Learning Measurement Matrices for Redundant Dictionaries

Chinmay Hegde
Massachusetts Institute of Technology
Email: chinmay@csail.mit.edu

Aswin Sankaranarayanan
Carnegie Mellon University
Email: saswin@ece.cmu.edu

Richard Baraniuk
Rice University
Email: richb@rice.edu

Sparsity as a powerful instrument in signal processing is now commonplace. However, it is also well known that certain classes of signals do *not* admit a sparse expansion in an orthonormal basis (e.g., a mixture of spikes and sinuoids is non-sparse in either the canonical or Fourier basis). Therefore, it is typical to use an *overcomplete basis*, or a *redundant dictionary*, for representing such complicated signals. Mathematically, $\mathbf{x} \in \mathbb{R}^N$ is k-sparse in a dictionary $\mathbf{D} \in \mathbb{R}^{n \times N}$, where $n < N$, if $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha} \in \mathbb{R}^N$ contains only k nonzeros.

Compressive Sensing (CS) [1] encompasses the development of efficient techniques for sampling and reconstruction of sparse signals. A signal $\mathbf{x} \in \mathbb{R}^n$ may be sampled by inner products with $m < n$ vectors; therefore, $\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}\mathbf{D}\boldsymbol{\alpha}$, where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ is the *measurement matrix*. Here, the matrix $\mathbf{\Phi}\mathbf{D}$ is sometimes called the *holographic basis*. Signal reconstruction can be performed using a slew of algorithms; see, for example, [2]. It is generally accepted that the *mutual coherence* μ (or the maximum absolute off-diagonal entry of the Gram matrix) of the holographic basis plays an important role in reconstruction performance; smaller values of μ typically lead to better reconstruction.

An important consideration is the choice of measurement matrix $\mathbf{\Phi}$. The typical CS approach offers a very simple, *universal* solution: construct $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ elementwise by *randomly* drawing from a Gaussian (or Bernoulli) probability distribution. Remarkably, if $\mathbf{x}$ is k-sparse, then with high probability, $m = \mathcal{O}(k \log n)$ samples suffice in order to ensure efficient, stable reconstruction of $\boldsymbol{\alpha}$ (and consequently, $\mathbf{x}$) from $\mathbf{y}$; in other words, m needs only to be linear in the sparsity level k and logarithmic in the actual signal length n.

While randomized constructions of measurement matrices are agnostic to the dictionary $\mathbf{D}$ under consideration, the question remains whether one can do better, i.e., whether one can construct a hypothetical $\mathbf{\Phi}$ with an even fewer number of rows m by leveraging the intrinsic structure of $\mathbf{D}$. In this work, we answer this question in the affirmative. We develop an algorithmic framework for *learning* measurement matrices $\mathbf{\Phi}$ that are well-tuned to the dictionary under consideration. Our framework can be viewed as a variant of *NuMax* [3], a new convex optimization framework for designing near-isometric linear embeddings of high-dimensional point clouds.

A brief sketch of our approach is as follows. Consider a sparsifying dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_N]$ with unit-norm columns. We seek a matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$, with as few rows as possible, such that the mutual coherence of the holographic basis $\mathbf{\Phi}\mathbf{D}$ is at most a scalar parameter $\mu > 0$. To avoid numerical degeneracies, we also impose the constraints that the columns of $\mathbf{\Phi}\mathbf{D}$ themselves be approximately unit-norm. Define $\mathbf{P} = \mathbf{\Phi}^T\mathbf{\Phi}$ so that $\text{rank}(\mathbf{P}) = m$. Then, $\mathbf{P}$ can be posed as the solution to the problem:

$$\begin{aligned} \text{minimize} \quad & \text{rank}(\mathbf{P}) \\ \text{subject to} \quad & |\mathbf{d}_i^T \mathbf{P} \mathbf{d}_j| \leq \mu, \quad i \neq j, \\ & \mathbf{d}_i^T \mathbf{P} \mathbf{d}_i \geq 1 - \mu, \quad \mathbf{P} \succeq 0. \end{aligned} \tag{1}$$

Note that (1) consists of inequality constraints that are *linear* in the optimization variable $\mathbf{P}$. Since rank minimization is NP-hard, we relax (1) to obtain a *nuclear norm* minimization problem, subject to linear inequality constraints, over the cone of positive semidefinite (PSD) matrices. This problem is convex and can be solved very efficiently, for example, using a simple modification of Algorithm 1 in [3]. We dub this modified algorithm *NuMax-Dict*. A simple matrix square root of the optimal $\mathbf{P}^*$ reveals the desired $\mathbf{\Phi}$, modulo an orthogonal transformation. The choice of the parameter μ is important; larger values of μ result in optimal matrices $\mathbf{P}^*$ of smaller rank, and consequently, a smaller number m of rows in $\mathbf{\Phi}$.

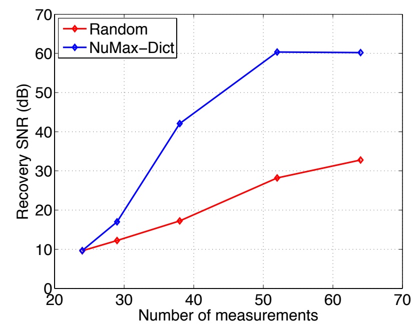

Fig. 1. *CS recovery performance for random projections versus the matrix produced by our proposed algorithm (NuMax-Dict). NuMax-Dict far outperforms random Gaussian projections in terms of recovered signal SNR.*

Figure 1 illustrates the benefits of CS recovery using our approach. We consider a generic dictionary $\mathbf{D}$ populated with i.i.d. Gaussian entries with parameters $n = 64$, $N = 128$ and solve the nuclear norm relaxation of (1) using NuMax-Dict for a given parameter μ. We generate 500 coefficient vectors $\boldsymbol{\alpha}$ with $k = 5$ nonzeros and random amplitudes, form signals $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, and obtain m compressive samples using both the learned measurement matrix $\mathbf{\Phi}$ as well as a random Gaussian measurement matrix with m rows. We reconstruct the signals using orthogonal matching pursuit (OMP) [2] and record the error in signal recovery in terms of SNR. We repeat this experiment for different values of input μ. As illustrated in Fig. 1, the learned matrix using our proposed NuMax-Dict algorithm outperforms conventional random sampling over a wide range of measurement regimes. Our proposed formulation (1) can be adapted to other notions of matrix coherence, such as average coherence. Further, it is applicable to a number of practical settings, such as the compressive acquisition of images that are sparse in large-scale redundant dictionaries. We explore these avenues further in the full version of this work.

References

[1] R. Baraniuk, "Compressive sensing," *IEEE Sig. Proc. Mag.*, vol. 24, no. 4, pp. 118–120, 124, 2007.

[2] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[3] C. Hegde, A. Sankaranarayanan, W. Yin, and R. Baraniuk, "A convex approach for learning near-isometric linear embeddings," Nov. 2012, preprint, available at http://dsp.rice.edu/publications.