

Master's Mathematics
Track: Biomedical Mathematics
Internship report

Compressive sensing for multimode fiber optical imaging

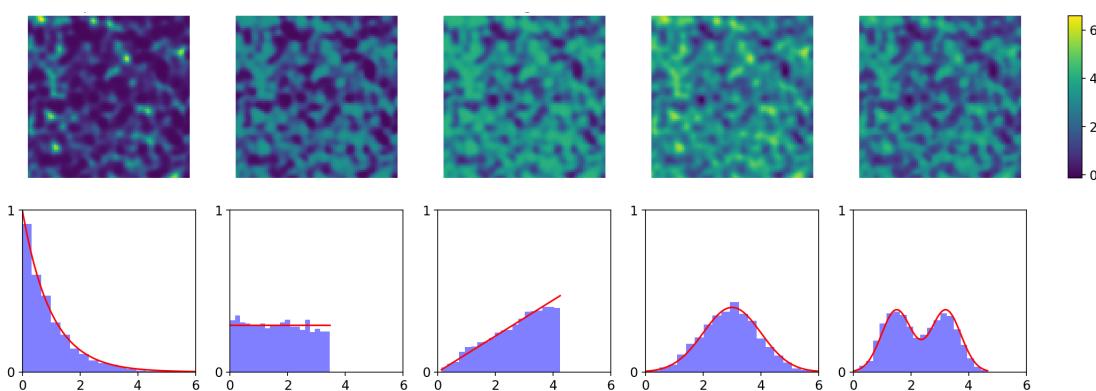
by
Emma Erkočević
April 17, 2025

First supervisor: dr. Svetlana Dubinkina

Second reader: prof.dr. Tristan van Leeuwen

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

Host Organization: Centrum Wiskunde & Informatica



Department of Mathematics
Faculty of Science

Summary

This thesis studies the recently developed optical microscopy technique known as multimode fiber compressive imaging in the mathematical context of compressed sensing. The microscopy technique constitutes an inverse problem that reconstructs images of a sample by illuminating it with random patterns emerging from multimode fibers. It provides high-resolution images from few illuminations and hence aligns with the compressive sensing framework for solving underdetermined linear systems. Using the theoretical principles of compressed sensing and numerical simulations, we explore potential improvements. First, we investigate to what extent correlations present in the random speckle patterns affect performance. We devise a model that emulates the correlations and derive an upper bound on the smoothing length based on the restricted isometry property. Upon estimating the smoothing length in realistic patterns and numerically solving simulated problem sets with basis pursuit denoising, we conclude that the correlation effect is negligible as long as the speckle size is sufficiently small. Next, we examine speckle patterns with variously distributed light intensities, as well as reconstruction schemes other than conventional ℓ_1 -optimization. We start by analyzing the ℓ_2 -robust nullspace property and \mathcal{M}^+ -criterion for random exponential arrays and infer from compressive sensing literature that the optical experiment is likely to benefit from nonnegative least-squares when ground truths are sparse in the original domain. We then utilize an inversion method to create patterns with specific correlation lengths and intensity distributions, and conclude from numerical experiments that all the distributions perform equally well in recovering arbitrary sparse signals.

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. Lyubov Amitonova

Host Organization: Centrum Wiskunde & Informatica

Date: April 17, 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	15
3. Effect of column-wise correlation in the measurement matrix	17
3.1. Model description	17
3.2. Theoretical reconstruction guarantee	18
3.3. Numerical results	20
3.3.1. Correlation effect across signal classes	22
3.4. Two-dimensional setting	23
4. Nonnegative measurement matrices	27
4.1. Random exponential matrices	27
4.1.1. ℓ_2 -robust nullspace property	27
4.1.2. \mathcal{M}^+ -criterion	29
4.2. Other distributions	30
4.2.1. Model description	30
4.2.2. Numerical results	32
5. Conclusion	35
Bibliography	37
A. Appendix	41
A.1. Restricted isometry property for random matrices	41
A.2. Concentration of measure inequalities	42

1. Introduction

In 2018, a novel optical microscopy technique called multimode fiber compressive imaging (MMFCI) was developed at the Advanced Research Center for Nanolithography (AR-CNL) [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 is an inverse problem that fits into the framework of compressive sensing (CS). Mathematically, CS offers principles for retrieving 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 [20]. From a signal processing perspective, compressed sensing provided a new sampling paradigm, allowing signal reconstruction from less data than conventionally deemed necessary [13]. 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 [11], 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. Lyubov Amitonova of ARCNL to ensure this work would have experimental value as well.

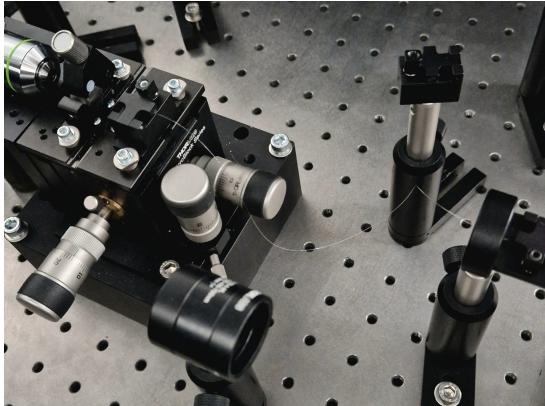


Figure 1.1.: Glass fiber in the ARCNL laboratory.

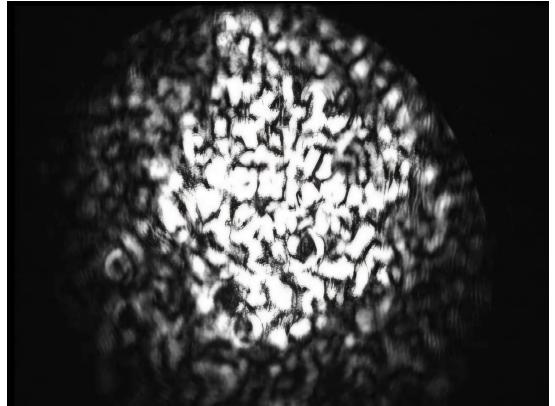


Figure 1.2.: Example of a speckle pattern [21].

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 [32]. 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 [33].

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 [21]. In other words, the elementary mathematical representation of a single speckle pixel (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 [31].

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 $d = \lambda/(2NA)$, 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.

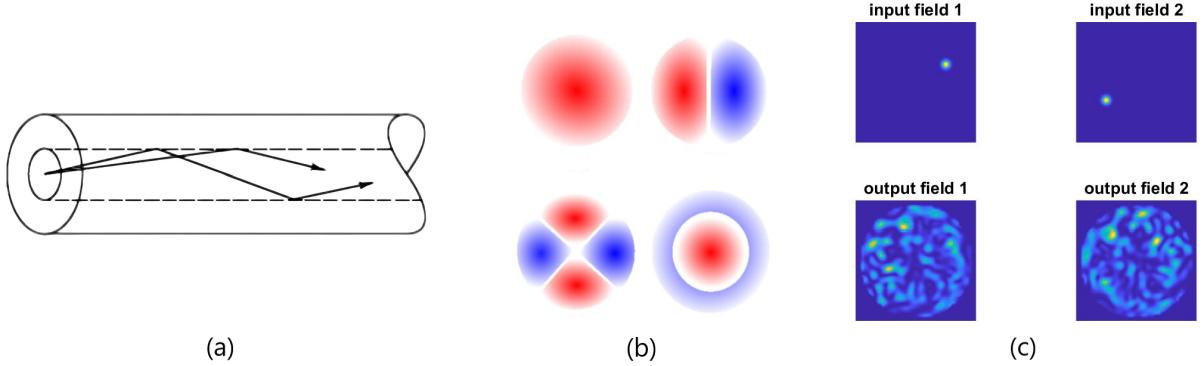


Figure 2.1.: (a) Simple illustration of a multimode fiber with two modes [39]. (b) Four mode intensity profiles at a fiber’s cross-section [23]. (c) Speckle from two input fields (MATLAB simulation of ARCNL).

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

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

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 nonzeros. The reconstruction is carried out by minimizing the ℓ_1 -norm of \mathbf{x} – or of $\nabla\mathbf{x}$ to preserve edges, known as total variation regularization – subject to (2.2), 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 [20]. Conventional sensing approaches sample according to Nyquist, after which the data often is converted to a compressed form [25]. Compressive sensing is based on the idea that data can be acquired and compressed simultaneously, employing a rule of thumb given by $m \gtrsim 2s \log(n^2/m)$ [19]. Indeed, the experiment of Figure 2.2(a) obeys this rule, as $m = 961$, $n^2 = 160000$ and the disks occupy $s = 208$ nonzero pixels. In the following section, we describe the mathematical theory of CS relevant to multimode fiber imaging.

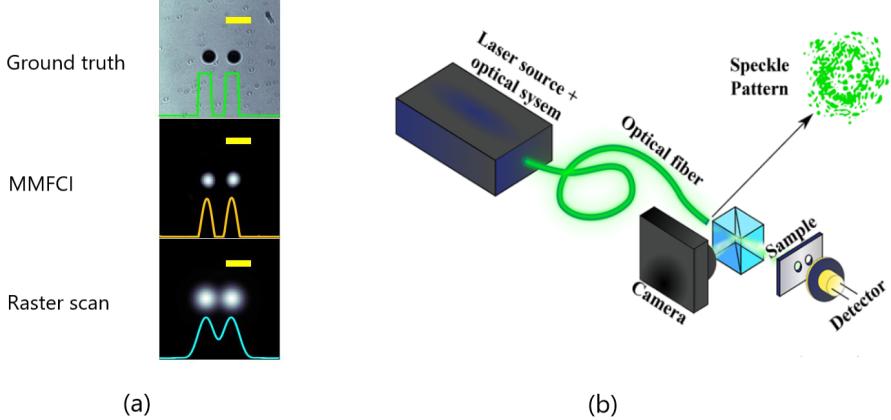


Figure 2.2.: (a) Experimental reconstruction with MMFCI of a 1.9-by-1.9 mm sample with two disks of diameter $\approx 54.1 \mu\text{m}$, along with a raster-scan diffraction-limited reconstruction [2]. (b) Basic schematic overview of the experimental set-up, consisting of the following components [30]: 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

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

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 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 inverse 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 [29]. 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 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 nonzeros. 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 [13]. 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 nonsparse 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 such that submatrices formed by columns of \mathbf{A} are well-conditioned [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 $\sqrt{(N-m)/(m(N-1))}$ [14]. Generally, low values indicate stable recovery. 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 [18]). 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.

Matrices in compressed sensing often contain randomness. Consequently, many of the proof techniques that we will encounter are based on notions from random matrix theory. A random matrix $\mathbf{R} \in \mathbb{R}^{m \times N}$ 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 $\mathbf{R}(\omega)$ with $\omega \in \Omega^{mN}$ by selecting each entry \mathbf{R}_{ij} as a realization of R [9]. Unless stated otherwise, the term random matrix in this thesis assumes independent and identically distributed entries, e.g. a Gaussian matrix consists of i.i.d. entries $\mathbf{R}_{ij} \sim \mathcal{N}(\mu, \sigma^2)$, a 0/1-Bernoulli matrix of i.i.d. $\mathbf{R}_{ij} \sim \text{Ber}(p)$, and an exponential matrix of i.i.d. $\mathbf{R}_{ij} \sim \text{Exp}(\lambda)$.

Furthermore, many properties of random variables extend to random matrices. For example, a random matrix's expected value is defined as the matrix of expected values, $[\mathbb{E}(\mathbf{R})]_{ij} := \mathbb{E}(\mathbf{R}_{ij})$, and hence linearity of the expectation operator also holds for random matrices [37]. Proof techniques commonly use bounds on the probability that a linear combination of the matrix's entries deviates from the matrix's mean –

known as concentration of measure inequalities – combined with union bounds of the form $\mathbb{P}(\cup_{k=1}^{\infty} B_k) \leq \sum_{k=1}^{\infty} \mathbb{P}(B_k)$ with $\{B_k\}_{k=1}^{\infty}$ a countable collection of events [20]. The concentration inequalities generally rely on the tail behavior of the probability distribution, requiring the tails to decay at least as fast as those of a Gaussian distribution (sub-Gaussian) or an exponential distribution (subexponential).

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

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

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 . 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 [14].

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 nonnegative 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.

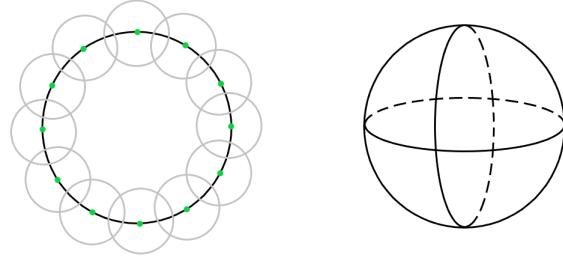


Figure 2.3.: (Left) s -dimensional unit sphere X_S , along with its covering Q_S in green; (right) union of s -dimensional unit spheres Σ_s .

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 (2.4) is interpreted as follows: 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 $[1 - \delta_s, 1 + \delta_s]$. In addition, a classic bound on the eigenvalues of a square matrix, the Gershgorin circle theorem, is invoked:

Theorem 2.3 (Theorem 3 of [8]). *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, and thus the Gram matrix $\mathbf{A}_S^T \mathbf{A}_S$ has a unit diagonal. By the definition of mutual coherence, 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 means 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) [27]. 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 $(\mathbf{v}_S)_i := v_i \mathbb{1}_{i \in S}$ – 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. that $\|\mathbf{v}\|_2 = 1$ 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} . 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 $T_{\rho, s} := \{\mathbf{v} \in \mathbb{R}^N : \|\mathbf{v}\|_2 = 1, \|\mathbf{v}_{S_{\max}}\|_2 > \frac{\rho}{\sqrt{s}} \|\mathbf{v}_{\bar{S}_{\max}}\|_1\}$. In other words, the problem reduces 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 as (2.8), and whose application only requires row-wise independence of the matrix. Its generic description is as follows:

Theorem 2.5 (Theorem 8 of [27]). *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 $Q_\xi(E, \mathbf{a}) = \inf_{\mathbf{u} \in E} \mathbb{P}(|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi)$ 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 aim is to show a lower bound on the *marginal tail function* Q_ξ and an upper bound on the *mean empirical width* W_m .

This approach of showing the ℓ_2 -robust NSP with Mendelson's small ball method has been applied successfully to 0/1-Bernoulli matrices [27], as well as Gaussian matrices with a nonzero mean [36]. Our application concerns matrices whose entries follow an exponential distribution. Formally establishing a lower bound on Q_ξ and an upper bound on W_m is computationally tedious and beyond the scope of this thesis; in Section 4.1, we will outline the broad idea and provide numerical approximations of W_m .

2.2.3. Sparse recovery

An intuitive approach to retrieve a sparse solution of (2.3) is to minimize the ℓ_0 -norm of \mathbf{x} . However, this approach amounts to an NP-hard combinatorial problem. Even though the solution of ℓ_0 -minimization can be approximated by means of greedy algorithms which identify the support of the signal based locally optimal choices, CS recovery schemes commonly employ the ℓ_1 -norm [5]. To understand why, write the ℓ_0 -norm as the limit of ℓ_p -norms as $p \rightarrow 0$, i.e. $\|\mathbf{x}\|_0 := \lim_{p \rightarrow 0} \|\mathbf{x}\|_p^p = \lim_{p \rightarrow 0} \sum_{i=1}^N |x_i|^p$, and notice in Figure 2.4 that any ℓ_p -norm with $0 < p < 1$ is concave. In addition, the ℓ_p -norm no longer induces sparsity when $p > 1$ because it lacks the sharp shape needed to drive coefficients to zero, as illustrated in Figure 2.5. Thus, the value of p that yields convex optimization while still inducing sparsity is 1. This technique is referred to as convex relaxation.

In this thesis, we employ a standard ℓ_1 -optimization scheme known as basis pursuit denoising (BPDN):

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

where the threshold η typically represents an upper bound on the noise \mathbf{e} , i.e. $\|\mathbf{e}\|_2 \leq \eta$, in noisy systems as (2.2). Under the assumption that \mathbf{A} fulfills a requirement such as the RIP or ℓ_2 -robust NSP, this scheme ensures stable recovery according to the guarantees stated in the following section. Furthermore, the BPDN scheme can alternatively be expressed as ℓ_1 -regularization, given by $\min \|\mathbf{y} - \mathbf{Ax}\|_2 + \varepsilon \|\mathbf{x}\|_1$ with $\varepsilon \geq 0$ the regularization parameter. Due to the simultaneous minimization of the data fidelity and regularization term, ℓ_1 -regularization goes a step further than BPDN, which only minimizes $\|\mathbf{x}\|_1$ within a fixed range of $\|\mathbf{y} - \mathbf{Ax}\|_2$. However, for appropriate values of ε , BPDN and ℓ_1 -regularization are equivalent [12]. It is also worth noting that many ℓ_1 -minimization algorithms (e.g. FISTA of [10]) converge without monotonic decrease of their objective functions.

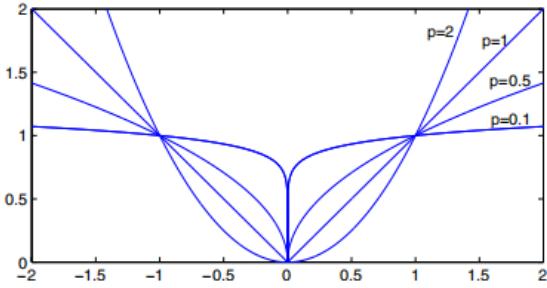


Figure 2.4.: $|x_i|^p$ for different p [5].

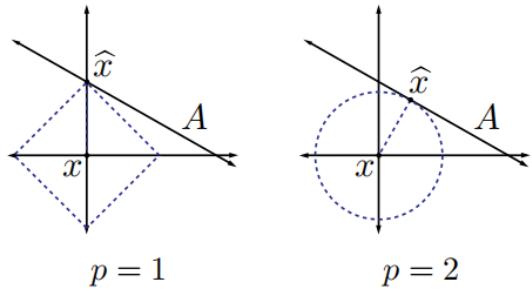


Figure 2.5.: ℓ_1 vs. ℓ_2 optimization [14].

The numerical experiments¹ will be conducted using CVXPY, a Python-embedded modeling language for convex optimization problems [16]. It lets the user formulate the problem in a way that follows the mathematics; for example, one provides the BPDN scheme as follows:

```
x = cp.Variable(N)
objective = cp.Minimize(cp.norm(x, 1))
constraints = [cp.norm(A @ x - y, 2) <= eta]
problem = cp.Problem(objective, constraints)
problem.solve()
```

The module then calls an open source convex optimization solver (CLARABEL, OSQP, SCS, or another solver if installed separately) suitable for the problem. We note that the particular solvers and their strategies are outside the focus of this thesis.

Furthermore, in scenarios where the signal is nonnegative, one can resort to basic nonnegative least-squares regression (NNLS),

$$\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - \mathbf{Ax}\|_2,$$

provided that the matrix fulfills the ℓ_2 -robust NSP and admits a strictly positive linear combination of its rows. The latter is known as the \mathcal{M}^+ -criterion and requires the existence of a $\mathbf{t} \in \mathbb{R}^m$ such that $\mathbf{w} := \mathbf{A}^T \mathbf{t}$ is strictly positive. Given these conditions, NNLS enables stable recovery of compressible (approximately sparse) signals from noisy systems (2.2). Section III-B of [27] discusses the scheme's recovery guarantees and their proofs in detail. NNLS, as opposed to BPDN, does not need an a-priori bound on the noise, and its recovery error scales proportionally to $\|\mathbf{e}\|_2$. The \mathcal{M}^+ -criterion has readily been proven for 0/1-Bernoulli [27] and shifted Gaussian matrices [36]; in Section 4.1, we prove the criterion for exponential matrices. Thus, for ground truths as in Figure 2.2(a) that are sparse in the original domain, MMFCI is likely to benefit from employing NNLS. It is worth noting that NNLS has already been applied in optical microscopy (cf. Chapter II-9 of [28]). We will focus solely on demonstrating the \mathcal{M}^+ -criterion and leave NNLS-based reconstructions for future research.

¹All of the thesis' code is provided on https://github.com/emmaerkocevic/compressive_sensing_fiber_optics/tree/main.

2.2.4. Reconstruction guarantees

The conventional RIP-based guarantee for recovering compressible signals \mathbf{x} with BPDN from noisy measurements is as follows:

Theorem 2.6 (Theorem 1.9 of [14]). *Assume that \mathbf{A} satisfies the RIP of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$, and let $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \eta$. Then the solution $\hat{\mathbf{x}}$ to (2.9) obeys*

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

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 restricted vectors $\mathbf{h}_{T_0}, \mathbf{h}_{T_1}, \mathbf{h}_{T_2}, \dots$, each of sparsity at most s : T_0 is the index set of the s largest entries of \mathbf{x} ; T_1 the index set of the s largest entries of $\mathbf{h}_{T_0^c}$ with $T_0^c = \{1, \dots, N\} \setminus T_0$; T_2 the index set 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 [13] for the detailed computations. Ultimately, the bounds are combined to yield the upper bound from the theorem. It is worth noting that the proof still applies in cases where \mathbf{h} does not have a unique decomposition. 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.

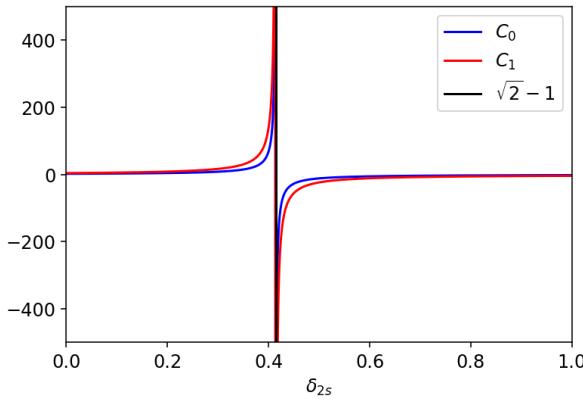


Figure 2.6.: Behavior of C_0 and C_1 in Theorem 2.6.

To conclude, let us state an alternative version of the above guarantee that relies on the ℓ_2 -robust NSP:

Theorem 2.7 (Theorem 4.22 of [20]). *Suppose that the matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies the ℓ_2 -robust nullspace property of order s with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $\mathbf{x} \in \mathbb{R}^N$, a solution $\hat{\mathbf{x}}$ of (2.9) with $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$ and $\|\mathbf{e}\|_2 \leq \eta$ approximates the vector \mathbf{x} with ℓ_p -error*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_p \leq C \frac{\|\mathbf{x} - \mathbf{x}_s\|_1}{s^{1-1/p}} + D s^{1/p-1/2} \eta, \quad 1 \leq p \leq 2,$$

for some constants $C, D > 0$ depending only on ρ and τ .

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 $\mathbf{y} = \mathbf{Ax}$ 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 nonnegative 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 nontrivial covariance matrix, say Σ . We define the smoothing operator \mathbf{S}_L implicitly through Σ . Hereto we use that products of the form \mathbf{Bz} with \mathbf{B} a deterministic matrix and $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ a random vector are $\mathcal{N}(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\mathbf{V}\mathbf{B}^T)$ [38]. 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 $\mathbf{a}_k \sim \mathcal{N}(\mathbf{0}, \Sigma = \mathbf{S}_L^T\mathbf{S}_L)$. 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 [22], 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.

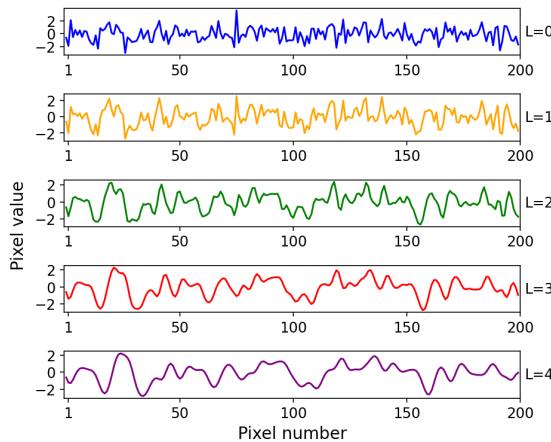


Figure 3.1.: 1D speckle (from the Gaussian distribution) for different L , with $L = 0$ no smoothing.

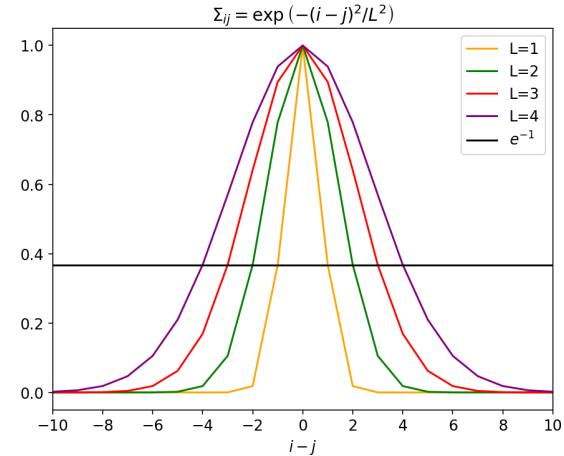


Figure 3.2.: RBF smoothing kernel for different L , with correlation at least e^{-1} considered significant.

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 $\mathbf{e} = \mathbf{0}$ and $\mathbf{x} = \mathbf{x}_s$:

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

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

Effectively, we have to verify whether our matrix \mathbf{GS}_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{GS}_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 [24]. 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. We constructed \mathbf{S}_L such that 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 (s, L) -plane is displayed in Figure 3.3. This bound is rather strict, an observation we will return to shortly.

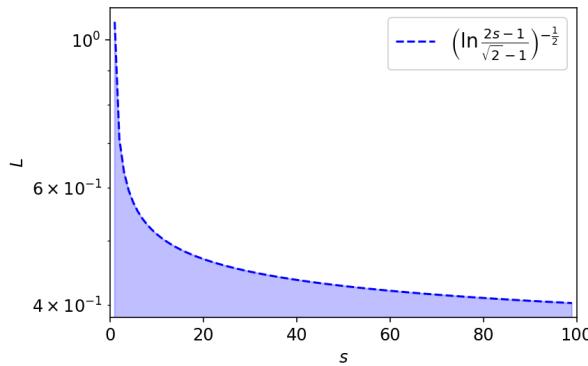


Figure 3.3.: Upper bound on L in terms of s .

3.3. Numerical results

This section presents results from numerical experiments with simulated data 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} of fixed length $N = 200$ by randomly selecting s positions to put nonzero values, and drawing the nonzero 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 $\mathbf{y} = \mathbf{Ax}$. For every s and m , we then have a problem instance (\mathbf{y}, \mathbf{A}) whose solution we estimate using CVXPY's BPDN with a varying $\eta = 10^{-3}\|\mathbf{Ax}\|_2$ and no specified maximum number of iterations. This varying η enables us to compare the outcomes across different \mathbf{A} , allowing for a relative error $\|\mathbf{y} - \mathbf{Ax}\|_2/\|\mathbf{Ax}\|_2$ of at most 0.1%. For every reconstruction $\hat{\mathbf{x}}$, we compute the normalized mean-square error (NMSE)

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

and call the recovery a success if $e \leq 0.05$, so that reconstructions resembling at least 95% 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 $L > 2$.

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 [19]. 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 [19], a computation which is outside the scope of this thesis.

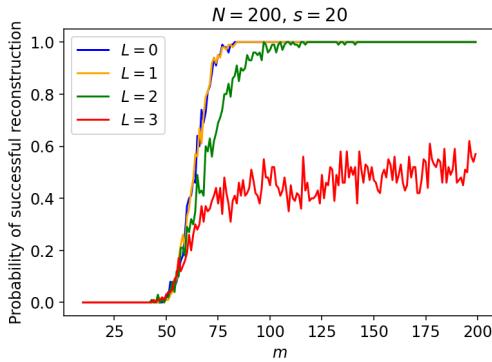


Figure 3.4.: $p_{success}$ across number of measurements for $L = 0, 1, 2, 3$.

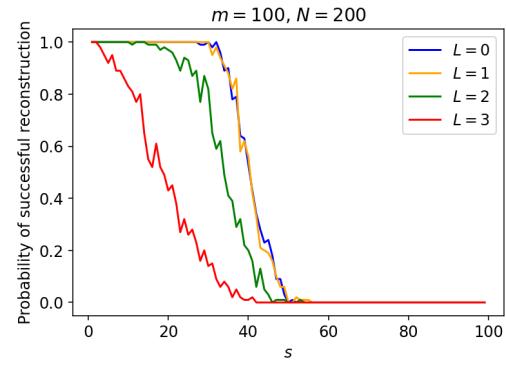


Figure 3.5.: $p_{success}$ across sparsity of the ground truth for $L = 0, 1, 2, 3$.

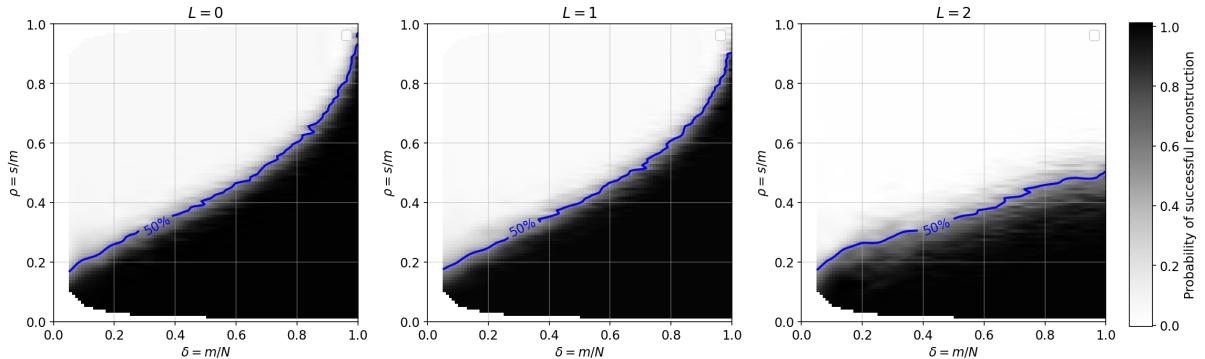


Figure 3.6.: Probability phase transitions for $L = 0, 1, 2$.

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 [17], since the theory typically establishes worst-case bounds for recovery of all s -sparse signals rather than a subset of the signals. In the following, we briefly examine how numerical results are affected when constraints on the signal \mathbf{x} are imposed, after which we extend our analysis 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 nonzero signal entries. We expect signals whose nonzeros 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{G}_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. These results indicate that one can overcome the smoothing effect by imposing a separation distance constraint in the signal.

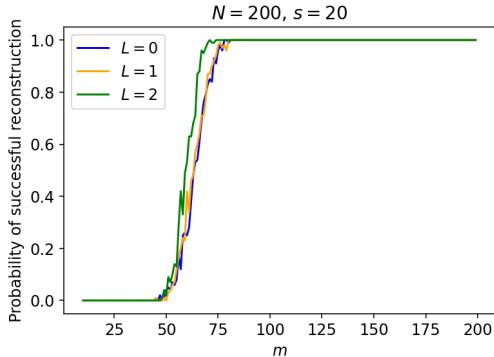


Figure 3.7.: p_{success} across number of measurements for $L = 0, 1, 2$, considering only signals \mathbf{x}_L .

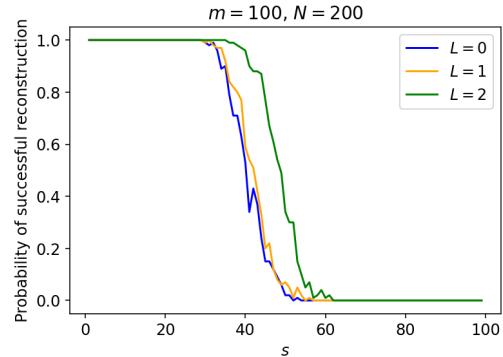


Figure 3.8.: p_{success} across sparsity of the ground truth for $L = 0, 1, 2$, considering only signals \mathbf{x}_L .

Next, we remove the separation distance constraint and consider nonnegative – one-sided Gaussian – signals $\mathbf{x}_{\geq 0}$, where nonzeros 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 $\mathbf{x} \geq 0$ to `constraints` in CVXPY; apart from this modification, we keep the exact same reconstruction scheme used thus far. According to the literature, the outcomes should improve [19]. 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.

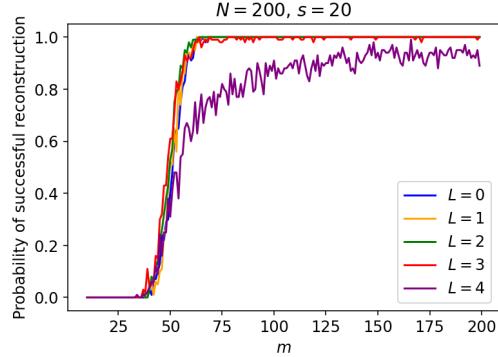


Figure 3.9.: p_{success} across number of measurements for varying L , considering only signals $\mathbf{x} \geq 0$.

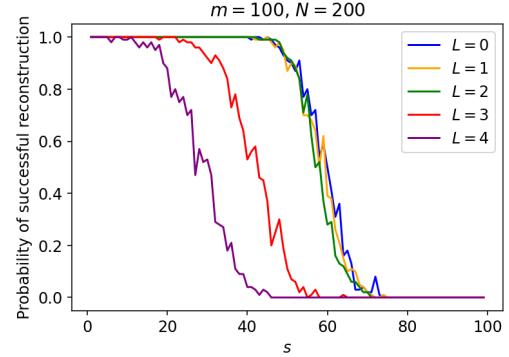


Figure 3.10.: p_{success} across sparsity of the ground truth for varying L , considering only signals $\mathbf{x} \geq 0$.

3.4. Two-dimensional setting

This section extends the model to n -by- n speckle patterns by modifying the RBF kernel to smooth in two directions. We verify that the correlation induced by the model coincides with that underlying speckle patterns simulated from a multimode fiber, to then estimate the smoothing length in fiber-simulated patterns and assess the extent to which it affects performance.

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 pixels. Since the kernel induces correlations directly from the coordinate system underlying the two-dimensional array, it automatically accounts for the array's boundaries and thus requires no separate boundary conditions. As illustrated in Figures 3.12(a) and 3.12(d), the matrix $\tilde{\Sigma} \in \mathbb{R}^{n^2 \times n^2}$ is symmetric with decaying values around a unit diagonal, with each row i containing the correlations of pixel i with all other pixels. 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, and to applying `gaussian_filter` from Python's `scipy.ndimage`, with the additional benefit of direct control over the extent of smoothing in terms of pixels.

Next, we simulate speckle patterns from a multimode fiber by specifying a fiber object and light wave in the Python module pyMMF [34], which then calculates the collection of mode profiles $\{\Psi_k(x, y)\}_{k=1}^M$. Given the profiles, we compute the speckle field (2.1) in a simplified way by randomly selecting $E_k \in [0, 1]$ and $\phi_k \in [-\pi, \pi]$ to obtain intensities $|E_{out}|^2$ as displayed in Figure 3.11(c). Since pyMMF calculates the mode profiles by default for a round fiber, we apply a square mask to the circular pattern; although the module offers the possibility to use square fibers, we found that the mode computation is limited to only six modes due to a bug.

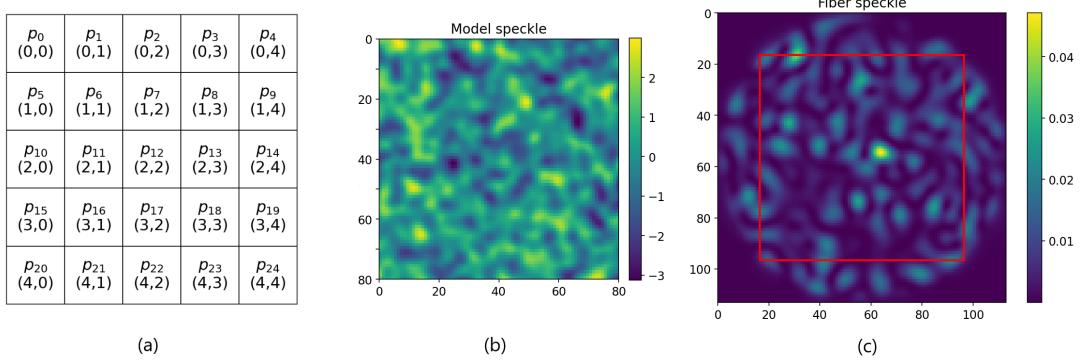


Figure 3.11.: (a) Coordinate system underlying the distance metric in the 2D RBF kernel for a 5-by-5 array. (b) 80-by-80 pattern obtained by applying \tilde{S}_L with $L = \sqrt{13}$ to a standard Gaussian baseline. (c) Simulated pattern from propagating light with wavelength $1 \mu\text{m}$ through 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.

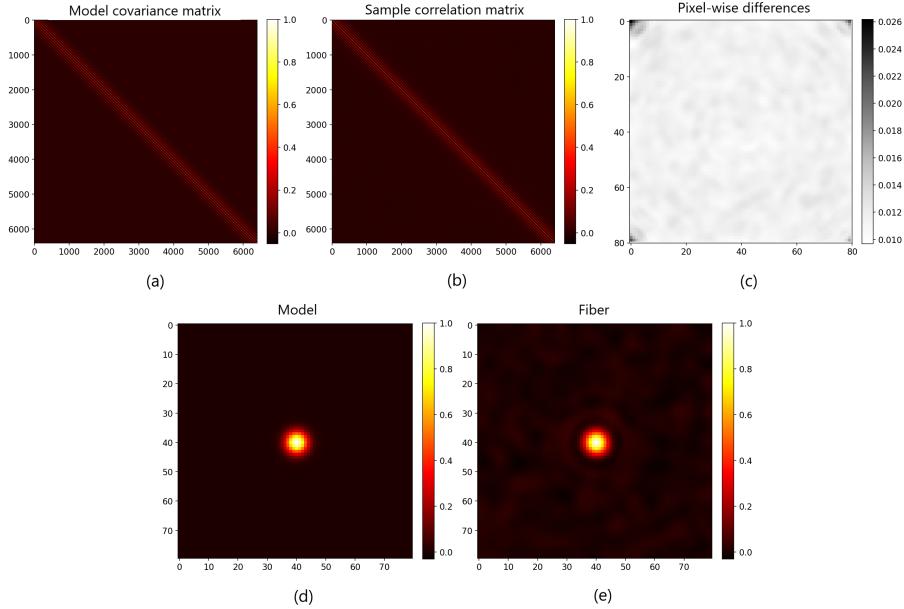


Figure 3.12.: (a) 2D RBF kernel covariance matrix with $n^2 = 80^2$ and $L = \sqrt{13}$; (b) sample correlation matrix from 10^4 cropped fiber-simulated 80-by-80 patterns; (c) pixel-wise average differences (3.4) as gray values in an 80-by-80 array; (d) $\tilde{\Sigma}_i$ corresponding to the central pixel $i = n/2 \cdot n + n/2$ reshaped into an 80-by-80 grid; (e) $\bar{\Sigma}_i$ corresponding to the central pixel $i = n/2 \cdot n + n/2$ reshaped into an 80-by-80 grid.

To verify that the correlation induced by the model resembles that underlying the example 80-by-80 pattern in Figure 3.11(c), we fix $\{\Psi_k(x, y)\}_{k=1}^M$ and generate 10^4 cropped speckle fields with the values of E_k, ϕ_k chosen at random. Next, we flatten every field and calculate the sample correlation matrix $\bar{\Sigma}$ of the collection of random vectors. Each entry $\bar{\Sigma}_{ij}$ estimates the correlation between pixels i and j in the fiber-simulated pattern, and hence carries the same meaning as $\tilde{\Sigma}_{ij}$. As shown in Figures 3.12(b) and 3.12(e), the shape of $\bar{\Sigma}$ resembles that of $\tilde{\Sigma}$, so we may estimate the length of the smoothing by minimizing the Frobenius norm $\|\tilde{\Sigma} - \bar{\Sigma}\|_F := \sqrt{\sum_{i=0}^{n^2-1} \sum_{j=0}^{n^2-1} (\tilde{\Sigma}_{ij} - \bar{\Sigma}_{ij})^2}$ with respect to L . Basic minimization using `scipy.optimize` yields $L \approx \sqrt{13}$. To confirm that the model and fiber correlations coincide for $L \approx \sqrt{13}$, we display in Figure 3.12(c) for each pixel i the average difference

$$\sqrt{\frac{1}{n^2} \sum_{j=0}^{n^2-1} (\tilde{\Sigma}_{ij} - \bar{\Sigma}_{ij})^2} \quad (3.4)$$

as a gray value and observe that these values are close to zero. We note that the corners show slightly higher values, likely because those areas generally contain less correlation, making imprecisions in the sample correlation more pronounced.

Finally, let us assess the impact of this smoothing with numerical experiments analogous to those in Section 3.3. To reduce the computational burden, we downscale the problem from $n \times n = 80 \times 80$ to $n \times n = 20 \times 20$ and study the effect of smoothing over $L = \frac{\sqrt{13}}{4} \approx 1$. We generate one-sided Gaussian s -sparse ground truths $\mathbf{x}_{\geq 0} \in \mathbb{R}^{n^2}$, measurement matrices $\mathbf{A} = \mathbf{G}\tilde{\mathbf{S}}_L \in \mathbb{R}^{m \times n^2}$ with \mathbf{G} standard Gaussian, and noise-free measurements $\mathbf{y} = \mathbf{Ax}_{\geq 0} \in \mathbb{R}^m$. Using CVXPY's BPDN with `x>=0` added to `constraints`, a varying $\eta = 10^{-3} \|\mathbf{Ax}_{\geq 0}\|_2$, 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 3.13 and 3.14 respectively. Since smoothing over $L = 1$ does not affect performance, we conclude that the smoothing effect is negligible for the example patterns in Figures 3.11(b) and 3.11(c). In other words, as long as the speckle's size is sufficiently small, which is achieved when the fiber excites sufficiently many modes, the pattern's smooth structure poses no concern. 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 the intensities for the model pattern follow a mean-zero normal distribution, while those for the fiber pattern are exponentially distributed. Therefore, in the following chapter we will refine our model to follow specific target distributions.

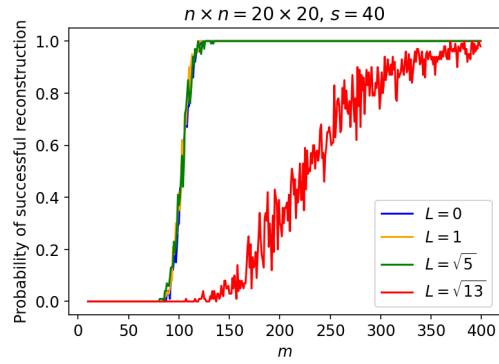


Figure 3.13.: p_{success} across the number of measurements for varying L .

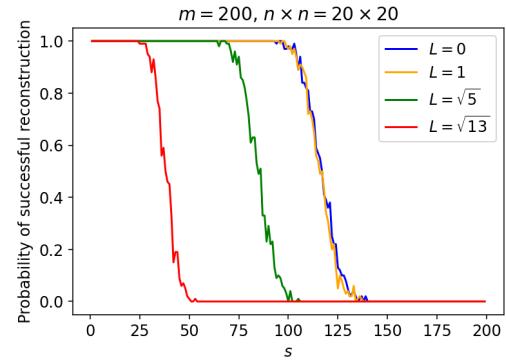


Figure 3.14.: p_{success} across the sparsity of the ground truth for varying L .

4. Nonnegative measurement matrices

In this chapter, we first analyze from a theoretical perspective the validity of random exponential matrices $\mathbf{A} \in \mathbb{R}^{m \times N}$ composed of i.i.d. entries $a_{ij} \sim \text{Exp}(\lambda)$ with rate parameter $\lambda > 0$. Following a procedure similar to that for 0/1-Bernoulli matrices [27] and shifted Gaussian matrices [36], we address the broad idea behind showing the ℓ_2 -robust nullspace property in Section 4.1.1 and prove the \mathcal{M}^+ -criterion in Section 4.1.2. The remainder of the chapter then studies measurement matrices constructed from n -by- n patterns, which contain a smoothing induced by the 2D RBF kernel as well as exponentially distributed intensities. Moreover, as it is physically possible to customize speckle intensity distributions, we simultaneously consider the other distributions displayed in Figure 4.1. Section 4.2.1 describes how we generate patterns with an assigned smoothing length and intensity distribution, followed by a numerical experiment in Section 4.2.2, which will resolve whether the optical experiment could benefit from customized light.

4.1. Random exponential matrices

4.1.1. ℓ_2 -robust nullspace property

Recall from Section 2.2.2 that establishing the ℓ_2 -robust NSP amounts to bounding the marginal tail function and the mean empirical width, given by

$$Q_\xi(T_{\rho,s}, \mathbf{a}) = \inf_{\mathbf{u} \in T_{\rho,s}} \mathbb{P}(|\langle \mathbf{a}, \mathbf{u} \rangle| \geq \xi) \quad \text{and} \quad W_m(T_{\rho,s}, \mathbf{a}) = \mathbb{E} \left(\sup_{\mathbf{u} \in T_{\rho,s}} \langle \mathbf{h}, \mathbf{u} \rangle \right)$$

respectively, where $\mathbf{a} \in \mathbb{R}^N$ now represents a random exponential vector. Generally, the lower bound on Q_ξ is proven using the Paley-Zygmund inequality – known as $\mathbb{P}(R > t) \geq (\mathbb{E}[R] - t)^2 / \mathbb{E}[R^2]$ with $0 \leq t \leq \mathbb{E}[R]$ and R any nonnegative random variable – in combination with moment calculations [17]. Since these moment calculations are rather lengthy, we leave the lower bound’s formal derivation for future research.

For the upper bound on W_m , it is common to use that $W_m(T_{\rho,s}, \mathbf{a}) \leq \frac{3}{\rho} W_m(\Sigma_s^2, \mathbf{a})$, where $\Sigma_s^2 := \{\mathbf{v} \in \mathbb{R}^N : \|\mathbf{v}\|_0 \leq s, \|\mathbf{v}\|_2 = 1\}$ is the collection of s -sparse unit vectors, and bound $W_m(\Sigma_s^2, \mathbf{a})$ in terms of N , m and s . As detailed in Section IV-B of [27], the inequality holds because $T_{\rho,s} \subset \frac{3}{\rho} \text{conv}(\Sigma_s^2)$ (with ‘conv’ denoting the convex hull) and because the supremum of $\langle \mathbf{h}, \mathbf{u} \rangle$ over $\text{conv}(\Sigma_s^2)$ is attained at Σ_s^2 . Again, due to computational complexity, we refrain from analytical derivations; instead, we numerically estimate the quantity $W_m(\Sigma_s^2, \mathbf{a})$ for varying values of m and s .

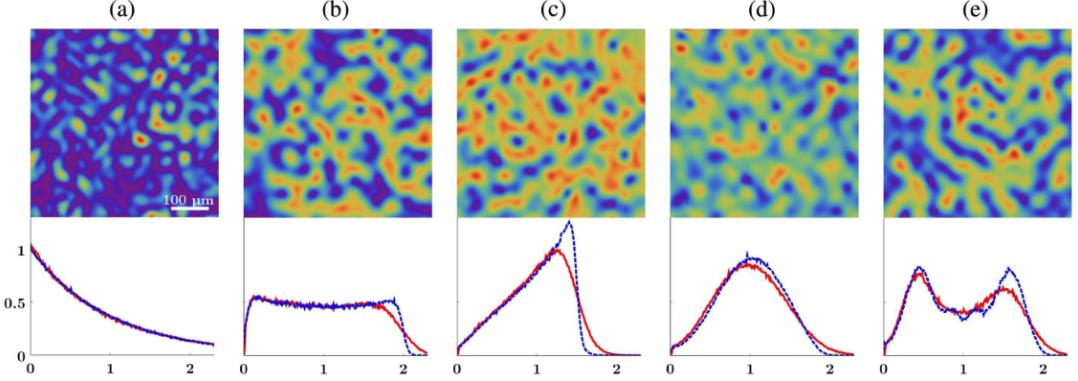


Figure 4.1.: $504\text{-by-}504 \mu\text{m}$ samples with (a) an exponential intensity distribution and (b)-(e) customized intensity distributions, together with their density functions from experimental data (red) and numerical simulations (blue) [11].

More specifically, we fix $N = 100$, $\lambda = 1/2$, vary $m = 1, 2, \dots, 100$, $s = 1, 2, \dots, 100$, and employ a greedy approach for the approximation. For every value of m and s , we randomly generate a Rademacher sequence $\{\epsilon_k\}_{k=1}^m$ of values ± 1 and an exponential matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, based on which we calculate the Rademacher-weighted sum $\mathbf{h} = \frac{1}{\sqrt{m}} \sum_{k=1}^m \epsilon_k \mathbf{a}_k$ of the matrix's rows \mathbf{a}_k . Next, instead of computing $\langle \mathbf{h}, \mathbf{u} \rangle$ for all $\mathbf{u} \in \Sigma_s^2$ to find the supremum, we consider only $\mathbf{u}_{opt} \in \Sigma_s^2$ that produces the largest inner product for the given \mathbf{h} . Hereto we retrieve the index set $S_{opt} \subset \{1, \dots, N\}$ of the s largest (in modulus) components of \mathbf{h} and define the entries of \mathbf{u}_{opt} as $u_i := \text{sgn}(h_i) \mathbf{1}_{i \in S_{opt}} / \|\mathbf{u}_{opt}\|_2$. Based on 100 randomly generated \mathbf{h} for each combination of m and s , we estimate $\hat{W}_m(\Sigma_s^2, \mathbf{a})$ as the average of the inner products $\langle \mathbf{h}, \mathbf{u}_{opt} \rangle$ to obtain the results in Figure 4.2. For comparison, we include the approximations for Bernoulli matrices, whose mean empirical width is known to admit an upper bound. We see that the approximations do not blow up (except for extremely low measurement numbers) and follow a similar shape to the Bernoulli estimates. Furthermore, although we only display the outcomes at $\lambda = 1/2$, we noticed that larger values of λ led to larger $\hat{W}_m(\Sigma_s^2, \mathbf{a})$ and vice versa, but that the overall shape of the curves was maintained.

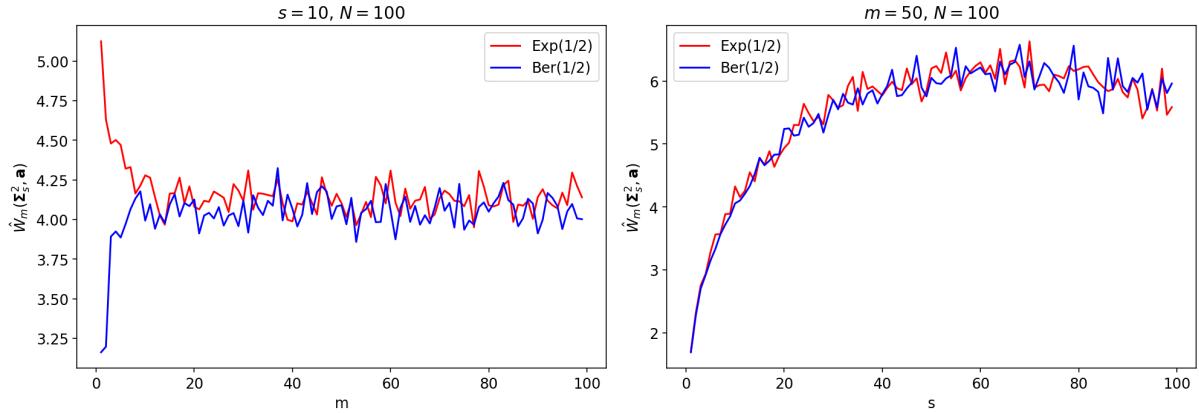


Figure 4.2.: $\hat{W}_m(\Sigma_s^2, \mathbf{a})$ for $\text{Exp}(1/2)$ and $\text{Ber}(1/2)$ matrices.

4.1.2. \mathcal{M}^+ -criterion

To establish the \mathcal{M}^+ -criterion for exponential matrices, we set $\lambda = 1/\langle I \rangle$ with $\langle I \rangle$ a parameter representing the average light intensity, so that $\mathbb{E}(a_{ij}) = \langle I \rangle$ for each entry. Furthermore, the following corollary – a particular concentration inequality called the Bernstein inequality for subexponential random variables – will be central to our proof:

Corollary 2 (Corollary 7.32 of [20]). *Let c_1, \dots, c_m be independent mean-zero subexponential random variables, i.e. $\mathbb{P}(|c_k| \geq t) \leq \beta e^{-\kappa t}$ for some constants $\beta, \kappa > 0$ and for all $t \geq 0$, $k = 1, \dots, m$. Then*

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

There are different variations of this corollary depending on the probability distribution under consideration. For example, the proofs of the \mathcal{M}^+ -criterion for Bernoulli and shifted Gaussian matrices use versions that assume bounded and sub-Gaussian random variables respectively (cf. Corrolary 3 and Theorem A.3 in the appendix).

Let us now state and prove our result:

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} fulfills the \mathcal{M}^+ -criterion with probability $\geq 1 - 2N \exp(-em/(4e^2 + 4e + 16))$.*

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 $\mathbf{t} := \frac{1}{\langle I \rangle m} \mathbf{1}_m$. From 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 establish that \mathbf{w} is strictly positive, we show that it is concentrated around its nonzero mean. More specifically, we prove that w.h.p. $|w_i - 1| < \frac{1}{2}$ for all $i = 1, \dots, N$ by using Corollary 2 to bound $\mathbb{P}(|w_i - 1| \geq \frac{1}{2})$ from above.

First, we use that $|w_i - 1| = |\langle \mathbf{e}_i, \mathbf{w} \rangle - 1|$ with \mathbf{e}_i the i th standard basis vector in \mathbb{R}^N , $\mathbf{w} = \frac{1}{\langle I \rangle m} \sum_{k=1}^m \mathbf{a}_k$ with \mathbf{a}_k the k th row of \mathbf{A} and $\langle I \rangle m = \sum_{k=1}^m \mathbb{E}(a_{ki})$ to write

$$|\langle \mathbf{e}_i, \mathbf{w} \rangle - 1| = \left| \frac{1}{\langle I \rangle m} \sum_{k=1}^m \langle \mathbf{e}_i, \mathbf{a}_k \rangle - \frac{\langle I \rangle m}{\langle I \rangle m} \right| = \frac{1}{\langle I \rangle m} \left| \sum_{k=1}^m (a_{ki} - \mathbb{E}(a_{ki})) \right|. \quad (4.1)$$

We then define $c_k := a_{ki} - \mathbb{E}(a_{ki})$; since the a_{ki} are independent and exponentially distributed, the c_k constitute a set of independent mean-zero subexponential random variables. Thus, we can apply Corrolary 2.

To find the expressions of β and κ in $\mathbb{P}(|c_k| \geq t) \leq \beta e^{-\kappa t}$, we write $\mathbb{P}(|c_k| \geq t) = \mathbb{P}(a_{ki} \geq t + \langle I \rangle) + \mathbb{P}(a_{ki} \leq -t + \langle I \rangle)$ and distinguish the following cases:

1. For $t > \langle I \rangle$, the term $\mathbb{P}(a_{ki} \leq -t + \langle I \rangle)$ vanishes, and so

$$\mathbb{P}(|c_k| \geq t) = \mathbb{P}(a_{ki} \geq t + \langle I \rangle) = \int_{t+\langle I \rangle}^{\infty} \frac{1}{\langle I \rangle} e^{-\frac{x}{\langle I \rangle}} dx = e^{-1} e^{-\frac{t}{\langle I \rangle}}.$$

2. For $0 \leq t \leq \langle I \rangle$, we have

$$\mathbb{P}(|c_k| \geq t) = \int_{t+\langle I \rangle}^{\infty} \frac{1}{\langle I \rangle} e^{-\frac{x}{\langle I \rangle}} dx + \int_0^{-t+\langle I \rangle} \frac{1}{\langle I \rangle} e^{-\frac{x}{\langle I \rangle}} dx = e^{-1} e^{-\frac{t}{\langle I \rangle}} - e^{-1} e^{\frac{t}{\langle I \rangle}} + 1,$$

which we claim to be bounded from above by $M e^{-\frac{t}{\langle I \rangle}}$ with $M > 0$ a constant. Indeed, the claim holds by the extreme value theorem, as $f(t) := e^{-1} - e^{-1} e^{\frac{2t}{\langle I \rangle}} + e^{\frac{t}{\langle I \rangle}}$ is a continuous function on the closed and bounded interval $[0, \langle I \rangle]$ and thus attains a maximum M . Basic calculus shows that M is given by the critical value $e^{-1} + e/4$, since $f''(t) < 0$ on $[0, \langle I \rangle]$ and the critical value exceeds the function's boundary values.

From Cases 1 and 2 it follows that $\mathbb{P}(|c_k| \geq t) \leq (e^{-1} + e/4)e^{-\frac{t}{\langle I \rangle}}$ for all $t \geq 0$, i.e. $\beta = e^{-1} + e/4$ and $\kappa = \frac{1}{\langle I \rangle}$.

Lastly, using (4.1) and applying Corrolary 2 results in

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

union bounding then confirms that $|w_i - 1| < \frac{1}{2}$ holds simultaneously for all $i = 1, \dots, N$ with probability $\geq 1 - 2N \exp(-em/(4e^2 + 4e + 16))$. \square

4.2. Other distributions

4.2.1. Model description

A key advantage of the normally distributed baselines \mathbf{g} used in Chapter 3 is their closure under linear combinations, ensuring that the products $\mathbf{g}\tilde{\mathbf{S}}_L$ remain 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. Therefore, we maintain the smoothing of Gaussian baselines to obtain patterns as in Figure 3.11(b), whose distributions we then transform to be exponential, uniform, triangular, shifted normal and bimodal.

The transformation is conducted by means of the *inversion method* [15], 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 $U \sim \text{Unif}[0, 1]$. 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

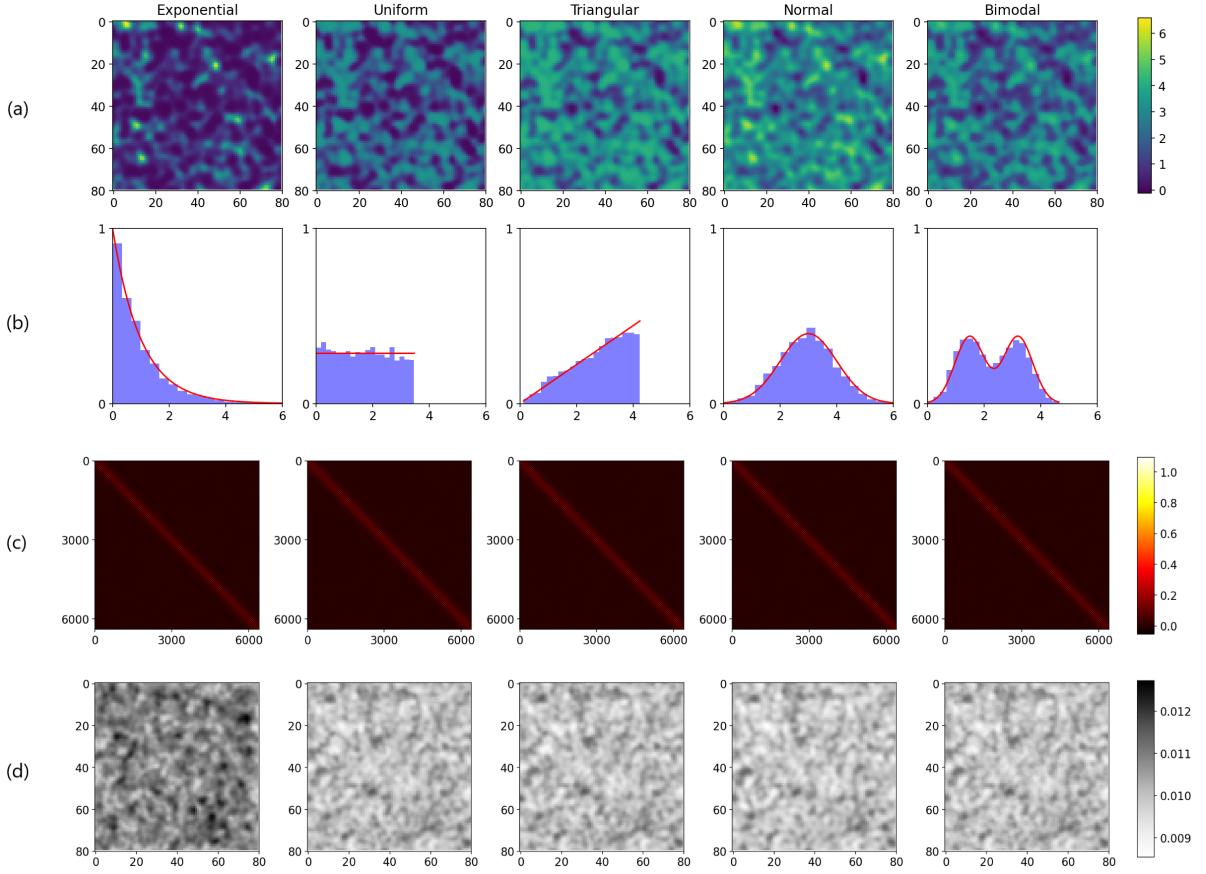


Figure 4.3.: (a) 80-by-80 patterns formed by applying $F_R^{-1} \circ \Phi$ entry-wise to $\mathbf{g}\tilde{\mathbf{S}}_L$ with $L = \sqrt{13}$. (b) Histograms of the intensities overlaid with the target density functions. (c) Sample covariance matrices for each pattern. (d) Pixel-wise average differences (3.4) between each sample covariance matrix and the 2D RBF kernel covariance matrix. (*Distributions from left to right*) $\text{Exp}(\lambda = 1)$, $\text{Unif}[a = 0, b = \sqrt{12}]$, $\text{Triangular}(a = 0, b = \sqrt{18}, c = \sqrt{18})$, $\mathcal{N}(\mu = 3, \sigma^2 = 1)$, bimodal according to (4.2).

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. Figures 4.3(a) and 4.3(b) display the outcomes of this procedure.

To ensure a fair comparison of the different distributions' performance in Section 4.2.2, we set the parameters such that each target distribution has unit variance. Hereto we use that $\text{Var}(R) = 1/\lambda^2$ for $R \sim \text{Exp}(\lambda)$, $\text{Var}(R) = (b - a)^2/12$ for $R \sim \text{Unif}[a, b]$, and $\text{Var}(R) = (a^2 + b^2 + c^2 - ab - ac - bc)/18$ for R a triangular random variable with minimum value a , maximum value b and mode c . For the normal distribution, we pick $\mu = 3$ so that the majority of the density (around 99.7%) is nonnegative. The bimodal distribution is treated in the following paragraph as it requires more explanation. Furthermore, to

verify that the transformation does not distort the smoothing, we generate 10^4 patterns for each distribution and calculate the resulting sample covariance matrices $\bar{\Sigma}$ as well as the pixel-wise average differences (3.4). Indeed, Figures 4.3(c) and 4.3(d) confirm that the sample covariance matrices maintain the shape of $\bar{\Sigma}$ and that the pixel-wise average differences are low. It is worth noting that although the distributions have a common variance, their expected values $\mathbb{E}(R)$ differ. For now, we do not consider this difference as problematic, since it has been shown for Gaussian matrices that the value of the mean does not influence performance [36]. Moreover, we remark that negative values in the normal and bimodal patterns can be removed through resampling.

Bimodal distribution

Since Python does not have a built-in bimodal distribution, we construct the density as a mixture of two normal distributions:

$$f(x) = 0.5g_1(x) + 0.5g_2(x) \quad (4.2)$$

with g_1 and g_2 the densities of $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ respectively. We then choose the values of μ_1 , μ_2 and σ such that the density has two peaks and unit variance.

The parameter values are determined as follows. Assume that $\mu_2 > \mu_1 > 0$ and use that two peaks occur when $\sigma \leq (\mu_1 - \mu_2)/2$ [35]. This requirement can be expressed equivalently as $\sigma = c \cdot d/2$ with $d = \mu_2 - \mu_1 > 0$ and $0 < c \leq 1$ a constant. Then use that the variance of f is given by $\sigma_f^2 = (\mu_2 - \mu_1)^2/4 + \sigma^2$ [26] to write

$$\sigma_f^2 = (1 + c^2) \cdot \frac{d^2}{4}.$$

Thus, the problem of finding suitable μ_1 , μ_2 and σ amounts to solving $(1 + c^2) \cdot d^2/4 = 1$ with $d > 0$ and $0 < c \leq 1$, which yields

$$c = \sqrt{\frac{4 - d^2}{d^2}} \text{ and } \sqrt{2} < d < 2.$$

In other words, for a given location μ_1 of the first peak and distance d between the peaks, the density requires $\mu_2 = \mu_1 + d$ and $\sigma = c \cdot d/2 = \frac{1}{2}\sqrt{4 - d^2}$. For the bimodal density in Figure 4.3(a), we choose $d = (\sqrt{2} + 2)/2$ and $\mu_1 = 1.5$.

Given the values of μ_1 , μ_2 and σ , the distribution function $F(x) = 0.5\Phi_1(x; \mu_1, \sigma) + 0.5\Phi_2(x; \mu_2, \sigma)$ is directly implemented in Python in terms of the built-in Gaussian distribution functions. Finally, since a closed-form expression is not immediate, the quantile function F^{-1} is numerically approximated through linear interpolation over 1000 values of the distribution function.

4.2.2. Numerical results

Having constructed differently distributed patterns with a common variance and smoothing length, let us examine how well they perform in recovering signals. For every distribution F_R in Figure 4.3, we generate measurement matrices $\mathbf{A} = (F_R^{-1} \circ \Phi)(\mathbf{G}\tilde{\mathbf{S}}_L) \in \mathbb{R}^{m \times n^2}$

of m flattened $n \times n = 20 \times 20$ patterns with $L = 1$. The s -sparse ground truths $\mathbf{x}_{\geq 0} \in \mathbb{R}^{n^2}$ are one-sided Gaussian and the measurements $\mathbf{y} = \mathbf{Ax}_{\geq 0} \in \mathbb{R}^m$ noise-free. For varying $m = 10, \dots, 400$ and $s = 1, \dots, 200$, we use CVXPY's BPDN with $\eta = 10^{-6} \|\mathbf{Ax}_{\geq 0}\|_2$ and the additional argument $\mathbf{x} \geq 0$ in `constraints`. Calling $\hat{\mathbf{x}}$ with NMSE (3.3) at most 0.05 successful, we obtain the outcomes in Figures 4.4 and 4.5. As a sanity check, we also present the outcomes for the mean-zero normal distribution ($\mathbf{A} = \mathbf{G}\tilde{\mathbf{S}}_L$) to confirm that altering the mean does not affect performance of the normal distribution. We conclude that each distribution F_R performs equally well in reconstructing arbitrary sparse signals. This finding, however, does not exclude the possibility that customized intensities could be beneficial for specific samples, e.g. bimodal intensities for the sample in Figure 2.2(a).

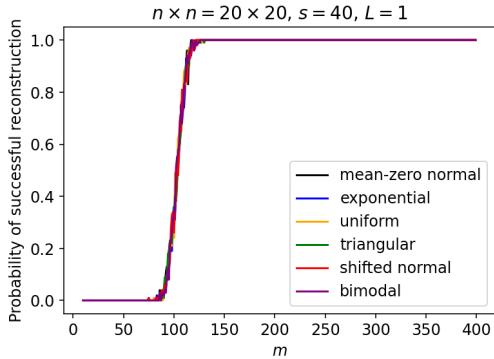


Figure 4.4.: p_{success} across the number of measurements for varying distributions.

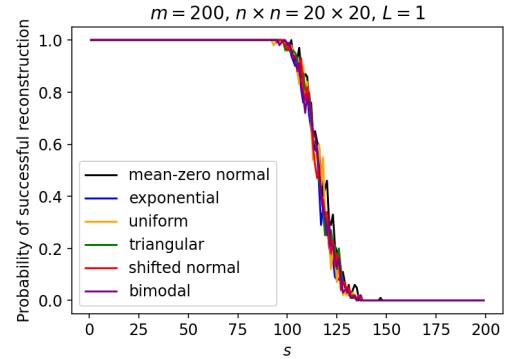


Figure 4.5.: p_{success} across the sparsity of the ground truth for varying distributions.

CVXPY versus SPGL1

It is worth mentioning that we initially used PyLops' SPGL1 algorithm, designed for large-scale one-norm regularized least-squares [12], for the numerical experiments. However, in cases where the matrix contained relatively high correlations or large values, the algorithm's outcomes were inaccurate due to a linesearch error. Figure 4.6, a numerical experiment for shifted Gaussian matrices replicated from the literature (cf. Figure 3 of [36]), illustrates the inaccurate results. For Gaussian matrices with a smaller mean of 5, it was possible to overcome the issue by adjusting SPGL1's optimality tolerances to `opt_tol=1e-6*eta` and `bp_tol=1e-6*eta`. It remains an open question how well CVXPY performs in recovering images of size at least 256-by-256 pixels and what algorithm would be optimal.

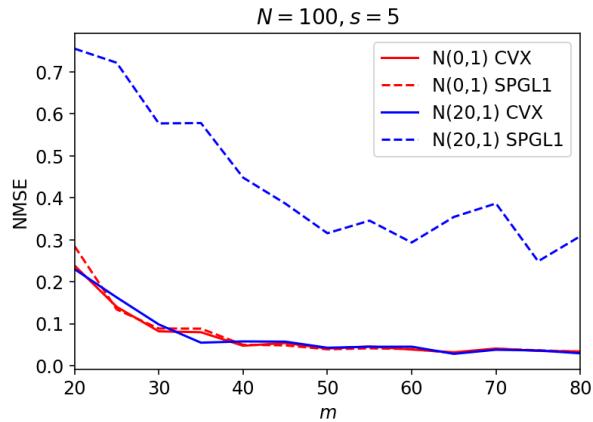


Figure 4.6.: Reconstruction errors for Gaussian matrices with CVXPY and SPGL1.

5. Conclusion

This thesis interpreted MMFCI through the mathematical framework of compressed sensing, with the aim of establishing that the experiment aligns with CS theory and exploring possible improvements. In particular, we focused on the effect of correlation in the measurement matrix, the validity and performance of various intensity distributions, and alternatives to conventional ℓ_1 -optimization. Our findings are as follows.

Sections 3.1-3.3 considered the smoothing of one-dimensional speckle patterns. We created a model that induces correlations over L entries in the rows of a random Gaussian matrix by applying a smoothing operator resulting from Cholesky decomposing an RBF kernel covariance matrix. We then proved Proposition 1 on the RIP of a Gaussian matrix multiplied by a deterministic matrix and calculated the mutual coherence of the smoothing operator, in order to derive the theoretical upper bound (3.2) on the smoothing length. By reconstructing randomly generated sparse vectors from noise-free measurements simulated with our model across varying measurement numbers, sparsity levels and correlation lengths, we identified a gap between numerical results and the theoretical bound. Furthermore, numerical experiments demonstrated that the correlation effect can be overcome by imposing a separation distance constraint on the signal.

In Section 3.4, we extended the model to two-dimensional speckle patterns by adapting the RBF kernel, and found that the correlations induced by the model accurately represent those underlying patterns from fiber simulations. While simulating speckle from fibers, we discovered a bug in the pyMMF module that restricts mode computation for square fibers. Lastly, we estimated the correlation length in fiber speckle patterns, and concluded from numerical experiments that the smoothing effect is negligible as long as the speckle size is sufficiently small.

Section 4.1 analyzed random exponential matrices from a theoretical perspective. Compressive sensing literature shows that the RIP is inconvenient for nonnegative matrices and that instead a weaker property – the ℓ_2 -robust NSP – can be used. We outlined the broad idea behind establishing the ℓ_2 -robust NSP for exponential matrices, in particular by numerically approximating the mean empirical width. Moreover, CS theory demonstrates that BPDN can be replaced by the simpler NNLS scheme provided that the signal is nonnegative and the matrix satisfies the \mathcal{M}^+ -criterion in addition to the ℓ_2 -robust NSP. Motivated by this result, we proved Theorem 1 on the \mathcal{M}^+ -criterion for exponential matrices. These theoretical considerations suggest NNLS as a promising alternative to ℓ_1 -optimization for scenarios where MMFCI reconstructs images that are sparse in the original domain.

In Section 4.2, we explored the performance of measurement matrices constructed from patterns with different intensity distributions. We generated the patterns in Figure 4.3(a) by applying the inversion method to the model of Section 3.4, allowing us to generate two-dimensional speckle with a specified smoothing length and intensity distribution. Upon verifying that the smoothing length is preserved under the transformation, we conducted numerical experiments and found that the intensity distributions perform equally well in recovering arbitrary sparse signals. Moreover, during the experiments we discovered that Python’s SPGL1 algorithm yields linesearch errors in cases where the measurement matrix contains relatively high correlations or large values.

There remains much to explore in the future. For example, formal bounds on the marginal tail function and mean empirical width could be derived to further solidify the theoretical basis of the experiment. From an applied standpoint, it is valuable to study how well different intensity distributions recover specific two-dimensional samples, with the model of Section 4.2 offering a solid foundation. Moreover, examining the effect of noise, particularly the potential benefit of NNLS as it does not require a noise bound, constitutes a noteworthy research direction. As mentioned in the discussion on SPGL1, it is still an open question which algorithm is optimal. Another promising direction is the prediction of speckle patterns with machine learning to further reduce the measurement number. Lastly, it is of great relevance – especially for metrology applications involving checkerboard-like images – to study the recovery of signals that are not sparse in the original domain but, for instance, in the wavelet domain.

Acknowledgments

I would like to express my thanks to everyone involved in the project for making my first real encounter with research a rewarding experience. It was interesting to witness how expertise from various backgrounds can be integrated to explore shared questions. First and foremost, thank you to my supervisors dr. Svetlana Dubinkina and prof.dr. Tristan van Leeuwen for the guidance as well as the freedom you provided. Your curiosity in our weekly discussions and seeing the way you navigate through papers was very motivating and insightful. Also many thanks to dr. Lyubov Amitonova; your experimentalist perspective during our monthly meetings encouraged me a lot to work out the mathematics. Special thanks to dr. Sjoerd Dirksen for introducing the ℓ_2 -robust nullspace property and our discussion at Utrecht University. I would also like to thank Alexander Skorikov for helping with the SPGL1 coding issues, as well as dr. Lyubov Amitonova's PhD students: Aleksandra Ivanina for showing the ARCNL lab and explaining the experiment, and Kian Goeloe and Maximilian Lipp for helping with fiber speckle simulation and looking into the pyMMF bug. Lastly, of course, I thank my family and friends.

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] Amir Beck and Marc Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In: *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202.
- [11] Nicholas Bender et al. “Customizing speckle intensity statistics”. In: *Optica* 5.5 (2018), pp. 595–600.
- [12] 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.
- [13] Emmanuel J. Candès. “The restricted isometry property and its implications for compressed sensing”. In: *Comptes Rendus Mathematique* 346.9–10 (2008), pp. 589–592.
- [14] Mark A. Davenport et al. “Introduction to compressed sensing”. In: *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012, pp. 1–64.

- [15] Gerhard Derflinger, Wolfgang Hörmann, and Josef Leydold. “Random variate generation by numerical inversion when only the density is known”. In: *ACM Trans. Model. Comput. Simul.* 20.4 (2010), pp. 1–25.
- [16] Steven Diamond and Stephen Boyd. “CVXPY: A Python-embedded modeling language for convex optimization”. In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.
- [17] Sjoerd Dirksen, Guillaume Lecué, and Holger Rauhut. “On the gap between RIP-properties and sparse recovery conditions”. In: *IEEE Transactions on Information Theory* (2016).
- [18] 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.
- [19] David L. Donoho and Jared Tanner. “Precise undersampling theorems”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 913–924.
- [20] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [21] Joseph W. Goodman. *Speckle phenomena in optics: theory and applications, Second edition*. SPIE, 2020.
- [22] 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.
- [23] Apostolos Karadimitrakis, Mérouane Debbah, and Aris Moustakas. “Optical MIMO: Results and analysis”. In: *11th International Symposium on Wireless Communications Systems*. Barcelona, Spain, 2014, pp. 966–970.
- [24] Shiva P. Kasiviswanathan and Mark Rudelson. “Restricted isometry property under high correlations”. In: *arXiv* (2019).
- [25] 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.
- [26] Tae-Hwan Kim and Halbert White. “On more robust estimation of skewness and kurtosis”. In: *Finance Research Letters* 1.1 (2004), pp. 56–73.
- [27] Richard Kueng and Peter Jung. “Robust nonnegative sparse recovery and the nullspace property for 0/1 measurements”. In: *IEEE Transactions on Information Theory* 64.2 (2018), pp. 689–703.
- [28] Heung-No Lee. “*Introduction to compressed sensing: with coding theoretic perspective*”, *Lecture notes*.
- [29] Tristan van Leeuwen and Christopher Brune. “*Inverse problems and imaging*”, *Lecture notes*. 2025. URL: https://tristanvanleeuwen.github.io/IP_and_Im_Lectures/intro.html.
- [30] 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.

- [31] Zhouping Lyu. “High-resolution imaging through a multimode fiber: from raster-scanning to compressive sensing”. Doctoral Dissertation. VU Amsterdam, 2024.
- [32] Fedor Mitschke. *Fiber optics: physics and technology, second edition*. Springer, 2016.
- [33] Rudiger Paschotta. *Field guide to lasers*. SPIE Press, 2008.
- [34] Sébastien M. Popoff. *pyMMF*. 2025. URL: <https://github.com/wavefrontshaping/pyMMF>.
- [35] Mark F. Schilling, Ann E. Watkins, and William Watkins. “Is human height bimodal?” In: *The American Statistician* 56.3 (2002), pp. 223–229.
- [36] Yonathan Shadmi, Peter Jung, and Giuseppe Caire. “Sparse non-negative recovery from biased subgaussian measurements using NNLS”. In: *IEEE International Symposium on Information Theory* (2019), pp. 2079–2083.
- [37] James H. Steiger. “*Random vectors, random matrices, and their expected Values*”, *Lecture notes*.
- [38] 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>.
- [39] Yusuke Yamada. “Textile-integrated polymer optical fibers for healthcare and medical applications”. In: *Biomedical Physics Engineering Express* 6 (2020), p. 062001.

A. Appendix

A.1. Restricted isometry property for random matrices

Lemma A.1 (Lemma 5.1 of [9]). *Let $\mathbf{A} \in \mathbb{R}^{m \times N}$ 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 $|Q_S| \leq (\frac{12}{\delta})^s$. 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

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

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 $\|\mathbf{Ax}\|_2 \leq (1 + \delta^*)\|\mathbf{x}\|_2$ for all $\mathbf{x} \in X_S$ and show that $\delta^* \leq \delta$. Write $\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\mathbf{q}\|_2 + \|\mathbf{A}(\mathbf{x} - \mathbf{q})\|_2 \leq 1 + \frac{\delta}{2} + (1 + \delta^*)\frac{\delta}{4}$ 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 (Theorem 5.2 of [9]). *Suppose m, N and $0 < \delta < 1$ are given. If the distribution generating $\mathbf{A} \in \mathbb{R}^{m \times N}$ 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

A.2. Concentration of measure inequalities

Corrolary 3 (Corrolary 7.31 of [20]). *Let c_1, \dots, c_m be independent random variables with zero mean such that $|c_k| < K$ almost surely for $k \in \{1, \dots, m\}$ and some constant $K > 0$. Furthermore, assume $\mathbb{E}|c_k|^2 \leq \beta_k^2$ for constants $\beta_k > 0$, $k \in \{1, \dots, m\}$. Then, for all $t > 0$,*

$$\mathbb{P}\left(\left|\sum_{k=1}^m c_k\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2/2}{\beta^2 + Kt/3}\right),$$

where $\beta := \sum_{k=1}^m \beta_k^2$.

Theorem A.3 (Theorem 7.27 of [20]). *Let X_1, \dots, X_m be a sequence of independent mean-zero sub-Gaussian random variables satisfying $\mathbb{E}[\exp(\theta X_k)] \leq \exp(\beta\theta^2)$ with sub-Gaussian parameter β and for all $\theta \in \mathbb{R}$. For $\mathbf{b} \in \mathbb{R}^m$, the random variable $X := \sum_{k=1}^m b_k X_k$ is sub-Gaussian, i.e.*

$$\mathbb{E}(\exp(\theta X)) \leq \exp(\beta \|\mathbf{b}\|_2^2 \theta^2),$$

and

$$\mathbb{P}\left(\left|\sum_{k=1}^m b_k X_k\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{4\beta \|\mathbf{b}\|_2^2}\right) \quad \text{for all } t > 0.$$