NEAR-ISOMETRIC LINEAR EMBEDDINGS OF MANIFOLDS

Chinmay Hegde, Aswin C. Sankaranarayanan, and Richard G. Baraniuk

ECE Department, Rice University, Houston TX 77005

ABSTRACT

We propose a new method for linear dimensionality reduction of manifold-modeled data. Given a training set $\mathcal{X}$ of Q points belonging to a manifold $\mathcal{M} \subset \mathbb{R}^N$, we construct a linear operator $P : \mathbb{R}^N \to \mathbb{R}^M$ that approximately preserves the norms of all $\binom{Q}{2}$ pairwise difference vectors (or secants) of $\mathcal{X}$. We design the matrix P via a trace-norm minimization that can be efficiently solved as a semi-definite program (SDP). When $\mathcal{X}$ comprises a sufficiently dense sampling of $\mathcal{M}$, we prove that the optimal matrix P preserves all pairs of secants over $\mathcal{M}$. We numerically demonstrate the considerable gains using our SDP-based approach over existing linear dimensionality reduction methods, such as principal components analysis (PCA) and random projections.

Index Terms— Adaptive sampling, Linear Dimensionality Reduction, Whitney's Theorem

1. INTRODUCTION

In many signal processing applications, we seek low-dimensional representations (or *embeddings*) of data that can be modeled as elements of a high-dimensional ambient space. The classical approach to construct such embeddings is principal components analysis (PCA), which involves *linearly* mapping the original N-dimensional data into the K-dimensional subspace spanned by the dominant eigenvectors of the data covariance matrix; typically, $K \ll N$. Over the last decade, more sophisticated *nonlinear* data embedding (or "manifold learning") methods have also emerged; see, for example, [1–4].

Linear dimensionality reduction is advantageous in several aspects. First, linear dimensionality reduction methods are marked by their *computational efficiency*: techniques such as PCA can be very efficiently performed using a singular value decomposition (SVD) of a linear transformation of the data. A second key appeal is their *generalizability*: linear methods produce smooth, globally defined mappings that can be easily applied to unseen, out-of-sample test data points.

Nevertheless, existing linear dimensionality reduction methods such as PCA are marked by the important shortcoming that the produced embedding potentially *distorts pairwise distances* between sample data points. This phenomenon is exacerbated when the data arises from a nonlinear submanifold of the signal space [5]. Due to this behaviour, two distinct points in the ambient signal space are often be mapped to a single point in the low-dimensional embedding space. This hampers the application of PCA-like techniques to some important problems such as reconstruction and parameter estimation of manifold-modeled signals.

An intriguing alternative to PCA is the method of *random projections*. Consider $\mathcal{X}$, a cloud of Q points in a high-dimensional Euclidean space $\mathbb{R}^N$. The Johnson-Lindenstrauss lemma [6] states that $\mathcal{X}$ can be linearly mapped to a subspace of dimension $M = \mathcal{O}(\log Q)$ with minimal distortion of the pairwise distances between the Q points (in other words, the mapping is *near-isometric*). Further, this linear mapping can be easily implemented in practice. One simply constructs a matrix $\Phi \in \mathbb{R}^{M \times N}$ with $M \ll N$ whose elements are randomly drawn from certain probability distributions. Then, with high probability, Φ is approximately isometric under a certain lower-bound on M [5, 7].

The method of random projections can be extended to more general signal classes beyond finite point clouds. For example, random linear projections provably preserve the isometric structure of compact, differentiable low-dimensional manifolds [8, 9], as well as the isometric structure of the set of sparse signals [10, 11]. Random projections are conceptually simple and useful in many signal processing applications. Yet, they too suffer from certain shortcomings. Their theoretical guarantees are probabilistic and asymptotic. Further, the mapping itself is independent of the data and/or task under consideration and hence cannot leverage the presence of labeled/unlabeled training data.

In this paper, we propose a novel method for *deterministic* construction of *linear, near-isometric* embeddings of data arising from a nonlinear manifold $\mathcal{M}$. Given a set of training points $\mathcal{X}$ belonging to $\mathcal{M}$, we consider the *secant manifold* $\mathcal{S}(\mathcal{X})$ that consists of all pairwise difference vectors of $\mathcal{X}$, normalized to lie on the unit sphere. Next, we formulate an affine-rank minimization problem (3) to construct a matrix Ψ that preserves the norms of all the vectors in $\mathcal{S}(\mathcal{X})$ up to a desired distortion parameter δ. This problem is known to be NP-hard, and so we perform a convex relaxation to obtain a trace-norm minimization (4), which can be efficiently solved as a semi-definite program (SDP). Once calculated, the matrix Ψ represents an approximately isometric linear embedding over *all* pairwise secants of the original manifold $\mathcal{M}$, given a sufficiently high number of training points (see Proposition 2 below).

We show that our approach is particularly useful for efficient signal reconstruction and parameter estimation using a very small number of samples. Further, by carefully pruning the secant set $\mathcal{S}(\mathcal{X})$, we can extend our approach to other, more general inference tasks (such as supervised binary classification). Several numerical experiments in Section 4 demonstrate the advantages of our approach over PCA as well as random projections.

2. MANIFOLD EMBEDDINGS

The basis for our approach stems from the observation that any smooth K-dimensional submanifold $\mathcal{M}$ can be mapped down to a Euclidean space of dimension $2K + 1$ such that the geometric structure of $\mathcal{M}$ is preserved. This is succinctly captured by two celebrated results in differential topology.

Theorem 1 (Whitney) [12] *Let $\mathcal{M}$ be a K-dimensional compact manifold. Then there exists a $\mathcal{C}^\infty$ embedding of $\mathcal{M}$ in $\mathbb{R}^{2K+1}$.*

Theorem 2 (Nash) [12] *Let $\mathcal{M}$ be a K-dimensional Riemannian manifold. Then there exists a $\mathcal{C}^1$ isometric embedding of $\mathcal{M}$ in $\mathbb{R}^{2K+1}$.*

An important notion in the proofs of Theorems 1 and 2 is the normalized *secant manifold* of $\mathcal{M}$:

$$\mathcal{S}(\mathcal{M}) = \left\{ \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|_2}, \ x, x' \in \mathcal{M}, \mathbf{x} \neq x' \right\}. \quad (1)$$

The secant manifold $\mathcal{S}(\mathcal{M})$ forms a $2K$-dimensional submanifold of the $(N-1)$-dimensional unit sphere in $\mathbb{R}^N$. Any element of the unit sphere can be equated to a projection from $\mathbb{R}^N$ to $\mathbb{R}^{N-1}$. Therefore, by choosing a projection direction that does *not* belong to $\mathcal{S}(\mathcal{M})$, one can map $\mathcal{M}$ into $\mathbb{R}^{N-1}$ injectively (i.e., without overap). If $2K \leq N - 1$, this is always possible.

Theorem 1 (Whitney's (weak) Embedding Theorem) is based upon this intuition. Whitney's Theorem shows that there exists a low-dimensional projective embedding of $\mathcal{M}$ in $\mathbb{R}^{2K+1}$ such that no two points of $\mathcal{M}$ are mapped to the same point in $\mathbb{R}^{2K+1}$. Theorem 2 (Nash's Embedding Theorem) makes a stronger claim: the manifold $\mathcal{M}$ can in fact be *isometrically* mapped into $\mathbb{R}^{2K+1}$, i.e., the distance between every pair of distinct points on $\mathcal{M}$ is preserved by the mapping. This mapping is nonlinear and has been shown to be the solution of a system of nonlinear partial differential equations [12]. Indeed, isometry (as a notion of stability) is a crucial and desirable property for practical applications.

Unfortunately, the proofs of both Theorem 1 and Theorem 2 are *non-constructive* and thus do not lend themselves to easy implementation in practice. In Section 3, we take some initial steps in this direction. We develop an efficient computational framework to produce a low-dimensional linear embedding that isometrically preserves the secant set of a manifold $\mathcal{M}$.

Our proposed method is not the first attempt to develop stable linear embeddings for secant manifolds. Specifically, we build upon and improve the *Whitney Reduction Network* (WRN) approach for computing auto-associative graphs [13]. The WRN approach is a greedy, heuristic technique algorithmically very similar to PCA. Our approach follows a completely different path: the optimization formulation (4) is based on convex programming and is guaranteed to produce a near-isometric, linear embedding. Further, our constructed linear embedding enjoys provable global isometry guarantees over the entire manifold $\mathcal{M}$.

3. ISOMETRIC LINEAR EMBEDDINGS

Given a manifold $\mathcal{M} \subset \mathbb{R}^N$, we wish to find a linear embedding $\mathcal{P} : \mathbb{R}^N \to \mathbb{R}^M$, $M \ll N$, such that the embedded manifold $\mathcal{P}\mathcal{M}$ is isometric (or approximately isometric) to $\mathcal{M}$, i.e., for $\mathbf{x}, \mathbf{x}' \in \mathcal{M}$, $\mathbf{x} \neq \mathbf{x}'$, $\mathcal{P}$ satisfies

$$1 - \delta \leq \frac{\|\mathcal{P}\mathbf{x} - \mathcal{P}\mathbf{x}'\|_2^2}{\|\mathbf{x} - \mathbf{x}'\|_2^2} \leq 1 + \delta \quad (2)$$

where $\delta > 0$ represents the distortion factor. As a practical assumption, suppose that we are given access to the training data $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Q\}$ sampled from $\mathcal{M}$. We first form the secant set $\mathcal{S}(\mathcal{X})$ using (1) to obtain a set of $Q' = \binom{Q}{2}$ unit vectors

$$\mathcal{S}(\mathcal{X}) = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{Q'}\}.$$

We seek a *measurement matrix* $\Psi \in \mathbb{R}^{M \times N}$ of low rank that preserves the norms of the secants of $\mathcal{X}$ up to a distortion parameter $\delta > 0$, such that $\left| \|\Psi\mathbf{v}_i - 1\|_2 \right|^2 < \delta$, for all secants $\mathbf{v}_i$ in $\mathcal{S}(\mathcal{X})$. Following [8], we will refer to δ as the desired *isometry constant*.

Define $P \doteq \Psi^T\Psi$. Clearly, P is a positive semi-definite, symmetric matrix. Following the technique in [14], we can formulate P as the solution to the matrix recovery problem:

$$\begin{aligned} \text{minimize} \quad & \text{rank}(P) \quad (3) \\ \text{subject to} \quad & P \succeq 0, \ P = P^T, \\ & 1 - \delta \leq \mathbf{v}_i^T P \mathbf{v}_i \leq 1 + \delta, \ \mathbf{v}_i \in \mathcal{S}(\mathcal{X}). \end{aligned}$$

In general, rank minimization is NP-hard, and so we instead propose solving a trace-norm relaxation of (3):

$$\begin{aligned} \text{minimize} \quad & \text{Tr}(P) \quad (4) \\ \text{subject to} \quad & P \succeq 0, \ P = P^T, \\ & 1 - \delta \leq \mathbf{v}_i^T P \mathbf{v}_i \leq 1 + \delta, \ \mathbf{v}_i \in \mathcal{S}(\mathcal{X}). \end{aligned}$$

The matrix recovery problem (4) consists of a linear objective function as well as linear inequality constraints and hence can be efficiently solved via a semi-definite program (SDP) [15]. The only inputs to (4) are the Q training vectors $\mathbf{v}_i$ sampled from the manifold and the desired isometry constant $\delta > 0$.

We observe that the optimization in (4) is *always feasible*. This is because the identity matrix I always satisfies the constraints in (4) regardless of the underlying manifold $\mathcal{M}$, the training set $\mathcal{X}$, and the distortion parameter δ. Therefore, by invoking the main result of [16], we obtain that the rank of the optimum P^* is no greater than $\sqrt{2Q'}$. We summarize this more precisely as follows.

Proposition 1 *Let r^* be the rank of the optimum to the semi-definite program (4). Then,*

$$r^* \leq \left\lceil \frac{\sqrt{8|\mathcal{S}(\mathcal{X})| + 1} - 1}{2} \right\rceil$$

Once the solution $P^* = \mathbf{U}\Lambda\mathbf{U}^T$ to (4) is found, the desired linear embedding Ψ can be calculated using a simple matrix square root:

$$\Psi = \Lambda_M^{1/2}\mathbf{U}_M^T, \quad (5)$$

where $\Lambda_M = \text{diag}\{\lambda_1, \ldots, \lambda_M\}$ denotes the M leading (non-zero) eigenvalues of P^* for any $M \leq r^*$. In this manner, we obtain a low-rank matrix $\Psi \in \mathbb{R}^{M \times N}$ that is approximately isometric on the secant set of the training samples $\mathcal{S}(\mathcal{X})$, up to a distortion δ. We claim that if the training set $\mathcal{X}$ represents a sufficiently dense sampling of $\mathcal{M}$, then Ψ is approximately isometric for the *entire manifold* $\mathcal{M}$. This statement is a direct consequence of the ϵ-cover construction method in Section 3.2.2 in [8]. We summarize this claim as follows, with proof omitted for brevity.

Proposition 2 *Suppose the training set $\mathcal{X}$ is an ϵ-cover of $\mathcal{M}$, i.e., for every $\mathbf{m} \in \mathcal{M}$, there exists an $\mathbf{x} \in \mathcal{X}$ such that $\min_{\mathbf{x} \in \mathcal{X}} d_{\mathcal{M}}(\mathbf{m}, \mathbf{x}) \leq \epsilon$. Then, the optimum Ψ obtained by solving (4) and (5) satisfies the approximate isometry condition (2) on $\mathcal{M}$ with distortion factor δ, where $\delta = C\epsilon$ and C is a constant that depends only on the volume and curvature of $\mathcal{M}$.*

Summarizing, the SDP (4) results in a measurement matrix $\Psi \in \mathbb{R}^{M \times N}, M \ll N$, that preserves approximate isometry over all pairwise secants of a given manifold. We will dub Ψ as the *SDP Secant* measurement matrix.

Task adaptivity. We observe that the matrix inequality constraints in (4) are derived by enforcing an approximate isometry condition on *all* pairwise secants $\{\mathbf{v}_i\}_{i=1}^{Q'}$. However, this can often prove to be too restrictive. For example, consider a supervised classification scenario where the signal of interest $\mathbf{x}$ can arise from one of two classes that are modeled by low-dimensional manifolds $\mathcal{M}, \mathcal{M}'$. The goal is to infer the signal class of $\mathbf{x}$ from the linear embedding $P\mathbf{x}$. Here, the measurement matrix Ψ should be designed such that signals from different classes are mapped to sufficiently distant points in the low-dimensional space. However, it does not matter if signals from the same class are mapped to the same point. Geometrically speaking, we seek P that preserves merely the *inter-class secants* $\mathcal{S}(\mathcal{M}, \mathcal{M}') = \left\{ \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|_2}, \ x \in \mathcal{M}, \ x' \in \mathcal{M} \right\}$ up to an isometry constant δ. This represents a reduced set of linear constraints in (4). Therefore, the solution space to (4) is larger, and the optimal P^* will necessarily be of lower rank, thus leading to a further reduction in the dimension of the embedding space.

4. EXPERIMENTS

We demonstrate the applicability of our approach using simulated data. We solve the SDP optimization (4) using a Lagrange multiplier method similar to the Singular Value Thresholding algorithm [15]. First, we consider a synthetic manifold $\mathcal{M}$ that comprises binary images of shifts of a white disk on a black background (see Fig. 2 (a)). We construct a training database of $Q' = 900$ secants from this manifold using (1), normalize the secants, and solve (4) with desired isometry constant $\delta = 0.05$ to obtain a positive semi-definite symmetric matrix P^* (for this problem, $N = 256$). We then form a rank-M matrix according to (5), compute linear embeddings of 1000 test secants, and calculate the empirical isometry constants on the test secants.

We repeat the empirical estimation of isometry constants obtained by projecting on to the PCA basis functions learned
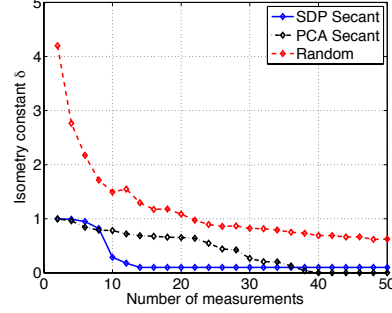

Fig. 1: *Empirical isometry constant δ vs. number of measurements M using various types of embeddings. The SDP secant approach ensures global approximate isometry using the fewest number of measurements.*

on the Q' secants (dubbed as *PCA secants*), as well as random Gaussian projections. Figure 1 plots the variation of the isometry constant δ with the number of measurements M averaged over all the test secants. Our proposed SDP Secant embedding Ψ achieves the desired isometry on the secants using the fewest number of measurements. Interestingly, both the SDP Secant embeddings as well as the PCA Secant embeddings far outperform the random projection approach.

Next, we consider a supervised classification problem, where our classes consist of binary images of shifts of a translating disk and a translating square (some examples are shown in Fig. 2). We construct a training dataset of $Q' = 2000$ *inter-class* secants and obtain a measurement matrix Ψ_{inter} via our proposed SDP Secant approach. Using a small number of measurements of a test signal, we estimate the class label using a Generalized Maximum Likelihood Classification (GMLC) approach following the framework in [17]. We repeat the above classification experiment using measurement basis functions learned via PCA on the inter-class secants, as well as random projections. Figure 2(c) plots the variation of the number of measurements M vs. the probability of error. Again, we observe that the SDP approach outperform PCA as well as random projections.

Finally, we test our approach on a more challenging dataset. We consider the CVDomes dataset, a collection of simulated X-band radar scattering databases for civilian vehicles [18]. The database for each vehicle class consists of (complex-valued) short-time signatures of length $N = 128$ for the entire $360°$ azimuth, sampled every $0.0625°$. We model each of the two classes as a one-dimensional manifold, construct $Q' = 2000$ training inter-class secants as well as , and compute the SDP Secant measurement matrix Ψ_{inter} using our proposed SDP approach. Additionally, we obtain a different measurement matrix Ψ_{joint} from a training dataset of $Q' = 2000$ vectors that comprise both inter- and intra-class secants. Given a new test signal, we acquire M linear measurements and perform maximum likelihood classification. From Fig. 3, we infer that the SDP approach using inter-class secants yields the best classification performance among the different measurement techniques. In particular, Ψ_{inter} produces the best classification performance, proving the potential benefits of task-adaptivity using our approach.
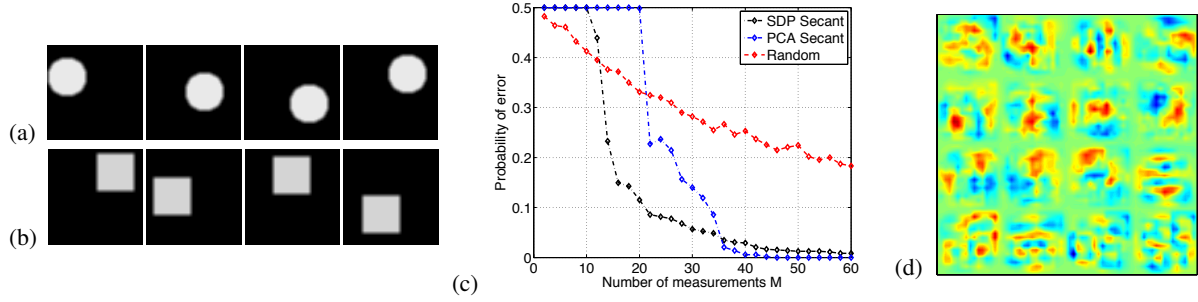
Fig. 2: *Binary classification from low-dimensional linear embeddings. (a,b) The signals of interest comprise shifted images of a white disk/square on a black background. We observe M linear measurements of a test image using different matrices, and classify the observed samples using a GMLC approach. (c) Observed probability of classification error as a function of M. Our SDP Secant approach yields high classification rates using very few measurements. (d) Illustrations of 16 linear embedding basis functions obtained by our proposed approach, tiled in a 4x4 grid.*
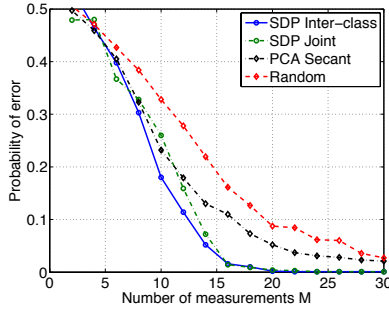
Fig. 3: *Classification of a Toyota Camry versus a Nissan Maxima using M linear measurements of length-128 radar signatures. The SDP approach produces $> 95\%$ classification rates using only a small number of measurements $M \leq 15$.*

5. DISCUSSION

We have taken some initial steps towards a new method for constructing near-isometric, low-dimensional linear embeddings of manifold-modeled signals and images. Our proposed SDP Secant embedding preserves the norms of all pairwise secants of a high-resolution sampling of the manifold and is calculated via a novel semi-definite programming formulation (4). Our method can be easily adapted to perform inference tasks such as classification. We have empirically demonstrated that our method is superior to current linear embedding methods, including PCA and random projections.

Two key challenges remain. First, our proposed approach relies on the efficiency of the trace norm as a valid proxy for the matrix rank in the objective function in (4). The precise conditions under which the optimum of (4) equals the optimum of (3) need to be carefully studied. Second, current SDP solvers do not perform very well beyond signal sizes $N > 500$ and constraints $Q' > 2000$. However, in many important situations, the signals are very high-dimensional and the amount of available training data is large. It is likely that specialized high-performance solvers for (4) need to be developed. We defer both issues to future research.

6. REFERENCES

[1] J. Tenenbaum, V.de Silva, and J. Landford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[2] D. Donoho and C. Grimes, "Hessian Eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Natl. Acad. Sci.*, vol. 100, no. 10, pp. 5591–5596, 2003.

[3] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[4] K. Weinberger and L. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Computer Vision*, vol. 70, no. 1, pp. 77–90, 2006.

[5] D. Achlioptas, "Database-friendly random projections," in *Proc. Symp. Principles of Database Systems (PODS)*, Santa Barbara, CA, May 2001.

[6] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Modern Anal. and Prob.*, New Haven, CT, Jun. 1982.

[7] P. Indyk, "Stable distributions, pseudorandom generators, embeddings and data stream computation," *J. ACM*, vol. 53, no. 3, pp. 307–323, 2006.

[8] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[9] K. Clarkson, "Tighter bounds for random projections of manifolds," in *Proc. Symp. Comp. Geom.* ACM, 2008, pp. 39–48.

[10] E. Candès, "Compressive sampling," in *Proc. Int. Congress of Math.*, Madrid, Spain, Aug. 2006.

[11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[12] M. Hirsch, *Differential topology*, Springer, 1976.

[13] D. Broomhead and M. Kirby, "The Whitney Reduction Network: A method for computing autoassociative graphs," *Neural Comput.*, vol. 13, pp. 2595–2616, 2001.

[14] E. Candes, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Arxiv preprint arXiv:1109.4499*, Sept. 2011.

[15] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Math. Prog.*, vol. 128, no. 1, pp. 321–353, 2011.

[16] A. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete and Comput. Geometry*, vol. 13, no. 1, pp. 189–202, 1995.

[17] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symp. Elec. Imag.: Comp. Imag.*, San Jose, CA, Jan. 2007.

[18] K. Dungan, C. Austin, J. Nehrbass, and L. Potter, "Civilian vehicle radar data domes," in *Proc. SPIE Algo. Synth. Aperture Radar Imagery*, Apr. 2010.