold.

The parameter α trades data fidelity, the first term, and sparsity, the second term. The choice of α is important practically and is an open area of research. With α too small, the solution coefficient vector $\mathbf{a}^{(i)}$ is not sparse and the heuristic is not meaningful; the guiding graph strays away from good search paths. With α too large, the algorithm incorrectly terminates early with all zero coefficients in the solution. In this work, we choose the parameter subjectively and can usually set it once for a given problem size. We keep α constant for all iterations of the graph-structured algorithm. Generally, solutions in step (2) of the algorithm are not very sensitive to small perturbations of α. It is possible, however, for a small change in α to cause the number of non-zero elements in the solution to change, but such a change in solution is not necessarily accompanied by a change in the heuristic and stopping criterion. In all examples in this paper, the p of the ℓ_p relaxation is 0.1; for the highly redundant dictionary that is employed, a small value of p results in suitable sparsity.

The search-based procedure we have presented is greedy, but not in the same way as matching pursuit and related algorithms [6], [14]–[16]. A commitment is not made to include an atom in the representation until the final iteration when the stopping criterion is met, and also, atoms within a guiding graph are considered jointly. As the guiding graph slides downwards, any subset of fine-scale atoms can start contributing to the representation. This behavior discourages the assignment of a coarse-scale atom to represent what would be better represented using a few close fine-scale atoms. In some later iteration, a matching pursuit-like algorithm includes a fine-scale atom with a negative coefficient to cancel extra energy from the coarse-scale atom included earlier. An example of this behavior is given in Fig. 4. For a particular signal $\mathbf{g}$ and an overcomplete dictionary of boxcar-shaped atoms, solutions are obtained using both the graph-structured algorithm presented in this section and the basic matching pursuit algorithm [6], and compared. Both the graph-structured algorithm and matching pursuit produce solutions that sum to approximate $\mathbf{g}$, but the decomposition of the graph-structured algorithm is more atomic.

The algorithm for dictionaries without molecular decomposition is straightforward; its operation in dictionaries with $L > 1$ molecules, which we discuss in Section III, is more interesting. Before reaching that point however, we illustrate the application of this method to anisotropy characterization in SAR.

C. Application to Wide-Angle SAR

Spotlight-mode SAR has an interpretation as a tomographic observation process [20]. As mentioned in Section I, SAR uses a radar mounted on an aircraft to collect measurements. From one point along the aircraft's flight path, the radar transmits a modulated signal in a certain direction, illuminating a portion of the ground known as the ground patch, and receives back scattered energy, which depends on the characteristics of the ground patch. Radar signals are similarly transmitted and received at many points along the flight path. The radar antenna continually changes its look direction to always illuminate the same ground patch. The geometry of data collection in spotlight-mode SAR is illustrated in Fig. 5. Coordinates on the ground plane x, range, and y, cross-range, are centered in the ground patch. Measurements are taken at equally spaced aspect angles θ as the aircraft traverses the flight path. The ground patch, with radius R, is shaded.

The scattering from the ground patch under observation is manifested as an amplitude scaling and phase shift that can be expressed as a complex number at each point. Thus, scattering from the entire ground patch can be characterized by a complex-valued function of two spatial variables $s(x, y)$, which is referred to as the scattering function. Due to the design of the radar signal and the physics of the observation process, the collection of received signals is not $s(x, y)$ directly. Procedures for obtaining

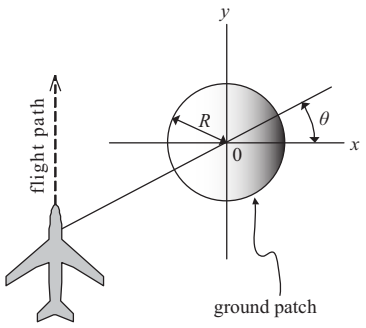


Fig. 5. Ground plane geometry in spotlight-mode SAR.

$s(x, y)$ from the measurements are known as image formation. In wide-angle SAR, measurements come from vastly different viewpoints and consequently, scattering behavior shows dependence on θ, referred to as anisotropy, as well as on (x, y) [21]. For example, a mirror-like flat metal sheet reflects strongly when viewed straight on, but barely reflects from an oblique angle. The relationship between the measurements g, obtained over a finite bandwidth of frequencies and over a range of aspect angles, and the anisotropic scattering function $s(x, y, \theta)$ is given by:

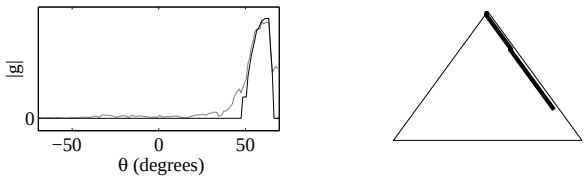
$$g(f, \theta) = \iint\limits_{x^2+y^2 \leq R^2} s\,(x, y, \theta)\, e^{-j\frac{4\pi f}{c}(x\cos\theta + y\sin\theta)} dx\, dy, \tag{2}$$

where c is the speed at which electromagnetic radiation propagates. The set of aspect angles θ is inherently discrete, because pulses are transmitted from a discrete set of points along the flight path. The measurements are sampled in frequency f to allow digital processing. The collection of measurements $g(f, \theta)$ is known as the phase history.

The scattering response of objects such as vehicles on the ground is well-approximated by the superposition of responses from point scattering centers when using frequencies and aperture lengths commonly employed in SAR [22]. The anisotropic scattering from a single point-scatterer takes the form $s(x, y, \theta) = s_0(\theta) \cdot \delta(x - x_0, y - y_0)$ and the measurement model is:

$$g(f, \theta) = s_0(\theta) e^{-j\frac{4\pi f}{c}(x_0\cos\theta + y_0\sin\theta)}. \tag{3}$$

The phenomenon of anisotropy often manifests as large magnitude scattering in a contiguous interval of θ and small, close to zero magnitude scattering elsewhere. Consequently, the dictionary described in Section II-A containing all widths and all shifts of contiguous intervals is well-suited for obtaining



(a) (b)

Fig. 6. Single point-scatterer example: (a) aspect-dependent scattering magnitude measurement (gray line) and solution (black line); (b) search path of graph-structured algorithm.

parsimonious representations of anisotropic scattering. An overcomplete expansion is as follows:

$$g(f,\theta) = \sum_{m=1}^{M} a_m b_m(\theta) e^{-j\frac{4\pi f}{c}(x_0\cos\theta + y_0\sin\theta)}. \tag{4}$$

Atoms are $\phi_m(\theta) = b_m(\theta)e^{-j\frac{4\pi f}{c}(x_0\cos\theta+y_0\sin\theta)}$, where $b_m(\theta)$ are dilations and translations of a common pulse shape. We can use boxcar pulses, Hamming pulses, or other shapes that we expect to encounter. Anisotropy of narrow angular extent comes from physical objects distributed in space and anisotropy of wide angular extent comes from physical objects localized in space; hence the atoms provide a directly meaningful physical interpretation. Appropriately stacking the measurements at different frequencies, we have the sparse signal representation problem with a non-molecular hierarchical dictionary and can obtain solutions using the graph-structured algorithm described above.

D. Anisotropy Characterization of Single Point-Scatterer

We now show anisotropy characterization on SAR phase history measurements from XPatch, a state-of-the-art electromagnetic prediction package, using the graph-structured heuristic method described in this section. A scene containing a single scatterer is measured at $N = 140$ aspect angles spaced one degree apart. The scattering magnitude as a function of aspect angle is the gray line plotted in Fig. 6a. (The line shows the measurements at one particular frequency within the frequency band covered by the radar pulse; frequency dependence is minimal and scattering magnitude at all frequencies is nearly the same.)

Using boxcar pulses for atoms in the overcomplete dictionary and a guiding graph of size $G = 32$, we obtain a sparse approximation for the aspect-dependent scattering given by the black line in Fig. 6a. The search path of the graph-structured algorithm is shown in Fig. 6b. The line indicates the location of the root node of the guiding graph within the full molecular graph. When the stopping criterion is met,

the atom at the root of the guiding graph is of width 34 samples. The finest atoms that contribute to the approximation have width 4 samples. The sparse solution has 14 non-zero coefficients out of a possible $M = 9870$ coefficients for $N = 140$.

From the solution, it is possible to infer physical properties about the object being imaged because thin anisotropy corresponds to objects of large physical size and wide anisotropy to objects of small physical size. Sparsity and the particular overcomplete dictionary are important because they allow this characterization directly by identifying the coarsest non-zero coefficient.

III. Algorithm for Molecular Dictionaries

In the previous section, we described a search-based algorithm for dictionaries whose atoms have a hierarchy, but did not consider dictionaries that have a molecular decomposition into subdictionaries. In this section, the heuristic algorithm is extended by applying it to dictionaries with $L > 1$ molecules, each individually having a hierarchical structure of atoms. We have L coexisting molecular graphs and thus not just one search, but L simultaneous searches. As we shall see, these searches are not performed independently, but rather interact and influence each other. For joint anisotropy characterization and image formation, the L molecules correspond to L different point-scatterers or spatial locations in the ground patch being imaged.

A. Molecular Dictionaries

Overcomplete dictionaries composed of molecules are fairly common, arising in one of two ways. The first is as the union of two or more orthogonal bases and the second, through dependence on some parameter that takes the same value for one subset of atoms, another value for a subset disjoint from the first, and so on.

An example of the first instance is a dictionary made up of the union of an orthogonal basis of lapped cosines and an orthogonal basis of discrete wavelets that provides atoms to represent tonal and transient components in audio signals [11]; the same idea is used for images as well, taking two different bases together as an overcomplete dictionary, one for periodic textures and one for edges [23]. An example in audio of the second instance is molecules whose atoms share a common fundamental frequency [12]. In the radar imaging example in Section III-D, atoms within molecules share a common (x, y) location and different molecules correspond to different spatial locations.

The two types of decompositions into molecules present different properties. In the first type, different molecules aim to represent very different phenomena and are incoherent from each other, whereas in

the second, the molecules correspond to different instances of the same phenomenon and may be highly coherent. In this work, we consider dictionaries whose molecules all have hierarchical structure that permits the construction of molecular graphs, regardless of decomposition type. We use simultaneous searches on all molecular graphs; the difficulty of the problem increases as the coherence between molecules increases.

B. Interacting Searches on Multiple Graphs

The general framework for the graph-structured algorithm with dictionaries containing more than one molecule is the same as for dictionaries without molecules, but with a few key differences. Here the dictionary is of the form $\left[\begin{array}{cccc}\mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_L\end{array}\right]$ with each molecule $\mathbf{\Phi}_l$ having a molecular graph. We assume that all atoms in the dictionary are distinct and that molecules do not share atoms. L guiding graphs iterate through the L molecular graphs, one guiding graph per molecular graph. The vector of coefficients $\mathbf{a}$ also partitions as $\left[\begin{array}{cccc}\mathbf{a}_1^T & \mathbf{a}_2^T & \cdots & \mathbf{a}_L^T\end{array}\right]^T$. L searches are performed simultaneously, as follows.

(1) Initialization: Let $i \leftarrow 1$ and for all molecules $l = 1, \ldots, L$, $\mathbf{\Phi}_l^{(i)} \leftarrow$ atoms from the top G levels of molecular graph l. $\mathbf{\Phi}^{(i)} \leftarrow \left[\begin{array}{ccc}\mathbf{\Phi}_1^{(i)} & \cdots & \mathbf{\Phi}_L^{(i)}\end{array}\right]$.

(2) Find a sparse $\mathbf{a}^{(i)}$ such that $\mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$ approximates $\mathbf{g}$.

(3) For all $l = 1, \ldots, L$, calculate weighted sum of bottom row coefficient magnitudes: $\mu_l \leftarrow \sum_{m=1}^{G} m|a^{(i)}_{l,\frac{1}{2}G^2-\frac{1}{2}G+m}|$.

(4) If $\sum_{l=1}^{L} \mu_l = 0$ then stop. Otherwise, $i \leftarrow i+1$. For all $l = 1, \ldots, L$, if $\mu_l = 0$, then $\mathbf{\Phi}_l^{(i)} \leftarrow \mathbf{\Phi}_l^{(i-1)}$. Else if bottom row nodes are leaves of molecular graph l or both children of guiding graph l have been visited before, then $\mathbf{\Phi}_l^{(i)} \leftarrow$ atoms from the highest unvisited guiding graph. Else, $\mathbf{\Phi}_l^{(i)} \leftarrow (\mu_l < \frac{G+1}{2} \sum_{m=1}^{G} |a^{(i)}_{l,\frac{1}{2}G^2-\frac{1}{2}G+m}|$ and left child unvisited ? atoms from the left child guiding graph : atoms from the right child guiding graph). Iterate to step (2).

Let us emphasize that although the L searches are performed simultaneously, they are not performed independently. The searches are coupled because the inverse problem is solved jointly for all molecules on every iteration; contributions to the reconstruction of $\mathbf{g}$ from all of the molecules interact. There is no notion of molecules when solving the smaller inverse problem $\mathbf{g} \approx \mathbf{\Phi}^{(i)}\mathbf{a}^{(i)}$. The molecular structure only comes into play after $\mathbf{a}^{(i)}$ has been solved, and the heuristics, stopping criteria, and $\mathbf{\Phi}_l^{(i)}$ updates are to be calculated. Since we consider all molecules jointly rather than one at a time as matching pursuit-like algorithms would do, we see similar advantages of the formulation presented here to those seen in Fig. 4 for the single molecule case.

The dictionary used in calculating the heuristic and stopping criterion has $\mathcal{O}(G^2)$ atoms per molecule and $\mathcal{O}(G^2L)$ atoms for L molecules, instead of $\mathcal{O}(N^2L)$ atoms used if one were to solve the full inverse problem. However, the graph-structured algorithm requires $\mathcal{O}(N^2)$ iterations, whereas solving the full inverse problem at once requires just one iteration. G is a small constant that is fairly independent of N. For joint anisotropy characterization and image formation, L and N may be in the thousands. The realistic example given in Section III-E would have eighty-nine million atoms if the full problem were solved at once, but the graph-structured approach allows us to only consider a small fraction of them. In the following section, we discuss variations to the algorithm presented thus far that further reduce computation or memory requirements.

C. Algorithmic Variations

The graph-structured algorithm described thus far uses the full hill-climbing search including back-tracking, taking steps of single levels per iteration based on a heuristic employing guiding graphs taking the form of G-level molecular graphs. A number of variations to the basic algorithm may be made; we present a few here, but many others are also possible. Algorithms that use one variation or use a few variations together can be used to solve the sparse signal representation problem. Depending on the size of the problem and the requirements of the application, one algorithm can be selected from this suite of possible algorithms.

1) Hill-climbing without back-tracking: Hill-climbing search always finds the goal node because of back-tracking. In a first variation, we limit the search to disallow back-tracking. This reduces the iterations from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, but results in a greedier method. If, on a particular example, hill-climbing with back-tracking were to terminate on the first pass down molecular graphs before reaching leaves, then the same operation would be achieved whether the original algorithm or the variation were used. In practice, we often observe termination on the first downward search, including in the example seen in Section II-D and an example presented below in Section III-D.

2) Modified molecular graph: Molecular graphs are structured such that in hill-climbing without back-tracking, one wrong step eliminates many nearby nodes and paths because each node has only two children. The graph may be modified to increase the number of children per node to four for interior nodes and three for nodes on the edges of the graph, consequently not disallowing as many nodes and paths per search step.

A modified heuristic to go along with this modified graph is to use the G coefficients in level G of the guiding graph as before, but instead of determining whether the center of mass of the coefficient

magnitudes is in the left half or the right half, determining which quarter it is in. If the left-most quadrant, then the search proceeds to the node in the next level that is two to the left of the current node. If the middle left quadrant, then the next node is one to the left in the next level, and so on. With these additional edges, search without back-tracking is less greedy with no additional cost, since calculating this modified heuristic is no more costly than calculating the original heuristic.

3) Modified guiding graph and larger steps: The guiding graph need not be a G-level molecular graph; for example, the graph may be thinned and include the top node, nodes in level G, and nodes in a few intermediate levels rather than all intermediate levels, further reducing the number of atoms in $\mathbf{\Phi}^{(i)}$. These atoms are sufficient for calculating the heuristic and stopping condition. Also, searches may take larger steps than moving guiding graphs down just one level per iteration.

4) Removal of stopped molecules: The graph-structured algorithm reduces the number of atoms per molecule from $\mathcal{O}(N^2)$ to $\mathcal{O}(G^2)$, but does nothing to reduce the number of molecules L. A further variation to the hill-climbing search without back-tracking may be introduced that reduces the average-case dependence of the number of atoms on L. It is observed that, despite interactions among contributions from different molecules, once the search on a particular molecule stops it does not restart in general, but may occasionally restart after a few iterations. It is thus natural to consider fixing the contribution from a molecule upon finding its coefficients.

In the algorithm, this implies that once the stopping criterion is met at molecule l, the signal $\mathbf{g}$ is updated to be $\mathbf{g}' = \mathbf{g} - \mathbf{\Phi}_l \mathbf{a}_l$, and $\mathbf{\Phi}_l$ is removed from $\mathbf{\Phi}$, thereby reducing the number of atoms in $\mathbf{\Phi}$. We perform the removal some iterations after the stopping criterion is met and maintained to allow for a possible restart. This variation, though distinct, has some similarity to matching pursuit.

D. Joint Anisotropy Characterization and Image Formation

The problem of joint anisotropy characterization and image formation in wide-angle SAR takes the problem of characterizing anisotropy of a single point-scatterer seen in Section II and extends it to doing so for all points in the ground patch. In other words, whereas standard image formation attempts to recover $s(x, y)$ assuming no dependence on θ, we aim to recover $s(x, y, \theta)$.

The observation model from more than one point is a superposition of terms like (3):

$$g(f, \theta) = \sum_{l=1}^{L} s_l(\theta) e^{-j \frac{4\pi f}{c}(x_l \cos\theta + y_l \sin\theta)}. \tag{5}$$

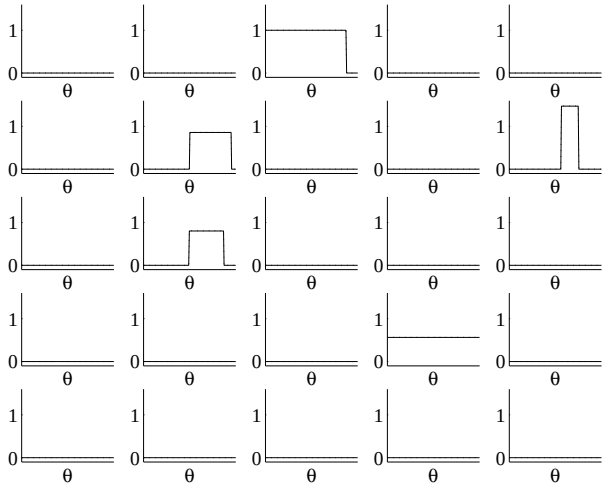


Fig. 7. Scattering magnitude at each spatial location.

The observation model (5) lends itself to an overcomplete expansion of the form:

$$g(f,\theta) = \sum_{l=1}^{L} \sum_{m=1}^{M} a_{lm} b_m(\theta) e^{-j\frac{4\pi f}{c}(x_l \cos\theta + y_l \sin\theta)}, \tag{6}$$

in a similar manner to the single point-scatterer case. Here the dictionary is naturally decomposed into molecules, with each molecule corresponding to a different spatial location (x_l, y_l). We can thus use the methods described above for joint anisotropy characterization and image formation [24].

When performing joint anisotropy characterization and image formation, a grid of pixels in the image to be reconstructed or points of interest identified through preprocessing may be used as the spatial locations (x_l, y_l). We now present an example with $L = 25$ spatial locations in a five by five grid, with rows and columns spaced one meter apart. Unlike Section II-D which uses XPatch data, the synthetic data in this example is matched to the dictionary for illustrative purposes.

This example has $N = 160$ aspect angles equally spaced over a 110° aperture. Fig. 7 shows the scattering magnitude at each of the 25 spatial locations arranged as in an image; five of the spatial locations contain boxcar-shaped scattering and the other twenty do not have scatterers. The coherent sum of the scatterers is the phase history measurement $g(f,\theta)$, plotted in Fig. 8 for one frequency.

We recover a signal representation from the phase history measurements using the basic algorithm for molecular dictionaries with guiding graphs of size $G = 8$ and boxcar-shaped atoms. The search paths for the different locations are shown in Fig. 9. The overcomplete dictionary for $N = 160$, $L = 25$ has $322,000$ atoms. In the solution of the sparse signal representation problem, contributions come from

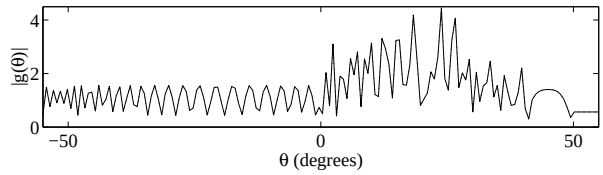


Fig. 8. Phase history measurement magnitude.

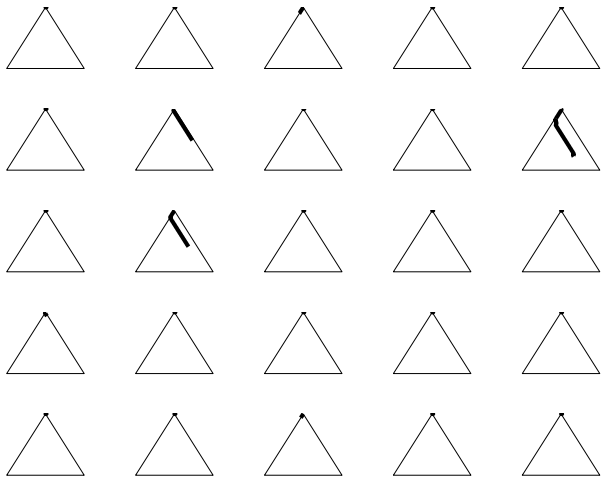


Fig. 9. Search paths of basic algorithm for molecular dictionaries.

exactly the five atoms used to generate the synthetic data; the coefficient values are also recovered. If the solution were to be overlaid on Fig. 7 and Fig. 8, it would not be distinguishable. Looking at the search paths, despite not containing scatterers, a couple of molecules initially iterate nonetheless, but in the end correctly give all zero coefficients. This effect is a result of the interaction between different molecules. The algorithm operates correctly on this synthetic example; a larger example on XPatch data is given below and others may be found in [24], [25].

E. Approaches to Wide-Angle SAR and a Realistic Example

To conclude this section, a large, realistic example with XPatch data is presented. The scene being imaged contains a backhoe-loader, illustrated in Fig. 10a [26]; measurements are taken at $N = 1541$ equally-spaced angles over an aperture ranging from -10° to 100°. $L = 75$ spatial locations are identified from a composite subaperture image using the method of [27], for which anisotropy is then jointly characterized. The full dictionary for this example has $M = 89,108,325$ atoms. We apply the graph-

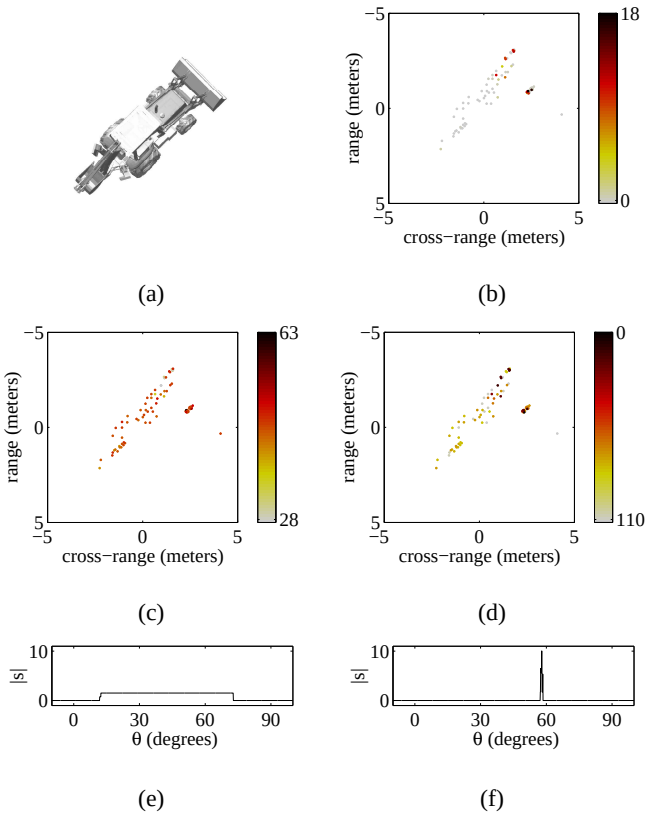


Fig. 10. Backhoe-loader example: (a) illustration of the scene; $L = 75$ spatial locations of interest shaded according to (b) maximum magnitude, (c) center angle of anisotropy (degrees), and (d) angular extent of anisotropy (degrees) in solution; (e)-(f) aspect-dependent scattering solution for two spatial locations.

structured algorithm with all of the variations listed in Section III-C to the problem and obtain seventy-five functions of aspect angle.

The magnitudes of two of these functions are plotted in Fig. 10e and Fig. 10f. In order to provide spatial visualization of the scattering behavior, the magnitude, center angle of anisotropy, and angular extent of anisotropy for each of the spatial locations is indicated by the shading of the markers in Fig. 10b-d.

In the magnitude visualization, light gray is small magnitude and black is high magnitude. Points corresponding to the front bucket of the backhoe-loader have high magnitude. In the visualization of center angle, the left side of the front bucket has responses closer to -10° (light gray) and the right side of the front bucket has responses closer to $+100^\circ$ (black). In the angular extent visualization, it can be seen that narrow and wide anisotropy is distributed, but the points on the front bucket with high magnitude also have narrow extent. Overall, one can note from the visualizations that the front bucket

flashes on its two sides and the other parts of the backhoe-loader have scattering with smaller magnitude and wider anisotropy.

Through joint anisotropy characterization and image formation, we obtain much more information than a simple image would provide, namely an entire dimension of aspect-dependence. The reflectivities of scatterers with narrow angular persistence, which are lost in Fourier-based image formation, are obtained. The formulation presented here solves for the anisotropy of all spatial locations within one system of equations, taking interactions among scattering centers into account.

The formulation is more flexible than parametric methods for anisotropy characterization such as [28], [29]. Also, solutions have more detail in aspect angle than subaperture methods such as [30]–[33], in which the measurements are divided into smaller segments covering only parts of the wide-angle aperture. Consequently, using the method presented here, angular persistence information can be extracted as in Fig. 10d, which is not possible from subaperture methods. Also, since data from the full wide-angle aperture is used here throughout, cross-range resolution is not reduced as it is with subaperture methods.

IV. Dictionary Refinement

In Section II and Section III, the dictionary $\mathbf{\Phi}$ is known and fixed, but this need not always be the case. A more ambitious goal is to find the best dictionary under some criteria and an optimally sparse representation jointly. The idea of learning overcomplete dictionaries has been applied in the case that one has many examples of signals $\mathbf{g}$, much more than the number of atoms in $\mathbf{\Phi}$, and a dictionary is to be determined that is able to most sparsely represent all of the signals, usually for compression tasks [34], [35]. In inverse problems, where the interest is in extracting physical meaning from the obtained sparse representation for each input signal $\mathbf{g}$, rather than compression of an entire signal class, it is of interest to look at the best dictionary for each input rather than the best dictionary to represent an entire set of training signals. At this point, one could conclude that a dictionary with $\phi_1 = \mathbf{g}$ is optimal and stop. However, we would like to consider dictionaries derived from a parameterized observation model and only consider parameterized atoms, not arbitrary atoms. In this section we propose and demonstrate a formulation for joint optimization to achieve a sparse coefficient vector and optimal parameter settings for a dictionary with parameterized atoms or molecules.

A. Joint Dictionary and Sparse Coefficient Optimization

We begin with a dictionary whose atoms depend on a set of parameters $\boldsymbol{\eta}$; each parameter may or may not be shared by atoms or molecules. Furthermore, we consider the ℓ_p relaxation to the sparse

signal representation problem mentioned in Section II-B [19]. The optimization problem at hand then is to minimize the following cost function:

$$J(\mathbf{a}, \boldsymbol{\eta}) = \|\mathbf{g} - \boldsymbol{\Phi}(\boldsymbol{\eta})\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p, \quad p < 1, \tag{7}$$

jointly determining a dictionary $\boldsymbol{\Phi}(\boldsymbol{\eta})$ and coefficients $\mathbf{a}$.

To carry out the joint minimization, we take a coordinate descent approach, alternately optimizing over the coefficients and dictionary parameters. The two optimizations are:

$$\mathbf{a}^{(t+1)} = \arg\min_{\mathbf{a}} \left\| \mathbf{g} - \boldsymbol{\Phi}\left(\boldsymbol{\eta}^{(t)}\right)\mathbf{a} \right\|_2^2 + \alpha \left\|\mathbf{a}\right\|_p^p. \tag{8}$$

$$\begin{aligned} \boldsymbol{\eta}^{(t+1)} &= \arg\min_{\boldsymbol{\eta}} \left\| \mathbf{g} - \boldsymbol{\Phi}\left(\boldsymbol{\eta}\right)\mathbf{a}^{(t+1)} \right\|_2^2 + \alpha \left\|\mathbf{a}^{(t+1)}\right\|_p^p \\ &= \arg\min_{\boldsymbol{\eta}} \left\| \mathbf{g} - \boldsymbol{\Phi}\left(\boldsymbol{\eta}\right)\mathbf{a}^{(t+1)} \right\|_2^2. \end{aligned} \tag{9}$$

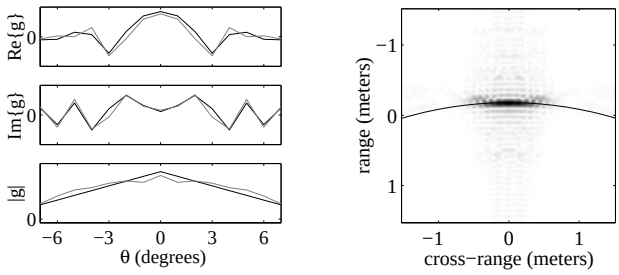
The application will guide the particular initialization for $\boldsymbol{\eta}$. The non-convex minimization (8) may be performed using the graph-structured algorithms of Section II and Section III, or using quasi-Newton optimization [19].

The minimization (9) may be recognized as nonlinear least-squares; many techniques exist in the literature including the trust-region reflective Newton algorithm that we use [36]. Linear inequality constraints on the parameter vector $\boldsymbol{\eta}$ may be handled within this framework. Termination of the procedure is upon the change in $\boldsymbol{\eta}$ falling below a small constant.

B. Characterization of Migratory Scattering Centers

We demonstrate joint dictionary parameter and sparse representation optimization on the characterization of a phenomenon in wide-angle SAR imaging different from anisotropy. Certain scattering mechanisms migrate as a function of aspect angle θ in wide-angle imaging [37], [38]. Migration occurs when radar signals bounce back from the closest surface of a physical object, but the closest surface of the object is different from different viewing angles; the physical object is not really moving, but appears to move in the measurement domain. By accounting for this effect in solving the inverse problem, a physically meaningful, parsimonious description can be extracted.

For example, considering a circular cylinder, the point of reflection on the surface closest to the radar can be parameterized as a function of θ around the center of the cylinder (x_c, y_c) using the radius of the cylinder η. When $\theta = 0$, the scatterer appears to be at $(x_c - \eta, y_c)$, which we define as $(\bar{x}, \bar{y})$. The



(a) (b)

Fig. 11. Tophat example: (a) aspect-dependent scattering measurement (gray line) and solution (black line); (b) conventionally formed image with migration solution overlaid.

observation model for migratory point scatterers is:

$$g\left(f,\theta\right)=\sum_{l=1}^{L}s_l(\theta)e^{-j\frac{4\pi f}{c}((\bar{x}_l+\eta_l)\cos\theta+\bar{y}_l\sin\theta-\eta_l)}. \tag{10}$$

A dictionary expansion for the observation model is:

$$g\left(f,\theta\right)=\sum_{l=1}^{L}\sum_{m=1}^{M}a_{lm}b_m(\theta)e^{-j\frac{4\pi f}{c}((\bar{x}_l+\eta_l)\cos\theta+\bar{y}_l\sin\theta-\eta_l)}. \tag{11}$$

In this instance, the atoms are parameterized by the radius η, and moreover, all atoms in molecule l share a common radius η_l; hence $\boldsymbol{\eta}$ is an L-vector of parameters. The inverse problem is to jointly recover the anisotropy and radius of migration of all scatterers in the ground patch.

The radius is constrained to be non-negative, i.e. $\boldsymbol{\eta}\geq\mathbf{0}$. Most scatterers are not migratory, and thus we initialize $\boldsymbol{\eta}$ with all zeroes. Often in practice, the coefficient vector $\mathbf{a}$ retains its sparsity structure on every iteration because even for $\boldsymbol{\eta}=\mathbf{0}$, characterized anisotropy may be close to correct, or at least have the correct support. The procedure may be envisioned as simultaneously inflating L balloons.

As an example, we look at data from XPatch of a scene containing a tophat that exhibits circular migratory scattering. In the aperture with $N=15$ aspect angles spaced one degree apart, the tophat also has anisotropy, as shown in Fig. 11a. The magnitudes as well as the real and imaginary parts of the measurements are shown, as migratory scattering affects phase, not magnitude. An image of the scene formed using the polar format algorithm, a conventional method based on the inverse Fourier transform, is shown in Fig. 11b.

After identifying the spatial location with largest magnitude in the conventionally formed image, the coordinate descent described in this section is applied with $L=1$. A raised triangle shape is used for the atoms. The solution has radius 5.314 meters and anisotropy as plotted in Fig. 11a. The circular migration

of radius 5.314 meters is overlaid on and matches well with the conventional image in Fig. 11b. Coordinate descent to jointly optimize over radius and anisotropy is effective with realistic data seen here, and with several scatterers in a scene ($L > 1$), see [25]. By allowing for a non-zero radius, image formation is not simply pixel-based but more region-based. Although point scatterers can be equated to spatial locations, if information about migration is considered, the scatterer is more of an object-level construct.

We have looked at characterizing the migration of scatterers when the migration is circular in shape. Circles are an important subset of migratory scattering because many man-made objects contain scatterers with circular migration. However, any shape defined by a radius function $\eta(\theta)$ around a center is easily expressed in the observation model:

$$g\left(f,\theta\right) = \sum_{l=1}^{L} s_l(\theta) e^{-j\frac{4\pi f}{c}\left(\left(\bar{x}_l + \eta_l(0)\right)\cos\theta + \bar{y}_l \sin\theta - \eta_l(\theta)\right)}. \tag{12}$$

Under this model, η_l is not constant across all angles, so a length L vector of parameters is not sufficient. One option is to take a functional form for $\eta_l(\theta)$ with more degrees of freedom than just a constant function, such as a polynomial, and lengthen the parameter vector $\boldsymbol{\eta}$. Another option is to locally, i.e. in small segments of θ, approximate $\eta_l(\theta)$ with pieces of circles [25]. The phenomenon of migratory scattering, which has rarely been explored in the literature, is a source of information that can be mined for details about object shape and size.

V. Simultaneous Sparsity in Multiple Domains

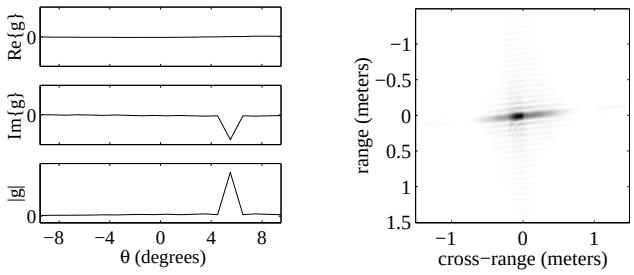
In the previous sections, we use an overcomplete dictionary $\mathbf{\Phi}$ to represent a signal $\mathbf{g}$, assuming that a sparse representation exists and then finding it. Our assumption in those sections is that $\mathbf{g}$ is sparse in the domain of the atoms. In this section, reverting to a known and fixed dictionary, we look at signals that are sparse in the domain of that known and fixed dictionary, but are also sparse in one or more other domains. The goal is to develop a formulation that recovers parsimonious representations, semantically interpretable in the case of inverse problems, making use of sparsity in all domains. Note that in the end, solutions will still be representations in terms of the atoms of the dictionary.

A. Additional Sparsity Terms

For sparsity in the domain of the dictionary, the ℓ_p relaxation as an objective function is:

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \alpha\|\mathbf{a}\|_p^p, \quad p < 1. \tag{13}$$

Let us assume that $\mathbf{g}$ is also sparse in a transformed domain and that that sparsity is to be exploited as well. First note that taking an orthonormal transformation of both the signal $\mathbf{g}$ and dictionary $\mathbf{\Phi}$ does not



(a) (b)

Fig. 12. Glint example: (a) aspect-dependent scattering measurement; (b) conventionally formed image.

change the cost function. Also, the dictionary $\mathbf{\Phi}$ is fixed; consequently, we keep the data fidelity term as is, and append additional sparsity terms.

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi}\mathbf{a}\|_2^2 + \sum_i \alpha_i \|\mathbf{F}_i(\mathbf{a})\|_p^p. \tag{14}$$

The functions $\mathbf{F}_i(\mathbf{a})$ return vectors related to the domain in which sparsity is to be favored. For the domain of the dictionary atoms, $\mathbf{F}_i$ is an identity operation. For domains that are transformations of the original domain, $\mathbf{F}_i$ is constructed as follows.

The operation $\mathbf{F}_i$ is the composition of three simpler operations. First, since the coefficients themselves have no particular meaning until paired with their corresponding atoms, initially $\mathbf{F}_i$ takes the coefficients through the atoms ϕ_m. Thereafter, the second operation is transformation to another domain. Finally, further operations in the transformed domain may follow. If all $\mathbf{F}_i(\mathbf{a})$ are linear, i.e. matrix-vector products, then the cost function may be optimized using quasi-Newton optimization [19] or the graph-structured algorithm using quasi-Newton optimization in each iteration. A concrete application given below constructs such $\mathbf{F}_i$.

B. Parsimonious Representation Recovery of Glint Anisotropy

Scattering behavior known as glint is produced by long, flat metal plates and is not migratory, has very narrow anisotropy, and corresponds to a line segment in the x-y domain oriented at the same angle as the center angle of the anisotropy. Fig. 12a shows aspect-dependent scattering of glint anisotropy from XPatch data and Fig. 12b shows a conventionally formed image. A parsimonious representation ought to explain scattering with a single scattering center, not with a collection of scatterers located on the line segment. We apply the formulation (14) both to favor sparsity among atoms and to favor sparsity along lines [38].

To favor sparsity among atoms, $\mathbf{F}_1$ is the identity. We now find a domain in which sparsity along a line can be favored. The normal parameter space of the Hough transform, the ρ-θ plane, and image space, the x-y plane, are related by the property that a set of points lying on the same line in image space corresponds to a set of sinusoids that intersect at a common point in parameter space [39]. Thus sparsity among scatterers in individual ρ-θ cells achieves the goal of sparsity among points on a line.

In [40], a Hough space sparsifying regularization approach is employed to enhance and detect straight lines in positive real-valued images by imposing sparsity when taking the image data to the ρ-θ plane. Parameter space cells with small counts are suppressed and cells with large counts are enhanced; thus, non-line features are suppressed and line features are enhanced in image space. The goals in our work are different and consequently, the sparsity terms are of a different flavor as well.

The range profile domain in SAR, a one-dimensional inverse Fourier transform of the phase history measurement domain, is equivalent to the parameter space of the Hough transform. It follows that for sparsity among scatterers in cell (ρ_k, θ_n), a sparsity term of the form $|||\mathbf{L}_{kn}\mathbf{a}|||_p^p$ is used, where $\mathbf{L}_{kn}$ is a linear operator that is a composition of a block-diagonal version of the dictionary to bring the coefficients to the phase history domain, a discrete Fourier transform operator to go to the range profile domain, and a selection operator to select cell (ρ_k, θ_n). The resulting vector $\mathbf{L}_{kn}\mathbf{a}$ is of length L. Favoring sparsity in all range profile cells, the overall sparsity cost function is:

$$J(\mathbf{a}) = \|\mathbf{g} - \mathbf{\Phi a}\|_2^2 + \alpha_1 \|\mathbf{a}\|_p^p + \alpha_2 \sum_{k=1}^{K} \sum_{n=1}^{N} |||\mathbf{L}_{kn}\mathbf{a}|||_p^p. \tag{15}$$

The parameters α_1 and α_2 control the influence of the two sparsity terms. When $\alpha_2 = 0$, the cost function reduces to (13).

We solve the inverse problem with $L = 24$ pixels of interest identified by having large magnitude in the conventional image Fig. 12b. These 24 pixels are along a diagonal line more or less. The measurements are at $N = 20$ aspect angles over a 19° aperture with the glint at 5.5°.

Let us define two counts related to the sparsity of the solution and look at their behavior as α_1 and α_2 are varied. We define L_{A} as the number of molecules out of the possible $L = 24$ that have at least one non-zero coefficient in the solution. Also, M_{A} is defined as the average number of non-zero coefficients per molecule in those molecules that have at least one non-zero coefficient. The maximum possible value of M_{A} is M, which is 210 for $N = 20$. When L_{A} is zero, M_{A} is defined to be zero. Solutions are obtained using the quasi-Newton method to minimize (15).

The two counts L_{A} and M_{A} are given in Table I for different values of α_1 and α_2. First, it should be noted that when α_1 and α_2 get too large, all of the coefficients go to zero. The main thing to take

TABLE I

L_{A} AND M_{A} AS A FUNCTION OF THE PARAMETERS α_1 AND α_2.

$(L_{\mathrm{A}}, M_{\mathrm{A}})$	$\alpha_1 = 0$	$\alpha_1 = 10$	$\alpha_1 = 20$	$\alpha_1 = 30$	$\alpha_1 = 40$
$\alpha_2 = 0$	$(24, 39)$	$(24, 1)$	$(24, 1)$	$(24, 1)$	$(0, 0)$
$\alpha_2 = 5$	$(2, 39)$	$(3, 1)$	$(3, 1)$	$(3, 1)$	$(0, 0)$
$\alpha_2 = 10$	$(1, 39)$	$(2, 1)$	$(4, 1)$	$(2, 1)$	$(0, 0)$
$\alpha_2 = 15$	$(1, 39)$	$(4, 1)$	$(3, 1)$	$(1, 1)^*$	$(0, 0)$
$\alpha_2 = 20$	$(1, 39)$	$(4, 1)$	$(3, 1)$	$(1, 1)^*$	$(0, 0)$
$\alpha_2 = 25$	$(1, 39)$	$(4, 1)$	$(2, 1)$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 30$	$(1, 39)$	$(2, 1)$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 35$	$(1, 39)$	$(1, 1)^*$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$
$\alpha_2 = 40$	$(1, 39)$	$(1, 1)^*$	$(1, 1)^*$	$(0, 0)$	$(0, 0)$

note of is that when $\alpha_2 = 0$, $L_{\mathrm{A}} = 24$, i.e. all spatial locations provide contributions to the solution, but as α_2 increases, sparsity along a line is a greater influence and the number of contributing spatial locations decreases to one. Sparsity among atoms is not enough for the solution on XPatch data to be parsimonious in the number of spatial locations, sparsity along a line is also required.

It can be seen that when $\alpha_1 = 0$, 39 atoms per spatial location contribute, not very sparse. For larger α_1, just one atom per spatial location contributes. Considering the behavior of L_{A} and M_{A} together, we note that the two sparsity terms are fairly orthogonal; the main effect of sparsity among atoms is on the number of atoms per spatial location and the main effect of sparsity along a line is on the number of spatial locations, as per the design objective.

A sparse and physically interpretable approximation ought to assign all of the scattering to the leaf atom at 5.5° of a single spatial location. Such a solution with one non-zero coefficient is recovered for the (α_1, α_2) pairs marked with an asterisk in Table I.

Through the example it has been seen that both types of sparsity are necessary to recover a solution that represents the scattering as coming from a single point and with very thin anisotropy explained by a single atom. With this representation, spatial properties about the object being imaged, such as orientation and physical extent, may be inferred. Although the same object-level inferences could have been made with $\alpha_2 = 0$, in that case, L such objects would be indicated rather than one, which does not make physical sense. Points have more meaning than just pixels with aspect-dependent amplitudes.

VI. Conclusion

We looked at methods of obtaining sparse signal representations and approximations from overcomplete dictionaries with hierarchical structures within subdictionaries, focusing on the context of coherent inverse problems with physically interpretable dictionary elements. We developed a heuristic method of solution for such problems that takes advantage of the structure by relating the problem to search on graphs. We also took a step back from the classic sparse signal representation problem to consider dictionary refinement as well as obtaining solutions simultaneously sparse in multiple domains. Under dictionary refinement, a coordinate descent approach was developed to jointly optimize parameterized atoms and coefficients, whereas under simultaneous sparsity, an extended sparsifying cost function was minimized.

The methods were demonstrated on various facets of wide-angle SAR, but are general enough to transfer to other applications with appropriate dictionaries. In the SAR context, starting from the same low-level measurements used by conventional image formation techniques, we have taken a step farther in scene understanding while also taking into account phenomena such as anisotropy that cause inaccuracies in conventional methods. We have started to move away from a pixel representation to more of an object-level representation through the use of a physically meaningful dictionary.

Appendix

Two experimental results are given as empirical validation for the search heuristic and stopping criterion described in Section II-B. We show that solutions from subdictionaries do in fact have non-zero coefficients for atoms most 'similar' to the signal $\mathbf{g}$, particularly when $\mathbf{g}$ is not contained in the subdictionary. For the experiments, the molecular graph has $N = 400$ levels and the guiding graph has $G = 8$ levels. Keeping the guiding graph fixed within the molecular graph, the behavior of the solution $\mathbf{a}$ is observed as the signal $\mathbf{g}$ is varied. Quasi-Newton optimization is used to obtain the sparse solution coefficients $\mathbf{a}$.

In the first experiment, with results in Fig. 13, the guiding graph is fixed with root at the left-most node of level 200 in the molecular graph. The true signal $\mathbf{g}$ is varied from coarse to fine support. In terms of the molecular graph, the true coefficient is varied, starting at the root node, through all nodes along the left edge of the graph, to the left-most node of level 400. In the plot, the row in the molecular graph which contains $\mathbf{g}$ is plotted on the horizontal axis. The magnitudes of the 36 coefficients in $\mathbf{a}$ are indicated by shading (white is zero); each horizontal strip is for one of the coefficients. Most coefficients are zero for all $\mathbf{g}$ due to sparsity. In the regime where the guiding graph is below the true coefficient, the coefficient of the guiding graph root node is non-zero. In the regime where the guiding graph covers the

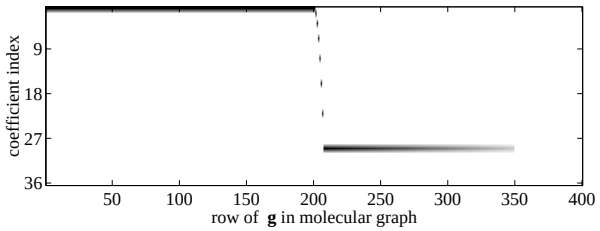


Fig. 13. Coefficient magnitudes in 8-level guiding graph as signal **g** is varied from coarse to fine.

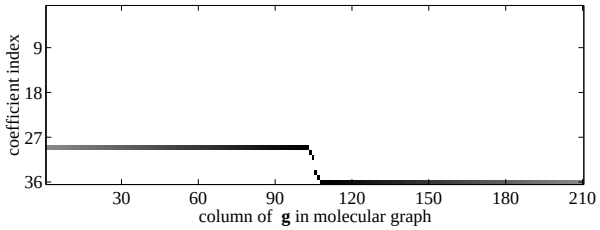


Fig. 14. Coefficient magnitudes in 8-level guiding graph as signal **g** is shifted from left to right.

true coefficient, the correct coefficient is non-zero. When the guiding graph is above the true coefficient, the coefficient of the bottom left node, the node in the last level closest to the truth, is non-zero and others are zero. It should be noted that the influence of the finest signals does not reach up to make any guiding graph node coefficients non-zero (a consequence of regularization).

In the experiment yielding the results of Fig. 14, the guiding graph is fixed with root at the center node of level 200 instead of the left-most node. The true node is varied from left to right across the molecular graph at level 210, three levels below the bottom of the guiding graph. This figure is organized in the same manner as Fig. 13, but the horizontal axis indicates the column of **g** in the molecular graph. From these results, first it is apparent that only coefficients in the last level of the guiding graph are non-zero, reconfirming results from the previous experiment. Second, it can be seen that when the truth is to the left of the guiding graph, the left-most node of level G is non-zero. Similarly, when the truth is to the right, the right node is non-zero; when the truth is underneath the 8-level graph, nodes in the interior of the last level are non-zero.

ACKNOWLEDGMENT

XPatch data provided by R. Bhalla. Polar format algorithm implementation provided by R. L. Moses.

REFERENCES

[1] D. M. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[2] J.-J. Fuchs and B. Delyon, "Minimum L_1-norm reconstruction function for oversampled signals: Applications to time-delay estimation," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1666–1673, July 2000.

[3] M. Çetin, W. C. Karl, and A. S. Willsky, "Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging," *Optical Engineering*, vol. 45, no. 1, p. 017003, Jan. 2006.

[4] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 48, no. 3, pp. 600–616, Mar. 1997.

[5] B. D. Jeffs and M. Gunsay, "Restoration of blurred star field images by maximally sparse optimization," *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 202–211, Apr. 1993.

[6] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[8] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. National Acad. Sciences*, vol. 100, no. 5, pp. 2197–2202, Mar. 4 2003.

[9] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, Montréal, May 2004, pp. 793–796.

[10] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[11] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio Speech Language Processing*, vol. 14, no. 5, pp. 1808–1816, Sept. 2006.

[12] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[13] L. Granai and P. Vandergheynst, "Sparse decomposition over multi-component redundant dictionaries," in *Proc. IEEE Multimedia Signal Processing Workshop*, Siena, Italy, Sept. 2004, pp. 494–497.

[14] A. Shoa and S. Shirani, "Tree structure search for matching pursuit," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Genoa, Italy, Sept. 2005, pp. 908–911.

[15] C. La and M. N. Do, "Signal reconstruction using sparse tree representations," *Proc. SPIE*, vol. 5914, p. 591410W, Sept. 2005.

[16] P. Jost, P. Vandergheynst, and P. Frossard, "Tree-based pursuit: Algorithm and properties," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4685–4697, Dec. 2006.

[17] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wave packets," in *Proc. Int. Conf. Wavelets*, Marseille, France, May 1989.

[18] S. Jaggi, W. C. Karl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *J. Appl. Comp. Harmonic Analysis*, vol. 5, pp. 428–449, 1998.

[19] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[20] D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE*, vol. 71, no. 8, pp. 917–925, Aug. 1983.

[21] R. L. Moses, L. C. Potter, and M. Çetin, "Wide angle SAR imaging," *Proc. SPIE*, vol. 5427, pp. 164–175, Apr. 2004.

[22] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, no. 2, pp. 116–130, Feb. 1962.

[23] F. G. Meyer, A. Z. Averbuch, and R. R. Coifman, "Multilayered image representation: Application to image compression," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1072–1080, Sept. 2002.

[24] K. R. Varshney, M. Çetin, J. W. Fisher, III, and A. S. Willsky, "Joint image formation and anisotropy characterization in wide-angle SAR," *Proc. SPIE*, vol. 6237, p. 62370D, Apr. 2006.

[25] K. R. Varshney, "Joint anisotropy characterization and image formation in wide-angle synthetic aperture radar," Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2006.

[26] "Backhoe data dome and Visual-D challenge problem," Available at Air Force Research Laboratory Sensor Data Management System (https://www.sdms.afrl.af.mil/main.php), 2004.

[27] M. A. Koets and R. L. Moses, "Feature extraction using attributed scattering center models on SAR imagery," *Proc. SPIE*, vol. 3721, pp. 104–115, Apr. 1999.

[28] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Processing*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[29] L. C. Trintinalia, R. Bhalla, and H. Ling, "Scattering center parameterization of wide-angle backscattered data using adaptive Gaussian representation," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 11, pp. 1664–1668, Nov. 1997.

[30] R. D. Chaney, A. S. Willsky, and L. M. Novak, "Coherent aspect-dependent SAR image formation," *Proc. SPIE*, vol. 2230, pp. 256–274, Apr. 1994.

[31] L. R. Flake, S. C. Ahalt, and A. K. Krishnamurthy, "Detecting anisotropic scattering with hidden Markov models," *IEE P. Radar, Son., Nav.*, vol. 144, no. 2, pp. 81–86, Apr. 1997.

[32] A. J. Kim, J. W. Fisher, III, and A. S. Willsky, "Detection and analysis of anisotropic scattering in SAR data," *Multidimensional Systems and Signal Processing*, vol. 14, no. 1-3, pp. 49–82, Jan. 2003.

[33] M. Çetin and R. L. Moses, "SAR imaging from partial-aperture data with frequency-band omissions," *Proc. SPIE*, vol. 5808, pp. 32–43, Mar. 2005.

[34] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comp.*, vol. 15, no. 2, pp. 349–396, Feb. 2003.

[35] M. Aharon, M. Elad, and A. M. Bruckstein, "On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them," *Linear Algebra Appl.*, vol. 416, no. 1, pp. 48–67, July 2006.

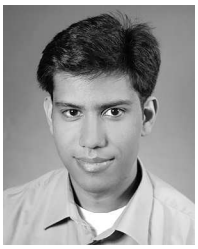
[36] T. F. Coleman and Y. Li, "A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J. Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.

[37] R. Bhalla, A. M. Raynal, H. Ling, J. Moore, and V. J. Velten, "Angular description for 3D scattering centers," *Proc. SPIE*, vol. 6237, p. 623706, Apr. 2006.

[38] K. R. Varshney, M. Çetin, J. W. Fisher, III, and A. S. Willsky, "Wide-angle SAR image formation with migratory scattering centers and regularization in Hough space," in *Proc. Adaptive Sensor Array Processing Workshop*, Lexington, Massachusetts, June 2006.

[39] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[40] N. Aggarwal and W. C. Karl, "Line detection in images through regularized Hough transform," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 582–591, Mar. 2006.



Kush R. Varshney (S'00) was born in Syracuse, New York in 1982. He received the B.S. degree in electrical and computer engineering with honors from Cornell University, Ithaca, New York, in 2004. He received the S.M. degree in electrical engineering and computer science in 2006 and is now pursuing the Ph.D. degree at the Massachusetts Institute of Technology, Cambridge.

He is a National Science Foundation Graduate Research Fellow and a research assistant with the Stochastic Systems Group in the Laboratory for Information and Decision Systems. He was a visiting student at Laboratoire de Mathématiques Appliquées aux Systèmes at École Centrale, Paris in 2006. He was an intern at Sun Microsystems during 2002-2003 and at Sensis Corporation in 2001. His research interests include statistical signal processing, image processing, and machine learning.

Mr. Varshney is a member of Eta Kappa Nu and Tau Beta Pi, and a student member of SIAM.

Müjdat Çetin (S'98–M'02) received the Ph.D. degree from Boston University, Boston, Massachusetts, in 2001, in electrical engineering.

From 2001 to 2005, he was with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge. Since September 2005 he has been an Assistant Professor at Sabancı University, İstanbul, Turkey. He has served in various organizational capacities, including special session organizer, session chair, and technical program committee member for a number of conferences including the IEEE International Conference on Acoustics, Speech, and Signal Processing; the SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery; the IEEE Statistical Signal Processing Workshop; the IEEE International Conference on Image Processing; and the EURASIP European Signal Processing Conference. He has also served as the Technical Program Co-chair for the 2006 IEEE Turkish Conference on Signal Processing and Communications Applications. His research interests include statistical signal and image processing, inverse problems, computer vision, data fusion, wireless sensor networks, biomedical information processing, radar imaging, brain computer interfaces, and sensor array signal processing.

John W. Fisher, III (M'90) received the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 1997.

He is currently a Principal Research Scientist in the Computer Science and Artificial Intelligence Laboratory and affiliated with the Laboratory for Information and Decision Systems, both at the Massachusetts Institute of Technology, Cambridge. Prior to joining the Massachusetts Institute of Technology he was affiliated with the Electronic Communications Laboratory at the University of Florida from 1987 to 1997, during which time he conducted research in the areas of ultra-wideband radar for ground and foliage penetration applications, radar signal processing, and automatic target recognition algorithms. His current area of research focus includes information theoretic approaches to signal processing, multi-modal data fusion, machine learning and computer vision.

Alan S. Willsky (S'70–M'73–SM'82–F'86) joined the Massachusetts Institute of Technology, Cambridge, in 1973 and is the Edwin Sibley Webster Professor of Electrical Engineering and Acting Director of the Laboratory for Information and Decision Systems.

He was a founder of Alphatech, Inc. and Chief Scientific Consultant, a role in which he continues at BAE Systems Advanced Information Technologies. From 1998 to 2002 he served on the U.S. Air Force Scientific Advisory Board. He has received several awards including the 1975 American Automatic Control Council Donald P. Eckman Award, the 1979 ASCE Alfred Noble Prize, the 1980 IEEE Browder J. Thompson Memorial Award, the IEEE Control Systems Society Distinguished Member Award in 1988, the 2004 IEEE Donald G. Fink Prize Paper Award, and Doctorat Honoris Causa from Université de Rennes in 2005.

Dr. Willsky has delivered numerous keynote addresses and is coauthor of the text *Signals and Systems*. His research interests are in the development and application of advanced methods of estimation, machine learning, and statistical signal and image processing.