1 Uncertainty Relations and Sparse Signal Recovery

Erwin Riegler and Helmut Bölcskei

1.1 Abstract

This chapter provides a principled introduction to uncertainty relations underlying sparse signal recovery. We start with the seminal work by Donoho and Stark, 1989, which defines uncertainty relations as upper bounds on the operator norm of the band-limitation operator followed by the time-limitation operator, generalize this theory to arbitrary pairs of operators, and then develop—out of this generalization—the coherence-based uncertainty relations due to Elad and Bruckstein, 2002, as well as uncertainty relations in terms of concentration of 1-norm or 2-norm. The theory is completed with the recently discovered set-theoretic uncertainty relations which lead to best possible recovery thresholds in terms of a general measure of parsimony, namely Minkowski dimension. We also elaborate on the remarkable connection between uncertainty relations and the "large sieve", a family of inequalities developed in analytic number theory. It is finally shown how uncertainty relations allow to establish fundamental limits of practical signal recovery problems such as inpainting, declipping, super-resolution, and denoising of signals corrupted by impulse noise or narrowband interference. Detailed proofs are provided throughout the chapter.

1.2 Introduction

The uncertainty principle in quantum mechanics says that certain pairs of physical properties of a particle, such as position and momentum, can be known to within a limited precision only [1]. Uncertainty relations in signal analysis [2–5] state that a signal and its Fourier transform can not both be arbitrarily well concentrated; corresponding mathematical formulations exist for square-integrable or integrable functions [6, 7], for vectors in $(\mathbb{C}^m, \|\cdot\|_2)$ or $(\mathbb{C}^m, \|\cdot\|_1)$ [6–10], and for finite abelian groups [11, 12]. These results feature prominently in many areas of the mathematical data sciences. Specifically, in compressed sensing [6–9, 13, 14] uncertainty relations lead to sparse signal recovery thresholds, in Gabor and Wilson frame theory [15] they characterize limits on the time-frequency localization of frame elements, in communications [16] they play a fundamental role in the design of pulse shapes for orthogonal frequency division multiplexing (OFDM) systems [17], in the theory of partial differential equations they serve to characterize existence and smoothness properties of solutions [18], and in coding theory they help to understand questions around the existence of good cyclic codes [19].

This chapter provides a principled introduction to uncertainty relations underlying sparse signal recovery, starting with the seminal work by Donoho and Stark [6], ranging over the Elad-Bruckstein coherence-based uncertainty relation for general pairs of orthonormal bases [8], later extended to general pairs of dictionaries [10], to the recently discovered set-theoretic uncertainty relation [13] which leads to information-theoretic recovery thresholds for general notions of parsimony. We also elaborate on the remarkable connection [7] between uncertainty relations for signals and their Fourier transforms—with concentration measured in terms of support—and the "large sieve", a family of inequalities involving trigonometric polynomials, originally developed in the field of analytic number theory [20, 21].

Uncertainty relations play an important role in data science beyond sparse signal recovery, specifically in the sparse signal separation problem, which comprises numerous practically relevant applications such as (image or audio signal) inpainting, declipping, super-resolution, and the recovery of signals corrupted by impulse noise or by narrowband interference. We provide a systematic treatment of the sparse signal separation problem and develop its limits out of uncertainty relations for general pairs of dictionaries as introduced in [10]. While the flavor of these results is that beyond certain thresholds something is not possible, for example a nonzero vector can not be concentrated with respect to two different orthonormal bases beyond a certain limit, uncertainty relations can also reveal that something unexpected is possible. Specifically, we demonstrate that signals that are sparse in certain bases can be recovered in a stable fashion from partial and noisy observations.

In practice one often encounters more general concepts of parsimony, such as, e.g., manifold structures and fractal sets. Manifolds are prevalent in the data sciences, e.g., in compressed sensing [22–27], machine learning [28], image processing [29, 30], and handwritten digit recognition [31]. Fractal sets find application in image compression and in modeling of Ethernet traffic [32]. In the last part of this chapter, we develop an information-theoretic framework for sparse signal separation and recovery, which applies to arbitrary signals of "low description complexity". The complexity measure our results are formulated in, namely Minkowski dimension, is agnostic to signal structure and goes beyond the notion of sparsity in terms of the number of nonzero entries or concentration in 1-norm or 2-norm. The corresponding recovery thresholds are information-theoretic in the sense of applying to arbitrary signal structures, and provide results of best possible nature that are, however, not constructive in terms of recovery algorithms.

To keep the exposition simple and to elucidate the main conceptual aspects, we restrict ourselves to the finite-dimensional cases $(\mathbb{C}^m, \|\cdot\|_2)$ and $(\mathbb{C}^m, \|\cdot\|_1)$ throughout. References to uncertainty relations for the infinite-dimensional case will be given wherever possible and appropriate. Some of the results in this chapter have not been reported before in the literature. Detailed proofs will be provided for most of the statements with the goal of allowing the reader to acquire a technical working knowledge that can serve as a basis for further own research.

The chapter is organized as follows. In Sections 1.3 and 1.4, we derive uncertainty relations for vectors in $(\mathbb{C}^m, \|\cdot\|_2)$ and $(\mathbb{C}^m, \|\cdot\|_1)$, respectively, discuss the connection to the large sieve, present applications to noisy signal recovery problems, and establish a fundamental relation between uncertainty relations for sparse vectors and null-space properties of the accompanying dictionary matrices. Section 1.5 is devoted to understanding the role of uncertainty relations in sparse signal separation problems. In Section 1.6, we generalize the classical sparsity notion as used in compressed sensing to a more comprehensive concept of description complexity, namely, lower modified Minkowski dimension, which in turn leads to a set-theoretic null-space property and corresponding recovery thresholds. Section 1.7 presents a large sieve inequality in $(\mathbb{C}^m, \|\cdot\|_2)$ one of our results in Section 1.3 is based on. Section 1.8 lists infinite-dimensional counterparts—available in the literature—to some of the results in this chapter. In Section 1.9, we provide a proof of the set-theoretic null-space property stated in Section 1.6. Finally, Section 1.10 contains results on operator norms used frequently in this chapter.

Notation. For $\mathcal{A} \subseteq \{1, \dots, m\}$, $\mathbf{D}_\mathcal{A}$ denotes the $m \times m$ diagonal matrix with diagonal entries $(\mathbf{D}_\mathcal{A})_{i,i} = 1$ for $i \in \mathcal{A}$, and $(\mathbf{D}_\mathcal{A})_{i,i} = 0$ else. With $\mathbf{U} \in \mathbb{C}^{m\times m}$ unitary and $\mathcal{A} \subseteq \{1, \dots, m\}$, we define the orthogonal projection $\mathbf{\Pi}_\mathcal{A}(\mathbf{U}) = \mathbf{U}\mathbf{D}_\mathcal{A}\mathbf{U}^*$ and set $\mathcal{W}^{\mathbf{U},\mathcal{A}} = \operatorname{range}(\mathbf{\Pi}_\mathcal{A}(\mathbf{U}))$. For $\mathbf{x} \in \mathbb{C}^m$ and $\mathcal{A} \subseteq \{1, \dots, m\}$, we let $\mathbf{x}_\mathcal{A} = \mathbf{D}_\mathcal{A}\mathbf{x}$. With $\mathbf{A} \in \mathbb{C}^{m\times m}$, $|||\mathbf{A}|||_1 = \max_{\mathbf{x}\colon \|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1$ refers to the operator 1-norm, $|||\mathbf{A}|||_2 = \max_{\mathbf{x}\colon \|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$ designates the operator 2-norm, $\|\mathbf{A}\|_2 = \sqrt{\operatorname{tr}(\mathbf{A}\mathbf{A}^*)}$ is the Frobenius norm, and $\|\mathbf{A}\|_1 = \sum_{i,j=1}^m |A_{i,j}|$. The $m\times m$ DFT matrix $\mathbf{F}$ has entry $(1/\sqrt{m})e^{-2\pi jkl/m}$ in its k-th row and l-th column for $k, l \in \{1, \dots, m\}$. For $x \in \mathbb{R}$, we set $[x]_+ = \max(x, 0)$. The vector $\mathbf{x} \in \mathbb{C}^m$ is said to be s-sparse if it has at most s nonzero entries. The open ball in $(\mathbb{C}^m, \|\cdot\|_2)$ of radius ρ centered at $\mathbf{u} \in \mathbb{C}^m$ is denoted by $\mathcal{B}_m(\mathbf{u}, \rho)$ and $V_m(\rho)$ refers to its volume. The indicator function on the set $\mathcal{A}$ is $\chi_\mathcal{A}$. We use the convention $0 \cdot \infty = 0$.

1.3 Uncertainty Relations in $(\mathbb{C}^m, \|\cdot\|_2)$

Donoho and Stark [6] define uncertainty relations as upper bounds on the operator norm of the band-limitation operator followed by the time-limitation operator. We adopt this elegant concept and extend it to refer to an upper bound on the operator norm of a general orthogonal projection operator (replacing the band-limitation operator) followed by the "time-limitation operator" $\mathbf{D}_\mathcal{P}$ as an uncertainty relation. More specifically, let $\mathbf{U} \in \mathbb{C}^{m\times m}$ be a unitary matrix, $\mathcal{P}, \mathcal{Q} \subseteq \{1, \dots, m\}$, and consider the orthogonal projection $\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})$ onto the subspace $\mathcal{W}^{\mathbf{U},\mathcal{Q}}$ which is spanned by $\{\mathbf{u}_i : i \in \mathcal{Q}\}$. Let[1] $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = |||\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})|||_2$. In the setting of [6] $\mathbf{U}$ would correspond to the DFT matrix $\mathbf{F}$ and $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F})$ is the operator 2-norm of the band-limitation operator followed by the time-limitation operator, both in finite dimensions. By Lemma 1.22 we have

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_{\mathcal{P}}\|_2}{\|\mathbf{x}\|_2}. \tag{1.1}$$

An uncertainty relation in $(\mathbb{C}^m, \|\cdot\|_2)$ is an upper bound of the form $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \le c$ with $c \ge 0$, and states that $\|\mathbf{x}_{\mathcal{P}}\|_2 \le c\|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$. $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})$ hence quantifies how well a vector supported on $\mathcal{Q}$ in the basis $\mathbf{U}$ can be concentrated on $\mathcal{P}$. Note that an uncertainty relation in $(\mathbb{C}^m, \|\cdot\|_2)$ is nontrivial only if $c < 1$. Application of Lemma 1.23 now yields

$$\frac{\|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\|_2}{\sqrt{\operatorname{rank}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))}} \le \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \le \|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\|_2, \tag{1.2}$$

where the upper bound constitutes an uncertainty relation and the lower bound will allow us to assess its tightness. Next, note that

$$\|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\|_2 = \sqrt{\operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))} \tag{1.3}$$

and

$$\begin{aligned} \operatorname{rank}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})) &= \operatorname{rank}(\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}\mathbf{U}^*) \qquad (1.4)\\ &\le \min(|\mathcal{P}|, |\mathcal{Q}|), \qquad (1.5)\end{aligned}$$

where (1.5) follows from $\operatorname{rank}(\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}) \le \min(|\mathcal{P}|, |\mathcal{Q}|)$ and [33, Property (c), Chapter 0.4.5]. When used in (1.2) this implies

$$\sqrt{\frac{\operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))}{\min(|\mathcal{P}|, |\mathcal{Q}|)}} \le \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \le \sqrt{\operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))}. \tag{1.6}$$

Particularizing to $\mathbf{U} = \mathbf{F}$, we obtain

$$\begin{aligned} \sqrt{\operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{F}))} &= \sqrt{\operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{F}\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*)} \qquad (1.7)\\ &= \sqrt{\sum_{i\in\mathcal{P}}\sum_{j\in\mathcal{Q}} |\mathbf{F}_{i,j}|^2} \qquad (1.8)\\ &= \sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}}, \qquad (1.9)\end{aligned}$$

so that (1.6) reduces to

$$\sqrt{\frac{\max(|\mathcal{P}|, |\mathcal{Q}|)}{m}} \le \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \le \sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}}. \tag{1.10}$$

[1] We note that, for general unitary $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times m}$, unitary invariance of $\|\cdot\|_2$ yields $|||\mathbf{\Pi}_{\mathcal{P}}(\mathbf{A})\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{B})|||_2 = |||\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})|||_2$ with $\mathbf{U} = \mathbf{A}^*\mathbf{B}$. The situation where both the band-limitation and the time-limitation operator are replaced by general orthogonal projection operators can hence be reduced to the case considered here.

There exist sets $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$ that saturate both bounds in (1.10), e.g., $\mathcal{P} = \{1\}$ and $\mathcal{Q} = \{1, \ldots, m\}$, which yields $\sqrt{\max(|\mathcal{P}|, |\mathcal{Q}|)/m} = \sqrt{|\mathcal{P}||\mathcal{Q}|/m} = 1$ and therefore $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = 1$. An example of sets $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$ saturating only the lower bound in (1.10) is as follows. Take n to divide m and set

$$\mathcal{P} = \left\{ \frac{m}{n}, \frac{2m}{n}, \ldots, \frac{(n-1)m}{n}, m \right\} \tag{1.11}$$

and

$$\mathcal{Q} = \{l+1, \ldots, l+n\} \tag{1.12}$$

with $l \in \{1, \ldots, m\}$ and $\mathcal{Q}$ interpreted circularly in $\{1, \ldots, m\}$. Then, the upper bound in (1.10) is

$$\sqrt{\frac{|\mathcal{P}||\mathcal{Q}|}{m}} = \frac{n}{\sqrt{m}}, \tag{1.13}$$

whereas the lower bound becomes

$$\sqrt{\frac{\max(|\mathcal{P}|, |\mathcal{Q}|)}{m}} = \sqrt{\frac{n}{m}}. \tag{1.14}$$

Thus, for $m \to \infty$ with fixed ratio m/n, the upper bound in (1.10) tends to infinity whereas the corresponding lower bound remains constant. The following result states that the lower bound in (1.10) is tight for $\mathcal{P}$ and $\mathcal{Q}$ as in (1.11) and (1.12), respectively. This implies a lack of tightness of the uncertainty relation $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{|\mathcal{P}||\mathcal{Q}|/m}$ by a factor of $\sqrt{n}$. The large sieve-based uncertainty relation developed in the next section will be seen to remedy this problem.

LEMMA 1.1 *[6, Theorem 11] Let n divide m and consider*

$$\mathcal{P} = \left\{ \frac{m}{n}, \frac{2m}{n}, \ldots, \frac{(n-1)m}{n}, m \right\} \tag{1.15}$$

and

$$\mathcal{Q} = \{l+1, \ldots, l+n\} \tag{1.16}$$

with $l \in \{1, \ldots, m\}$ and $\mathcal{Q}$ interpreted circularly in $\{1, \ldots, m\}$. Then, $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = \sqrt{n/m}$.

Proof We have

$$\begin{aligned}
\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) &= |||\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{F})\mathbf{D}_{\mathcal{P}}|||_2 && (1.17)\\
&= |||\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*\mathbf{D}_{\mathcal{P}}|||_2 && (1.18)\\
&= \max_{\mathbf{x}:\, \|\mathbf{x}\|_2=1} \|\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*\mathbf{D}_{\mathcal{P}}\mathbf{x}\|_2 && (1.19)\\
&= \max_{\mathbf{x}:\, \mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*\mathbf{x}_{\mathcal{P}}\|_2}{\|\mathbf{x}\|_2} && (1.20)\\
&= \max_{\substack{\mathbf{x}:\, \mathbf{x}=\mathbf{x}_{\mathcal{P}}\\ \mathbf{x}\neq\mathbf{0}}} \frac{\|\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*\mathbf{x}\|_2}{\|\mathbf{x}\|_2}, && (1.21)
\end{aligned}$$

where in (1.17) we applied Lemma 1.22 and in (1.18) we used unitary invariance of $\|\cdot\|_2$. Next, consider an arbitrary but fixed $\mathbf{x} \in \mathbb{C}^m$ with $\mathbf{x} = \mathbf{x}_{\mathcal{P}}$ and define $\mathbf{y} \in \mathbb{C}^n$ according to $y_s = x_{ms/n}$ for $s = 1, \ldots, n$. It follows that

$$\|\mathbf{D}_{\mathcal{Q}}\mathbf{F}^*\mathbf{x}\|_2^2 = \frac{1}{m}\sum_{q\in\mathcal{Q}}\Big|\sum_{p\in\mathcal{P}} x_p\, e^{\frac{2\pi jpq}{m}}\Big|^2 \tag{1.22}$$

$$= \frac{1}{m}\sum_{q\in\mathcal{Q}}\Big|\sum_{s=1}^{n} x_{ms/n}\, e^{\frac{2\pi jsq}{n}}\Big|^2 \tag{1.23}$$

$$= \frac{1}{m}\sum_{q\in\mathcal{Q}}\Big|\sum_{s=1}^{n} y_s\, e^{\frac{2\pi jsq}{n}}\Big|^2 \tag{1.24}$$

$$= \frac{n}{m}\|\mathbf{F}^*\mathbf{y}\|_2^2 \tag{1.25}$$

$$= \frac{n}{m}\|\mathbf{y}\|_2^2, \tag{1.26}$$

where $\mathbf{F}$ in (1.25) is the $n\times n$ DFT matrix and in (1.26) we used unitary invariance of $\|\cdot\|_2$. With (1.22)–(1.26) and $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2$ in (1.21), we get $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = \sqrt{n/m}$. □

1.3.1 Uncertainty Relations Based on the Large Sieve

The uncertainty relation in (1.6) is very crude as it simply upper-bounds the operator 2-norm by the Frobenius norm. For $\mathbf{U} = \mathbf{F}$ a more sophisticated upper bound on $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F})$ was reported in [7, Theorem 12]. The proof of this result establishes a remarkable connection to the so-called "large sieve", a family of inequalities involving trigonometric polynomials originally developed in the field of analytic number theory [20, 21]. We next present a slightly improved and generalized version of [7, Theorem 12].

THEOREM 1.2 *Let $\mathcal{P} \subseteq \{1, \ldots, m\}$, $l, n \in \{1, \ldots, m\}$, and*

$$\mathcal{Q} = \{l+1, \ldots, l+n\} \tag{1.27}$$

with $\mathcal{Q}$ interpreted circularly in $\{1, \ldots, m\}$. For $\lambda \in (0, m]$, we define the circular Nyquist density $\rho(\mathcal{P}, \lambda)$ according to

$$\rho(\mathcal{P}, \lambda) = \frac{1}{\lambda}\max_{r\in[0,m)} |\widetilde{\mathcal{P}} \cap (r, r+\lambda)|, \tag{1.28}$$

where $\widetilde{\mathcal{P}} = \mathcal{P} \cup \{m + p : p \in \mathcal{P}\}$. Then,

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{\left(\frac{\lambda(n-1)}{m} + 1\right)\rho(\mathcal{P}, \lambda)} \tag{1.29}$$

for all $\lambda \in (0, m]$.

Proof If $\mathcal{P} = \emptyset$, then $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = 0$ as a consequence of $\mathbf{\Pi}_{\emptyset}(\mathbf{F}) = \mathbf{0}$ and (1.29) holds trivially. Suppose now that $\mathcal{P} \neq \emptyset$, consider an arbitrary but fixed $\mathbf{x} \in$

$\mathcal{W}^{\mathbf{F},\mathcal{Q}}$ with $\|\mathbf{x}\|_2 = 1$, and set $\mathbf{a} = \mathbf{F}^*\mathbf{x}$. Then, $\mathbf{a} = \mathbf{a}_\mathcal{Q}$ and, by unitarity of $\mathbf{F}$, $\|\mathbf{a}\|_2 = 1$. We have

$$
\begin{aligned}
|x_p|^2 &= |(\mathbf{F}\mathbf{a})_p|^2 &(1.30)\\
&= \frac{1}{m}\bigg|\sum_{q\in\mathcal{Q}} a_q e^{-\frac{2\pi jpq}{m}}\bigg|^2 &(1.31)\\
&= \frac{1}{m}\bigg|\sum_{k=1}^{n} a_k e^{-\frac{2\pi jpk}{m}}\bigg|^2 &(1.32)\\
&= \frac{1}{m}\Big|\psi\Big(\frac{p}{m}\Big)\Big|^2 \quad \text{for } p \in \{1,\dots,m\}, &(1.33)
\end{aligned}
$$

where we defined the 1-periodic trigonometric polynomial $\psi(s)$ according to

$$\psi(s) = \sum_{k=1}^{n} a_k e^{-2\pi jks}. \tag{1.34}$$

Next, let ν_t denote the unit Dirac measure centered at $t \in \mathbb{R}$ and set $\mu = \sum_{p\in\mathcal{P}} \nu_{p/m}$ with 1-periodic extension outside $[0, 1)$. Then,

$$
\begin{aligned}
\|\mathbf{x}_\mathcal{P}\|_2^2 &= \frac{1}{m}\sum_{p\in\mathcal{P}}\Big|\psi\Big(\frac{p}{m}\Big)\Big|^2 &(1.35)\\
&= \frac{1}{m}\int_{[0,1)} |\psi(s)|^2 \mathrm{d}\mu(s) &(1.36)\\
&\le \Big(\frac{n-1}{m} + \frac{1}{\lambda}\Big) \sup_{r\in[0,1)} \mu\Big(\Big(r, r+\frac{\lambda}{m}\Big)\Big) &(1.37)
\end{aligned}
$$

for all $\lambda \in (0, m]$, where (1.35) is by (1.30)–(1.33) and in (1.37) we applied the large sieve inequality Lemma 1.20 with $\delta = \lambda/m$ and $\|\mathbf{a}\|_2 = 1$. Now,

$$
\begin{aligned}
&\sup_{r\in[0,1)} \mu\Big(\Big(r, r+\frac{\lambda}{m}\Big)\Big) &(1.38)\\
&= \sup_{r\in[0,m)} \sum_{p\in\mathcal{P}} (\nu_p((r, r+\lambda)) + \nu_{m+p}((r, r+\lambda))) &(1.39)\\
&= \max_{r\in[0,m)} |\widetilde{\mathcal{P}} \cap (r, r+\lambda)| &(1.40)\\
&= \lambda\, \rho(\mathcal{P}, \lambda) \quad \text{for all } \lambda \in (0, m], &(1.41)
\end{aligned}
$$

where in (1.39) we used the 1-periodicity of μ. Using (1.38)–(1.41) in (1.37) yields

$$\|\mathbf{x}_\mathcal{P}\|_2^2 \le \Big(\frac{\lambda(n-1)}{m} + 1\Big)\rho(\mathcal{P}, \lambda) \quad \text{for all } \lambda \in (0, m]. \tag{1.42}$$

As $\mathbf{x} \in \mathcal{W}^{\mathbf{F},\mathcal{Q}}$ with $\|\mathbf{x}\|_2 = 1$ was arbitrary, we conclude that

$$\Delta^2_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{F},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_\mathcal{P}\|_2^2}{\|\mathbf{x}\|_2^2} \tag{1.43}$$

$$\leq \left(\frac{\lambda(n-1)}{m} + 1\right)\rho(\mathcal{P},\lambda) \quad \text{for all } \lambda \in (0, m], \tag{1.44}$$

thereby finishing the proof. □

Theorem 1.2 slightly improves upon [7, Theorem 12] by virtue of applying to more general sets $\mathcal{Q}$ and defining the circular Nyquist density in (1.28) in terms of open intervals $(r, r+\lambda)$.

We next apply Theorem 1.2 to specific choices of $\mathcal{P}$ and $\mathcal{Q}$. First, consider $\mathcal{P} = \{1\}$ and $\mathcal{Q} = \{1, \dots, m\}$, which were shown to saturate the upper and the lower bound in (1.10) leading to $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) = 1$. Since $\mathcal{P}$ consists of a single point, $\rho(\mathcal{P},\lambda) = 1/\lambda$ for all $\lambda \in (0, m]$. Thus, Theorem 1.2 with $n = m$ yields

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{\frac{m-1}{m} + \frac{1}{\lambda}} \quad \text{for all } \lambda \in (0, m]. \tag{1.45}$$

Setting $\lambda = m$ in (1.45) yields $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq 1$.

Next, consider $\mathcal{P}$ and $\mathcal{Q}$ as in (1.11) and (1.12), respectively, which, as already mentioned, have the uncertainty relation in (1.10) lacking tightness by a factor of $\sqrt{n}$. Since $\mathcal{P}$ consists of points spaced m/n apart, we get $\rho(\mathcal{P},\lambda) = 1/\lambda$ for all $\lambda \in (0, m/n]$. The upper bound (1.29) now becomes

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{\frac{n-1}{m} + \frac{1}{\lambda}} \quad \text{for all } \lambda \in \left(0, \frac{m}{n}\right]. \tag{1.46}$$

Setting $\lambda = m/n$ in (1.46) yields

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \sqrt{(2n-1)/m} \leq \sqrt{2}\sqrt{n/m}, \tag{1.47}$$

which is tight up to a factor of $\sqrt{2}$ (cf. Lemma 1.1). We hasten to add, however, that the large sieve technique applies to $\mathbf{U} = \mathbf{F}$ only.

1.3.2 Coherence-based Uncertainty Relation

We next present an uncertainty relation that is of simple form and applies to general unitary $\mathbf{U}$. To this end, we first introduce the concept of coherence of a matrix.

DEFINITION 1.3 For $\mathbf{A} = (\mathbf{a}_1 \dots \mathbf{a}_n) \in \mathbb{C}^{m\times n}$ with columns $\|\cdot\|_2$-normalized to 1, the coherence is defined as $\mu(\mathbf{A}) = \max_{i\neq j} |\mathbf{a}_i^* \mathbf{a}_j|$.

We have the following coherence-based uncertainty relation valid for general unitary $\mathbf{U}$.

LEMMA 1.4 *Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ be unitary and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \dots, m\}$. Then,*

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \sqrt{|\mathcal{P}||\mathcal{Q}|}\,\mu([\mathbf{I}\ \ \mathbf{U}]). \tag{1.48}$$

Proof The claim follows from

$$\Delta^2_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \le \operatorname{tr}(\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}\mathbf{U}^*) \tag{1.49}$$

$$= \sum_{k\in\mathcal{P}}\sum_{l\in\mathcal{Q}} |\mathbf{U}_{k,l}|^2 \tag{1.50}$$

$$\le |\mathcal{P}||\mathcal{Q}| \max_{k,l} |\mathbf{U}_{k,l}|^2 \tag{1.51}$$

$$= |\mathcal{P}||\mathcal{Q}|\, \mu^2([\mathbf{I}\ \ \mathbf{U}]), \tag{1.52}$$

where (1.49) is by (1.6) and in (1.52) we used the definition of coherence. □

Since $\mu([\mathbf{I}\ \ \mathbf{F}]) = 1/\sqrt{m}$, Lemma 1.4 particularized to $\mathbf{U} = \mathbf{F}$ recovers the upper bound in (1.10).

1.3.3 Concentration Inequalities

As mentioned at the beginning of this chapter, the classical uncertainty relation in signal analysis quantifies how well concentrated a signal can be in time and frequency. In the finite-dimensional setting considered here this amounts to characterizing the concentration of $\mathbf{p}$ and $\mathbf{q}$ in $\mathbf{p} = \mathbf{F}\mathbf{q}$. We will actually study the more general case obtained by replacing $\mathbf{I}$ and $\mathbf{F}$ by unitary $\mathbf{A} \in \mathbb{C}^{m\times m}$ and $\mathbf{B} \in \mathbb{C}^{m\times m}$, respectively, and will ask ourselves how well concentrated $\mathbf{p}$ and $\mathbf{q}$ in $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ can be. Rewriting $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ according to $\mathbf{p} = \mathbf{U}\mathbf{q}$ with $\mathbf{U} = \mathbf{A}^*\mathbf{B}$, we now show how the uncertainty relation in Lemma 1.4 can be used to answer this question. Let us start by introducing a measure for concentration in $(\mathbb{C}^m, \|\cdot\|_2)$.

DEFINITION 1.5 Let $\mathcal{P} \subseteq \{1,\dots,m\}$ and $\varepsilon_{\mathcal{P}} \in [0,1]$. The vector $\mathbf{x} \in \mathbb{C}^m$ is said to be $\varepsilon_{\mathcal{P}}$-concentrated if $\|\mathbf{x} - \mathbf{x}_{\mathcal{P}}\|_2 \le \varepsilon_{\mathcal{P}}\|\mathbf{x}\|_2$.

The fraction of 2-norm an $\varepsilon_{\mathcal{P}}$-concentrated vector exhibits outside $\mathcal{P}$ is therefore no more than $\varepsilon_{\mathcal{P}}$. In particular, if $\mathbf{x}$ is $\varepsilon_{\mathcal{P}}$-concentrated with $\varepsilon_{\mathcal{P}} = 0$, then $\mathbf{x} = \mathbf{x}_{\mathcal{P}}$ and $\mathbf{x}$ is $|\mathcal{P}|$-sparse. The zero vector is trivially $\varepsilon_{\mathcal{P}}$-concentrated for all $\mathcal{P} \subseteq \{1,\dots,m\}$ and $\varepsilon_{\mathcal{P}} \in [0,1]$.

We next derive a lower bound on $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})$ for unitary matrices $\mathbf{U}$ that relate $\varepsilon_{\mathcal{P}}$-concentrated vectors $\mathbf{p}$ to $\varepsilon_{\mathcal{Q}}$-concentrated vectors $\mathbf{q}$ through $\mathbf{p} = \mathbf{U}\mathbf{q}$. The formal statement is as follows.

LEMMA 1.6 *Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ be unitary and $\mathcal{P},\mathcal{Q} \subseteq \{1,\dots,m\}$. Suppose that there exist a nonzero $\varepsilon_{\mathcal{P}}$-concentrated $\mathbf{p} \in \mathbb{C}^m$ and a nonzero $\varepsilon_{\mathcal{Q}}$-concentrated $\mathbf{q} \in \mathbb{C}^m$ such that $\mathbf{p} = \mathbf{U}\mathbf{q}$. Then,*

$$\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \ge [1 - \varepsilon_{\mathcal{P}} - \varepsilon_{\mathcal{Q}}]_+. \tag{1.53}$$

Proof We have

$$\begin{aligned}
\|\mathbf{p}-\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}_{\mathcal{P}}\|_2 &\leq \|\mathbf{p}-\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}\|_2+\|\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}_{\mathcal{P}}-\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}\|_2 && (1.54)\\
&\leq \|\mathbf{p}-\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}\|_2+|||\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})|||_2\|\mathbf{p}_{\mathcal{P}}-\mathbf{p}\|_2 && (1.55)\\
&\leq \|\mathbf{p}-\mathbf{U}\mathbf{D}_{\mathcal{Q}}\mathbf{U}^*\mathbf{p}\|_2+\|\mathbf{p}_{\mathcal{P}}-\mathbf{p}\|_2 && (1.56)\\
&= \|\mathbf{q}-\mathbf{q}_{\mathcal{Q}}\|_2+\|\mathbf{p}_{\mathcal{P}}-\mathbf{p}\|_2, && (1.57)\\
&\leq \varepsilon_{\mathcal{Q}}\|\mathbf{q}\|_2+\varepsilon_{\mathcal{P}}\|\mathbf{p}\|_2 && (1.58)\\
&= (\varepsilon_{\mathcal{P}}+\varepsilon_{\mathcal{Q}})\|\mathbf{p}\|_2, && (1.59)
\end{aligned}$$

where in (1.57) we made use of the unitary invariance of $\|\cdot\|_2$. It follows that

$$\begin{aligned}
\|\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}_{\mathcal{P}}\|_2 &\geq [\|\mathbf{p}\|_2-\|\mathbf{p}-\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p}_{\mathcal{P}}\|_2]_+ && (1.60)\\
&\geq \|\mathbf{p}\|_2[1-\varepsilon_{\mathcal{P}}-\varepsilon_{\mathcal{Q}}]_+, && (1.61)
\end{aligned}$$

where (1.60) is by the reverse triangle inequality and in (1.61) we used (1.54)–(1.59). Since $\mathbf{p}\neq\mathbf{0}$ by assumption, (1.60)–(1.61) implies

$$\left\|\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{D}_{\mathcal{P}}\frac{\mathbf{p}}{\|\mathbf{p}\|_2}\right\|_2 \geq [1-\varepsilon_{\mathcal{P}}-\varepsilon_{\mathcal{Q}}]_+, \tag{1.62}$$

which in turn yields $|||\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{D}_{\mathcal{P}}|||_2 \geq [1-\varepsilon_{\mathcal{P}}-\varepsilon_{\mathcal{Q}}]_+$. This concludes the proof as $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = |||\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{D}_{\mathcal{P}}|||_2$ by Lemma 1.22. □

Combining Lemma 1.6 with the uncertainty relation Lemma 1.4 yields the announced result stating that a nonzero vector can not be arbitrarily well concentrated with respect to two different orthonormal bases.

COROLLARY 1 *Let $\mathbf{A},\mathbf{B}\in\mathbb{C}^{m\times m}$ be unitary and $\mathcal{P},\mathcal{Q}\subseteq\{1,\dots,m\}$. Suppose that there exist a nonzero $\varepsilon_{\mathcal{P}}$-concentrated $\mathbf{p}\in\mathbb{C}^m$ and a nonzero $\varepsilon_{\mathcal{Q}}$-concentrated $\mathbf{q}\in\mathbb{C}^m$ such that $\mathbf{A}\mathbf{p}=\mathbf{B}\mathbf{q}$. Then,*

$$|\mathcal{P}||\mathcal{Q}| \geq \frac{[1-\varepsilon_{\mathcal{P}}-\varepsilon_{\mathcal{Q}}]_+^2}{\mu^2([\mathbf{A}\ \ \mathbf{B}])}. \tag{1.63}$$

Proof Let $\mathbf{U}=\mathbf{A}^*\mathbf{B}$. Then, by Lemmata 1.4 and 1.6, we have

$$[1-\varepsilon_{\mathcal{P}}-\varepsilon_{\mathcal{Q}}]_+ \leq \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \sqrt{|\mathcal{P}||\mathcal{Q}|}\,\mu([\mathbf{I}\ \ \mathbf{U}]). \tag{1.64}$$

The claim now follows by noting that $\mu([\mathbf{I}\ \ \mathbf{U}])=\mu([\mathbf{A}\ \ \mathbf{B}])$. □

For $\varepsilon_{\mathcal{P}}=\varepsilon_{\mathcal{Q}}=0$, we recover the well-known Elad-Bruckstein result.

COROLLARY 2 *[8, Theorem 1] Let $\mathbf{A},\mathbf{B}\in\mathbb{C}^{m\times m}$ be unitary. If $\mathbf{A}\mathbf{p}=\mathbf{B}\mathbf{q}$ for nonzero $\mathbf{p},\mathbf{q}\in\mathbb{C}^m$, then $\|\mathbf{p}\|_0\|\mathbf{q}\|_0\geq 1/\mu^2([\mathbf{A}\ \ \mathbf{B}])$.*

1.3.4 Noisy Recovery in $(\mathbb{C}^m,\|\cdot\|_2)$

Uncertainty relations are typically employed to prove that something is not possible. For example, by Corollary 1 there is a limit on how well a nonzero vector can be concentrated with respect to two different orthonormal bases. Donoho and Stark [6] noticed that uncertainty relations can also be used to show that something unexpected is possible. Specifically, [6, Section 4] considers a noisy signal recovery problem, which we now translate to the finite-dimensional setting. Let $\mathbf{p}, \mathbf{n} \in \mathbb{C}^m$ and $\mathcal{P} \subseteq \{1, \ldots, m\}$, set $\mathcal{P}^c = \{1, \ldots, m\} \backslash \mathcal{P}$, and suppose that we observe $\mathbf{y} = \mathbf{p}_{\mathcal{P}^c} + \mathbf{n}$. Note that the information contained in $\mathbf{p}_{\mathcal{P}}$ is completely lost in the observation. Without structural assumptions on $\mathbf{p}$, it is therefore not possible to recover information on $\mathbf{p}_{\mathcal{P}}$ from $\mathbf{y}$. However, if $\mathbf{p}$ is sufficiently sparse with respect to an orthonormal basis and $|\mathcal{P}|$ is sufficiently small, it turns out that all entries of $\mathbf{p}$ can be recovered in a linear fashion to within a precision determined by the noise level. This is often referred to in the literature as stable recovery [6]. The corresponding formal statement is as follows.

LEMMA 1.7 *Let* $\mathbf{U} \in \mathbb{C}^{m \times m}$ *be unitary,* $\mathcal{Q} \subseteq \{1, \ldots, m\}$, $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$, *and consider*

$$\mathbf{y} = \mathbf{p}_{\mathcal{P}^c} + \mathbf{n}, \tag{1.65}$$

where $\mathbf{n} \in \mathbb{C}^m$ *and* $\mathcal{P}^c = \{1, \ldots, m\} \backslash \mathcal{P}$ *with* $\mathcal{P} \subseteq \{1, \ldots, m\}$. *If* $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1$, *then there exists a matrix* $\mathbf{L} \in \mathbb{C}^{m \times m}$ *such that*

$$\|\mathbf{L}\mathbf{y} - \mathbf{p}\|_2 \leq C \|\mathbf{n}_{\mathcal{P}^c}\|_2 \tag{1.66}$$

with $C = 1/(1 - \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}))$. *In particular,*

$$|\mathcal{P}||\mathcal{Q}| < \frac{1}{\mu^2([\mathbf{I} \;\; \mathbf{U}])} \tag{1.67}$$

is sufficient for $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1$.

Proof For $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1$, it follows that (cf. [33, p. 301]) $(\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))$ is invertible with

$$\begin{aligned} |||(\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}|||_2 &\leq \frac{1}{1 - |||\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})|||_2} && (1.68) \\ &= \frac{1}{1 - \Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})}. && (1.69) \end{aligned}$$

We now set $\mathbf{L} = (\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}\mathbf{D}_{\mathcal{P}^c}$ and note that

$$\begin{aligned} \mathbf{L}\mathbf{p}_{\mathcal{P}^c} &= (\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}\mathbf{p}_{\mathcal{P}^c} && (1.70) \\ &= (\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}(\mathbf{I} - \mathbf{D}_{\mathcal{P}})\mathbf{p} && (1.71) \\ &= (\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}(\mathbf{I} - \mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))\mathbf{p} && (1.72) \\ &= \mathbf{p}, && (1.73) \end{aligned}$$

where in (1.72) we used $\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\mathbf{p} = \mathbf{p}$, which is by assumption. Next, we upper-bound $\|\mathbf{Ly}-\mathbf{p}\|_2$ according to

$$\|\mathbf{Ly}-\mathbf{p}\|_2 = \|\mathbf{L}\mathbf{p}_{\mathcal{P}^c}+\mathbf{Ln}-\mathbf{p}\|_2 \tag{1.74}$$

$$= \|\mathbf{Ln}\|_2 \tag{1.75}$$

$$\leq |||(\mathbf{I}-\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U}))^{-1}|||_2 \|\mathbf{n}_{\mathcal{P}^c}\|_2 \tag{1.76}$$

$$\leq \frac{1}{1-\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})}\|\mathbf{n}_{\mathcal{P}^c}\|_2, \tag{1.77}$$

where in (1.75) we used (1.70)–(1.73). Finally, Lemma 1.4 implies that (1.67) is sufficient for $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{U})<1$. □

We next particularize Lemma 1.7 for $\mathbf{U}=\mathbf{F}$,

$$\mathcal{P} = \left\{\frac{m}{n}, \frac{2m}{n}, \dots, \frac{(n-1)m}{n}, m\right\} \tag{1.78}$$

with n dividing m, and

$$\mathcal{Q} = \{l+1, \dots, l+n\} \tag{1.79}$$

with $l \in \{1,\dots,m\}$ and $\mathcal{Q}$ interpreted circularly in $\{1,\dots,m\}$. This means that $\mathbf{p}$ is n-sparse in $\mathbf{F}$ and we are missing n entries in the noisy observation $\mathbf{y}$. From Lemma 1.1 we know that $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F})=\sqrt{n/m}$. Since n divides m by assumption, stable recovery of $\mathbf{p}$ is possible for $n \leq m/2$. In contrast, the coherence-based uncertainty relation in Lemma 1.4 yields $\Delta_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq \frac{n}{\sqrt{m}}$, and would hence suggest that $n^2<m$ is needed for stable recovery.

1.4 Uncertainty Relations in $(\mathbb{C}^m, \|\cdot\|_1)$

We introduce uncertainty relations in $(\mathbb{C}^m, \|\cdot\|_1)$ following the same story line as in Section 1.3. Specifically, let $\mathbf{U}=(\mathbf{u}_1 \dots \mathbf{u}_m) \in \mathbb{C}^{m\times m}$ be a unitary matrix, $\mathcal{P},\mathcal{Q} \subseteq \{1,\dots,m\}$, and consider the orthogonal projection $\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})$ onto the subspace $\mathcal{W}^{\mathbf{U},\mathcal{Q}}$, which is spanned by $\{\mathbf{u}_i : i \in \mathcal{Q}\}$. Let[2] $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = |||\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})|||_1$. By Lemma 1.22 we have

$$\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_{\mathcal{P}}\|_1}{\|\mathbf{x}\|_1}. \tag{1.80}$$

An uncertainty relation in $(\mathbb{C}^m, \|\cdot\|_1)$ is an upper bound of the form $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq c$ with $c \geq 0$ and states that $\|\mathbf{x}_{\mathcal{P}}\|_1 \leq c\|\mathbf{x}\|_1$ for all $\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}$. $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U})$ hence quantifies how well a vector supported on $\mathcal{Q}$ in the basis $\mathbf{U}$ can be concentrated

[2] In contrast to the operator 2-norm, the operator 1-norm is not invariant under unitary transformations so that we do not have $|||\mathbf{\Pi}_{\mathcal{P}}(\mathbf{A})\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{B})|||_1 \neq |||\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{A}^*\mathbf{B})|||_1$ for general unitary $\mathbf{A},\mathbf{B}$. This, however, does not constitute a problem as whenever we apply uncertainty relations in $(\mathbb{C}^m, \|\cdot\|_1)$, the case of general unitary $\mathbf{A},\mathbf{B}$ can always be reduced directly to $\mathbf{\Pi}_{\mathcal{P}}(\mathbf{I})=\mathbf{D}_{\mathcal{P}}$ and $\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{A}^*\mathbf{B})$, simply by rewriting $\mathbf{Ap}=\mathbf{Bq}$ according to $\mathbf{p}=\mathbf{A}^*\mathbf{Bq}$.

on $\mathcal{P}$, where now concentration is measured in terms of 1-norm. Again, an uncertainty relation in $(\mathbb{C}^m, \|\cdot\|_1)$ is nontrivial only if $c < 1$. Application of Lemma 1.24 yields

$$\frac{1}{m}\|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\|_1 \leq \Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq \|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{U})\|_1, \tag{1.81}$$

which constitutes the 1-norm equivalent of (1.2).

1.4.1 Coherence-based Uncertainty Relation

We next derive a coherence-based uncertainty relation for $(\mathbb{C}^m, \|\cdot\|_1)$, which comes with the same advantages and disadvantages as its 2-norm counterpart.

LEMMA 1.8 *Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ be a unitary matrix and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Then,*

$$\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) \leq |\mathcal{P}||\mathcal{Q}|\mu^2([\mathbf{I}\;\;\mathbf{U}]). \tag{1.82}$$

Proof Let $\tilde{\mathbf{u}}_i$ denote the column vectors of $\mathbf{U}^*$. It follows from Lemma 1.24 that

$$\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) = \max_{j\in\{1,\ldots,m\}} \|\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}\tilde{\mathbf{u}}_j\|_1. \tag{1.83}$$

With

$$\begin{aligned}
\max_{j\in\{1,\ldots,m\}} \|\mathbf{D}_{\mathcal{P}}\mathbf{U}\mathbf{D}_{\mathcal{Q}}\tilde{\mathbf{u}}_j\|_1 &\leq |\mathcal{P}| \max_{i,j\in\{1,\ldots,m\}} |\tilde{\mathbf{u}}_i^*\mathbf{D}_{\mathcal{Q}}\tilde{\mathbf{u}}_j| && (1.84)\\
&\leq |\mathcal{P}||\mathcal{Q}| \max_{i,j,k\in\{1,\ldots,m\}} |\mathbf{U}_{i,k}||\mathbf{U}_{j,k}| && (1.85)\\
&\leq |\mathcal{P}||\mathcal{Q}|\mu^2([\mathbf{I}\;\;\mathbf{U}]), && (1.86)
\end{aligned}$$

this establishes the proof. □

For $\mathcal{P} = \{1\}$, $\mathcal{Q} = \{1, \ldots, m\}$, and $\mathbf{U} = \mathbf{F}$, the upper bounds on $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{F})$ in (1.81) and (1.82) coincide and equal 1. We next present an example where (1.82) is sharper than (1.81). Let m be even, $\mathcal{P} = \{m\}$, $\mathcal{Q} = \{1, \ldots, m/2\}$, and $\mathbf{U} = \mathbf{F}$. Then, (1.82) becomes $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq 1/2$, whereas

$$\begin{aligned}
\|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{F})\|_1 &= \frac{1}{m}\sum_{l=1}^{m}\left|\sum_{k=1}^{m/2} e^{\frac{2\pi j l k}{m}}\right| && (1.87)\\
&= \frac{1}{2} + \frac{1}{m}\sum_{l=1}^{m-1}\left|\frac{1-e^{\pi j l}}{1-e^{\frac{2\pi j l}{m}}}\right| && (1.88)\\
&= \frac{1}{2} + \frac{2}{m}\sum_{l=1}^{m/2}\frac{1}{\left|1-e^{\frac{2\pi j(2l-1)}{m}}\right|} && (1.89)\\
&= \frac{1}{2} + \frac{1}{m}\sum_{l=1}^{m/2}\frac{1}{\sin\!\left(\frac{\pi(2l-1)}{m}\right)}. && (1.90)
\end{aligned}$$

Applying Jensen's inequality [34, Theorem 2.6.2] to (1.90) and using $\sum_{l=1}^{\frac{m}{2}}(2l-1) = (m/2)^2$ then yields $\|\mathbf{D}_{\mathcal{P}}\mathbf{\Pi}_{\mathcal{Q}}(\mathbf{F})\|_1 \geq 1$, which shows that (1.81) is trivial.

For $\mathcal{P}$ and $\mathcal{Q}$ as in (1.11) and (1.12), respectively, (1.82) becomes $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{F}) \leq n^2/m$, which for fixed ratio n/m increases linearly in m and becomes trivial for $m \geq (m/n)^2$. A more sophisticated uncertainty relation based on a large sieve inequality exists for strictly band-limited (infinite) ℓ_1-sequences [7, Theorem 14]; a corresponding finite-dimensional result does not seem to be available.

1.4.2 Concentration Inequalities

Analogously to Section 1.3.3, we next ask how well concentrated a given signal vector can be in two different orthonormal bases. Here we, however, consider a different measure of concentration accounting for the fact that we deal with the 1-norm.

DEFINITION 1.9 Let $\mathcal{P} \subseteq \{1, \ldots, m\}$ and $\varepsilon_{\mathcal{P}} \in [0, 1]$. The vector $\mathbf{x} \in \mathbb{C}^m$ is said to be $\varepsilon_{\mathcal{P}}$-concentrated if $\|\mathbf{x} - \mathbf{x}_{\mathcal{P}}\|_1 \leq \varepsilon_{\mathcal{P}}\|\mathbf{x}\|_1$.

The fraction of 1-norm an $\varepsilon_{\mathcal{P}}$-concentrated vector exhibits outside $\mathcal{P}$ is therefore no more than $\varepsilon_{\mathcal{P}}$. In particular, if $\mathbf{x}$ is $\varepsilon_{\mathcal{P}}$-concentrated for $\varepsilon_{\mathcal{P}} = 0$, then $\mathbf{x} = \mathbf{x}_{\mathcal{P}}$ and $\mathbf{x}$ is $|\mathcal{P}|$-sparse. The zero vector is trivially $\varepsilon_{\mathcal{P}}$-concentrated for all $\mathcal{P} \subseteq \{1, \ldots, m\}$ and $\varepsilon_{\mathcal{P}} \in [0, 1]$. In the remainder of Section 1.4, concentration is with respect to the 1-norm according to Definition 1.9.

We are now ready to state the announced result on the concentration of a vector in two different orthonormal bases.

LEMMA 1.10 *Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times m}$ be unitary and $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$. Suppose that there exist a nonzero $\varepsilon_{\mathcal{P}}$-concentrated $\mathbf{p} \in \mathbb{C}^m$ and a nonzero $\mathbf{q} \in \mathbb{C}^m$ with $\mathbf{q} = \mathbf{q}_{\mathcal{Q}}$ such that $\mathbf{Ap} = \mathbf{Bq}$. Then,*

$$|\mathcal{P}||\mathcal{Q}| \geq \frac{1 - \varepsilon_{\mathcal{P}}}{\mu^2([\mathbf{A} \;\; \mathbf{B}])}. \tag{1.91}$$

Proof Rewriting $\mathbf{Ap} = \mathbf{Bq}$ according to $\mathbf{p} = \mathbf{A}^*\mathbf{Bq}$, it follows that $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$ with $\mathbf{U} = \mathbf{A}^*\mathbf{B}$. We have

$$\begin{aligned} 1 - \varepsilon_{\mathcal{P}} &\leq \frac{\|\mathbf{p}_{\mathcal{P}}\|_1}{\|\mathbf{p}\|_1} && (1.92)\\ &\leq \Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) && (1.93)\\ &\leq |\mathcal{P}||\mathcal{Q}|\mu^2([\mathbf{I} \;\; \mathbf{U}]), && (1.94) \end{aligned}$$

where (1.92) is by $\varepsilon_{\mathcal{P}}$-concentration of $\mathbf{p}$, (1.93) follows from (1.80) and $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$, and in (1.94) we applied Lemma 1.8. The proof is concluded by noting that $\mu([\mathbf{I} \;\; \mathbf{U}]) = \mu([\mathbf{A} \;\; \mathbf{B}])$.

□

For $\varepsilon_{\mathcal{P}} = 0$, Lemma 1.10 recovers Corollary 2.

1.4.3 Noisy Recovery in $(\mathbb{C}^m, \|\cdot\|_1)$

We next consider a noisy signal recovery problem akin to that in Section 1.3.4. Specifically, we investigate recovery—through 1-norm minimization—of a sparse signal corrupted by $\varepsilon_\mathcal{P}$-concentrated noise.

LEMMA 1.11 *Let*

$$\mathbf{y} = \mathbf{p} + \mathbf{n}, \tag{1.95}$$

where $\mathbf{n} \in \mathbb{C}^m$ *is* $\varepsilon_\mathcal{P}$*-concentrated to* $\mathcal{P} \subseteq \{1, \ldots, m\}$ *and* $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$ *for* $\mathbf{U} \in \mathbb{C}^{m\times m}$ *unitary and* $\mathcal{Q} \subseteq \{1, \ldots, m\}$*. Denote*

$$\mathbf{z} = \underset{\mathbf{w}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}}{\operatorname{argmin}}\,(\|\mathbf{y} - \mathbf{w}\|_1). \tag{1.96}$$

If $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1/2$*, then* $\|\mathbf{z} - \mathbf{p}\|_1 \le C\varepsilon_\mathcal{P}\|\mathbf{n}\|_1$ *with* $C = 2/(1 - 2\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}))$*. In particular,*

$$|\mathcal{P}||\mathcal{Q}| < \frac{1}{2\mu^2([\mathbf{I}\ \ \mathbf{U}])} \tag{1.97}$$

is sufficient for $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1/2$*.*

Proof Set $\mathcal{P}^c = \{1, \ldots, m\}\setminus\mathcal{P}$ and let $\mathbf{q} = \mathbf{U}^*\mathbf{p}$. Note that $\mathbf{q}_\mathcal{Q} = \mathbf{q}$ as a consequence of $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$, which is by assumption. We have

$$\begin{aligned}
\|\mathbf{n}\|_1 &= \|\mathbf{y} - \mathbf{p}\|_1 &(1.98)\\
&\ge \|\mathbf{y} - \mathbf{z}\|_1 &(1.99)\\
&= \|\mathbf{n} - \tilde{\mathbf{z}}\|_1 &(1.100)\\
&= \|(\mathbf{n} - \tilde{\mathbf{z}})_\mathcal{P}\|_1 + \|(\mathbf{n} - \tilde{\mathbf{z}})_{\mathcal{P}^c}\|_1 &(1.101)\\
&\ge \|\mathbf{n}_\mathcal{P}\|_1 - \|\mathbf{n}_{\mathcal{P}^c}\|_1 + \|\tilde{\mathbf{z}}_{\mathcal{P}^c}\|_1 - \|\tilde{\mathbf{z}}_\mathcal{P}\|_1 &(1.102)\\
&= \|\mathbf{n}\|_1 - 2\|\mathbf{n}_{\mathcal{P}^c}\|_1 + \|\tilde{\mathbf{z}}\|_1 - 2\|\tilde{\mathbf{z}}_\mathcal{P}\|_1 &(1.103)\\
&\ge \|\mathbf{n}\|_1(1 - 2\varepsilon_\mathcal{P}) + \|\tilde{\mathbf{z}}\|_1(1 - 2\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U})), &(1.104)
\end{aligned}$$

where in (1.100) we set $\tilde{\mathbf{z}} = \mathbf{z} - \mathbf{p}$, in (1.102) we applied the reverse triangle inequality, and in (1.104) we used that $\mathbf{n}$ is $\varepsilon_\mathcal{P}$-concentrated and $\tilde{\mathbf{z}} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$, owing to $\mathbf{z} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$ and $\mathbf{p} \in \mathcal{W}^{\mathbf{U},\mathcal{Q}}$, together with (1.80). This yields

$$\begin{aligned}
\|\mathbf{z} - \mathbf{p}\|_1 &= \|\tilde{\mathbf{z}}\|_1 &(1.105)\\
&\le \frac{2\varepsilon_\mathcal{P}}{1 - 2\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U})}\|\mathbf{n}\|_1. &(1.106)
\end{aligned}$$

Finally, (1.97) implies $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1/2$ thanks to (1.82). □

Note that for $\varepsilon_\mathcal{P} = 0$, i.e., the noise vector is supported on $\mathcal{P}$, we can recover $\mathbf{p}$ from $\mathbf{y} = \mathbf{p} + \mathbf{n}$ perfectly provided that $\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{U}) < 1/2$. For the special case $\mathbf{U} = \mathbf{F}$, this is guaranteed by

$$|\mathcal{P}||\mathcal{Q}| < \frac{m}{2}, \tag{1.107}$$

and perfect recovery of $\mathbf{p}$ from $\mathbf{y} = \mathbf{p} + \mathbf{n}$ amounts to the finite-dimensional version of what is known as Logan's phenomenon [6, Section 6.2].

1.4.4 Coherence-based Uncertainty Relation for Pairs of General Matrices

In practice, one is often interested in sparse signal representations with respect to general (i.e., possibly redundant or incomplete) dictionaries. The purpose of this section is to provide a corresponding general uncertainty relation. Specifically, we consider representations of a given signal vector $\mathbf{s}$ according to $\mathbf{s} = \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$, where $\mathbf{A} \in \mathbb{C}^{m\times p}$ and $\mathbf{B} \in \mathbb{C}^{m\times q}$ are general matrices, $\mathbf{p} \in \mathbb{C}^p$, and $\mathbf{q} \in \mathbb{C}^q$. We start by introducing the notion of mutual coherence for pairs of matrices.

DEFINITION 1.12 For $\mathbf{A} = (\mathbf{a}_1 \dots \mathbf{a}_p) \in \mathbb{C}^{m\times p}$ and $\mathbf{B} = (\mathbf{b}_1 \dots \mathbf{b}_q) \in \mathbb{C}^{m\times q}$, both with columns $\|\cdot\|_2$-normalized to 1, the mutual coherence $\bar{\mu}(\mathbf{A}, \mathbf{B})$ is defined as $\bar{\mu}(\mathbf{A}, \mathbf{B}) = \max_{i,j} |\mathbf{a}_i^* \mathbf{b}_j|$.

The general uncertainty relation we are now ready to state is in terms of a pair of upper bounds on $\|\mathbf{p}_\mathcal{P}\|_1$ and $\|\mathbf{q}_\mathcal{Q}\|_1$ for $\mathcal{P} \subseteq \{1, \dots, p\}$ and $\mathcal{Q} \subseteq \{1, \dots, q\}$.

THEOREM 1.13 *Let $\mathbf{A} \in \mathbb{C}^{m\times p}$ and $\mathbf{B} \in \mathbb{C}^{m\times q}$, both with column vectors $\|\cdot\|_2$-normalized to 1, and consider $\mathbf{p} \in \mathbb{C}^p$ and $\mathbf{q} \in \mathbb{C}^q$. Suppose that $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$. Then, we have*

$$\|\mathbf{p}_\mathcal{P}\|_1 \le |\mathcal{P}| \left(\frac{\mu(\mathbf{A})\|\mathbf{p}\|_1 + \bar{\mu}(\mathbf{A}, \mathbf{B})\|\mathbf{q}\|_1}{1 + \mu(\mathbf{A})} \right) \tag{1.108}$$

for all $\mathcal{P} \subseteq \{1, \dots, p\}$ and, by symmetry,

$$\|\mathbf{q}_\mathcal{Q}\|_1 \le |\mathcal{Q}| \left(\frac{\mu(\mathbf{B})\|\mathbf{q}\|_1 + \bar{\mu}(\mathbf{A}, \mathbf{B})\|\mathbf{p}\|_1}{1 + \mu(\mathbf{B})} \right) \tag{1.109}$$

for all $\mathcal{Q} \subseteq \{1, \dots, q\}$.

Proof Since (1.109) follows from (1.108) simply by replacing $\mathbf{A}$ by $\mathbf{B}$, $\mathbf{p}$ by $\mathbf{q}$, $\mathcal{P}$ by $\mathcal{Q}$, and noting that $\bar{\mu}(\mathbf{A}, \mathbf{B}) = \bar{\mu}(\mathbf{B}, \mathbf{A})$, it suffices to prove (1.108). Let $\mathcal{P} \subseteq \{1, \dots, p\}$ and consider an arbitrary but fixed $i \in \{1, \dots, p\}$. Multiplying $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ from the left by $\mathbf{a}_i^*$ and taking absolute values results in

$$|\mathbf{a}_i^* \mathbf{A}\mathbf{p}| = |\mathbf{a}_i^* \mathbf{B}\mathbf{q}|. \tag{1.110}$$

The left-hand side of (1.110) can be lower-bounded according to

$$|\mathbf{a}_i^* \mathbf{A}\mathbf{p}| = \Big| p_i + \sum_{\substack{k=1\\k\neq i}}^{p} \mathbf{a}_i^* \mathbf{a}_k p_k \Big| \tag{1.111}$$

$$\geq |p_i| - \Big| \sum_{\substack{k=1\\k\neq i}}^{p} \mathbf{a}_i^* \mathbf{a}_k p_k \Big| \tag{1.112}$$

$$\geq |p_i| - \sum_{\substack{k=1\\k\neq i}}^{p} |\mathbf{a}_i^* \mathbf{a}_k| |p_k| \tag{1.113}$$

$$\geq |p_i| - \mu(\mathbf{A}) \sum_{\substack{k=1\\k\neq i}}^{p} |p_k| \tag{1.114}$$

$$= (1 + \mu(\mathbf{A}))|p_i| - \mu(\mathbf{A})\|\mathbf{p}\|_1, \tag{1.115}$$

where (1.112) is by the reverse triangle inequality and in (1.114) we used Definition 1.3. Next, we upper-bound the right-hand side of (1.110) according to

$$|\mathbf{a}_i^* \mathbf{B}\mathbf{q}| = \Big| \sum_{k=1}^{q} \mathbf{a}_i^* \mathbf{b}_k q_k \Big| \tag{1.116}$$

$$\leq \sum_{k=1}^{q} |\mathbf{a}_i^* \mathbf{b}_k| |q_k| \tag{1.117}$$

$$\leq \bar{\mu}(\mathbf{A}, \mathbf{B}) \|\mathbf{q}\|_1, \tag{1.118}$$

where the last step is by Definition 1.12. Combining the lower bound (1.111)–(1.115) and the upper bound (1.116)–(1.118) yields

$$(1 + \mu(\mathbf{A}))|p_i| - \mu(\mathbf{A})\|\mathbf{p}\|_1 \leq \bar{\mu}(\mathbf{A}, \mathbf{B})\|\mathbf{q}\|_1. \tag{1.119}$$

Since (1.119) holds for arbitrary $i \in \{1, \dots, p\}$, we can sum over all $i \in \mathcal{P}$ and get

$$\|\mathbf{p}_\mathcal{P}\|_1 \leq |\mathcal{P}| \bigg(\frac{\mu(\mathbf{A})\|\mathbf{p}\|_1 + \bar{\mu}(\mathbf{A}, \mathbf{B})\|\mathbf{q}\|_1}{1 + \mu(\mathbf{A})} \bigg). \tag{1.120}$$

$\square$

For the special case $\mathbf{A} = \mathbf{I} \in \mathbb{C}^{m \times m}$ and $\mathbf{B} \in \mathbb{C}^{m \times m}$ with $\mathbf{B}$ unitary, we have $\mu(\mathbf{A}) = \mu(\mathbf{B}) = 0$ and $\bar{\mu}(\mathbf{I}, \mathbf{B}) = \mu([\mathbf{I}\ \ \mathbf{B}])$, so that (1.108) and (1.109) simplify to

$$\|\mathbf{p}_\mathcal{P}\|_1 \leq |\mathcal{P}|\, \mu([\mathbf{I}\ \ \mathbf{B}])\, \|\mathbf{q}\|_1 \tag{1.121}$$

and

$$\|\mathbf{q}_\mathcal{Q}\|_1 \leq |\mathcal{Q}|\, \mu([\mathbf{I}\ \ \mathbf{B}])\, \|\mathbf{p}\|_1, \tag{1.122}$$

respectively. Thus, for arbitrary but fixed $\mathbf{p} \in \mathcal{W}^{\mathbf{B},\mathcal{Q}}$ and $\mathbf{q} = \mathbf{B}^*\mathbf{p}$, we have $\mathbf{q}_{\mathcal{Q}} = \mathbf{q}$ so that (1.121) and (1.122) taken together yield

$$\|\mathbf{p}_{\mathcal{P}}\|_1 \leq |\mathcal{P}||\mathcal{Q}|\,\mu^2([\mathbf{I}\;\;\mathbf{B}])\,\|\mathbf{p}\|_1. \tag{1.123}$$

As $\mathbf{p}$ was assumed to be arbitrary, by (1.80) this recovers the uncertainty relation

$$\Sigma_{\mathcal{P},\mathcal{Q}}(\mathbf{B}) \leq |\mathcal{P}||\mathcal{Q}|\mu^2([\mathbf{I}\;\;\mathbf{B}]) \tag{1.124}$$

in Lemma 1.8.

1.4.5 Concentration Inequalities for Pairs of General Matrices

We next refine the result in Theorem 1.13 to vectors that are concentrated in 1-norm according to Definition 1.9. The formal statement is as follows.

COROLLARY 3 *Let $\mathbf{A} \in \mathbb{C}^{m\times p}$ and $\mathbf{B} \in \mathbb{C}^{m\times q}$, both with column vectors $\|\cdot\|_2$-normalized to 1, $\mathcal{P} \subseteq \{1,\dots,p\}$, $\mathcal{Q} \subseteq \{1,\dots,q\}$, $\mathbf{p} \in \mathbb{C}^p$, and $\mathbf{q} \in \mathbb{C}^q$. Suppose that $\mathbf{Ap} = \mathbf{Bq}$. Then, the following statements hold.*

1 If $\mathbf{q}$ is $\varepsilon_{\mathcal{Q}}$-concentrated, then,

$$\|\mathbf{p}_{\mathcal{P}}\|_1 \leq \frac{|\mathcal{P}|}{1+\mu(\mathbf{A})}\left(\mu(\mathbf{A}) + \frac{\bar{\mu}^2(\mathbf{A},\mathbf{B})|\mathcal{Q}|}{[(1+\mu(\mathbf{B}))(1-\varepsilon_{\mathcal{Q}}) - \mu(\mathbf{B})|\mathcal{Q}|]_+}\right)\|\mathbf{p}\|_1. \tag{1.125}$$

2 If $\mathbf{p}$ is $\varepsilon_{\mathcal{P}}$-concentrated, then,

$$\|\mathbf{q}_{\mathcal{Q}}\|_1 \leq \frac{|\mathcal{Q}|}{1+\mu(\mathbf{B})}\left(\mu(\mathbf{B}) + \frac{\bar{\mu}^2(\mathbf{A},\mathbf{B})|\mathcal{P}|}{[(1+\mu(\mathbf{A}))(1-\varepsilon_{\mathcal{P}}) - \mu(\mathbf{A})|\mathcal{P}|]_+}\right)\|\mathbf{q}\|_1. \tag{1.126}$$

3 If $\mathbf{p}$ is $\varepsilon_{\mathcal{P}}$-concentrated, $\mathbf{q}$ is $\varepsilon_{\mathcal{Q}}$-concentrated, $\bar{\mu}(\mathbf{A},\mathbf{B}) > 0$, and $(\mathbf{p}^T\;\mathbf{q}^T)^T \neq \mathbf{0}$, then,

$$|\mathcal{P}||\mathcal{Q}| \geq \frac{[(1+\mu(\mathbf{A}))(1-\varepsilon_{\mathcal{P}}) - \mu(\mathbf{A})|\mathcal{P}|]_+[(1+\mu(\mathbf{B}))(1-\varepsilon_{\mathcal{Q}}) - \mu(\mathbf{B})|\mathcal{Q}|]_+}{\bar{\mu}^2(\mathbf{A},\mathbf{B})}. \tag{1.127}$$

Proof By Theorem 1.13, we have

$$\|\mathbf{p}_{\mathcal{P}}\|_1 \leq |\mathcal{P}|\left(\frac{\mu(\mathbf{A})\|\mathbf{p}\|_1 + \bar{\mu}(\mathbf{A},\mathbf{B})\|\mathbf{q}\|_1}{1+\mu(\mathbf{A})}\right) \tag{1.128}$$

and

$$\|\mathbf{q}_{\mathcal{Q}}\|_1 \leq |\mathcal{Q}|\left(\frac{\mu(\mathbf{B})\|\mathbf{q}\|_1 + \bar{\mu}(\mathbf{A},\mathbf{B})\|\mathbf{p}\|_1}{1+\mu(\mathbf{B})}\right). \tag{1.129}$$

Suppose now that $\mathbf{q}$ is $\varepsilon_{\mathcal{Q}}$-concentrated, i.e., $\|\mathbf{q}_{\mathcal{Q}}\|_1 \geq (1-\varepsilon_{\mathcal{Q}})\|\mathbf{q}\|_1$. Then, (1.129) implies that

$$\|\mathbf{q}\|_1 \leq \frac{|\mathcal{Q}|\bar{\mu}(\mathbf{A},\mathbf{B})}{[(1+\mu(\mathbf{B}))(1-\varepsilon_{\mathcal{Q}}) - \mu(\mathbf{B})|\mathcal{Q}|]_+}\|\mathbf{p}\|_1. \tag{1.130}$$

Using (1.130) in (1.128) yields (1.125). The relation (1.126) follows from (1.125) by swapping the roles of $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{p}$ and $\mathbf{q}$, and $\mathcal{P}$ and $\mathcal{Q}$, and upon noting that $\bar{\mu}(\mathbf{A},\mathbf{B}) = \bar{\mu}(\mathbf{B},\mathbf{A})$. It remains to establish (1.127). Using $\|\mathbf{p}_{\mathcal{P}}\|_1 \geq (1-\varepsilon_{\mathcal{P}})\|\mathbf{p}\|_1$ in (1.128) and $\|\mathbf{q}_{\mathcal{Q}}\|_1 \geq (1-\varepsilon_{\mathcal{Q}})\|\mathbf{q}\|_1$ in (1.129) yields

$$\|\mathbf{p}\|_1[(1+\mu(\mathbf{A}))(1-\varepsilon_{\mathcal{P}}) - \mu(\mathbf{A})|\mathcal{P}|]_+ \leq \bar{\mu}(\mathbf{A},\mathbf{B})\|\mathbf{q}\|_1|\mathcal{P}| \tag{1.131}$$

and

$$\|\mathbf{q}\|_1[(1+\mu(\mathbf{B}))(1-\varepsilon_{\mathcal{Q}}) - \mu(\mathbf{B})|\mathcal{Q}|]_+ \leq \bar{\mu}(\mathbf{A},\mathbf{B})\|\mathbf{p}\|_1|\mathcal{Q}|, \tag{1.132}$$

respectively. Suppose first that $\mathbf{p} = \mathbf{0}$. Then, $\mathbf{q} \neq \mathbf{0}$ by assumption, and (1.132) becomes

$$[(1+\mu(\mathbf{B}))(1-\varepsilon_{\mathcal{Q}}) - \mu(\mathbf{B})|\mathcal{Q}|]_+ = 0. \tag{1.133}$$

In this case (1.127) holds trivially. Similarly, if $\mathbf{q} = \mathbf{0}$, then $\mathbf{p} \neq \mathbf{0}$ again by assumption, and (1.131) becomes

$$[(1+\mu(\mathbf{A}))(1-\varepsilon_{\mathcal{P}}) - \mu(\mathbf{A})|\mathcal{P}|]_+ = 0. \tag{1.134}$$

As before, (1.127) holds trivially. Finally, if $\mathbf{p} \neq \mathbf{0}$ and $\mathbf{q} \neq \mathbf{0}$, then we multiply (1.131) by (1.132) and divide the result by $\bar{\mu}^2(\mathbf{A},\mathbf{B})\|\mathbf{p}\|_1\|\mathbf{q}\|_1$ which yields (1.127). □

Corollary 3 will be used in Section 1.5 to derive recovery thresholds for sparse signal separation. The lower bound on $|\mathcal{P}||\mathcal{Q}|$ in (1.127) is [9, Theorem 1] and states that a nonzero vector can not be arbitrarily well concentrated with respect to two different general matrices $\mathbf{A}$ and $\mathbf{B}$. For the special case $\varepsilon_{\mathcal{Q}} = 0$ and $\mathbf{A}$ and $\mathbf{B}$ unitary, and hence $\mu(\mathbf{A}) = \mu(\mathbf{B}) = 0$ and $\bar{\mu}(\mathbf{A},\mathbf{B}) = \mu([\mathbf{A}\ \ \mathbf{B}])$, (1.127) recovers Lemma 1.10.

Particularizing (1.127) to $\varepsilon_{\mathcal{P}} = \varepsilon_{\mathcal{Q}} = 0$ yields the following result.

COROLLARY 4 *[10, Lemma 33] Let $\mathbf{A} \in \mathbb{C}^{m\times p}$ and $\mathbf{B} \in \mathbb{C}^{m\times q}$, both with column vectors $\|\cdot\|_2$-normalized to 1, and consider $\mathbf{p} \in \mathbb{C}^p$ and $\mathbf{q} \in \mathbb{C}^q$ with $(\mathbf{p}^T\ \mathbf{q}^T)^T \neq \mathbf{0}$. Suppose that $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$. Then, $\|\mathbf{p}\|_0\|\mathbf{q}\|_0 \geq f_{\mathbf{A},\mathbf{B}}(\|\mathbf{p}\|_0, \|\mathbf{q}\|_0)$, where*

$$f_{\mathbf{A},\mathbf{B}}(u,v) = \frac{[1+\mu(\mathbf{A})(1-u)]_+[1+\mu(\mathbf{B})(1-v)]_+}{\bar{\mu}^2(\mathbf{A},\mathbf{B})}. \tag{1.135}$$

Proof Let $\mathcal{P} = \{i \in \{1,\ldots,p\} : p_i \neq 0\}$ and $\mathcal{Q} = \{i \in \{1,\ldots,q\} : q_i \neq 0\}$, so that $\mathbf{p}_{\mathcal{P}} = \mathbf{p}$, $\mathbf{q}_{\mathcal{Q}} = \mathbf{q}$, $|\mathcal{P}| = \|\mathbf{p}\|_0$, and $|\mathcal{Q}| = \|\mathbf{q}\|_0$. The claim now follows directly from (1.127) with $\varepsilon_{\mathcal{P}} = \varepsilon_{\mathcal{Q}} = 0$. □

If $\mathbf{A}$ and $\mathbf{B}$ are both unitary, then $\mu(\mathbf{A}) = \mu(\mathbf{B}) = 0$ and $\bar{\mu}(\mathbf{A},\mathbf{B}) = \mu([\mathbf{A}\ \ \mathbf{B}])$, and Corollary 4 recovers the Elad-Bruckstein result in Corollary 2.

Corollary 4 admits the following appealing geometric interpretation in terms of a null-space property, which will be seen in Section 1.6 to pave the way to an extension of the classical notion of sparsity to a more general concept of parsimony.

LEMMA 1.14 *Let* $\mathbf{A} \in \mathbb{C}^{m \times p}$ *and* $\mathbf{B} \in \mathbb{C}^{m \times q}$*, both with column vectors* $\|\cdot\|_2$*-normalized to* 1*. Then, the set (which actually is a finite union of subspaces)*

$$\mathcal{S} = \left\{ \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} : \mathbf{p} \in \mathbb{C}^p,\ \mathbf{q} \in \mathbb{C}^q,\ \|\mathbf{p}\|_0 \|\mathbf{q}\|_0 < f_{\mathbf{A},\mathbf{B}}(\|\mathbf{p}\|_0, \|\mathbf{q}\|_0) \right\} \tag{1.136}$$

with $f_{\mathbf{A},\mathbf{B}}$ *defined in* (1.135) *intersects the kernel of* $[\mathbf{A}\ \ \mathbf{B}]$ *trivially, i.e.,*

$$\ker([\mathbf{A}\ \ \mathbf{B}]) \cap \mathcal{S} = \{\mathbf{0}\}. \tag{1.137}$$

Proof The statement of this lemma is equivalent to the statement of Corollary 4 through a chain of equivalences between the following statements:

1 $\ker([\mathbf{A}\ \ \mathbf{B}]) \cap \mathcal{S} = \{\mathbf{0}\}$;
2 if $(\mathbf{p}^T\ -\mathbf{q}^T)^T \in \ker([\mathbf{A}\ \ \mathbf{B}])\backslash\{\mathbf{0}\}$, then $\|\mathbf{p}\|_0\|\mathbf{q}\|_0 \geq f_{\mathbf{A},\mathbf{B}}(\|\mathbf{p}\|_0, \|\mathbf{q}\|_0)$;
3 if $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ with $(\mathbf{p}^T\ \mathbf{q}^T)^T \neq \mathbf{0}$, then $\|\mathbf{p}\|_0\|\mathbf{q}\|_0 \geq f_{\mathbf{A},\mathbf{B}}(\|\mathbf{p}\|_0, \|\mathbf{q}\|_0)$,

where 1 $\Leftrightarrow$ 2 is by definition of $\mathcal{S}$, 2 $\Leftrightarrow$ 3 follows from the fact that $\mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ with $(\mathbf{p}^T\ \mathbf{q}^T)^T \neq \mathbf{0}$ is equivalent to $(\mathbf{p}^T\ -\mathbf{q}^T)^T \in \ker([\mathbf{A}\ \ \mathbf{B}])\backslash\{\mathbf{0}\}$, and 3 is the statement in Corollary 4. □

1.5 Sparse Signal Separation

Numerous practical signal recovery tasks can be cast as sparse signal separation problems of the following form. We want to recover $\mathbf{y} \in \mathbb{C}^p$ with $\|\mathbf{y}\|_0 \leq s$ and/or $\mathbf{z} \in \mathbb{C}^q$ with $\|\mathbf{z}\|_0 \leq t$ from the noiseless observation

$$\mathbf{w} = \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{z}, \tag{1.138}$$

where $\mathbf{A} \in \mathbb{C}^{m \times p}$ and $\mathbf{B} \in \mathbb{C}^{m \times q}$. Here, s and t are the sparsity levels of $\mathbf{y}$ and $\mathbf{z}$ with corresponding ambient dimensions p and q, respectively. Prominent applications include (image) inpainting, declipping, super-resolution, the recovery of signals corrupted by impulse noise, and the separation of (e.g., audio or video) signals into two distinct components [9, Section I]. We next briefly describe some of these problems.

1 *Clipping:* Non-linearities in power-amplifiers or in analog-to-digital converters often cause signal clipping or saturation [35]. This effect can be cast into the signal model (1.138) by setting $\mathbf{B} = \mathbf{I}$, identifying $\mathbf{s} = \mathbf{A}\mathbf{y}$ with the signal to be clipped, and setting $\mathbf{z} = (g_a(\mathbf{s}) - \mathbf{s})$ with $g_a(\cdot)$ realizing entry-wise clipping of the amplitude to the interval $[0, a]$. If the clipping level a is not too small, then $\mathbf{z}$ will be sparse, i.e., $t \ll q$.
2 *Missing entries:* Our framework also encompasses super-resolution [36,37] and inpainting [38] of, e.g., images, audio, and video signals. In both these applications only a subset of the entries of the (full-resolution) signal vector $\mathbf{s} = \mathbf{A}\mathbf{y}$ is available and the task is to fill in the missing entries, which are accounted for by writing $\mathbf{w} = \mathbf{s} + \mathbf{z}$ with $z_i = -s_i$ if the i-th entry of $\mathbf{s}$ is missing and

$z_i = 0$ else. If the number of entries missing is not too large, then $\mathbf{z}$ is sparse, i.e., $t \ll q$.

3 *Signal separation:* Separation of (audio, image, or video) signals into two structurally distinct components also fits into the framework described above. A prominent example is the separation of texture from cartoon parts in images (see [39,40] and references therein). The matrices $\mathbf{A}$ and $\mathbf{B}$ are chosen to allow for sparse representations of the two distinct features. Note that here $\mathbf{Bz}$ no longer plays the role of undesired noise and the goal is to recover both $\mathbf{y}$ and $\mathbf{z}$ from the observation $\mathbf{w} = \mathbf{Ay} + \mathbf{Bz}$.

The first two examples above demonstrate that in many practically relevant applications the locations of the possibly nonzero entries of one of the sparse vectors, say $\mathbf{z}$, may be known. This can be accounted for by removing the columns of $\mathbf{B}$ corresponding to the other entries, which results in $t = q$, i.e., the sparsity level of $\mathbf{z}$ equals the ambient dimension. We next show how Corollary 3 can be used to state a sufficient condition for recovery of $\mathbf{y}$ from $\mathbf{w} = \mathbf{Ay} + \mathbf{Bz}$ when $t = q$. For recovery guarantees in the case where the sparsity levels of both $\mathbf{y}$ and $\mathbf{z}$ are strictly smaller than their corresponding ambient dimensions, we refer to [9, Theorem 8].

THEOREM 1.15 *[9, Theorem 4, Theorem 7] Let $\mathbf{y} \in \mathbb{C}^p$ with $\|\mathbf{y}\|_0 \leq s$, $\mathbf{z} \in \mathbb{C}^q$, $\mathbf{A} \in \mathbb{C}^{m \times p}$, and $\mathbf{B} \in \mathbb{C}^{m \times q}$, both with column vectors $\|\cdot\|_2$-normalized to 1 and $\bar{\mu}(\mathbf{A}, \mathbf{B}) > 0$. Suppose that*

$$2sq < f_{\mathbf{A},\mathbf{B}}(2s, q) \tag{1.139}$$

with

$$f_{\mathbf{A},\mathbf{B}}(u, v) = \frac{[1 + \mu(\mathbf{A})(1 - u)]_+ [1 + \mu(\mathbf{B})(1 - v)]_+}{\bar{\mu}^2(\mathbf{A}, \mathbf{B})}. \tag{1.140}$$

Then, $\mathbf{y}$ can be recovered from $\mathbf{w} = \mathbf{Ay}+\mathbf{Bz}$ by either of the following algorithms:

$$\text{(P0)} \quad \begin{cases} \textit{minimize } \|\tilde{\mathbf{y}}\|_0 \\ \textit{subject to } \mathbf{A}\tilde{\mathbf{y}} \in \{\mathbf{w} + \mathbf{B}\tilde{\mathbf{z}} : \tilde{\mathbf{z}} \in \mathbb{C}^q\}. \end{cases} \tag{1.141}$$

$$\text{(P1)} \quad \begin{cases} \textit{minimize } \|\tilde{\mathbf{y}}\|_1 \\ \textit{subject to } \mathbf{A}\tilde{\mathbf{y}} \in \{\mathbf{w} + \mathbf{B}\tilde{\mathbf{z}} : \tilde{\mathbf{z}} \in \mathbb{C}^q\}. \end{cases} \tag{1.142}$$

Proof We provide the proof for (P1) only. The proof for recovery through (P0) is very similar and can be found in [9, Appendix B].

Let $\mathbf{w} = \mathbf{Ay} + \mathbf{Bz}$ and suppose that (P1) delivers $\tilde{\mathbf{y}} \in \mathbb{C}^p$. This implies $\|\tilde{\mathbf{y}}\|_1 \leq \|\mathbf{y}\|_1$ and the existence of a $\tilde{\mathbf{z}} \in \mathbb{C}^q$ such that

$$\mathbf{A}\tilde{\mathbf{y}} = \mathbf{w} + \mathbf{B}\tilde{\mathbf{z}}. \tag{1.143}$$

On the other hand, we also have

$$\mathbf{Ay} = \mathbf{w} - \mathbf{Bz}. \tag{1.144}$$

Subtracting (1.144) from (1.143) yields

$$\mathbf{A}(\underbrace{\tilde{\mathbf{y}}-\mathbf{y}}_{=\mathbf{p}}) = \mathbf{B}(\underbrace{\tilde{\mathbf{z}}+\mathbf{z}}_{=\mathbf{q}}). \tag{1.145}$$

We now set

$$\mathcal{U} = \{i \in \{1,\dots,p\} : y_i \neq 0\} \tag{1.146}$$

and

$$\mathcal{U}^c = \{1,\dots,p\}\backslash\mathcal{U} \tag{1.147}$$

and show that $\mathbf{p}$ is $\varepsilon_{\mathcal{U}}$-concentrated (with respect to 1-norm) for $\varepsilon_{\mathcal{U}} = 1/2$, i.e.,

$$\|\mathbf{p}_{\mathcal{U}^c}\|_1 \leq \frac{1}{2}\|\mathbf{p}\|_1. \tag{1.148}$$

We have

$$\begin{aligned}
\|\mathbf{y}\|_1 &\geq \|\tilde{\mathbf{y}}\|_1 && (1.149)\\
&= \|\mathbf{y}+\mathbf{p}\|_1 && (1.150)\\
&= \|\mathbf{y}_{\mathcal{U}}+\mathbf{p}_{\mathcal{U}}\|_1 + \|\mathbf{p}_{\mathcal{U}^c}\|_1 && (1.151)\\
&\geq \|\mathbf{y}_{\mathcal{U}}\|_1 - \|\mathbf{p}_{\mathcal{U}}\|_1 + \|\mathbf{p}_{\mathcal{U}^c}\|_1 && (1.152)\\
&= \|\mathbf{y}\|_1 - \|\mathbf{p}_{\mathcal{U}}\|_1 + \|\mathbf{p}_{\mathcal{U}^c}\|_1, && (1.153)
\end{aligned}$$

where (1.151) follows from the definition of $\mathcal{U}$ in (1.146), and in (1.152) we applied the reverse triangle inequality. Now, (1.149)–(1.153) implies $\|\mathbf{p}_{\mathcal{U}}\|_1 \geq \|\mathbf{p}_{\mathcal{U}^c}\|_1$. Thus, $2\|\mathbf{p}_{\mathcal{U}^c}\|_1 \leq \|\mathbf{p}_{\mathcal{U}}\|_1 + \|\mathbf{p}_{\mathcal{U}^c}\|_1 = \|\mathbf{p}\|_1$, which establishes (1.148). Next, set $\mathcal{V} = \{1,\dots,q\}$ and note that $\mathbf{q}$ is trivially $\varepsilon_{\mathcal{V}}$-concentrated (with respect to 1-norm) for $\varepsilon_{\mathcal{V}} = 0$. Suppose, towards a contradiction, that $\mathbf{p} \neq \mathbf{0}$. Then, we have

$$\begin{aligned}
&2sq && (1.154)\\
&\geq 2|\mathcal{U}||\mathcal{V}| && (1.155)\\
&\geq \frac{[(1+\mu(\mathbf{A})) - 2\mu(\mathbf{A})|\mathcal{U}|]_+[1+\mu(\mathbf{B})(1-|\mathcal{V}|)]_+}{\bar{\mu}^2(\mathbf{A},\mathbf{B})} && (1.156)\\
&\geq \frac{[(1+\mu(\mathbf{A})) - 2s\mu(\mathbf{A})]_+[1+\mu(\mathbf{B})(1-q)]_+}{\bar{\mu}^2(\mathbf{A},\mathbf{B})}, && (1.157)
\end{aligned}$$

where (1.156) is obtained by applying Part 3 of Corollary 3 with $\mathbf{p}$ $\varepsilon_{\mathcal{U}}$-concentrated for $\varepsilon_{\mathcal{U}} = 1/2$ and $\mathbf{q}$ $\varepsilon_{\mathcal{V}}$-concentrated for $\varepsilon_{\mathcal{V}} = 0$. But (1.154)—(1.157) contradicts (1.139). Hence, we must have $\mathbf{p} = \mathbf{0}$, which yields $\tilde{\mathbf{y}} = \mathbf{y}$. □

We next provide an example showing that, as soon as (1.139) is saturated, recovery through (P0) or (P1) can fail. Take $m = n^2$ with n even, $\mathbf{A} = \mathbf{F} \in \mathbb{C}^{m\times m}$, and $\mathbf{B} \in \mathbb{C}^{m\times\sqrt{m}}$ containing every $\sqrt{m}$-th column of the $m \times m$ identity matrix, i.e.,

$$B_{k,l} = \begin{cases} 1 & \text{if } k = \sqrt{m}\, l \\ 0 & \text{else} \end{cases} \tag{1.158}$$

for all $k \in \{1, \dots, m\}$ and $l \in \{1, \dots, \sqrt{m}\}$. For every $a \in \mathbb{N}$ dividing m, we define the vector $\mathbf{d}^{(a)} \in \mathbb{C}^m$ with components

$$d_l^{(a)} = \begin{cases} 1 & \text{if } l \in \left\{a, 2a, \dots, \left(\frac{m}{a} - 1\right)a, m\right\} \\ 0 & \text{else.} \end{cases} \tag{1.159}$$

Straightforward calculations now yield

$$\mathbf{F}\mathbf{d}^{(a)} = \frac{\sqrt{m}}{a}\mathbf{d}^{(m/a)} \tag{1.160}$$

for all $a \in \mathbb{N}$ dividing m. Suppose that $\mathbf{w} = \mathbf{F}\mathbf{y} + \mathbf{B}\mathbf{z}$ with

$$\mathbf{y} = \mathbf{d}^{(2\sqrt{m})} - \mathbf{d}^{(\sqrt{m})} \in \mathbb{C}^m \tag{1.161}$$

$$\mathbf{z} = (1 \dots 1)^T \in \mathbb{C}^{\sqrt{m}}. \tag{1.162}$$

Evaluating (1.139) for $\mathbf{A} = \mathbf{F}$, $\mathbf{B}$ as defined in (1.158), and $q = \sqrt{m}$ results in $s < \sqrt{m}/2$. Now, $\mathbf{y}$ in (1.161) has $\|\mathbf{y}\|_0 = \sqrt{m}/2$ and thus just violates the threshold $s < \sqrt{m}/2$. We next show that this slender violation is enough for the existence of an alternative pair $\tilde{\mathbf{y}} \in \mathbb{C}^m$, $\tilde{\mathbf{z}} \in \mathbb{C}^{\sqrt{m}}$ satisfying $\mathbf{w} = \mathbf{F}\tilde{\mathbf{y}} + \mathbf{B}\tilde{\mathbf{z}}$ with $\|\tilde{\mathbf{y}}\|_0 = \|\mathbf{y}\|_0$ and $\|\tilde{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1$. Thus, neither (P0) nor (P1) can distinguish between $\mathbf{y}$ and $\tilde{\mathbf{y}}$. Specifically, we set

$$\tilde{\mathbf{y}} = \mathbf{d}^{(2\sqrt{m})} \in \mathbb{C}^m \tag{1.163}$$

$$\tilde{\mathbf{z}} = \mathbf{0} \in \mathbb{C}^m \tag{1.164}$$

and note that $\|\tilde{\mathbf{y}}\|_0 = \|\mathbf{y}\|_0 = \|\tilde{\mathbf{y}}\|_1 = \|\mathbf{y}\|_1 = \sqrt{m}/2$. It remains to establish that $\mathbf{w} = \mathbf{F}\tilde{\mathbf{y}} + \mathbf{B}\tilde{\mathbf{z}}$. To this end, first note that

$$\mathbf{w} = \mathbf{F}\mathbf{y} + \mathbf{B}\mathbf{z} \tag{1.165}$$

$$= \frac{1}{2}\mathbf{d}^{(\sqrt{m}/2)} - \mathbf{d}^{(\sqrt{m})} + \mathbf{B}\mathbf{z} \tag{1.166}$$

$$= \frac{1}{2}\mathbf{d}^{(\sqrt{m}/2)}, \tag{1.167}$$

where (1.166) follows from (1.160) and (1.167) is by (1.158). Finally, again using (1.160), we find that

$$\mathbf{F}\tilde{\mathbf{y}} + \mathbf{B}\tilde{\mathbf{z}} = \frac{1}{2}\mathbf{d}^{(\sqrt{m}/2)}, \tag{1.168}$$

which completes the argument.

The threshold $s < \sqrt{m}/2$ constitutes a special instance of the so-called "square-root bottleneck" [41] all coherence-based deterministic recovery thresholds suffer from. The square-root bottleneck says that the number of measurements, m, has to scale at least quadratically in the sparsity level s. It can be circumvented by considering random models for either the signals or the measurement matrices [42–45] leading to thresholds of the form $m \propto s \log p$ and applying with high probability. Deterministic linear recovery thresholds, i.e., $m \propto s$, have, to the best of our knowledge, first been reported in [46] for the DFT measurement matrix under positivity constraints on the vector to be recovered. Further instances of deterministic linear recovery thresholds were discovered in the context of spectrum-blind sampling [47, 48] and system identification [49].

1.6 The Set-Theoretic Null-Space Property

The notion of sparsity underlying the theory developed so far is that of either the number of nonzero entries or of concentration in terms of 1-norm or 2-norm. In practice, one often encounters more general concepts of parsimony, such as manifold or fractal set structures. Manifolds are prevalent in data science, e.g., in compressed sensing [22–27], machine learning [28], image processing [29, 30], and handwritten digit recognition [31]. Fractal sets find application in image compression and in modeling of Ethernet traffic [32]. Based on the null-space property established in Lemma 1.14, we now extend the theory to account for more general notions of parsimony. To this end, we first need a suitable measure of "description complexity" that goes beyond the concepts of sparsity and concentration. Formalizing this idea requires an adequate dimension measure, which, as it turns out, is lower modified Minkowski dimension. We start by defining Minkowski dimension and modified Minkowski dimension.

DEFINITION 1.16 [50, Section 3.1][3] For $\mathcal{U} \subseteq \mathbb{C}^m$ nonempty, the lower and upper Minkowski dimension of $\mathcal{U}$ is defined as

$$\underline{\dim}_{\mathrm{B}}(\mathcal{U}) = \liminf_{\rho \to 0} \frac{\log N_{\mathcal{U}}(\rho)}{\log \frac{1}{\rho}} \tag{1.169}$$

and

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) = \limsup_{\rho \to 0} \frac{\log N_{\mathcal{U}}(\rho)}{\log \frac{1}{\rho}}, \tag{1.170}$$

respectively, where

$$N_{\mathcal{U}}(\rho) = \min \left\{ k \in \mathbb{N} : \mathcal{U} \subseteq \bigcup_{i \in \{1,\ldots,k\}} \mathcal{B}_m(\mathbf{u}_i, \rho),\ \mathbf{u}_i \in \mathcal{U} \right\} \tag{1.171}$$

is the covering number of $\mathcal{U}$ for radius $\rho > 0$. If $\underline{\dim}_{\mathrm{B}}(\mathcal{U}) = \overline{\dim}_{\mathrm{B}}(\mathcal{U})$, this common value, denoted by $\dim_{\mathrm{B}}(\mathcal{U})$, is the Minkowski dimension of $\mathcal{U}$.

[3] Minkowski dimension is sometimes also referred to as box-counting dimension, which is the origin of the subscript B in the notation $\dim_{\mathrm{B}}(\cdot)$ used henceforth.

DEFINITION 1.17 [50, Section 3.3] For $\mathcal{U} \subseteq \mathbb{C}^m$ nonempty, the lower and upper modified Minkowski dimension of $\mathcal{U}$ is defined as

$$\underline{\dim}_{\mathrm{MB}}(\mathcal{U}) = \inf\left\{ \sup_{i\in\mathbb{N}} \underline{\dim}_{\mathrm{B}}(\mathcal{U}_i) : \mathcal{U} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{U}_i \right\} \tag{1.172}$$

and

$$\overline{\dim}_{\mathrm{MB}}(\mathcal{U}) = \inf\left\{ \sup_{i\in\mathbb{N}} \overline{\dim}_{\mathrm{B}}(\mathcal{U}_i) : \mathcal{U} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{U}_i \right\}, \tag{1.173}$$

respectively, where in both cases the infimum is over all possible coverings $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of $\mathcal{U}$ by nonempty compact sets $\mathcal{U}_i$. If $\underline{\dim}_{\mathrm{MB}}(\mathcal{U}) = \overline{\dim}_{\mathrm{MB}}(\mathcal{U})$, this common value, denoted by $\dim_{\mathrm{MB}}(\mathcal{U})$, is the modified Minkowski dimension of $\mathcal{U}$.

For further details on (modified) Minkowski dimension, we refer the interested reader to [50, Section 3].

We are now ready to extend the null-space property in Lemma 1.14 to the following set-theoretic null-space property.

THEOREM 1.18 *Let $\mathcal{U} \subseteq \mathbb{C}^{p+q}$ be nonempty with $\underline{\dim}_{\mathrm{MB}}(\mathcal{U}) < 2m$, and let $\mathbf{B} \in \mathbb{C}^{m\times q}$ with $m \geq q$ be a full-rank matrix. Then, $\ker[\mathbf{A}\ \ \mathbf{B}] \cap (\mathcal{U}\backslash\{\mathbf{0}\}) = \emptyset$ for Lebesgue a.a. $\mathbf{A} \in \mathbb{C}^{m\times p}$.*

Proof See Section 1.9. □

The set $\mathcal{U}$ in this set-theoretic null-space property generalizes the finite union of linear subspaces $\mathcal{S}$ in Lemma 1.14. For $\mathcal{U} \subseteq \mathbb{R}^{p+q}$, the equivalent of Theorem 1.18 was reported previously in [13, Proposition 1]. The set-theoretic null-space property can be interpreted in geometric terms as follows. If $p + q \leq m$, then $[\mathbf{A}\ \ \mathbf{B}]$ is a tall matrix so that the kernel of $[\mathbf{A}\ \ \mathbf{B}]$ is $\{\mathbf{0}\}$ for Lebesgue-a.a. matrices $\mathbf{A}$. The statement of the theorem holds trivially in this case. If $p+q > m$, then the kernel of $[\mathbf{A}\ \ \mathbf{B}]$ is a $(p+q-m)$-dimensional subspace of the ambient space $\mathbb{C}^{p+q}$ for Lebesgue-a.a. matrices $\mathbf{A}$. The theorem therefore says that, for Lebesgue-a.a. $\mathbf{A}$, the set $\mathcal{U}$ intersects the subspace $\ker([\mathbf{A}\ \ \mathbf{B}])$ at most trivially if the sum of $\dim\ker([\mathbf{A}\ \ \mathbf{B}])$ and[4] $\underline{\dim}_{\mathrm{MB}}(\mathcal{U})/2$ is strictly smaller than the dimension of the ambient space. What is remarkable here is that the notions of Euclidean dimension (for the kernel of $[\mathbf{A}\ \ \mathbf{B}]$) and of lower modified Minkowski dimension (for the set $\mathcal{U}$) are compatible. We finally note that, by virtue of the chain of equivalences in the proof of Lemma 1.14, the set-theoretic null-space property in Theorem 1.18 leads to a set-theoretic uncertainty relation, albeit not in the form of an upper bound on an operator norm; for a detailed discussion of this equivalence the interested reader is referred to [13].

[4] The factor 1/2 stems from the fact that (modified) Minkowski dimension "counts real dimensions". For example, the modified Minkowski dimension of an n-dimensional linear subspace of $\mathbb{C}^m$ is $2n$ [26, Example II.2].

We next put the set-theoretic null-space property in Theorem 1.18 in perspective with the null-space property in Lemma 1.14. Fix the sparsity levels s and t, consider the set

$$\mathcal{S}_{s,t} = \left\{ \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} : \mathbf{p} \in \mathbb{C}^p, \mathbf{q} \in \mathbb{C}^q, \|\mathbf{p}\|_0 \le s, \|\mathbf{q}\|_0 \le t \right\}, \tag{1.174}$$

which is a finite union of $(s+t)$-dimensional linear subspaces, and, for the sake of concreteness, let $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = \mathbf{F}$ of size $q \times q$. Lemma 1.14 then states that the kernel of $[\mathbf{I} \;\; \mathbf{F}]$ intersects $\mathcal{S}_{s,t}$ trivially provided that

$$m > st, \tag{1.175}$$

which leads to a recovery threshold in the signal separation problem that is quadratic in the sparsity levels s and t [9, Theorem 8]. To see what the set-theoretic null-space property gives, we start by noting that, by [26, Example II.2], $\dim_{\mathrm{MB}}(\mathcal{S}_{s,t}) = 2(s+t)$. Theorem 1.18 hence states that, for Lebesgue a.a. matrices $\mathbf{A} \in \mathbb{C}^{m \times p}$, the kernel of $[\mathbf{A} \;\; \mathbf{B}]$ intersects $\mathcal{S}_{s,t}$ trivially, provided that

$$m > s + t. \tag{1.176}$$

This is striking as it says that, while the threshold in (1.175) is quadratic in the sparsity levels s and t and, therefore, suffers from the square-root bottleneck, the threshold in (1.176) is linear in s and t.

To understand the operational implications of the observation just made, we demonstrate how the set-theoretic null-space property in Theorem 1.18 leads to a sufficient condition for the recovery of vectors in sets of small lower modified Minkowski dimension.

LEMMA 1.19 *Let $\mathcal{S} \subseteq \mathbb{C}^{p+q}$ be nonempty with $\underline{\dim}_{\mathrm{MB}}(\mathcal{S} \ominus \mathcal{S}) < 2m$, where $\mathcal{S} \ominus \mathcal{S} = \{\mathbf{u} - \mathbf{v} : \mathbf{u}, \mathbf{v} \in \mathcal{S}\}$, and let $\mathbf{B} \in \mathbb{C}^{m \times q}$, with $m \ge q$, be a full-rank matrix. Then, $[\mathbf{A} \;\; \mathbf{B}]$ is one-to-one on $\mathcal{S}$ for Lebesgue a.a. $\mathbf{A} \in \mathbb{C}^{m \times p}$.*

Proof Follows immediately from the set-theoretic null-space property in Theorem 1.18 and linearity of $[\mathbf{A} \;\; \mathbf{B}]$. □

To elucidate the implications of Lemma 1.19, consider $\mathcal{S}_{s,t}$ defined in (1.174). Since $\mathcal{S}_{s,t} \ominus \mathcal{S}_{s,t}$ is again a finite union of linear subspaces of dimensions no larger than $\min(p, 2s) + \min(q, 2t)$, where the $\min(\cdot, \cdot)$-operation accounts for the fact that the dimension of a linear subspace can not exceed the dimension of its ambient space, we have [26, Example II.2]

$$\dim_{\mathrm{MB}}(\mathcal{S}_{s,t} \ominus \mathcal{S}_{s,t}) = 2(\min(p, 2s) + \min(q, 2t)). \tag{1.177}$$

Application of Lemma 1.19 now yields that, for Lebesgue a.a. matrices $\mathbf{A} \in \mathbb{C}^{m \times p}$, we can recover $\mathbf{y} \in \mathbb{C}^p$ with $\|\mathbf{y}\|_0 \le s$ and $\mathbf{z} \in \mathbb{C}^q$ with $\|\mathbf{z}\|_0 \le t$ from $\mathbf{w} = \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{z}$ provided that $m > \min(p, 2s) + \min(q, 2t)$. This qualitative behavior (namely, linear in $s + t$) is best possible as it can not be improved even if the support sets of $\mathbf{y}$ and $\mathbf{z}$ were known prior to recovery. We emphasize, however, that the statement in Lemma 1.19 guarantees injectivity of $[\mathbf{A}\ \ \mathbf{B}]$ only absent computational considerations for recovery.

1.7 A Large Sieve Inequality in $(\mathbb{C}^m, \|\cdot\|_2)$

We present a slightly improved and generalized version of the large sieve inequality stated in [7, Equation (32)].

LEMMA 1.20 *Let μ be a 1-periodic, σ-finite measure on $\mathbb{R}$, $n \in \mathbb{N}$, $\varphi \in [0,1)$, $\mathbf{a} \in \mathbb{C}^n$, and consider the 1-periodic trigonometric polynomial*

$$\psi(s) = e^{j2\pi\varphi} \sum_{k=1}^{n} a_k e^{-2\pi jks}. \tag{1.178}$$

Then,

$$\int_{[0,1)} |\psi(s)|^2 \mathrm{d}\mu(s) \le \left(n - 1 + \frac{1}{\delta}\right) \sup_{r\in[0,1)} \mu((r, r+\delta)) \|\mathbf{a}\|_2^2 \tag{1.179}$$

for all $\delta \in (0,1]$.

Proof Since

$$|\psi(s)| = \left| \sum_{k=1}^{n} a_k e^{-2\pi jks} \right|, \tag{1.180}$$

we can assume, without loss of generality, that $\varphi = 0$. The proof now follows closely the line of argumentation in [51, pp. 185–186] and in the proof of [7, Lemma 5]. Specifically, we make use of the result in [51, p. 185] saying that, for every $\delta > 0$, there exists a function $g \in L^2(\mathbb{R})$ with Fourier transform

$$G(s) = \int_{-\infty}^{\infty} g(t) e^{-2\pi jst} \mathrm{d}t \tag{1.181}$$

such that $\|G\|_2^2 = n - 1 + 1/\delta$, $|g(t)|^2 \ge 1$ for all $t \in [1, n]$, and $G(s) = 0$ for all $s \notin [-\delta/2, \delta/2]$. With this g, consider the 1-periodic trigonometric polynomial

$$\theta(s) = \sum_{k=1}^{n} \frac{a_k}{g(k)} e^{-2\pi jks} \tag{1.182}$$

and note that

$$\int_{-\delta/2}^{\delta/2} G(r)\theta(s-r)\mathrm{d}r = \sum_{k=1}^{n} \frac{a_k}{g(k)} e^{-2\pi jks} \int_{-\infty}^{\infty} G(r) e^{2\pi jkr} \mathrm{d}r \tag{1.183}$$

$$= \sum_{k=1}^{n} a_k e^{-2\pi jks} \tag{1.184}$$

$$= \psi(s) \quad \text{for all } s \in \mathbb{R}. \tag{1.185}$$

We now have

$$\int_{[0,1)} |\psi(s)|^2 \mathrm{d}\mu(s) = \int_{[0,1)} \left| \int_{-\delta/2}^{\delta/2} G(r)\theta(s-r)\mathrm{d}r \right|^2 \mathrm{d}\mu(s) \tag{1.186}$$

$$\le \|G\|_2^2 \int_{[0,1)} \left(\int_{-\delta/2}^{\delta/2} |\theta(s-r)|^2 \mathrm{d}r \right) \mathrm{d}\mu(s) \tag{1.187}$$

$$= \|G\|_2^2 \int_{[0,1)} \left(\int_{s-\delta/2}^{s+\delta/2} |\theta(r)|^2 \mathrm{d}r \right) \mathrm{d}\mu(s) \tag{1.188}$$

$$= \|G\|_2^2 \int_{-1}^{2} \mu\big((r-\delta/2, r+\delta/2) \cap [0,1)\big) |\theta(r)|^2 \mathrm{d}r \tag{1.189}$$

$$= \|G\|_2^2 \sum_{i=-1}^{1} \int_{0+i}^{1+i} \mu\big((r-\delta/2, r+\delta/2) \cap [0,1)\big) |\theta(r)|^2 \mathrm{d}r \tag{1.190}$$

$$= \|G\|_2^2 \sum_{i=-1}^{1} \int_{0}^{1} \mu\big((r-\delta/2, r+\delta/2) \cap [i,1+i)\big) |\theta(r)|^2 \mathrm{d}r \tag{1.191}$$

$$= \|G\|_2^2 \int_{0}^{1} \mu\big((r-\delta/2, r+\delta/2) \cap [-1,2)\big) |\theta(r)|^2 \mathrm{d}r \tag{1.192}$$

$$= \|G\|_2^2 \int_{0}^{1} \mu\big((r-\delta/2, r+\delta/2)\big) |\theta(r)|^2 \mathrm{d}r \tag{1.193}$$

for all $\delta \in (0,1]$, where (1.186) follows from (1.183)–(1.185), in (1.187) we applied the Cauchy-Schwartz inequality [52, Theorem 1.37], (1.189) is by Fubini's theorem [53, Theorem 1.14] (recall that μ is σ-finite by assumption) upon noting that

$$\{(r,s) : s \in [0,1), r \in (s-\delta/2, s+\delta/2)\} \tag{1.194}$$

$$= \{(r,s) : r \in [-1,2), s \in (r-\delta/2, r+\delta/2) \cap [0,1)\} \tag{1.195}$$

for all $\delta \in (0,1]$, in (1.191) we used the 1-periodicity of μ and θ, and (1.192) is by σ-additivity of μ. Now,

$$\int_0^1 \mu\big((r-\delta/2, r+\delta/2)\big) |\theta(r)|^2 \mathrm{d}r \le \sup_{r \in [0,1)} \mu((r, r+\delta)) \int_0^1 |\theta(r)|^2 \mathrm{d}r \tag{1.196}$$

$$= \sup_{r \in [0,1)} \mu((r, r+\delta)) \sum_{k=1}^{n} \frac{|a_k|^2}{|g(k)|^2} \tag{1.197}$$

$$\le \sup_{r \in [0,1)} \mu((r, r+\delta)) \|\mathbf{a}\|_2^2 \tag{1.198}$$

for all $\delta > 0$, where (1.198) follows from $|g(t)|^2 \ge 1$ for all $t \in [1,n]$. Using (1.196)–(1.198) and $\|G\|_2^2 = n - 1 + 1/\delta$ in (1.193) establishes (1.179). $\square$

Lemma 1.20 is a slightly strengthened version of the large sieve inequality [7, Equation (32)]. Specifically, in (1.179) it is sufficient to consider open intervals $(r, r+\delta)$, whereas [7, Equation (32)] requires closed intervals $[r, r+\delta]$. Thus, the upper bound in [7, Equation (32)] can be strictly larger than that in (1.179) whenever μ has mass points.

1.8 Uncertainty Relations in L_1 and L_2

The following table contains a list of infinite-dimensional counterparts—available in the literature—to results in this chapter. Specifically, these results apply to band-limited L_1- and L_2-functions and correspond to $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = \mathbf{F}$ in our setting.

	L_2 analog	L_1 analog
Upper bound in (1.10)	[6, Lemma 2]	
Corollary 1	[6, Theorem 2]	
Lemma 1.7	[6, Theorem 4]	
Lemma 1.8		[6, Lemma 3]
Lemma 1.11		[7, Lemma 2]
Lemma 1.20		[7, Theorem 4]

1.9 Proof of Theorem 1.18

By definition of lower modified Minkowski dimension, there exists a covering $\{\mathcal{U}_i\}_{i\in\mathbb{N}}$ of $\mathcal{U}$ by nonempty compact sets $\mathcal{U}_i$ satisfying $\underline{\dim}_{\mathrm{B}}(\mathcal{U}_i) < 2m$ for all $i \in \mathbb{N}$. The countable subadditivity of Lebesgue measure λ now implies

$$\lambda(\{\mathbf{A} \in \mathbb{C}^{m\times p} : \ker[\mathbf{A}\ \ \mathbf{B}] \cap (\mathcal{U}\backslash\{\mathbf{0}\}) \neq \emptyset\}) \tag{1.199}$$

$$\leq \sum_{i=1}^{\infty} \lambda(\{\mathbf{A} \in \mathbb{C}^{m\times p} : \ker[\mathbf{A}\ \ \mathbf{B}] \cap (\mathcal{U}_i\backslash\{\mathbf{0}\}) \neq \emptyset\}). \tag{1.200}$$

We next establish that every term in the sum on the right-hand side of (1.200) equals zero. Take an arbitrary but fixed $i \in \mathbb{N}$. Repeating the steps in [13, Equation (10)–(14)] shows that it is sufficient to prove that

$$\mathbb{P}[\ker([\mathbf{A}\ \ \mathbf{B}]) \cap \mathcal{V} \neq \emptyset] = 0 \tag{1.201}$$

with

$$\mathcal{V} = \left\{ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} : \mathbf{u} \in \mathbb{C}^p, \mathbf{v} \in \mathbb{C}^q, \|\mathbf{u}\|_2 > 0 \right\} \cap \mathcal{U}_i \tag{1.202}$$

and $\mathbf{A} = (\mathbf{A}_1 \dots \mathbf{A}_m)^*$, where the random vectors $\mathbf{A}_i$ are independent and uniformly distributed on $\mathcal{B}_p(\mathbf{0}, r)$ for arbitrary but fixed $r > 0$. Suppose, towards a contradiction, that (1.201) is false. This implies

$$0 = \liminf_{\rho \to 0} \frac{\log \mathbb{P}[\ker([\mathbf{A} \ \ \mathbf{B}]) \cap \mathcal{V} \neq \emptyset]}{\log \frac{1}{\rho}} \tag{1.203}$$

$$\leq \liminf_{\rho \to 0} \frac{\log \sum_{i=1}^{N_{\mathcal{V}}(\rho)} \mathbb{P}[\ker([\mathbf{A} \ \ \mathbf{B}]) \cap \mathcal{B}_{p+q}(\mathbf{c}_i, \rho) \neq \emptyset]}{\log \frac{1}{\rho}}, \tag{1.204}$$

where we have chosen $\{\mathbf{c}_i : i = 1, \dots, N_{\mathcal{V}}(\rho)\} \subseteq \mathcal{V}$ such that

$$\mathcal{V} \subseteq \bigcup_{i=1}^{N_{\mathcal{V}}(\rho)} \mathcal{B}_{p+q}(\mathbf{c}_i, \rho) \tag{1.205}$$

with $N_{\mathcal{V}}(\rho)$ denoting the covering number of $\mathcal{V}$ for radius $\rho > 0$ (cf. (1.171)). Now let $i \in \{1, \dots, N_{\mathcal{V}}(\rho)\}$ be arbitrary but fixed and write $\mathbf{c}_i = (\mathbf{u}_i^T \ \ \mathbf{v}_i^T)^T$. It follows that

$$\|\mathbf{A}\mathbf{u}_i + \mathbf{B}\mathbf{v}_i\|_2 = \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{c}_i\|_2 \tag{1.206}$$

$$\leq \|[\mathbf{A} \ \ \mathbf{B}](\mathbf{x} - \mathbf{c}_i)\|_2 + \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{x}\|_2 \tag{1.207}$$

$$\leq \|[\mathbf{A} \ \ \mathbf{B}]\|_2 \|(\mathbf{x} - \mathbf{c}_i)\|_2 + \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{x}\|_2 \tag{1.208}$$

$$\leq (\|[\mathbf{A}\|_2 + \|\mathbf{B}\|_2)\rho + \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{x}\|_2 \tag{1.209}$$

$$\leq (r\sqrt{m} + \|\mathbf{B}\|_2)\rho + \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{x}\|_2 \quad \text{for all } \mathbf{x} \in \mathcal{B}_{p+q}(\mathbf{c}_i, \rho), \tag{1.210}$$

where in the last step we made use of $\|\mathbf{A}_i\|_2 \leq r$ for $i = 1, \dots, m$. We now have

$$\mathbb{P}[\ker([\mathbf{A} \ \ \mathbf{B}]) \cap \mathcal{B}_{p+q}(\mathbf{c}_i, \rho) \neq \emptyset] \tag{1.211}$$

$$\leq \mathbb{P}[\exists \, \mathbf{x} \in \mathcal{B}_{p+q}(\mathbf{c}_i, \rho) : \|[\mathbf{A} \ \ \mathbf{B}]\mathbf{x}\|_2 < \rho] \tag{1.212}$$

$$\leq \mathbb{P}[\|\mathbf{A}\mathbf{u}_i + \mathbf{B}\mathbf{v}_i\|_2 < \rho(1 + r\sqrt{m} + \|\mathbf{B}\|_2)] \tag{1.213}$$

$$\leq \rho^{2m} \frac{C(p, m, r)}{\|\mathbf{u}_i\|_2^{2m}} (1 + r\sqrt{m} + \|\mathbf{B}\|_2)^{2m}, \tag{1.214}$$

where (1.213) is by (1.206)–(1.210), and in (1.214) we applied the concentration of measure result Lemma 1.21 below (recall that $\mathbf{c}_i = (\mathbf{u}_i^T \ \ \mathbf{v}_i^T)^T \in \mathcal{V}$ implies $\mathbf{u}_i \neq \mathbf{0}$) with $C(p, m, r)$ as in (1.219). Inserting (1.211)–(1.214) into (1.204) yields

$$0 \leq \liminf_{\rho \to \infty} \frac{\log(N_{\mathcal{V}}(\rho)\rho^{2m})}{\log \frac{1}{\rho}} \tag{1.215}$$

$$= \underline{\dim}_{\mathrm{B}}(\mathcal{V}) - 2m \tag{1.216}$$

$$< 0, \tag{1.217}$$

where (1.217) follows from $\underline{\dim}_{\mathrm{B}}(\mathcal{V}) \leq \underline{\dim}_{\mathrm{B}}(\mathcal{U}_i) < 2m$, which constitutes a contradiction. Therefore, (1.201) must hold. □

LEMMA 1.21 *Let* $\mathbf{A} = (\mathbf{A}_1 \dots \mathbf{A}_m)^*$ *with independent random vectors* $\mathbf{A}_i$ *uniformly distributed on* $\mathcal{B}_p(\mathbf{0}, r)$ *for* $r > 0$. *Then,*

$$\mathbb{P}[\|\mathbf{A}\mathbf{u} + \mathbf{v}\|_2 < \delta] \leq \frac{C(p,m,r)}{\|\mathbf{u}\|_2^{2m}} \delta^{2m} \tag{1.218}$$

with

$$C(p,m,r) = \left(\frac{\pi V_{p-1}(r)}{V_p(r)} \right)^m \tag{1.219}$$

for all $\mathbf{u} \in \mathbb{C}^p \backslash \{\mathbf{0}\}$, $\mathbf{v} \in \mathbb{C}^m$, *and* $\delta > 0$.

Proof Since

$$\mathbb{P}[\|\mathbf{A}\mathbf{u} + \mathbf{v}\|_2 < \delta] \leq \prod_{i=1}^{m} \mathbb{P}[|\mathbf{A}_i^* \mathbf{u} + v_i| < \delta] \tag{1.220}$$

owing to the independence of the $\mathbf{A}_i$ and as $\|\mathbf{A}\mathbf{u}+\mathbf{v}\|_2 < \delta$ implies $|\mathbf{A}_i^*\mathbf{u}+v_i| < \delta$ for $i = 1, \dots, m$, it is sufficient to show that

$$\mathbb{P}[|\mathbf{B}^* \mathbf{u} + v| < \delta] \leq \frac{D(p,r)}{\|\mathbf{u}\|_2^2} \delta^2 \tag{1.221}$$

for all $\mathbf{u} \in \mathbb{C}^p \backslash \{\mathbf{0}\}$, $v \in \mathbb{C}$, and $\delta > 0$, where the random vector $\mathbf{B}$ is uniformly distributed on $\mathcal{B}_p(\mathbf{0}, r)$ and

$$D(p,r) = \frac{\pi V_{p-1}(r)}{V_p(r)}. \tag{1.222}$$

We have

$$\begin{aligned}
\mathbb{P}[|\mathbf{B}^* \mathbf{u} + v| < \delta] &= \mathbb{P}\left[\frac{|\mathbf{B}^* \mathbf{u} + v|}{\|\mathbf{u}\|_2} < \frac{\delta}{\|\mathbf{u}\|_2} \right] && (1.223)\\
&= \mathbb{P}[|\mathbf{B}^* \mathbf{U}^* \mathbf{e}_1 + \tilde{v}| < \tilde{\delta}] && (1.224)\\
&= \mathbb{P}[|\mathbf{B}^* \mathbf{e}_1 + \tilde{v}| < \tilde{\delta}] && (1.225)\\
&= \frac{1}{V_p(r)} \int_{\mathcal{B}_p(\mathbf{0},r)} \chi_{\{b_1 : |b_1 + \tilde{v}| < \tilde{\delta}\}}(b_1) \mathrm{d}\mathbf{b} && (1.226)\\
&\leq \frac{1}{V_p(r)} \int_{|b_1+\tilde{v}| \leq \tilde{\delta}} \mathrm{d}b_1 \int_{\mathcal{B}_{p-1}(\mathbf{0},r)} \mathrm{d}(b_2 \dots b_p)^T && (1.227)\\
&= \frac{V_{p-1}(r)}{V_p(r)} \int_{|b_1+\tilde{v}| < \tilde{\delta}} \mathrm{d}b_1 && (1.228)\\
&= \frac{V_{p-1}(r)}{V_p(r)} \pi \tilde{\delta}^2 && (1.229)\\
&= \frac{\pi V_{p-1}(r)}{V_p(r) \|\mathbf{u}\|^2} \delta^2, && (1.230)
\end{aligned}$$

where the unitary matrix $\mathbf{U}$ in (1.224) has been chosen such that $\mathbf{U}(\mathbf{u}/\|\mathbf{u}\|_2) = \mathbf{e}_1 = (1\ 0\ \dots\ 0)^T \in \mathbb{C}^p$ and we set $\tilde{\delta} := \delta/\|\mathbf{u}\|_2$ and $\tilde{v} := v/\|\mathbf{u}\|_2$. Further, (1.225) follows from unitary invariance of the uniform distribution on $\mathcal{B}_p(\mathbf{0}, r)$, and in

(1.226) the factor $1/V_p(r)$ is owing to the assumption of a uniform probability density function on $\mathcal{B}_p(\mathbf{0}, r)$. $\square$

1.10 Results for $|||\cdot|||_1$ and $|||\cdot|||_2$

LEMMA 1.22 *Let $\mathbf{U} \in \mathbb{C}^{m\times m}$ be unitary, $\mathcal{P}, \mathcal{Q} \subseteq \{1, \ldots, m\}$, and consider the orthogonal projection $\mathbf{\Pi}_\mathcal{Q}(\mathbf{U}) = \mathbf{U}\mathbf{D}_\mathcal{Q}\mathbf{U}^*$ onto the subspace $\mathcal{W}^{\mathbf{U},\mathcal{Q}}$. Then,*

$$|||\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{D}_\mathcal{P}|||_2 = |||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_2. \tag{1.231}$$

Moreover, we have

$$|||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_2 = \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_\mathcal{P}\|_2}{\|\mathbf{x}\|_2} \tag{1.232}$$

and

$$|||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_1 = \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_\mathcal{P}\|_1}{\|\mathbf{x}\|_1}. \tag{1.233}$$

Proof The identity (1.231) follows from

$$\begin{aligned}
|||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_2 &= |||(\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U}))^*|||_2 &(1.234)\\
&= |||\mathbf{\Pi}^*_\mathcal{Q}(\mathbf{U})\mathbf{D}^*_\mathcal{P}|||_2 &(1.235)\\
&= |||\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{D}_\mathcal{P}|||_2, &(1.236)
\end{aligned}$$

where in (1.234) we used that $|||\cdot|||_2$ is self-adjoint [33, p. 309], $\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})^* = \mathbf{\Pi}_\mathcal{Q}(\mathbf{U})$, and $\mathbf{D}^*_\mathcal{P} = \mathbf{D}_\mathcal{P}$. To establish (1.232), we note that

$$\begin{aligned}
|||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_2 &= \max_{\mathbf{x}:\,\|\mathbf{x}\|_2=1} \|\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2 &(1.237)\\
&= \max_{\substack{\mathbf{x}:\,\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\neq\mathbf{0}\\ \|\mathbf{x}\|_2=1}} \|\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2 &(1.238)\\
&\le \max_{\substack{\mathbf{x}:\,\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\neq\mathbf{0}\\ \|\mathbf{x}\|_2=1}} \left\|\mathbf{D}_\mathcal{P}\frac{\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}}{\|\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2}\right\|_2 &(1.239)\\
&\le \max_{\mathbf{x}:\,\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\neq\mathbf{0}} \left\|\mathbf{D}_\mathcal{P}\frac{\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}}{\|\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2}\right\|_2 &(1.240)\\
&= \max_{\mathbf{x}\in\mathcal{W}^{\mathbf{U},\mathcal{Q}}\setminus\{\mathbf{0}\}} \frac{\|\mathbf{x}_\mathcal{P}\|_2}{\|\mathbf{x}\|_2} &(1.241)\\
&= \max_{\mathbf{x}:\,\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\neq\mathbf{0}} \left\|\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\frac{\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}}{\|\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2}\right\|_2 &(1.242)\\
&\le \max_{\mathbf{x}:\,\|\mathbf{x}\|_2=1} \|\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2 &(1.243)\\
&= |||\mathbf{D}_\mathcal{P}\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})|||_2, &(1.244)
\end{aligned}$$

where in (1.239) we used $\|\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2 \le \|\mathbf{x}\|_2$, which implies $\|\mathbf{\Pi}_\mathcal{Q}(\mathbf{U})\mathbf{x}\|_2 \le 1$

for all $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$. Finally, (1.233) follows by repeating the steps in (1.237)–(1.244) with $\|\cdot\|_2$ replaced by $\|\cdot\|_1$ at all occurrences. □

LEMMA 1.23 *Let $\mathbf{A} \in \mathbb{C}^{m\times n}$. Then,*

$$\frac{\|\mathbf{A}\|_2}{\sqrt{\mathrm{rank}(\mathbf{A})}} \leq |||\mathbf{A}|||_2 \leq \|\mathbf{A}\|_2. \tag{1.245}$$

Proof The proof is trivial for $\mathbf{A} = \mathbf{0}$. If $\mathbf{A} \neq \mathbf{0}$, set $r = \mathrm{rank}(\mathbf{A})$ and let $\sigma_1, \ldots, \sigma_r$ denote the nonzero singular values of $\mathbf{A}$ organized in decreasing order. Unitary invariance of $|||\cdot|||_2$ and $\|\cdot\|_2$ (cf. [33, Problem 5, p. 311]) yields $|||\mathbf{A}|||_2 = \sigma_1$ and $\|\mathbf{A}\|_2 = \sqrt{\sum_{i=1}^r \sigma_i^2}$. The claim now follows from

$$\sigma_1 \leq \sqrt{\sum_{i=1}^{r} \sigma_i^2} \leq \sqrt{r}\sigma_1. \tag{1.246}$$

□

LEMMA 1.24 *For $\mathbf{A} = (\mathbf{a}_1 \ldots \mathbf{a}_n) \in \mathbb{C}^{m\times n}$, we have*

$$|||\mathbf{A}|||_1 = \max_{j\in\{1,\ldots,n\}} \|\mathbf{a}_j\|_1 \tag{1.247}$$

and

$$\frac{1}{n}\|\mathbf{A}\|_1 \leq |||\mathbf{A}|||_1 \leq \|\mathbf{A}\|_1. \tag{1.248}$$

Proof The identity (1.247) is established in [33, p.294], and (1.248) follows directly from (1.247). □

References

[1] W. Heisenberg, *The Physical Principles of the Quantum Theory.* Chicago, IL: Chicago Press, 1930.

[2] W. G. Faris, "Inequalities and uncertainty principles," *J. Math. Phys.*, vol. 19, no. 2, pp. 461–466, Jan. 1978.

[3] M. G. Cowling and J. F. Price, "Bandwidth versus time concentration: The Heisenberg-Pauli-Weyl inequality," *SIAM J. Math. Anal.*, vol. 15, no. 1, pp. 151–165, Jan. 1984.

[4] J. J. Benedetto, *Wavelets: Mathematics and Applications.* New York, NY: CRC Press, 1994, ch. Frame decompositions, sampling, and uncertainty principle inequalities.

[5] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," *J. Fourier Anal. Appl.*, vol. 3, no. 3, pp. 207–238, 1997.

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

[7] D. L. Donoho and B. F. Logan, "Signal recovery and the large sieve," *SIAM J. Appl. Math.*, vol. 52, no. 2, pp. 577–591, Apr. 1992.

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

[9] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[10] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[11] A. Terras, *Fourier Analysis on Finite Groups and Applications.* New York, NY: Cambridge Univ. Press, 1999.

[12] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Math. Res. Lett.*, vol. 12, no. 1, pp. 121–127, 2005.

[13] D. Stotz, E. Riegler, E. Agustsson, and H. Bölcskei, "Almost lossless analog signal separation and probabilistic uncertainty relations," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5445–5460, Sep. 2017.

[14] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing.* Basel, Switzerland: Birkhäuser, 2013.

[15] K. Gröchenig, *Foundations of Time-Frequency Analysis.* Birkhäuser, 2001.

[16] D. Gabor, "Theory of communications," *J. Inst. Elec. Eng.*, vol. 96, pp. 429–457, 1946.

[17] H. Bölcskei, *Advances in Gabor analysis.* Birkhäuser, 2003, ch. Orthogonal frequency division multiplexing based on offset QAM.

[18] C. Fefferman, "The uncertainty principle," *Bull. Amer. Math. Soc.*, vol. 9, no. 2, pp. 129–206, Sep. 1983.

[19] S. Evra, E. Kowalski, and A. Lubotzky, "Good cyclic codes and the uncertainty principle," *arXiv:1703.01080*, 2017.

[20] E. Bombieri, *Le grand crible dans la théorie analytique des nombres.* Paris, France: Soc. Math. de France, 1974.

[21] H. L. Montgomery, *Twentieth Century Harmonic Analysis – A Celebration.* Dordrecht, Netherlands: Springer, 2001, ch. Harmonic analysis as found in analytic number theory.

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

[23] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[24] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Sig. Process.*, vol. 58, no. 6, Jun. 2010.

[25] E. J. Candès and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 4, no. 57, pp. 2342–2359, Apr. 2011.

[26] G. Alberti, H. Bölcskei, C. De Lellis, G. Koliander, and E. Riegler, "Lossless analog compression," *submitted to IEEE Trans. Inf. Theory, arXiv:1803.06887*, 2018.

[27] E. Riegler, D. Stotz, and H. Bölcskei, "Information-theoretic limits of matrix completion," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 1836–1840.

[28] A. J. Lzenman, "Introduction to manifold learning," *WIREs Comput. Stat.*, vol. 4, pp. 439–446, 2012.

[29] H. Lu, Y. Fainman, and R. Hecht-Nielsen, "Image manifolds," in *Proc. SPIE*, Apr. 1998, pp. 3307–3307.

[30] N. Sochen and Y. Y. Zeevi, "Representation of colored images by manifolds embedded in higher dimensional non-Euclidean space," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1998, pp. 166–170.

[31] G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 65–74, Jan. 1997.

[32] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge, U.K.: Cambridge Univ. Press, 1990.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.

[35] J. S. Abel and J. O. Smith III, "Restoring a clipped signal," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, Apr. 1991, pp. 1745–1748.

[36] S. G. Mallat and G. Yu, "Super-resolution with sparse mixing estimators," *IEEE Trans. Image Proc.*, vol. 19, no. 11, pp. 2889–2900, Nov. 2010.

[37] M. Elad and Y. Hel-Or, "Fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Proc.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

[38] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. of 27th Ann. Conf. Comp. Graph. Int. Tech.*, Jul. 2000, pp. 417–424.

[39] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, pp. 340–358, Nov. 2005.

[40] D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem," *Commun. Pure and Appl. Math.*, vol. 66, no. 1, pp. 1–47, Jan. 2013.

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

[42] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[43] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3104–3116, May 2013.

[44] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[45] J. A. Tropp, "On the linear independence of spikes and sines," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 838–858, 2008.

[46] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *J. R. Statist. Soc. B.*, vol. 54, no. 1, pp. 41–81, 1992.

[47] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, May 1996, pp. 1688–1691.

[48] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Areas Commun.*, vol. 4, no. 2, pp. 375–391, Feb. 2010.

[49] R. Heckel and H. Bölcskei, "Identification of sparse linear operators," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7985–8000, Dec. 2013.

[50] K. Falconer, *Fractal Geometry*, 1st ed. New York, NY: Wiley, 1990.

[51] J. D. Vaaler, "Some extremal functions in Fourier analysis," *Bull. Amer. Math. Soc.*, vol. 2, no. 12, pp. 183–216, 1985.

[52] C. Heil, *A Basis Theory Primer.* New York, NY: Springer, 2011.

[53] P. Mattila, *Geometry of Sets and Measures in Euclidean Space: Fractals and Rectifiability.* Cambridge, UK: Cambridge Univ. Press, 1999.