Finite Frames for Sparse Signal Processing

Waheed U. Bajwa and Ali Pezeshki

Abstract Over the last decade, considerable progress has been made towards developing new signal processing methods to manage the deluge of data caused by advances in sensing, imaging, storage, and computing technologies. Most of these methods are based on a simple but fundamental observation. That is, high-dimensional data sets are typically highly redundant and live on low-dimensional manifolds or subspaces. This means that the collected data can often be represented in a sparse or parsimonious way in a suitably selected finite frame. This observation has also led to the development of a new sensing paradigm, called compressed sensing, which shows that high-dimensional data sets can often be reconstructed, with high fidelity, from only a small number of measurements. Finite frames play a central role in the design and analysis of both sparse representations and compressed sensing methods. In this chapter, we highlight this role primarily in the context of compressed sensing for estimation, recovery, support detection, regression, and detection of sparse signals. The recurring theme is that frames with small spectral norm and/or small worst-case coherence, average coherence, or sum coherence are well-suited for making measurements of sparse signals.

Key words: approximation theory, coherence property, compressed sensing, detection, estimation, Grassmannian frames, model selection, regression, restricted isometry property, typical guarantees, uniform guarantees, Welch bound

---

Waheed U. Bajwa
Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, 94 Brett Rd, Piscataway, NJ 08854, USA e-mail: `waheed.bajwa@rutgers.edu`

Ali Pezeshki
Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523, USA e-mail: `ali.pezeshki@colostate.edu`

1 Introduction

It was not too long ago that scientists, engineers, and technologists were complaining about *data starvation*. In many applications, there never was sufficient data available to reliably carry out various inference and decision making tasks in real time. Technological advances during the last two decades, however, have changed all of that. So much in fact that *data deluge*, instead of data starvation, is now becoming a concern. If left unchecked, the rate at which data is being generated in numerous applications will soon overwhelm the associated systems' computational and storage resources.

During the last decade or so, there has been a surge of research activity in the signal processing and statistics communities to deal with the problem of data deluge. The proposed solutions to this problem rely on a simple but fundamental principle of *redundancy*. Massive data sets in the real world may live in high-dimensional spaces, but information embedded within these data sets almost always live near low-dimensional (often linear) manifolds. There are two ways in which the principle of redundancy can help us better manage the sheer abundance of data. First, we can represent the collected data in a parsimonious (or *sparse*) manner in carefully designed bases and frames. Sparse representations of data help reduce their (computational and storage) footprint and constitute an active area of research in signal processing [12]. Second, we can redesign the *sensing systems* to acquire only a small number of measurements by exploiting the low-dimensional nature of the signals of interest. The term *compressed sensing* has been coined for the area of research that deals with rethinking the design of sensing systems under the assumption that the signal of interest has a sparse representation in a *known* basis or frame [1, 16, 27].

There is a fundamental difference between the two aforementioned approaches to dealing with the data deluge; the former deals with the collected data while the latter deals with the collection of data. Despite this difference, however, there exists a great deal of mathematical similarity between the areas of sparse signal representation and compressed sensing. Our primary focus in this chapter will be on the compressed sensing setup and the role of finite frames in its development. However, many of the results discussed in this context can be easily restated for sparse signal representation. We will therefore use the generic term *sparse signal processing* in this chapter to refer to the collection of these results.

Mathematically, sparse signal processing deals with the case when a highly redundant frame $\Phi = (\varphi_i)_{i=1}^M$ in $\mathscr{H}^N$ is used to make (possibly noisy) measurements of sparse signals.[1] Consider an arbitrary signal $x \in \mathscr{H}^M$ that is K-sparse: $\|x\|_0 := \sum_{i=1}^M 1_{\{x_i \neq 0\}}(x) \leq K < N \ll M$. Instead of measuring x directly, sparse signal processing uses a small number of linear measurements of x, given by $y = \Phi x + n$, where $n \in \mathscr{H}^N$ corresponds to deterministic perturbation or stochastic noise. Given measurements y of x, the fundamental problems in sparse signal processing include: (*i*) recovering/estimating the sparse signal x, (*ii*) estimating x for linear regression,

[1] Sparse signal processing literature often uses the terms *sensing matrix*, *measurement matrix*, and *dictionary* for the frame Φ in this setting.

(*iii*) detecting the locations of the nonzero entries of x, and (*iv*) testing for the presence of x in noise. In all of these problems, certain geometrical properties of the frame Φ play crucial roles in determining the optimality of the end solutions. In this chapter, our goal is to make explicit these connections between the geometry of frames and sparse signal processing.

The four geometric measures of frames that we focus on in this chapter include the *spectral norm*, *worst-case coherence*, *average coherence*, and *sum coherence*. Recall that the spectral norm $\|\Phi\|$ of a frame Φ is simply a measure of its tightness and is given by the maximum singular value: $\|\Phi\| = \sigma_{max}(\Phi)$. The worst-case coherence μ_Φ, defined as

$$\mu_\Phi := \max_{\substack{i,j\in\{1,\dots,M\}\\ i\neq j}} \frac{|\langle\varphi_i,\varphi_j\rangle|}{\|\varphi_i\|\|\varphi_j\|}, \tag{1}$$

is a measure of the similarity between different frame elements. On the other hand, the average coherence is a new notion of frame coherence, introduced recently in [2, 3] and analyzed further in [4]. In words, the average coherence ν_Φ, defined as

$$\nu_\Phi := \frac{1}{M-1} \max_{i\in\{1,\dots,M\}} \left| \sum_{\substack{j=1\\ j\neq i}}^{M} \frac{\langle\varphi_i,\varphi_j\rangle}{\|\varphi_i\|\|\varphi_j\|} \right|, \tag{2}$$

is a measure of the spread of normalized frame elements $(\varphi_i/\|\varphi_i\|)_{i=1}^M$ in the unit ball. The sum coherence, defined as

$$\sum_{j=2}^{M}\sum_{i=1}^{j-1} \frac{|\langle\varphi_i,\varphi_j\rangle|}{\|\varphi_i\|\|\varphi_j\|}, \tag{3}$$

is a notion of coherence that arises in the context of detecting the presence of a sparse signal in noise [76, 77].

In the following sections, we show that different combinations of these geometric measures characterize the performance of a multitude of sparse signal processing algorithms. In particular, a theme that emerges time and again throughout this chapter is that frames with small spectral norm and/or small worst-case coherence, average coherence, or sum coherence are particularly well-suited for the purposes of making measurements of sparse signals.

Before proceeding further, we note that the signal x in some applications is sparse in the identity basis, in which case Φ represents the measurement process itself. In other applications, however, x can be sparse in some other orthonormal basis or an overcomplete dictionary Ψ. In this case, Φ corresponds to a composition of Θ, the frame resulting from the measurement process, and Ψ, the sparsifying dictionary, i.e., $\Phi = \Theta\Psi$. We do not make a distinction between the two formulations in this chapter. In particular, while the reported results are most readily interpretable in a physical setting for the former case, they are easily extendable to the latter case.

We note that this chapter provides an overview of only a small subset of current results in sparse signal processing literature. Our aim is simply to highlight the central role that finite frame theory plays in the development of sparse signal processing theory. We refer the interested reader to [34] and the references therein for a more comprehensive review of sparse signal processing literature.

2 Sparse Signal Processing: Uniform Guarantees and Grassmannian Frames

Recall the fundamental system of equations in sparse signal processing: $y = \Phi x + n$. Given the measurements y, our goal in this section is to specify conditions on the frame Φ and accompanying computational methods that enable reliable inference of the high-dimensional sparse signal x from the low-dimensional measurements y. There has been a lot of work in this direction in the sparse signal processing literature. Our focus in this section is on providing an overview of some of the key results in the context of performance guarantees for *every* K-sparse signal in $\mathscr{H}^M$ using a *fixed* frame Φ. It is shown in the following that *uniform performance guarantees* for sparse signal processing are directly tied to the worst-case coherence of frames. In particular, the closer a frame is to being a *Grassmannian frame*—defined as one that has the smallest worst-case coherence for given N and M—the better its performance is in the uniform sense.

2.1 Recovery of Sparse Signals via ℓ_0 Minimization

We consider the simplest of setups in sparse signal processing, corresponding to the recovery of a sparse signal x from noiseless measurements $y = \Phi x$. Mathematically speaking, this problem is akin to solving an *underdetermined* system of linear equations. Although an underdetermined system of linear equations has infinitely many solutions in general, one of the surprises of sparse signal processing is that recovery of x from y remains a well-posed problem for large classes of random and deterministic frames because of the underlying sparsity assumption. Since we are looking to solve y for a K-sparse x, an intuitive way of obtaining a candidate solution from y is to search for the sparsest solution $\widehat{x}_0$ that satisfies $y = \Phi\widehat{x}_0$. Mathematically, this solution criterion can be expressed in terms of the following ℓ_0 minimization program

$$\widehat{x}_0 = \arg\min_{z \in \mathscr{H}^M} \|z\|_0 \quad \text{subject to} \quad y = \Phi z. \tag{P_0}$$

Despite the apparent simplicity of (P_0), the conditions under which it can be claimed that $\widehat{x}_0 = x$ for any $x \in \mathscr{H}^M$ are not immediately obvious. Given that (P_0) is a highly nonconvex optimization, there is in fact little reason to expect that $\widehat{x}_0$

should be unique to begin with. It is because of these roadblocks that a rigorous mathematical understanding of (P_0) alluded the researchers for a long time. These mathematical challenges were eventually overcome through surprisingly elementary mathematical tools in [28, 41]. In particular, it is argued in [41] that a property termed *Unique Representation Property* (URP) of Φ is the key to understanding the behavior of the solution obtained from (P_0).

Definition 1 (Unique Representation Property). A frame $\Phi = (\varphi_i)_{i=1}^M$ in $\mathscr{H}^N$ is said to have the unique representation property of order K if any K frame elements of Φ are linearly independent.

It has been shown in [28, 41] that the URP of order $2K$ is both a necessary and a sufficient condition for the equivalence of $\widehat{x}_0$ and x.[2]

Theorem 1 ([28, 41]). *An arbitrary K-sparse signal x can be uniquely recovered from $y = \Phi x$ as a solution to (P_0) if and only if Φ satisfies the URP of order $2K$.*

The proof of Theorem 1 is simply an exercise in elementary linear algebra. It follows from the simple observation that K-sparse signals in $\mathscr{H}^M$ are mapped injectively into $\mathscr{H}^N$ if and only if the nullspace of Φ does not contain nontrivial $2K$-sparse signals. In order to understand the significance of Theorem 1, note that random frames with elements distributed uniformly at random on the unit sphere in $\mathscr{H}^N$ are almost surely going to have the URP of order $2K$ as long as $N \geq 2K$. This is rather powerful, since this signifies that sparse signals can be recovered from a number of random measurements that is only linear in the *sparsity* K of the signal, rather than the ambient dimension M. Despite this powerful result, however, Theorem 1 is rather opaque in the case of arbitrary (not necessarily random) frames. This is because the URP is a local geometric property of Φ and explicitly verifying the URP of order $2K$ requires a combinatorial search over all $\binom{M}{2K}$ possible collections of frame elements. Nevertheless, it is possible to replace the URP in Theorem 1 with the worst-case coherence of Φ, which is a global geometric property of Φ that can be easily computed in polynomial time. The key to this is the classical *Geršgorin Circle Theorem* [40] that can be used to relate the URP of a frame Φ to its worst-case coherence.

Lemma 1 (Geršgorin). *Let $t_{i,j}, i, j = 1,\ldots,M$, denote the entries of an $M \times M$ matrix T. Then every eigenvalue of T lies in at least one of the M circles defined below*

$$\mathscr{D}_i(T) = \Big\{ z \in \mathbb{C} : |z - t_{i,i}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{M} |t_{i,j}| \Big\},\ i = 1,\ldots,M. \tag{4}$$

The Geršgorin Circle Theorem seems to have first appeared in 1931 in [40] and its proof can be found in any standard text on matrix analysis such as [50]. This theorem allows one to relate the worst-case coherence of Φ to the URP as follows.

[2] Theorem 1 has been stated in [28] using the terminology of *spark*, instead of the URP. The spark of a frame Φ is defined in [28] as the smallest number of frame elements of Φ that are linearly dependent. In other words, Φ satisfies the URP of order K if and only if $\mathrm{spark}(\Phi) \geq K+1$.

Theorem 2 ([28]). *Let Φ be a unit-norm frame and $K \in \mathbb{N}$. Then Φ satisfies the URP of order K as long as $K < 1 + \mu_\Phi^{-1}$.*

The proof of this theorem follows by bounding the minimum eigenvalue of any $K \times K$ principal submatrix of the Gramian matrix G_Φ using Lemma 1. We can now combine Theorem 1 with Theorem 2 to obtain the following theorem that relates the worst-case coherence of Φ to the sparse signal recovery performance of (P_0).

Theorem 3. *An arbitrary K-sparse signal x can be uniquely recovered from $y = \Phi x$ as a solution to (P_0) provided*

$$K < \frac{1}{2}\left(1 + \mu_\Phi^{-1}\right). \tag{5}$$

Theorem 3 states that ℓ_0 minimization enables unique recovery of *every* K-sparse signal measured using a frame Φ as long as $K = O(\mu_\Phi^{-1})$.[3] This dictates that frames that have small worst-case coherence are particularly well suited for measuring sparse signals. It is also instructive to understand the fundamental limitations of Theorem 3. In order to do so, we recall the following fundamental lower bound on the worst-case coherence of unit-norm frames.

Lemma 2 (The Welch Bound [75]). *The worst-case coherence of any unit-norm frame $\Phi = (\varphi_i)_{i=1}^M$ in $\mathscr{H}^N$ satisfies the inequality $\mu_\Phi \geq \sqrt{\frac{M-N}{N(M-1)}}$.*

It can be seen from the Welch bound that $\mu_\Phi = \Omega(N^{-1/2})$ as long as $M > N$. Therefore, we have from Theorem 3 that even in the best of cases ℓ_0 minimization yields unique recovery of every sparse signal as long as $K = O(\sqrt{N})$. This implication is weaker than the $K = O(N)$ scaling that we observed earlier for random frames. A natural question to ask therefore is whether Theorem 3 is weak in terms of the relationship between K and μ_Φ. The answer to this question however is in the negative, since there exist frames such as union of identity and Fourier bases [30] and Steiner equiangular tight frames [36] that have certain collections of frame elements with cardinality $O(\sqrt{N})$ that are linearly dependent. We therefore conclude from the preceding discussion that Theorem 3 is tight from the frame-theoretic perspective and, in general, frames with small worst-case coherence are better suited for recovery of sparse signals using (P_0). In particular, this highlights the importance of Grassmannian frames in the context of sparse signal recovery in the uniform sense.

2.2 Recovery and Estimation of Sparse Signals via Convex Optimization and Greedy Algorithms

The implications of Sect. 2.1 are quite remarkable. We have seen that it is possible to recover an K-sparse signal x using a small number of measurements that is propor-

[3] Recall, with big-O notation, that $f(n) = O(g(n))$ if there exists positive C and n_0 such that for all $n > n_0$, $f(n) \leq Cg(n)$. Also, $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$, and $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

tional to μ_{Φ}^{-1}; in particular, for large classes of frames such as *Gabor frames* [3], we see that $O(K^2)$ number of measurements suffice to recover a sparse signal using ℓ_0 minimization. This can be significantly smaller than the $N = M$ measurements dictated by the classical signal processing when $K \ll M$. Despite this, however, sparse signal recovery using (P_0) is something that one cannot be expected to use for practical purposes. The reason for this is the computational complexity associated with ℓ_0 minimization; in order to solve (P_0), one needs to exhaustively search through all possible sparsity levels. The complexity of such exhaustive search is clearly exponential in M and it has been shown in [54] that (P_0) is in general an NP-hard problem. Alternate methods of solving $y = \Phi x$ for a K-sparse x that are also computationally feasible therefore has been of great interest to the practitioners. The recent interest in the literature on sparse signal processing partly stems from the fact that significant progress has been made by numerous researchers in obtaining various practical alternatives to (P_0). Such alternatives range from convex optimization-based methods [18, 22, 66] to greedy algorithms [25, 51, 55]. In this subsection, we review the performance guarantees of two such seminal alternative methods that are widely used in practice and once again highlight the role Grassmannian frames play in sparse signal processing.

2.2.1 Basis Pursuit

A common heuristic approach taken in solving nonconvex optimization problems is to approximate them with a convex problem and solve the resulting optimization program. A similar approach can be taken to *convexify* (P_0) by replacing the ℓ_0 "norm" in (P_0) with its closest convex approximation, the ℓ_1 norm: $\|z\|_1 = \sum_i |z_i|$. The resulting optimization program, which seems to have been first proposed as a heuristic in [59], can be formally expressed as follows:

$$\widehat{x}_1 = \arg\min_{z \in \mathcal{H}^M} \|z\|_1 \quad \text{subject to} \quad y = \Phi z. \tag{P_1}$$

The ℓ_1 minimization program (P_1) is termed as *Basis Pursuit* (BP) [22] and is in fact a linear optimization program [11]. To this date, a number of numerical methods have been proposed for solving BP in an efficient manner; we refer the reader to [72] for a survey of some of the numerical methods.

Even though BP has existed in the literature since at least [59], it is only in the last decade that results concerning its performance have been reported. Below, we present one such result that is expressed in terms of the worst-case coherence of the frame Φ [28, 42].

Theorem 4 ([28, 42]). *An arbitrary K-sparse signal x can be uniquely recovered from $y = \Phi x$ as a solution to (P_1) provided*

$$K < \frac{1}{2}\left(1 + \mu_{\Phi}^{-1}\right). \tag{6}$$

Algorithm 1 Orthogonal Matching Pursuit

Input: Unit-norm frame Φ and measurement vector y
Output: Sparse OMP estimate $\widehat{x}_{OMP}$
Initialize: $i = 0$, $\widehat{x}^0 = 0$, $\widehat{\mathscr{K}} = \emptyset$, and $r^0 = y$
while $\|r^i\| \geq \varepsilon$ **do**
  $i \leftarrow i+1$ {Increment counter}
  $z \leftarrow \Phi^* r^{i-1}$ {Form signal proxy}
  $\ell \leftarrow \arg\max_j |z_j|$ {Select frame element}
  $\widehat{\mathscr{K}} \leftarrow \widehat{\mathscr{K}} \cup \{\ell\}$ {Update the index set}
  $\widehat{x}^i_{\widehat{\mathscr{K}}} \leftarrow \Phi^\dagger_{\widehat{\mathscr{K}}} y$ and $\widehat{x}^i_{\widehat{\mathscr{K}}^c} \leftarrow 0$ {Update the estimate}
  $r^i \leftarrow y - \Phi \widehat{x}^i$ {Update the residue}
end while
return $\widehat{x}_{OMP} = \widehat{x}^i$

The reader will notice that the sparsity requirements in both Theorem 3 and Theorem 4 are the same. This does not mean however that (P_0) and (P_1) always yield the same solution. This is because the sparsity requirements in the two theorems are only sufficient conditions. Regardless, it is rather remarkable that one can solve an underdetermined system of equations $y = \Phi x$ for an K-sparse x in polynomial time as long as $K = O(\mu_\Phi^{-1})$. In particular, we can once again draw the conclusion from Theorem 4 that frames with small worst-case coherence in general and Grassmannian frames in particular are highly desirable in the context of recovery of sparse signals using BP.

2.2.2 Orthogonal Matching Pursuit

Basis pursuit is arguably a highly practical scheme for recovering an K-sparse signal x from the set of measurements $y = \Phi x$. In particular, depending upon the particular implementation, the computational complexity of convex optimization methods like BP for general frames is typically $O(M^3 + NM^2)$, which is much better than the complexity of (P_0), assuming P $\neq$ NP. Nevertheless, BP can be computationally demanding for large-scale sparse recovery problems. Fortunately, there do exist greedy alternatives to optimization-based approaches for sparse signal recovery. The oldest and perhaps the most well-known among these greedy algorithms goes by the name of *Orthogonal Matching Pursuit* (OMP) in the literature [51]. Note that just like BP, OMP has been in practical use for a long time, but it is only recently that its performance has been characterized by the researchers.

The OMP algorithm obtains an estimate $\widehat{\mathscr{K}}$ of the indices of the frame elements $\{\varphi_i : x_i \neq 0\}$ that contribute to the measurements $y = \sum_{i:x_i \neq 0} \varphi_i x_i$. The final OMP estimate $\widehat{x}_{OMP}$ then corresponds to a least-squares estimate of x using the frame elements $\{\varphi_i\}_{i \in \widehat{\mathscr{K}}}$: $\widehat{x}_{OMP} = \Phi^\dagger_{\widehat{\mathscr{K}}} y$, where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse. In order to estimate the indices, the OMP starts with an empty set and greedily expands that set by one additional frame element in each iteration. A formal description of the OMP algorithm is presented in Algorithm 1, in which $\varepsilon > 0$ is a stopping threshold. The power of OMP stems from the fact that if the estimate delivered by the algorithm has exactly K nonzeros then its computationally complexity is only $O(NMK)$, which is typically much better than the computationally complexity of $O(M^3 + NM^2)$ for convex optimization based approaches. We are now ready to state a theorem characterizing the performance of the OMP algorithm in terms of the worst-case coherence of frames.

Theorem 5 ([29, 68]). *An arbitrary K-sparse signal x can be uniquely recovered from $y = \Phi x$ as a solution to the OMP algorithm with $\varepsilon = 0$ provided*

$$K < \frac{1}{2}\left(1 + \mu_{\Phi}^{-1}\right). \tag{7}$$

Theorem 5 shows that the guarantees for the OMP algorithm in terms of the worst-case coherence match those for both (P_0) and BP; OMP too requires that $K = O(\mu_{\Phi}^{-1})$ in order for it to successfully recover a K-sparse x from $y = \Phi x$. It cannot be emphasized enough however that once $K = \Omega(\mu_{\Phi}^{-1})$, we start to see a difference in the empirical performance of (P_0), BP, and OMP. Nevertheless, the basic insight of Theorems 3–5 that frames with smaller worst-case coherence improve the recovery performance remains valid in all three cases.

2.2.3 Estimation of Sparse Signals

Our focus in this section has so far been on recovery of sparse signals from the measurements $y = \Phi x$. In practice, however, it is seldom the case that one obtains measurements of a signal without any additive noise. A more realistic model for measurement of sparse signals in this case can be expressed as $y = \Phi x + n$, where n represents either deterministic or random noise. In the presence of noise, one's objective changes from sparse signal recovery to sparse signal estimation; the goal being an estimate $\widehat{x}$ that is to close to the original sparse signal x in an ℓ_2-sense.

It is clear from looking at (P_1) that BP in its current form should not be used for estimation of sparse signals in the presence of noise, since $y \neq \Phi x$ in this case. However, a simple modification of the constraint in (P_1) allows us to gracefully handle noise in sparse signal estimation problems. The modified optimization program can be formally described as

$$\widehat{x}_1 = \arg \min_{z \in \mathscr{H}^M} \|z\|_1 \quad \text{subject to} \quad \|y - \Phi z\| \leq \varepsilon \tag{P_1^ε}$$

where ε is typically chosen to be equal to the noise magnitude: $\varepsilon = \|n\|$. The optimization (P_1^ε) is often termed as *Basis Pursuit with Inequality Constraint* (BPIC). It is easy to check that BPIC is also a convex optimization program, although it is no longer a linear program. Performance guarantees based upon the worst-case coherence for BPIC in the presence of deterministic noise alluded the researchers for quite some time. The problem was settled recently in [29], and the solution is summarized in the following theorem.

Theorem 6 ([29]). *Suppose that an arbitrary K-sparse signal x satisfies the sparsity constraint* $K < \frac{1+\mu_\Phi^{-1}}{4}$*. Given* $y = \Phi x + n$*, BPIC with* $\varepsilon = \|n\|$ *can be used to obtain an estimate* $\widehat{x}_1$ *of x such that*

$$\|x - \widehat{x}_1\| \leq \frac{2\varepsilon}{\sqrt{1-\mu_\Phi(4K-1)}}. \tag{8}$$

Theorem 6 states that BPIC with an appropriate ε results in a stable solution, *despite* the fact that we are dealing with an underdetermined system of equations. In particular, BPIC also handles sparsity levels that are $O(\mu_\Phi^{-1})$ and results in a solution that differs from the true signal x by $O(\|n\|)$.

In contrast with BP, OMP in its original form can be run for both noiseless sparse signal recovery and noisy sparse signal estimation. The only thing that changes in OMP in the latter case is the value of ε, which typically should also be set equal to the noise magnitude. The following theorem characterizes the performance of OMP in the presence of noise [29, 67, 69].

Theorem 7 ([29, 67, 69]). *Suppose that* $y = \Phi x + n$ *for an arbitrary K-sparse signal x and OMP is used to obtain an estimate* $\widehat{x}_{OMP}$ *of x with* $\varepsilon = \|n\|$*. Then the OMP solution satisfies*

$$\|x - \widehat{x}_{OMP}\| \leq \frac{\varepsilon}{\sqrt{1-\mu_\Phi(K-1)}} \tag{9}$$

provided x satisfies the sparsity constraint

$$K < \frac{1+\mu_\Phi^{-1}}{2} - \frac{\varepsilon \cdot \mu_\Phi^{-1}}{x_{min}}. \tag{10}$$

Here, x_{min} *denotes the smallest (in magnitude) nonzero entry of x:* $x_{min} = \min_{i:x_i \neq 0} |x_i|$*.*

It is interesting to note that unlike the case of sparse signal recovery, OMP in the noisy case does *not* have guarantees similar to that of BPIC. In particular, while the estimation error in OMP is still $O(\|n\|)$, the sparsity constraint in the case of OMP becomes restrictive as the smallest (in magnitude) nonzero entry of x decreases.

The estimation error guarantees provided in Theorem 6 and Theorem 7 are near-optimal for the case when the noise n follows an adversarial (or deterministic) model. This is since the noise n under the adversarial model can always be aligned with the signal x, making it impossible to guarantee an estimation error smaller than the size of n. However, if one is dealing with stochastic noise then it is possible to improve upon the estimation error guarantees for sparse signals. In order to do that, we first define a Lagrangian relaxation of (P_1^ε), which can be formally expressed as

$$\widehat{x}_{1,2} = \arg\min_{z \in \mathscr{H}^M} \frac{1}{2}\|y - \Phi x\| + \tau \|z\|_1. \tag{$P_{1,2}$}$$

The mixed-norm optimization program ($P_{1,2}$) goes by the name of *Basis Pursuit Denoising* (BPDN) [22] as well as *Least Absolute Shrinkage and Selection Operator* (LASSO) [66]. In the following, we state estimation error guarantees for both LASSO and OMP under the assumption of an *Additive White Gaussian Noise* (AWGN): $n \sim \mathcal{N}\left(0, \sigma^2 Id\right)$.

Theorem 8 ([6]). *Suppose that $y = \Phi x + n$ for an arbitrary K-sparse signal x, the noise n is distributed as $\mathcal{N}\left(0, \sigma^2 Id\right)$, and the LASSO is used to obtain an estimate $\widehat{x}_{1,2}$ of x with $\tau = 4\sqrt{\sigma^2 \log(M-K)}$. Then under the assumption that x satisfies the sparsity constraint $K < \frac{\mu_\Phi^{-1}}{3}$, the LASSO solution satisfies $support(\widehat{x}_{1,2}) \subset support(x)$ and*

$$\|x - \widehat{x}_{1,2}\| \leq \left(\sqrt{3} + 3\sqrt{4\log(M-K)}\right)^2 K\sigma^2 \tag{11}$$

with probability exceeding $\left(1 - \frac{1}{(M-K)^2}\right)\left(1 - e^{-K/7}\right)$.

A few remarks are in order now concerning Theorem 8. First, note that the results of the theorem hold with high probability since there exists a small probability that the Gaussian noise aligns with the sparse signal. Second, (11) shows that the estimation error associated with the LASSO solution is $O(\sqrt{\sigma^2 K \log M})$. This estimation error is within a logarithmic factor of the *best unbiased* estimation error $O(\sqrt{\sigma^2 K})$ that one can obtain in the presence of stochastic noise.[4] Ignoring the probabilistic aspect of Theorem 8, it is also worth comparing the estimation error of Theorem 6 with that of the LASSO. It is a tedious but simple exercise in probability to show that $\|n\| = \Omega(\sqrt{\sigma^2 M})$ with high probability. Therefore, if one were to apply Theorem 6 directly to the case of stochastic noise, then one obtains that the square of the estimation error scales linearly with the ambient dimension M of the sparse signal. On the other hand, Theorem 8 yields that the square of the estimation error scales linearly with the sparsity (modulo a logarithmic factor) of the sparse signal. This highlights the differences that exist between guarantees obtained under a deterministic noise model versus a stochastic (random) noise model.

We conclude this subsection by noting that it is also possible to obtain better OMP estimation error guarantees for the case of stochastic noise *provided* one inputs the sparsity of x to the OMP algorithm and modifies the halting criterion in Algorithm 1 from $\|r^i\| \geq \varepsilon$ to $i \leq K$ (i.e., the OMP is restricted to K iterations only). Under this modified setting, the guarantees for the OMP algorithm can be stated in terms of the following theorem.

Theorem 9 ([6]). *Suppose that $y = \Phi x + n$ for an arbitrary K-sparse signal x, the noise n is distributed as $\mathcal{N}\left(0, \sigma^2 Id\right)$, and the OMP algorithm is input the sparsity K of x. Then under the assumptions that x satisfies the sparsity constraint*

---

[4] It is worth pointing out here that if one is willing to tolerate some bias in the estimate, then the estimation error can be made smaller than $O(\sqrt{\sigma^2 K})$; see, e.g., [18, 31].

$$K < \frac{1+\mu_\Phi^{-1}}{2} - \frac{2\sqrt{\sigma^2 \log M} \cdot \mu_\Phi^{-1}}{x_{min}}, \tag{12}$$

the OMP solution obtained by terminating the algorithm after K iterations satisfies $support(\widehat{x}_{OMP}) = support(x)$ *and*

$$\|x - \widehat{x}_{OMP}\| \le 4\sqrt{\sigma^2 K \log M} \tag{13}$$

with probability exceeding $1 - \frac{1}{M\sqrt{2\pi \log M}}$. *Here,* x_{min} *again denotes the smallest (in magnitude) nonzero entry of x.*

2.3 Remarks

Recovery and estimation of sparse signals from a small number of linear measurements $y = \Phi x + n$ is an area of immense interest to a number of communities such as signal processing, statistics, and harmonic analysis. In this context, numerous reconstruction algorithms based upon either optimization techniques or greedy methods have been proposed in the literature. Our focus in this section has primarily been on two of the most well known methods in this regard, namely, BP (and BPIC and LASSO) and OMP. Nevertheless, it is important for the reader to realize that there exist other methods in the literature, such as the Dantzig Selector [18], CoSaMP [55], subspace pursuit [25], and IHT [7], that can also be used for recovery and estimation of sparse signals. These methods primarily differ from each other in terms of computational complexity and explicit constants, but offer error guarantees that appear very similar to the ones in Theorems 4–9.

We conclude this section by noting that our focus in here has been on providing uniform guarantees for sparse signals and relating those guarantees to the worst-case coherence of frames. The most important lesson of the preceding results in this regard is that there exist many computationally feasible algorithms that enable recovery/estimation of arbitrary K-sparse signals as long as $K = O(\mu_\Phi^{-1})$. There are two important aspects of this lesson. First, frames with small worst-case coherence are particularly well-suited for making observations of sparse signals. Second, even Grassmannian frames cannot be guaranteed to work well if $K = O(N^{1/2+\delta})$ for $\delta > 0$, which follows trivially from the Welch bound. This second observation seems overly restrictive, and there does exist literature based upon other properties of frames that attempts to break this "square-root" bottleneck. One such property, which has found widespread use in the compressed sensing literature, is termed as the *Restricted Isometry Property* (RIP) [14].

Definition 2 (Restricted Isometry Property). A unit-norm frame $\Phi = (\varphi_i)_{i=1}^M$ in $\mathscr{H}^N$ is said to have the RIP of order K with parameter $\delta_K \in (0,1)$ if for every K-sparse x, the following inequalities hold:

$$(1-\delta_K)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\delta_K)\|x\|_2^2. \tag{14}$$

The RIP of order K is essentially a statement concerning the minimum and maximum singular values of *all* $N \times K$ submatrices of Φ. However, even though the RIP has been used to provide guarantees for numerous sparse recovery/estimation algorithms such as BP, BPDN, CoSaMP, and IHT, explicit verification of this property for arbitrary frames appears to be computationally intractable. In particular, the only frames that are known to break the square-root bottleneck (using the RIP) for uniform guarantees are random (Gaussian, random binary, randomly subsampled partial Fourier, etc.) frames.[5] Still, it is possible to verify the RIP indirectly through the use of the Geršgorin Circle Theorem [5, 44, 71]. Doing so, however, yields results that match the ones reported above in terms of the sparsity constraint: $K = O(\mu_\Phi^{-1})$.

3 Beyond Uniform Guarantees: Typical Behavior

The square-root bottleneck in sparse recovery/estimation problems is hard to overcome in part because of our insistence that the results hold uniformly for all K-sparse signals. In this section, we take a departure from uniform guarantees and instead focus on *typical* behavior of various methods. In particular, we demonstrate in the following that the square-root bottleneck can be shattered by (*i*) imposing a statistical prior on the support and/or the nonzero entries of sparse signals and (*ii*) considering additional geometric measures of frames in conjunction with the worst-case coherence. In the following, we will focus on recovery, estimation, regression, and support detection of sparse signals using a multitude of methods. In all of these cases, we will assume that the support $\mathscr{K} \subset \{1,\dots,M\}$ of x is drawn uniformly at random from all $\binom{M}{K}$ size-K subsets of $\{1,\dots,M\}$. In some sense, this is the simplest statistical prior one can put on the support of x; in words, this assumption simply states that all supports of size K are equally likely.

3.1 Typical Recovery of Sparse Signals

In this section, we focus on typical recovery of sparse signals and provide guarantees for both ℓ_0 and ℓ_1 minimization (cf. (P_0) and (P_1)). The statistical prior we impose on the nonzero entries of sparse signals for this purpose however will differ for the two optimization schemes. We begin by providing a result for typical recovery of sparse signals using (P_0). The following theorem is due to Tropp and follows from combining results in [70] and [71].

---

[5] Recently Bourgain et al. in [10] have reported a deterministic construction of frames that satisfies the RIP of $K = O(N^{1/2+\delta})$. However, the constant δ in there is so small that the scaling can be considered $K = O(N^{1/2})$ for all practical purposes.

Theorem 10 ([70, 71]). *Suppose that $y = \Phi x$ for a K-sparse signal x whose support is drawn uniformly at random and whose nonzero entries have a jointly continuous distribution. Further, let the frame Φ be such that $\mu_\Phi \leq (c_1 \log M)^{-1}$ for numerical constant $c_1 = 240$. Then under the assumption that x satisfies the sparsity constraint*

$$K < \min\left\{\frac{\mu_\Phi^{-2}}{\sqrt{2}}, \frac{M}{c_2^2\|\Phi\|^2 \log M}\right\}, \tag{15}$$

the solution of (P_0) satisfies $\widehat{x}_0 = x$ with probability exceeding $1 - M^{-2\log 2}$. Here, $c_2 = 148$ is another numerical constant.

In order to understand the significance of Theorem 10, let us focus on the case of an approximately tight frame Φ: $\|\Phi\|^2 \approx \Theta(\frac{M}{N})$. In this case, ignoring the logarithmic factor, we have from (15) that ℓ_0 minimization can recover an K-sparse signal with high probability as long as $K = O(\mu_\Phi^{-2})$. This is in stark contrast to Theorem 3 that only allows $K = O(\mu_\Phi^{-1})$; in particular, Theorem 10 implies recovery of "most" K-sparse signals with $K = O(N/\log M)$ using frames such as Gabor frames. In essence, shifting our focus from uniform guarantees to typical guarantees allows us to break the square-root bottleneck for arbitrary frames.

Even though Theorem 10 allows us to obtain near-optimal sparse recovery results, it is still a statement about the computationally infeasible ℓ_0 optimization. We now shift our focus to the computationally tractable BP optimization and present guarantees concerning its typical behavior. Before proceeding further, it is worth pointing out that typicality in the case of ℓ_0 minimization is defined by a uniformly random support and a continuous distribution of the nonzero entries. In contrast, typicality in the case of BP will be defined in the following by a uniformly random support *but* nonzero entries whose phases are independent and uniformly distributed on the unit circle $\mathscr{C} = \{w \in \mathbb{C} : |w| = 1\}$.[6] The following theorem is once again due to Tropp and follows from combining results in [70] and [71].

Theorem 11 ([70, 71]). *Suppose that $y = \Phi x$ for a K-sparse signal x whose support is drawn uniformly at random and whose nonzero entries have independent phases distributed uniformly on $\mathscr{C}$. Further, let the frame Φ be such that $\mu_\Phi \leq (c_1 \log M)^{-1}$. Then under the assumption that x satisfies the sparsity constraint*

$$K < \min\left\{\frac{\mu_\Phi^{-2}}{16\log M}, \frac{M}{c_2^2\|\Phi\|^2 \log M}\right\}, \tag{16}$$

the solution of BP satisfies $\widehat{x}_1 = x$ with probability exceeding $1 - M^{-2\log 2} - M^{-1}$. Here, c_1 and c_2 are the same numerical constants specified in Theorem 10.

It is worth pointing out that there exists another variant of Theorem 11 that involves sparse signals whose nonzero entries are independently distributed with zero median. Theorem 11 once again provides us with a powerful typical behavior result.

[6] Recall the definition of the phase of a number $r \in \mathbb{C}$: $sgn(r) = \frac{r}{|r|}$.

Given approximately tight frames, it is possible to recover with high probability K-sparse signals using BP as long as $K = O(\mu_{\Phi}^{-2}/\log M)$. It is interesting to note here that unlike Section 2, which dictates the use of Grassmannian frames for best uniform guarantees, both Theorem 10 and Theorem 11 dictate the use of Grassmannian frames that are also approximately tight for best typical guarantees. Heuristically speaking, insisting on tightness of frames is what allows us to break the square-root bottleneck in the typical case.

3.2 Typical Regression of Sparse Signals

Instead of shifting the discussion to typical sparse estimation, we now focus on another important problem in the statistics literature, namely, sparse linear regression [32, 38, 66]. We will return to the problem of sparse estimation in Sect. 3.4. Given $y = \Phi x + n$ for a K-sparse vector $x \in \mathbb{R}^M$, the goal in sparse regression is to obtain an estimate $\widehat{x}$ of x such that the *regression error* $\|\Phi x - \Phi\widehat{x}\|_2$ is small. It is important to note that the only nontrivial result that can be provided for sparse linear regression is in the presence of noise, since the regression error in the absence of noise is always zero. Our focus in this section will be once again on the AWGN n with variance σ^2 and we will restrict ourselves to the LASSO solution (cf. $(P_{1,2})$). The following theorem provides guarantees for the typical behavior of LASSO as reported in a recent work of Candès and Plan [15].

Theorem 12 ([15]). *Suppose that $y = \Phi x + n$ for a K-sparse signal $x \in \mathbb{R}^M$ whose support is drawn uniformly at random and whose nonzero entries are jointly independent with zero median. Further, let the noise n be distributed as $\mathcal{N}(0, \sigma^2 Id)$, the frame Φ be such that $\mu_{\Phi} \leq (c_3 \log M)^{-1}$, and x satisfies the sparsity constraint $K \leq \frac{M}{c_4\|\Phi\|^2 \log M}$ for some positive numerical constants c_3 and c_4. Then the solution $\widehat{x}_{1,2}$ of LASSO computed with $\tau = 2\sqrt{2\sigma^2 \log M}$ satisfies*

$$\|\Phi x - \Phi\widehat{x}\|_2 \leq c_5 \sqrt{2\sigma^2 K \log M} \tag{17}$$

with probability at least $1 - 6M^{-2\log 2} - M^{-1}(2\pi \log M)^{-1/2}$. Here, the constant c_5 may be taken as $8(1+\sqrt{2})^2$.

There are two important things to note about Theorem 12. First, it states that the regression error of the LASSO is $O(\sqrt{\sigma^2 K \log M})$ with very high probability. This regression error is in fact very close to the near-ideal regression error of $O(\sqrt{\sigma^2 K})$. Second, the performance guarantees of Theorem 12 are a strong function of $\|\Phi\|$ but only a weak function of the worst-case coherence μ_{Φ}. In particular, Theorem 12 dictates that the sparsity level accommodated by the LASSO is primarily a function of $\|\Phi\|$ provided μ_{Φ} is not too large. If, for example, Φ was an approximately tight frame, then the LASSO can handle $K \approx O(N/\log M)$ regardless of the value of μ_{Φ}, provided $\mu_{\Phi} = O(1/\log M)$. In essence, the above theorem signifies the use of approximately tight frames with small-enough coherence in regression problems.

We conclude this subsection by noting that some of the techniques used in [15] to prove this theorem can in fact be used to also relax the dependence of BP on μ_Φ and obtain BP guarantees that primarily require small $\|\Phi\|$.

3.3 Typical Support Detection of Sparse Signals

It is often the case in many signal processing and statistics applications that one is interested in obtaining locations of the nonzero entries of a sparse signal x from a small number of measurements. This problem of *support detection* or *model selection* is of course trivial in the noiseless setting; exact recovery of sparse signals in this case implies exact recovery of the signal support: $support(\widehat{x}) = support(x)$. Given $y = \Phi x + n$ with nonzero noise n, however, the support detection problem becomes nontrivial. This is because a small estimation error in this case does not necessarily imply a small support detection error. Both exact support detection ($support(\widehat{x}) = support(x)$) and partial support detection ($support(\widehat{x}) \subset support(x)$) in the case of deterministic noise are very challenging (perhaps impossible) tasks. In the case of stochastic noise, however, both these problems become feasible, and we alluded to them in Theorem 8 and Theorem 9 in the context of uniform guarantees. In this subsection, we now focus on typical support detection in order to overcome the square-root bottleneck.

3.3.1 Support Detection Using the LASSO

The LASSO is arguably one of the standard tools used for support detection by the statistics and signal processing communities. Over the years, a number of theoretical guarantees have been provided for the LASSO support detection in [53, 73, 79]. The results reported in [53, 79] established that the LASSO asymptotically identifies the correct support under certain conditions on the frame Φ and the sparse signal x. Later, Wainwright in [73] strengthened the results of [53, 79] and made explicit the dependence of exact support detection using the LASSO on the smallest (in magnitude) nonzero entry of x. However, apart from the fact that the results reported in [53, 73, 79] are only asymptotic in nature, the main limitation of these works is that explicit verification of the conditions (such as the *irrepresentable condition* of [79] and the *incoherence condition* of [73]) that an arbitrary frame Φ needs to satisfy is computationally intractable for $K = \Omega(\mu_\Phi^{-1-\delta}), \delta > 0$.

The support detection results reported in [53, 73, 79] suffer from the square-root bottleneck because of their focus on uniform guarantees. Recently, Candès and Plan reported typical support detection results for the LASSO that overcome the square-root bottleneck of the prior work in the case of exact support detection [15].

Theorem 13 ([15]). *Suppose that $y = \Phi x + n$ for a K-sparse signal $x \in \mathbb{R}^M$ whose support is drawn uniformly at random and whose nonzero entries are jointly independent with zero median. Further, let the noise n be distributed as $\mathcal{N}(0, \sigma^2 Id)$,* *the frame Φ be such that $\mu_\Phi \le (c_6 \log M)^{-1}$, and let x satisfy the sparsity constraint $K \le \frac{M}{c_7 \|\Phi\|^2 \log M}$ for some positive numerical constants c_6 and c_7. Finally, let $\mathscr{K}$ be the support of x and suppose that*

$$\min_{i \in \mathscr{K}} |x_i| > 8\sqrt{2\sigma^2 \log M}. \tag{18}$$

Then the solution $\widehat{x}_{1,2}$ of LASSO computed with $\tau = 2\sqrt{2\sigma^2 \log M}$ satisfies

$$support(\widehat{x}_{1,2}) = support(x) \quad and \quad sgn(\widehat{x}_{\mathscr{K}}) = sgn(x_{\mathscr{K}}) \tag{19}$$

with probability at least $1 - 2M^{-1}\left((2\pi \log M)^{-1/2} + KM^{-1}\right) - O(M^{-2\log 2})$.

This theorem states that if the nonzero entries of the sparse signal x are significant in the sense that they roughly lie (modulo the logarithmic factor) above the noise floor σ, then the LASSO successfully carries out exact support detection for sufficiently sparse signals. Of course if any nonzero entry of the signal lies below the noise floor, then it is impossible to tell that entry apart from the noise itself. Theorem 13 is nearly optimal for exact model selection in this regard. In terms of the sparsity constraints, the statement of this theorem matches that of Theorem 12. Therefore, we once again see that frames that are approximately tight and have worst-case coherence that is not too large are particularly well-suited for sparse signal processing when used in conjunction with the LASSO.

3.3.2 Support Detection Using One-Step Thresholding

Although the support detection results reported in Theorem 13 are near optimal, it is desirable to investigate alternative solutions to the problem of typical support detection. This is because:

1. The LASSO requires the minimum singular value of the subframe of Φ corresponding to the support $\mathscr{K}$ to be bounded away from zero [15, 53, 73, 79]. While this is a plausible condition for the case when one is interested in estimating x, it is arguable whether this condition is necessary for the case of support detection.
2. Theorem 13 still lacks guarantees for $K = \Omega(\mu_\Phi^{-1-\delta}), \delta > 0$ in the case of deterministic nonzero entries of x.
3. Computational complexity of the LASSO for arbitrary frames tends to be $O(M^3 + NM^2)$. This makes the LASSO computationally demanding for large-scale model-selection problems.

In light of these concerns, a few researchers recently revisited the much older (and oft-forgotten) method of thresholding for support detection [2, 3, 37, 39, 57, 61]. The *One-Step Thresholding* (OST) algorithm, described in Algorithm 2, has computational complexity of only $O(NM)$ and it has been known to be nearly optimal for $M \times M$ orthonormal bases [31]. In this subsection, we focus on a recent result of Bajwa et al. [2, 3] concerning typical support detection using OST. The

Algorithm 2 The One-Step Thresholding (OST) Algorithm for Support Detection

Input: Unit-norm frame Φ, measurement vector y, and a threshold $\lambda > 0$

Output: Estimate of signal support $\widehat{\mathscr{K}} \subset \{1,\dots,M\}$

$z \leftarrow \Phi^* y$ {Form signal proxy}

$\widehat{\mathscr{K}} \leftarrow \{i \in \{1,\dots,M\} : |z_i| > \lambda\}$ {Select indices via OST}

forthcoming theorem in this regard relies on a notion of the *coherence property*, defined below.

Definition 3 (The Coherence Property [2, 3]). We say a unit-norm frame Φ satisfies the coherence property if

$$\text{(CP-1)} \quad \mu_\Phi \le \frac{0.1}{\sqrt{2\log M}} \qquad \text{and} \qquad \text{(CP-2)} \quad \nu_\Phi \le \frac{\mu_\Phi}{\sqrt{N}}.$$

In words, (CP-1) roughly states that the frame elements of Φ are not too similar, while (CP-2) roughly states that the frame elements of a unit-norm Φ are somewhat distributed within the N-dimensional unit ball. Note that the coherence property (*i*) does not require the singular values of the submatrices of Φ to be bounded away from zero and (*ii*) can be verified in polynomial time since it simply requires checking $\|G_\Phi - Id\|_{\max} \le (200\log M)^{-1/2}$ and $\|(G_\Phi - Id)1\|_\infty \le \|G_\Phi - Id\|_{\max}(M-1)N^{-1/2}$.

The implications of the coherence property are described in the following theorem. Before proceeding further, however, we first define some notation. We use $\mathsf{SNR} \doteq \|x\|^2/\mathbb{E}[\|n\|^2]$ to denote the *signal-to-noise ratio* associated with the support detection problem. Also, we use $x_{(\ell)}$ to denote the ℓ-th largest (in magnitude) nonzero entry of x. We are now ready to state the typical support detection performance of the OST algorithm.

Theorem 14 ([3]). *Suppose that $y = \Phi x + n$ for a K-sparse signal $x \in \mathbb{C}^M$ whose support $\mathscr{K}$ is drawn uniformly at random. Further, let $M \ge 128$, the noise n be distributed as complex Gaussian with mean 0 and covariance $\sigma^2 Id$, $n \sim \mathscr{CN}(0, \sigma^2 Id)$, and the frame Φ satisfy the coherence property. Finally, fix a parameter $t \in (0,1)$ and choose the threshold*

$$\lambda = \max\left\{\frac{1}{t}10\mu_\Phi\sqrt{N\cdot\mathsf{SNR}}, \frac{1}{1-t}\sqrt{2}\right\}\sqrt{2\sigma^2\log M}.$$

Then, under the assumption that $K \le N/(2\log M)$, the OST algorithm (Algorithm 2) guarantees with probability exceeding $1 - 6M^{-1}$ that $\widehat{\mathscr{K}} \subset \mathscr{K}$ and $\left|\mathscr{K} \setminus \widehat{\mathscr{K}}\right| \le (K-L)$, where L is the largest integer for which the following inequality holds:

$$x_{(L)} > \max\{c_8\sigma, c_9\mu_\Phi\|x\|\}\sqrt{\log M}. \tag{20}$$

Here, $c_8 \doteq 4(1-t)^{-1}$, $c_9 \doteq 20\sqrt{2}t^{-1}$, and the probability of failure is with respect to the true model $\mathscr{K}$ and the Gaussian noise n.

Algorithm 3 One-Step Thresholding (OST) for Sparse Signal Reconstruction

Input: Unit-norm frame Φ, measurement vector y, and a threshold $\lambda > 0$
Output: Sparse OST estimate $\widehat{x}^{OST}$

$\widehat{x}^{OST} \leftarrow 0$	{Initialize}
$z \leftarrow \Phi^* y$	{Form signal proxy}
$\widehat{\mathscr{K}} \leftarrow \{i : \lvert z_i \rvert > \lambda\}$	{Select indices via OST}
$\widehat{x}^{OST}_{\widehat{\mathscr{K}}} \leftarrow (\Phi_{\widehat{\mathscr{K}}})^{\dagger} y$	{Reconstruct signal via least-squares}

In order to put the significance of Theorem 14 into perspective, we recall the thresholding results obtained by Donoho and Johnstone [31]—which form the basis of ideas such as the wavelet denoising—for the case of $M \times M$ orthonormal bases. It was established in [31] that if Φ is an orthonormal basis, then hard thresholding the entries of $\Phi^* y$ at $\lambda = \Theta\left(\sqrt{\sigma^2 \log M}\right)$ results in oracle-like performance in the sense that one recovers (with high probability) the locations of all the nonzero entries of x that are above the noise floor (modulo $\log M$).

Now the first thing to note regarding Theorem 14 is the intuitively pleasing nature of the proposed threshold. Specifically, assume that Φ is an orthonormal basis and notice that, since $\mu_\Phi = 0$, the threshold $\lambda = \Theta\left(\max\left\{\mu_\Phi \sqrt{N \cdot \mathsf{SNR}}, 1\right\}\sqrt{\sigma^2 \log M}\right)$ proposed in the theorem reduces to the threshold proposed in [31] *and* Theorem 14 guarantees that thresholding recovers (with high probability) the locations of all the nonzero entries of x that are above the noise floor. The reader can convince oneself of this assertion by noting that $x_{(\ell)} = \Omega\left(\sqrt{\sigma^2 \log M}\right) \Rightarrow \ell \in \widehat{\mathscr{K}}$ in the case of orthonormal bases. Now consider instead frames that are not necessarily orthonormal but which satisfy $\mu_\Phi = O\left(N^{-1/2}\right)$ and $\nu_\Phi = O\left(N^{-1}\right)$. Then we have from the theorem that OST identifies (with high probability) the locations of the nonzero entries of x whose energies are greater than both the noise variance (modulo $\log M$) and the average energy per nonzero entry: $x^2_{(\ell)} = \Omega\left(\max\{\sigma^2 \log M, \|x\|^2/K\}\right) \Rightarrow \ell \in \widehat{\mathscr{K}}$. It is then easy to see in this case that if either the noise floor is high enough or the nonzero entries of x are roughly of the same magnitude then the simple OST algorithm leads to recovery of the locations of all the nonzero entries that are above the noise floor. Stated differently, the OST in certain cases has the oracle property in the sense of Donoho and Johnstone [31] *without* requiring the frame Φ to be an orthonormal basis.

3.4 Typical Estimation of Sparse Signals

Our goal in this section is to provide typical guarantees for the reconstruction of sparse signals from noisy measurements $y = \Phi x + n$, where the entries of the noise vector $n \in \mathbb{C}^N$ are independent, identical complex-Gaussian random variables with mean zero and variance σ^2. The reconstruction algorithm we analyze here is an extension of the OST algorithm described earlier for support detection. This

OST algorithm for reconstruction is described in Algorithm 3, and has been recently analyzed in [4]. The following theorem is due to Bajwa et al. [4] and shows that the OST algorithm leads to near-optimal reconstruction error for certain important classes of sparse signals.

Before a formal statement of the theorem, however, we need to define some more notation. We use $\mathscr{T}_\sigma(t) := \{i : |x_i| > \frac{2\sqrt{2}}{1-t}\sqrt{2\sigma^2 \log M}\}$ for any $t \in (0,1)$ to denote the locations of all the entries of x that, roughly speaking, lie above the *noise floor* σ. Also, we use $\mathscr{T}_\mu(t) := \{i : |x_i| > \frac{20}{t}\mu_\Phi \|x\|\sqrt{2\log M}\}$ to denote the locations of entries of x that, roughly speaking, lie above the *self-interference floor* $\mu_\Phi\|x\|$. Finally, we also need a stronger version of the coherence property for reconstruction guarantees.

Definition 4 (The Strong Coherence Property [3]). We say a unit norm frame Φ satisfies the *strong coherence property* if

$$\text{(SCP-1)} \quad \mu_\Phi \le \tfrac{1}{164 \log M} \qquad \text{and} \qquad \text{(SCP-2)} \quad \nu_\Phi \le \tfrac{\mu_\Phi}{\sqrt{N}}.$$

Theorem 15 ([4]). *Take a unit-norm frame Φ which satisfies the strong coherence property, pick $t \in (0,1)$, and choose $\lambda = \sqrt{2\sigma^2 \log M}\ \max\{\frac{10}{t}\mu_\Phi\sqrt{N\,\mathsf{SNR}}, \frac{\sqrt{2}}{1-t}\}$. Further, suppose $x \in \mathbb{C}^M$ has support $\mathscr{K}$ drawn uniformly at random from all possible K-subsets of $\{1,\dots,M\}$. Then provided*

$$K \le \frac{M}{c_{10}^2\|\Phi\|^2 \log M}, \tag{21}$$

Algorithm 3 produces $\widehat{\mathscr{K}}$ such that $\mathscr{T}_\sigma(t) \cap \mathscr{T}_\mu(t) \subseteq \widehat{\mathscr{K}} \subseteq \mathscr{K}$ and $\widehat{x}^{OST}$ such that

$$\|x - \widehat{x}^{OST}\| \le c_{11}\sqrt{\sigma^2|\widehat{\mathscr{K}}|\log M} + c_{12}\|x_{\mathscr{K}\setminus\widehat{\mathscr{K}}}\| \tag{22}$$

with probability exceeding $1 - 10M^{-1}$. Finally, defining $T := |\mathscr{T}_\sigma(t) \cap \mathscr{T}_\mu(t)|$, we further have

$$\|x - \widehat{x}\| \le c_{11}\sqrt{\sigma^2 K \log M} + c_{12}\|x - x_T\| \tag{23}$$

in the same probability event. Here, $c_{10} = 37\mathrm{e}$, $c_{11} = \frac{2}{1-\mathrm{e}^{-1/2}}$, and $c_{12} = 1 + \frac{\mathrm{e}^{-1/2}}{1-\mathrm{e}^{-1/2}}$ are numerical constants.

A few remarks are in order now for Theorem 15. First, if Φ satisfies the strong coherence property *and* Φ is nearly tight, then OST handles sparsity that is almost linear in N: $K = O(N/\log M)$ from (21). Second, the ℓ_2 error associated with the OST algorithm is the near-optimal (modulo the log factor) error of $\sqrt{\sigma^2 K \log M}$ *plus* the best T-term approximation error caused by the inability of the OST algorithm to recover signal entries that are smaller than $O(\mu_\Phi\|x\|\sqrt{2\log M})$. In particular, if the K-sparse signal x, the worst-case coherence μ_Φ, and the noise n together satisfy $\|x - x_T\| = O(\sqrt{\sigma^2 K \log M})$, then the OST algorithm succeeds with a near-optimal ℓ_2 error of $\|x - \widehat{x}\| = O(\sqrt{\sigma^2 K \log M})$. To see why this error is near-optimal, note that a K-dimensional vector of random entries with mean zero and variance σ^2 has expected squared norm $\sigma^2 K$; in here, the OST pays an additional log factor to find the locations of the K nonzero entries among the entire M-dimensional signal. It is important to recognize that the optimality condition $\|x - x_T\| = O(\sqrt{\sigma^2 K \log M})$ depends on the signal class, the noise variance, and the worst-case coherence of the frame; in particular, the condition is satisfied whenever $\|x_{\mathscr{K} \setminus \mathscr{T}_\mu(t)}\| = O(\sqrt{\sigma^2 K \log M})$, since

$$\|x - x_T\| \leq \|x_{\mathscr{K} \setminus \mathscr{T}_\sigma(t)}\| + \|x_{\mathscr{K} \setminus \mathscr{T}_\mu(t)}\| = O\Big(\sqrt{\sigma^2 K \log M}\Big) + \|x_{\mathscr{K} \setminus \mathscr{T}_\mu(t)}\|. \tag{24}$$

We conclude this subsection by stating a lemma from [4] that provides classes of sparse signals which satisfy $\|x_{\mathscr{K} \setminus \mathscr{T}_\mu(t)}\| = O(\sqrt{\sigma^2 K \log M})$ given sufficiently small noise variance and worst-case coherence.

Lemma 3. *Take a unit-norm frame Φ with worst-case coherence $\mu_\Phi \leq \frac{c_{13}}{\sqrt{N}}$ for some $c_{13} > 0$, and suppose that $K \leq \frac{M}{c_{14}^2 \|\Phi\|^2 \log M}$ for some $c_{14} > 0$. Fix a constant $\beta \in (0,1]$, and suppose the magnitudes of βK nonzero entries of x are some $\alpha = \Omega(\sqrt{\sigma^2 \log M})$, while the magnitudes of the remaining $(1-\beta)K$ nonzero entries are not necessarily same, but are smaller than α and scale as $O(\sqrt{\sigma^2 \log M})$. Then $\|x_{\mathscr{K} \setminus \mathscr{T}_\mu(t)}\| = O(\sqrt{\sigma^2 K \log M})$, provided $c_{13} \leq \frac{t c_{14}}{20\sqrt{2}}$.*

In words, Lemma 3 states that OST is near-optimal for those K-sparse signals whose entries above the noise floor have roughly the same magnitude. This subsumes a very important class of signals that appears in applications such as multi-label prediction [47], in which all the nonzero entries take values $\pm\alpha$.

4 Finite Frames for Detecting the Presence of Sparse Signals

In the previous sections, we discussed the role of frame theory in recovering and estimating sparse signals in different settings. We now consider a different problem, namely the problem of detecting the presence of a sparse signal in noise. In the simplest form, the problem is to decide whether an observed data vector is a realization from a hypothesized noise-only model or from a hypothesized signal-plus-noise model, where in the latter model the signal is sparse but the indices and the values of its nonzero elements are unknown. The problem is a binary hypothesis test of the form

$$\begin{cases} \mathscr{H}_0 : y = \Phi n \\ \mathscr{H}_1 : y = \Phi(x+n) \end{cases}, \tag{25}$$

where $x \in \mathbb{R}^M$ is a deterministic but unknown K-sparse signal, the measurement matrix $\Phi = \{\varphi_i\}_{i=1}^M$ is a frame for $\mathbb{R}^N$, $N \leq M$, which we get to design, and $n \in \mathbb{R}^M$ is a white Gaussian noise vector with covariance matrix $\mathbb{E}[nn^T] = (\sigma_n^2/M)Id$.

We assume here that the number of measurements N allowed for detection is fixed and pre-specified. We wish to decide whether the measurement vector $y \in \mathbb{R}^N$ belongs to model $\mathscr{H}_0$ or $\mathscr{H}_1$. This problem is fundamentally different from that of estimating a sparse signal, as the objective in detection typically is to maximize the probability of detection, while maintaining a low false alarm rate, or to minimize the total error probability or a Bayes risk, rather than to find the sparsest signal that fits a linear observation model. Unlike the signal estimation problem, the detection of sparse signals has received very little attention so far, with notable exceptions being [45, 56, 74]. But in particular, the design of optimal or near-optimal compressive measurement matrices for detection of sparse signals has only been scarcely addressed [76, 77]. In this section, we provide an overview of selected results by Zahedi et al. [76, 77], concerning the necessary and sufficient conditions for a frame Φ to optimize a measure of detection performance.

We look at the general problem of designing the measurement frame Φ to maximize the measurement signal-to-noise ratio (SNR), under $\mathscr{H}_1$, which is given by

$$\mathsf{SNR} = \frac{\|\Phi x\|^2}{\sigma_n^2/M}. \tag{26}$$

This is motivated by the fact that for the class of linear log-likelihood ratio detectors, where the log-likelihood ratio is a linear function of the data, the detection performance is improved by increasing SNR. In particular, for a Neyman-Pearson detector (see, e.g., [60]) with false alarm rate $P_F \leq \gamma$, the probability of detection

$$P_d = Q(Q^{-1}(\gamma) - \sqrt{\mathsf{SNR}}) \tag{27}$$

is monotonically increasing in SNR, where $Q(\cdot)$ is the Q-function, given by

$$Q(z) = \int_z^\infty e^{-w^2/2} dw. \tag{28}$$

In addition, maximizing SNR leads to maximum detection probability at a prespecified false alarm rate in an energy detector, which simply tests the energy of the measured vector y against a threshold. Without loss of generality, we assume that $\sigma_n^2 = 1$ and $\|x\|^2 = 1$, and we design Φ to maximize the measured signal energy $\|\Phi x\|^2$. To avoid coloring the noise vector n, that is, to keep the noise vector white, we constrain the measurement frame Φ to be Parseval, or tight with frame bound equal to one. That is, we only consider frames for which the frame operator $S_\Phi = \Phi\Phi^T$ is identity. From here on we simply refer to these frames as tight frames, but it is understood that all tight frames we consider in this section are in fact Parseval.

In solving the problem, one approach is to assume a value for the sparsity level K and design the measurement frame Φ based on this assumption. This approach, however, runs the risk that the true sparsity level might be different. An alternative approach is not to assume any specific sparsity level. Instead, when designing Φ, we prioritize the level of importance of different values of sparsity. In other words, we first find a set of solutions that are optimal for a K_1-sparse signal. Then, within this set, we find a subset of solutions that are also optimal for K_2-sparse signals. We follow this procedure until we find a subset that contains a family of optimal solutions for sparsity levels K_1, K_2, K_3, $\cdots$. This approach is known as a *lexicographic optimization* method (see, e.g., [33, 43, 48]). The measurement frame design naturally depends on one's assumptions about the unknown vector x. In the subsequent section, we review two different design problems, namely a *worst-case SNR design* and an *average SNR design*, following the developments of [76, 77].

We note that lexicographic optimizations have been employed earlier in [46] in the design of frames that have maximal robustness to erasures of frame coefficients. The analysis used in deriving the main results for the worst-case SNR design is similar in nature to that used in [46].

4.1 Wort-Case SNR Design

In the worst-case design for a sparsity level K, we consider the vector x that minimizes the SNR among all K-sparse signals and design the frame Φ to maximize this minimum SNR. Of course, when minimizing the SNR with respect to x, we have to find the minimum SNR with respect to both the locations and the values of the nonzero entries in x. To combine this with the lexicographic approach, we design the matrix Φ to maximize the *worst-case* detection SNR, where the worst-case is taken over all subsets of size K_i of elements of x, where K_i is the sparsity level considered at the ith level of lexicographic optimization. This is a design for robustness with respect to the worst sparse signal that can be produced.

Consider the Kth step of the lexicographic approach. In this step, the vector x is assumed to have up to K nonzero entries, and we assume $\|x\|^2 = 1$. But otherwise, we do not impose any constraints on the locations and the values of the nonzero entries of x. We wish to maximize the minimum (worst-case) SNR, produced by assigning the worst possible locations and values to the nonzero entries of the K-sparse vector x. Since we assume $\sigma_n^2 = 1$, this corresponds to a worst-case design for maximizing the signal energy $\|\Phi x\|^2$.

Let $\mathscr{B}_0$ be the set containing all $(N \times M)$ tight frames. We recursively define the set $\mathscr{B}_K$, $K = 1, 2, \ldots$, as the set of solutions to the following worst-case optimization problem [77]:

$$\begin{aligned} \max_{\Phi} \min_{x} \quad & \|\Phi x\|^2, \\ \text{s.t.} \quad & \Phi \in \mathscr{B}_{K-1}, \\ & \|x\| = 1, \\ & x \text{ is } K\text{-sparse}. \end{aligned} \tag{29}$$

The optimization problem for the Kth stage (29) involves a worst-case objective restricted to the set of solutions $\mathscr{B}_{K-1}$ from the $(K-1)$th problem. So, $\mathscr{B}_K \subset \mathscr{B}_{K-1} \subset \cdots \subset \mathscr{B}_0$.

Now let $\Omega = \{1, 2, \ldots, M\}$, and define Ω_K to be $\Omega_K = \{\omega \subset \Omega : |\omega| = K\}$. For any $\mathscr{T} \in \Omega_K$, let $x_{\mathscr{T}}$ be the subvector of size $(K \times 1)$ that contains all the components of x corresponding to indices in $\mathscr{T}$. Similarly, given a frame Φ, let $\Phi_{\mathscr{T}}$ be the

$(N \times K)$ submatrix consisting of all columns of Φ whose indices are in $\mathscr{T}$. Note that the vector $x_{\mathscr{T}}$ may have zero entries and hence is not necessarily the same as the support of x. Given $\mathscr{T} \in \Omega_K$, the product Φx can be replaced by $\Phi_{\mathscr{T}} x_{\mathscr{T}}$ instead. To consider the worst-case design, for any $\mathscr{T}$ we need to consider the $x_{\mathscr{T}}$ that minimizes $\|\Phi_{\mathscr{T}} x_{\mathscr{T}}\|^2$ and then also find the worst $\mathscr{T} \in \Omega_K$. Using this notation and after some simple algebra, the worst-case problem (29) can be posed as the following max-min problem [77]

$$(\mathscr{P}_K) \quad \begin{cases} \max_{\Phi} \min_{\mathscr{T}} \lambda_{\min}(\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}}), \\ \text{s.t.} \quad \Phi \in \mathscr{B}_{K-1}, \\ \qquad \mathscr{T} \in \Omega_K, \end{cases} \tag{30}$$

where $\lambda_{\min}(\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}})$ denotes the smallest eigenvalue of the frame sub-Gramian $G_{\Phi_{\mathscr{T}}} = \Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}}$.

To solve the worst-case design problem, we first find the solution set $\mathscr{B}_1$ for problem $(\mathscr{P}_1)$. Then, we find a subset $\mathscr{B}_2 \subset \mathscr{B}_1$ as the solution for $(\mathscr{P}_2)$. We continue this procedure for general sparsity level K.

Sparsity Level $K = 1$**.** If $K = 1$, then any $\mathscr{T}$ such that $|\mathscr{T}| = 1$ can be written as $\mathscr{T} = \{i\}$ with $i \in \Omega$, and $\Phi_{\mathscr{T}} = \varphi_i$ consists of only the ith column of Φ. Therefore, $\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}} = \|\varphi_i\|^2$, and $\mathscr{P}_1$ simplifies to

$$\begin{array}{ll} \max_{\Phi} \min_{i} & \|\varphi_i\|^2, \\ \quad \text{s.t.} & \Phi \in \mathscr{B}_0, \\ & i \in \Omega. \end{array} \tag{31}$$

We have the following result.

Theorem 16 ([77]). *The optimal value of the objective function of the max-min problem* (31) *is N/M, and a necessary and sufficient condition for $\hat{\Phi} \in \mathscr{B}_0$ to lie in the solution set $\mathscr{B}_1$ is for $\hat{\Phi} = \{\hat{\varphi}_i\}_{i=1}^M$ to be an equal-norm tight frame with $\|\hat{\varphi}_i\| = \sqrt{N/M}$, for $i = 1, 2, \ldots, M$.*

Sparsity Level $K = 2$**.** The next step is to solve $(\mathscr{P}_2)$. Given $\mathscr{T} \in \Omega_2$, the matrix $\Phi_{\mathscr{T}}$ consists of two columns, say, φ_i and φ_j. So, the matrix $\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}}$ in the max-min problem $(\mathscr{P}_2)$ is a (2×2) matrix:

$$\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}} = \begin{bmatrix} \langle \varphi_i, \varphi_i \rangle & \langle \varphi_i, \varphi_j \rangle \\ \langle \varphi_i, \varphi_j \rangle & \langle \varphi_j, \varphi_j \rangle \end{bmatrix}.$$

The solution for this case must lie among the family of optimal solutions for $K = 1$. In other words, the optimal solution $\hat{\Phi}$ must be an equal-norm tight frame with $\|\hat{\varphi}_i\| = \sqrt{N/M}$, for $i = 1, 2, \ldots, M$. Therefore, we have

$$\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}} = (N/M) \begin{bmatrix} 1 & \cos \alpha_{ij} \\ \cos \alpha_{ij} & 1 \end{bmatrix},$$

where α_{ij} is the angle between vectors φ_i and φ_j. The minimum possible eigenvalue of this matrix is

$$\lambda_{\min}(\Phi_{\mathscr{T}}^T \Phi_{\mathscr{T}}) = (N/M)(1-\mu_{\Phi}). \tag{32}$$

where μ_{Φ} is the worst-case coherence of the frame $\Phi = \{\varphi_i\}_{i=1}^M \in \mathscr{B}_1$, as defined in (1).

Now, let $\mu_{\min}$ be the minimum worst-case coherence

$$\mu_{\min} = \min_{\Phi \in \mathscr{B}_1} \mu_{\Phi} \tag{33}$$

for all frames in $\mathscr{B}_1$. We refer to the element of $\mathscr{B}_1$ that has the worst-case coherence $\mu_{\min}$ as a *Grassmannian equal-norm tight frame*.

We have the following theorem.

Theorem 17 ([77]). *The optimal value of the objective function of the max-min problem $(\mathscr{P}_2)$ is $(N/M)(1-\mu_{\min})$. A frame $\hat{\Phi}$ is in $\mathscr{B}_2$ if and only if the columns of $\hat{\Phi}$ form an equal-norm tight frame with norm values $\sqrt{N/M}$ and $\mu_{\hat{\Phi}} = \mu_{\min}$. In other words, the solution to $(\mathscr{P}_2)$ is an $N \times M$ Grassmannian equal-norm tight frame.*

Sparsity Level $K > 2$. We now consider the case where $K > 2$. In this case, $\mathscr{T} \in \Omega_K$ can be written as $\mathscr{T} = \{i_1, i_2, \ldots, i_K\} \subset \Omega$. From the previous results, we know that an optimal frame $\hat{\Phi} \in \mathscr{B}_K$ must be a Grassmannian equal-norm tight frame, with norms $\sqrt{N/M}$ and worst-case coherence $\mu_{\min}$.Taking this into account, the $(K \times K)$ matrix $\hat{\Phi}_{\mathscr{T}}^T \hat{\Phi}_{\mathscr{T}}$ in $(\mathscr{P}_K)$, $K > 2$, can be written as $\hat{\Phi}_{\mathscr{T}}^T \hat{\Phi}_{\mathscr{T}} = (N/M)[Id + A_{\mathscr{T}}]$ where $A_{\mathscr{T}}$ is given by

$$A_{\mathscr{T}} = \begin{bmatrix} 0 & \cos\hat{\alpha}_{i_1 i_2} & \ldots & \cos\hat{\alpha}_{i_1 i_k} \\ \cos\hat{\alpha}_{i_1 i_2} & 0 & \ldots & \cos\hat{\alpha}_{i_2 i_k} \\ \vdots & \vdots & \ddots & \vdots \\ \cos\hat{\alpha}_{i_1 i_k} & \cos\hat{\alpha}_{i_2 i_k} & \ldots & 0 \end{bmatrix}, \tag{34}$$

and $\cos\hat{\alpha}_{i_h i_f}$ is the cosine of the angle between frame elements $\hat{\varphi}_{i_h}$ and $\hat{\varphi}_{i_f}$, $i_h \neq i_f \in \mathscr{T}$. It is easy to see that

$$\lambda_{\min}(\hat{\Phi}_{\mathscr{T}}^T \hat{\Phi}_{\mathscr{T}}) = (N/M)(1 + \lambda_{\min}(A_{\mathscr{T}})). \tag{35}$$

So, the problem $(\mathscr{P}_K)$, $K > 2$, simplifies to

$$(\mathscr{P}_K) \quad \begin{cases} \max\limits_{\Phi} \min\limits_{\mathscr{T}} \lambda_{\min}(A_{\mathscr{T}}), \\ \quad \text{s.t.} \quad \Phi \in \mathscr{B}_{K-1}, \\ \qquad\qquad \mathscr{T} \in \Omega_K. \end{cases} \tag{36}$$

Solving the above problem however is not trivial. But we can at least bound the optimum value. Given $\mathscr{T} \in \Omega_K$, let $\hat{\delta}_{i_h i_f}$ and $\Delta_{\min}$ be

$$\hat{\delta}_{i_h i_f} = \mu_{\min} - |\cos \alpha_{i_h i_f}|, \quad i_h \neq i_f \in \mathscr{T}, \tag{37}$$

$$\Delta_{\min} = \min_{\mathscr{T} \in \Omega_K} \sum_{i_h \neq i_f \in \mathscr{T}} \hat{\delta}_{i_h i_f}. \tag{38}$$

Also, define $\hat{\Delta}$ in the following way:

$$\hat{\Delta} = \min_{\mathscr{T} \in \Omega_K} \sum_{i_h \neq i_f \in \mathscr{T}} \hat{\delta}_{i_h i_f}.$$

We have the following theorem.

Theorem 18 ([77]). *The optimal value of the objective function of the max-min problem* $(\mathscr{P}_K)$ *for* $K > 2$ *lies between* $(N/M)(1 - \binom{K}{2}\mu_{\min} + \Delta_{\min})$ *and* $(N/M)(1 - \mu_{\min})$.

Before we conclude the worst-case SNR design, a few remarks are in order.

1. Examples of uniform tight frames and their methods of construction can be found in [8, 13, 19, 20], and the references therein.
2. In the case where $K = 2$, $\hat{\Phi}^T_{\mathscr{T}} \hat{\Phi}_{\mathscr{T}}$ associated with the frame $\hat{\Phi}$ identified in Theorem 17, has the largest minimum eigenvalue $(N/M)(1 - \mu_{\min})$ and the smallest maximum eigenvalue $(N/M)(1 + \mu_{\min})$ among all $\Phi \in \mathscr{B}_1$ and $\mathscr{T} \in \Omega_2$. This means that the solution $\hat{\Phi}$ to $(\mathscr{P}_2)$ is an RIP matrix of order 2 with optimal RIC $\delta_2 = \mu_{\min}$.
3. In general, the minimum worst-case coherence $\mu_{\min}$ of the solution $\hat{\Phi}$ to $(\mathscr{P}_K)$, $K \geq 2$, is bounded below by the Welch bound (see Lemma 2). However, when $1 \leq N \leq M - 1$ and

$$M \leq \min\{N(N+1)/2, (M-N)(M-N+1)/2\} \tag{39}$$

the Welch bound can be met [64]. For such a case, all frame angles are equal and the solution to $(\mathscr{P}_K)$ for $K \geq 2$ is an *equiangular equal-norm tight frame*. Such frames are *Grassmannian line packings* (see, e.g., [8, 21, 24, 49, 52, 58, 63, 64, 65]).

4.2 Average-Case Design

Let us now assume that in (25) the locations of nonzero entries of x are random but their values are deterministic and unknown. We wish to find the frame Φ that maximizes the expected value of the minimum SNR. The expectation is taken with respect to a random index set with uniform distribution over the set of all possible subsets of size K_i of the index set $\{1, 2, \ldots, M\}$ of elements of x. The minimum SNR, whose expected value we wish to maximize, is calculated with respect to the values of the entries of the vector x for each realization of the random index set.

Let $\mathscr{T}_K$ be a random variable that is uniformly distributed over Ω_K. Then $p_{\mathscr{T}_K}(t) = 1/\binom{M}{K}$ is the probability that $\mathscr{T}_K = t$ for $t \in \Omega_K$. Our goal is to find a measurement frame Φ that maximizes the expected value of the minimum SNR, where the expectation is taken with respect to the random $\mathscr{T}_K$, and the minimum is taken with respect to the entries of the vector x on $\mathscr{T}_K$. Taking into account the simplifying steps used earlier for the worst-case problem and also adopting the lexicographic approach, the problem of maximizing the average SNR can then be formulated in the following way.

Let $\mathscr{N}_0$ be the set containing all $(N \times M)$ tight frames. Then for $K = 1, 2, \ldots$, recursively define the set $\mathscr{N}_K$ as the solution set to the following optimization problem:

$$\begin{cases} \max_{\Phi} \mathbb{E}_{\mathscr{T}_K} \min_{x_K} & \|\Phi_{\mathscr{T}_K} x_K\|^2, \\ \quad\text{s.t.} & \Phi \in \mathscr{N}_{K-1}, \\ & \|x_K\| = 1, \end{cases} \tag{40}$$

where $\mathbb{E}_{\mathscr{T}_K}$ is the expectation with respect to $\mathscr{T}_K$. As before, the $(N \times K)$ matrix $\Phi_{\mathscr{T}_K}$ is a submatrix of Φ whose column indices are in $\mathscr{T}_K$. The above problem can be simplified to the following [77]:

$$(\mathscr{F}_K) \qquad \begin{cases} \max_{\Phi} \ \mathbb{E}_{\mathscr{T}_K} \lambda_{\min}(\Phi_{\mathscr{T}_K}^T \Phi_{\mathscr{T}_K}), \\ \text{s.t.} \ \ \Phi \in \mathscr{N}_{K-1}. \end{cases} \tag{41}$$

To solve the lexicographic problems $(\mathscr{F}_K)$, we follow the same method we used earlier for the worst-case problem, i.e., we begin by solving problem $(\mathscr{F}_1)$. Then, from the solution set $\mathscr{N}_1$, we find optimal solutions for the problem $(\mathscr{F}_2)$, and so on.

Sparsity Level $K = 1$. Assume that the signal x is 1-sparse. So, there are $\binom{M}{1} = M$ different possibilities to build the matrix $\Phi_{\mathscr{T}_1}$ from the matrix Φ. The expectation in problem $(\mathscr{F}_1)$ can be written as:

$$\mathbb{E}_{\mathscr{T}_1} \lambda_{\min}(\Phi_{\mathscr{T}_1}^T \Phi_{\mathscr{T}_1}) = \sum_{t \in \Omega_1} p_{\mathscr{T}_1}(t) \lambda_{\min}(\Phi_t^T \Phi_t) = \sum_{i=1}^{M} p_{\mathscr{T}_1}(\{i\}) \|\varphi_i\|^2 = \frac{N}{M}. \tag{42}$$

The following result holds.

Theorem 19 ([77]). *The optimal value of the objective function of problem $(\mathscr{F}_1)$ is N/M. This value is obtained by using any $\Phi \in \mathscr{N}_0$, i.e., any tight frame.*

Theorem 19 shows that unlike the worst-case problem, any tight frame is an optimal solution for the problem $(\mathscr{F}_1)$. Next, we study the case where the signal x is 2-sparse.

Sparsity Level $K = 2$. For problem $(\mathscr{F}_2)$, the expected value term $\mathbb{E}_{\mathscr{T}_2} \lambda_{\min}(\Phi_{\mathscr{T}_2}^T \Phi_{\mathscr{T}_2})$ is equal to

$$\sum_{t \in \Omega_2} p_{\mathscr{T}_2}(t) \lambda_{\min}(\Phi_t^T \Phi_t) = \frac{2}{M(M-1)} \sum_{j=2}^{M} \sum_{i=1}^{j-1} \lambda_{\min}(\Phi_{\{i,j\}}^T \Phi_{\{i,j\}}). \tag{43}$$

In general, solving the family of problems $(\mathscr{F}_K)$, $K = 2, 3, \ldots$, is not trivial. However, if we constrain ourselves to the class of equal-norm tight frames, which also arise in solving the worst-case problem, we can establish necessary and sufficient conditions for optimality. These conditions are different from those for the worst-case problem and as we will show next the optimal solution here is an equal-norm tight frame for which a cumulative measure of coherence is minimal.

Let $\mathscr{M}_1$ be defined as $\mathscr{M}_1 = \{\Phi : \Phi \in \mathscr{N}_1, \|\varphi_i\| = \sqrt{N/M}, \forall i \in \Omega\}$. Also, for $K = 2, 3, \ldots$, recursively define the set $\mathscr{M}_K$ as the solution set to the following optimization problem:

$$(\mathscr{F}_K^{'}) \quad \begin{cases} \max\limits_{\Phi} \ \mathbb{E}_{\mathscr{T}_K} \lambda_{\min}(\Phi_{\mathscr{T}_K}^T \Phi_{\mathscr{T}_K}), \\ \text{s.t.} \ \ \Phi \in \mathscr{M}_{K-1}. \end{cases} \tag{44}$$

We will concentrate on solving the above family of problems instead of $(\mathscr{F}_K)$, $K = 2, 3, \ldots$. We have the following results.

Theorem 20 ([77]). *The frame Φ is in $\mathscr{M}_2$ if and only if the sum coherence of Φ, i.e., $\sum_{j=2}^{M} \sum_{i=1}^{j-1} |\langle \varphi_i, \varphi_j \rangle| / (\|\varphi_i\| \|\varphi_j\|)$, is minimized.*

Theorem 20 shows that for problem $(\mathscr{F}_2^{'})$, angles between elements of the equal-norm tight frame Φ should be designed in a different way than for the worst-case problem. For example, an equiangular tight frame of $M = 2N$ in N dimensions, with vectors of equal norm $\sqrt{1/2}$, has worst-case coherence $1/(2\sqrt{2N-1})$ and sum coherence $N\sqrt{2N-1}/2$, while two copies of an orthonormal basis form a frame with worst-case coherence $1/2$ and sum coherence $N/2$. While it is not clear whether copies of orthonormal bases form tight frames with minimal sum coherence, this example certainly illustrates that Grassmannian frames do not, in general, result in minimal sum coherence. To the best of our knowledge, no general method for constructing tight frames with minimal sum coherence has been proposed so far.

The following lemma provides bounds on the sum coherence of an equal-norm tight frame.

Lemma 4 ([77]). *For an equal-norm tight frame Φ with norm values $\sqrt{N/M}$, the following inequalities hold:*

$$c|(M/N-1) - 2(M-1)\mu_\Phi^2| \leq \sum_{j=2}^{M} \sum_{i=1}^{j-1} |\langle \varphi_i, \varphi_j \rangle| \leq c(M-1)\mu_\Phi^2,$$

where

$$c = \left(\frac{(N/M)^2}{1 - 2(N/M)} \right) \left(\frac{M(M-2)}{2} \right).$$

Sparsity Level $K > 2$**.** Similar to the worst-case problem, solving problems $(\mathscr{F}_K^{'})$ for $K > 2$ is not trivial. This is because the solution sets for these problems all lie in $\mathscr{M}_2$, and $(\mathscr{F}_2^{'})$ is still an open problem. The following lemma provides a lower bound for the optimal objective function of $(\mathscr{F}_K^{'})$, $K > 2$.

Lemma 5 ([77]). *The optimal value of the objective function for problem* ($\mathscr{F}_K'$)*, $K > 2$, is bounded below by* $(N/M)(1-(K(K-1)/2)\mu_\Phi)$.

We conclude this section by giving a summary. In the worst-case SNR problem, the optimal measurement matrix is a Grassmannian equal-norm tight frame for most—and a equal-norm tight frame for all—sparse signals. In the average SNR problem, we limited ourselves to the class of equal-norm tight frames and showed that the optimal measurement frame is an equal-norm tight frame that has minimum sum coherence.

5 Other Topics

As mentioned earlier, this chapter covers only a small subset of the results in the sparse signal processing literature. Our aim has been to simply highlight the central role that finite frames and their geometric measures, such as spectral norm, worst-case coherence, average coherence, and sum coherence, play in the development of sparse signal processing methods. But many developments, which also involve finite frames, have not been covered. For example, there is a large body of work on signal processing of *compressible* signals. These are signals that are not sparse, but their entries decay in magnitude according to a particular power law. Many of the results covered in this chapter on estimating sparse signals have counterparts for compressible signals. The reader is referred to [17, 23, 26, 27] for examples of such results. Another example is the estimation and recovery of *block-sparse* signals, where the nonzero entries of the signal to be estimated are either clustered or the signal has a sparse representation in a fusion frame. Again, the majority of the results on the estimation and recovery of sparse signals can be extended to block-sparse signals. The reader is referred to [9, 35, 62, 78] and the references therein.

References

1. IEEE signal processing mag., special issue on compressive sampling (2008)
2. Bajwa, W.U., Calderbank, R., Jafarpour, S.: Model selection: Two fundamental measures of coherence and their algorithmic significance. In: Proc. IEEE Intl. Symp. Information Theory (ISIT'10), pp. 1568–1572. Austin, TX (2010)
3. Bajwa, W.U., Calderbank, R., Jafarpour, S.: Why Gabor frames? Two fundamental measures of coherence and their role in model selection. J. Commun. Netw. **12**(4), 289–307 (2010)
4. Bajwa, W.U., Calderbank, R., Mixon, D.G.: Two are better than one: Fundamental parameters of frame coherence. Appl. Comput. Harmon. Anal. (2011). DOI doi:10.1016/j.acha.2011.09.005. URL `http://arxiv.org/abs/1103.0435`. In press
5. Bajwa, W.U., Haupt, J., Raz, G., Nowak, R.: Compressed channel sensing. In: Proc. 42nd Annu. Conf. Information Sciences and Systems (CISS'08), pp. 5–10. Princeton, NJ (2008)
6. Ben-Haim, Z., Eldar, Y.C., Elad, M.: Coherence-based perfromance guarantees for estimating a sparse vector under random noise. IEEE Trans. Signal Processing **58**(10), 5030–5043 (2010)
7. Blumensath, T., Davies, M.E.: Iterative hard thresholding for compressed sensing. Appl. Comput. Harmon. Anal. **27**(3), 265–274 (2009)

8. Bodmann, B.G., Paulsen, V.I.: Frames, graphs and erasures. Linear Algebra and its Applications **404**, 118–146 (2005)
9. Boufounos, P., Kutynio, G., Rahut, H.: Sparse recovery from combined fusion frame measurements. IEEE Trans. Info. Theory **57**(6), 3864–3876 (2011)
10. Bourgain, J., Dilworth, S.J., Ford, K., Konyagin, S.V., Kutzarova, D.: Breaking the k^2 barrier for explicit RIP matrices. In: Proc. 43rd Annu. ACM Symp. Theory Computing (STOC'11), pp. 637–644. San Jose, California (2011)
11. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge, U.K. (2004)
12. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review **51**(1), 34–81 (2009)
13. Calderbank, R., Casazza, P., Heinecke, A., Kutyniok, G., Pezeshki, A.: Sparse fusion frames: existence and construction. Advances in Computational Mathematics **35**, 1–31 (2011)
14. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. In: C. R. Acad. Sci., Ser. I, vol. 346, pp. 589–592. Paris (2008)
15. Candès, E.J., Plan, Y.: Near-ideal model selection by ℓ_1 minimization. Ann. Statist. **37**(5A), 2145–2177 (2009)
16. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)
17. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory **52**(12), 5406–5425 (2006)
18. Candès, E.J., Tao, T.: The Dantzig selector: Statistical estimation when p is much larger than n. Ann. Statist. **35**(6), 2313–2351 (2007)
19. Casazza, P., Fickus, M., Mixon, D., Wang, Y., Zhou, Z.: Constructing tight fusion frames. Applied and Computational Harmonic Analysis **30**, 175–187 (2011)
20. Casazza, P., Leon, M.: Existence and construction of finite tight frames. J. Concr. Appl. Math **4**(3), 277–289 (2006)
21. Casazza, P.G., Kovačević, J.: Equal-norm tight frames with erasures. Appl. Comput. Harmon. Anal. **18**(2–4), 387–430 (2003)
22. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Scientific Comput. **20**(1), 33–61 (1998)
23. Cohen, A., Dahmen, W., Devore, R.A.: Compressed sensing and best k-term approximation. J. Amer. Math. Soc. **22**(1), 211–231 (2009)
24. Conway, J.H., Hardin, R.H., Sloane, N.J.A.: Packing lines, planes, etc.: Packings in Grassmannian spaces. Experimental Mathematics **5**(2), 139–159 (1996)
25. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction. IEEE Trans. Inform. Theory **55**(5), 2230–2249 (2009)
26. Devore, R.A.: Nonlinear approximation. In: A. Iserles (ed.) Acta Numerica, vol. 7, pp. 51–150. Cambridge University Press, Cambridge, U.K. (1998)
27. Donoho, D.L.: Compressed sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)
28. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. Proc. Natl. Acad. Sci. **100**(5), 2197–2202 (2003)
29. Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory **52**(1), 6–18 (2006)
30. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inform. Theory **47**(7), 2845–2862 (2001)
31. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
32. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Statist. **32**(2), 407–451 (2004)
33. Ehrgott, M.: Multicriteria Optimization, 2nd edn. Springer (2005)
34. Eldar, Y., Kutyniok, G.: Compressed Sensing: Theory and Applications. Cambridge University Press (in press)

35. Eldar, Y.C., Kuppinger, P., Bölcskei, H.: Block-sparse signals: Uncertainty relations and efficient recovery. IEEE Trans. Signal Processing **58**(6), 3042–3054 (2010)
36. Fickus, M., Mixon, D.G., Tremain, J.C.: Steiner equiangular tight frames. Linear Algebra Appl. (2011). DOI 10.1016/j.laa.2011.06.027
37. Fletcher, A.K., Rangan, S., Goyal, V.K.: Necessary and sufficient conditions for sparsity pattern recovery. IEEE Trans. Inform. Theory **55**(12), 5758–5772 (2009)
38. Foster, D.P., George, E.I.: The risk inflation criterion for multiple regression. Ann. Statist. **22**(4), 1947–1975 (1994)
39. Genovese, C., Jin, J., Wasserman, L.: Revisiting marginal regression (2009). URL `http://arxiv.org/abs/0911.4080`. Submitted for publication (arXiv:0911.4080v1 [math.ST])
40. Geršgorin, S.A.: Über die Abgrenzung der Eigenwerte einer Matrix. Izv. Akad. Nauk SSSR Ser. Fiz.-Mat. **6**, 749–754 (1931)
41. Gorodnitsky, I.F., Rao, B.D.: Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Trans. Signal Processing **45**(3), 600–616 (1997)
42. Gribonval, R., Nielsen, M.: Sparse representations in unions of bases. IEEE Trans. Inform. Theory **49**(12), 3320–3325 (2003)
43. Hajek, B., Seri, P.: Lex-optimal online multiclass scheduling with hard deadlines. Mathematics of Operations Research **30**(3), 562–596 (2005)
44. Haupt, J., Bajwa, W.U., Raz, G., Nowak, R.: Toeplitz compressed sensing matrices with applications to sparse channel estimation. IEEE Trans. Inform. Theory **56**(11), 5862–5875 (2010)
45. Haupt, J., Nowak, R.: Compressive sampling for signal detection. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 3, pp. III–1509–III–1512 (2007)
46. Holmes, R.B., Paulsen, V.I.: Optimal frames for erasures. Linear Algebra and its Applications **377**(15), 31–51 (2004)
47. Hsu, D., Kakade, S., Langford, J., Zhang, T.: Multi-label prediction via compressed sensing. In: Advances in Neural Information Processing Systems, pp. 772–780 (2009)
48. Isermann, H.: Linear lexicographic optimization. OR Spectrum **4**(4), 223–228 (1982)
49. Kutyniok, G., Pezeshki, A., Calderbank, R., Liu, T.: Robust dimension reduction, fusion frames, and Grassmannian packings. Appl. Comput. Harmon. Anal. **26**(1), 64–76 (2009)
50. Lancaster, P., Tismenetsky, M.: The Theory of Matrices, second edn. Academic Press, Orlando, FL (1985)
51. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Processing **41**(12), 3397–3415 (1993)
52. Malozemov, V.N., Pevnyi, A.B.: Equiangular tight frames. Journal of Mathematical Sciences **157**(6), 789–815 (2009)
53. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. Ann. Statist. **34**(3), 1436–1462 (2006)
54. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**(2), 227–234 (1995)
55. Needell, D., Tropp, J.A.: CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**(3), 301–321 (2009)
56. Paredes, J., Wang, Z., Arce, G., Sadler, B.: Compressive matched subspace detection. In: Proc. 17th European Signal Processing Conference, pp. 120–124. Glasgow, Scotland (2009)
57. Reeves, G., Gastpar, M.: A note on optimal support recovery in compressed sensing. In: Proc. 43rd Asilomar Conf. Signals, Systems and Computers. Pacific Grove, CA (2009)
58. Renes, J.: Equiangular tight frames from Paley tournaments. Linear Algebra Appl. **426**(2–3), 497–501 (2007)
59. Santosa, F., Symes, W.W.: Linear inversion of band-limited reflection seismograms. SIAM J. Sci. Statist. Comput. **7**(4), 1307–1330 (1986)
60. Scharf, L.L.: Statistical Signal Processing. Addison-Wesley, Cambridge, MA (1991)
61. Schnass, K., Vandergheynst, P.: Average performance analysis for thresholding. IEEE Signal Processing Lett. **14**(11), 828–831 (2007)

62. Stojnic, M., Parvaresh, F., Hassibi, B.: On the representaton of block-sparse signals with an optimal number of meausrements. IEEE Trans. Signal Processing **57**(8), 3075–3085 (2009)
63. Strohmer, T.: A note on equiangular tight frames. Linear Algebra and its Applications **429**(1), 326–330 (2008)
64. Strohmer, T., Heath Jr., R.W.: Grassmannian frames with applications to coding and communication. Appl. Comput. Harmon. Anal. **14**(3), 257–275 (2003)
65. Sustik, M., Tropp, J.A., Dhillon, I.S., Heath Jr., R.W.: On the existence of equiangular tight frames. Linear Algebra Appl. **426**(2–3), 619–635 (2007)
66. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B **58**(1), 267–288 (1996)
67. Tropp, J., Gilbert, A., Muthukrishnan, S., Strauss, M.: Improved sparse approximation over quasiincoherent dictionaries. In: Proc. IEEE Conf. Image Processing (ICIP'03), pp. 37–40 (2003)
68. Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory **50**(10), 2231–2242 (2004)
69. Tropp, J.A.: Just relax: Convex programming methods for identifying sparse signals in noise. IEEE Trans. Inform. Theory **52**(3), 1030–1051 (2006)
70. Tropp, J.A.: Norms of random submatrices and sparse approximation. In: C. R. Acad. Sci., Ser. I, vol. 346, pp. 1271–1274. Paris (2008)
71. Tropp, J.A.: On the conditioning of random subdictionaries. Appl. Comput. Harmon. Anal. **25**, 1–24 (2008)
72. Tropp, J.A., Wright, S.J.: Computational methods for sparse solution of linear inverse problems. Proc. IEEE **98**(5), 948–958 (2010)
73. Wainwright, M.J.: Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). IEEE Trans. Inform. Theory **55**(5), 2183–2202 (2009)
74. Wang, Z., Arce, G., Sadler, B.: Subspace compressive detection for sparse signals. In: IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 3873–3876 (2008)
75. Welch, L.: Lower bounds on the maximum cross correlation of signals. IEEE Trans. Inform. Theory **20**(3), 397–399 (1974)
76. Zahedi, R., Pezeshki, A., Chong, E.K.P.: Robust measurement design for detecting sparse signals: Equiangular uniform tight frames and Grassmannian packings. In: Proc. 2010 American Control Conference (ACC). Baltimore, MD (2010)
77. Zahedi, R., Pezeshki, A., Chong, E.K.P.: Measurement design for detecting sparse signals. Physical Communications (2011). DOI 10.1016/j.phycom.2011.09.007
78. Zelnik-Manor, L., Rosenblum, K., Eldar, Y.C.: Sensing matrix optimization for block-sparse decoding. IEEE Trans. Signal Processing **59**(9), 4300–4312 (2011)
79. Zhao, P., Yu, B.: On model selection consistency of lasso. J. Machine Learning Res. **7**, 2541–2563 (2006)

Index

basis pursuit, 7–10, 12–16
  denoising, *see* LASSO
  with inequality constraint, 9, 10, 12

coherence, *see* frame coherence
coherence property, 18
  strong coherence property, 20
compressed sensing, 2, 12
  square-root bottleneck, *see* square-root bottleneck, sparse signal processing

frame coherence
  average coherence, 3, 29
  sum coherence, 3, 28, 29
  worst-case coherence, 3–9, 12, 13, 15, 17, 20, 21, 25, 26, 28, 29

Gabor frame, 7, 14
Geršgorin circle theorem, 5, 13
Grassmannian frame, 4, 6–8, 12, 15, 25, 28
Grassmannian line packings, 26

LASSO, 11, 12, 15–17

Neyman-Pearson detector, 22

one-step thresholding, 17–21
orthogonal matching pursuit, 8–12

restricted isometry property, 12

spark, *see* unique representation property
sparse signal processing, 2–4, 7, 17, 29
  square-root bottleneck, 12–16
sparse signals
  estimation
    average guarantees, 19–21
    uniform guarantees, 9–12
  recovery
    average guarantees, 13–15
    uniform guarantees, 4–9
  regression
    average guarantees, 15–16
  signal detection
    average design, 26–29
    worst-case SNR design, 23–26
  support detection
    average guarantees, 16–19

unique representation property, 5, 6

Welch bound, 6, 12, 26