Path Orthogonal Matching Pursuit for k-Sparse Image Reconstruction

Tegan H. Emerson
Optical Sciences Division
U.S. Naval Research Laboratory
Washington, DC, USA
tegan.emerson@nrl.navy.mil

Timothy Doster
Optical Sciences Division
U.S. Naval Research Laboratory
Washington, DC, USA
timothy.doster@nrl.navy.mil

Colin Olson
Optical Sciences Division
U.S. Naval Research Laboratory
Washington, DC, USA
colin.olson@nrl.navy.mil

***Abstract*—We introduce a path-augmentation step to the standard orthogonal matching pursuit algorithm. Our augmentation may be applied to any algorithm that relies on the selection and sorting of high-correlation atoms during an analysis or identification phase by generating a "path" between the two highest-correlation atoms. Here we investigate two types of path: a linear combination (Euclidean geodesic) and a construction relying on an optimal transport map (2-Wasserstein geodesic). We test our extension by generating k-sparse reconstructions of faces using an eigen-face dictionary learned from a subset of the data. We show that our method achieves lower reconstruction error for fixed sparsity levels than either orthogonal matching pursuit or generalized orthogonal matching pursuit.**

***Index Terms*—matching pursuit, basis mismatch, optimal transport, k-sparse representation, signal reconstruction**

I. Introduction

Compressive sensing (CS) is a methodology that enables higher-resolution digital sampling of natural phenomena by leveraging good signal models to reconstruct signals that are undersampled according to classical Nyquist sampling theory [3], [9], [12]. In this case, "good" models are those that can sparsely represent signals as linear combinations of relatively few atoms drawn from a 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. CS theory predicts, to a level of probabilistic certainty, successful reconstruction of an undersampled signal when the underlying true signal satisfies upper limits on sparsity relative to a given dictionary [4], [8], [17]. As a result, significant effort has been spent on designing dictionaries or developing algorithms that are capable of learning dictionaries that are highly representative of the expected signal class [1], [11].

A persistent problem exists, however, because even if the underlying signal model could perfectly represent the signal with a single atom, the atoms must be discretely sampled from the model and therefore, with high probability, will fail to represent any given signal component exactly. For example, a 1-D sinusoidal signal composed of a single tone is well-represented by a sinusoidal signal 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 [24]. Guarantees on successful reconstruction of undersampled signals begin to fail when the assumption of sparsity is violated which can lead to the introduction of artifacts at best or complete failure to reconstruct at worst. This "basis mismatch" problem has been considered in the literature [7], [10], [14], [29], [31] with some success in the case of 1-D sinusoidal signals where a search over the frequencies residing between the two Fourier atoms with the largest correlation can find the exact representative atom [27].

Here, we propose two methods for constructing better exemplars from an underlying dictionary: 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) [18]. This "path-based" augmentation can be applied to any reconstruction algorithm that relies on the selection and sorting of high-correlation atoms during an analysis or identification phase. In particular, we consider the matching pursuit (MP) family of algorithms [6], [23] which contains a number of algorithm variations predicated on the selection of high-correlation atoms.

We illustrate here that augmenting MP with our path-based modification leads to lower reconstruction errors for fixed sparsity levels. We demonstrate these improvements by constructing sparse face representations from a learned eigen-face dictionary. Such an unstructured dictionary, although not state-of-the-art, works well for illustrating and testing concepts on an easily understood data set. We compare our method to traditional MP algorithm variations and show significantly reduced k-sparse reconstruction errors.

We begin with a brief summary of matching pursuit and describe paths between atoms in Section II. In Section III we present the proposed algorithm Path Orthogonal Matching Pursuit (POMP). Next, in Section IV we describe a popular face data set and the experiment it is used in to evaluate algorithm performance. Results of the experiment are presented in Section V. Finally, we conclude with a discussion of the results as well as directions of future work in Section VI.

II. Background and Related Work

A. Matching Pursuit

For a fixed degree of sparsity, consideration of all possible atom combinations of that order is computationally intractable.

ISBN 978-90-827970-1-5

A popular work-around of this combinatorial optimization problem is an algorithm called *Matching Pursuit (MP)* [23]. 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. The 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. This residual image is then fit in the same greedy way, updated, and the process repeats. k iterations of MP yields a k-sparse representation with some associated final error/residual $\mathbf{R}_k$.

Several variations of MP exist [13], [22], [25], [26], [28], but in general 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 (as well as OMP derivatives) is that in OMP the residual is updated by projecting the image onto the orthogonal complement of the span of the current support which has been shown to produce better results.

B. Paths Between Atoms

A fundamental component of our approach is the construction of a path between the two closest dictionary atoms. A *path* is a smooth map from the closest dictionary atom to the second-closest dictionary atom that is parameterized by a variable $t \in [0, 1]$. Explicitly, a path p is defined as

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

$$\text{s.t.} \quad p(\mathbf{D}_1, \mathbf{D}_2, 0) = \mathbf{D}_1 \quad \text{and} \quad p(\mathbf{D}_1, \mathbf{D}_2, 1) = \mathbf{D}_2. \tag{2}$$

An example of a familiar path is the line segment with endpoints $\mathbf{D}_1$ and $\mathbf{D}_2$ given by

$$\mathbf{D} = (1 - t)\mathbf{D}_1 + (t)\mathbf{D}_2, \tag{3}$$

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

Another path between images that has gained notoriety in image processing is induced by an Optimal Transport (OT) map. Two main versions of OT exist: (1) the Monge OT problem 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 OT problem which allows for intensities at starting pixels to be split among multiple destination pixels [32].

Paths produced by OT between images have yielded impressive results and insights in image registration and warping [16], super-resolution of low-resolution face images [21], and cell morphology [5]. In recent years a computationally-efficient approximation to the solution of Monge OT based on the Radon Cumulative Distribution Transform (RCDT) has been developed and has shown reduced computational time and increased performance on multiple tasks in machine learning, signal processing, and image classification [19], [20]. Due to the success of these methods, we use the path induced by the RCDT approximate solution to Monge OT.

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

$$\mathbf{V} = \begin{bmatrix} \vec{v}_{1,1} & \cdots & \vec{v}_{1,m} \\ \vdots & \ddots & \vdots \\ \vec{v}_{n,1} & \cdots & \vec{v}_{n,m} \end{bmatrix} \tag{4}$$

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

$$p_V(\mathbf{D}_1, \mathbf{D}_2, t) = t\mathbf{V}(\mathbf{D}_1, \mathbf{D}_2) \tag{5}$$

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 path can be thought of as a set of pixel-wise linear trajectories approximating a globally nonlinear path between images (the 2-Wasserstein geodesic).

Other viable paths between the atoms exist but at present we narrow the focus to the two path types defined above. Given a path between two dictionary atoms we search along the path for a novel atom which is closer to the test signal than either of the path end-points. If such a novel atom exists it takes the place of the single best atom in OMP. Details of the proposed algorithm are presented in the subsequent section. An example of samples along linear and OT paths between two images, as well as the angle formed between the path samples and the test image, are shown in Fig. 1.

III. The Algorithm: POMP

Path Orthogonal Matching Pursuit (POMP) is a modification to the well-known OMP algorithm. Instead of finding a single nearest dictionary element and removing its contribution, the two closest dictionary elements are chosen at each iteration. A path is formed which moves between the two closest dictionary atoms and an optimal atom is found along the path. Here, optimality is defined as having the largest-magnitude inner product with the test image, i.e., the smallest angle between the pair of images when considered as vectors. In OMP, only the magnitude of the inner product is considered. 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 choose to work with atoms sharing the same signed inner product with the residual. Two atoms with opposite signed inner products can still be used by multiplying either by -1.

Let $\mathbf{D}_1$ be the dictionary atom having the largest-magnitude inner product with the test image, $\mathbf{T}$. Define $\mathbf{D}_2$ to be the second-closest dictionary atom to our test image. The simplest, and perhaps most natural, form of a path is linear. In Section II we describe a globally linear path (3) and a pixel-wise linear path (5). When the linear path is used to choose an optimal atom within an iteration of POMP we refer to it as L-POMP.

ISBN 978-90-827970-1-5 © EURASIP 2018

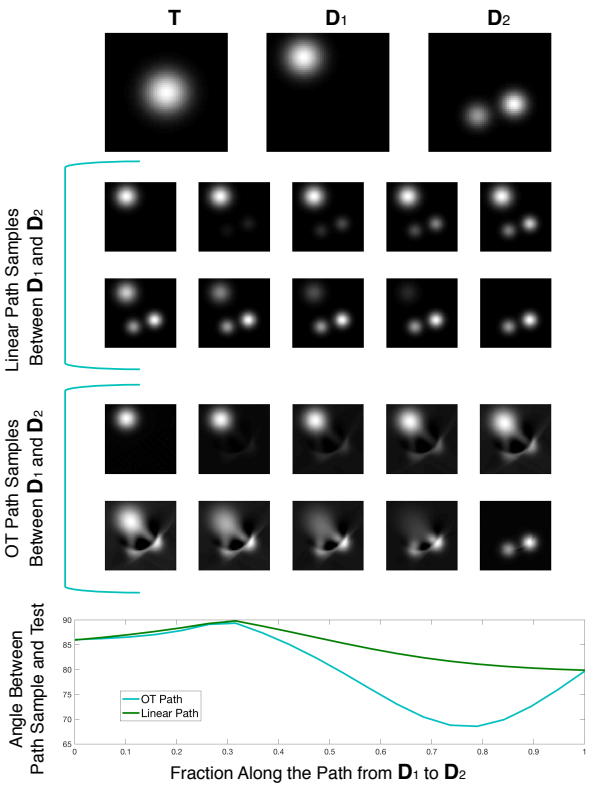

Fig. 1. Examples of samples along the paths between two image atoms and the angle between the labeled test image and the path samples. Images along the linear path (Euclidean geodesic) are characterized by the presence of intensity that matches the combined support from $\mathbf{D}_1$ and $\mathbf{D}_2$ and simply shifts intensity magnitude from the support represented by $\mathbf{D}_1$ to the support represented by $\mathbf{D}_2$ as t increases. Conversely, the intensity support shifts between images along the OT path (2-Wasserstein geodesic) as t increases.

The 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

$$\theta_t = \cos^{-1}\left(\frac{\langle \mathbf{D}_t, \mathbf{T}\rangle}{||\mathbf{D}_t||_F||\mathbf{T}||_F}\right), \tag{6}$$

where $t \in [0,1]$ parameterizes the distance along the path from $\mathbf{D}_1$ to $\mathbf{D}_2$ and $\|A\|_F$ 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 given by

$$\mathbf{D}_{t^*} = p(\mathbf{D}_1, \mathbf{D}_2, t^*) \quad \text{where} \quad t^* = \arg\min_{t\in[0,1]} \theta_t. \tag{7}$$

Pseudocode for POMP is provided in Algorithm 1. 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 same procedure is performed for the D_2 in lines 8-9. The signs of the inner products of the first and second closest (identified in line 8) atoms are identified and matched, in lines 10. The optimal atom along the path between the two nearest neighbors is selected in lines 11-12 and is then appended to the support in line 13. Residual updates and updating of indexing variables are performed in lines 14-17. It should be noted that standard OMP is simply lines 1-6 and 13-19 where $D^* = D_1$.

It is important to note that without further constraints there is no guarantee of a nontrivial minimum angle being found along the path due to the general non-convexity of the inner product [15]. With added constraints on the equations governing $\mathbf{D}_t$ it may be possible to prove the existence of a nontrivial minimum along the path. A rigorous study of these necessary and sufficient conditions is a focus of ongoing and future work. Explicit formulas for determining the optimal path parameter may also exist.

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. For this reason when an orthogonal dictionary is used, a linear path-based MP algorithm will be equivalent to a linear path-based OMP algorithm. When guarantees about orthogonality along a path cannot be made, the reconstruction using MP can differ from the OMP reconstruction. The proposed algorithm can be seamlessly combined with the many OMP variants as well.

Algorithm 1: 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, $\{d_k^1\}_{k=1}^K$, $\{d_k^2\}_{k=1}^K$ the vectors of first and 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 $\quad\quad \mathbf{D}_2 \leftarrow \arg\max_{\mathbf{D}\in\mathcal{D}\setminus\mathbf{D}_1} |\langle \mathbf{D}, \mathbf{T}\rangle|$;
9 $\quad\quad d_k^2 \leftarrow \text{index}(\mathbf{D}_2)$;
10 $\quad\quad \mathbf{D}_2 \leftarrow \text{sgn}(\langle D_1, T\rangle)\text{sgn}(\langle D_2, T\rangle)\mathbf{D}_2$;
11 $\quad\quad t_k \leftarrow \arg\min_{t\in[0,1]} \theta_t$;
12 $\quad\quad \mathbf{D}^* \leftarrow \text{path}(\mathbf{D}_1, \mathbf{D}_2, t_n)$;
13 $\quad\quad \mathbf{S} \leftarrow \text{augment}(\mathbf{S}, \mathbf{D}^*)$;
14 $\quad\quad \mathbf{P}_S \leftarrow \mathbf{S}(\mathbf{S}^\top\mathbf{S})^{-1}\mathbf{S}^\top$;
15 $\quad\quad \mathbf{X} \leftarrow \mathbf{P}_S\mathbf{T}$;
16 $\quad\quad \mathbf{R}_{k+1} \leftarrow (\mathbf{I} - \mathbf{P}_S)\mathbf{T}$;
17 $\quad\quad k \leftarrow k + 1$;
18 $\quad$ **end**
19 **end**

IV. Data and Experimental Design

The data consist of 10 images (with varying lighting conditions and facial expressions and orientations) of 40 subjects

ISBN 978-90-827970-1-5 © EURASIP 2018

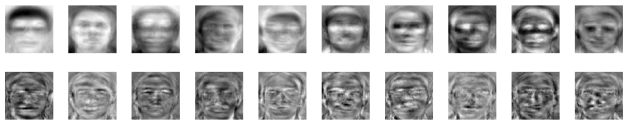

Fig. 2. Examples from the Eigen-dictionary built from 300 randomly chosen faces. Top row contains the first 10 eigen-faces (global facial structures) while the bottom row contains the $40-50^{th}$ eigen-faces (finer scale/local structures).

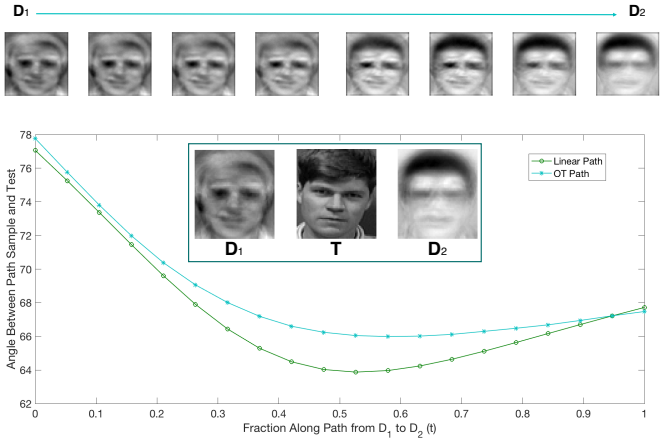

Fig. 3. The top set of images are discrete samples along the OT path between atoms $\mathbf{D}_1$ and $\mathbf{D}_2$. In the bottom plot we see the angle between samples along the two paths and the test image labeled $\mathbf{T}$.

for a total of 400 images [2]. We generate an eigen-face (EF) dictionary by randomly selecting 300 images and performing principal component analysis to get a dictionary of size 300. The remaining 100 images will be designated as the testing set. Although there may be an example of a subject's face in both the training and testing sets, all results presented are based on reconstruction of images not used in the generation of the eigen-dictionary. As a further challenge, all 10 images of one subject were placed in the test set.

Beginning in the late 1980s the use of EFs for tasks in facial recognition, detection, and reconstruction became widespread. The canonical paper on the use of EFs for image reconstruction is [30] to which we refer interested readers for details on construction of eigen-dictionaries. A dictionary comprised of EFs captures both global and local features. The first (dominant) EFs correspond to global structures while later EFs represent finer-scale structures. Thus, such a dictionary is appropriate for an application of OMP since the residuals being fit in consecutive iterations can be thought of as fitting finer-scale structures which may be appropriately represented by the dictionary. Although EFs are not state-of-the-art the authors believe them to be a sufficient proof of concept. Examples of the EF dictionary atoms are shown in Fig. 2.

A novel face from the testing set is approximated as a linear combination of dictionary atoms. The terms "EF" and "dictionary atom" will be used interchangeably when referring to the facial image reconstruction challenge. Performance is measured by the relative error of the signal estimate corresponding to a fixed sparsity level. All of the path-based results shown were generated by comparing the image being fit to 20 discrete samples along a given path (linear or OT). Examples of samples along a path between two dictionary elements can be seen in Fig. 3 as well as the angles between the test image (residual fit on the first iteration) along the two paths.

For each test image four MP variations are performed and the relative error is computed at each iteration/sparsity level. The four variations implemented are OMP, Generalized OMP (GOMP) [33] with two atoms amended to the support at each iteration, L-POMP, and OT-POMP. The choice to compare to GOMP with two atoms is due to it being the most conceptually similar existing OMP variant to the proposed method.

V. Results

Evaluation of the performance of the four variations of MP (OMP, GOMP, L-POMP, OT-POMP) is measured in terms of the reconstruction error for the $k-$sparse approximation, i.e. $||\mathbf{R}_k||_F$. In Fig. 4 we see the relative error as a function of the first 100 iterations of the OMP variants averaged over 100 test images. The performance of the path-based approaches show dramatic improvement over both OMP and GOMP. For several iterations L-POMP achieves lower reconstruction error than OT-POMP but the distinction is negligible when compared to the performance differential relative to standard OMP. Note that it can take OMP 2-3 times as many iterations to achieve the same error as L-POMP or OT-POMP.

The face data set analyzed is noise-free and all of the faces are well-registered on a naturally occurring coordinate system. In the absence of these ideal circumstances, it is likely that the difference in performance between L-POMP and OT-POMP will diverge. As an example we look to the experiment presented in Fig. 1. The test image in this setting is a single large Gaussian being compared to two other images consisting of either one or two different Gaussians. Included in the figure are samples along the OT and linear paths between the two atoms. For the linear path the intensity of the Gaussians in the upper left corner and bottom right merely fade in and out as t increases, the intensity support remains unchanged. Alternatively, on the OT path intensity support moves from the upper-left corner through the center of the image to the bottom right. Angles formed between the path images and the test image are also shown. A nontrivial minimum is found along the OT path but not along the linear path. These results are suggestive that in compression or reconstruction problems where there is no inherited coordinate system (such as one might find with a learned dictionary generated by k-SVD) OT-POMP could greatly improve the quality of reconstruction.

VI. Discussion

POMP reconstructs facial images with lower relative errors than both the OMP and GOMP on the considered data set. When averaged over 100 trials there is as much as a 2-3 fold reduction in the number of iterations needed to obtain the same relative error using a path-based method.

ISBN 978-90-827970-1-5 © EURASIP 2018

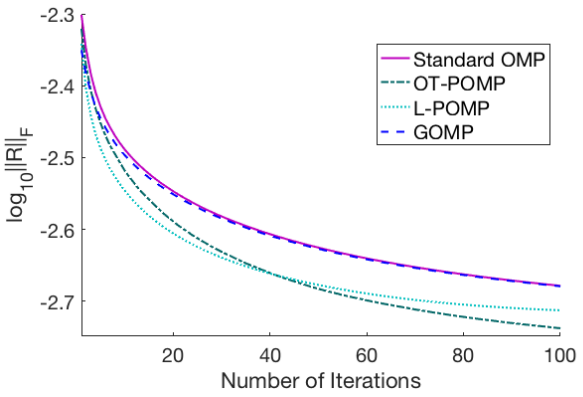

Fig. 4. The relative reconstruction error as a function of iterations for the four algorithms considered. Results over 100 iterations are shown.

The two versions of POMP do not demonstrate significant performance differences relative to one another in the facial reconstruction task for low sparsity levels. However, the similarity in performance between L-POMP and OT-POMP may be specific to the data set as a toy example indicates that there can be substantial differences between the two approaches.

Application of POMP to image denoising and patch-based reconstructions will allow for a more comprehensive comparison to existing MP-based algorithms. Additionally, we intend to explore path-based searches as an alternative to domain adaptation and dictionary augmentation algorithms. Furthermore, we intend to perform a detailed computational complexity analysis of POMP so that the gains of implementation may be appropriately qualified.

We plan to develop theory for when novel minimums along the path exist. Necessary and/or sufficient conditions under which there exist closed-form solutions for determination of both the optimal path parameter value and its corresponding atom are underway and may lead to reduced computational expense. Future work will also include extensions of the path-based approach for tasks in compressed sensing.

Acknowledgements: The authors would like to acknowledge support from the DoD LUCI program and the Naval Research Laboratory Jerome and Isabella Karle Distinguished Scholar Fellowship.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. rmk-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[2] AT&T Laboratories Cambridge Database of Faces. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[3] R. Baraniuk. Compressive sensing. *IEEE Signal Process. Mag.*, 24(4):118–121, 2007.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[5] S. Basu, S. Kolouri, and G. Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proc. Natl. Acad. Sci.*, 111(9):3448–3453, 2014.

[6] F. Bergeaud and S. Mallat. Matching pursuit of images. In *Image Processing, 1995. Proceedings., International Conference on*, volume 1, pages 53–56. IEEE, 1995.

[7] P. Boufounos, V. Cevher, A. Gilbert, Y. Li, and M. Strauss. What's the frequency, Kenneth?: Sublinear Fourier sampling off the grid. *Lecture Notes in Computer Science*, 7408:61–72, 2012.

[8] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[9] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.

[10] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59(5):2182–2195, 2011.

[11] R. Coifman and M. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory*, 38(2):713–718, 1992.

[12] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[13] D. 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*, 58(2):1094–1121, 2012.

[14] C. Ekanadham, D. Tranchina, and E. Simoncelli. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE Trans. Signal Process.*, 59(10):4735–4744, 2011.

[15] G. Golub and C. Van Loan. *Matrix Computations*, volume 3. JHU Press, 2012.

[16] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[17] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Trans. Inf. Theory*, 52(9):4036–4048, 2006.

[18] S. Kolouri, S. Park, M. Thorpe, D. Slepčev, and G. Rohde. Transport-based analysis, modeling, and learning from signal and data distributions. *preprint*, arXiv:1609.04767, 2017.

[19] S. Kolouri, S. R. Park, and G. Rohde. The Radon Cumulative Distribution Transform and its application to image classification. *IEEE Trans. Image Process.*, 25(2):920–934, 2016.

[20] S. Kolouri, S. R. Park, M. Thorpe, D. Slepčev, and G. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Process. Mag.*, 34(4):43–59, 2017.

[21] S. Kolouri and G. Rohde. Transport-based single frame super resolution of very low resolution face images. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4876–4884, 2015.

[22] S. Kwon, J. Wang, and B. Shim. Multipath matching pursuit. *IEEE Trans. Inf. Theory*, 60(5):2986–3001, 2014.

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

[24] C. McLaughlin, J. Nichols, and F. Bucholtz. Basis mismatch in a compressively sampled photonic link. *IEEE Photon. Technol. Lett.*, 25(23):2297–2300, 2013.

[25] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[26] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2009.

[27] J. Nichols, A. Oh, and R. Willett. Reducing basis mismatch in harmonic signal recovery via alternating convex search. *IEEE Signal Process. Lett.*, 21(8):1007–1011, 2014.

[28] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44. IEEE, 1993.

[29] N. Rao, P. Shah, S. Wright, and R. Nowak. A greedy forward-backward algorithm for atomic norm constrained minimization. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5885–5889. IEEE, 2013.

[30] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *JOSA A*, 4(3):519–524, 1987.

[31] G. Tang, B. Bhaskar, P. Shah, and B. Recht. Compressive sensing off the grid. In *50th Annual Allerton Conference on Communication, Control, and Computing*, pages 778–785. IEEE, 2012.

[32] C. Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2008.

[33] J. Wang, S. Kwon, and B. Shim. Generalized orthogonal matching pursuit. *IEEE Trans. Signal Process.*, 60(12):6202–6216, 2012.

ISBN 978-90-827970-1-5 © EURASIP 2018