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 sinusoids 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, \( x \in \mathbb{R}^n \) is \( k \)-sparse in a dictionary \( D \in \mathbb{R}^{n \times N} \), where \( n < N \), if \( x = D\alpha \) where \( \alpha \in \mathbb{R}^N \) contains only \( k \) nonzeros.

Compressive Sensing (CS) [1] encompasses the development of efficient techniques for sampling and reconstruction of sparse signals. A signal \( x \in \mathbb{R}^n \) may be sampled by inner products with \( m < n \) vectors; therefore, \( y = \Phi x = \Phi D\alpha \), where \( \Phi \in \mathbb{R}^{m \times n} \) is the measurement matrix. Here, the matrix \( \Phi D \) 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 \( \mu \) (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 \( \mu \) typically lead to better reconstruction.

An important consideration is the choice of measurement matrix \( \Phi \). The typical CS approach offers a very simple, universal solution: construct \( \Phi \in \mathbb{R}^{m \times n} \) elementwise by randomly drawing from a Gaussian (or Bernoulli) probability distribution. Remarkably, if \( x \) is \( k \)-sparse, then with high probability, \( m = \mathcal{O}(k \log n) \) samples suffice in order to ensure efficient, stable reconstruction of \( \alpha \) (and consequently, \( x \)) from \( y \); 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 \( D \) under consideration, the question remains whether one can do better, i.e., whether one can construct a hypothetical \( \Phi \) with an even fewer number of rows \( m \) by leveraging the intrinsic structure of \( D \). In this work, we answer this question in the affirmative. We develop an algorithmic framework for learning measurement matrices \( \Phi \) 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 \( D = [d_1, \ldots, d_N] \) with unit-norm columns. We seek a matrix \( \Phi \in \mathbb{R}^{m \times n} \), with as few rows as possible, such that the mutual coherence of the holographic basis \( \Phi D \) is at most a scalar parameter \( \mu > 0 \). To avoid numerical degeneracies, we also impose the constraints that the columns of \( \Phi D \) themselves be approximately unit-norm. Define \( P = \Phi^* \Phi \) so that \( \text{rank}(P) = m \). Then, \( P \) can be posed as the solution to the problem:

\[
\begin{align*}
\text{minimize} & \quad \text{rank}(P) \\
\text{subject to} & \quad |d_i^T P d_j| \leq \mu, \quad i \neq j, \\
& \quad d_i^T P d_i \geq 1 - \mu, \quad P \succeq 0.
\end{align*}
\]

Note that (1) consists of inequality constraints that are linear in the

---

**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.

**References**

