Path-Based Dictionary Augmentation: A Framework for Improving k-Sparse Image Processing

Tegan H. Emerson, *Member, IEEE*, Colin C. Olson, *Member, IEEE*, and Timothy Doster, *Member, IEEE*

***Abstract*— We have previously shown that augmenting orthogonal matching pursuit (OMP) with an additional step in the identification stage of each pursuit iteration yields improved k-sparse reconstruction and denoising performance relative to baseline OMP. At each iteration a "path" or geodesic, is generated between the two dictionary atoms that are most correlated with the residual and from this path a new atom that has a greater correlation to the residual than either of the two bracketing atoms is selected. Here, we provide new computational results illustrating improvements in sparse coding and denoising on canonical datasets using both learned and structured dictionaries. The two methods of constructing a path are investigated for each dictionary type: the Euclidean geodesic formed by a linear combination of the two atoms and the 2-Wasserstein geodesic corresponding to the optimal transport map between the atoms. We prove here the existence of a higher-correlation atom in the Euclidean case under assumptions on the two bracketing atoms and introduce algorithmic modifications to improve the likelihood that the bracketing atoms meet those conditions. Although, we demonstrate our augmentation on OMP alone, in general it may be applied to any reconstruction algorithm that relies on the selection and sorting of high-similarity atoms during an analysis or identification phase.**

***Index Terms*— Matching pursuit, path augmentation, image denoising, image reconstruction, k-Sparse.**

I. Introduction

A FREQUENT goal within signal/image processing is to reconstruct or compress the information contained in a signal by representing it as a linear combination of a set of reference signals. In the most general case, this reference set is a (possibly overcomplete) *dictionary* composed of signal *atoms* drawn from some underlying signal model. "Good" models are those that can sparsely represent signals as linear combinations of relatively few atoms drawn from the dictionary. Signals that can be represented to within some acceptable error tolerance using at most k atoms are defined as *k-sparse* relative to that dictionary. Reconstruction algorithms designed to decompose the signal into a linear combination of atoms can be designed to terminate based on an error threshold or a fixed degree of sparsity.

For fixed sparsity, consideration of all possible atom combinations of that order is computationally intractable other than for a limited set of problems. A popular and successful approach to this combinatorial optimization problem is a greedy algorithm called *Matching Pursuit (MP)* [3], [21]. Standard MP begins by greedily searching for the best reconstruction produced from a single atom where "best" is determined by the magnitude of the inner product between the signal and the dictionary atoms. This optimal atom is scaled by the length of the projection of the signal onto the space spanned by the optimal atom and is then subtracted from the original signal to yield a residual. The residual image is then fit in the same greedy way, updated, and the process repeats such that k iterations of MP yields a k-sparse representation with some associated final error/residual $\mathbf{R}_k$.

One problem persists, however, because even if the underlying signal model could perfectly represent the signal the atoms must be discretely sampled from the model and will in general fail to represent any given signal component exactly. For example, a 1D sinusoidal signal composed of a single tone is well-represented by a sinusoidal model, but if the frequency of the signal falls between the discrete Fourier frequencies of a given Fourier basis then the number of non-zero Fourier coefficients can actually be quite large [22].

This problem of "basis mismatch" has been considered in the literature [4], [6], [10], [28], [29] and mitigated with some success in the case of 1D sinusoidal signals where a search over the frequencies between the two most similar Fourier basis elements can find the exact representative element [25]. But reconstruction is more difficult for images because relatively few local image regions can be exactly represented by an atom from a single structured (parameterized) dictionary (e.g., Fourier, wavelet). Unstructured, learned dictionaries constructed from ensembles of images generally produce higher-sparsity representations but still suffer from basis mismatch and in such cases there is no model parameter that one can modify to find some better atom residing between the two atoms with which the image patch has the largest correlation.

Overcomplete dictionaries attempt to reduce basis mismatch error by increasing the number of signal exemplars such that any given signal is more likely to be well-represented by at least one member of the dictionary. Unfortunately, there is a trade-off between the degree of overcompleteness and the uniqueness required to ensure sparsity [8]. In addition to increasing the computational burden at each identification stage, increasing overcompleteness essentially reduces the likelihood that there will be just a single representative atom and thus suffers from diminishing returns.

Manuscript received July 17, 2018; revised April 30, 2019; accepted June 26, 2019. Date of publication July 15, 2019; date of current version November 4, 2019. This work was supported in part by the OSD LUCI Fellowship and in part by the Jerome and Isabella Karle Distinguished Scholar Fellowship Program. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dipti P. Mukherjee. *(Corresponding author: Colin C. Olson.)*

The authors are with the Naval Research Laboratory, Washington, DC 20375 USA (e-mail: colin.olson@nrl.navy.mil).

Digital Object Identifier 10.1109/TIP.2019.2927331

We introduced "path-based" augmentation of the matching pursuit algorithm in [12], [13] as a compromise between orthonormal bases and overcomplete dictionaries. In essence, this augmentation can be applied to any reconstruction algorithm that relies on the selection and sorting of high-similarity atoms during an analysis or identification phase. We considered two methods for constructing better image exemplars from a learned dictionary as a means of mitigating basis mismatch: a linear combination of the two most-correlated atoms (Euclidean geodesic) and a construction relying on the optimal transport map between said atoms (2-Wasserstein geodesic) [16]. Computational results were provided for denoising and k-sparse reconstruction using learned dictionaries on a specific short-wave infrared (SWIR) dataset [13] and for improved k-sparse reconstruction of faces [12] using an open-source eigenface dataset [1].

Here, we illustrate that augmenting MP with our path-based modification leads to better general image reconstruction and denoising performance relative to traditional orthogonal matching pursuit (OMP) on canonical datasets using both structured (DCT) and unstructured learned (kSVD) dictionaries. Although learned dictionaries are expected to perform better than their structured cousins, structured dictionaries are of interest in superresolution and compressive sensing applications where high-resolution datasets are difficult and/or expensive to acquire (e.g., infrared images of Navy interest).

In addition, we provide a proof of the existence of an atom on the Euclidean geodesic that is maximally similar to the signal as long as certain assumptions regarding the endpoints of the geodesic are satisfied. Based on these results, we modify the algorithm to increase the probability that said conditions will indeed be satisfied during each analysis stage of the algorithm.

We begin with a general description of path-based augmentation and detail the algorithm as it applies to OMP in Section 2. The proof of the existence of an atom on the Euclidean geodesic that is maximally similar to the signal is provided in Section 3. Numerical experiments and their results are presented in Section 4 before concluding with a discussion in Section 5.[1]

II. Path-Augmented Matching Pursuit

A. Orthogonal Matching Pursuit

As previously mentioned, MP is a greedy search based approach to solving the combinatorial optimization problem to identify the best sparse representation. Due to the success of the original algorithm several variations of MP have been developed [5], [7], [9], [20], [23], [24], [26], [30], [32]. In general each of these variations consist of three steps. The *identification* step in an iteration of an MP-based algorithm refers to determining which atom(s) is(are) closest to the current residual. *Augmentation* is used to describe the step of adding the atom(s) identified to the support of the reconstruction. Finally, each pursuit-type algorithm is concluded by a *residual update*. The fundamental difference between MP and OMP is that the residual in OMP is updated by projecting the image onto the orthogonal complement of the span of the current support. Note that when the dictionary consists of pairwise orthogonal atoms, MP and OMP are equivalent. Here we only consider augmenting basic OMP, the pseudocode for which is provided in Algorithm 1.

Algorithm 1 Orthogonal Matching Pursuit

Input: $\mathbf{T}$, the test image and $\mathcal{D}$, the dictionary, K the number of iterations/sparsity level.

Output: $\mathbf{X}$, the image estimate, $\mathbf{S}$, the support of the reconstruction, and $\{d_k\}_{k=1}^{K}$, the vector of atom indices.

begin
  $\mathbf{R}_1 \leftarrow \mathbf{T}$;
  $\mathbf{S} = []$;
  $k \leftarrow 1$;
  while $k \leq K$ **do**
    $\mathbf{D}^* \leftarrow \arg\max_{\mathbf{D}\in\mathcal{D}} |\langle \mathbf{D}, \mathbf{T}\rangle|$;
    $d_k \leftarrow \text{index}(\mathbf{D}^*)$;
    $\mathbf{S} \leftarrow \text{augment}(\mathcal{S}, \mathbf{D}^*)$;
    $\mathbf{P}_S \leftarrow \mathbf{S}(\mathbf{S}^\top\mathbf{S})^{-1}\mathbf{S}^\top$;
    $\mathbf{X} \leftarrow \mathbf{P}_S\mathbf{T}$;
    $\mathbf{R}_{k+1} \leftarrow (\mathbf{I} - \mathbf{P}_S)\mathbf{T} = \mathbf{T} - \mathbf{X}$;
    $k \leftarrow k + 1$;
  end
end

B. Paths Between Atoms

Our approach can be differentiated from other existing variants of OMP by the construction of a path between the two most-similar dictionary atoms and then performing a secondary indentification step. In the OMP case, similarity is quantified by the inner product between a dictionary atom and the test image (residual) at that stage. We recall that the largest-magnitude inner product is equivalent to the closest vector as determined by the angle, θ, between the vectorized images.

We define a *path* as a smooth map between two dictionary atoms, $\{\mathbf{D}_1, \mathbf{D}_2\}$, that is parameterized by a variable $t \in [0, 1]$. Explicitly, a path p is defined as

$$p(\mathbf{D}_1, \mathbf{D}_2, t) : \mathbf{D}_1 \to \mathbf{D}_2, \tag{1}$$

$$\text{s.t. } p(\mathbf{D}_1, \mathbf{D}_2, 0) = \mathbf{D}_1 \quad \text{and} \tag{2}$$

$$p(\mathbf{D}_1, \mathbf{D}_2, 1) = \mathbf{D}_2. \tag{3}$$

Although a path can be created between any arbitrary dictionary atoms, here we exclusively consider paths between the two atoms that are most similar to the current test signal (i.e., have the largest inner product or the smallest angle). That is to say in a primary identification step we identify the two most similar dictionary atoms and form a path between them.

[1]The numerical results for the Maritime imagery will be appearing in the proceedings of a workshop in the 2018 IEEE Conference on Computer Vision and Pattern Recognition [13]. The algorithm will also be appearing in the proceedings of the 2018 European Signal Processing Conference [12]. All theoretical results are novel and have not appeared in previously published or currently accepted papers.

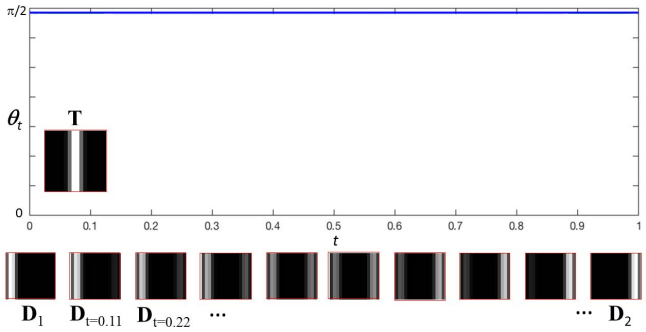

Fig. 1. Samples along the linear path between two bar atoms $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown below a plot displaying the angle between the test image, $\mathbf{T}$, and the path samples, $\mathbf{D}_t$. Samples are generated using Equation 4 for a discrete set of t sampled with uniform spacing from [0, 1]. All atoms along the path are equally similar, or dissimilar, to the test atom as there is never any intensity overlap between a path atom and the test atom.

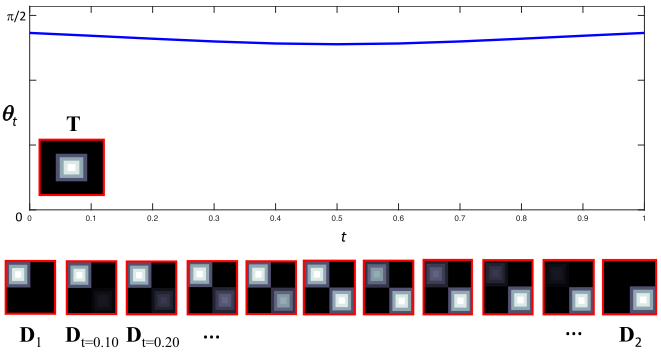

Fig. 2. Samples along the linear path between two Gaussian atoms $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown below a plot displaying the angle between the test image, $\mathbf{T}$, and the path samples, $\mathbf{D}_t$. Samples are generated using Equation 4 for a discrete set of t sampled with uniform spacing from [0, 1]. An optimal atom along the linear path occurs at $t = 0.50$ due to maximal intensity overlap. The location of this optimal atom could also have been predicted according to Theorem 3.4 after noting that $\cos(\theta_1) = \cos(\theta_2)$ and $\cos(\theta_{1,2}) = \pi/2$.

Given a path between two atoms we search for a novel atom which is more similar to the test signal than either of the path end-points. Here, we "search" by drawing samples from a path that correspond to a discrete set of t sampled with uniform spacing from [0, 1] and test their similarity to the test image. An alternate formulation might seek a local optimum by performing a line search along the geodesic. We further note that in some cases there is a closed form for identifying the optimal sample along the path as will be shown in Section III. If a novel atom that is more similar to the test image exists it is added to the current support of the test signal. If a more similar atom is not found on the geodesic, the most similar of the original two atoms identified in the primary identification step will be selected instead.

A variety of viable paths between the atoms exist and some will be considered in later work. At present we narrow the focus to a globally linear path (Euclidean geodesic) and a globally nonlinear path (2-Wasserstein geodesic) approximated by pixel-wise linear trajectories.

1) Linear Path: Perhaps the most familiar path is a line segment with endpoints $\mathbf{D}_1$ and $\mathbf{D}_2$ given by

$$\mathbf{D}_t = (1-t)\mathbf{D}_1 + (t)\mathbf{D}_2, \tag{4}$$

for $t \in [0, 1]$. This path is the Euclidean geodesic between $\mathbf{D}_1$ and $\mathbf{D}_2$, that is, the shortest path in Euclidean space.

Examples drawn from the linear path between two image atoms $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown in Figure 1 and Figure 2 for atoms with "bar" and Gaussian structures, respectively. Note that in both cases the sampled images are characterized by the presence of intensity that matches the combined support from $\mathbf{D}_1$ and $\mathbf{D}_2$ and simply shifts intensity magnitude from the support represented by $\mathbf{D}_1$ to the support represented by $\mathbf{D}_2$ as t increases. As such, all samples along the Euclidean geodesic between the two bar atoms are equidistant from and orthogonal to the test image. That is, there is no overlap in the intensity support between $\mathbf{T}$, either atom, or any linear combination of the atoms and therefore the angle between $\mathbf{T}$ and any sample from the geodesic will be 90 degrees. In the Gaussian example, the most similar sample from the geodesic is the endpoint, $\mathbf{D}_2$.

2) Optimal Transport Path: An alternate path can be constructed by an Optimal Transport (OT) map that transforms one image into the other while minimizing the transport energy of the map [31]. Two main versions of OT exist: (1) the Monge formulation in which all the intensity located at a pixel in $\mathbf{D}_1$ must be mapped to a single pixel in $\mathbf{D}_2$, and (2) the Kantorovich formulation which allows for intensities at starting pixels to be split among multiple destination pixels [31].

Within both Monge and Kantorovich OT formulations there are additional sub-versions for different combinations of objective function and constraints. Specifically, different formulations of the OT problem produce constant-speed geodesics with respect to different distance measures. Paths produced by OT between images have shown improved performance in image registration and warping [15], super-resolution of low-resolution face images [19], and cell morphology [2]. We generate paths corresponding to an approximate solution to the Monge OT using a recently developed, computationally-efficient approximation based on the Radon Cumulative Distribution Transform (RCDT) [17], [18].

A solution to the Monge OT problem between images yields a vector field of direction vectors that implicitly indicate the terminal location (in $\mathbf{D}_2$) of intensity from a given pixel in $\mathbf{D}_1$. Let

$$\mathbf{V} = \begin{bmatrix} \vec{v}_{1,1} & \vec{v}_{1,2} & \cdots & \vec{v}_{1,m} \\ \vec{v}_{2,1} & \vec{v}_{2,2} & \cdots & \vec{v}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_{n,1} & \vec{v}_{n,2} & \cdots & \vec{v}_{n,m} \end{bmatrix} \tag{5}$$

where $\vec{v}_{j,k}$ is the velocity vector for the intensity of the pixel indexed by (j, k) in $\mathbf{D}_1$. Let $p_V(\mathbf{D}_1, \mathbf{D}_2, t)$ be the path induced by $\mathbf{V}$. We define

$$p_V(\mathbf{D}_1, \mathbf{D}_2, t) \ : \ \mathbf{D}_1 \rightarrow \mathbf{D}_2 \text{ s.t.} \tag{6}$$

$$p_V(\mathbf{D}_1, \mathbf{D}_2, t) = t\mathbf{V}(\mathbf{D}_1, \mathbf{D}_2) \tag{7}$$

where $t\mathbf{V}(\mathbf{D}_1, \mathbf{D}_2)$ indicates movement of the intensity in $\mathbf{D}_1$ a partial step (of size t) in the directions given by OT. This OT path can be thought of as a set of linear path approximations to

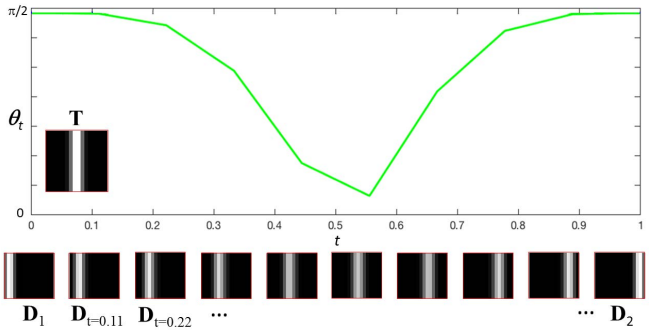

Fig. 3. Samples along the 2-Wasserstein geodesic between two bar atoms $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown below a plot displaying the angle between the test image, $\mathbf{T}$, and the path samples, $\mathbf{D}_t$. Samples are generated using Equation 5 and Equation 7 for a discrete set of t sampled with uniform spacing from [0, 1]. The minimum angle occurs at the value of t where the intensity support of the geodesic sample and the test image maximally overlap.

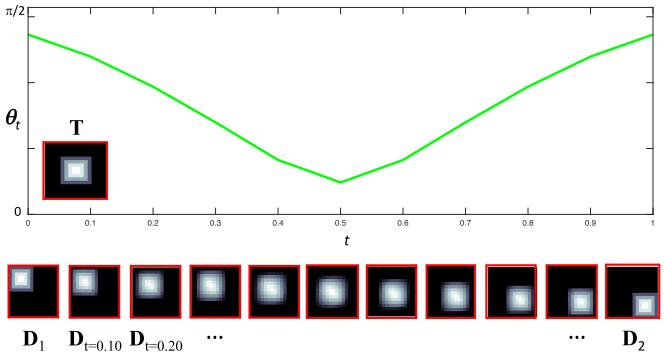

Fig. 4. Samples along the 2-Wasserstein geodesic between two Gaussian atoms $\mathbf{D}_1$ and $\mathbf{D}_2$ are shown below a plot displaying the angle between the test image, $\mathbf{T}$, and the path samples, $\mathbf{D}_t$. Samples are generated using Equation 5 and Equation 7 for a discrete set of t sampled with uniform spacing from [0, 1]. The minimum angle occurs at the value of t where the intensity support of the geodesic sample and the test image maximally overlap.

a globally nonlinear path between images (the 2-Wasserstein geodesic).

In other words, we will transform $\mathbf{D}_1$ into $\mathbf{D}_2$ by transporting pixel intensities at constant speed from starting pixel locations (j, k). The velocity map, $\mathbf{V}$, provides both the speed and direction (i.e., velocity) that the intensity at pixel location (j, k) in $\mathbf{D}_1$ will need to move over a total transport time given by $t = 1$ to produce the final image $\mathbf{D}_2$. Intermediate images along this geodesic are generated by selecting $t < 1$ such that each pixel intensity starting in $\mathbf{D}_1$ is only partially transported along its trajectory at constant velocity toward its final position in $\mathbf{D}_2$.

Examples of samples along the OT path between two image atoms with bar and Gaussian structures are shown in Figure 3 and Figure 4, respectively. Contrary to the linear case, the intensity support for samples along the geodesic is seen to shift between the support of the two endpoint images. As a result, there is a clear minimum angle (maximum similarity) between the bar test image and one of the geodesic samples in Figure 3 when their intensity supports maximally overlap.

The effect of this shift in support can also be clearly seen in Figure 4 where images sampled from the geodesic show a smooth melding from the single Gaussian intensity spot of $\mathbf{D}_1$ into the two Gaussian intensity spots in $\mathbf{D}_2$. In both examples, a sample from the geodesic is found to be more similar to the test image.

C. The Algorithm: POMP

Path Orthogonal Matching Pursuit (POMP) is a modification to the well-known OMP algorithm described in Section II-A. Instead of finding a single nearest dictionary element and removing its contribution, two dictionary elements are chosen at each iteration. A path is formed which moves between the closest dictionary atom and a secondary dictionary atom (which meets criteria presented in Section III) and an optimal atom is found along the path where optimality is defined as having the largest-magnitude inner product with the test image.

A positive inner product can be interpreted as two images sharing more in-phase (same sign) intensities than out-of-phase intensities. When considered from the phase perspective, it is reasonable to pick two atoms whose inner products share the same sign with the test image. As will be shown in Section III the assumption of shared signs is a sufficient condition to guarantee the existence of a novel optimal atom for the linear path given an orthonormal dictionary. If $sgn(\langle \mathbf{T}, \mathbf{D}_1 \rangle) \neq sgn(\langle \mathbf{T}, \mathbf{D}_2 \rangle)$, then we set $\mathbf{D}_2 = -\mathbf{D}_2$ so that the signs of the inner products are in agreement. We adopt the same sign heuristic for the OT-based path although this has not been theoretically shown to be a sufficient condition to guarantee improvement.

Let $\mathbf{D}_1$ be the dictionary atom having the largest-magnitude inner product with the test image. Define $\mathbf{D}_2$ to be a second-closest dictionary atom. The simplest, and perhaps most natural, form of a path is linear. When the globally-linear path (Equation 4) is used within an iteration of POMP we refer to it as L-POMP. The pixel-wise linear path resulting from solving the OT problem between $\mathbf{D}_1$ and $\mathbf{D}_2$ selected within an iteration of POMP is denoted by OT-POMP.

For samples along the paths between $\mathbf{D}_1$ and $\mathbf{D}_2$, the angle between $\mathbf{T}$ may be computed by

$$\theta_t = \cos^{-1} \frac{\langle \mathbf{D}_t, \mathbf{T} \rangle}{||\mathbf{D}_t||_F ||\mathbf{T}||_F}, \tag{8}$$

where $t \in [0, 1]$ parameterizes the distance along the path from $\mathbf{D}_1$ to $\mathbf{D}_2$ and $\|A\|_F = \sqrt{\sum_i \sum_j |a_{i,j}|^2}$ is the Frobenius norm. Let $p(\mathbf{D}_1, \mathbf{D}_2, t)$ be the path from $\mathbf{D}_1$ to $\mathbf{D}_2$. At each iteration the optimal atom is

$$\mathbf{D}_{t_\star} = p(\mathbf{D}_1, \mathbf{D}_2, t_\star) \quad \text{where} \quad t_\star = \arg\min_{t \in [0,1]} \theta_t. \tag{9}$$

Pseudocode for POMP is provided in Algorithm 2. Lines 2-4 initialize variables. In line 6 we find the optimal atom and in line 7 we add its index to the list of support atoms. The signs of the inner products of the first and second closest (identified in line 9) atoms are identified and matched, if necessary, in lines 10-14. The optimal atom along the path between the two nearest neighbors is selected in lines 16-17 and is

Algorithm 2 Path Orthogonal Matching Pursuit

Input: $\mathbf{T}$, the test image and $\mathcal{D}$, the dictionary, and K the number of iterations/sparsity level.

Output: $\mathbf{X}$, the image estimate, $\mathbf{S}$, the support of the reconstruction, $\{d_k^1\}_{k=1}^K$, the vectors of first closest atom indices, $\{d_k^2\}_{k=1}^K$, the vector of second closest atom indices, and $\{t_k\}_{k=1}^K$, the vector containing the path parameter values.

1 **begin**
2 $\quad \mathbf{R}_1 \leftarrow \mathbf{T}$;
3 $\quad \mathbf{S} = []$;
4 $\quad k \leftarrow 1$;
5 $\quad$ **while** $k \leq K$ **do**
6 $\quad\quad \mathbf{D}_1 \leftarrow \arg\max_{\mathbf{D}\in\mathcal{D}} |\langle \mathbf{D}, \mathbf{T}\rangle|$;
7 $\quad\quad d_k^1 \leftarrow \text{index}(\mathbf{D}_1)$;
8 $\quad\quad s \leftarrow \text{sgn}(\langle \mathbf{D_1}, \mathbf{T}\rangle)$;
9 $\quad\quad \mathbf{D}_2 \leftarrow \arg\max_{\mathbf{D}\in\mathcal{D}\setminus\mathbf{D}_1} |\langle \mathbf{D}, \mathbf{T}\rangle|$;
10 $\quad\quad$ **if** $\text{sgn}(\langle \mathbf{D_1}, \mathbf{T}\rangle) \neq \text{sgn}(\langle \mathbf{D_1}, \mathbf{T}\rangle)$ **then**
11 $\quad\quad\quad \mathbf{D}_2 \leftarrow -\mathbf{D}_2$
12 $\quad\quad$ **else**
13 $\quad\quad\quad \mathbf{D}_2 \leftarrow \mathbf{D}_2$
14 $\quad\quad$ **end**
15 $\quad\quad d_k^2 \leftarrow \text{index}(\mathbf{D}_2)$;
16 $\quad\quad t_k \leftarrow \arg\min_{t\in[0,1]} \theta_t$;
17 $\quad\quad \mathbf{D}^* \leftarrow \text{path}(\mathbf{D}_1, \mathbf{D}_2, t_n)$;
18 $\quad\quad \mathbf{S} \leftarrow \text{augment}(\mathbf{S}, \mathbf{D}^*)$;
19 $\quad\quad \mathbf{P}_S \leftarrow \mathbf{S}(\mathbf{S}^\top\mathbf{S})^{-1}\mathbf{S}^\top$;
20 $\quad\quad \mathbf{X} \leftarrow \mathbf{P}_S\mathbf{T}$;
21 $\quad\quad \mathbf{R}_{k+1} \leftarrow (\mathbf{I} - \mathbf{P}_S)\mathbf{T}$;
22 $\quad\quad k \leftarrow k + 1$;
23 $\quad$ **end**
24 **end**

then appended to the support in line 18. Residual updates and updating of indexing variables are performed in lines 19-22.

It is important to note that without further constraints, there is no guarantee of a novel/nontrivial minimum angle being found along the path, that is, $\mathbf{D}_* = \mathbf{D}_t$ for $t \in (0, 1)$. As long as the path is continuous for $t \in [0, 1]$ a minimum will exist since the composition of continuous functions is also continuous (inner product composed with the path). This minimum may occur at $t = 1$ (nontrivial) and may not be unique. Consider $|\langle \mathbf{T}, \mathbf{D}_t\rangle|$ for some $t \in [0, 1]$. The desired inner product can be written as

$$\langle \mathbf{T}, \mathbf{D}_t\rangle = \frac{1}{2}\begin{bmatrix}\mathbf{T}\\ \mathbf{D}_t\end{bmatrix}^\top \begin{bmatrix}\mathbf{0} & \mathbf{I}\\ \mathbf{I} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{T}\\ \mathbf{D}_t\end{bmatrix}. \tag{10}$$

This is an indefinite quadratic form, that is, the characteristic matrix has eigenvalues $\{-1, 1\}$ with each one having the same multiplicity [14]. As a result, the inner product is not generally convex. However, with added constraints on the equations governing $\mathbf{D}_t$ it may be possible to prove convexity. Current results guaranteeing existence of non-trivial atoms along the Euclidean geodesic between atoms are presented in Section III. A rigorous study of these necessary and sufficient properties of the path form are the focus of ongoing and future work.

Regardless, the proposed algorithm can be seamlessly combined with OMP as well as many of its derivatives. When the dictionary consists of pairwise orthogonal atoms, MP and OMP are equivalent. If the dictionary atoms are orthogonal then the linear combination of two of the atoms will also be orthogonal to all other atoms. Thus, a linear path-based MP algorithm will be equivalent to an OMP algorithm when an orthogonal dictionary is used. When guarantees about orthogonality along a path cannot be made, the reconstruction can differ between MP and OMP implementations.

III. Linear Path Theoretical Results

We will show a set of sufficient conditions under which there is guaranteed to be an improvement in reconstruction error using a linear path orthogonal matching pursuit. The more involved proofs have been placed in Appendix . In standard matching pursuit a "closest" dictionary atom is chosen for the current residual. Previously we used bold face capitals to denote the matrix form of images and dictionary elements. Moving forward we will use lower case letters to denote the vectorized version of a matrix. This convention is appropriate since $||\mathbf{A}||_F = ||a||$ when $a = vec(\mathbf{A})$.

Let r be the current residual and let d_1 be the best dictionary element, i.e. $|\langle r, d_1\rangle| > |\langle r, d\rangle|$ for all $d \in \mathcal{D}$. Recall the inner product between two vectors can be equivalently written as $\langle r, d\rangle = r^\top d = ||r||\ ||d|| \cos(\theta)$ where $r, d \in \mathbb{R}^N$, θ is the angle between the two vectors, and $||\cdot||$ is the usual 2-norm. When the dictionary is comprised of unit length vectors and the current residual has also been normalized we then have $\langle r, d\rangle = \cos(\theta)$. Let d_1 be chosen as above. Let $d_2 \in \mathcal{D}$ and consider the linear path between d_1 and d_2 defined by $d_t = td_1 + (1 - t)d_2$ for $t \in [0, 1]$.

Another way to define the "closest" dictionary atom is as the dictionary atom d_1 such that the length of the orthogonal projection of r onto d_1 is the longest, i.e., $||\mathbf{P}_{d_1} r||^2 \geq ||\mathbf{P}_d r||^2$ for all $d \in \mathcal{D}$ where $\mathbf{P}_d = d(d^\top d)^{-1}d^\top$ is the orthogonal projection matrix onto the vector d. Consequently, in order for our presented algorithm to have merit there must be some atom (d_t) along the t-parameterized path between d_1 (as defined) and some $d_2 \in \mathcal{D}$ satisfying that the length of the projection of r onto d_t is greater than the length of the projection onto d_1. This alternative way of considering closeness will play a critical role in the proof of our theorem.

Before we prove our theorem, we will first state two lemmas. Proofs of these lemmas are omitted but can be readily verified using linear algebra.

Lemma 3.1: If $d_1, d_2 \in \mathcal{D}$, $||d_1|| = ||d_2|| = 1$, given the linear path between d_1 and d_2 defined by $d_t = td_1 + (1-t)d_2$ for $t \in [0, 1]$, then the orthogonal projection matrix onto d_t is given by

$$\mathbf{P}_{d_t} = \frac{t^2 d_1 d_1^\top + t(1-t)(d_1 d_2^\top + d_2 d_1^\top) + (1-t)^2 d_2 d_2^\top}{t^2 + 2t(1-t)\cos\theta_{1,2} + (1-t)^2}$$

This lemma falls out directly from the definition of the orthogonal projection matrix onto d_t given by

$\mathbf{P}_{d_t} = d_t(d_t^\top d_t)^{-1}d_t^\top$. Using this definition and the definition of d_t together with properties of vector transpose one can verify the result.

Lemma 3.2: If $||\mathbf{P}_{d_1}r||^2 > ||\mathbf{P}_{d_2}r||^2$ *then* $|\langle r, d_1\rangle| > |\langle r, d_2\rangle|$.

The proof of this lemma relies simply on the definition $||v||^2 = \langle v, v\rangle$. Armed with these two lemmas we now state and prove our main result.

Theorem 3.3 Straddle Theorem: Let $||d_1|| = ||d_2|| = ||r|| = 1$, $\langle r, d_1\rangle = r^\top d_1 = \cos\theta_1, \langle r, d_2\rangle = r^\top d_2 = \cos\theta_2$ *and* $\langle d_1, d_2\rangle = d_1^\top d_2 = \cos\theta_{1,2}$. *If* $|\langle r, d_1\rangle| > |\langle r, d\rangle|$ *for all* $d \in \mathcal{D}$, $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$ *and* $\cos\theta_{1,2} \in [0, \cos\theta_1/\cos\theta_2)$, *then there exists* $t_\star \in (0, 1)$ *such that* $|\langle r, d_{t_\star}\rangle| > |\langle r, d_1\rangle|$.

The proof largely relies on a few trigonometric identities and algebraic manipulations. The proof of Theorem 3.3 is presented in Appendix A for interested readers. We note that the condition that $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$ can be trivially satisfied for any dictionary containing two or more elements. In the case that $\text{sgn}(\cos\theta_1) \neq \text{sgn}(\cos\theta_2)$, i.e. $(\text{sgn}(\langle r, d_1\rangle) = -\text{sgn}(\langle r, d_2\rangle)$ for all $d_2 \in \mathcal{D}$ such that $d_2 \neq d_1)$, multiply any dictionary element by -1 and there is now an atom with $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$. Moreover, to make the ratio $\cos\theta_2/\cos\theta_1$ as large as possible we employ the heuristic of picking the two closest atoms by measure of the magnitude of the inner product and multiplying the second closest atom by -1 if necessary to match the inner-product sign of the closest atom. This heuristic guarantees the first assumption will be met and increases the probability that the second assumption is met.

To help build intuition we look to Figure 5 which contains two examples in 2D: one where the assumptions of Theorem 3.3 are met and one where they are not. As previously stated it is always possible to satisfy the condition $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$. The difficulty can arise in ensuring the pair $\{d_1, d_2\}$ satisfy $\cos\theta_{1,2} < \cos\theta_2/\cos\theta_1$. In two dimensions, this inequality can be interpreted as d_1 and d_2 must "straddle" the vector r. Furthermore, consider that we want to find a $t \in (0, 1)$ such that

$$|\langle r, d_1\rangle| < |\langle r, td_1 + (1-t)d_2\rangle|. \tag{11}$$

From properties of the inner product this can be rewritten as

$$|\langle r, d_1\rangle| < |\langle r, d_2 + t(d_1 - d_2)\rangle|. \tag{12}$$

The vectors in green in Figure 5 are the vectors $d_1 - d_2$ for each scenario and the black dashed lines correspond to $d_2 + t(d_1 - d_2)$ for multiple ts. Thus, in the top picture it is apparent that there are multiple ts with the desired property while in the bottom there are none.

Keeping this intuition in mind we can now consider some special cases of Theorem 3.3. These cases are summarized in the following two corollaries and theorem. First we begin with the special case of an orthonormal dictionary. This setting is possible when the problem is under-determined or when the dictionary is at most a spanning set.

Corollary 3.1 Orthonormal Dictionary Case: If $\mathcal{D}$ *is an orthonormal set, then there will always be a* $t \in (0, 1)$ *such that* d_t *has the desired property.*

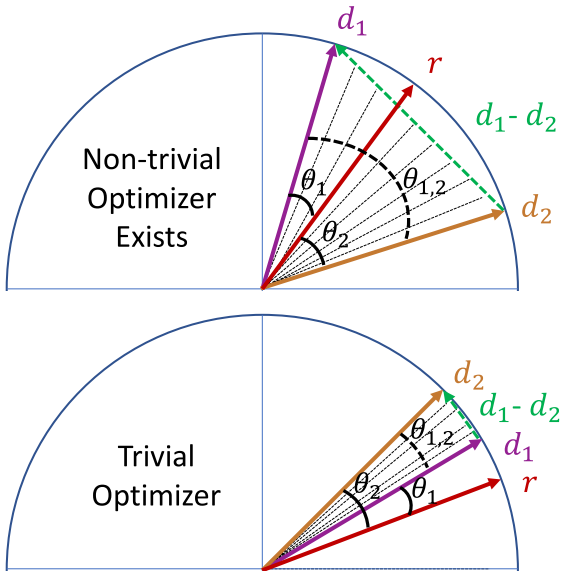

Fig. 5. 2D examples where nontrivial optimizer does/does not exist, that is, the conditions of Theorem 3.3 are/are not satisfied. Dashed black lines correspond to various values of t.

In the case that $\mathcal{D}$ is an orthonormal set we have that $\cos\theta_{1,2} = 0$ for all pairs $\{d_1, d_2\}$ and thus $\cos\theta_{1,2} \in [0, \alpha)$ and the assumptions of Theorem 3.3 are trivially satisfied and the existence of a t with the desired properties follows.

Corollary 3.2 Positive Maximizer: For each pair $\{d_1, d_2\} \in \mathcal{D}$ *if* $|\langle r, d_1\rangle| > |\langle r, d_2\rangle|$ *and* $sgn(\cos\theta_1) = sgn(\cos\theta_2)$ *there is a single positive* t *(given by a closed form) that*

$$t = \frac{\cos(2\theta_1) - \cos(2\theta_2)}{\cos(2\theta_1) - \cos(2\theta_2) + 4\cos(\theta_1)(\cos(\theta_2) - \cos(\theta_1)\cos(\theta_3))} \tag{13}$$

maximizes $||\mathbf{P}_{d_{t_\star}}r||^2 - ||\mathbf{P}_{d_1}r||^2$. *Explicitly*

$$t_\star = \frac{(\cos\theta_1 - \cos\theta_2\cos\theta_{1,2})\left(\csc\left(\frac{\theta_{1,2}}{2}\right)\right)^2}{2(\cos\theta_1 + \cos\theta_2)}. \tag{14}$$

The proof or Corollary 3.2 is provided in Appendix B. The result is shown by maximizing, over t, *the difference in projection lengths between the most similar atom and atoms along the parameterized path. Two critical points are identified, one of which is shown to correspond to a* $t > 0$. *As a last theoretical result we state our final theorem.*

Theorem 3.4 Orthogonal Closed-Form Solution: Let

$$||d_1|| = ||d_2|| = ||r|| = 1,\ \langle r, d_1\rangle = r^\top d_1 = \cos\theta_1,$$
$$\langle r, d_2\rangle = r^\top d_2 = \cos\theta_2,\ \textit{and}\ \langle d_1, d_2\rangle = d_1^\top d_2 = \cos\theta_{1,2}.$$

If $|\langle r, d_1\rangle| \geq |\langle r, d\rangle|$ *for all* $d \in \mathcal{D}$, $sgn(\cos\theta_1) = sgn(\cos\theta_2)$ *and* $\cos\theta_{1,2} = 0$, *then*

$$t_\star = \frac{\cos\theta_1}{\cos\theta_1 + \cos\theta_2} \in (0, 1) \tag{15}$$

will yield the greatest improvement for the current iteration.

The proof of this final results amounts to showing that for $\cos\theta_{1,2} = 0$ the maximizer stated in Corollary 3.2 is less than

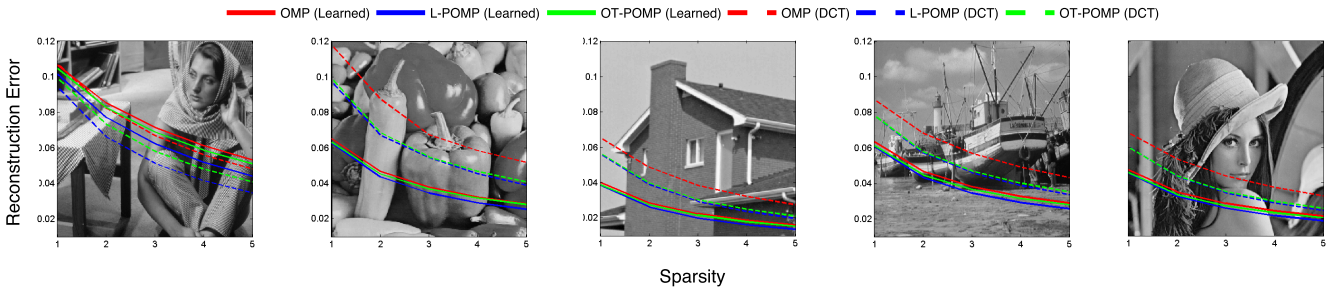

Fig. 6. Reconstruction error, $||\mathbf{R}_k||_F/||\mathbf{T}||_F$, as a function of sparsity k using OMP and the two path-augmented variants using either DCT or kSVD dictionaries. Standard images from left to right are: Lena, Barbara, Peppers, House, and Boat.

one. This proof is straight forward when recalling that for $\cos\theta_{1,2} = 0$, we also have $\csc\left(\frac{\theta_{1,2}}{2}\right)^2 = 2$.

Practically speaking, Theorems 3.3 and 3.4 verify the necessity of picking two atoms with the same signed inner product to the current residual. In the case of an orthonormal dictionary (possible when the dictionary is at most a spanning set) the guaranteed maximizer can be readily computed and the d_2 yielding the greatest improvement can be chosen. Alternatively, when the dictionary is unstructured/learned the assumptions of the Theorem 3.3 can be checked and any $t = m \in (0, 1)$ can be chosen for m defined as in Eq. 20.

IV. Computational Results

Computational results are shown for both $k-$sparse signal reconstruction examples and image denoising using a fixed number of OMP iterations. We test our method on two data sets: one canonical set of images and one comprised of SWIR maritime imagery. We consider both reconstruction and denoising using two types of dictionaries: structured (DCT) and unstructured (learned) overcomplete dictionaries. Overcomplete dictionaries cannot be orthonormalized and so the theoretical results of Section III cannot be leveraged, that is, there is no known closed-form solution for the optimal t along the path. Consequently, for both path types we compare the current residual to 20 discrete samples along the path to check for a closer atom. This implementation is sub-optimal but provides sufficient proof of concept.

A. Canonical Image Data Set

1) $k-$Sparse Reconstructions: We begin by considering the reconstruction error between a pristine original image and a patch-based k-sparse reconstruction using a dictionary. In particular, we consider the DCT and kSVD dictionaries composed of 256 8×8 image atoms that were provided in support of [11]. Each dictionary is used to reconstruct 8x8 patches overlapped with a stride of four where the error between the original and the reconstructed image is given by $||\mathbf{R}_k||_F/||\mathbf{T}||_F$, where k is the number of atoms included in the reconstruction. Reconstruction errors as a function of k for all algorithms and images are shown in Figure 7.

Figure 7 shows that the path-augmented approach yields lower reconstruction error for all tested sparsity levels and images. Reconstructions produced from the DCT dictionary are nearly always worse than the learned dictionary but the differences in performance between the three algorithms are greater in the case of the DCT dictionary. In the case of the DCT dictionary the reconstructions obtained using the proposed heuristic show OT-POMP performing better than L-POMP. However, in the case of the learned dictionary L-POMP achieves better fidelity than OT-POMP. We strongly suspect that OT-POMP will outperform L-POMP once we identify the appropriate heuristic to identify the two bracketing atoms for OT-POMP. For example, Figure 1 shows a case where no trivial minimum can be found along the linear path (due to mismatched signs) yet there exists a much closer atom along the OT path than at either endpoint. Potential modifications for improving POMP results are presented in Section V.

TABLE I

Starting PSNR Values for the Five Images and Denoised PSNR Values for Three Methods Using DCT Dictionaries, Results Reported for $k = 1$. See Figure 11 for a Representation of the Imagery

Image	Original	OMP	L-POMP	OT-POMP
Barbara	22.096	25.611	26.513	**26.680**
Peppers	22.112	24.574	25.948	**26.555**
Boat	22.096	26.507	27.213	**27.854**
Lena	22.096	28.831	29.445	**30.001**
House	22.112	28.369	29.196	**29.759**

2) Denoising: Denoising experiments were also performed where white Gaussian noise with $\sigma = 20$ was added to each pristine image. Patches of the noisy image, of size 8×8, are then estimated using 5 iterations of the indicated algorithm. Both DCT and the globally trained learned dictionary are tested. A denoised image is then constructed by stitching together the OMP estimated patches. Performance is measured using output PSNR as is standard practice. The noise level presented was one of the middle levels tested in [11]. Results of DCT and kSVD denoising are shown in Tables I and II, respectively. An example of the different denoising methods can be seen in Figure 8.

Path augmentation of OMP improves PSNR for each image after patch fitting for a fixed number of OMP iterations. Average improvement in PSNR using POMP is 1.3dB, across all five images, on the structured dictionary. When POMP outperforms OMP using the learned dictionary the improvement

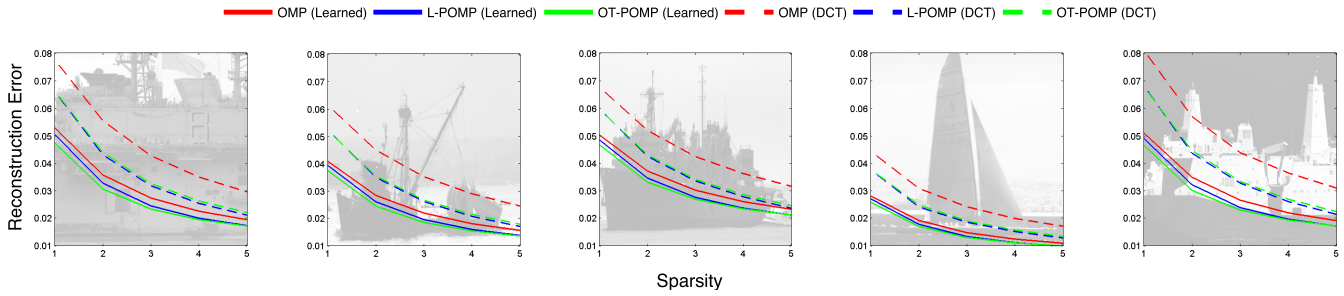

Fig. 7. Reconstruction error, $||\mathbf{R}_k||_F/||\mathbf{T}||_F$, as a function of sparsity k using OMP and the two path-augmented variants using either DCT dictionary (left image set) or the learned kSVD dictionary (right image set).

TABLE II
STARTING PSNR VALUES FOR THE FIVE IMAGES AND DENOISED PSNR VALUES FOR THREE METHODS USING LEARNED DICTIONARIES. RESULTS REPORTED FOR $k = 1$. SEE FIGURE 11 FOR A REPRESENTATION OF THE IMAGERY

Image	Original	OMP	L-POMP	OT-POMP
Barbara	22.096	25.440	**25.726**	25.596
Peppers	22.112	28.704	**28.827**	28.813
Boat	22.096	28.497	28.621	**28.665**
Lena	22.096	**30.650**	30.539	30.535
House	22.112	**31.049**	30.840	30.823

Fig. 8. Denoising example. Clockwise from upper left, noisy image ($\sigma = 20$), OMP denoised, L-POMP denoised, OT-POMP denoised with DCT dictionary, $k = 1$.

Fig. 9. A subsample of the training data set used to create the learned dictionary.

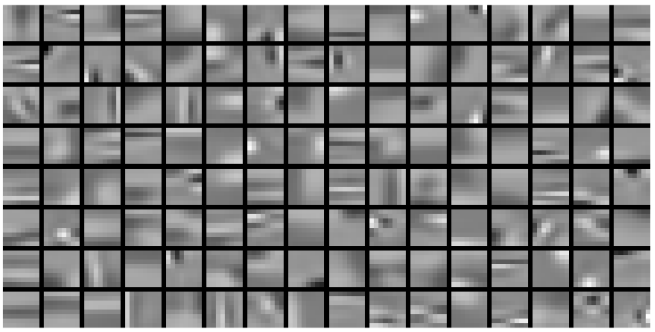

Fig. 10. The kSVD dictionary learned from a corpus of SWIR maritime images.

is 0.2dB but overall PSNR is greater for denoising performed on the learned rather than structured DCT dictionary. Table I shows that OT-POMP performs better on the DCT dictionary than the learned dictionary. Using a learned but spanning (not overcomplete) dictionary may improve POMP performance. In the L-POMP case this would guarantee the existence of a nontrivial minimizer along the path.

B. SWIR Maritime Imagery

To construct a learned dictionary one must identify a collection of a training images that roughly represent the expected statistical qualities of the imagery under test. We choose a set of 85 broadband SWIR images of marine vessels including Naval (military), fishing, cargo, and sailing. Each 16-bit image is of size 1024×1280–a subset of which can be seen in Figure 9. This corpus of images was divided randomly such that 80 images were designated to train the dictionary and 5 images were withheld for testing. We randomly selected 5,000 patches of size 8×8 from each training image to generate the kSVD dictionary. Over 100 iterations of the kSVD algorithm 256 atoms were selected to minimize the representation error using OMP and a target sparsity of 5 atoms to yield the learned dictionary shown in Figure 10.

Computational results are shown for both k-sparse signal reconstruction (see Section IV-B.1) and image denoising (see Section IV-B.2 using a fixed number of OMP iterations). Both applications are tested on the five test images shown as backgrounds in Figure 11. For both algorithms we compare

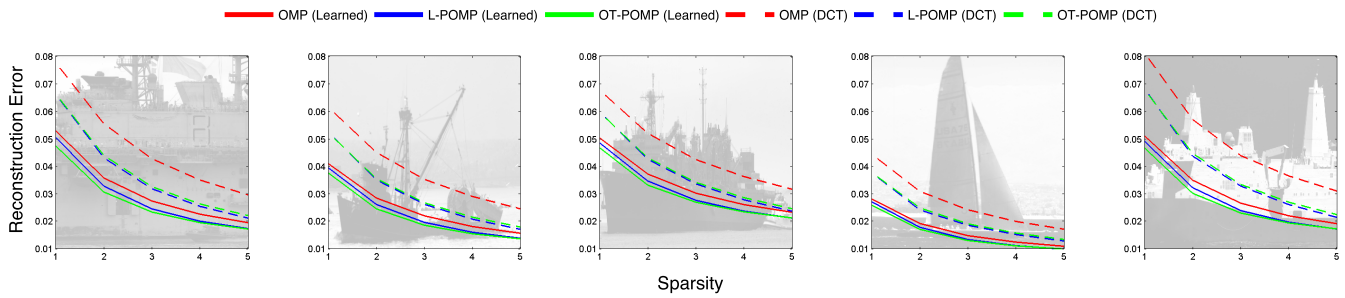

Fig. 11. Relative reconstruction error as a function of iterations for the three algorithms considered. Results over 5 iterations are shown. Background of each plot is the test image under consideration (note that the alpha value has been lowered for graph readability). Images are numbered 1-5 left-to-right to match with the numbering in Table IV.

the current residual to five discrete samples along the path to check for a closer atom.

1) $k-$Sparse Reconstructions: As was done for the canonical image data set, we consider the reconstruction error between the original image and a patch-based k-sparse reconstruction using the dictionary. The dictionary is used to reconstruct 8x8 patches overlapped with a stride of one where the error between the original and the reconstructed image is given by $||\mathbf{R}_k||_F/||\mathbf{T}||_F$, where k is the number of atoms included in the reconstruction. Reconstruction errors as a function of k for all algorithms and images are shown in Figure 11; the background of each plot is the image being sparsely reconstructed.

Figure 11 shows that the path-augmented approach yields lower reconstruction error for all tested sparsity levels and images. In each case the reconstructions obtained using the proposed heuristic show L-POMP and OT-POMP performing better than OMP. It is also clear that OT-POMP outperforms L-POMP over the first few iterations (sparsity level) but after approximately 5 the two versions converge in performance. We have noticed–though the results are omitted here–that both L-POMP and OT-OMP continue to outperform OMP for decreasing sparsity levels but L-POMP begins to overtake OT-POMP in performance. This is believed to occur due to additional possibility of errors introduced by the estimation algorithm used to construct the OT geodesic, which is not a problem in the linear case.

2) Denoising: Denoising experiments were also performed where additive white Gaussian noise (AWGN) with $\sigma = 2,500$ (images are 16-bit) was added to each pristine image. Patches of the noisy image, of size 8×8, are then estimated using several iterations of the indicated algorithm. A denoised image is then constructed by stitching together the OMP estimated patches. Performance is measured using output PSNR as is standard practice. Results of denoising are shown in Table IV for a sparsity of 1. Path augmentation of OMP improves PSNR for each image after patch fitting for a fixed number of OMP iterations. Average improvement in PSNR using L-POMP is ~ 0.2dB and using OT-POMP is ~ 1.1dB.

TABLE III

STARTING PSNR VALUES FOR FIVE IMAGES AND DENOISED PSNR VALUES FOR THE THREE METHODS USING THE DCT DICTIONARY. SEE FIGURE 11 FOR A REPRESENTATION OF THE IMAGERY

Image	Original	OMP	L-POMP	OT-POMP
1	28.378	32.754	33.488	**34.619**
2	28.352	31.376	32.273	**32.428**
3	28.369	30.849	31.548	**32.676**
4	28.374	32.299	33.112	**33.273**
5	28.364	31.333	32.326	**32.508**

TABLE IV

STARTING PSNR VALUES FOR FIVE IMAGES AND DENOISED PSNR VALUES FOR THE THREE METHODS USING A LEARNED DICTIONARY. SEE FIGURE 11 FOR A REPRESENTATION OF THE IMAGERY

Image	Original	OMP	L-POMP	OT-POMP
1	28.378	33.831	34.239	**34.786**
2	28.352	31.996	32.205	**33.232**
3	28.369	30.541	30.767	**31.525**
4	28.374	33.246	33.397	**34.645**
5	28.364	32.314	32.506	**33.629**

V. DISCUSSION

A novel modification to the well-known matching pursuit algorithm has been proposed using path-based dictionary expansion/augmentation at each iteration. Theoretical results guarantee improved reconstruction error after a fixed number of iterations under certain assumptions. Furthermore, in the case of an orthonormal, under-determined dictionary the theoretical assumptions are trivially satisfied. Existing theory has been proved for the case of a linear path-based approach and extensions to generic paths are underway because it is likely that the current heuristic (same-sign sufficient condition) is sub-optimal for the OT-based path.

This work has demonstrated added benefit to a path-based OMP approach for both $k-$sparse signal reconstruction and image denoising. Results show improved output PSNR using the path-augmented approach on both structured and unstructured dictionaries. Additionally, for all considered sparsity levels there is a reduced reconstruction error obtained using the proposed modification. We note that this augmentation can be integrated into any MP-type approach.

From a mathematical perspective image reconstruction and denoising (as considered in this paper) are based on the assumption that the image patches belong to some underlying manifold. Learned dictionaries are a way of trying to produce a denser sampling of the global, underlying manifold. The path augmented approach, however, uses paths to more densely sample the manifold in local regions. In future work we hope

to exploit this interpretation to make statements about the distributions of the points sampled from the manifold using path-based approaches as well as learned samplings. Further, this interpretation will contextualize our research relative to other existing manifold models like that of [27].

Rigorous comparison to performance on overcomplete dictionaries, including statistical statements of when the assumptions of proven theory will hold for overcompleteness, is a point of continuing research. We also seek to formulate our approach as an alternative to overcomplete learned dictionaries for addressing the basis mismatch problem. Additionally, error bounds and convergence rates for k-sparse signal reconstruction will be considered as well as the application of POMP to compressive sensing.

APPENDIX

A. Proof of Theorem 3.3

Theorem 1.1 Straddle Theorem: Let $||d_1|| = ||d_2|| = ||r|| = 1$, $\langle r, d_1\rangle = r^\top d_1 = \cos\theta_1, \langle r, d_2\rangle = r^\top d_2 = \cos\theta_2$ *and* $\langle d_1, d_2\rangle = d_1^\top d_2 = \cos\theta_{1,2}$. *If* $|\langle r, d_1\rangle| > |\langle r, d\rangle|$ *for all* $d \in \mathcal{D}$, $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$ *and* $\cos\theta_{1,2} \in [0, \cos\theta_1/\cos\theta_2)$, *then there exists* $t_\star \in (0, 1)$ *such that* $|\langle r, d_{t_\star}\rangle| > |\langle r, d_1\rangle|$.

Proof: Let $||d_1|| = ||d_2|| = ||r|| = 1$, $\langle r, d_1\rangle = r^\top d_1 = \cos\theta_1$, $\langle r, d_2\rangle = r^\top d_2 = \cos\theta_2$ and $\langle d_1, d_2\rangle = d_1^\top d_2 = \cos\theta_{1,2}$. Assume that $|\langle r, d_1\rangle| > |\langle r, d\rangle|$ for all $d \in \mathcal{D}$, $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$ and $\cos\theta_{1,2} \in [0, \cos\theta_1/\cos\theta_2)$. Define $\alpha = \cos\theta_2/\cos\theta_1$. Note that $0 \le \alpha < 1$ is implied by $|\langle r, d_1\rangle| > |\langle r, d_2\rangle|$ together with $\text{sgn}(\cos\theta_1) = \text{sgn}(\cos\theta_2)$. Further, we have $1 - \alpha^2 > 0$. Using these facts together with our assumptions we have

$$\cos\theta_{1,2} < \alpha \tag{16}$$
$$0 < 2\alpha - 2\cos\theta_{1,2} \tag{17}$$
$$0 < 1 - \alpha^2 < 2\alpha - 2\cos\theta_{1,2} + 1 - \alpha^2 \tag{18}$$
$$0 < \frac{1-\alpha^2}{2\alpha - 2\cos\theta_{1,2} + 1 - \alpha^2} < 1. \tag{19}$$

From the property of reals we have that there exists some number m such that

$$\frac{1-\alpha^2}{2\alpha - 2\cos\theta_{1,2} + 1 - \alpha^2} < m < 1. \tag{20}$$

Set $t_\star = m \in (0, 1)$. Consequently,

$$1 - \alpha^2 < t_\star(2\alpha - 2\cos\theta_{1,2} + 1 - \alpha^2) \tag{21}$$
$$\implies 0 < t_\star(2\alpha - 2\cos\theta_{1,2} + 1 - \alpha^2) + \alpha^2 - 1 \tag{22}$$
$$\implies 0 < 2t_\star(\alpha - \cos\theta_{1,2}) + (1 - t_\star)(\alpha^2 - 1). \tag{23}$$

Rewriting in terms of the angles we have

$$0 < 2t_\star\left(\frac{\cos\theta_2}{\cos\theta_1} - \cos\theta_{1,2}\right) + (1 - t_\star)\left(\frac{(\cos\theta_2)^2}{(\cos\theta_1)^2} - 1\right). \tag{24}$$

Multiplying through by $(\cos\theta_1)^2$ yields

$$0 < 2t_\star(\cos\theta_2\cos\theta_1 - \cos\theta_{1,2}(\cos\theta_1)^2) + (1 - t_\star)((\cos\theta_2)^2 - (\cos\theta_1)^2). \tag{25}$$

Expanding and writing in terms of vector multiplication we have

$$0 < t_\star(r^\top d_2 d_1^\top r + r^\top d_1 d_2^\top r) - 2t_\star\cos\theta_{1,2}(r^\top d_1 d_1^\top r) + (1 - t_\star)(r^\top d_2 d_2^\top r - r^\top d_1 d_1^\top). \tag{26}$$

Multiply through by the positive value

$$\frac{1 - t_\star}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2} \tag{27}$$

to produce

$$0 < \frac{t_\star(1 - t_\star)(r^\top d_2 d_1^\top r + r^\top d_1 d_2^\top r)}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2} - \frac{2t_\star(1 - t_\star)\cos\theta_{1,2}(r^\top d_1 d_1^\top r)}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2} + \frac{(1 - t_\star)^2(r^\top d_2 d_2^\top r - r^\top d_1 d_1^\top)}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2}. \tag{28}$$

Next, add $0 = t_\star^2 r^\top d_1 d_1^\top r - t_\star^2 r^\top d_1 d_1^\top$ to the numerator and regroup terms to get

$$0 < \frac{t_\star^2 r^\top d_1 d_1^\top r + t_\star(1 - t_\star)(r^\top d_2 d_1^\top r + r^\top d_1 d_2^\top r)}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2} + \frac{(1 - t_\star)^2(r^\top d_2 d_2^\top r)}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2} - \frac{(t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2) r^\top d_1 d_1^\top}{t_\star^2 + 2t_\star(1 - t_\star)\cos\theta_{1,2} + (1 - t_\star)^2}. \tag{29}$$

From Lemma 3.1 we have that the first two lines of Eq. 29 can be simplified to $r^\top \mathbf{P}_{d_{t_\star}} r$ and the last line simplifies to $r^\top \mathbf{P}_{d_1} r$. Thus, we have

$$0 < r^\top \mathbf{P}_{d_{t_\star}} r - r^\top \mathbf{P}_{d_1} r \tag{30}$$
$$0 < ||\mathbf{P}_{d_{t_\star}} r||^2 - ||\mathbf{P}_{d_1} r||^2 \tag{31}$$
$$\implies ||\mathbf{P}_{d_1} r||^2 < ||\mathbf{P}_{d_{t_\star}} r||^2 \tag{32}$$

Finally, using Lemma 3.2 we have

$$|\langle r, d_1\rangle| < |\langle r, d_{t_\star}\rangle|. \tag{33}$$

Thus, we have that there exists a $t_\star \in (0, 1)$ such that $d_{t_\star}$ is "closer" to r than d_1. □

B. Proof of Corollary 3.2

Corollary 1.1 Positive Maximizer: For each pair $\{d_1, d_2\} \in \mathcal{D}$ *if* $|\langle r, d_1\rangle| > |\langle r, d_2\rangle|$ *and* $sgn(\cos\theta_1) = sgn(\cos\theta_2)$ *there is a single positive* t *(given by a closed form) that maximizes* $||\mathbf{P}_{d_{t_\star}} r||^2 - ||\mathbf{P}_{d_1} r||^2$. *Explicitly*

$$t_\star = \frac{\left(\cos\theta_1 - \cos\theta_2\cos\theta_{1,2}\right)\left(\csc\left(\frac{\theta_{1,2}}{2}\right)\right)^2}{2\left(\cos\theta_1 + \cos\theta_2\right)}. \tag{34}$$

Proof: We seek to maximize $||\mathbf{P}_{d_t} r||^2 - ||\mathbf{P}_{d_1} r||^2$ over all feasible t. Using Lemma 3.1 together with Eq. 30 we can write

$$f(t) = ||\mathbf{P}_{d_t} r||^2 - ||\mathbf{P}_{d_1} r||^2 \tag{35}$$
$$f(t) = r^\top \mathbf{P}_{d_t} r - r^\top \mathbf{P}_{d_t} r. \tag{36}$$

Equivalently,

$$f(t) = \frac{t^2(\cos\theta_1)^2 + 2t(1-t)\cos\theta_1\cos\theta_2}{t^2 + 2t(1-t)\cos\theta_{1,2} + (1-t)^2} + \frac{(1-t)^2(\cos\theta_2)^2}{t^2 + 2t(1-t)\cos\theta_{1,2} + (1-t)^2} + (\cos\theta_1)^2. \quad (37)$$

The function $f(t)$ is differentiable and has two critical points given by

$$t_1 = \frac{1}{1 - \cos\theta_1 \sec\theta_2} \text{ and} \quad (38)$$

$$t_2 = \frac{\left(\cos\theta_1 - \cos\theta_2\cos\theta_{1,2}\right)\left(\csc\left(\frac{\theta_{1,2}}{2}\right)\right)^2}{2\left(\cos\theta_1 + \cos\theta_2\right)}. \quad (39)$$

This can be verified by hand or through the use of software. When $\operatorname{sgn}(\cos\theta_1) = \operatorname{sgn}(\cos\theta_2)$ and $|\langle r,\, d_1\rangle| > |\langle r,\, d_2\rangle|$ it can be seen that $t_1 < 0$ and $t_2 > 0$. Further, it can be shown that for all values of $\cos\theta_1, \cos\theta_2$, and $\cos\theta_{1,2}$ that $f''(t_2) < 0$ implying that t_2 is a local maximum of $f(t)$ and the global maximum when $t \geq 0$. □

ACKNOWLEDGMENT

This work was supported in part through an OSD LUCI Fellowship and a Jerome and Isabella Karle Distinguished Scholar Fellowship Program.

REFERENCES

[1] *AT&T Laboratories Cambridge Database of Faces*. Accessed: Sep. 15, 2017. [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[2] S. Basu, S. Kolouri, and G. K. Rohde, "Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 9, pp. 3448–3453, 2014.

[3] F. Bergeaud and S. Mallat, "Matching pursuit of images," in *Proc. Int. Conf. Image Process.*, vol. 1, Oct. 1995, pp. 53–56.

[4] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss, "What's the frequency, Kenneth?: Sublinear Fourier sampling off the grid," in *Proc. Int. Workshop Approximation Algorithms Combinat. Optim.*, in Lecture Notes in Computer Science, vol. 7408, 2012, pp. 61–72.

[5] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[6] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[7] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

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

[9] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[10] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[11] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[12] T. H. Emerson, T. Doster, and C. Olson, "Path orthogonal matching pursuit for k-sparse image reconstruction," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1955–1959.

[13] T. Doster, T. Emerson, and C. Olson, "Path orthogonal matching pursuit for sparse reconstruction and denoising of SWIR maritime imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 1242–12427.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2012.

[15] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *Int. J. Comput. Vis.*, vol. 60, no. 3, pp. 225–240, 2004.

[16] S. Kolouri, S. Park, M. Thorpe, D. Slepčev, and G. K. Rohde, "Transport-based analysis, modeling, and learning from signal and data distributions," 2016, *arXiv:1609.04767*. [Online]. Available: https://arxiv.org/abs/1609.04767

[17] S. Kolouri, S. R. Park, and G. K. Rohde, "The radon cumulative distribution transform and its application to image classification," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 920–934, Feb. 2016.

[18] S. Kolouri, S. R. Park, M. Thorpe, D. Slepčev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 43–59, Jul. 2017.

[19] S. Kolouri and G. K. Rohde, "Transport-based single frame super resolution of very low resolution face images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4876–4884.

[20] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

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

[22] C. V. McLaughlin, J. M. Nichols, and F. Bucholtz, "Basis mismatch in a compressively sampled photonic link," *IEEE Photon. Technol. Lett.*, vol. 25, no. 23, pp. 2297–2300, Dec. 1, 2013.

[23] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[24] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.

[25] J. M. Nichols, A. K. Oh, and R. M. Willett, "Reducing basis mismatch in harmonic signal recovery via alternating convex search," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 1007–1011, Aug. 2014.

[26] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Nov. 1993, pp. 40–44.

[27] G. Peyré, "Manifold models for signals and images," *Comput. Vis. Image Understand.*, vol. 113, no. 2, pp. 249–260, 2009.

[28] N. Rao, P. Shah, S. Wright, and R. Nowak, "A greedy forward-backward algorithm for atomic norm constrained minimization," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 5885–5889.

[29] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressive sensing off the grid," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2012, pp. 778–785.

[30] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[31] C. Villani, *Optimal Transport: Old and New*, vol. 338. New York, NY, USA: Springer, 2008.

[32] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

Tegan H. Emerson received the B.S. degree in mathematics from Oregon State University in 2011 and the M.S. and Ph.D. degrees in mathematics from Colorado State University in 2013 and 2017, respectively. She is currently a Jerome and Isabella Karle Distinguished Scholar Fellow and a Mathematician in applied optics with the Naval Research Laboratory. Her research interests include geometric and topological data analysis, dimensionality reduction, algorithms for image and video processing, and optimization. She was a recipient of the Best Paper Award at the 2016 Workshop on Hyperspectral Imaging and Signal Processing: Evolutions in Remote Sensing. She was a member of the American delegation with the Heidelberg Laureate Forum in 2014.

Colin C. Olson received the B.S. degree in mechanical engineering from Colorado State University in 2003 and the M.S. and Ph.D. degrees in structural engineering from the University of California at San Diego in 2005 and 2008, respectively, where he was a recipient of the Los Alamos National Laboratory and National Defense Science and Engineering Graduate Fellowships. He was a National Research Council Postdoctoral Fellow with the Naval Research Laboratory, where he is currently a Staff Research Scientist with the Optical Sciences Division. He was awarded the DoD Laboratory University Collaboration Initiative (LUCI) grant in 2017 to support work in image processing and compressed sensing. His research interests include computational imaging, reinforcement learning, nonlinear dynamics, novel solutions to inverse problems, and the application of machine learning and deep neural networks to signal processing and image analysis.

Timothy Doster received the B.S. degree in computational mathematics and the M.S. degree in applied and computational mathematics from the Rochester Institute of Technology in 2008 and 2009, respectively, and the Ph.D. degree in applied mathematics and scientific computing from the University of Maryland, College Park, in 2014. From 2014 to 2016, he was a Jerome and Isabella Karle Distinguished Scholar Fellow before becoming a Permanent Research Scientist in applied optics with the Naval Research Laboratory. His research interests include deep, machine, and reinforcement learning as well as computational imaging, hyperspectral filter design, target and anomaly detection, and data fusion.