

Master's Mathematics
Track: Biomedical Mathematics

Internship report

Compressive sensing for multimode fiber optical imaging

by

Emma Erkočević

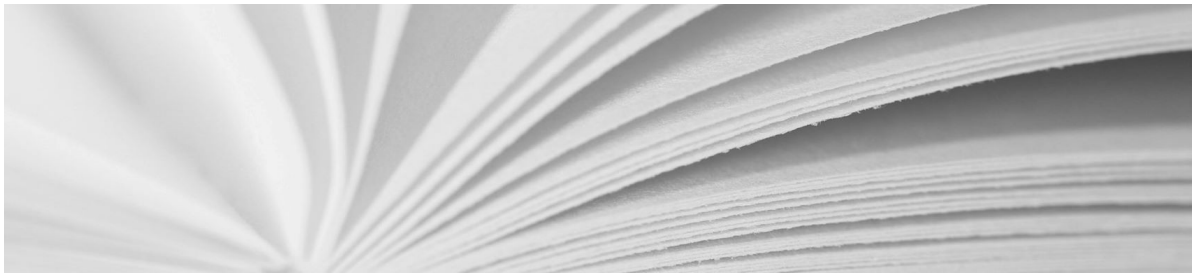
April 3, 2025

First supervisor: dr. Svetlana Dubinkina

Second reader: prof.dr. Tristan van Leeuwen

External Supervisor: prof.dr. Tristan van Leeuwen and
dr. Lyuba Amitonova

Host Organization: CWI Amsterdam



Department of Mathematics
Faculty of Science



Summary

Title: Compressive sensing for multimode fiber optical imaging

Author: Emma Erkočević, e.erkocevic@student.vu.nl, 2664982

First supervisor: dr. Svetlana Dubinkina

Second reader: prof.dr. Tristan van Leeuwen

External Supervisor: prof.dr. Tristan van Leeuwen and
dr. Lyuba Amitonova

Host Organization: CWI Amsterdam

Date: April 3, 2025

Department of Mathematics

Vrije Universiteit Amsterdam

de Boelelaan 1111, 1081 HV Amsterdam

<http://www.math.vu.nl/>

Contents

1. Introduction	4
2. Background	6
2.1. Optical experiment	6
2.2. Fundamentals of compressive sensing theory	8
2.2.1. Problem statement	8
2.2.2. Matrix properties	10
2.2.3. Sparse recovery	13
2.2.4. Reconstruction guarantees	14
3. Effect of column-wise correlation in the measurement matrix	16
3.1. Model description	16
3.2. Theoretical reconstruction guarantee	17
3.3. Numerical results	19
3.3.1. Correlation effect across signal classes	21
3.4. Two-dimensional setting	23
4. Non-negative measurement matrices	26
4.1. Random exponential matrices	26
4.1.1. ℓ_2 -robust nullspace property	26
4.1.2. \mathcal{M}^+ -criterion	26
4.2. Other distributions	27
4.2.1. Model description	28
4.2.2. Numerical experiment	30
5. Conclusion	31
Bibliography	31
A. Appendix	35

1. Introduction

In 2018, a novel optical microscopy technique called multimode fiber compressive imaging (MMFCI) was developed at the Advanced Research Center for Nanolithography (ARCNL) [7]. The technique is based on light propagation through ultra-thin glass fibers (Figure 1.1) and enables imaging without lenses by utilizing a phenomenon of light scattering. A laser illuminates one end of the fiber, which transports the light while scrambling it, such that at the other end a random *speckle pattern* emerges (Figure 1.2). Experiments show that by illuminating a sample with such patterns, a high-resolution image of the sample can be reconstructed from remarkably few illuminations. The method has large potential in endoscopy where the fiber acts both as a compact, minimally invasive probe and as the imaging device [2]. Furthermore, the high speed and resolution of MMFCI make it a promising approach for visualizing rapid interactions at a fine scale, such as in the brains of freely behaving organisms [6].

This type of image recovery from a small number of measurements which have a random structure fits into the framework of compressive sensing (CS). Mathematically, CS is an inverse problem that retrieves unique solutions to underdetermined linear systems by assuming the solution contains many zero elements and the system exhibits randomness. Hereto it typically employs recovery schemes based on the ℓ_1 -norm. Since its introduction in 2004, extensive theory has been developed on what matrix properties and optimization techniques are required, primarily based on linear algebra, random matrix theory and convex analysis [16]. From a signal processing perspective, compressed sensing provided a new sampling paradigm, allowing signal reconstruction from less data than conventionally deemed necessary [12]. The method has been applied to various imaging modalities, such as MRI, where the lower sampling rate led to a 32% reduction in scanning time [4]. The mathematical theory of CS in relation to MMFCI is still an open area of study.

This thesis interprets the multimode fiber experiment in the mathematical framework of compressive sensing. Using theoretical guarantees and numerical simulations, our aim is to verify that the experiment aligns with CS theory and to explore possible improvements. We focus on three main questions. First, we analyze to which extent the smooth structure present in speckle patterns – as shown in Figure 1.2 – affects the reconstruction quality, since CS theory assumes that there is no correlation within individual measurements. Next, motivated by the ability to experimentally customize the probability distribution of speckle light intensity [10], we study which distribution is most suitable. To conclude, we inquire whether there are reconstruction schemes tailored to the experimental set-up that are more efficient than conventional ℓ_1 -norm-based approaches currently employed by the experiment.

In Chapter 2, we describe the existing knowledge that underlies our work. Section 2.1 is dedicated to the optical experiment, providing a brief explanation of multimode fibers and speckle patterns, as well as the experiment’s physical set-up and mathematical formulation. Section 2.2 covers the fundamentals of compressive sensing theory, focusing on measurement properties, recovery schemes and guarantees relevant to our application. In Chapter 3, we devise a model to study the effect of the speckle patterns’ smooth structure, analytically derive an upper bound on the amount of correlation, and compare this bound with numerical simulations. Chapter 4 adapts the model to follow different probability distributions, examines which distribution is preferable, and explores the use of alternative reconstruction schemes. Chapter 5 summarizes our findings and suggests directions for further research.

This project was carried out as an internship at Centrum Wiskunde & Informatica (CWI), the national research institute for mathematics and computer science in the Netherlands. I was a trainee in the Computational Imaging group, where aspects of mathematics, computer science and physics are combined to develop the next generation of imaging [1]. Moreover, the project was in collaboration with ARCNL’s Nanoscale Imaging and Metrology group, which developed MMFCI and expects new technological insights to be based on computational algorithms [6]. In addition to regular meetings with university supervisor dr. Svetlana Dubinkina and internship supervisor prof.dr. Tristan van Leeuwen, monthly meetings took place with dr. Lyuba Amitonova of ARCNL to ensure this work would have experimental value as well.

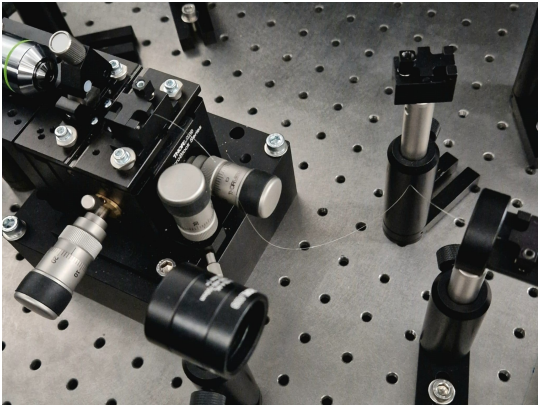


Figure 1.1.: Glass fiber in the ARCNL laboratory.

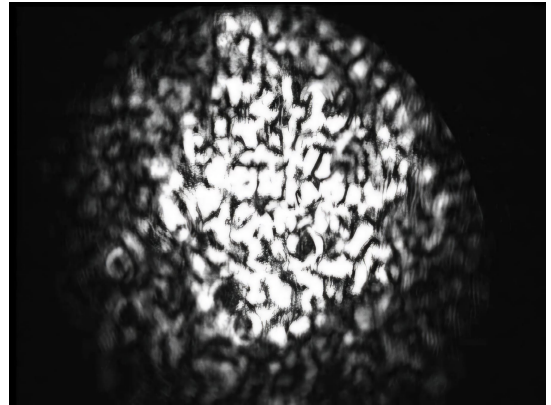


Figure 1.2.: Example of a speckle pattern [17].

2. Background

2.1. Optical experiment

An optical fiber is a transporter of light, consisting of a core surrounded by a cladding. Traveling light remains within the fiber because the core and cladding have different refractive indices, i.e. different measures of bending light rays, so that waves arriving at the core-cladding boundary are reflected back into the core. This phenomenon is referred to as total internal reflection [26]. For the experiment, a particular class of fibers called multimode fibers (MMFs) is used, where the core contains multiple paths – modes – for light to propagate, as shown in Figure 2.1(a). The number of modes depends on properties of the fiber such as its radius and refractive index, as well as the light’s wavelength. At the fiber’s cross-section, each mode has a light intensity profile as illustrated in Figure 2.1(b), which is constant along the fiber’s propagation axis and orthogonal to the profiles of the other modes [27].

The formation of scattered light is a consequence of the different modes in a multimode fiber interfering with each other. Essentially, a point in a speckle pattern is the result of a random walk made up of independently phased complex components [17]. In other words, the elementary mathematical representation of a single speckle point (x, y) at the fiber output is

$$E_{\text{out}}(x, y) = \sum_{k=1}^M E_k e^{j\phi_k} \Psi_k(x, y), \quad (2.1)$$

where j denotes the imaginary unit, each term represents a mode, each phase ϕ_k the mode’s propagation constant, and each $\Psi_k(x, y)$ the mode’s intensity profile. Moreover, the expression of E_k is rather involved and outside the scope of this thesis. Depending on the position of the laser beam at the fiber input, a different speckle pattern emerges, as displayed in Figure 2.1(c). The intensity of the input beam usually follows a Gaussian distribution and that of the output speckle an exponential distribution, with intensities computed by squaring the amplitude of the complex field [25].

It is possible with a method called wavefront shaping to reverse the mode scrambling and obtain focused output spots. However, this approach requires lenses, making the set-up bulky. Additionally, such imaging – called raster-scan imaging – requires as many spots as the number of pixels in the final image, whereas compressive imaging achieves reconstruction with significantly fewer measurements. Moreover, MMFCI overcomes the diffraction limit $d = \lambda/(2\text{NA})$, a common constraint on the smallest resolvable feature size d in terms of the light’s wavelength λ and the fiber’s numerical aperture NA – a dimensionless quantity describing the range of angles at which light can enter the fiber.

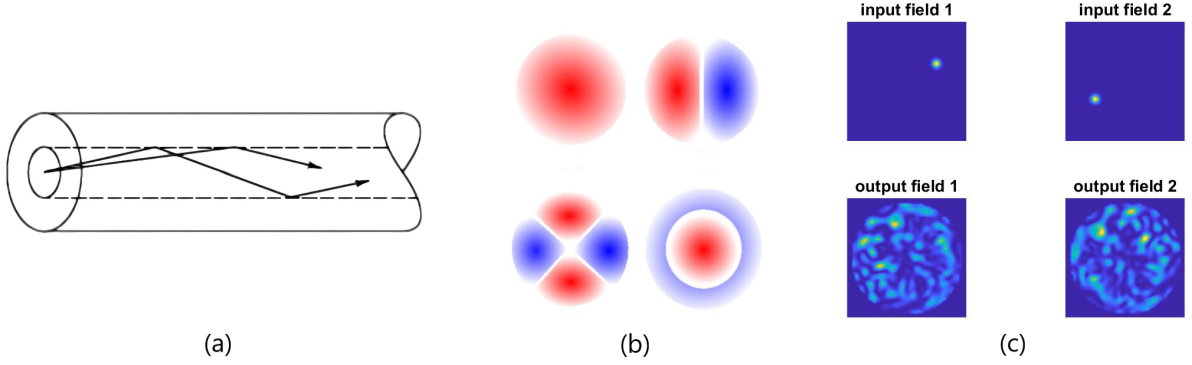


Figure 2.1.: (a) Simple illustration of a multimode fiber with two modes [32]. (b) Four mode intensity profiles at a fiber's cross-section [19]. (c) Speckle from two input fields (MATLAB simulation of ARCNL).

Approaches referred to as super-resolution microscopy have surpassed this limit as well, but they are rather slow. To summarize, the MMFCI design is compact and performs fast at a high spatial resolution. For example, it reconstructed a 400-by-400 pixel image of the sample in Figure 2.2(a) from 961 measurements, with a spatial resolution 2.5 times better than the diffraction limit and three orders of magnitude faster than state-of-the-art diffraction limited imaging [2].

Figure 2.2(b) presents an overview of the experimental set-up, describing the components and measurement protocol. The mathematical formulation is as follows. Say m speckle patterns of n -by- n pixels are measured, then these patterns are flattened and stored as rows in a matrix $\mathbf{A} \in \mathbb{R}^{m \times n^2}$. Similarly, the intensity measurements are stored in a vector $\mathbf{y} \in \mathbb{R}^m$. The problem is to reconstruct a flattened version $\mathbf{x} \in \mathbb{R}^{n^2}$ of the sample from the system

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \quad (2.2)$$

where $\mathbf{e} \in \mathbb{R}^m$ denotes a noise vector. Compressive sensing enables recovery from few measurements ($m \ll n^2$) by assuming that \mathbf{x} is sparse, meaning it contains only a small number s of non-zeros. The reconstruction is carried out by minimizing the ℓ_1 -norm of \mathbf{x} subject to (2.1), a standard approach to enforce sparsity. Lastly, \mathbf{x} is reshaped to form an n -by- n image.

To conclude, we note that compressed sensing overcomes an essential principle in signal processing called the Nyquist sampling rate, which states that accurate signal recovery requires a sampling frequency of at least twice the highest frequency present in the signal [16]. Conventional sensing approaches sample according to Nyquist, after which the data often is converted to a compressed form [21]. Compressive sensing is based on the idea that data can be acquired and compressed simultaneously, employing a rule of thumb given by $m \gtrsim 2s \log(n^2/m)$ [15]. Indeed, the experiment of Figure 2.2(a) obeys this rule, as $m = 961$, $n^2 = 160000$ and the disks occupy $s = 208$ non-zero pixels. The following section studies the mathematical theory of CS underlying our analysis.

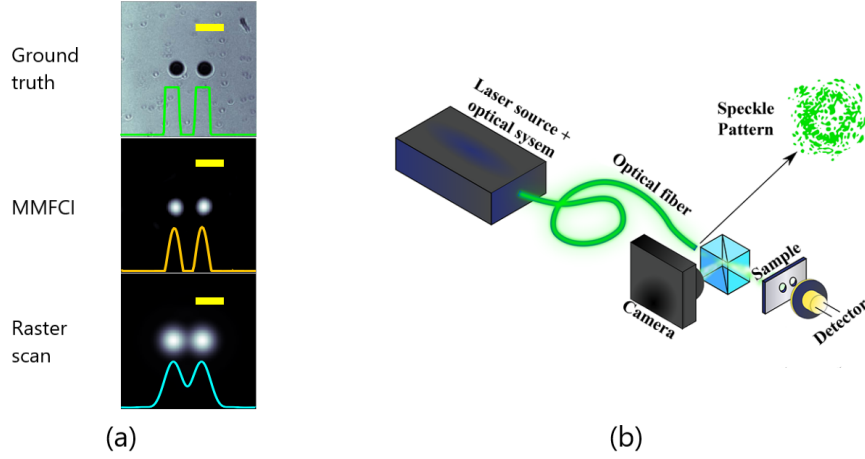


Figure 2.2.: (a) Experimental reconstruction with MMFCI of a 1.9-by-1.9 mm sample with two disks of diameter $\approx 54.1 \mu\text{m}$, along with a raster-scan diffraction-limited reconstruction [2]. (b) Basic schematic overview of the experimental set-up, consisting of the following components [24]: a laser illuminating the fiber at a specific position; an MMF that propagates and scrambles the light; a camera recording the speckle pattern before the light reaches the sample; a two-dimensional sample, typically a glass plate with two round beads; a detector which registers the total amount of light behind the sample, ignoring spatial information. The fiber is illuminated sequentially at different locations on its input and for every location, the speckle pattern and overall light intensity are registered.

2.2. Fundamentals of compressive sensing theory

2.2.1. Problem statement

Let us first outline the basic architecture of a compressive sensing problem. Consider the system of linear equations

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.3)$$

where $\mathbf{x} \in \mathbb{R}^N$ represents a signal, $\mathbf{y} \in \mathbb{R}^m$ measurements and $\mathbf{A} \in \mathbb{R}^{m \times N}$ is the forward operator – typically called the measurement matrix. The compressive sensing problem is an inverse problem that aims to reconstruct \mathbf{x} from \mathbf{y} in the case where the number of measurements is significantly smaller than the signal length, i.e. $m \ll N$. For the problem to be well-posed in the sense of Hadamard, there must exist a unique solution \mathbf{x} that depends continuously on the data \mathbf{y} , such that the solution is stable and not significantly changed by small perturbations in the data [23]. Without any additional information, (2.3) describes an underdetermined linear system, which is ill-posed with infinitely many solutions. So, to ensure a unique solution that can be stably recovered, compressive sensing is built on on two essential assumptions: *sparsity* of the signal and *incoherence* of the measurement matrix.

Formally, a signal \mathbf{x} is called s -sparse if it has at most s non-zeros. Using the ℓ_0 -“norm” defined as $\|\mathbf{x}\|_0 := \#\{i : x_i \neq 0\}$, one writes $\|\mathbf{x}\|_0 \leq s$ [5]. The collection of all s -sparse signals is represented by $\Sigma_s := \{\mathbf{x} : \|\mathbf{x}\|_0 \leq s\}$. Additionally, a signal that is not truly sparse but can be well approximated by a sparse signal is referred to as compressible, and its sparse approximation – \mathbf{x}_s – is typically obtained by setting all but the s largest entries to zero [12]. Many signals do not immediately appear sparse but have a sparse representation when transformed to an appropriate basis. The system then takes the form $\mathbf{y} = \Phi \mathbf{f}$ with Φ the measurement matrix and \mathbf{f} the original non-sparse signal. With \mathbf{x} denoting the sparse representation of \mathbf{f} in some basis Ψ , one writes $\mathbf{f} = \Psi \mathbf{x}$ and reformulates the problem as (2.3) with $\mathbf{A} = \Phi \Psi$. In other words, the signal is transformed to a sparse domain, reconstructed in this domain, and transformed back to the original domain. A widely used sparsifying transformation for natural images is the discrete wavelet transform (DWT). This thesis, however, focuses on signals that are sparse in the original domain.

With (2.3) restricted to $\mathbf{x} \in \Sigma_s$, a solution is unique provided that \mathbf{A} maps s -sparse vectors injectively. In other words, any subcollection of $2s$ columns of \mathbf{A} must be linearly independent. To ensure stability in the sense of Hadamard besides uniqueness, an extra dissimilarity measure on the columns of \mathbf{A} is required [8]. The following definition provides such a measure:

Definition 2.1. The *mutual coherence* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is the largest absolute inner product between any of its two columns $\mathbf{a}_i, \mathbf{a}_j$:

$$\mu(\mathbf{A}) := \max_{1 \leq i < j \leq N} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2}.$$

Here the value of μ represents the extent to which the matrix’s columns are correlated, with the lowest attainable value given by the Welch bound $\sqrt{(N-m)/(m(N-1))}$. Generally, low values indicate that submatrices formed by columns of \mathbf{A} are well-conditioned, implying stable recovery [33]. Although mutual coherence is convenient due to its computational simplicity, reconstruction guarantees directly based on it tend to have large gaps between theoretical bounds and numerical results (cf. Theorem 3.1 in [14]). Therefore, we will resort to other matrix properties such as the restricted isometry property and ℓ_2 -robust nullspace property, discussed in Section 2.2.2.

Measurement matrices often achieve incoherence by exhibiting randomness. Consequently, many of the proof techniques that we will encounter are based on notions from random matrix theory. A random matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$ is a matrix whose entries are randomly sampled from a probability distribution. More formally, given a probability measure space (Ω, ρ) and a random variable R on Ω , one can form a random matrix $\mathbf{X}(\omega)$ with $\omega \in \Omega^{mN}$ by selecting each entry \mathbf{X}_{ij} as a realization of R [9]. In this thesis, the term random matrix will imply independent and identically distributed entries, i.e. a Gaussian matrix will consist of i.i.d. entries $\mathbf{X}_{ij} \sim \mathcal{N}(\mu, \sigma^2)$, a 0/1-Bernoulli matrix of i.i.d. $\mathbf{X}_{ij} \sim \text{Ber}(p)$, an exponential matrix of i.i.d. $\mathbf{X}_{ij} \sim \text{Exp}(\lambda)$, and so on.

Many properties of random variables extend naturally to random matrices. For instance, a random matrix's expected value is defined as the matrix of expected values, i.e. $[\mathbb{E}(\mathbf{X})]_{ij} := \mathbb{E}(\mathbf{X}_{ij})$, and hence linearity of the expectation operator also holds for random matrices [30]. Furthermore, many proof techniques use concentration of measure inequalities (also referred to as Bernstein, Hoeffding or Chernoff inequalities), which bound the probability that a linear combination of the matrix's entries deviates from the matrix's mean, as well as union bounds ($P(\cup_{k=1}^{\infty} B_k) \leq \sum_{k=1}^{\infty} P(B_k)$ with $\{B_k\}_{k=1}^{\infty}$ a countable set of events) [16]. The concentration inequalities commonly rely on properties of the probability distribution such as the tails decaying at least as fast as those of a Gaussian distribution (sub-Gaussian random variable) or an exponential distribution (subexponential random variable).

To conclude, we state the two fundamental questions of compressive sensing which will be addressed repeatedly throughout the thesis. First, what sensing matrices are suitable and how are they designed? Second, how can the true signal be recovered and what are efficient techniques?

2.2.2. Matrix properties

This section first discusses the restricted isometry property, a typically used criterion that ensures each submatrix of $2s$ columns of \mathbf{A} is well-conditioned, and central to our analysis in Chapter 3. Next, we turn to the ℓ_2 -robust nullspace property, a weaker condition that is sufficient when \mathbf{A} consists of nonnegative entries, as will be the case in Chapter 4.

Restricted isometry property

Definition 2.2. A matrix \mathbf{A} satisfies the *restricted isometry property* (RIP) of order s if there exists a $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s) \|\mathbf{x}\|_2^2 \quad (2.4)$$

holds for all $\mathbf{x} \in \Sigma_s$.

The definition effectively says that \mathbf{A} acts as a near-isometry on sparse vectors, i.e. if \mathbf{A} has the RIP of order $2s$, it approximately preserves (up to δ_{2s}) the distance between any pair of s -sparse vectors [8]. The smaller the value of δ_s , the better \mathbf{A} is at preserving distances; consequently, recovery guarantees usually involve an upper bound on δ_s . An alternative interpretation of (2.4) is that each Gram matrix $\mathbf{A}_S^T \mathbf{A}_S$ with \mathbf{A}_S the submatrix made up of s columns of \mathbf{A} indexed by $S \subset \{1, \dots, N\}$ has all eigenvalues in the interval $[1 - \delta_s, 1 + \delta_s]$. Furthermore, the bounds in (2.4) do not necessarily need to be symmetric as a matrix satisfies the RIP whenever a scaling of the matrix exists that has the RIP [13].

Generally, the RIP is challenging to verify as it amounts to checking (2.4) for $\binom{N}{s}$ vectors. For certain random matrices (Gaussian matrices in particular), however, neat proofs exist which omit the computational complexity of the problem by instead using standard tools from random matrix theory such as concentration of measure inequalities and union bounds [9]. Moreover, shortcuts have been devised to demonstrate the RIP for deterministic matrices [8].

RIP for random matrices We now outline the classic proof of the RIP for random matrices [9]. The main result states that for given m, N and $0 < \delta < 1$, matrices \mathbf{A} that have $\mathbb{E}(\|\mathbf{Ax}\|_2^2) = \|\mathbf{x}\|_2^2$ and

$$\mathbb{P}(|\|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2| \geq \epsilon \|\mathbf{x}\|_2^2) \leq 2e^{-mc_0(\epsilon)} \quad (2.5)$$

with $c_0(\epsilon) > 0$, $0 < \epsilon < 1$ and $\mathbf{x} \in \mathbb{R}^N$, satisfy with high probability (w.h.p.) the RIP of any order $s \leq c_1 m / \log(N/s)$, where the constant $c_1 > 0$ is independent of s, m and N . The expectation is computed as $\mathbb{E}(\|\mathbf{Ax}\|_2^2) = \mathbf{x}^T \mathbb{E}(\mathbf{A}^T \mathbf{A}) \mathbf{x}$ with $\mathbb{E}(\mathbf{A}^T \mathbf{A})$ the matrix formed by the means of the entries of $\mathbf{A}^T \mathbf{A}$, and (2.5) is referred to as a concentration of measure inequality. Thus, random matrices that have $\|\mathbf{Ax}\|_2^2$ strongly concentrated around their mean $\|\mathbf{x}\|_2^2$ are bound to satisfy the RIP.

A chief example of matrices obeying (2.5) are those composed of i.i.d. entries $a_{ij} \sim \mathcal{N}(0, \frac{1}{m})$ [3]. For now, we briefly verify that $\mathbb{E}(\|\mathbf{Ax}\|_2^2) = \|\mathbf{x}\|_2^2$, which amounts to checking $\mathbb{E}(\mathbf{A}^T \mathbf{A}) = \mathbf{I}_N$. For any entry (k, l) of $\mathbb{E}(\mathbf{A}^T \mathbf{A})$, it holds that

$$\mathbb{E}([\mathbf{A}^T \mathbf{A}]_{kl}) = \sum_{i=1}^m \mathbb{E}(a_{ik} a_{il}) = \begin{cases} \sum_{i=1}^m \text{Var}(a_{ik}) + \mathbb{E}(a_{ik})^2 & \text{for } k = l; \\ \sum_{i=1}^m \mathbb{E}(a_{ik}) \mathbb{E}(a_{il}) & \text{for } k \neq l, \end{cases} \quad (2.6)$$

and so the diagonal ($k = l$) is given by $\sum_{i=1}^m 1/m = 1$ and the off-diagonal ($k \neq l$) by 0. Furthermore, (2.6) shows that general random matrices with i.i.d. entries have $\mathbb{E}(\|\mathbf{Ax}\|_2^2) = \|\mathbf{x}\|_2^2$ only if the distribution has zero mean, an important observation as we turn to non-negative random matrices in Chapter 4.

The structure of the proof in [9] is as follows. First of all, since \mathbf{A} is linear, it suffices to consider only unit vectors \mathbf{x} . The RIP is then proven in two steps: (i) for fixed s and (ii) for all $s \leq c_1 m / \log(N/s)$. More specifically, a lemma establishes that (2.4) holds w.h.p. on any fixed s -dimensional unit sphere X_S comprised of all vectors in \mathbb{R}^N that are zero outside of the s positions $S \subset \{1, \dots, N\}$, starting from a finite covering Q_S and then extending to the entire sphere, as illustrated on the left in Figure 2.3. A theorem, in turn, extends the result to the union Σ_s of s -dimensional unit spheres in \mathbb{R}^N , depicted on the right in Figure 2.3. This way, the RIP is shown for all s -sparse \mathbf{x} without the computational difficulty. In Appendix A, we present the lemma and theorem and offer insights into their proofs.

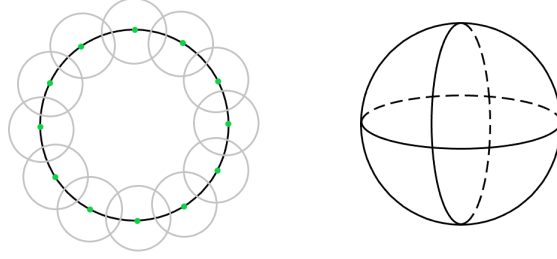


Figure 2.3.: (Left) s -dimensional unit sphere X_S , along with its covering Q_S in green; (right) union of s -dimensional unit spheres Σ_s .

RIP for deterministic matrices In the following, we briefly describe a standard procedure in CS theory to show the RIP for deterministic matrices [8]. Given some deterministic matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ and sparsity level s , the goal is to find an upper bound on the constant δ_s . Hereto a classic bound on the eigenvalues of a square matrix, the Gershgorin circle theorem, is invoked:

Theorem 2.3. *For each eigenvalue λ of a matrix $\mathbf{B} \in \mathbb{R}^{s \times s}$, there is an index $i \in \{1, \dots, s\}$ such that $|\lambda - \mathbf{B}_{ii}| \leq \sum_{j=1, j \neq i}^s |\mathbf{B}_{ij}|$.*

It is assumed w.l.o.g. that \mathbf{A} has unit-norm columns, so that the Gram matrix $\mathbf{A}_S^T \mathbf{A}_S$ of s columns indexed by $S \subset \{1, \dots, N\}$ has a unit diagonal. Moreover, the off-diagonal of $\mathbf{A}_S^T \mathbf{A}_S$ is bounded from above by $\mu(\mathbf{A})$. From the Gershgorin circle theorem with $\mathbf{B} := \mathbf{A}_S^T \mathbf{A}_S - \mathbf{I}_s$ it follows that every eigenvalue λ of $\mathbf{A}_S^T \mathbf{A}_S - \mathbf{I}_s$ obeys

$$|\lambda| \leq \sum_{j=1}^s \mu(\mathbf{A}) = (s-1)\mu(\mathbf{A})$$

regardless of the choice of S . Consequently, all eigenvalues of $\mathbf{A}_S^T \mathbf{A}_S$ lie in $1 \pm (s-1)\mu(\mathbf{A})$, which implies that $\delta_s \leq (s-1)\mu(\mathbf{A})$. In other words, deterministic matrices \mathbf{A} are $(s, (s-1)\mu(\mathbf{A}))$ -RIP, a result that will be used in Section 3.2.

ℓ_2 -robust nullspace property

In the case of nonnegative random matrices, the individual entries do not have a zero mean as required by the RIP. In this scenario, reconstruction guarantees typically use the ℓ_2 -robust nullspace property (ℓ_2 -robust NSP). For a vector $\mathbf{v} \in \mathbb{R}^N$ and a collection of indices $S \subset \{1, \dots, N\}$, the property considers the restriction of \mathbf{v} to S – defined as $(\mathbf{v}_S)_i := \mathbf{v}_i \mathbb{1}_{i \in S}$ – and the restriction of \mathbf{v} to $\bar{S} = \{1, \dots, N\} \setminus S$:

Definition 2.4. A matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies the ℓ_2 -robust nullspace property of order s with parameters $\rho \in (0, 1)$ and $\tau > 0$ if

$$\|\mathbf{v}_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|\mathbf{v}_{\bar{S}}\|_1 + \tau \|\mathbf{A}\mathbf{v}\|_2 \quad \forall \mathbf{v} \in \mathbb{R}^N \quad (2.7)$$

holds for all $S \subset \{1, \dots, N\}$ with $|S| \leq s$.

The procedure of showing that a matrix has the ℓ_2 -robust NSP is as follows. One may assume w.l.o.g. $\|\mathbf{v}\|_2 = 1$ as (2.7) is scale-invariant, and consider only the support set S_{\max} consisting of the s largest (in modulus) entries of \mathbf{v} [22]. In addition, since vectors \mathbf{v} with $\|\mathbf{v}_{S_{\max}}\|_2 \leq \frac{\rho}{\sqrt{s}}\|\mathbf{v}_{\bar{S}_{\max}}\|_1$ fulfill (2.7) by default, it suffices to consider only the collection $T_{\rho,s} := \{\mathbf{v} \in \mathbb{R}^N : \|\mathbf{v}_{S_{\max}}\|_2 > \frac{\rho}{\sqrt{s}}\|\mathbf{v}_{\bar{S}_{\max}}\|_1\}$. In other words, the problem boils down to establishing that

$$\inf_{\mathbf{v} \in T_{\rho,s}} \|\mathbf{A}\mathbf{v}\|_2 > \frac{1}{\tau}. \quad (2.8)$$

This inequality is commonly verified with Mendelson's small ball method, a general purpose tool for establishing lower bounds on infima, and whose main requirement is row-wise independence of the matrix.

In general, Mendelson's small ball method is defined as follows:

Theorem 2.5. Fix $E \subset \mathbb{R}^N$ and let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be independent copies of a random vector $\mathbf{a} \in \mathbb{R}^N$. Set $\mathbf{h} = \frac{1}{\sqrt{m}} \sum_{k=1}^m \epsilon_k \mathbf{a}_k$, where $\epsilon_1, \dots, \epsilon_m$ is a Rademacher sequence. For $\xi > 0$, define $Q_\xi(E, \mathbf{a}) = \inf_{\mathbf{u} \in E} P(|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi)$ and $W_m(E, \mathbf{a}) = \mathbb{E}(\sup_{\mathbf{u} \in E} \langle \mathbf{h}, \mathbf{u} \rangle)$. Then, for any $\xi > 0$ and $t \geq 0$, the following is true with probability at least $1 - e^{-2t^2}$:

$$\inf_{\mathbf{v} \in E} \left(\sum_{k=1}^m |\langle \mathbf{a}_k, \mathbf{v} \rangle|^2 \right)^{1/2} \geq \xi \sqrt{m} Q_{2\xi}(E, \mathbf{a}) - \xi t - 2W_m(E, \mathbf{a}).$$

In the case of (2.8), $E = T_{\rho,s}$, the set $\{\mathbf{a}_k\}_{k=1}^m$ constitutes the rows of the matrix \mathbf{A} , and the inequality is ensured by showing a lower bound on the *marginal tail function* Q_ξ and an upper bound on the *empirical width* W_m . **The lower bound on Q_ξ is typically proven by To establish an upper bound on W_m , it is used that ... so that an upper bound on ... suffices.**

This approach of showing the ℓ_2 -robust NSP with Mendelson's small ball method has been applied successfully to 0/1-Bernoulli matrices [22], as well as Gaussian matrices with a nonzero mean [29]. For our application, we are interested in matrices whose entries follow an exponential distribution. Formally deriving a lower bound on Q_ξ and an upper bound on W_m involves tedious computations that are beyond the scope of this thesis. Instead, we will broadly outline the validity of exponential matrices in Section 4.1, supported by numerical results.

2.2.3. Sparse recovery

...still have to write...

- introduce recovery w.r.t. the ℓ_0 -norm as a straightforward approach to recover sparse signals, and mention some common greedy algorithms
- explain that due to NP-hardness typically the ℓ_1 -norm is used instead (convex relaxation)

- state the BPDN, ℓ_1 -regularization, TV-regularization recovery schemes
- describe the CVX algorithm and the problems encountered with SPGL1
- for nonnegative matrices and signals, may resort to NNLS: outline advantages (simpler problem, no additional knowledge on noise needed), show that NNLS is equivalent to BPDN when sigma goes to zero, explain \mathcal{M}^+ -criterion (i.e. row space of the matrix must intersect the positive orthant)

CS recovery schemes usually employ the ℓ_1 -norm to induce sparsity in the solution. Let us briefly explain why. A straightforward approach to retrieve a sparse solution of (2.3) would be to minimize the ℓ_0 -norm of \mathbf{x} . More formally, this “norm” is the limit of ℓ_p -norms as $p \rightarrow 0$, i.e. $\|\mathbf{x}\|_0 := \lim_{p \rightarrow 0} \|\mathbf{x}\|_p^p = \lim_{p \rightarrow 0} \sum_{i=1}^n |x_i|^p$. However, as illustrated in Figure 2.4, any ℓ_p -norm with $0 \leq p < 1$ is concave, and hence results in computationally expensive optimization [5]. The lowest value of p that yields convex optimization while still inducing sparsity is 1. This approach is referred to as convex relaxation. Furthermore, the ℓ_p -norm no longer induces sparsity when $p > 1$ because it lacks the sharp shape needed to drive coefficients to zero, as depicted in Figure 2.5.

A common recovery scheme that will appear repeatedly throughout the thesis is basis pursuit denoising (BPDN):

$$\min \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma, \quad (2.9)$$

where σ typically represents an upper bound on the noise \mathbf{e} in (2.2), i.e. $\|\mathbf{e}\|_2 \leq \sigma$. Another formulation is ℓ_1 -regularization, given by $\min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \varepsilon \|\mathbf{x}\|_1$, which simultaneously minimizes the data fidelity and regularization term. In this sense, ℓ_1 -regularization goes a step further than BPDN, which only minimizes $\|\mathbf{x}\|_1$ within a fixed range of $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$. However, for appropriate values of the regularization parameter $\varepsilon \geq 0$, BPDN and ℓ_1 -regularization are equivalent [11].

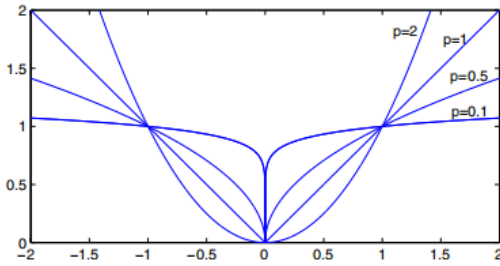


Figure 2.4.: $|x_i|^p$ for different p [5].

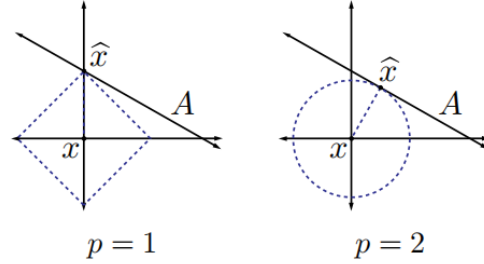


Figure 2.5.: ℓ_1 vs. ℓ_2 optimization [cite].

2.2.4. Reconstruction guarantees

conventional RIP-based BPDN guarantee for noisy measurements and generic signals \mathbf{x} :

Theorem 2.6. Assume that \mathbf{A} satisfies the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$, and let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \sigma$. Then the solution $\hat{\mathbf{x}}$ to (2.9) obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \frac{\|\mathbf{x} - \mathbf{x}_s\|_1}{\sqrt{s}} + C_1 \sigma,$$

where

$$C_0 = 2 \frac{1 - (1 - \sqrt{2})\delta_{2s}}{1 - (1 + \sqrt{2})\delta_{2s}}, \quad C_1 = 4 \frac{\sqrt{1 + \delta_{2s}}}{1 - (1 + \sqrt{2})\delta_{2s}}.$$

The proof aims to bound the size of the difference $\mathbf{h} := \hat{\mathbf{x}} - \mathbf{x}$. Hereto it decomposes \mathbf{h} into a sum of vectors $\mathbf{h}_{T_0}, \mathbf{h}_{T_1}, \mathbf{h}_{T_2}, \dots$, each of sparsity at most s : T_0 is the set of indices of the s largest entries of \mathbf{x} ; T_1 the indices of the s largest entries of $\mathbf{h}_{T_0^c}$; T_2 the indices of the next s largest entries of $\mathbf{h}_{T_0^c}$, and so on. The size of \mathbf{h} is then split up as $\|\mathbf{h}\|_2 \leq \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 + \|\mathbf{h}_{T_0 \cup T_1}\|_2$ and it is shown that (i) the size of \mathbf{h} outside of $T_0 \cup T_1$ is bounded by that of \mathbf{h} on $T_0 \cup T_1$ and (ii) $\|\mathbf{h}_{T_0 \cup T_1}\|_2$ is small. The upper bounds on $\|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2$ and $\|\mathbf{h}_{T_0 \cup T_1}\|_2$ are determined by repeatedly applying basic functional analysis tools, such as (reverse) triangle inequalities, Cauchy-Schwarz and bilinearity of inner products, as well as inequalities induced by the RIP; see [12] for the detailed computations. Ultimately, the bounds are combined to yield the upper bound from the theorem. Moreover, the requirement $\delta_{2s} < \sqrt{2} - 1$ ensures that C_0 and C_1 are positive. As shown in Figure 2.6, the constants blow up as δ_{2s} approaches $\sqrt{2} - 1$ and subsequently only attain negative values.

specific BPDN reconstruction guarantee based on the ℓ_2 -robust nullspace property: ...

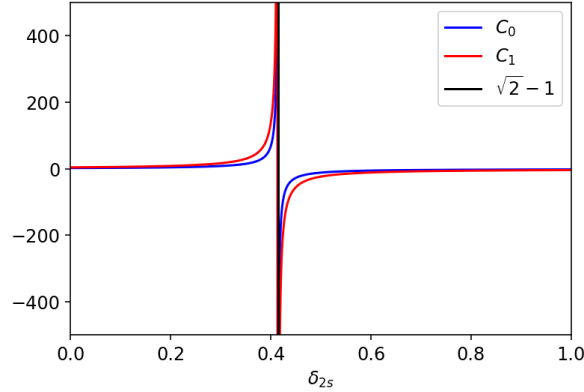


Figure 2.6.: Behavior of C_0 and C_1 in Theorem 2.6.

3. Effect of column-wise correlation in the measurement matrix

It has been demonstrated that when the distance between the input scanning beams is at least the diffraction limit of the MMF, the produced speckle patterns are mutually uncorrelated [2]. Thus, we may assume that the matrix in (2.2) has independent rows. Its columns, however, are correlated due to the smooth structure within each speckle pattern. In this chapter, we devise a model to examine the extent to which this correlation affects the reconstruction quality. Hereto we first consider one-dimensional speckle, which we then extend to patterns that are smoothed in two dimensions. Moreover, we investigate how the correlation effect changes when we impose constraints on the signal.

3.1. Model description

We study the case in which the measurement matrix \mathbf{A} has correlation within its rows, or equivalently, between its columns. Therefore, we set $\mathbf{A} := \mathbf{G}\mathbf{S}_L$, where \mathbf{G} is a random matrix with i.i.d. entries and \mathbf{S}_L an operator that performs a smoothing over a length L . For simplicity, we consider one-dimensional speckle patterns, as shown in Figure 3.1. The signal \mathbf{x} is one-dimensional as well, and the measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$ noise-free. Our goal is to find the maximum L for which the signal can be reconstructed, using the BPDN scheme (2.9).

We choose \mathbf{G} random Gaussian, i.e. $\mathbf{g}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with \mathbf{g}_k denoting the k th row of \mathbf{G} . Even though this distribution does not optimally reflect the experimental set-up where the intensities are non-negative and exponentially distributed, it serves as a reliable baseline for which recovery is guaranteed and allows us to focus entirely on the smoothing effect. Moreover, the normal distribution is closed under linear combinations such that the rows of \mathbf{A} are normally distributed as well, but with a non-trivial covariance matrix, say $\mathbf{\Sigma}$. We define the smoothing operator \mathbf{S}_L implicitly through $\mathbf{\Sigma}$. Hereto we use that products of the form $\mathbf{B}\mathbf{z}$ with \mathbf{B} a deterministic matrix and $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ a random vector are $\mathcal{N}(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{V}\mathbf{B}^T)$ [31]. In our case, the k th row of \mathbf{A} is given by $\mathbf{a}_k = \mathbf{g}_k\mathbf{S}_L$, and so $\mathbf{a}_k^T = \mathbf{S}_L^T\mathbf{g}_k^T$, i.e. $\mathbf{B} := \mathbf{S}_L^T$ and $\mathbf{z} := \mathbf{g}_k^T \sim \mathcal{N}(\boldsymbol{\mu} = \mathbf{0}, \mathbf{V} = \mathbf{I})$. It follows that $\mathbf{a}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} = \mathbf{S}_L^T\mathbf{S}_L)$. Thus the smoothing operator is interpreted through the Cholesky decomposition of a given covariance matrix.

In particular, we pick the radial basis function (RBF) kernel to constitute the entries of the covariance matrix. It is given by $\Sigma_{ij} = \exp(-(i-j)^2/L^2)$ and is a commonly used positive-definite kernel [18], so that taking the Cholesky decomposition is justified. This Σ is symmetric with unit diagonal and consists of correlation values. It performs a local smoothing in the rows of \mathbf{A} , with entries of \mathbf{a}_k located close to each other strongly correlated and the correlation decaying exponentially for entries that are further apart, as shown in Figure 3.2. Furthermore, larger values of L imply correlation over longer distances, and $\Sigma \rightarrow \mathbf{I}$ as L vanishes. We may view L as the number of pixels across which there is correlation, considering correlation below $e^{-1} \approx 0.368$ (which occurs when $|i-j| > L$) to be insignificant.

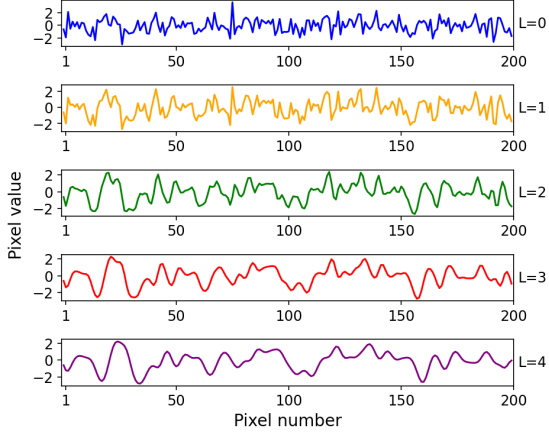


Figure 3.1.: 1D speckle (from the Gaussian distribution) for different L , with $L = 0$ no smoothing.

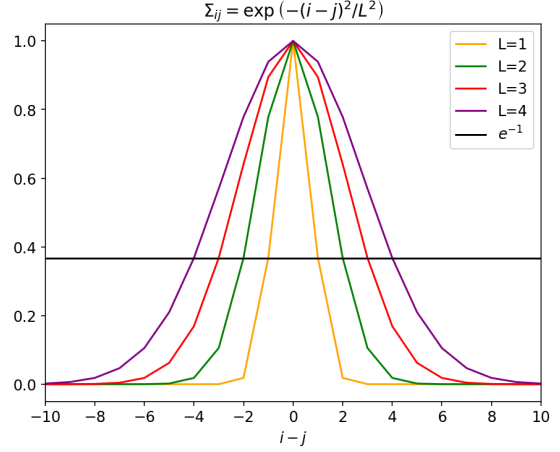


Figure 3.2.: RBF smoothing kernel for different L , with correlation at least e^{-1} considered significant.

3.2. Theoretical reconstruction guarantee

In this section, we derive an upper bound on L such that the reconstruction is theoretically justified. Our starting point is a particular case of the traditional guarantee for BPDN that was given in Theorem 2.6, with $\epsilon = \mathbf{0}$ and $\mathbf{x} = \mathbf{x}_s$:

Corrolary 1. *Assume that \mathbf{A} satisfies the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$, and let $\mathbf{y} = \mathbf{A}\mathbf{x}$. Then the solution $\hat{\mathbf{x}}$ to (2.9) obeys*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq 4 \frac{\sqrt{1 + \delta_{2s}}}{1 - (1 + \sqrt{2})\delta_{2s}} \sigma.$$

Effectively, we have to verify whether our matrix $\mathbf{G}\mathbf{S}_L$ satisfies the assumption of attaining the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$. We know that the assumption holds for vanishing L since $\mathbf{G}\mathbf{S}_L$ then approaches a random Gaussian matrix, and aim to find the maximum L for which it remains true.

In general, matrix products of the shape \mathbf{GX} with \mathbf{G} Gaussian and \mathbf{X} deterministic satisfy the RIP if and only if \mathbf{X} has the RIP [20]. The implication from right to left is of our interest; in the following, we provide a brief proof:

Proposition 1. *Let \mathbf{X} be some deterministic matrix and \mathbf{G} random Gaussian. Then the product \mathbf{GX} satisfies the RIP if \mathbf{X} has the RIP.*

Proof. Suppose \mathbf{X} has the RIP, then there is a $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s) \|\mathbf{v}\|_2^2 \leq \|\mathbf{Xv}\|_2^2 \leq (1 + \delta_s) \|\mathbf{v}\|_2^2 \quad (3.1)$$

for all s -sparse \mathbf{v} . Since \mathbf{G} is random Gaussian, we have that $\|\mathbf{Gz}\|_2^2$ is strongly concentrated around its mean $\|\mathbf{z}\|_2^2$ for any \mathbf{z} , which - using the concentration inequality (2.5) - we may write as $(1 - \epsilon) \|\mathbf{z}\|_2^2 \leq \|\mathbf{Gz}\|_2^2 \leq (1 + \epsilon) \|\mathbf{z}\|_2^2$ w.h.p. and $0 < \epsilon < 1$. Set $\mathbf{z} := \mathbf{Xv}$ to obtain $(1 - \epsilon) \|\mathbf{Xv}\|_2^2 \leq \|\mathbf{GXv}\|_2^2 \leq (1 + \epsilon) \|\mathbf{Xv}\|_2^2$, and in turn use (3.1) to conclude that there is a $\delta_s \in (0, 1)$ such that

$$(1 - (\delta_s + \epsilon - \delta_s \epsilon)) \|\mathbf{v}\|_2^2 \leq \|\mathbf{GXv}\|_2^2 \leq (1 + \delta_s + \epsilon + \delta_s \epsilon) \|\mathbf{v}\|_2^2$$

for all s -sparse \mathbf{v} and $0 < \epsilon < 1$. Lastly, pick ϵ sufficiently small such that both $\delta_s + \epsilon - \delta_s \epsilon$ and $\delta_s + \epsilon + \delta_s \epsilon$ lie in $(0, 1)$. \square

It remains to find when \mathbf{S}_L has the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$. Recall from Section 2.2.2 that any deterministic \mathbf{X} is $(s, \delta_s = (s - 1)\mu(\mathbf{X}))$ -RIP. By the way we constructed \mathbf{S}_L , its mutual coherence can be elegantly derived; with \mathbf{s}_i denoting the i th column of \mathbf{S}_L , it follows that

$$\begin{aligned} \mu(\mathbf{S}_L) &= \max_{1 \leq i \neq j \leq N} \frac{|\mathbf{s}_i^T \mathbf{s}_j|}{\|\mathbf{s}_i\|_2 \|\mathbf{s}_j\|_2} = \max_{1 \leq i \neq j \leq N} \frac{|\Sigma_{ij}|}{\sqrt{\Sigma_{ii} \Sigma_{jj}}} \\ &= \max_{1 \leq i \neq j \leq N} \exp\left(-\frac{(i - j)^2}{L^2}\right) = \exp(-L^{-2}). \end{aligned}$$

As a result, the BPDN guarantee holds as long as $(2s - 1) \exp(-L^{-2}) < \sqrt{2} - 1$. In other words, reconstruction is guaranteed for correlation lengths

$$L < \left(\ln \frac{2s - 1}{\sqrt{2} - 1} \right)^{-\frac{1}{2}}; \quad (3.2)$$

the corresponding region in the (L, s) -plane is displayed in Figure 3.3. This bound is rather strict, an observation we will return to shortly.

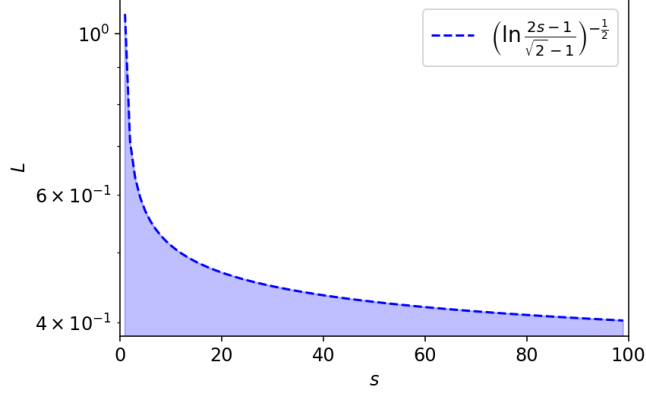


Figure 3.3.: Upper bound on L in terms of s .

3.3. Numerical results

This section presents results we obtained from simulations in Python (footnote with GitHub link) and compares them with the theoretical bound (3.2). The analysis centers around figures that assess for different cases of L the performance across varying number of taken measurements and sparsity levels of the ground truth.

More specifically, we generate s -sparse vectors \mathbf{x}^\dagger of fixed length $N = 200$ by randomly selecting s positions to put non-zero values, and drawing the non-zero values from the standard normal distribution. Similarly, we create matrices $\mathbf{A} = \mathbf{G}\mathbf{S}_L$ of size $m \times 200$ and simulate noise-free measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^\dagger$. For every s and m , we then have a problem instance (\mathbf{y}, \mathbf{A}) whose solution we estimate using CVX's BPDN with a varying $\sigma = 0.001\|\mathbf{A}\mathbf{x}^\dagger\|_2$ and no limit on the maximum number of iterations. This varying σ enables us to compare the outcomes across different \mathbf{A} , allowing for a relative error $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2 / \|\mathbf{A}\mathbf{x}^\dagger\|_2$ of at most 0.1%. For every reconstruction $\hat{\mathbf{x}}$, we compute the normalized mean-square error

$$e = \frac{\|\hat{\mathbf{x}} - \mathbf{x}^\dagger\|_2}{\|\mathbf{x}^\dagger\|_2} \quad (3.3)$$

and call the recovery a success if $e \leq 0.05$, so that reconstructions resembling at least 95% percent of the original signal are regarded successful. For every specific problem instance (\mathbf{y}, \mathbf{A}) , we run 100 Monte-Carlo reconstructions and define the probability of successful recovery as the proportion of the 100 estimates falling within the error range.

In the first experiment, we fix $s = 20$ and let m range in $\{10, 11, \dots, 200\}$ to obtain the probability of success across number of measurements, as shown in Figure 3.4. The experiment was conducted for different cases of L , with $L = 0$ representing the ideal scenario of $\mathbf{A} = \mathbf{G}$. The figure demonstrates that the correlation-free setting requires at least around 80 measurements to successfully recover a 200-long, 20-sparse signal with high probability. The reconstruction quality remains the same when correlation

over one pixel is induced and declines for correlation over two pixels, requiring around 120 measurements at least. As the correlation length exceeds two pixels, successful reconstruction becomes entirely hopeless, with even square matrices failing to reconstruct well.

Next, we set $m = 100$ and let s vary in $\{1, 2, \dots, 100\}$ to get the outcomes in Figure 3.5 on the probability of success across the ground truth's sparsity levels. Analogous to the measurement results, smoothing over one pixel does not affect the recovery quality, with $L = 0, 1$ both requiring the sparsity to be at most around 35. The performance drops at $L = 2$ and becomes rather poor for $L > 2$.

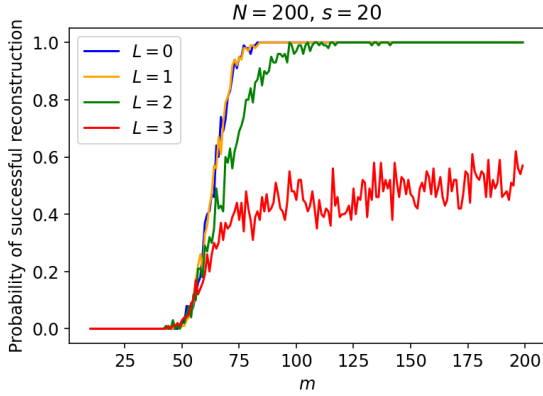


Figure 3.4.: p_{success} across number of measurements for $L = 0, 1, 2, 3$.

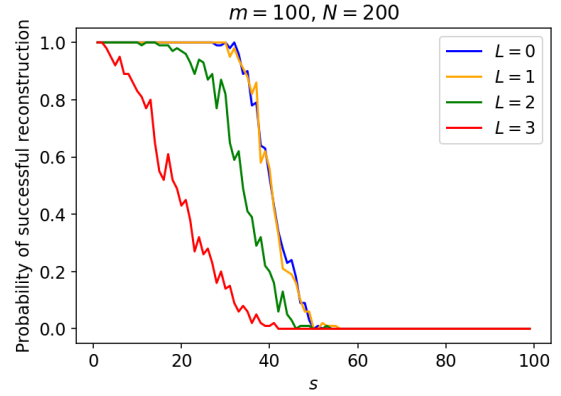


Figure 3.5.: p_{success} across sparsity of the ground truth for $L = 0, 1, 2, 3$.

Lastly, we vary the values of m and s simultaneously to obtain *probability phase transitions*, commonly used in compressed sensing to analyze performance [5]. In particular, we consider the fractions $\delta := m/N$ and $\rho := s/m$, and compute p_{success} across the points (δ, ρ) . Hereto we follow a similar procedure as outlined in [15]. The signal length $N = 200$ remains fixed and we vary m from 10 to 200 in steps of 10. For each value of m , we vary $s = 1, \dots, m$. Then, for every combination of m and s , the successes across 100 runs are counted to determine p_{success} , resulting in a collection of 3-tuples $(\delta, \rho, p_{\text{success}})$. Upon linear interpolation to get an evenly spaced (δ, ρ) -grid, we depict the probabilities in the (δ, ρ) -plane by associating with every point in the plane a gray value corresponding to the value of p_{success} (high/low probabilities in black/white, respectively).

Our outcomes are depicted in Figure 3.6. We considered the cases $L = 0, 1, 2$ as they appeared relevant from the previous measurement and sparsity results. As is expected to occur with probability phase transitions, a clear transition between a success region (in black) and a failure region (in white) is visible. In accordance with the measurement and sparsity results, we observe similar performance for $L = 0, 1$ and a drop in performance for $L = 2$. Additionally, we plotted the 50% level curves which form a boundary

between the success and failure regions. In general, such boundaries can be determined analytically by using combinatorial geometry [15], a computation which is outside the scope of this thesis.

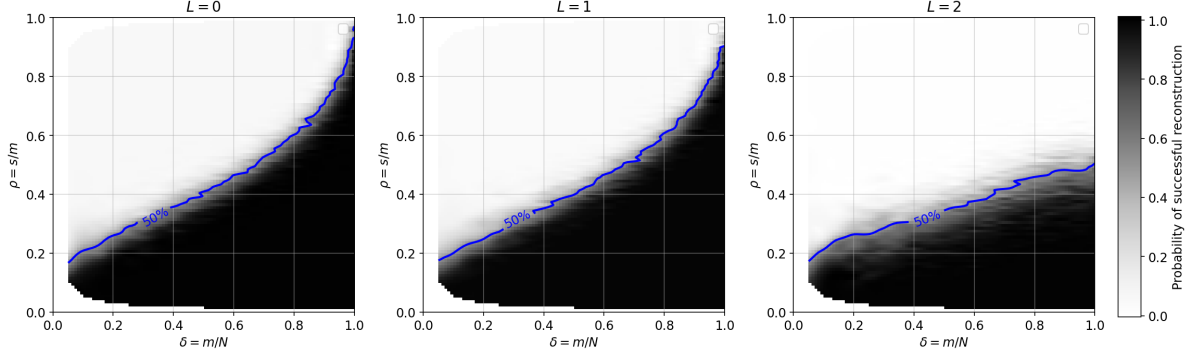


Figure 3.6.: Probability phase transitions for $L = 0, 1, 2$.

We draw two main conclusions from these numerical experiments. First of all, the allowed correlation appears rather limited: Figures 3.4, 3.5 and 3.6 all indicate a decline in reconstruction quality as the correlation length is increased from one pixel to two pixels. We should, however, keep in mind that a signal length of 200 is rather small compared to the number of pixels in actual speckle. Secondly, the theoretical bound (3.2) compared to the numerical outcomes is even more restrictive: e.g. at $s = 20$, the bound allows L at most ≈ 0.469 . Gaps between theory and numerical results or applications are common in compressed sensing, since the theory typically establishes worst-case bounds for recovery of all s -sparse signals rather than a subset of the signals [cite]. In the following, we briefly examine how numerical results are affected when constraints on the signal \mathbf{x} are imposed, after which we extend the numerical experiment to speckle patterns that are smoothed in two dimensions.

3.3.1. Correlation effect across signal classes

The first constraint is on the minimum number of zeros between any two non-zero signal entries. We expect signals whose non-zeros are separated by many zeros to suffer less from the measurement matrix's column-wise correlation. In particular, we consider for every L used to create \mathbf{GS}_L only signals \mathbf{x}_L with separation distance (number of zeros) at least L . To ensure that all s nonzero signal entries are located at least L zero entries away from each other, we randomly draw s positions from the set $\{0, \dots, N - L(s - 1)\}$, sort the selected positions in increasing order, and add to each position i the value $L \cdot i$. Moreover, we consider L at most 2 to prevent that the product $L(s - 1)$ exceeds the signal length N at higher sparsity levels. Applying the same recovery procedure from before on \mathbf{x}_L yields the outcomes in Figures 3.7 and 3.8. There are no significant improvements for $L = 1$, as signals $\mathbf{x}_{L=1}$ are not much different from constraint-free \mathbf{x} . The case of $L = 2$, however, shows significant improvement, even outperforming the correlation-free

case. Thus, one can overcome the smoothing effect by imposing a separation distance constraint in the signal.

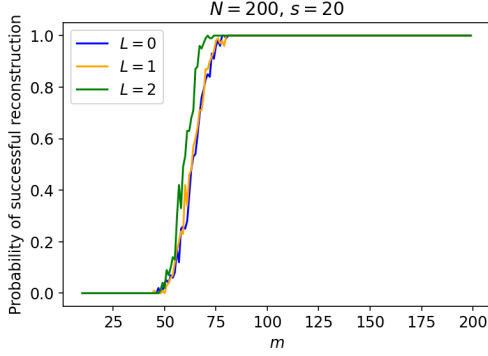


Figure 3.7.: $p_{success}$ across number of measurements for $L = 0, 1, 2$, considering only \mathbf{x}_L .

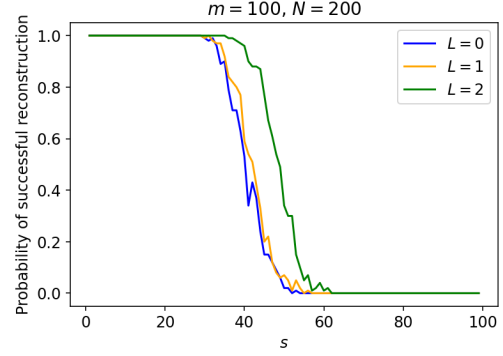


Figure 3.8.: $p_{success}$ across sparsity of the ground truth for $L = 0, 1, 2$, considering only \mathbf{x}_L .

Next, we remove the separation distance constraint and consider non-negative signals $\mathbf{x}_{\geq 0}$, where non-zeros are defined as the absolute values of Gaussian random variables. This restriction resembles more closely the setup of the physics experiment which involves light intensities. We also enforce non-negativity in the recovery process by adding the constraint $\mathbf{x}_{\geq 0}$ to CVX; apart from this modification, we keep the exact same reconstruction scheme used thus far. According to the literature, the outcomes should improve [15]. Comparing Figures 3.9 and 3.10 with Figures 3.4 and 3.5 respectively, we see improvement for $L = 0, 1, 2$, requiring a minimum number of measurements of ± 65 and a maximum sparsity of around 50. The case of $L = 3$ also shows better results, although in the sparsity plot a drop in performance between a correlation length of 2 and 3 is visible. When $L = 4$, the outcomes are rather poor, but still outperform those obtained for $L = 3$ in the constraint-free case. In conclusion, imposing non-negativity improves results overall, but the separation distance restriction is better at reducing correlation effects.

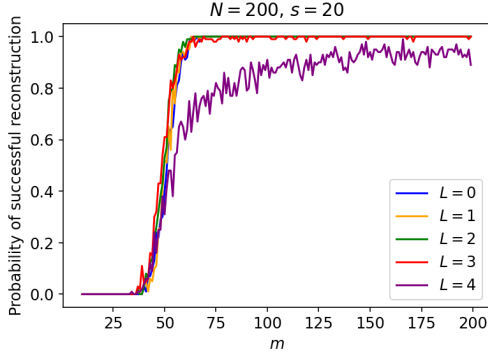


Figure 3.9.: p_{success} across number of measurements for $L = 0, 1, 2, 3, 4$, considering only signals $\mathbf{x}_{\geq 0}$.

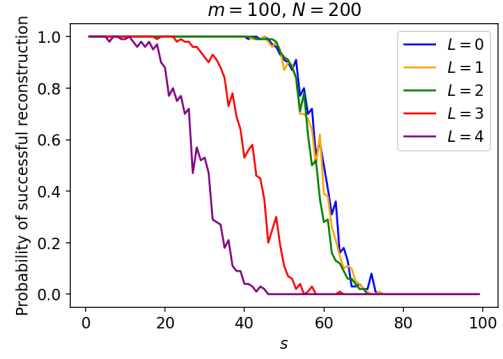


Figure 3.10.: p_{success} across sparsity of the ground truth for $L = 0, 1, 2, 3, 4$, considering only signals $\mathbf{x}_{\geq 0}$.

3.4. Two-dimensional setting

In this section, we extend the model to n -by- n speckle patterns by modifying the RBF kernel to smooth in two directions. We then simulate speckle patterns from a multimode fiber and inquire whether their correlation structure coincides with that of our model. Moreover, we find the smoothing length underlying the fiber speckle patterns and study its effect on the reconstruction quality.

To adapt the kernel, we introduce the Euclidean distance $d_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\|_2$ by assigning to each pixel $i = 0, 1, \dots, n^2 - 1$ a position $\mathbf{p}_i := (q_i, r_i)$ with q_i the quotient and r_i the remainder of i divided by n , as illustrated in Figure 3.11(a). The two-dimensional RBF kernel is then defined as $\tilde{\Sigma}_{ij} = \exp(-d_{ij}^2/L^2)$ with $L = \sqrt{L_x^2 + L_y^2}$ composed of the horizontal and vertical smoothing lengths $L_x, L_y \in \mathbb{Z}_{\geq 0}$. The matrix $\tilde{\Sigma} \in \mathbb{R}^{n^2 \times n^2}$ consists of correlations whose values decay around a unit diagonal. As a result, applying the Cholesky factor $\tilde{\mathbf{S}}_L$ to an n^2 -dimensional row vector $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and reshaping the product $\mathbf{g}\tilde{\mathbf{S}}_L$ into an n -by- n array yields a pattern as shown in Figure 3.11(b). This local smoothing of a noisy baseline is similar to cutting off high frequencies in the Fourier domain, a common technique for emulating speckle.

Next, we simulate a fiber speckle pattern, shown in Figure 3.11(c). Hereto we use the module pyMMF [28], which calculates for a given fiber object and light beam the collection of mode profiles $\{\Psi_k(x, y)\}_{k=1}^M$. Based on the profiles, we compute the speckle field according to (2.1) with amplitudes $E_k \in [0, 1]$ and phases $\phi_k \in [-\pi, \pi]$ randomly selected, and display the intensities $|E_{\text{out}}|^2$. We note that randomly drawing the amplitudes and phases simplifies the physics, but suffices for studying the correlation effect. Furthermore, we apply a square mask to the pattern, which is circular since the mode profiles are by default calculated for a round fiber. Although pyMMF offers the possibil-

ity to use square fibers, we found that a bug in the module limits the mode computation to only six modes.

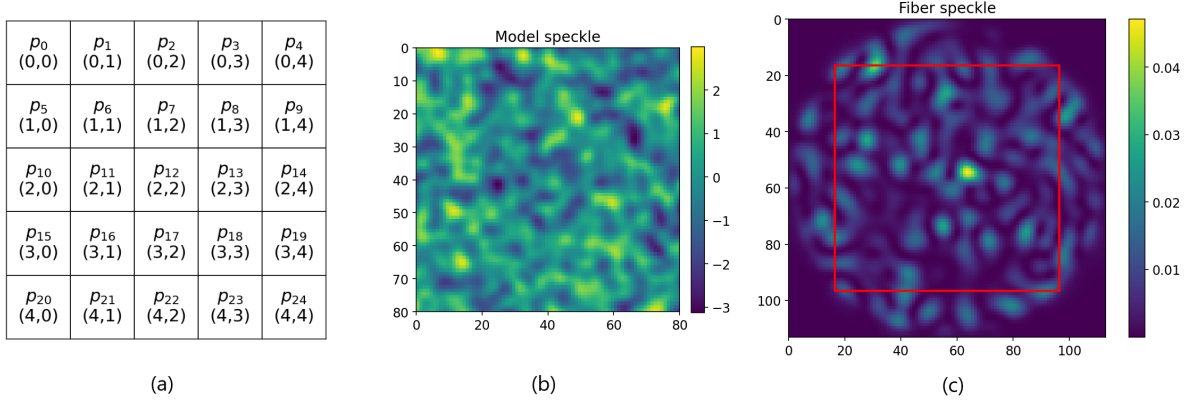


Figure 3.11.: (a) Coordinate system underlying the distance metric in the 2D RBF kernel for a 5-by-5 array. (b) 80-by-80 pattern obtained by applying $\tilde{\mathbf{S}}_L$ with $L = \sqrt{13}$ to a standard Gaussian baseline. (c) Simulated pattern from propagating a light beam with wavelength $1 \mu\text{m}$ through a fiber with core radius $20 \mu\text{m}$, NA 0.22, core refractive index $1.46 \mu\text{m}$ and 197 modes, as well as a square 80-by-80 mask.

Before retrieving the fiber speckle’s smoothing length, we verify that its correlation structure resembles that induced by our model. Hereto we fix the set $\{\Psi_k(x, y)\}_{k=1}^M$ of Figure 3.11(c) and generate 5000 cropped speckle fields by randomly choosing E_k and ϕ_k . Next, we flatten every field to obtain a collection of vectors from which we calculate a sample correlation matrix $\bar{\Sigma}$. The entries of $\bar{\Sigma}$ estimate the correlations between the pixels in the fiber speckle pattern, and are thus identical in meaning to the entries of $\tilde{\Sigma}$. Figure (add heatmap of sample correlation matrix) illustrates that $\bar{\Sigma}$ consists of decreasing values around a unit diagonal, and so it resembles $\tilde{\Sigma}$. We can therefore estimate smoothing length in the fiber speckle by minimizing the Frobenius norm

$$\|\bar{\Sigma} - \tilde{\Sigma}\|_F = \sqrt{\sum_{i=1}^{n^2} \sum_{j=1}^{n^2} |\bar{\Sigma}_{ij} - \tilde{\Sigma}_{ij}|^2}$$

with respect to L . This minimization yields $L \approx \sqrt{13}$ for 80-by-80 patterns such as in Figure 3.11(c). To confirm that the correlations coincide for $L \approx \sqrt{13}$, Figure (add figure) displays for each pixel i the normalized mean-square difference $\|\tilde{\Sigma}_i - \bar{\Sigma}_i\|_2 / \|\tilde{\Sigma}_i\|_2$ between the model’s and fiber’s correlations. The low gray values imply that the model’s smoothing indeed accurately reflects the smooth structure present in fiber speckle patterns.

Let us now perform numerical experiments as in Section 3.3. First, we reduce the computational burden by downscaling the problem from $n \times n = 80 \times 80$ to $n \times n = 20 \times 20$

and studying the effect of smoothing over $L = \frac{\sqrt{13}}{4} \approx 1$ instead. Again, we generate s -sparse ground truths $\mathbf{x}^\dagger \in \mathbb{R}^{n^2}$ by randomly choosing indices to place nonzeros, drawing nonzero values from $\mathcal{N}(0, 1)$ and taking absolute values to more closely resemble reality. Furthermore, the measurement matrices $\mathbf{G}\tilde{\mathbf{S}}_L \in \mathbb{R}^{m \times n^2}$ have a standard Gaussian baseline \mathbf{G} and the measurements $\mathbf{y} = \mathbf{G}\tilde{\mathbf{S}}_L\mathbf{x}^\dagger \in \mathbb{R}^m$ are noise-free. Using CVX's BPDN with $\mathbf{x} \geq 0$, a varying $\sigma = 0.001\|\mathbf{G}\tilde{\mathbf{S}}_L\mathbf{x}^\dagger\|_2$, and deeming reconstructions $\hat{\mathbf{x}}$ with normalized mean-square errors (3.3) at most 0.05 successful, we obtain the results across $m = 10, 11, \dots, 200$ and $s = 1, 2, \dots, 100$ shown in Figures 3.12 and 3.13 respectively. Since smoothing over $L = 1$ does not affect performance, we conclude that the smoothing effect is negligible for speckle patterns such as in Figures 3.11(b) and 3.11(c). Moreover, we see that a considerable amount of smoothing is allowed, with even $L = \sqrt{13}$ reconstructing accurately at high measurement numbers and low sparsity levels.

Having analyzed the effect of correlation, we remark that even though the smoothing lengths in Figures 3.11(b) and 3.11(c) coincide, the intensity values are noticeably different. This difference is due to the fact that our model follows a mean-zero normal distribution, while natural fiber speckle is exponentially distributed. Therefore, in the following chapter we will refine our model to follow specific target distributions.

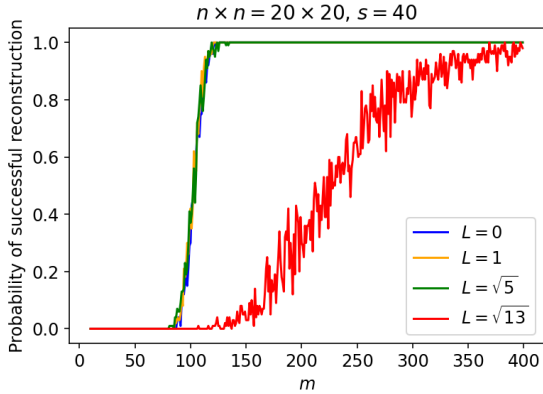


Figure 3.12.: $p_{success}$ across the number of measurements for varying L .

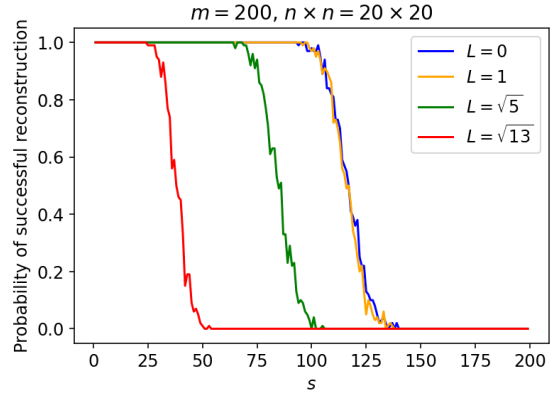


Figure 3.13.: $p_{success}$ across the sparsity of the ground truth for varying L .

4. Non-negative measurement matrices

4.1. Random exponential matrices

4.1.1. ℓ_2 -robust nullspace property

- numerically demonstrate upper bound on empirical width
- reason why achieving a lower bound on marginal tail function is possible

4.1.2. \mathcal{M}^+ -criterion

In this section, we establish the \mathcal{M}^+ -criterion for random exponential matrices, using a similar procedure as for 0/1-Bernoulli matrices [22] and shifted Gaussian matrices [29]. We define a random exponential matrix as $\mathbf{A} \in \mathbb{R}^{m \times N}$ composed of i.i.d. entries $a_{ij} \sim \text{Exp}(\lambda = 1/\langle I \rangle)$ with $\langle I \rangle$ a parameter representing the average light intensity, and have that $\mathbb{E}(a_{ij}) = \langle I \rangle$. Furthermore, our proof will invoke Corollary 7.32 from [16], which is a particular instance of a Bernstein inequality for subexponential random variables:

Corollary 2. *Let c_1, \dots, c_m be independent mean-zero subexponential random variables, i.e. $P(|c_k| \geq t) \leq \beta e^{-\kappa t}$ for some constants $\beta, \kappa > 0$ for all $t > 0$, $k = 1, \dots, m$. Then*

$$P\left(\left|\sum_{k=1}^m c_k\right| \geq t\right) \leq 2 \exp\left(-\frac{(\kappa t)^2/2}{2\beta m + \kappa t}\right). \quad (4.1)$$

Our result is as follows:

Theorem 4.1. *Suppose that $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a random exponential matrix with parameter $\lambda = 1/\langle I \rangle > 0$, then $\mathbf{A} \in \mathcal{M}^+$ with probability $\geq 1 - 2n \exp(-em/(4e + 16))$.*

Proof. We must show the existence of a $\mathbf{t} \in \mathbb{R}^m$ such that $\mathbf{w} := \mathbf{A}^T \mathbf{t}$ is strictly positive. Call $\mathbf{1}_m$ the vector of ones in \mathbb{R}^m and $\mathbf{1}_{N \times m}$ the matrix of ones in $\mathbb{R}^{N \times m}$, and set $\mathbf{t} := \frac{1}{\langle I \rangle m} \mathbf{1}_m$. Using the linearity of the expectation operator and $\mathbb{E}(\mathbf{A}^T) = \langle I \rangle \mathbf{1}_{N \times m}$, it follows that $\mathbb{E}(\mathbf{w}) = \mathbb{E}(\mathbf{A}^T) \mathbf{t} = \langle I \rangle \mathbf{1}_{N \times m} \frac{1}{\langle I \rangle m} \mathbf{1}_m = \mathbf{1}_n$. To demonstrate strict positivity, we establish that \mathbf{w} is concentrated around its nonzero mean. More specifically, we ensure that w.h.p. $|\mathbf{w}_i - 1| < \frac{1}{2}$ for each $i = 1, \dots, n$ by bounding $P(|\mathbf{w}_i - 1| \geq \frac{1}{2})$ from above with Corollary 2.

First, we rewrite $|\mathbf{w}_i - 1|$ by expressing \mathbf{w} as $\frac{1}{\langle I \rangle m} \sum_{k=1}^m \mathbf{a}_k$ with \mathbf{a}_k the k th row of \mathbf{A} :

$$|\langle \mathbf{e}_i, \mathbf{w} \rangle - 1| = \left| \frac{1}{\langle I \rangle m} \sum_{k=1}^m \langle \mathbf{e}_i, \mathbf{a}_k \rangle - \frac{\langle I \rangle m}{\langle I \rangle m} \right| = \frac{1}{\langle I \rangle m} \left| \sum_{k=1}^m (a_{ki} - \mathbb{E}(a_{ki})) \right|, \quad (4.2)$$

where \mathbf{e}_i is the i th standard basis vector in \mathbb{R}^n and $\langle I \rangle m = \sum_{k=1}^m \mathbb{E}(a_{ki})$ in the second equality. Next, we invoke Corollary 2 on $c_k := a_{ki} - \mathbb{E}(a_{ki})$. Since the a_{ki} are independent and exponentially distributed, the c_k are independent, mean-zero and subexponential as required. We find the expressions of β and κ in (4.1) by writing

$$\begin{aligned} P(|c_k| \geq t) &= P(|a_{ki} - \langle I \rangle| \geq t) = P(a_{ki} \geq t + \langle I \rangle) + P(a_{ki} \leq -t + \langle I \rangle) \\ &= \int_{t+\bar{I}}^{\infty} \frac{1}{\langle I \rangle} e^{-x/\langle I \rangle} dx + \int_0^{-t+\langle I \rangle} \frac{1}{\langle I \rangle} e^{-x/\langle I \rangle} dx \\ &= e^{-1} e^{-t/\langle I \rangle} - e^{-1} e^{t/\langle I \rangle} + 1 \leq e^{-1} e^{-t/\langle I \rangle} \quad \forall t \geq \langle I \rangle. \end{aligned}$$

The change of variables $\tilde{t} := t - \langle I \rangle$ yields $P(|c_k| \geq \tilde{t}) \leq e^{-1} e^{-\tilde{t}/\langle I \rangle}$ for all $\tilde{t} > 0$, and so $\beta = e^{-1}$ and $\kappa = 1/\langle I \rangle$. From the corollary and (4.2) it follows that

$$P\left(|\langle \mathbf{e}_i, \mathbf{w} \rangle - 1| \geq \frac{1}{2}\right) = P\left(\left|\sum_{k=1}^m c_k\right| \geq \frac{\langle I \rangle m}{2}\right) \leq 2 \exp\left(-\frac{em}{4e + 16}\right).$$

Lastly, we union bound over $i = 1, \dots, n$ to conclude that all components of \mathbf{w} are positive with probability $\geq 1 - 2n \exp(-em/(4e + 16))$. \square

...mention that \mathcal{M}^+ -criterion for other distributions generally requires bounded/sub-Gaussian/subexponential random variables...

4.2. Other distributions

Having analyzed the validity of random exponential matrices from a theoretical perspective, we turn to their numerical performance. In particular, we consider exponentially distributed patterns that also contain a smoothing induced by the two-dimensional RBF kernel from Section 3.4. As it is physically possible to customize speckle intensity distributions (Figure 4.1), we simultaneously examine whether the optical experiment could benefit from customized light by also conducting numerical experiments for these other distributions. Section 4.2.1 first describes how we generate patterns with an assigned smoothing length and intensity distribution, followed by numerical results in Section 4.2.2.

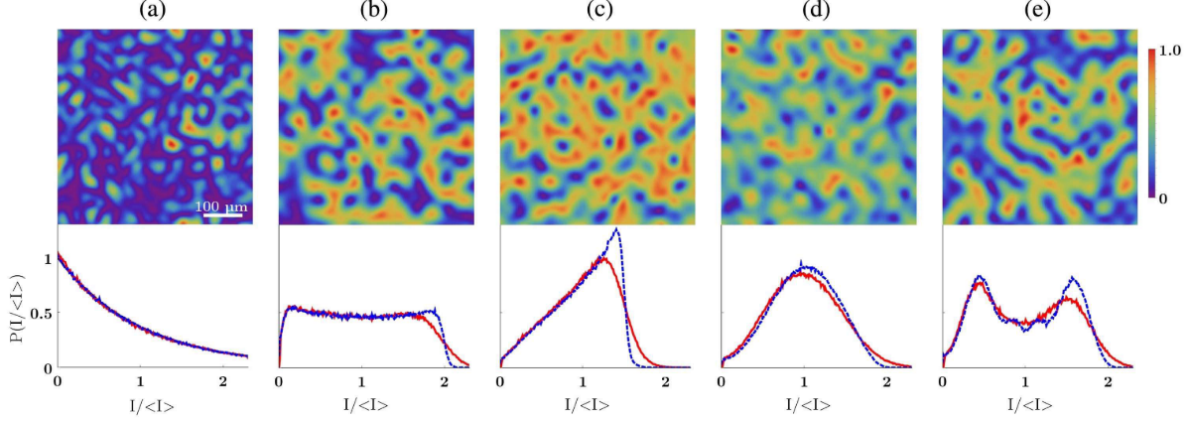


Figure 4.1.: ...

4.2.1. Model description

A key advantage of using a normally distributed baseline \mathbf{g} is that the distribution is closed under linear combinations and thus the product $\mathbf{g}\tilde{\mathbf{S}}_L$ remains Gaussian. Distributions (a)-(c) in Figure 4.1 do not possess this stability property, i.e. applying the smoothing operator to an exponential, uniform or triangular baseline will alter the distribution. To achieve other distributions while maintaining the smoothing, we therefore continue to smooth in the Gaussian domain, after which we transform the pattern to each target distribution.

We transform to other distributions by means of a method called inverse transform sampling [cite], which says that one can generate realizations r of a random variable R with a target distribution function F_R by applying the inverse distribution function F_R^{-1} to realizations u of $U \sim \text{Unif}[0, 1]$. In our case, we start with a random vector $\mathbf{g}\tilde{\mathbf{S}}_L \in \mathbb{R}^{n^2}$ composed of entries $(\mathbf{g}\tilde{\mathbf{S}}_L)_i \sim \mathcal{N}(0, 1)$ which are mutually correlated through the smoothing operator. Applying the Gaussian distribution function Φ entry-wise to $\mathbf{g}\tilde{\mathbf{S}}_L$ results in the vector \mathbf{u} made up of $u_i \sim \text{Unif}[0, 1]$. Next, we apply the inverse distribution function F_R^{-1} of the target distribution entry-wise to \mathbf{u} to obtain the vector \mathbf{r} whose entries follow F_R . In other words, applying the operator $F_R^{-1} \circ \Phi$ entry-wise to $\mathbf{g}\tilde{\mathbf{S}}_L$ allows us to directly control the distribution of the speckle. Except for the bimodal distribution, the inverse distribution functions – also called quantile functions – are readily available in Python. Figure 4.2 displays the outcomes of this procedure.

Since the goal is to compare the different distributions' performance, the parameters of the targets F_R are chosen such that they all have a unit variance: $\lambda = 1$ in (i) as $\text{Var}(R) = 1/\lambda^2$ for $R \sim \text{Exp}(\lambda)$; $a = 0$ and $b = \sqrt{12}$ in (ii) as $\text{Var}(R) = (b - a)^2/12$ for $R \sim \text{Unif}[a, b]$; $a = 0$ and $b = c = \sqrt{18}$ in (iii) as $\text{Var}(R) = (a^2 + b^2 + c^2 - ab - ac - bc)/18$ for R triangular with lower limit a , upper limit b and mode c . For $\mathcal{N}(\mu, 1)$ in (iv), we pick $\mu = 3$ so that the majority of the density function ($\approx 99.7\%$ lies within 3σ from μ

[cite]) is nonnegative and resample potential negative values. The bimodal distribution in (v) will be discussed shortly as it is somewhat more intricate. Moreover, we must verify that the transformations do not distort the smoothing. Therefore, similarly as in Section 3.4, we calculate for each distribution a sample covariance matrix $\tilde{\Sigma}$ from 5000 patterns (Figure ...), as well as the normalized mean-square difference (...) with the model $\tilde{\Sigma}$ (Figure ...). From these sample covariance matrices and normalized mean-square differences we conclude that the smoothing is overall maintained. Lastly, we remark that the mean values $\mathbb{E}(R)$ are different across the targets F_R ; however, since it has been established that shifts in Gaussian distributions do not influence performance [29], we will not concern ourselves with this difference for now.

The bimodal density function in (v) is a mixture of two normal distributions, i.e. $f_R(x) := 0.5g_1(x) + 0.5g_2(x)$ with g_1 and g_2 the densities of $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ respectively. For suitable values of μ_1 , μ_2 and σ , the density f_R has two peaks and a unit variance. In particular, two peaks occur when $\sigma \leq (\mu_1 - \mu_2)/2$ with $\mu_2 > \mu_1 > 0$, which we can express equivalently as

$$\sigma = 0.5 \cdot c \cdot d \quad (4.3)$$

with $d := \mu_2 - \mu_1 > 0$, $\mu_1 > 0$ and $0 < c \leq 1$ a constant. Moreover, the variance of the Gaussian mixture is $\sigma_{f_R} = 0.25(\mu_2 - \mu_1)^2 + \sigma^2$, or $\sigma_{f_R} = 0.25(1 + c^2)d^2$ by using (...). Setting $\sigma_{f_R} = 1$, we find that $c = \sqrt{(4 - d^2)/d^2}$ and $\sqrt{2} < d < 2$. As a result, we can entirely define the bimodal density function with unit variance in terms of the location μ_1 of the first peak and the distance d between the peaks. To make sure that the peaks are located not too close and not too far from each other, we set $d = (2 + \sqrt{2})/2 \approx \dots$, and to make sure that the majority of the density is nonnegative, we pick $\mu_1 = 1.5$, and resample potential negative values. Based on this idea of a Gaussian mixture, we implement the bimodal quantile function in Python by defining the bimodal distribution function through the built-in normal distribution function, and numerically approximate the inverse distribution function by

...still have to write...

- for each target distribution, we set a variance of 1 so that we can later compare the different distributions' performance; also add a comparison of the sample covariance matrices of the different distributions to show that the transformation does not distort the smoothing length - check
- explain how you obtain the bimodal distribution as a mixture of two Gaussian distributions
- mention how the different distributions obey the boundedness/sub-Gaussian/subexponential requirement from Section 4.1, so that from a theoretical point of view we expect the distributions to perform similar

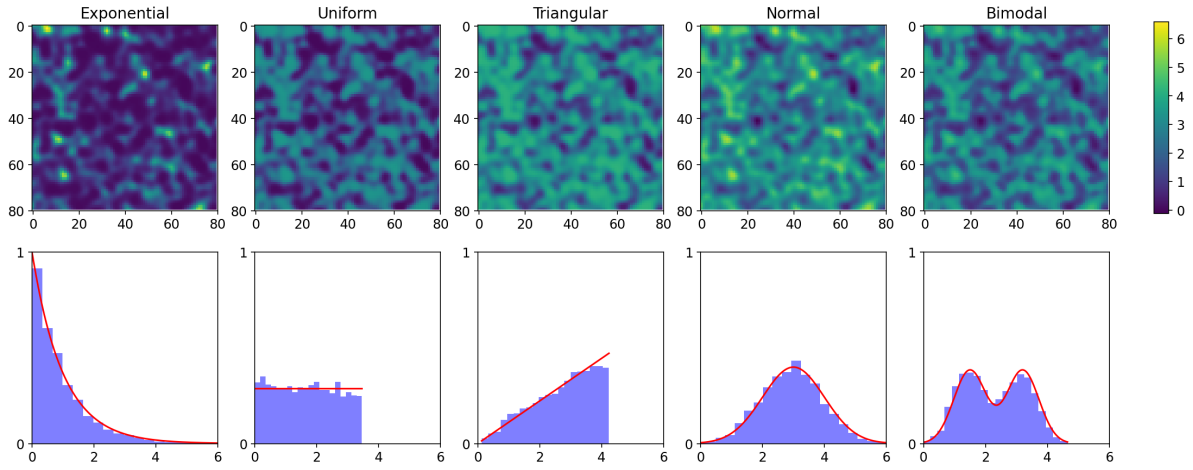


Figure 4.2.: ...

4.2.2. Numerical experiment

- add plot of the simulations across the different distributions and conclude that for generic ground truths the use of the distribution does not impact performance
- thus, exponentially distributed light is justified and there is no motivation to use customize the laser light in general
- however, for specific samples it might be useful to use different light, e.g. it might be beneficial to use bimodal light for the sample with two beads

5. Conclusion

- summarize findings: constructed a model to generate speckle with specified smoothing length and distribution function, established theoretical upper bound on smoothing, found by means of numerical experiments that smoothing affects reconstruction quality, looked at different signal classes, compared the model's smoothing with fiber speckle and found that if the speckle is reasonably small the smooth structure does not affect performance, established that NNLS can be used, showed that for general ground truths distribution of light does not matter
- suggest directions for further research, in particular reconstruction of images that are not sparse in the original domain, and do a little analysis with a checkerboard ground truth and the wavelet transform

Bibliography

- [1] URL: <https://www.cwi.nl/en/groups/computational-imaging/>.
- [2] Ksenia Abrashitova and Lyubov V. Amitonova. “High-speed label-free multimode-fiber-based compressive imaging beyond the diffraction limit”. In: *Optics Express* 30.7 (2022), pp. 10456–10469.
- [3] Dimitris Achlioptas. “Database-friendly random projections: Johnson-Lindenstrauss with binary coins”. In: *Journal of Computer and System Sciences* 66.4 (2003), pp. 671–687.
- [4] Pohchoo Seow et al. “Evaluation of compressed SENSE on image quality and reduction of MRI acquisition time: a clinical validation study”. In: *Academic Radiology* 31.3 (2024), pp. 956–965.
- [5] David L. Donoho Alfred M. Bruckstein and Michael Elad. “From sparse solutions of systems of equations to sparse modelling of signals and images”. In: *SIAM Review* 51.1 (2009), pp. 34–81.
- [6] Lyubov V. Amitonova. “Multimode fiber endoscopes for computational brain imaging”. In: *Neurophotonics* 11.S1 (2024), S11509-1–7.
- [7] Lyubov V. Amitonova and Johannes F. de Boer. “Compressive imaging through a multimode fiber”. In: *Optics Letters* 43.21 (2018), pp. 5427–5430.
- [8] Afonso S. Bandeira et al. “The road to deterministic matrices with the restricted isometry property”. In: *Journal of Fourier Analysis and Applications* 19 (2013), pp. 1123–1149.
- [9] Richard Baraniuk et al. “A simple proof of the restricted isometry property for random matrices”. In: *Constructive Approximation* 28 (2008), pp. 253–263.
- [10] Nicholas Bender et al. “Customizing speckle intensity statistics”. In: *Optica* 5.5 (2018), pp. 595–600.
- [11] Ewout van den Berg and Michael P. Friedlander. “Probing the Pareto frontier for basis pursuit solutions”. In: *SIAM Journal on Scientific Computing* 31.2 (2008), pp. 890–912.
- [12] Emmanuel J. Candès. “The restricted isometry property and its implications for compressed sensing”. In: *Comptes Rendus Mathématique* 346.9–10 (2008), pp. 589–592.
- [13] Mark A. Davenport et al. “Introduction to compressed sensing”. In: *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012, pp. 1–64.

- [14] David L. Donoho, Michael Elad, and Vladimir N. Temlyakov. “Stable recovery of sparse overcomplete representations in the presence of noise”. In: *IEEE Transactions on Information Theory* 52.1 (2006), pp. 6–18.
- [15] David L. Donoho and Jared Tanner. “Precise undersampling theorems”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 913–924.
- [16] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [17] Joseph W. Goodman. *Speckle phenomena in optics: theory and applications, Second edition*. SPIE, 2020.
- [18] Koji Tsuda Jean-Philippe Vert and Bernhard Schölkopf. “A primer on kernel methods”. In: *Kernel Methods in Computational Biology*. The MIT Press, 2004, pp. 35–70.
- [19] Apostolos Karadimitrakakis, Mérouane Debbah, and Aris Moustakas. “Optical MIMO: Results and analysis”. In: *11th International Symposium on Wireless Communications Systems*. Barcelona, Spain, 2014, pp. 966–970.
- [20] Shiva P. Kasiviswanathan and Mark Rudelson. “Restricted isometry property under high correlations”. In: ? (2019).
- [21] Velat Kilic, Trac D. Tran, and Mark A. Foster. “Compressed sensing in photonics: a tutorial”. In: *Journal of the Optical Society of America B* 40.1 (2023), pp. 28–52.
- [22] Richard Kueng and Peter Jung. “Robust nonnegative sparse recovery and the nullspace property for 0/1 measurements”. In: ? ?? (2017), ?
- [23] Tristan van Leeuwen and Christopher Brune. “Inverse problems and imaging”, *Lecture notes*. 2025. URL: https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html.
- [24] Wei Li, Ksenia Abrashitova, and Lyubov V. Amitonova. “Super-resolution multi-mode fiber imaging with an untrained neural network”. In: *Optics Letters* 48.13 (2023), pp. 3363–3366.
- [25] Zhouping Lyu. “High-resolution imaging through a multimode fiber: from raster-scanning to compressive sensing”. Doctoral Dissertation. VU Amsterdam, 2024.
- [26] Fedor Mitschke. *Fiber optics: physics and technology, Second edition*. Springer, 2016.
- [27] Rudiger Paschotta. *Field Guide to Lasers*. SPIE Press, 2008.
- [28] Sébastien M. Popoff. *pyMMF*. 2025. URL: <https://github.com/wavefrontshaping/pyMMF>.
- [29] Yonathan Shadmi, Peter Jung, and Giuseppe Caire. “Sparse non-negative recovery from biased subgaussian measurements using NNLS”. In: ? ?? (2019), ?
- [30] James H. Steiger. *Random Vectors, Random Matrices, and Their Expected Values*. URL: <https://www.statpower.net/Content/313/Lecture%20Notes/MatrixExpectedValue.pdf>.

- [31] Marco Taboga. “*Linear combinations of normal random variables*”, *Lectures on probability theory and mathematical statistics*. 2021. URL: <https://www.statlect.com/probability-distributions/normal-distribution-linear-combinations>.
- [32] Yusuke Yamada. “Textile-Integrated Polymer Optical Fibers for Healthcare and Medical Applications”. In: *Biomedical Physics Engineering Express* 6 (2020), p. 062001.
- [33] Lawrence J. Zhang. “A Friendly Introduction to Compressed Sensing”. Undergraduate Honors Thesis. University of California San Diego, 2021.

A. Appendix

(still a bit messy)

Lemma A.1. *Let \mathbf{A} ($m \times N$) be a random matrix, drawn according to any distribution that satisfies (2.5). Then, for any S with $|S| = s < m$ and any $0 < \delta < 1$, (2.4) holds for all $\mathbf{x} \in X_S$ with probability $\geq 1 - 2(\frac{12}{\delta})^s e^{-c_0(\frac{\delta}{2})^m}$.*

Proof. Choose a cover Q_S of X_S such that $Q_S \subset X_S$, $\|\mathbf{q}\|_2 = 1$ for all $\mathbf{q} \in Q_S$ and $\min_{\mathbf{q} \in Q_S} \|\mathbf{x} - \mathbf{q}\|_2 \leq \frac{\delta}{4}$ for all $\mathbf{x} \in X_S$. Such a cover can be chosen with $|Q_S| \leq (\frac{12}{\delta})^s$ [9]. Then apply (2.5) with $\epsilon := \frac{\delta}{2}$ to get $\mathbb{P}(|\|\mathbf{A}\mathbf{q}\|_2^2 - \|\mathbf{q}\|_2^2| \geq \frac{\delta}{2}\|\mathbf{q}\|_2^2) \leq 2e^{-mc_0(\frac{\delta}{2})}$ for each fixed $\mathbf{q} \in Q_S$, and union bound over $\mathbf{q} \in Q_S$ to find that

$$(1 - \frac{\delta}{2})\|\mathbf{q}\|_2^2 \leq \|\mathbf{A}\mathbf{q}\|_2^2 \leq (1 + \frac{\delta}{2})\|\mathbf{q}\|_2^2$$

with probability $\geq 1 - 2(\frac{12}{\delta})^s e^{-c_0(\frac{\delta}{2})^m}$ across all $\mathbf{q} \in Q_S$.

Subsequently, extend the RIP to all $\mathbf{x} \in X_S$. Hereto define δ^* as the smallest number such that $\|\mathbf{A}\mathbf{x}\|_2 \leq (1 + \delta^*)\|\mathbf{x}\|_2$ for all $\mathbf{x} \in X_S$ and show that $\delta^* \leq \delta$. Write $\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{q}\|_2 + \|\mathbf{A}(\mathbf{x} - \mathbf{q})\|_2 \leq 1 + \frac{\delta}{2} + (1 + \delta^*)\frac{\delta}{4}$ and use that by definition $1 + \delta^* \leq 1 + \frac{\delta}{2} + (1 + \delta^*)\frac{\delta}{4}$. Upon rewriting, the latter inequality yields $\delta^* \leq \delta$, and the upper bound of the RIP is proven. The lower bound follows from applying the reverse triangle inequality on $\|\mathbf{A}(\mathbf{x} - \mathbf{q} + \mathbf{q})\|_2$ and using $\delta^* \leq \delta$. \square

Theorem A.2. *Suppose m, n and $0 < \delta < 1$ are given. If the distribution generating \mathbf{A} ($m \times n$) satisfies (2.5), then there are constants $c_1, c_2 > 0$ depending only on δ such that the RIP holds with the prescribed δ and any $s \leq c_1 m / \log(n/s)$ with probability $\geq 1 - 2e^{-c_2 m}$.*

Proof. Lemma A.1 established that for each X_S , \mathbf{A} fails to satisfy (2.4) with probability $\leq 2(\frac{12}{\delta})^s e^{-c_0(\frac{\delta}{2})^m}$. In total, there are $\binom{n}{s} \leq (\frac{en}{s})^s$ such X_S , so (2.4) fails to hold with probability $\leq 2(\frac{en}{s})^s (\frac{12}{\delta})^s e^{-c_0(\frac{\delta}{2})^m} = 2e^{-c_0(\frac{\delta}{2})^m + s(1 + \log(\frac{n}{s} + \log(\frac{12}{\delta}))}$. For $s \leq c_1 m / \log(n/s)$, the exponent may be bounded from above by $-c_2 m$ with $c_2 \leq (c_0(\frac{\delta}{2}) - c_1(1 + (1 + \log(\frac{12}{\delta}))/\log(\frac{n}{s})))$. To conclude, $c_1 > 0$ can be chosen sufficiently small such that $c_2 > 0$. \square