The Sparsity Gap:
Uncertainty Principles Proportional to Dimension

Joel A. Tropp
Computing and Mathematical Sciences
California Institute of Technology
Pasadena, CA 91125–5000
Email: jtropp@acm.caltech.edu

Abstract—In an incoherent dictionary, most signals that admit a sparse representation admit a unique sparse representation. In other words, there is no way to express the signal without using strictly more atoms. This work demonstrates that sparse signals typically enjoy a higher privilege: each nonoptimal representation of the signal requires far more atoms than the sparsest representation—unless it contains many of the same atoms as the sparsest representation. One impact of this finding is to confer a certain degree of legitimacy on the particular atoms that appear in a sparse representation. This result can also be viewed as an uncertainty principle for random sparse signals over an incoherent dictionary.

I. INTRODUCTION

The purpose of this paper is to develop a new class of uncertainty principles for sparse representation that hold even when the sparsity level approaches the ambient dimension. We begin with a discussion of the background and related results before moving on to the new contributions.

A. Sparse Representation in Dictionaries

Let $\mathbf{\Phi}$ be an $m \times N$ matrix with normalized columns:

$$\|\boldsymbol{\varphi}_j\|_2 = 1, \quad j = 1, 2, \ldots, N.$$

We refer to $\mathbf{\Phi}$ as a *dictionary* and to its columns as *atoms*. Assume the atoms span the ambient space $\mathbb{C}^m$.

There are two simple geometric quantities associated with a dictionary. The first is a measure of *redundancy*:

$$\rho = \|\mathbf{\Phi}\|^2,$$

where $\|\cdot\|$ denotes the spectral, or $\ell_2 \to \ell_2$ operator norm, of a matrix. We always have $\rho \geq N/m$. Equality holds if and only if $\mathbf{\Phi}$ is a *tight frame*. The second quantity is the *coherence*:

$$\mu = \max_{j \neq k} |\langle \boldsymbol{\varphi}_j, \, \boldsymbol{\varphi}_k \rangle|.$$

The coherence is small when the angle between each pair of atoms is large. Strohmer and Heath [1] have observed that

$$\mu \geq \sqrt{\frac{N - m}{m(N - 1)}}. \tag{1}$$

In the typical case $N \geq 2m$, the inequality (1) indicates that the coherence cannot be very small: $\mu \gtrsim m^{-1/2}$.

Let S be a subset of $\{1, 2, \ldots, N\}$, and define $\mathbf{\Phi}_S$ to be the column submatrix of $\mathbf{\Phi}$ whose columns are listed in S. We say

that S is *linearly independent* if it lists a linearly independent family of atoms. Note that $\mathbf{\Phi}_S$ is injective if and only if S is linearly independent. Suppose that a signal $\boldsymbol{u} \in \mathbb{C}^m$ can be written as

$$\boldsymbol{u} = \mathbf{\Phi}\boldsymbol{x} \quad \text{where supp}(\boldsymbol{x}) \subset S.$$

We call the vector $\boldsymbol{x}$ a *representation* of the signal $\boldsymbol{u}$, and we say that $\boldsymbol{u}$ can be *represented with* S. When S is linearly independent, $\boldsymbol{x}$ is the unique representation of $\boldsymbol{u}$ over S.

In a redundant dictionary ($N > m$), each signal has an infinity of representations. The sparse representation problem asks us to express $\boldsymbol{u}$ with the fewest number of atoms:

$$\min \|\boldsymbol{z}\|_0 \qquad \text{subject to} \qquad \boldsymbol{u} = \mathbf{\Phi}\boldsymbol{z}, \tag{2}$$

where $\|\cdot\|_0$ counts the number of nonzero components in its argument. If $\boldsymbol{x}$ is a minimizer of this mathematical program, the set $S = \text{supp}(\boldsymbol{x})$ must be linearly independent. Otherwise, we could remove an atom to obtain a sparser representation. As a result, when studying sparse representation, we focus on linearly independent sets of atoms.

B. Uniqueness of Sparse Representations

One might wonder when the problem (2) has a unique solution. The sparse approximation literature took up this inquiry about ten years ago, although one can trace some of the ideas to the late 1980s [2]. The early research led to the following result for deterministic signals.

Proposition 1: Assume that

$$|S| < \frac{1}{2}(\mu^{-1} + 1). \tag{3}$$

If a signal $\boldsymbol{u} = \mathbf{\Phi}\boldsymbol{x}$ with $\text{supp}(\boldsymbol{x}) \subset S$, then $\boldsymbol{x}$ is the unique minimizer of (2).

Donoho and Huo established this result for dictionaries consisting of two orthonormal bases [3]; Gribonval and Nielsen proved that it holds for *every* dictionary [4]. Subsequently, Donoho and Elad showed that Proposition 1 follows from a more general result, phrased in terms of the *Kruskal rank*, or *spark*, of the dictionary [5]. Another line of work [6], [4] sharpened the condition (3) for dictionaries consisting of multiple orthonormal bases. See [7] for a detailed discussion of the spikes and sines dictionary.

The requirement (3) is very stringent: it typically demands that the sparsity level $|S| \lesssim \sqrt{m}$. In spite of this apparent

shortcoming, the condition (3) cannot be improved in general. For example, when m is a perfect square, the Dirac comb can be represented perfectly using $\sqrt{m}$ spikes or $\sqrt{m}$ sines [2]. To move past the square-root threshold, we must place additional restrictions on the sparse signals we are willing to consider.

A natural approach is to introduce some randomness. Let S be linearly independent, and let $\boldsymbol{x} \in \mathbb{C}^S$ be a random vector whose distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{C}^S$. We say that a random signal of the form $\boldsymbol{u} = \boldsymbol{\Phi}_S \boldsymbol{x}$ is *generic*, and we refer to the (unique) representation of $\boldsymbol{u}$ over S as the *natural representation*.

To obtain interesting uniqueness results for generic signals, we impose some additional hypotheses. We say that the dictionary $\boldsymbol{\Phi}$ is a *weakly incoherent tight frame* if

$$\|\boldsymbol{\Phi}\|^2 = \frac{N}{m} \quad \text{and} \quad \mu \leq \frac{c}{\log N}, \tag{4}$$

where c is an absolute constant. Assume the sparsity level

$$s \leq \frac{cm}{\log N}. \tag{5}$$

In this setting, we have the following result.

Proposition 2: Assume the dictionary $\boldsymbol{\Phi}$ satisfies (4) and the sparsity s satisfies (5). Draw a uniformly random set S of s atoms from the dictionary. Except with probability $O(N^{-1})$, the following statement holds.

Let $\boldsymbol{u} = \boldsymbol{\Phi}_S \boldsymbol{x}$ be a generic signal. With probability one, the natural representation of $\boldsymbol{u}$ is the unique minimizer of (2).

Roughly speaking, Proposition 2 states that a generic sparse signal over a random set of atoms is unlikely to have any other representation that is equally sparse—even when the sparsity level is nearly proportional to the ambient dimension.

Candès and Romberg established the first theorem of this type in the specific case of the spikes and sines dictionary [8]. Using different methods, the present author showed that analogous results hold for any strongly incoherent dictionary [9, Sec. 7]. The extension to weakly incoherent dictionaries requires additional ideas from [10, Sec. 5].

C. Uncertainty Principles

Historically, the sparse approximation community has viewed uniqueness through the lens of *uncertainty principles*. Suppose that a signal has two (different) representations:

$$\boldsymbol{u} = \boldsymbol{\Phi}_S \boldsymbol{x} = \boldsymbol{\Phi}_T \boldsymbol{y}.$$

The Donoho–Elad dictionary uncertainty principle [5, Thms. 3 and 5] states[1] that

$$|S| + |T| > \mu^{-1}. \tag{6}$$

In particular, if a signal $\boldsymbol{u}$ can be represented with a set S that satisfies (3), then every alternative representation requires strictly more atoms.

Since the coherence usually satisfies $\mu \gtrsim m^{-1/2}$, the dictionary uncertainty principle only operates in the regime of very

[1]Donoho and Elad express their uncertainty principle [5, Thm. 3] in terms of the Kruskal rank of the dictionary, which is notoriously difficult to estimate. The result quoted here provides the best general bound.

sparse representations: $|S| \lesssim \sqrt{m}$. Except for very structured (or very random) dictionaries, it does not seem possible to obtain dictionary uncertainty principles for arbitrary signals that hold at sparsity levels near the ambient dimension.

D. The Sparsity Gap

This paper describes uncertainty principles for *generic signals*. It is easy to appreciate why generic signals might behave better than adversarially chosen signals. If there are small sets S and T of atoms for which $\text{range}(\boldsymbol{\Phi}_S)$ and $\text{range}(\boldsymbol{\Phi}_T)$ intersect, there exists a signal that has sparse representations over both S and T. (Witness the Dirac comb!) On the other hand, it is hard for a generic signal to have two sparse representations because $\text{range}(\boldsymbol{\Phi}_T)$ rarely *contains* $\text{range}(\boldsymbol{\Phi}_S)$! This fact offers a plausible route to reach uncertainty principles at sparsity levels far greater than $\sqrt{m}$.

Our first result extends the dictionary uncertainty principle (6) to generic signals. The proof appears in Section IV.

Theorem A (Sparsity Gap under Strong Incoherence): Suppose that S is linearly independent, and draw a generic signal $\boldsymbol{u}$ in $\text{range}(\boldsymbol{\Phi}_S)$. Then, almost surely, we cannot represent $\boldsymbol{u}$ with a set T disjoint from S *unless*

$$|S| + |T| > \mu^{-1} \sqrt{|S|}.$$

When $|S| = 1$, our result coincides with the dictionary uncertainty principle (6), but it becomes increasingly strict requirements as the sparsity level $|S|$ increases! Indeed, an equivalent condition is

$$|T| > \left(\frac{\mu^{-1}}{\sqrt{|S|}} - 1 \right) \cdot |S|,$$

so a generic signal that uses $|S| \ll \mu^{-2}$ atoms cannot be represented with any disjoint set T of atoms unless $|T| \gg |S|$. In the extreme case where $\mu^{-2} \sim m$, we obtain an uncertainty principle that operates at sparsity levels proportional to the ambient dimension!

Our second result is an uncertainty principle that parallels Proposition 2, just as Theorem A parallels Proposition 1. This proof appears in Section V.

Theorem B (Sparsity Gap under Weak Incoherence): Assume that $\boldsymbol{\Phi}$ is a weakly incoherent tight frame that satisfies (4), and assume further that $N > 2m$. Suppose that S is a randomly chosen set of s atoms, where s satisfies (5). Except with probability $O(N^{-1})$, the following holds.

Draw a generic signal $\boldsymbol{u}$ in $\text{range}(\boldsymbol{\Phi}_S)$. Then, almost surely, $\boldsymbol{u}$ cannot be represented with a set T disjoint from S *unless*

$$|T| > \left(1 + \frac{2}{\rho} \right) \cdot |S|.$$

The redundancy $\rho = N/m$, by hypothesis.

In words, Theorem B considers a generic signal over a random set of atoms. It is likely that every (disjoint) alternative representation requires a constant factor more atoms than the natural representation, where the extra factor decreases as the dictionary becomes more redundant. We see that there is

typically a *sparsity gap* between the natural representation and the the sparsest representation that uses different atoms.

This result provides an interesting guarantee for a huge class of dictionaries because of the weak bound for the incoherence. On the other hand, it holds for a smaller class of signals than Theorem A because we have randomized the set of atoms in addition to choosing generic coefficients.

II. RANK AND FILE

Although a generic signal has many representations aside from the natural one, there is a large class of representations that we can almost surely rule out. As a first step toward our main results, we develop an algebraic condition that describes which representations can and cannot occur.

To motivate the discussion, let us recall a standard argument for establishing dictionary uncertainty principles. Suppose that both S and T are linearly independent. A few moments of thought reveals that the following conditions are equivalent:

1) We have $\mathrm{range}(\mathbf{\Phi}_S) \cap \mathrm{range}(\mathbf{\Phi}_T) = \emptyset$.
2) The matrix $\mathbf{\Phi}_R$ has full rank, where $R = S \cup T$.

For a fixed set S, suppose that Condition 2) holds whenever $|R| < r_\star$. We conclude that, if there exists a signal that has representations over both S and T, then $|S| + |T| \geq r_\star$. Read the paper [5] to see this argument in action.

We can extend this methodology by *quantifying* the rank of the matrix $\mathbf{\Phi}_R$. These bounds allow us to count how many extra atoms are needed to represent a generic sparse signal.

Lemma 3: Suppose that both S and T are linearly independent. The following conditions are equivalent.

1) We have $\mathrm{range}(\mathbf{\Phi}_S) \cap \mathrm{range}(\mathbf{\Phi}_T) \subsetneq \mathrm{range}(\mathbf{\Phi}_S)$.
2) We have $|T| < \mathrm{rank}(\mathbf{\Phi}_R)$, where $R = S \cup T$.

Proof: Define the subspaces $\mathscr{S} = \mathrm{range}(\mathbf{\Phi}_S)$ and $\mathscr{T} = \mathrm{range}(\mathbf{\Phi}_T)$, which implies that $\mathscr{S} + \mathscr{T} = \mathrm{range}(\mathbf{\Phi}_R)$. Note that $\mathscr{S} \cap \mathscr{T}$ is a proper subspace of $\mathscr{S}$ if and only if

$$\dim(\mathscr{S} \cap \mathscr{T}) < \dim(\mathscr{S}) \tag{7}$$

The algebra of subspaces yields

$$\dim(\mathscr{S} \cap \mathscr{T}) = \dim(\mathscr{S}) + \dim(\mathscr{T}) - \dim(\mathscr{S} + \mathscr{T}).$$

Therefore, the condition (7) is equivalent with

$$\dim(\mathscr{T}) < \dim(\mathscr{S} + \mathscr{T}).$$

Since T is linearly independent, $\dim(\mathscr{T}) = |T|$. Meanwhile,

$$\dim(\mathscr{S} + \mathscr{T}) = \dim(\mathrm{range}(\mathbf{\Phi}_R)) = \mathrm{rank}(\mathbf{\Phi}_R).$$

This is the required conclusion. ∎

Let us translate the previous result from the language of subspaces to the language of probability.

Corollary 4: Suppose that both S and T are linearly independent. The following conditions are equivalent.

1) A generic signal $\boldsymbol{u} = \mathbf{\Phi}_S \boldsymbol{x}$ almost surely has no representation of the form $\boldsymbol{u} = \mathbf{\Phi}_T \boldsymbol{y}$.
2) We have $|T| < \mathrm{rank}(\mathbf{\Phi}_R)$, where $R = S \cup T$.

Proof: Lemma 3 states that Condition 2) is the same as

$$\mathrm{range}(\mathbf{\Phi}_S) \cap \mathrm{range}(\mathbf{\Phi}_T) \subsetneq \mathrm{range}(\mathbf{\Phi}_S), \tag{8}$$

so we prove that Condition 1) is the same as (8). To that end, let $\boldsymbol{u} = \mathbf{\Phi}_S \boldsymbol{x}$ be a generic signal, which means that $\boldsymbol{x}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{C}^S$. Let ν denote the Lebesgue measure on $\mathrm{range}(\mathbf{\Phi}_S)$.

First, assume (8) holds. A proper subspace has zero Lebesgue measure, so

$$\nu(\mathrm{range}(\mathbf{\Phi}_S) \cap \mathrm{range}(\mathbf{\Phi}_T)) = 0.$$

The set S is linearly independent, so $\mathbf{\Phi}_S$ is injective. As a result, the distribution of $\boldsymbol{u}$ is absolutely continuous with respect to ν. It follows immediately that

$$\mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\mathbf{\Phi}_S) \cap \mathrm{range}(\mathbf{\Phi}_T)\} = 0.$$

We conclude that

$$\mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\mathbf{\Phi}_T)\} = 0$$

because the signal $\boldsymbol{u} \in \mathrm{range}(\mathbf{\Phi}_S)$.

Conversely, suppose (8) is false. Then $\mathrm{range}(\mathbf{\Phi}_S) \subset \mathrm{range}(\mathbf{\Phi}_T)$, so the signal $\boldsymbol{u}$ can be represented over T. ∎

It is convenient to remove the assumption of linear independence from the previous result.

Corollary 5: Suppose that S is linearly independent, and let T be any other set of atoms. Assume that

$$|T| < \mathrm{rank}(\mathbf{\Phi}_R), \quad \text{where } R = S \cup T.$$

Draw a generic signal $\boldsymbol{u} = \mathbf{\Phi}_S \boldsymbol{x}$. Then

$$\mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\mathbf{\Phi}_T)\} = 0.$$

Proof: When T is linearly independent, the claim follows directly from Corollary 4. Otherwise, extract a maximal linear independent subset T' from T, and write $R' = S \cup T'$. Apply the result to T' to obtain the statement

$$|T'| < \mathrm{rank}(\mathbf{\Phi}_{R'}) \quad \Longrightarrow \quad \mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\mathbf{\Phi}_{T'})\} = 0.$$

Since T' is maximal, $\mathrm{rank}(\mathbf{\Phi}_{R'}) = \mathrm{rank}(\mathbf{\Phi}_R)$ and also $\mathrm{range}(\mathbf{\Phi}_{T'}) = \mathrm{range}(\mathbf{\Phi}_T)$. To complete the proof, note that the hypothesis $|T| < \mathrm{rank}(\mathbf{\Phi}_R)$ implies $|T'| < \mathrm{rank}(\mathbf{\Phi}_R)$ because $T' \subset T$. ∎

III. ANALYTIC RANK ESTIMATES

The main challenge is that we only possess analytic/geometric information about the dictionary, encapsulated in the redundancy ρ and the coherence μ. But the rank is fundamentally an algebraic quantity. Our approach will be to construct *analytic* estimates for the rank that we can compute from the data at hand.

A. Schatten Norms

A primary tool is the Schatten class of matrix norms. Let $\boldsymbol{A}$ be a matrix, and write $\boldsymbol{\sigma}(\boldsymbol{A})$ for the vector of singular values of $\boldsymbol{A}$, arranged in weakly decreasing order. The *Schatten p-norm* is defined as

$$\|\boldsymbol{A}\|_{S_p} = \|\boldsymbol{\sigma}(\boldsymbol{A})\|_p,$$

where $\|\cdot\|_p$ is the usual ℓ_p vector norm. In particular, S_2 is the Frobenius norm, and S_∞ is the spectral norm. The norm S_1 is often called the *trace norm* because

$$\|\boldsymbol{A}\|_{S_1} = \mathrm{trace}(\boldsymbol{A}) \quad \text{when } \boldsymbol{A} \text{ is psd.}$$

The term *psd* abbreviates *positive semidefinite*. For general matrices, the Frobenius norm is the only Schatten-class norm computable directly from the matrix entries:

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \left[\sum\nolimits_{jk} |a_{jk}|^2\right]^{1/2}.$$

B. Rank Bounds via Norm Ratios

A simple but powerful method for estimating rank is to compare two different Schatten norms of the same matrix.

Lemma 6: Suppose that $p < q$. For each matrix $\boldsymbol{A}$,

$$\mathrm{rank}(\boldsymbol{A}) \geq \left[\frac{\|\boldsymbol{A}\|_{S_p}}{\|\boldsymbol{A}\|_{S_q}}\right]^{pq/(q-p)}.$$

Proof: For each vector $\boldsymbol{x} \in \mathbb{C}^r$, we have the inequality

$$\frac{\|\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_q} \leq r^{1/p-1/q}.$$

Indeed, one can use Lagrange multipliers to verify that the left-hand side is maximized when $\boldsymbol{x}$ is a constant vector.

Suppose that $\mathrm{rank}(\boldsymbol{A}) = r$. Then the vector $\boldsymbol{\sigma}$ of nonzero singular values of $\boldsymbol{A}$ lies in $\mathbb{C}^r$. By definition of the Schatten norms,

$$\frac{\|\boldsymbol{A}\|_{S_p}}{\|\boldsymbol{A}\|_{S_q}} = \frac{\|\boldsymbol{\sigma}\|_p}{\|\boldsymbol{\sigma}\|_q} \leq r^{1/p-1/q}.$$

Take the $(1/p - 1/q)$ root and simplify the exponent to reach the conclusion. ∎

The following simple corollary is fantastically useful.

Corollary 7: Let $\boldsymbol{A}$ be a matrix. Then

$$\mathrm{rank}(\boldsymbol{A}) \geq \frac{\|\boldsymbol{A}\|_{S_1}^2}{\|\boldsymbol{A}\|_{\mathrm{F}}^2} \quad \text{and} \quad \mathrm{rank}(\boldsymbol{A}) \geq \frac{\|\boldsymbol{A}\|_{\mathrm{F}}^2}{\|\boldsymbol{A}\|^2}.$$

Alon has applied the first estimate in his work on extremal combinatorics [11]. The second estimate arises in a paper of Bourgain and Tzafriri on restricted invertibility [12].

C. A Schur Complement Rank Identity

Suppose that $\boldsymbol{X}$ is a psd matrix, partitioned so that its diagonal blocks are square:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^* & \boldsymbol{C} \end{bmatrix}.$$

Provided that the block $\boldsymbol{A}$ is nonsingular, the *Schur complement* of $\boldsymbol{A}$ in $\boldsymbol{X}$ is the matrix

$$\boldsymbol{X}/\boldsymbol{A} = \boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^*.$$

For our purposes, the relevant fact is that

$$\mathrm{rank}(\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{A}) + \mathrm{rank}(\boldsymbol{X}/\boldsymbol{A}). \tag{9}$$

See [13, Sec. 2] for more Schur complement identities.

IV. Sparsity Gap under Strong Incoherence

Corollary 5 indicates that we can obtain uncertainty principles for generic signals by developing lower bounds on the rank of a subdictionary $\boldsymbol{\Phi}_R$. This section describes the simplest approach to this problem, which proceeds via Corollary 7. This method is most effective when the coherence μ is small.

Let R be a set of atoms. Since

$$\mathrm{rank}(\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R) = \mathrm{rank}(\boldsymbol{\Phi}_R), \tag{10}$$

we may as well work with the Gram matrix of $\boldsymbol{\Phi}_R$. This substitution allows us to exploit geometric information about the dictionary. Indeed, the diagonal entries of $\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R$ equal one because the atoms have unit ℓ_2 norm, and the off-diagonal entries are bounded in magnitude by μ because they contain the inner products between distinct atoms.

Lemma 8: Let R be a set of r atoms. Then

$$\mathrm{rank}(\boldsymbol{\Phi}_R) \geq \frac{r}{1 + (r-1)\mu^2}.$$

Proof: Relation (10) and Corollary 7 imply that

$$\mathrm{rank}(\boldsymbol{\Phi}_R) = \mathrm{rank}(\boldsymbol{\Phi}^*\boldsymbol{\Phi}_R) \geq \frac{\|\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R\|_{S_1}^2}{\|\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R\|_{\mathrm{F}}^2}. \tag{11}$$

Owing to the properties of the Gram matrix,

$$\|\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R\|_{S_1}^2 = (\mathrm{trace}\,\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R)^2 = r^2$$

and

$$\|\boldsymbol{\Phi}_R^* \boldsymbol{\Phi}_R\|_{\mathrm{F}}^2 \leq r + r(r-1)\mu^2$$

Introduce these bounds into (11) to complete the proof. ∎

When the coherence μ is small, Lemma 8 provides excellent estimates for the rank of $\boldsymbol{\Phi}_R$. By combining this bound with Corollary 5, we obtain our first main result.

Theorem 9: Suppose that S indexes a linearly independent set of s atoms and that T lists t atoms. Assume that the size δ of their intersection

$$\delta = |S \cap T| < s\left[1 - \frac{t-1}{s} \cdot \frac{t\mu^2}{1 - t\mu^2}\right]. \tag{12}$$

Let $\boldsymbol{u} = \boldsymbol{\Phi}_S \boldsymbol{x}$ be a generic signal in $\mathrm{range}(\boldsymbol{\Phi}_S)$. Then

$$\mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\boldsymbol{\Phi}_T)\} = 0.$$

A fortiori, there is zero probability that $\boldsymbol{u}$ can be represented using *any* set T of t atoms whose overlap with S equals δ.

Proof: Define the set $R = S \cup T$. Observe that

$$r := |R| = |S| + |T| - |S \cap T| = s + t - \delta.$$

Corollary 5 states that $t < \mathrm{rank}(\boldsymbol{\Phi}_R)$ implies

$$\mathbb{P}\{\boldsymbol{u} \in \mathrm{range}(\boldsymbol{\Phi}_T)\} = 0. \tag{13}$$

According to Lemma 8,

$$\frac{r}{1 + (r-1)\mu^2} \leq \mathrm{rank}(\boldsymbol{\Phi}_R).$$

Therefore, another sufficient condition for (13) is

$$t < \frac{r}{1 + (r-1)\mu^2} = \frac{s + t - \delta}{1 + (s + t - \delta - 1)\mu^2}.$$

Solving this relation for δ results in the bound

$$\delta < s - \frac{(t-1)t\mu^2}{1 - t\mu^2}.$$

If δ satisfies this condition, then (13) is in force.

Since only a finite number of sets T meet the hypotheses of the theorem, there is zero chance that $\boldsymbol{u}$ is representable using *any* such set T of atoms. ∎

In words, Theorem 9 states that a generic signal constructed using a set of s atoms almost surely has no representation using another set of t atoms unless there is a substantial overlap between the two sets.

For example, suppose $\mu \leq m^{-1/2}$, and let $\boldsymbol{u}$ be a generic signal in $\text{range}(\boldsymbol{\Phi}_S)$, where $|S| \leq m/3$. Take $t = s$ in the overlap condition (12) to see that, almost surely, *every* representation of $\boldsymbol{u}$ with s atoms requires at least $s/2$ atoms from S!

To obtain an uncertainty principle, note that the relation (12) is quadratic in t. By reverting this inequality, we obtain a sufficient condition on t as a function of s and δ.

Corollary 10: With the notation of Theorem 9, assume

$$t < (s - \delta - 1)\left[\sqrt{\left(1 + \frac{1}{s - \delta - 1}\right)\frac{\mu^{-2}}{s - \delta - 1} + \frac{1}{4}} - 1\right].$$

Then $\mathbb{P}\{\boldsymbol{u} \in \text{range}(\boldsymbol{\Phi}_T)\} = 0$.

After a considerable amount of algebra, Corollary 10 simplifies to a new uncertainty principle.

Corollary 11 (Uncertainty Principle for Generic Signals): Suppose S is linearly independent. A generic signal $\boldsymbol{u}$ in $\text{range}(\boldsymbol{\Phi}_S)$ almost surely has no representation over a set T with overlap $\delta = |S \cap T|$ *unless*

$$|S| + |T| > \delta + \mu^{-1}\sqrt{|S| - \delta}.$$

Theorem A follows from Corollary 11 by setting $\delta = 0$.

V. SPARSITY GAP UNDER WEAK INCOHERENCE

In this section, we use a different technique to bound the rank of $\boldsymbol{\Phi}_R$ below. This approach relies on the Schur complement identity (9) and the second part of Corollary 7. We also require some information about the properties of random sets of atoms, drawn from [9], [10]. The conclusions are most interesting for weakly incoherent dictionaries.

Assume S is linearly independent, which implies that the Gram matrix $\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S$ is invertible. Let V be an index set disjoint from S, and write $R = S \cup V$. Then the Gram matrix has the block structure

$$\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R = \begin{bmatrix} \boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S & \boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_V \\ \boldsymbol{\Phi}_V^*\boldsymbol{\Phi}_S & \boldsymbol{\Phi}_V^*\boldsymbol{\Phi}_V \end{bmatrix}.$$

The identity (9) shows that we can control the rank of the Gram matrix by controlling the rank of the northwest block and its Schur complement. To state the result, we recall that

$$\boldsymbol{P}_S = \boldsymbol{\Phi}_S(\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S)^{-1}\boldsymbol{\Phi}_S^* \qquad (14)$$

is the orthogonal projector onto $\text{range}(\boldsymbol{\Phi}_S)$.

Lemma 12: Suppose that S is linearly independent; let V be disjoint from S; and define $R = S \cup V$. Then

$$\text{rank}(\boldsymbol{\Phi}_R) = |S| + \text{rank}((\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V).$$

Proof: The Schur complement identity (9) shows that

$$\text{rank}(\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R) = \text{rank}(\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S) + \text{rank}(\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R/\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S).$$

The set S is linearly independent, so

$$\text{rank}(\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S) = \text{rank}(\boldsymbol{\Phi}_S) = |S|.$$

Next, recall the definition of the Schur complement:

$$\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R/\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S = \boldsymbol{\Phi}_V^*\boldsymbol{\Phi}_V - \boldsymbol{\Phi}_V^*\boldsymbol{\Phi}_S(\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S)^{-1}\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_V.$$

Identify the orthogonal projector onto $\text{range}(\boldsymbol{\Phi}_S)$ to see that

$$\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R/\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S = \boldsymbol{\Phi}_V^*(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V.$$

We conclude that

$$\text{rank}(\boldsymbol{\Phi}_R^*\boldsymbol{\Phi}_R/\boldsymbol{\Phi}_S^*\boldsymbol{\Phi}_S) = \text{rank}((\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V)$$

because $(\mathbf{I} - \boldsymbol{P}_S)^2 = (\mathbf{I} - \boldsymbol{P}_S)$. ∎

The next result provides a lower bound on the second term in Lemma 12. Here and elsewhere, the dagger † denotes the pseudoinverse of a matrix.

Lemma 13: Suppose that S is linearly independent, and let V be disjoint from S. Then

$$\text{rank}((\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V) \geq$$
$$\rho^{-1}|V|\left(1 - \left\|\boldsymbol{\Phi}_S^\dagger\right\|^2 \max_{v \notin S}\|\boldsymbol{\Phi}_S^*\boldsymbol{\varphi}_v\|_2^2\right).$$

Proof: Corollary 7 states that

$$\text{rank}((\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V) \geq \frac{\|(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V\|_\text{F}^2}{\|(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V\|^2}. \qquad (15)$$

The spectral norm satisfies the bound

$$\|(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V\|^2 \leq \|\boldsymbol{\Phi}_V\|^2 \leq \rho$$

because $\boldsymbol{\Phi}_V$ is a column submatrix of $\boldsymbol{\Phi}$. With a little more work, we can produce a lower bound on the Frobenius norm.

$$\|(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\Phi}_V\|_\text{F}^2 = \sum_{v \in V}\|(\mathbf{I} - \boldsymbol{P}_S)\boldsymbol{\varphi}_v\|_2^2$$
$$= \sum_{v \in V}(1 - \|\boldsymbol{P}_S\boldsymbol{\varphi}_v\|_2^2)$$
$$\geq \sum_{v \in V}\left(1 - \left\|\boldsymbol{\Phi}_S^\dagger\right\|^2 \|\boldsymbol{\Phi}_S^*\boldsymbol{\varphi}_v\|_2^2\right)$$
$$\geq |V|\left(1 - \left\|\boldsymbol{\Phi}_S^\dagger\right\|^2 \max_{v \notin S}\|\boldsymbol{\Phi}_S^*\boldsymbol{\varphi}_v\|_2^2\right).$$

The first inequality follows by expanding the projector via (14) and invoking the usual operator norm bound. Introduce the two norm estimates into (15) to complete the argument. ∎

Lemma 13 is interesting because a *random* set S of atoms from a *weakly incoherent* dictionary usually leads to small values for the mysterious quantities in the rank bound.

Proposition 14: Suppose that S is a uniformly random subset of $\{1, 2, \ldots, N\}$ with cardinality s. For $\beta \geq 1$,

$$\mathbb{P}\left\{\max_{v \notin S} \|\boldsymbol{\Phi}_S^* \boldsymbol{\varphi}_v\|_2 > \mathrm{C}\beta\left[\mu\sqrt{\log N} + 2\sqrt{\frac{\rho s}{N}}\right]\right\} \leq 2N^{-\beta}.$$

and

$$\mathbb{P}\left\{\|\boldsymbol{\Phi}_S^* \boldsymbol{\Phi}_S - \mathbf{I}\| < \mathrm{C}\beta\left[\mu\log N + \sqrt{\frac{\rho s \log N}{N}}\right]\right\} \leq 2N^{-\beta}.$$

The number C is an absolute constant.

Proposition 14 follows from the results in [10, Sec. 5] after some standard arguments. We use Markov's inequality to convert the moment bounds into tail bounds. Then we invoke a simple decoupling argument (see [7, Lem. 14] for a model) to transfer the result for a random set with *expected* cardinality s to a random set with exact cardinality s.

To take advantage of Proposition 14, we restrict our attention to a specific class of dictionaries. Assume that $\boldsymbol{\Phi}$ is a weakly incoherent tight frame, as defined in (4). Suppose furthermore that the sparsity s satisfies (5). If we fix a sufficiently small value for constant c and take $\beta = 1$, Proposition 14 ensures that

$$\max_{v \notin S} \|\boldsymbol{\Phi}_S^* \boldsymbol{\varphi}_v\|_2 \leq 1/2 \quad \text{and} \quad \|\boldsymbol{\Phi}_S^\dagger\| \leq \sqrt{2},$$

except with probability $\mathrm{O}(N^{-1})$. The upshot of this discussion is the following bound.

Corollary 15: Assume the hypotheses of Lemma 13 are in force. Let $R = S \cup V$. Then

$$\mathrm{rank}(\boldsymbol{\Phi}_R) \geq |S| + \frac{2m\,|V|}{N}.$$

We reach our final major result by introducing this bound into Corollary 5.

Theorem 16: Assume that $\boldsymbol{\Phi}$ satisfies the weak incoherence conditions (4), and assume further that $N > 2m$. Draw a random set S of s atoms, where s satisfies (5). Except with probability $\mathrm{O}(N^{-1})$, the following result holds.

Suppose that T lists t atoms. Let $\delta = |S \cap T|$ be the size of the overlap with S, and assume that

$$t < \left(s - \frac{2\delta m}{N}\right)\left(1 - \frac{2m}{N}\right)^{-1}.$$

Draw a generic signal $\boldsymbol{u} = \boldsymbol{\Phi}_S \boldsymbol{x}$. Then

$$\mathbb{P}\left\{\boldsymbol{u} \in \mathrm{range}(\boldsymbol{\Phi}_T)\right\} = 0.$$

A fortiori, there is zero probability that $\boldsymbol{u}$ has a representation over *any* such set T of atoms.

Proof: Draw a random set S of s atoms. Corollary 15 guarantees that, with probability $\mathrm{O}(N^{-1})$, the following result holds. For any given set T of t atoms, define $V = T \setminus S$ and select $R = S \cup V$. Then

$$\frac{2m\,|V|}{N} \leq \mathrm{rank}(\boldsymbol{\Phi}_R). \tag{16}$$

Now, suppose that T is a specific set of t atoms. According to Corollary 5, the condition $t < \mathrm{rank}(\boldsymbol{\Phi}_R)$ implies that a generic signal in $\mathrm{range}(\boldsymbol{\Phi}_S)$ almost surely has no representation using the atoms in T. It follows from (16) that

$$t < s + \frac{2m\,|V|}{N}$$

is a stricter sufficient condition that $\boldsymbol{u}$ almost surely cannot be represented with T. Introduce the identity $|V| = t - \delta$ into the last relation and rearrange to obtain the equivalent condition

$$t < \left(s - \frac{2\delta m}{N}\right)\left(1 - \frac{2m}{N}\right)^{-1}.$$

As usual, we may take a union bound over the finite number of admissible T to obtain a uniform result. $\blacksquare$

The message may be clearer if we make additional simplifications to the sufficient condition in Theorem 16. First, subtract and add $2sm/N$ in the first parenthesis to reach

$$t < \left[s\left(1 - \frac{2m}{N}\right) + \frac{2(s - \delta)m}{N}\right]\left(1 - \frac{2m}{N}\right)^{-1}.$$

Since $1 < (1 - 2m/N)^{-1}$, we find that a further sufficient condition is

$$t \leq s + \frac{2(s - \delta)m}{N} \tag{17}$$

In words, there is zero probability that a generic signal in $\mathrm{range}(\boldsymbol{\Phi}_S)$ has a representation using t atoms unless t is somewhat larger than s or the alternative representation uses many atoms from S. Take $\delta = 0$ in (17) to reach Theorem B.

References

[1] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comp. Harmonic Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.

[2] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, June 1989.

[3] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.

[4] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[5] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197–2202, March 2003.

[6] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[7] J. A. Tropp, "On the linear independence of spikes and sines," *J. Fourier Anal. Appl.*, vol. 14, pp. 838–858, 2008.

[8] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Comput. Math*, vol. 6, pp. 227–254, 2006.

[9] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comp. Harmonic Anal.*, vol. 25, pp. 1–24, 2008.

[10] ——, "Norms of random submatrices and sparse approximation," *C. R. Acad. Sci. Paris Ser. I Math.*, vol. 346, pp. 1271–1274, 2008.

[11] N. Alon, "Problems and results in extremal combinatorics," *Discrete Math.*, vol. 273, pp. 31–53, 2003.

[12] J. Bourgain and L. Tzafriri, "Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis," *Israel J. Math*, vol. 57, no. 2, pp. 137–224, 1987.

[13] C. Page, G. P. H. Styan, B.-Y. Wang, and F. Zhang, "Hua's matrix equality and Schur complements," *Intl. J. Info. Sys. Sci.*, vol. 3, no. 1, pp. 1–18, 2007.