ON TWO RANDOM MODELS IN DATA ANALYSIS

On two Random Models in Data Analysis

Dissertation

zur Erlangung des
mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsstudiengang *Mathematical Sciences*
der Georg August University School of Science (GAUSS)

vorgelegt von

David James

aus Fulda

Göttingen, 2016

Betreuungsausschuss

Prof. Dr. Felix Krahmer[1]
Prof. Dr. D. Russell Luke (NAM[3])

Mitglieder der Prüfungskommission

Referent:	Prof. Dr. Felix Krahmer[1]
Korreferent:	Prof. Dr. Matthias Hein[2]
Weitere Mitglieder:	Prof. Dr. D. Russell Luke (NAM[3])
	Prof. Dr. Gerlind Plonka-Hoch (NAM[3])
	Prof. Dr. Anja Sturm (IMS[4])
	Prof. Dr. Stephan Waack (IFI[5])

[1] Fakultät für Mathematik,
(Lehrstuhl M15, Technische Universität München)

[2] Fakultät für Mathematik und Informatik
(Fachrichtung Informatik, Universität des Saarlandes)

[3] Institut für Numerische und Angewandte Mathematik,
[4] Institut für Mathematische Stochastik,
[5] Institut für Informatik,
(jeweils Fakultät für Mathematik und Informatik, Georg-August-Universität Göttingen)

Tag der mündlichen Prüfung: 12. Januar 2017

Für Steffi, Linus und Ella.

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

Douglas Adams, The Restaurant at the End of the Universe

Abstract

In this thesis, we study two random models with various applications in data analysis. For our first model, we investigate subspaces spanned by biased random vectors. The underlying random model is motivated by applications in computational biology, where one aims at computing a low-rank matrix factorization involving a binary factor. In a random model with adjustable expected sparsity of the binary factor, we show for a large class of random binary factors that the corresponding factorization problem is uniquely solvable with high probability. In data analysis, such uniqueness results are of particular interest; ambiguous solutions often lack interpretability and do not give an insight into the structure of the underlying data. For proving uniqueness in this random model, small ball probability estimates are a key ingredient. Since to the best of our knowledge, there are no such estimate suitable for our application, we prove an extension of the famous Lemma of Littlewood and Offord. Hereby, we also discover a connection between the matrix factorization problem at hand and the notion of Sperner families.

In the second part of this thesis, we will investigate a model for randomized ultrasonic data in nondestructive testing. Here, we aim at accelerating the data acquisition process by superposing ultrasonic measurements with random time shifts. To this end, we will first study the effects of randomized ultrasonic measurements in the context of the Synthetic Aperture Focusing Technique (SAFT), a widely used defect imaging method. By adapting SAFT to our random data model, we will significantly improve its performance for randomized data. In this way, for sparse defects and with high probability, we achieve better defect reconstructions as with SAFT applied to deterministic ultrasonic data acquired in the same amount of time.

Acknowledgement

After three and a half years of working on this thesis, I want to thank everyone who made it possible.

First of all, I would like to thank my advisor Felix Krahmer for his genuine support and many inspiring discussions. Whenever I felt lost, he was there to help me back on my way. Thank you for your confidence and guidance. Moreover, I am indebted to my co-supervisor Russell Luke, who was very supportive and always had a sympathetic ear. Many thanks to Matthias Hein for helpful discussions and challenging (but solvable) problems, and for taking the position as co-referee without hesitation.

Moreover, I highly acknowledge the financial support of the *Federal Ministry of Education and Research* (BMBF), and the Research Training Group (GRK 2088) *Discovering structure in complex data: Statistics meets Optimization and Inverse Problems* for travel support and interesting workshops.

I owe my thanks to Gerlind Plonka-Hoch and all my (former) colleagues from the *Mathematical Signal and Image Processing* research group. You provided a warm and pleasant atmosphere, interesting talks and discussions in our research seminar, and a regular supply of tea and cake. Besides, I want to thank the research group for *Applied Numerical Analysis and Optimization and Data Analysis* at TU Munich. During my numerous visits, I always felt very welcome.

For providing experimental data and parts of their software, I would like to thank Oliver Nemitz and his colleagues from *Salzgitter Mannesmann Forschung GmbH*. In this context, I am also very thankful for fruitful and encouraging discussions with Martin Spies and Hans Rieder from *Fraunhofer Institute for Nondestructive Testing IZFP*. You all contributed to my understanding of nondestructive testing.

I highly appreciate the proofreading by Robert Beinert, Sina Bittens, Renato Budinich, and Christian Kruschel. My special thanks go to Robert Beinert, who was also of great help for typesetting this thesis.

I want to thank my family, my parents, Marion and Peter, and my sister Lisa, for encouraging and supporting me, especially in the last months of my work on this thesis. Last but not least, I thank my wife Steffi, and my children, Linus and Ella, for their love, support, and endless patience. You are my true source of inspiration and motivation.

Contents

ON TWO RANDOM MODELS IN DATA ANALYSIS

Chapter I.

Subspaces Spanned by Biased Random Vectors

1. Introduction

The Span of Random Binary Matrices

Bernoulli random matrices have recently gained a lot of attention. Arguably, the most prominent problem in this field is to estimate the probability that an $n \times n$ Bernoulli random matrix is singular as n goes to infinity. There has been tremendous progress in proving the conjecture that this probability is dominated by the probability that two columns or rows coincide [KKS95, BVW10, TV07]. A closely related problem concerns the investigation of the span of Bernoulli random matrices. Motivated by an application in neural networks [KS87], this problem was first investigated by Odlyzko in [Odl88]. He found that the probability that the linear span of the columns of a rectangular $N \times n$ Bernoulli matrix does not contain any $\{\pm 1\}$-vector besides its columns is dominated by the probability of the corresponding event for just three of its columns.

Theorem 1.1 (Odlyzko [Odl88]). *Let T be an $N \times n$ random matrix whose entries are independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 1] = 1/2$ and $\mathbb{P}[\epsilon = -1] = 1/2$. If*

$$n \leq \left(1 - \frac{10}{\log(N)}\right) N,$$

then the probability P that there exists a vector $x \in \mathbb{R}^n$ with at least two non-vanishing entries such that Tx is a $\{\pm 1\}$-vector can be bounded from above by

$$P \leq 4\binom{n}{3}\left(\frac{3}{4}\right)^N + \mathcal{O}\left(\left(\frac{7}{10}\right)^N\right),$$

as N goes to infinity.

The result, however, only treats the case where all entries of T are unbiased, i.e., they attain the values ±1 with equal probability. Motivated by applications related to matrix factorization (see below), we aim to transfer this result to the case of biased Bernoulli random variables; that is, random variables where the values ±1 are attained with unequal probabilities. We will show that the asymptotic behavior, i.e., the dominance by the probability that there exists a linear combination of three columns resulting in a $\{\pm 1\}$-vector, carries over to biased Bernoulli random matrices. The main result of this chapter reads as follows:

Theorem 1.2. *Let T be an $N \times n$ random matrix whose entries are independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 1] = p$ and $\mathbb{P}[\epsilon = -1] = 1 - p$. If there exists $\delta \in (0,1)$ with*

$$\min\{p, 1-p\} \geq N^{-(1-\delta)},$$

then there exists an absolute constant $C > 0$ depending only on δ such that for

$$n \leq \left(1 - \frac{C}{\log(N)}\right) N,$$

the probability P that there exists a vector $x \in \mathbb{R}^n$ with at least two non-vanishing entries such that Tx is a $\{\pm 1\}$-vector can be bounded from above by

$$P \leq 4\binom{n}{3}(1 - p(1-p))^N + o\left((1 - p(1-p))^N\right),$$

as N goes to infinity.

Note that for $p = 1/2$, we recover the asymptotic behavior of Theorem 1.1. Our result also covers the observation that the probability that there exists a vector x with two or more non-vanishing entries such that $Tx \in \{\pm 1\}^N$ is dominated by the corresponding event for just three columns of T. To see this, one can check that $(1 - p(1-p))^N$ is exactly the probability that $Tx \in \{\pm 1\}^N$ for a vector x in $\mathbb{R}^n$ with the entries $1, 1$ and -1 on its support of length 3. We will later see that for vectors x whose support set is of cardinality larger than 4 or equal to 2, this probability is of higher order.

A New Littlewood-Offord-type Inequality

The proof of Theorem 1.1 relied on estimating small ball probabilities of the form

$$\mathbb{P}\left[\left|\sum_{j=1}^{n} \epsilon_j x_j\right| = 1\right], \tag{1.1}$$

where x is a vector of non-vanishing entries and $n \geq 2$. In the unbiased case, the probability in (1.1) can be treated via the Lemma of LITTLEWOOD and OFFORD, which was proven by ERDŐS in [Erd45].

Theorem 1.3 (LITTLEWOOD-OFFORD [Erd45]). *Let $x \in \mathbb{R}^n$ be a vector with $\min_j |x_j| \geq c > 0$ and let ϵ_j, $1 \leq j \leq n$ be independent copies of a BERNOULLI random variable ϵ taking the values $+1$ and -1 with equal probability. Then for any open interval I of length at most $2c$, it holds that*

$$\mathbb{P}\left[\sum_{j=1}^{n} \epsilon_j x_j \in I\right] \leq \binom{n}{\lfloor n/2 \rfloor} 2^{-n}. \tag{1.2}$$

In contrast to the original result of LITTLEWOOD and OFFORD [LO43], which was only optimal up to a logarithmic factor, the estimate in (1.2) is sharp; one can easily verify that the right hand side of (1.2) is indeed achieved. For that, choose all entries of x to have the same modulus $c > 0$ and I to be the open interval of length $2c$ centered at $y = 0$ for n even or $y = \pm c$ for n odd. ERDŐS' proof is based on a connection between random sums as in (1.2) and SPERNER families in combination with SPERNER's Lemma [Spe28]. In the case where all entries of x are positive, a generalization of the Lemma of LITTLEWOOD and OFFORD to biased random variables can be proven using the LYM-inequality [Bol65, Lub66, Meš63, Yam54] instead of SPERNER's Lemma. The LYM-inequality was proven by BOLLOBÁS , LUBELL, MESHALKIN, and YAMAMOTO; it is named by the initials of the latter three.

Theorem 1.4 (Biased LITTLEWOOD-OFFORD [LL70]). *Let $x \in \mathbb{R}^n$ be a real vector with $\min_j x_j \geq c > 0$, let further ϵ_j, $1 \leq j \leq n$ be independent copies of a BERNOULLI random variable ϵ with $\mathbb{P}[\epsilon = 1] = p$ and $\mathbb{P}[\epsilon = -1] = 1 - p$, and let $I \subset \mathbb{R}$ be an open interval of length at most $2c$. Then*

$$\mathbb{P}\left[\sum_{j=1}^{n} \epsilon_j x_j \in I\right] \leq \max_{0 \leq k \leq n} \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.3}$$

As in Theorem 1.3, the estimate in (1.3) also is sharp for vectors x with constant entries, but it is not applicable if one aims to bound a probability like the one in (1.1). This is due to two reasons: While the assumption in Theorem 1.4 that all entries of x have a positive sign can easily be dropped in the case of $p = 1/2$, the same is not true in the unbiased case. Additionally, the problem of estimating the absolute value of the sum in (1.1) can be solved for $p = 1/2$ using Theorem 1.4 via a union bound, but the same method does not give meaningful probability estimates when considering highly biased Bernoulli random variables with p close to 0 or 1. In order to handle the first issue, the following lemma was proven by Costello et al. in [CV08] using Chebyshev's inequality.

Lemma 1.5 ([CV08]). *Let $x \in \mathbb{R}^n$ be arbitrary with $\min_j |x_j| \geq c > 0$, let further ϵ_j, $1 \leq j \leq n$ be independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 1] = p$ and $\mathbb{P}[\epsilon = -1] = 1 - p$ and let I be an open interval of length at most $2c$. Then there exists an absolute constant $C > 0$ such that*

$$\mathbb{P}\left[\sum_{j=1}^{n} \epsilon_j x_j \in I\right] \leq \frac{C}{\sqrt{n\mu}},$$

where $\mu = \min\{p, q\}$.

In contrast to the Theorem 1.4, Lemma 1.5 also allows us to treat vectors x with varying signs.

Note that this bound is not meaningful for small n or values of p which are close to 0 or 1, due to the constant factor C, which is not sharp. The following tighter variant of Lemma 1.5 can be derived from Theorem 1.4.

Corollary 1.6. *Let $x \in \mathbb{R}^n$ with $\min_j |x_j| \geq c > 0$ have exactly n_+ strictly positive and n_- strictly negative entries and let ϵ_j, $1 \leq j \leq n$ be independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 1] = p$ and $\mathbb{P}[\epsilon = -1] = 1 - p$. Let $I \subset \mathbb{R}$ be an arbitrary open interval of length at most $2c$ and let $\bar{n} \leq \max\{n_-, n_+\}$. Then*

$$\mathbb{P}\left[\sum_{j=1}^{n} \epsilon_j x_j \in I\right] \leq \max_{0 \leq k \leq \bar{n}} \binom{\bar{n}}{k} p^k (1-p)^{\bar{n}-k}.$$

Corollary 1.6 follows from Theorem 1.4 mainly by conditioning on the random variables whose coefficients have the less frequent sign. Nevertheless, since we did not find this bound in the literature, a proof will be provided in Section 5. Note that Theorem 1.6 still does not give meaningful results when used together with a union bound to estimate (1.1) when p is close to 0 or 1. A main result of this chapter is the following symmetric version of Corollary 1.6, which takes into account that, for biased Bernoulli random variables, the sum in (1.1) typically does not attain the values ±1 with equal probability. As we will see, the proof is considerably more involved than the proof of Corollary 1.6.

Theorem 1.7. *Let $x \in \mathbb{R}^n$ with $\min_j |x_j| \geq c > 0$ have n_+ strictly positive and n_- strictly negative entries and let $\epsilon_j, 1 \leq j \leq n$ be independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 1] = p$ and $\mathbb{P}[\epsilon = -1] = 1 - p$. Let $I \subset \mathbb{R}$ be an open interval of length at most $2c$ and let $\bar{n} \leq \max\{n_-, n_+\}$. Then*

$$\mathbb{P}\left[\left|\sum_{j=1}^{n} \epsilon_j x_j\right| \in I\right] \leq \max_{0 \leq k \leq \bar{n}} \binom{\bar{n}}{k} \left(p^k (1-p)^{\bar{n}-k} + p^{\bar{n}-k}(1-p)^k\right).$$

The inequality is tight, as equality is achieved for odd n and vectors x with constant entries. While it is neither a strict inequality for even n nor for vectors x with varying sign, it still gives meaningful estimates for these cases even when n is small or the Bernoulli random variables ϵ_j are extremely biased.

Matrix Factorization with Binary Components

Low rank matrix factorization is an important tool in data analysis, which allows us to represent data as linear combinations of a small number of building blocks, often referred to as *components*. In matrix factorization with binary components, one aims to factor a given data matrix $D \in \mathbb{R}^{N \times n}$ into the product BA, where $B \in \{0,1\}^{N \times r}$ is a binary matrix, and $A \in \mathbb{R}^{r \times n}$ is an arbitrary matrix whose rows sum to 1 and $r \ll \min\{N, n\}$. To be more precise, matrix factorization with binary components considers the problem

$$\text{find } B \in \{0,1\}^{N \times r} \text{ and } A \in \mathbb{R}^{r \times n}, \ A^T 1_r = 1_n, \text{ such that } D = BA, \tag{1.4}$$

where 1_r, 1_n denote the vectors of length r, n, respectively, with all entries equal to 1. Motivated by numerous applications, such as blind source separation in wireless communications with binary source signals [Vee97], network inference from gene expression data [LBY⁺03, TCX12], unmixing of cell mixtures from DNA methylation signatures [HAK⁺12], or clustering with overlapping [BKG⁺05, SBK03], it gained a lot of attention in recent years. Similar factorization problems involving binary matrices have for example been studied in [SSU03, KB08, MGNR06, ZLDZ07, MMG⁺08]. Note that, if we additionally demand that all entries of A are non-negative, the problem (1.4) is an instance of the non-negative matrix factorization problem, see, e.g., [PT94, LS99].

In [SHL13], Slawski, Hein and Lutsik proposed an algorithm to solve this problem by computing the intersection of the affine hull of the data matrix D, i.e.,

$$\operatorname{aff}(D) = \left\{ Dx \middle| x \in \mathbb{R}^n,\ \sum_{j=1}^{n} x_j = 1 \right\}, \tag{1.5}$$

with the set of vertices $\{0,1\}^N$. Their algorithm provably finds the solution to (1.4) if A has full rank, the columns of B are affinely independent, i.e., $\forall x \in \mathbb{R}^r$, $\sum_j x_j = 0$, $Bx = 0$ implies that $x = 0$, and the uniqueness condition

$$\operatorname{aff}(B) \cap \{0,1\}^N = \{B_{:,1}, \dots, B_{:,n}\} \tag{1.6}$$

is satisfied. Here, $B_{:,j}$, $1 \leq j \leq n$ denotes the jth column of B. By a direct calculation, we can see that the union of both conditions on B is equivalent to

$$\nexists x \in \mathbb{R}^r,\ \sum_{j=1}^{r} x_j = 1,\ \|x\|_0 \geq 2\ :\ Bx \in \{0,1\}^N, \tag{1.7}$$

where $\|x\|_0$ denotes the number of nonzero entries of x. Combining this observation with properties of modulated symmetric Sperner-2 families, a notion that we will introduce in Definition 2.14 below, allows us to prove the following result.

Theorem 1.8. *Let B be an $N \times n$ random matrix whose entries are independent copies of a Bernoulli random variable ϵ with $\mathbb{P}[\epsilon = 0] = p$ and $\mathbb{P}[\epsilon = 1] = 1 - p$. If there exists $\delta \in (0,1)$ with*

$$\min\{p, 1-p\} \geq N^{-(1-\delta)},$$

then there exists an absolute constant $C > 0$ depending only on δ such that for

$$n \leq \left(1 - \frac{C}{\log(N)}\right) N,$$

it holds that

$$\mathbb{P}\left[\exists x \in \mathbb{R}^r,\ \sum_{j=1}^{r} x_j = 1,\ \|x\|_0 \geq 2\ :\ Bx \in \{0,1\}^N\right] \\ \leq 4\binom{n}{3}(1 - p(1-p))^N + o\left((1 - p(1-p))^N\right), \tag{1.8}$$

as N goes to infinity.

As we will see later, Theorem 1.8 is a direct consequence of Theorem 1.2. Together with (1.7), it implies that the matrix factorization problem with binary components can be solved for a large class of random matrices. Note that the parameter p in Theorem 1.8 now also allows us to model sparse matrices B with just a few non-zero entries, which often occur in practice in the matrix factorization problem (1.4).

Organization of the Chapter

In Section 2, we will first consider deterministic versions of Theorem 1.2 and Theorem 1.8, and show a deep relation between both problems and the notion of SPERNER-k families, which we will also introduce in that section. In Section 3, we will pass on to the random setting and consider Theorem 1.2 in terms of probabilities involving SPERNER families. Afterwards, we aim to bound these probabilities in Section 4 and Section 5, which also contains the proofs of Theorem 1.6 and Theorem 1.7.

Notation

Throughout this chapter, $[n]$ will denote the integers from 1 to n, and 2^n will denote the power set of $[n]$. The symmetric difference $A \Delta B$ of two sets $A, B \subset [n]$ is defined by $A \Delta B := (A \setminus B) \cup (B \setminus A)$. For two sets A, B, we will write $A \mathbin{\dot\cup} B$ instead of the union $A \cup B$ if the two sets are disjoint. Similarly, for N sets A_i, $\dot\bigcup_{i \in [N]} A_i$ will denote the the union of the sets A_i if they are pairwise disjoint. Subsets $\mathcal{A} \subset 2^n$ will be referred to as families and will be denoted by calligraphic capital letters. When dealing with matrices, we denote, for an $N \times n$ matrix T and arbitrary sets $R \subset [N]$ and $C \subset [n]$, by $T_{R,C}$ the submatrix of T which arises by restricting T to the rows indexed by R and the columns indexed by C. Furthermore, we will write $T_{:,C}$ instead of $T_{[N],C}$ and $T_{R,:}$ instead of $T_{R,[n]}$, and we will denote by $T_{i,j}$ the entry of T with row index $i \in [N]$ and column index $j \in [n]$. The restriction of an n-dimensional vector x to its entries indexed by a set $J \subset [n]$ will be denoted by x_J. For an arbitrary n-dimensional vector x, we will denote by $\operatorname{supp}(x) \subset [n]$ the set containing all indices $j \in [n]$ such that $|x_j| > 0$ and refer to it as the support of x. We call a vector $x \in \mathbb{R}^n$ *s-sparse* if $|\operatorname{supp}(x)| = s$; the ℓ_0-norm of x is defined via $\|x\|_0 := |\operatorname{supp}(x)|$. The sign pattern $\operatorname{sgn}(x)$ of an arbitrary $x \in \mathbb{R}^n$ with non-vanishing entries will refer to the vector in $\{\pm 1\}^n$ defined by

$$\operatorname{sgn}(x)_j = \begin{cases} 1 & x_j > 0, \\ -1 & x_j < 0. \end{cases}$$

For two sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$, we write $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. In order to to highlight that two random variables X, Y have the same probability distribution, we will write $X \sim Y$.

2. From the Span of Binary Matrices to Sperner Families

In this section, we will establish the connection between the deterministic version of Theorem 1.2 for $N \times n$ matrices T with values in $\{\pm 1\}$ and a special class of families $\mathcal{A} \subset 2^n$, which was first studied by Sperner in [Spe28]. Our result generalizes the connection between Sperner families and random sums as in Theorem 1.3 and Theorem 1.4, which was discovered by Erdős in [Erd45], in multiple ways. Later, our generalization will allow us to prove our main results, Theorem 1.2 and Theorem 1.7, and also yields a uniqueness condition for matrix factorization with binary components. We first recall the definition of a Sperner-k family.

Definition 2.1 (Sperner-k family [Spe28]). We call any family $\mathcal{A} \subset 2^n$ a *Sperner-k family* if there does not exist a chain of $k+1$ sets $A_1, \dots, A_{k+1} \in \mathcal{A}$ with

$$A_1 \subsetneq A_2 \subsetneq \dots \subsetneq A_{k+1}. \tag{2.1}$$

For notational brevity and historical reasons, Sperner-1 families will simply be called Sperner families.

Example 2.2. The family $\mathcal{A}_1 = \Big\{\{1,2\},\{2,3\},\{1,3\}\Big\} \subset 2^3$ is a Sperner family, since no set contained in $\mathcal{A}_1$ is a proper subset of another set contained in $\mathcal{A}_1$. The family $\mathcal{A}_2 = \mathcal{A}_1 \cup \Big\{\{1,2,3\}\Big\}$ is not a Sperner family. It holds for instance that $\{1,2\} \subset \{1,2,3\}$. $\mathcal{A}_2$ is however a Sperner-2 family, as the longest chain of inclusions $A_1 \subsetneq \dots \subsetneq A_k$ with sets $A_1, \dots, A_k \in \mathcal{A}_2$ is of length $k = 2$.

While Sperner only considered Sperner families for $k = 1$, Sperner-k families are well known to connect to this basis case via the following lemma.

Lemma 2.3. *A family $\mathcal{A} \subset 2^n$ is a Sperner-k family if and only if it is the union of k Sperner families.*

Proof. If the family $\mathcal{A}$ is the union of k Sperner families but not a Sperner-k family, there must exist a chain of $k+1$ subsets $A_1, \dots, A_{k+1} \in \mathcal{A}$ with

$$A_1 \subsetneq A_2 \subsetneq \dots \subsetneq A_{k+1},$$

and the pigeonhole principle implies that at least two of them have to be contained in the same Sperner-1 family, which yields a contradiction.

For the reverse implication, let $\mathcal{A} = \{A_1, \dots, A_\ell\}$ be a Sperner-k family. Without loss of generality we may assume that $|A_j| \le |A_{j+1}|$ for $j \in [l-1]$. We will now construct k Sperner-1 families in an iterative way. Let $\mathcal{A}_1^0, \dots, \mathcal{A}_k^0$ be empty families. For each $j \in [\ell]$ and each $i \in [k]$ define iteratively

$$\mathcal{A}_i^j = \begin{cases} \mathcal{A}_i^{j-1} & \text{if } \left(\exists B \in \mathcal{A}_i^{j-1} \text{ s. th. } B \subsetneq A_j\right), \\ \mathcal{A}_i^{j-1} \cup \{A_j\} & \text{if } \left(\nexists B \in \mathcal{A}_i^{j-1} \text{ s. th. } B \subsetneq A_j\right) \wedge \left(A_j \notin \mathcal{A}_{i'}^j \text{ for } 1 \le i' \le i-1\right). \end{cases}$$

By construction, each of the families A_i^ℓ, $i \in [k]$ is a Sperner-1 family. We now claim that $\mathcal{A} = \bigcup_{i=1}^k \mathcal{A}_i^\ell$. Suppose for contradiction that there exists a set $B_{k+1} \in \mathcal{A}$ which is not assigned to any family A_i^ℓ, $i \in [k]$. Hence, there exists $B_k \in \mathcal{A}_k^\ell$ such that $B_k \subsetneq B_{k+1}$. Since for arbitrary $i \in [k]$, $i \ne 1$ and arbitrary $B_i \in A_i^\ell$, there must exist a set $B_{i-1} \in A_{i-1}^\ell$ with $B_{i-1} \subsetneq B_i$, we can therefore construct a chain of inclusions as in (2.1), contradicting the assumption that $\mathcal{A}$ is a Sperner-k family. This completes the proof.

Note that Lemma 2.3 also implies that every Sperner-k family also is a Sperner-ℓ family for $\ell \ge k$. To establish the connection between Sperner families and binary matrices, we will introduce the following operation.

Definition 2.4. For any $A \subset [n]$ and any $\xi \in \{\pm 1\}^n$, define the *modulation*

$$A^\xi = \left\{j | j \in A, \xi_j = 1\right\} \cup \left\{j | j \notin A, \xi_j = -1\right\} \subset [n];$$

for any family $\mathcal{A} \subset 2^n$, $\mathcal{A} = \{A_1, \dots, A_m\}$, denote by $\mathcal{A}^\xi$ the family given by

$$\mathcal{A}^\xi = \left\{A_1^\xi, \dots, A_m^\xi\right\}.$$

Remark 2.5. By definition, it holds for arbitrary $\xi \in \{\pm 1\}^n$ that

$$\emptyset^\xi = \{j \mid \xi_j = -1\}.$$

Also note that for any $A \subset [n]$, it holds that $A^{-1} = A^c$, where -1 denotes the n-dimensional vector with constant entries equal to -1. For this reason, we will denote by $\mathcal{A}^{-1}$ the family of all sets complementing the sets of $\mathcal{A}$.

Remark 2.6. By definition, the operation $(\cdot)^\xi$ for a sign-pattern $\xi \in \{\pm 1\}^n$ is *union compatible*, i.e., for two families $\mathcal{A}, \mathcal{B} \subset 2^n$, it holds that

$$(\mathcal{A} \cup \mathcal{B})^\xi = \mathcal{A}^\xi \cup \mathcal{B}^\xi$$

and

$$(\mathcal{A} \setminus \mathcal{B})^{\xi} = \mathcal{A}^{\xi} \setminus \mathcal{B}^{\xi}.$$

Example 2.7. For the family $\mathcal{A}_1 \subset 2^3$ as in Example 2.2 and $\xi = (-1, 1, -1)^T$, it holds that

$$\mathcal{A}_1^{\xi} = \left\{\{1,2\}^{\xi}, \{2,3\}^{\xi}, \{1,3\}^{\xi}\right\} = \left\{\{2,3\}, \{1,2\}, \emptyset\right\},$$

which is not a Sperner family, since $\emptyset \subset \{1,2\}$.

We now present some useful properties of the operation defined in Definition 2.4.

Proposition 2.8. *Let $A \subset [n]$ and $\xi \in \{\pm 1\}^n$ be arbitrary. Then*

$$A^{\xi} = A \Delta \{j | \xi_j = -1\}.$$

Consequently, for $A, B \subset [n]$ and any $\xi \in \{\pm 1\}^n$, it holds that

$$A^{\xi} \Delta B^{\xi} = A \Delta B$$

and

$$(A^{\xi})^{\nu} = A^{\xi \nu} = (A^{\nu})^{\xi},$$

where $\xi \nu \in \{\pm 1\}^n$ denotes the entrywise product of ξ and ν.

Remark 2.9. It is a direct consequence of Definition 2.4, that for arbitrary sign pattern $\xi \in \{\pm 1\}^n$, all properties in Proposition 2.8 concerning a set $A \subset [n]$ can be lifted to analogous properties of a family $\mathcal{A} \subset 2^n$.

Proof of Proposition 2.8. First, we observe that, for any $A \subset [n]$ and $\xi \in \{\pm 1\}^n$, the set A^{ξ} can be written as

$$\begin{aligned} A^{\xi} &= \left\{j | j \in A, \xi_j = 1\right\} \cup \left\{j | j \notin A, \xi_j = -1\right\} \\ &= A \setminus \{j | \xi_j = -1\} \cup \{j | \xi_j = -1\} \setminus A \\ &= A \Delta \{j | \xi_j = -1\} = A \Delta N_{\xi}, \end{aligned}$$

where

$$N_{\xi} := \{j | \xi_j = -1\}$$

for arbitrary $\xi \in \{\pm 1\}^n$. This establishes the first claim of the proposition. By the associativity and commutativity of the symmetric difference, it follows for any $A, B \subset [n]$ and $\xi \in \{\pm 1\}^n$ that

$$A^{\xi} \Delta B^{\xi} = (A \Delta N_{\xi}) \Delta (B \Delta N_{\xi}) = A \Delta B \Delta (N_{\xi} \Delta N_{\xi}) = A \Delta B \Delta \emptyset = A \Delta B,$$

which establishes the second claim of the proposition. Furthermore, we observe that for any $A \subset [n]$ and any $\xi, \nu \in \{\pm 1\}^n$ it holds that

$$\begin{aligned}(A^{\xi})^{\nu} &= (A \Delta N_{\xi}) \Delta N_{\nu} = A \Delta (N_{\xi} \Delta N_{\nu}) \\ &= A \Delta \{j \mid \text{either } \xi_j = -1 \text{ or } \nu_j = -1\} = A \Delta \{j \mid (\xi \nu)_j = -1\} \\ &= A \Delta N_{\xi \nu} = A^{\xi \nu}.\end{aligned}$$

Since the entrywise product is commutative, the last claim now follows by interchanging the roles of ξ and ν. □

Based on the notion of modulation, we now introduce a variant of Sperner-k families, which will play an essential role in the proof of our main results.

Definition 2.10 (Symmetric Sperner-k family). Let $\mathcal{A} \subset 2^n$ be arbitrary. For even k, we call $\mathcal{A}$ a *symmetric* Sperner-k family if $\mathcal{A}$ admits a decomposition of the form

$$\mathcal{A} = \bigcup_{l=1}^{k/2} \left(\mathcal{A}_l \cup \mathcal{A}_l^{-1} \right),$$

where $\mathcal{A}_l \subset 2^n$ is a Sperner family for each $l \in [k/2]$.
For odd k, we call $\mathcal{A}$ a *symmetric* Sperner-k family if

$$\mathcal{A} = \mathcal{A}_0 \cup \bigcup_{l=1}^{\lfloor k/2 \rfloor} \left(\mathcal{A}_l \cup \mathcal{A}_l^{-1} \right),$$

where $\mathcal{A}_l \subset 2^n$ is a Sperner family for each $0 \leq l \leq \lfloor k/2 \rfloor$ and $\mathcal{A}_0$ additionally satisfies that $\mathcal{A}_0 = \mathcal{A}_0^{-1}$.

Note that by Lemma 2.3, every symmetric Sperner-k family is indeed a Sperner-k family.

Example 2.11. Considering again the family $\mathcal{A}_1 = \Big\{\{1,2\},\{2,3\},\{1,3\}\Big\} \subset 2^3$ defined in Example 2.2, it follows that

$$\mathcal{A}_3 = \mathcal{A}_1 \cup \mathcal{A}_1^{-1} = \Big\{\{1,2\},\{2,3\},\{1,3\}\Big\} \cup \Big\{\{3\},\{1\},\{2\}\Big\}$$

is a symmetric Sperner-2 family.

The next lemma establishes a link between $\{\pm 1\}$-valued matrices and Sperner families. It generalizes the observations of Erdős in [Erd45].

Lemma 2.12. *Let T be an $N \times n$ binary matrix with values in $\{\pm 1\}$ and $x \in \mathbb{R}^n$, $\min_j |x_j| \geq c > 0$ be a vector such that $Tx \in V^N$, where V is the union of k open intervals of length at most $2c$. Let $\mathcal{A} \subset 2^n$ be the family containing the sets*

$$A_i = \{j | T_{i,j} = 1\}, \qquad i \in [N]. \tag{2.2}$$

Then $\mathcal{A}^{\operatorname{sgn}(x)}$ is a Sperner*-k family. If V additionally is symmetric, i.e., $V = -V$, it follows that $\mathcal{A}^{\operatorname{sgn}(x)}$ is a subfamily of a symmetric* Sperner*-k family.*

Proof. Set $\xi := \operatorname{sgn}(x)$. We may assume that $k < N$, since otherwise the first assertion of the lemma is trivial. Suppose for contradiction that $\mathcal{A} = \{A_1^\xi, \dots, \mathcal{A}_n^\xi\}$ is not a Sperner-k family. Then, after a possible permutation of the indices, it must hold that $A_1^\xi \subsetneq A_2^\xi \subsetneq \cdots \subsetneq A_{k+1}^\xi$. We define for $1 \leq i \leq k+1$,

$$y_i := (Tx)_i = \left(\sum_{j \in A_i} x_j - \sum_{j \in A_i^c} x_j \right) \in V. \tag{2.3}$$

Since all entries of x which make a positive contribution to this sum are contained in A_i^ξ and all entries of x which make a negative contribution to this sum are contained in $(A_i^c)^\xi = A_i^{-\xi}$, one can write

$$y_i = \left(\sum_{j \in A_i^\xi} |x_j| - \sum_{j \in A_i^{-\xi}} |x_j| \right) \in V.$$

Recall that V is the union of k intervals of length at most $2c$. Therefore, by the pigeonhole principle, there must be $v < w$ such that y_v, y_w are contained in the same interval. This, in turn, implies that $|y_v - y_w| < 2c$. As $A_v^\xi \subsetneq A_w^\xi$, there exists a non-empty set $S \subset [n] \setminus A_v^\xi$ such that $A_v^\xi \cup S = A_w^\xi$, and thus $A_w^{-\xi} \cup S = A_v^{-\xi}$. It follows that

$$\begin{aligned} y_w &= \sum_{j \in A_w^\xi} |x_j| - \sum_{j \in A_w^{-\xi}} |x_j| \\ &= \sum_{j \in A_v^\xi} |x_j| + \sum_{j \in S} |x_j| - \sum_{j \in A_v^{-\xi}} |x_j| + \sum_{j \in S} |x_j| \\ &= \left(\sum_{j \in A_v^\xi} |x_j| - \sum_{j \in A_v^{-\xi}} |x_j| \right) + 2 \sum_{j \in S} |x_j| \\ &= y_v + 2 \sum_{j \in S} |x_j|, \end{aligned}$$

which translates to

$$y_w - y_v = 2\sum_{j\in S} |x_j| \geq 2|S| \min_j |x_j| \geq 2c,$$

contradicting our finding that $|y_v - y_w| < 2c$. The family $\mathcal{A}^{\xi}$ therefore must be a Sperner-k family, which proves the first part of the lemma.

It remains to show that, if V is symmetric, then $\mathcal{A}^{\xi}$ is contained in a symmetric Sperner-k family. To see this, let $Y = \{y_1, \dots, y_k\}$ be the set of the centers of the k open intervals of V. We can assume without loss of generality that they are distinct and that $Y = -Y$ is a symmetric set. Therefore, there exists a permutation π of $[k]$ such that $y_i = -y_{\pi(i)}$ and π has at most one fixpoint, i.e., a 1-cycle corresponding to $y_i = 0$. All remaining cycles have length 2. Hence, if k is even, one can write

$$V = \bigcup_{\ell=1}^{k/2} V_\ell \cup (-V_\ell), \tag{2.4}$$

where the sets V_ℓ, $\ell \in [k/2]$ are open intervals of length at most $2c$ whose centers are contained in the positive real axis. If k is odd, one can write

$$V = V_0 \cup \bigcup_{\ell=1}^{\lfloor k/2 \rfloor} V_\ell \cup (-V_\ell), \tag{2.5}$$

where V_0 is an open interval of length at most $2c$ centered at zero and the sets V_ℓ, $\ell \in [k]$ are intervals of length at most $2c$ whose centers are contained in the positive real axis. The same decomposition can now be applied to the matrix T, and for ℓ as in (2.4) or (2.5), we denote by $T^{(\ell)}$ the submatrix of T containing the maximum number of rows t of T such that $\langle t, x\rangle \in V_\ell$. By permuting the rows of each of the matrices $T^{(\ell)}$ and possibly adding further rows, we may assume that $T^{(\ell)} = -T^{(-\ell)}$. Denote by $\mathcal{A}^{(\ell)}$ the family which arises by applying the construction described in (2.2) to the matrix $T^{(\ell)}$. We now claim that $-T^{(\ell)} = T^{(-\ell)}$ also implies that $\mathcal{A}_{-\ell} = \mathcal{A}_\ell^{-1}$. This directly follows from the observation that, for arbitrary $U \subset \mathbb{R}$ and $B \subset [n]$, multiplying

$$\left(\sum_{j\in B} x_j - \sum_{j\in B^c} x_j\right) \in U,$$

by (-1) corresponds to exchanging the roles of B and $B^c = B^{-1}$.

The first part of the proof now implies that $\mathcal{A}_\ell^{\xi}$ is a Sperner-1 family for each $\ell \in \lfloor k/2 \rfloor$ and $\xi = \operatorname{sgn}(x)$. Since we only applied row permutations or added further rows in order to construct the matrices $T^{(\ell)}$ from T, the decomposition in (2.4), (2.5) resp., now translates to

$$\mathcal{A} \subset \bigcup_{\ell=1}^{k/2} \mathcal{A}_\ell \cup (\mathcal{A}_{-\ell}) = \bigcup_{\ell=1}^{k/2} \mathcal{A}_\ell \cup (\mathcal{A}_\ell^{-1}),$$

in the case where k is even, and

$$\mathcal{A} \subset \mathcal{A}_0 \cup \bigcup_{\ell=1}^{\lfloor k/2 \rfloor} \mathcal{A}_\ell \cup (\mathcal{A}_\ell^{-1}),$$

in the case where k is odd. As the operation $(\cdot)^\xi$ is union compatible, see Remark 2.6, this completes the proof. □

The assertions of Lemma 2.12 can also be transferred to binary matrices with values in $\{0,1\}$.

Lemma 2.13. *Let B be an $N \times n$ binary matrix with values in $\{0,1\}$ and $x \in \mathbb{R}^n$, $\min_j |x_j| \geq c > 0$ be a vector such that $Bx \in V^N$, where V is the union of k open intervals of length at most c. Let $\mathcal{A} \subset 2^n$ be the family containing the sets*

$$A_i = \left\{ j | B_{i,j} = 1 \right\}, \quad i \in [N]. \tag{2.6}$$

Then $\mathcal{A}^{\operatorname{sgn}(x)}$ is a Sperner-k family. If, in addition, the set

$$V' := V - \frac{1}{2} \sum_{j=1}^{n} x_j \tag{2.7}$$

is symmetric, it follows that $\mathcal{A}^{\operatorname{sgn}(x)}$ is a subfamily of a symmetric Sperner-k family.

Proof. Let T be the $N \times m$ matrix defined by

$$T_{i,j} := \begin{cases} -1 & B_{i,j} = 0, \\ 1 & B_{i,j} = 1. \end{cases}$$

Since, in matrix form,

$$B = \tfrac{1}{2}(1_{N\times n} + T), \tag{2.8}$$

where $1_{N\times n}$ is the $N \times n$ matrix where all entries are equal to 1, for arbitrary $x \in \mathbb{R}^n$ it follows that

$$(Bx)_i = \frac{1}{2} \left(\sum_{j=1}^{n} x_j + T_{i,:}x \right).$$

With $2V := \{2v|v \in V\}$, which is the union of k open intervals of length at most $2c$, it therefore holds for arbitrary $i \in [N]$ that

$$(Tx)_i \in 2V \quad \Leftrightarrow \quad (Bx)_i \in \left(V + \sum_{j=1}^{n} x_j \right). \tag{2.9}$$

The result now directly follows from Lemma 2.12 by noting that $2V$ on the right hand side of (2.9) is symmetric if and only if (2.7) holds. □

We will now introduce a second variant of Sperner families.

Definition 2.14. A family $\mathcal{A} \subset 2^n$ is a *modulated Sperner-k family* if there exist a sign pattern $\xi \in \{\pm 1\}^n$ and a Sperner-k family $\mathcal{B} \subset 2^n$ such that $\mathcal{A} = \mathcal{B}^{\xi}$. If $\mathcal{B}$ is a symmetric Sperner-k family, we call $\mathcal{A}$ a *modulated symmetric Sperner-k family.*

We will now prove a result which lays the foundation to the proof of Theorem 1.2. It is based on the following definition.

Definition 2.15. For arbitrary $\mathcal{A} \subset 2^n$ and $J \subset [n]$, let $\mathcal{A} \sqcap J \subset 2^J$ be defined as

$$\mathcal{A} \sqcap J = \{A \cap J | A \in \mathcal{A}\}.$$

Corollary 2.16. *Let T be an $N \times n$ binary matrix with values in $\{\pm 1\}$ and $\mathcal{A} \subset 2^n$ be the family defined in Lemma 2.12. If there exist an s-sparse vector $x \in \mathbb{R}^n$ with*

$$\min_{j \in \operatorname{supp}(x)} |x_j| \geq c$$

and a set $V \subset \mathbb{R}$ which is the union of k open intervals of length at most $2c$ such that $Tx \in V^n$, then there exists a set $J \subset [n]$, $|J| = s$ such that $\mathcal{A} \sqcap J$ is a modulated Sperner-k family. If V is symmetric, it follows that $\mathcal{A} \sqcap J$ is a modulated symmetric Sperner-k family.

Proof. We will present the proof only for the symmetric case; the general case is similar. Suppose that there exist an s-sparse $x \in \mathbb{R}^n$ with

$$\min_{j \in \operatorname{supp}(x)} |x_j| \geq c$$

and a symmetric set $V \subset \mathbb{R}$ which is the union of k open intervals of length at most $2c$ such that $Tx \in V^n$. Set $J = \operatorname{supp}(x)$. Then $T_{:,J}x_J \in V^N$ and Lemma 2.12 implies that $(A \sqcap J)^{\xi_J}$ is a subfamily of a symmetric Sperner-k family. Since $((A \sqcap J)^{\operatorname{sgn}(x_J)})^{\operatorname{sgn}(x_J)} = A \sqcap J$ (Proposition 2.8), it follows that $A \sqcap J$ is a modulated symmetric Sperner-k family. □

Remark 2.17. Note that for arbitrary s-sparse $x \in \mathbb{R}^n$, any discrete set V with $|V| = k$ is a subset of

$$\bigcup_{y \in V} (y - c, y + c),$$

where $c = \min_j |x_j|$. Therefore, the assertion of Corollary 2.16 also holds for s-sparse $x \in \mathbb{R}^n$ and discrete sets V with $|V| = k$.

With Corollary 2.16, we are able to derive the following condition implying (1.7). It can be read directly from the matrix B without considering the affine hull.

Theorem 2.18. *Let B be an $N \times n$ binary matrix with values in $\{0,1\}$ and let $\mathcal{A} \subset 2^n$ be the family containing the sets*

$$A_i = \left\{ j | B_{i,j} = 1 \right\}, \quad i \in \{1, \ldots, N\}.$$

If none of the families

$$\{\mathcal{A} \sqcap J, | J \subset \{1, \ldots, n\}, 2 \leq |J| \leq n\}$$

is a subfamily of a modulated symmetric Sperner-2 family, then

$$\nexists x \in \mathbb{R}^r, \ \sum_{j=1}^{r} x_j = 1, \ \|x\|_0 \geq 2 \ : \ Bx \in \{0,1\}^N. \tag{2.10}$$

Proof. We can write the affine hull $\operatorname{aff}(B)$ as

$$\operatorname{aff}(B) = \left\{ Bx \middle| x \in \mathbb{R}^n, \ \sum_{j=1}^{n} x_j = 1 \right\} = \bigcup_{s=0}^{n} \operatorname{aff}_s(B),$$

where

$$\operatorname{aff}_s(B) = \left\{ Bx \middle| x \in \mathbb{R}^n, \ \|x\|_0 = s, \ \sum_{j=1}^{n} x_j = 1 \right\}.$$

In order to show (2.10), it suffices to prove

$$\operatorname{aff}_s(B) \cap \{0,1\}^N = \emptyset \qquad \text{for all } s \text{ with } 2 \leq s \leq n. \tag{2.11}$$

Let $s \geq 2$ be arbitrary, and let $T \in \{\pm 1\}^{N \times n}$ as in the proof of Lemma 2.13. Corollary 2.16 together with Remark 2.17 now implies that there do not exist an s-sparse vector $x \in \mathbb{R}^n$ and a symmetric set V with $|V| = 2$ such that $Tx \in V^N$. In particular,

$$\emptyset = \{Tx | x \in \mathbb{R}^n,\ \|x\|_0 = s\} \cap \{\pm 1\}^N = \operatorname{aff}_s(B) \cap \{0,1\}^N;$$

the last equality holds by (2.9). As s was arbitrary, this now implies (2.11) and completes the proof. □

Remark 2.19. From now on, we will only consider $\{\pm 1\}$-valued binary matrices. By Lemma 2.13, all results for $\{\pm 1\}$-valued binary matrices carry over to $\{0,1\}$-valued binary matrices by adjusting the right hand sides accordingly.

3. The Span of Random Binary Matrices and the Lemma of Littlewood and Offord

In this section, we pass from the setting of deterministic binary $N \times n$ matrices to Bernoulli random matrices. Similarly as in the previous section, they induce random sets, which we will call Bernoulli random sets.

Definition 3.1.

- A *Bernoulli random vector* $\epsilon^{(n)}$ of parameter $p \in (0,1)$ is a random vector whose entries ϵ_j, $j \in [n]$ are independent copies of a Bernoulli random variable taking the values $1, -1$ with probability p, $(1-p)$, respectively.
- A *Bernoulli random matrix* $\mathcal{E}^{(N,n)}$ of parameter $p \in (0,1)$ is an $N \times n$ random matrix where each row is an independent copy of a Bernoulli random vector $\epsilon^{(n)}$ of parameter p.
- For any finite set J, the *Bernoulli random set* $S^{(J)}$ of parameter $p \in (0,1)$ is a random subset of J, such that for any $A \subset J$,

$$\mathbb{P}\left[S^{(J)} = A\right] = p^{|A|}(1-p)^{|J|-|A|}. \tag{3.1}$$

That is, each element is included with probability p.

Remark 3.2. For all random variables described above, we will sometimes omit the upper indices if they are clear from the context. Furthermore, we write q instead of $1-p$ and $S^{(n)}$ instead of $S^{([n])}$.

The connection between Bernoulli random vectors and matrices and Bernoulli random sets is evident from their definition.

Remark 3.3. Let $\epsilon^{(n)}$ be a Bernoulli random vector with parameter p. Then the random set

$$S^{(n)} = \left\{ j | \epsilon_j = 1 \right\} \subset [n]$$

is a Bernoulli random set with the same parameter; we will call $S^{(n)}$ the *Bernoulli random set corresponding to* $\epsilon^{(n)}$. Also note that, since for arbitrary finite set J and a Bernoulli random set $S^{(J)}$ of parameter p it holds that

$$\sum_{A \in 2^J} \mathbb{P}\left[S^{(J)} = A \right] = \sum_{s=1}^{|J|} \binom{|J|}{s} p^s q^{|J|-s} = (p+q)^{|J|} = 1,$$

it follows that (3.1) actually defines a probability distribution.

With the next lemma we can transfer the problem of bounding small ball probabilities as in (1.1) to the domain of Bernoulli random sets and (symmetric) Sperner-k families; it generalizes the ideas used by Erdős to prove the Lemma of Littlewood and Offord in [Erd45].

Lemma 3.4. *Let $\epsilon^{(n)}$ be a Bernoulli random vector with parameter p and let $S^{(n)}$ be the corresponding Bernoulli random set. If V is the union of k open intervals of length at most $2c$ and $x \in \mathbb{R}^n$ is an arbitrary vector with $\min_{j \in [n]} |x_j| \geq c > 0$, then*

$$\mathbb{P}\left[\langle \epsilon^{(n)}, x \rangle \in V \right] \leq \max_{\substack{\mathcal{A} \subset 2^n \\ \mathcal{A} \text{ Sperner-}k}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\operatorname{sgn}(x)} \right] \leq P_{k,n}(p),$$

where we set

$$P_{k,n}(p) := \max_{\xi \in \{\pm 1\}^n} \max_{\substack{\mathcal{A} \subset 2^n \\ \mathcal{A} \text{ Sperner-}k}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi} \right].$$

If V is symmetric, we only need to consider symmetric Sperner-k families; we denote the corresponding probability by $P_{\pm,k,n}(p)$.

Proof. We will only prove the theorem in the case where V is symmetric; the general case follows analogously. For $x \in \mathbb{R}^n$ with $\min_j |x_j| \geq c$, let E be the matrix whose rows are all vectors $e \in \{\pm 1\}^n$ with $\langle e, x\rangle \in V$. For the family $\mathcal{B} = \{B_i | i \in [N]\}$ with

$$B_i = \left\{ j | E_{i,j} = 1 \right\} \subset [n],$$

Lemma 2.12 now implies that $\mathcal{B}^{\operatorname{sgn}(x)}$ is a subfamily of a symmetric Sperner-k family. Since by construction, it holds that

$$\langle e, x\rangle \in V \quad \Leftrightarrow \quad e \text{ is a row of } E \quad \Leftrightarrow \quad \{j | e_j = 1\} \in \mathcal{B},$$

it follows with Remark 3.3 that

$$\mathbb{P}\left[\langle \epsilon^{(n)}, x\rangle \in V\right] = \mathbb{P}\left[S^{(n)} \in \mathcal{B}\right]. \tag{3.2}$$

Bearing in mind that the family $\mathcal{B}^{\operatorname{sgn}(x)}$ is a subfamily of a symmetric Sperner-k family and that $\left(\mathcal{A}^{\xi}\right)^{\xi} = \mathcal{A}$ for $\mathcal{A} \subset 2^n$ and $\xi \in \{\pm 1\}^n$, we can bound (3.2) from above by

$$\begin{aligned} \mathbb{P}\left[S^{(n)} \in \mathcal{B}\right] &= \mathbb{P}\left[S^{(n)} \in (\mathcal{B}^{\operatorname{sgn}(x)})^{\operatorname{sgn}(x)}\right] \\ &\leq \max_{\substack{\mathcal{A} \subset 2^n \\ \mathcal{A} \text{ symmetric Sperner-}k}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\operatorname{sgn}(x)}\right] \\ &\leq \max_{\xi \in \{\pm 1\}^n} \max_{\substack{\mathcal{A} \subset 2^n \\ \mathcal{A} \text{ symmetric Sperner-}k}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right]. \end{aligned}$$

This completes the proof. □

Remark 3.5. As in Remark 2.17, Lemma 3.4 implies for a Bernoulli random vector $\epsilon^{(n)}$ of parameter p, arbitrary s-sparse $x \in \mathbb{R}^n$, and arbitrary discrete set $V \subset \mathbb{R}$, $|V| = k$ that

$$\mathbb{P}\left[\langle \epsilon^{(n)}, x\rangle \in V\right] \leq \max_{\substack{\mathcal{A} \subset 2^J \\ \mathcal{A} \text{ Sperner-}k}} \mathbb{P}\left[S^{(J)} \in \mathcal{A}^{\operatorname{sgn}(x)}\right] \leq P_{k,s}(p),$$

where $J = \operatorname{supp}(x)$. If V is symmetric, we similarly obtain

$$\mathbb{P}\left[\langle \epsilon^{(n)}, x\rangle \in V\right] \leq \max_{\substack{\mathcal{A} \subset 2^J \\ \mathcal{A} \text{ symmetric Sperner-}k}} \mathbb{P}\left[S^{(J)} \in \mathcal{A}^{\operatorname{sgn}(x)}\right] \leq P_{\pm,k,s}(p).$$

The quantities $P_{k,n}(p)$ and $P_{\pm,k,n}(p)$ will play an important role in the remainder of the chapter. A key distinction between the general and the symmetric case is that for the latter case, the following basic montonicity property is no longer true in general, see Remark 3.11 below.

Lemma 3.6. *For arbitrary integers k and $m \leq n$ and arbitrary $p \in (0,1)$, it holds that*

$$P_{k,m}(p) \geq P_{k,n}(p).$$

Proof. By induction, it is enough to prove the statement for $m = n-1$. For arbitrary $\mathcal{A} \subset 2^n$, define

$$\mathcal{A}_0 = \{A \subset [n-1] | A \in \mathcal{A}\} \subset 2^{n-1} \quad \text{and} \quad \mathcal{A}_1 = \{A \subset [n-1] | A \cup \{n\} \in \mathcal{A}\} \subset 2^{n-1}.$$

If $\mathcal{A}$ is a Sperner-k family, then both $\mathcal{A}_0 \subset 2^{n-1}$ and $\mathcal{A}_1 \subset 2^{n-1}$ are Sperner-k families. Now let $S^{(n)}$ and $S^{(n-1)}$ be Bernoulli random sets with parameter p, $\xi \in \{\pm 1\}^n$ be a sign pattern and $\nu \in \{\pm 1\}^{n-1}$ the restriction of ξ to the first $n-1$ entries. If $\xi_n = 1$, we have

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = q \cdot \mathbb{P}\left[S^{(n-1)} \in \mathcal{A}_0^{\nu}\right] + p \cdot \mathbb{P}\left[S^{n-1} \in \mathcal{A}_1^{\nu}\right] \leq pP_{k,n-1} + qP_{k,n-1} = P_{k,n-1}. \tag{3.3}$$

If $\xi_n = -1$, inequality (3.3) holds true with interchanged roles of p and q. This completes the proof. □

The next definition is required in order to be able to transfer the ideas of Corollary 2.16 to the random matrix case.

Definition 3.7. Let $\mathfrak{F}_{k,n} \subset 2^{2^n}$ be the set of all *maximal modulated Sperner-k* families, that is, the set of all modulated Sperner-k families $\mathcal{A} \subset 2^n$ which are not a proper subfamily of any other modulated Sperner-k family.
Furthermore, denote by $\mathfrak{F}_{\pm,k,n} \subset 2^{2^n}$ the set of all *maximal modulated symmetric Sperner-k* families.

Remark 3.8. Definition 3.7 also enables us to rewrite the probability $P_{k,n}(p)$ in terms of the set $\mathfrak{F}_{k,n}$, since for a Bernoulli random set $S^{(n)}$ with parameter p it holds that

$$P_{k,n}(p) = \max_{\xi \in \{\pm 1\}^n} \max_{\substack{\mathcal{A} \subset 2^n \\ \mathcal{A} \text{ Sperner-}k}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = \max_{\mathcal{A} \in \mathfrak{F}_{k,n}} \mathbb{P}\left[S^{(n)} \in \mathcal{A}\right],$$

and similarly for $P_{\pm,k,n}(p)$ and $\mathfrak{F}_{\pm,k,n}$.

For arbitrary $p \in (0,1)$ we will now compute the probabilities $P_{1,2}(p)$ and $P_{\pm,2,2}(p)$, which we will need later.

Lemma 3.9. *For arbitrary* $p \in (0,1)$*, it holds that*

$$P_{\pm,2,2}(p) = P_{1,2}(p) = p^2 + q^2.$$

Proof. Let $p \in (0,1)$ be arbitrary and let $\mathfrak{F}$ be the set of all Sperner families $\mathcal{A} \subset 2^2$. Then

$$\mathfrak{F} = \Big\{\emptyset, \{\emptyset\}, \{\{1\}\}, \{\{2\}\}, \{\{1,2\}\}, \{\{1\},\{2\}\}\Big\}.$$

Direct calculations yield that the set of all modulated Sperner families of subsets of $\{1,2\}$ is given by

$$\mathfrak{F}' = \mathfrak{F} \cup \Big\{\{\emptyset, \{1,2\}\}\Big\}.$$

Consequently, the set of all maximal modulated Sperner families is given by

$$\mathfrak{F}_{1,2} = \Big\{\{\{1\},\{2\}\}, \{\emptyset, \{1,2\}\}\Big\}.$$

Next, we will compute $\mathfrak{F}_{\pm,2,2}$. By Definition 2.10, Definition 2.14 and Remark 2.6, every modulated symmetric Sperner-2 family of subsets of $[n]$ is of the form

$$\left(\mathcal{A} \cup \mathcal{A}^{-1}\right)^{\xi} = \mathcal{A}^{\xi} \cup \mathcal{A}^{-\xi},$$

where $\mathcal{A} \subset 2^n$ is a Sperner-1 family and $\xi \in \{\pm 1\}^n$ is a sign pattern. Since for $\mathcal{A} \in \mathfrak{F}_{1,2}$, one has $\mathcal{A} = \mathcal{A}^{-1}$ and hence

$$\mathfrak{F}_{\pm,2,2} = \Big\{\mathcal{A} \cup \mathcal{A}^{-1} \Big| \mathcal{A} \in \mathfrak{F}_{1,2}\Big\} = \mathfrak{F}_{1,2},$$

we can conclude that $P_{\pm,2,2}(p) = P_{1,2}(p)$. It remains to show that $P_{1,2}(p) = p^2 + q^2$. For that, note that

$$P_{1,2}(p) = \max_{\mathcal{A} \in \mathfrak{F}_{1,2}} \mathbb{P}\left[S^{(2)} \in \mathcal{A}\right] = \max\left\{2pq, p^2 + q^2\right\} = p^2 + q^2,$$

where the last equality is implied by

$$p^2 + q^2 - 2pq = (p-q)^2 \geq 0.$$

This completes the proof. □

Lemma 3.10. *We have*

$$|\mathfrak{F}_{\pm,2,3}| = 4,$$

and for arbitrary $p \in (0,1)$*, it holds that*

$$P_{\pm,2,3}(p) = 1 - pq.$$

Remark 3.11. While the probabilities $P_{k,n}(p)$ are non-increasing with respect to n (Lemma 3.6), the same does not necessarily hold true for $P_{\pm,k,n}(p)$. Lemma 3.9 and Lemma 3.10 imply for arbitrary $p \in (0,1)$ that

$$P_{\pm,2,3}(p) - P_{\pm,2,2}(p) = 1 - pq - (p^2 + q^2) = (p+q)^2 - pq - (p^2 + q^2) = pq > 0.$$

Proof of Lemma 3.10. Since the families

$$\mathcal{A}_1 = \Big\{\{1\},\{2\},\{3\}\Big\} \qquad \text{and} \qquad \mathcal{A}_1^{-1} = \Big\{\{2,3\},\{1,3\},\{1,2\}\Big\}$$

are Sperner families, the set

$$\mathfrak{F}' = \left\{\mathcal{A}_1^{\xi} \cup \mathcal{A}_1^{-\xi} \middle| \xi \in \{\pm 1\}^3\right\}$$

consists of modulated symmetric Sperner-2 families. We claim that $\mathfrak{F}' = \mathfrak{F}_{\pm,2,3}$. To this end, observe that by union compatibility, it holds for arbitrary $\xi \in \{\pm 1\}^3$ that

$$\left(\mathcal{A}_1 \cup \mathcal{A}_1^{-1}\right)^{\xi} = \left(2^3 \setminus \{\emptyset,\{1,2,3\}\}\right)^{\xi} = (2^3)^{\xi} \setminus \{\emptyset,\{1,2,3\}\}^{\xi} = (2^3) \setminus \{\emptyset,\{1,2,3\}\}^{\xi}.$$

Since $\{\emptyset^{\xi} | \xi \in \{\pm 1\}^n\} = 2^n$, it therefore follows that

$$\mathfrak{F}' = \left\{2^3 \setminus \{A, A^c\} \middle| A \in 2^3\right\} = \left\{2^3 \setminus \{A, A^c\} \middle| A \in 2^3,\ |A| \le 1\right\}, \tag{3.4}$$

where the last equality holds by symmetry of the set $\{\mathcal{A}, \mathcal{A}^{-1}\}$. Consequently, all the modulated symmetric Sperner-2 families contained in $\mathfrak{F}'$ are maximal. Indeed, for any $\mathcal{A} \in \mathfrak{F}'$, adding some $A \in 2^3 \setminus \mathcal{A}$ results in a family which is not symmetric, and adding both missing sets results in 2^3, which is not a modulated symmetric Sperner-2 family. The same arguments also show that there are no modulated symmetric Sperner-2 families of cardinality larger than 6. For a modulated symmetric Sperner-2 family $\mathcal{A} \in 2^3$ of cardinality smaller than 6, its complement must also be symmetric, which shows that $\mathcal{A}$ is a subfamily of some $\mathcal{A}' \in \mathfrak{F}'$. Hence, $\mathcal{A}$ cannot be maximal and one has $\mathfrak{F}_{\pm,2,3} = \mathfrak{F}'$. Since each $A \in 2^3$ with $|A| \le 1$ yields a different set $2^3 \setminus \{A, A^c\}$, (3.4) implies that $|\mathfrak{F}_{\pm,2,3}| = |\{A \subset 2^3 | |A| \leq 1\}| = 4$. For the second part, a direct calculation shows that

$$\begin{aligned} P_{\pm,2,3}(p) &= \max_{\mathcal{A} \in \mathfrak{F}_{\pm,2,3}} \mathbb{P}\left[S^{(3)} \in \mathcal{A}\right] \\ &= 1 - \min_{A \in 2^3} \mathbb{P}\left[S^{(3)} \in \{A, A^{-1}\}\right] = 1 - \min\left\{p^3 + q^3, p^2 q + pq^2\right\} \\ &= \max\left\{1 - p^3 - q^3, 1 - pq\right\} = 1 - pq, \end{aligned}$$

as

$$(1 - pq) - (1 - (p^3 + q^3)) = p^3 + q^3 - pq = 4p^2 - 4p + 1 = (2p - 1)^2 > 0.$$

This completes the proof. □

The next lemma transfers the ideas of Corollary 2.16 to the random matrix case. It will enable us to prove a more general version of Theorem 1.2 in terms of the quantities $P_{k,n}(p)$ and $P_{\pm,k,n}(p)$ for a constant number of columns.

Lemma 3.12. *Let $\mathcal{E}^{(N,n)}$ be a Bernoulli random matrix with parameter p, let further V be the union of k open intervals of length at most $2c$, and let ℓ be an integer. Then with probability P of at most*

$$\sum_{s=k}^{\ell} \binom{n}{s} \sum_{\mathcal{A} \in \mathfrak{F}_{k,s}} \left(\mathbb{P}\left[S^{(s)} \in \mathcal{A}\right]\right)^N, \tag{3.5}$$

there exists a vector $x \in \mathbb{R}^n$, $k \leq \|x\|_0 \leq \ell$ with

$$\min_{j \in \operatorname{supp}(x)} |x_j| \geq c \tag{3.6}$$

such that $\mathcal{E}^{(N,n)} x \in V^N$. If

$$s_1 = \operatorname*{argmax}_{k \leq s \leq \ell} P_{k,s}(p) \tag{3.7}$$

is unique, then the probability P is bounded by

$$|\mathfrak{F}_{k,s_1}| \binom{n}{s_1} \left(P_{k,s_1}(p)\right)^N + o\left(\left(P_{k,s_1}(p)\right)^N\right).$$

If V is symmetric, we may replace $\mathfrak{F}_{k,s}$ by $\mathfrak{F}_{\pm,k,s}$ and $P_{k,n}$ by $P_{\pm,k,n}$.

Proof. We again only establish the lemma for the symmetric case, as the general case is analogous. Note that each row of the random matrix $\mathcal{E}^{(N,n)}$ is an independent copy of a Bernoulli random vector $\epsilon^{(n)}$. The family $\mathcal{S}^{(n)} \subset 2^n$ containing the sets

$$S_i^{(n)} = \left\{ j \middle| \mathcal{E}_{i,j}^{(N,n)} = 1 \right\}, \qquad i \in [N],$$

therefore is a random family of independent Bernoulli random sets with the same parameter p as $\mathcal{E}^{(N,n)}$. Recall that V is a union of k open intervals of length at most $2c$ and symmetric. Now consider the families

$$\left\{ \mathcal{S}^{(n)} \sqcap J \middle| J \subset [N], k \leq |J| \leq \ell \right\}. \tag{3.8}$$

Applying a union bound over all $J \subset [n]$ with $k \leq |J| \leq \ell$, we can therefore estimate the probability P that $\mathcal{E}^{(N,n)} x \in V^n$ for some x with $k \leq \|x\|_0 \leq \ell$ using Corollary 2.16 as

$$\begin{aligned} P &\leq \mathbb{P}\left[\text{there exists a modulated symmetric Sperner-}k\text{ family in (3.8) }\right] \\ &\leq \sum_{\substack{J \subset 2^s \\ k \leq |J| \leq \ell}} \mathbb{P}\left[\left\{ S_i^{(n)} \sqcap J \middle| i \in [N] \right\} \text{ is a modulated symmetric Sperner-}k\text{ family} \right] \\ &\leq \sum_{s=k}^{\ell} \binom{n}{s} \mathbb{P}\left[\left\{ S_i^{(s)} \middle| i \in [N] \right\} \text{ is a modulated symmetric Sperner-}k\text{ family} \right] \\ &\leq \sum_{s=k}^{\ell} \binom{n}{s} \mathbb{P}\left[\text{there exists a } \mathcal{A} \in \mathfrak{F}_{\pm,k,n} \text{ such that } \left\{ S_i^{(s)} \middle| i \in [N] \right\} \subset \mathcal{A} \right]. \end{aligned}$$

The last inequality holds since every modulated symmetric Sperner-k family is a subfamily of a maximal modulated symmetric Sperner-k family. By another union bound, it follows that

$$\begin{aligned} P &\leq \sum_{s=k}^{\ell} \binom{n}{s} \sum_{\mathcal{A} \in \mathfrak{F}_{\pm,k,s}} \mathbb{P}\left[\left\{ S_i^{(s)} \middle| i \in [N] \right\} \subset \mathcal{A} \right] \\ &= \sum_{s=k}^{\ell} \binom{n}{s} \sum_{\mathcal{A} \in \mathfrak{F}_{\pm,k,s}} \left(\mathbb{P}\left[S^{(s)} \in \mathcal{A} \right] \right)^N, \end{aligned} \tag{3.9}$$

which establishes the first part of the lemma. For the second part, note that assumption (3.7) implies that there exists $Q < P_{\pm,k,s_1}$ such that

$$\max_{\substack{k \leq s \leq \ell \\ s \neq s_1}} P_{\pm,k,s} \leq Q.$$

Recall from Remark 3.8 that

$$\max_{\mathcal{A}\in\mathfrak{F}_{\pm,k,s}} \mathbb{P}\left[S^{(s)}\in\mathcal{A}\right] = P_{\pm,k,s},$$

and note that, for $0\le s\le \ell\le n\le N$, it holds that

$$\binom{n}{s}\le\binom{2n}{s}\le\binom{2n}{\ell}\le\left(\frac{2\mathrm{e}N}{\ell}\right)^{\ell},$$

where we used the well-known inequality

$$\binom{n}{k}\le\left(\frac{\mathrm{e}n}{k}\right)^{k}\quad\text{for all } k\le n. \tag{3.10}$$

Bounding $|\mathfrak{F}_{\pm,k,s}|$ by the size of the power set of 2^s, which is its superset, we can therefore estimate (3.9) as

$$P\le\sum_{s=k}^{\ell}\binom{n}{s}|\mathfrak{F}_{\pm,k,s}|P_{\pm,k,s}^{N}\le\binom{n}{s_1}|\mathfrak{F}_{\pm,k,s_1}|P_{\pm,k,s_1}^{N}+\ell\left(\frac{2\mathrm{e}N}{\ell}\right)^{\ell}2^{2^{\ell}}Q^{N}.$$

As $Q<P_{\pm,k,s_1}$, and hence

$$\ell\left(\frac{2\mathrm{e}N}{\ell}\right)^{\ell}2^{2^{\ell}}Q^{N}=o\left(P_{\pm,k,s_1}^{N}\right)$$

as $N\to\infty$, this completes the proof. □

The next lemma allows us to prove a generalized version of Theorem 1.2 for vectors x with $\|x\|_0\ge s_0$. In contrast to Lemma 3.12, it only considers discrete sets V, but allows the number of columns n to tend to infinity. The proof is based on ideas by Odlyzko in [Odl88].

Lemma 3.13. *Let $M\sim\mathcal{E}^{(N,n)}$ be a Bernoulli random matrix with parameter $p\in(0,1)$ and assume there exists a constant $\delta\in(0,1)$ with*

$$\mu:=\min\{p,q\}\ge N^{-(1-\delta)}. \tag{3.11}$$

Furthermore, assume that for constant integers k, s_0 there exists a Q with

$$Q\ \ge\ \sqrt{p^2+q^2}, \tag{3.12}$$

$$Q\ \ge\ \sqrt{P_{k,s}(p)}\quad\textit{for all}\quad s\ge s_0. \tag{3.13}$$

Then there exists an absolute constant $C > 0$, depending only on k and δ, such that for any set V of k distinct points, for

$$n \leq \left(1 - \frac{C}{\log(N)}\right) N \tag{3.14}$$

and for arbitrary $\bar{Q} > Q$, the probability P that there exists a vector $x \in \mathbb{R}^n$ with $\|x\|_0 \geq s_0$ such that $Mx \in V^N$ satisfies

$$P = o\left(\bar{Q}^N\right).$$

If V is symmetric, we can replace $P_{k,s}$ with $P_{\pm,k,s}$ in (3.13).

Remark 3.14. In the non-symmetric formulation of Lemma 3.13, we can replace assumption (3.13) with $Q \geq \sqrt{P_{k,s_0}(p)}$, as $P_{k,n}(p)$ is monotone with respect to n, see Lemma 3.6. This is not possible in the symmetric formulation, since, as noted in Remark 3.11, $P_{\pm,k,n}(p)$ is not monotone with respect to n.

Remark 3.15. Note that the number of columns n in Lemma 3.13 is allowed to tend to ∞, as long as (3.14) is satisfied. The lower bound on the minimum of p and q is not an artifact of the proof, but it is necessary in order to ensure that the probability P in Lemma 3.13 tends to 0 as $N, n \to \infty$. To see this, let $\mathcal{E} := \mathcal{E}^{(N,n)}$ be a Bernoulli random matrix of parameter $p = \frac{c}{N}$ for some constant $c > 0$ independent of N. Considering only s_0 fixed columns of $\mathcal{E}$, we have

$$\mathbb{P}\left[\mathcal{E}_{i,j} = -1 \, \forall i \in [N], \, j \in [s_0]\right] = q^{s_0 N} = \left(1 - \tfrac{c}{N}\right)^{s_0 N} \geq \tfrac{1}{2}\mathrm{e}^{-cs_0}, \tag{3.15}$$

for N large enough. Suppose now that there exists an index $\ell \in [\lceil n/s_0 \rceil]$ such that $\mathcal{E}_{i,j} = -1$ for all $i \in [N]$ and $j \in \{s_0\ell + 1, \ldots, s_0(\ell + 1)\}$, which happens by independence and (3.15) with probability at least

$$1 - \left(1 - \tfrac{1}{2}\mathrm{e}^{-cs_0}\right)^{\lceil \frac{n}{s_0} \rceil}. \tag{3.16}$$

Then, for arbitrary $y \in \mathbb{R}$ and the s_0-sparse vector $x \in \mathbb{R}^n$ defined via

$$x_j = \begin{cases} -\frac{y}{s_0} & j \in \{s_0\ell + 1, \ldots, s_0(\ell + 1)\}, \\ 0 & \text{else}, \end{cases}$$

we have $(\mathcal{E}x)_i = y$ for all $i \in [N]$. Therefore, the probability P in Lemma 3.13 must be larger than (3.16), which tends to 1 for $n \to \infty$.

Proof of Lemma 3.13. Again, we only establish the proof for symmetric V, the general case is analogous by replacing all occurring symmetric Sperner-k families with Sperner-k families. Throughout this proof, we will omit the argument p of $P_{\pm,k,n}$. Let

$M \sim \mathcal{E}^{(N,s)}$ be a random Bernoulli matrix with parameter p and let $Q_{\pm,N,s}$ be the probability that there exists a vector $x \in \mathbb{R}^s$ with non-vanishing entries such that $Mx \in V^N$. We can now apply a union bound over the possible supports, to estimate

$$P \leq \sum_{s=s_0}^{n} \binom{n}{s} Q_{\pm,N,s} \leq \sum_{s=s_0}^{\lceil N^{\varepsilon} \rceil} \binom{N}{s} Q_{\pm,N,s} + \sum_{s=\lceil N^{\varepsilon}\rceil+1}^{n} \binom{N}{s} Q_{\pm,N,s} =: S_1 + S_2, \tag{3.17}$$

with $\varepsilon = 1 - \delta/2 > 1/2$. To bound the terms in S_1 corresponding to the sparsity level $s_0 \leq s \leq \lceil N^{\varepsilon} \rceil$, we will decompose the matrix M into two parts. Let $\ell \in \mathbb{N}$ with $\ell \leq N - s$ be arbitrary. The first matrix, $M^{(1)} \sim \mathcal{E}^{(s+\ell,s)}$ consists of the first $s + \ell$ rows of M and the second matrix, $M^{(2)} \sim \mathcal{E}^{(N-s-\ell,s)}$ consists of the remaining rows. Let $Q_{s+\ell,s}$ be the probability that the matrix $M^{(1)} \sim \mathcal{E}^{(s+\ell,s)}$ does not have full rank; we consider this case separately. Suppose now that $M^{(1)}$ is injective. Then there exists $R \subset [s + \ell]$ with $|R| = s$, such that the restriction $\tilde{M}^{(1)}$ of $M^{(1)}$ to the rows indexed by R has full rank. For each of the k^s vectors $y \in V^s$, there hence exists a unique vector $x \in \mathbb{R}^s$ with $\tilde{M}^{(1)}x = y$. By invertibility of $\tilde{M}^{(1)}$, the case of $x \in \mathbb{R}^s$ with vanishing entries does not contribute to $Q_{\pm,N,s}$, and we can assume that x only has nonvanishing entries. By stochastic independence of the rows, we can therefore bound the probability that $M^{(2)}x \in V^{(N-s-\ell)}$ from above by $\tilde{P}_{\pm,k,s}^{N-(s+\ell)}$. Here, $\tilde{P}_{\pm,k,s}$ is the probability that for a Bernoulli random vector $\epsilon^{(n)}$ of parameter p, it holds that $\langle e^{(n)}, x \rangle \in V$. By Remark 3.5, $\tilde{P}_{\pm,k,s}$ can be estimated as

$$\tilde{P}_{\pm,k,s} \leq P_{\pm,k,s} \leq Q^2; \tag{3.18}$$

the last inequality holds by assumption (3.13). On the event that $M^{(1)}$ is injective, we apply a union bound over all $\binom{s+\ell}{s}$ candidates for a subset R and all k^s vectors $y \in V^N$. Combining this with the event that $M^{(1)}$ is not injective we can therefore bound

$$Q_{\pm,N,s} \leq \binom{s+\ell}{s} k^s \tilde{P}_{\pm,k,s}^{N-(s+\ell)} + Q_{s+\ell,s}. \tag{3.19}$$

We will now bound $Q_{s+\ell,s}$ using a union bound over all potential ranks of $M^{(1)}$ and an approach similar to the one above. Suppose that $M^{(1)}$ has rank r with $1 \leq r \leq s-1$, implying that there exists $C \subset [s]$ with $|C| = r$, $\tau \in [s] \setminus C$ and $x \in \mathbb{R}^r$ with $\operatorname{rank} M^{(1)}_{:,C} = r$ and

$$M^{(1)}_{:,C} x = M^{(1)}_{:,\tau} \qquad \Leftrightarrow \qquad M^{(1)}_{:,C\cup\{\tau\}} (x^T, -1)^T = 0, \tag{3.20}$$

where $M^{(1)}_{:,C\cup\{\tau\}}$ is the matrix which arises by adding the column $M^{(1)}_{:,\tau}$ on the right to the matrix $M^{(1)}_{:,C}$. The vector x cannot have any vanishing entries, since this would imply that $M^{(1)}$ had rank smaller than r. As $M^{(1)}_{:,C}$ has rank r, there exists $R \subset [s + \ell]$, $|R| = r$, such that $M^{(1)}_{R,C}$ is invertible; the vector x in (3.20) is therefore unique. Note that (3.20) also implies

$$M^{(1)}_{R^c,C\cup\{\tau\}} (x^T, -1)^T = 0. \tag{3.21}$$

Consequently,

$$\begin{aligned}
\mathbb{P}\Big[M^{(1)} \text{ not injective}\Big] &\leq \sum_{1\leq r\leq s-1} \mathbb{P}\left[\operatorname{rank}(M^{(1)}) = r\right] \\
&\leq \sum_{1\leq r\leq s-1} \mathbb{P}\left[\exists C \subset [s], |C| = r : (\operatorname{rank}(M^{(1)}_{:,C}) = r) \wedge (\forall \tau \in [s] \setminus C : \operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}})) = r\right] \\
&\leq \sum_{1\leq r\leq s-1} \mathbb{P}\Big[\exists C \subset [s], \exists R \subset [s+\ell], |C| = |R| = r, \forall \tau \in [s] \setminus C : \qquad (3.22) \\
&\qquad\qquad (\operatorname{rank}(M^{(1)}_{R,C}) = r) \wedge (\operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}}) = r)\Big] \\
&\leq \sum_{1\leq r\leq s-1} \sum_{\substack{C\subset[s]\\|C|=r}} \sum_{\substack{R\subset[s+\ell]\\|C|=r}} \mathbb{P}\left[(\operatorname{rank}(M^{(1)}_{R,C}) = r) \wedge (\operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}}) = r)\right],
\end{aligned}$$

the last line follows by a union bound over all choices of C, R while always choosing the smallest τ in $[s] \setminus C$. Conditioning on $M^{(1)}_{R,C}$, which is invertible, $\operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}}) = r$ can only hold true if the unique x in (3.20) also satisfies (3.21). By stochastic independence of the rows of $M^{(1)}_{R^c,C}$, we therefore have

$$\begin{aligned}
&\mathbb{P}\left[(\operatorname{rank}(M^{(1)}_{R,C}) = r) \wedge (\operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}}) = r)\right] \\
&\quad = \mathbb{E}\left[\mathbb{P}\left[(\operatorname{rank}(M^{(1)}_{R,C}) = r) \wedge (\operatorname{rank}(M^{(1)}_{:,C\cup\{\tau\}}) = r) | M^{(1)}_{R,C}\right]\right] \qquad (3.23) \\
&\quad \leq \tilde{P}^{s+\ell-r}_{1,r+1},
\end{aligned}$$

where $\tilde{P}_{1,r+1}$ is the maximal probability that for x as in (3.21), and a Bernoulli random vector $e^{(n)}$ of the parameter p, it holds that $\langle e^{(n)}, (x, -1)^T \rangle = 0$. By Remark 3.5, $\tilde{P}_{1,r+1}$ can be estimated as

$$\tilde{P}_{1,r+1} \leq P_{1,r+1} \leq P_{1,2} = p^2 + q^2 \leq Q^2; \qquad (3.24)$$

the inequalities hold by Lemma 3.6, Lemma 3.9, and (3.12). Bringing (3.22) and (3.23) together, we therefore arrive at

$$Q_{s+\ell,s} \leq \sum_{r=1}^{s-1} \binom{s}{r} \binom{s+\ell}{r} \tilde{P}^{s+\ell-r}_{1,r+1}. \qquad (3.25)$$

For $s \leq N^{\varepsilon}$ and $N \geq 2^{(1-\varepsilon)^{-1}}$ such that $N^{\varepsilon} \leq N/2$, we can therefore bound $Q_{s+\ell,s}$ from above by

$$Q_{s+\ell,s} \leq s2^s \binom{N}{s} Q^{2\ell}, \qquad (3.26)$$

since for $1 \leq r \leq s \leq N^{\varepsilon} \leq N/2$ and $\ell \leq N - s$, we have

$$\binom{s}{r} \leq 2^s, \quad \binom{s+\ell}{r} \leq \binom{N}{s}, \quad \tilde{P}^{s+\ell-k}_{1,k+1} \leq Q^{2(s+\ell-r)} \leq Q^{2\ell}.$$

Applying (3.19) together with (3.18), (3.26), and $n, s + \ell \leq N$, it follows that

$$\binom{n}{s} Q_{\pm,N,s} \leq \binom{N}{s}^2 k^s Q^{2(N-(s+\ell))} + s2^s \binom{N}{s}^2 Q^{2\ell} \tag{3.27}$$

$$\leq s(2k)^s \binom{N}{s}^2 \left(Q^{2(N-(s+\ell))} + Q^{2\ell} \right). \tag{3.28}$$

For $N \geq 2^{(1-\varepsilon)^{-1}}$, we can now choose $\ell = {}^N\!/2$ and, using (3.27) and (3.10), bound S_1 in (3.17) as

$$\begin{aligned} S_1 &= \sum_{s=s_0}^{\lceil N^\varepsilon \rceil} \binom{n}{s} Q_{\pm,N,s} \\ &\leq N^{2\varepsilon} (2k)^{N^\varepsilon} \binom{N}{N^\varepsilon}^2 \left(Q^{N-2N^\varepsilon} + Q^N \right) \\ &\leq N^{2\varepsilon} \left(2k\mathrm{e}^2 N^{2(1-\varepsilon)} \right)^{N^\varepsilon} \left(Q^{N-2N^\varepsilon} + Q^N \right) \\ &\leq 2N^{2\varepsilon} \left(4k\mathrm{e}^2 N^{2(1-\varepsilon)} \right)^{N^\varepsilon} Q^N =: D_N. \end{aligned} \tag{3.29}$$

In last inequality, we used (3.12) to bound

$$Q^{-2N^\varepsilon} \leq (p^2 + q^2)^{-N^\varepsilon} = (1 - 2pq)^{-N^\varepsilon} \leq ({}^1\!/2)^{-N^\varepsilon} = 2^{N^\varepsilon}. \tag{3.30}$$

Using monotonicity and inequality (3.30), we can now bound

$$\begin{aligned} &\log(S_1) - \log(\bar{Q}^N) \leq \log(D_N) - \log(\bar{Q}^N) \\ &\qquad \leq \log(2) + 2\varepsilon \log(N) + N^\varepsilon \left(\log(4k\mathrm{e}^2) + 2(1-\varepsilon)\log(N) \right) + N(\log(Q) - \log(\bar{Q})). \end{aligned} \tag{3.31}$$

Recalling that $\bar{Q} > Q$ and noting that $N^\varepsilon \log N = o(N)$, it follows that the right-hand side of (3.31) tends to $-\infty$ as $N \to \infty$, which implies that $S_1 = o\left(\bar{Q}^N\right)$.

To bound S_2, we will again decompose the matrix $M \sim \mathcal{E}^{(N,n)}$ into two parts, $M^{(1)}$ and $M^{(2)}$, consisting of the first $n + \ell$ rows of M, and the remaining rows, respectively, for some $\ell \leq N - n$ to be determined later. If $M^{(1)}$ is injective, so are all of its column restricted submatrices. Hence, on the event that $M^{(1)}$ is injective (the complementing event has probability at most $Q_{n+\ell,n}$), we obtain a bound similar to (3.19) but without $Q_{s+\ell,s}$. It follows that

$$S_2 = \sum_{s=\lceil N^\varepsilon \rceil + 1}^{n} \binom{n}{s} Q_{\pm,N,s} \leq Q_{n+\ell,n} + \sum_{s=\lceil N^\varepsilon \rceil + 1}^{n} \binom{n}{s} \binom{s+\ell}{s} k^s \tilde{P}_{\pm,k,s}^{N-(s+\ell)}. \tag{3.32}$$

Using Lemma 1.5 and a union bound over the k points in V, we can now bound

$$\tilde{P}_{\pm,k,s} \leq \tilde{P}_{k,s} \leq \frac{kC}{\sqrt{\mu s}}. \tag{3.33}$$

It follows from (3.32) that

$$\begin{aligned} S_2 &\le Q_{n+\ell,n} + N8^N \left(\frac{kC}{\mu^{1/2}N^{\varepsilon/2}}\right)^{N-(n+\ell)} \\ &\le Q_{n+\ell,n} + 9^N \left(\frac{kC}{\mu^{1/2}N^{\varepsilon/2}}\right)^{N-(s+\ell)}, \end{aligned}$$

provided N is large enough such that $\log(N)N^{-1} \le \log(9/8)$, which is satisfied for $N \ge 3 \cdot 10^6$. Applying (3.25) for $s = n$, it follows that

$$\begin{aligned} Q_{n+\ell,n} &\le \sum_{r=1}^{n-1} \binom{n}{r}(n-r)\binom{n+\ell}{r}\tilde{P}_{1,r+1}^{n+\ell-r} \\ &= \sum_{r=1}^{\lceil N^\varepsilon \rceil} \binom{n}{r}(n-r)\binom{n+\ell}{r}\tilde{P}_{1,r+1}^{n+\ell-r} + \sum_{r=\lceil N^\varepsilon \rceil+1}^{n} \binom{n}{r}(n-r)\binom{n+\ell}{r}\tilde{P}_{1,r+1}^{n+\ell-r} \\ &=: T_1 + T_2. \end{aligned}$$

Applying again (3.33), we find that

$$T_2 \le n^2 4^N \left(\frac{C}{\mu^{1/2}N^{\varepsilon/2}}\right)^{\ell} \le 8^N \left(\frac{C}{\mu^{1/2}N^{\varepsilon/2}}\right)^{\ell},$$

provided $N \ge 4$. Since $n + \ell \le N$ and $\tilde{P}_{1,r+1} \le Q^2$ by (3.24), T_1 can be bounded for large enough N by

$$\begin{aligned} T_1 &\le N^\varepsilon \cdot N\binom{N}{N^\varepsilon}^2 Q^{2(n+\ell-N^\varepsilon)} \\ &\le N^{(1+\varepsilon)} \left(\mathrm{e}N^{(1-\varepsilon)}\right)^{2N^\varepsilon} Q^{2(n+\ell-N^\varepsilon)} \\ &\le N^{(1+\varepsilon)} \left(2\mathrm{e}^2N^{2(1-\varepsilon)}\right)^{N^\varepsilon} Q^{2(n+\ell)}, \end{aligned}$$

where we used (3.10) in the second and (3.30) in the third inequality. If we choose ℓ such that $n + \ell \ge N/2$, we obtain, with D_N as in (3.29) and (3.31), that

$$T_1 \le N^{(1+\varepsilon)} \left(2\mathrm{e}^2N^{2(1-\varepsilon)}\right)^{N^\varepsilon} Q^N = \frac{D_N}{2N^{(2\varepsilon-1)}(2k)^{N^\varepsilon}} \le D_N = o(\bar{Q}^N), \tag{3.34}$$

as $\varepsilon = 1 - \delta/2 > 1/2$. Combining all these results, we therefore obtain for any choice of ℓ with $n + \ell \ge N/2$ that

$$S_2 \le o\left(\bar{Q}^N\right) + 9^N \left(\frac{kC}{\mu^{1/2}N^{\varepsilon/2}}\right)^{N-(n+\ell)} + 8^N \left(\frac{C}{\mu^{1/2}N^{\varepsilon/2}}\right)^{\ell}. \tag{3.35}$$

If $n \leq N/2$, we can choose $\ell = \max\{N/2 - n, N/4\}$, which implies via direct calculation that $n + \ell \geq N/2$, $\ell \geq N/4$ and $N - (n + \ell) \geq N/4$. Together with (3.35), we therefore have

$$S_2 \leq o\left(\bar{Q}^N\right) + 2 \cdot 9^N \left(\frac{kC}{\mu^{1/2} N^{\varepsilon/2}}\right)^{N/4} = o\left(\bar{Q}^N\right) + 2 \cdot \left(\frac{9^4 kC}{\mu^{1/2} N^{\varepsilon/2}}\right)^{N/4}.$$

With $\delta > 0$ as in the assumption, $\varepsilon = 1 - \delta/2 > 1/2$ now implies that $\mu^{1/2} N^{\varepsilon/2} \geq N^{\delta/4}$ and hence $S_2 = o\left(\bar{Q}^N\right)$ in the case where $n \leq N/2$. If $n > N/2$, we choose $\ell = \frac{N-n}{2}$, which implies that

$$N - (n + \ell) = \frac{N - n}{2} = \ell.$$

Suppose for a moment that

$$\left(\frac{kC}{N^{\delta/4}}\right)^{\ell/N} < \frac{1}{9\sqrt{2}}. \tag{3.36}$$

Then, from (3.35), it follows that

$$\begin{aligned} S_2 &\leq o\left(\bar{Q}^N\right) + 9^N \left(\frac{kC}{\mu^{1/2} N^{\varepsilon/2}}\right)^{\ell} + 8^N \left(\frac{C}{\mu^{1/2} N^{\varepsilon/2}}\right)^{N-(n+\ell)} \\ &\leq o\left(\bar{Q}^N\right) + 2 \cdot 9^N \left(\frac{kC}{N^{\delta/4}}\right)^{\ell} \\ &= o\left(\bar{Q}^N\right). \end{aligned}$$

To complete the proof, note that by the monotonicity of the logarithm, (3.36) is equivalent to

$$\ell > \frac{\log(9\sqrt{2})N}{\delta/4 \log(N) - \log(kC)},$$

which, for $N > (kC)^{4/\delta}$ and a sufficiently large constant $\tilde{C} > 0$ depending only on the constants δ and k, is implied by

$$\ell \geq \frac{\tilde{C}N}{2\log(N)}.$$

Since $\ell = \frac{N-n}{2}$, this inequality is equivalent to

$$n \leq \left(1 - \frac{\tilde{C}}{\log N}\right) N. \qquad \square$$

In this section, we reduced the problem of investigating the span of the columns of a random Bernoulli matrix $\mathcal{E}^{(N,n)}$ to the problem of bounding the quantities $P_{k,n}(p)$ and $P_{\pm,k,n}(p)$, which we aim to bound in the remainder of the chapter. In Section 4, we will bound $P_{\pm,k,n}(p)$ and $P_{\pm,k,n}(p)$ using cardinality estimates for Sperner-k families and a greedy method. Since these bounds are only applicable for certain values of p and n, we will reduce the problem of bounding $P_{k,n}(p)$ and $P_{\pm,k,n}(p)$ in Section 5 in a way that we can apply the LYM-inequality, which we will also introduce in that section. Whereas resulting bounds on $P_{\pm,k,n}(p)$ will be weaker than the bounds in Section 4, they will be monotone with respect to n.

4. Bounding $P_{k,n}$ and $P_{\pm,k,n}$ using Cardinality Estimates

In the case where $k = 1$, the problem of bounding the cardinality of a Sperner family was first addressed by Sperner.

Lemma 4.1 (Sperner's Lemma [Spe28]). *Let $\mathcal{A} \subset 2^n$ be a Sperner family, then*

$$|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

In particular, if n is even, the largest Sperner family of $[n]$ contains exactly the subsets of cardinality $n/2$ of $[n]$. If n is odd, the largest two Sperner families of $[n]$ are the families containing all subsets of cardinality $\lfloor n/2 \rfloor$ of $[n]$ or all subsets of cardinality $\lceil n/2 \rceil$ of $[n]$.

Sperner's Lemma can be generalized to Sperner-k families. The corresponding result is considered folklore without known reference, see, e.g., [EFK05].

Lemma 4.2. *Let $\mathcal{A} \subset 2^n$ be a Sperner-k family. Then*

$$|\mathcal{A}| \leq \sum_{i=\lfloor (n-k+1)/2 \rfloor}^{\lfloor (n+k-1)/2 \rfloor} \binom{n}{i}.$$

Equality holds if and only if $\mathcal{A}$ is the family of all sets A with $|A| \in [\lfloor \frac{(n-k+1)}{2} \rfloor, \lfloor \frac{(n+k-1)}{2} \rfloor]$ or the family of all sets A with $A \in [\lceil \frac{(n-k+1)}{2} \rceil, \lceil \frac{(n+k-1)}{2} \rceil]$.

In order to be able to handle symmetric Sperner-2 families, a refinement of Sperner's Lemma will be useful. The following lemma, which is due to Milner, gives an estimate for the cardinality of a special class of Sperner families.

Lemma 4.3 (Milner [Mil68]). *Let $\mathcal{A} \subset 2^n$ be a Sperner family which does not contain any complementing sets, i.e., sets $A \in \mathcal{A}$ with $A^c \in \mathcal{A}$. Then*

$$|\mathcal{A}| \le \binom{n}{\left\lfloor \frac{n-1}{2} \right\rfloor}.$$

We can now use Lemma 4.3 to estimate the cardinality of symmetric Sperner-2 families.

Corollary 4.4. *Let $\mathcal{A} \subset 2^n$ be a symmetric Sperner-2 family. Then*

$$|\mathcal{A}| \le 2\binom{n}{\lfloor (n-1)/2 \rfloor}.$$

Proof. Let $\mathcal{A} \subset 2^n$ be an arbitrary symmetric Sperner-2 family. By Definition 2.10, $\mathcal{A}$ is of the form

$$\mathcal{A} = \mathcal{B} \cup \mathcal{B}^{-1},$$

where $\mathcal{B} \subset 2^n$ is a Sperner family. We may assume that $\mathcal{B}$ does not contain any complementing sets. Lemma 4.3 now implies that

$$|\mathcal{A}| \le 2|\mathcal{B}| \le 2\binom{n}{\lfloor (n-1)/2 \rfloor}.$$

This completes the proof. □

For Bernoulli random sets of parameter $p = 1/2$, we have the following:

Lemma 4.5. *Let $\mathcal{A} \subset 2^n$ be a family and $S^{(n)}$ a Bernoulli random set with parameter $p = 1/2$. Then it holds for arbitrary sign pattern $\xi \in \{\pm 1\}^n$ that*

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = \frac{|\mathcal{A}|}{2^n}.$$

Proof. A Bernoulli random set $S^{(n)}$ of parameter $p = 1/2$ attains each set $A \in 2^n$ with the same probability 2^{-n}. It follows that

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = \frac{|\mathcal{A}^{\xi}|}{2^n} = \frac{|\mathcal{A}|}{2^n},$$

since for any family $\mathcal{A} \subset 2^n$ and any sign pattern $\xi \in \{\pm 1\}^n$, we have $|\mathcal{A}^{\xi}| = |\mathcal{A}|$. □

Together with the cardinality bounds above, it is now straightforward to estimate the probabilities $P_{\pm,k,n}(p)$ and $P_{k,n}(p)$ in the case of $p = 1/2$. Note that Theorem 1.3 was proven by Erdős in [Erd45] using Sperner's Lemma (Lemma 4.1) and the observation of Lemma 4.5. While this will eventually give rise to generalized Littlewood-Offord-type inequalities using Lemma 3.4, it also directly yields Theorem 1.2 for $p = 1/2$, which basically is the case considered by Odlyzko (Theorem 1.1).

Proof of Theorem 1.2 for $p = 1/2$. Since $p = 1/2$ is fixed, we will omit the argument p for $P_{\pm,k,n}$ and $P_{k,n}$. Note that $\{\pm 1\}$ is a symmetric set of two points. Lemma 3.9 and Lemma 3.10 now yield

$$P_{\pm,2,2} = 1/2, \quad P_{\pm,2,3} = 3/4.$$

For all $s \geq 4$, Lemma 3.6 and Lemma 4.5 together with Lemma 4.2 imply

$$P_{\pm,2,s} \leq P_{2,s} \leq P_{2,4} \leq \frac{\left(\binom{4}{1} + \binom{4}{2}\right)}{2^4} = 5/8 < 3/4 = P_{\pm,2,3}.$$

Similarly, it holds for all $s \geq 6$ that

$$P_{\pm,2,s} \leq P_{2,6} \leq \frac{\left(\binom{6}{2} + \binom{6}{3}\right)}{2^6} = 35/64. \tag{4.1}$$

Each of these bounds will now be used to bound one part of the probability. Denote by P_1 the probability that there exists a vector $x \in \mathbb{R}^n$ with $2 \leq \|x\|_0 \leq 5$ such that $M \sim \mathcal{E}^{(N,n)} x \in \{\pm 1\}^n$ and denote by P_2 the probability that there exists a vector $x \in \mathbb{R}^n$ with $6 \leq \|x\|_0 \leq n$. By the considerations above, it holds that

$$3 = \underset{2 \leq s \leq 5}{\operatorname{argmax}}\, P_{\pm,2,s}(1/2),$$

and the maximizer is unique. Observing that $|\mathfrak{F}_{\pm,2,3}| = 4$ (Lemma 3.10), Lemma 3.12 now implies that

$$P_1 \leq 4\binom{N}{3}(3/4)^N + o\left((3/4)^N\right).$$

On the other hand, applying Lemma 3.13 for $s_0 = 6$ and $Q = 7/10$ yields

$$P_2 = o\left((3/4)^N\right). \qquad \square$$

In the proof of Theorem 1.2, we did not have to fully exploit the symmetry of the set $\{\pm 1\}$, since we used the upper bound $P_{\pm,2,n}(1/2) \leq P_{2,n}(1/2)$ for $n \geq 4$. In order to be able to prove Theorem 1.2 for arbitrary $p \in (0,1)$, the symmetry of the set $\{\pm 1\}$ will be crucial.

Lemma 4.6. *Let $S^{(n)}$ be a Bernoulli random set with parameter $p \neq 1/2$ and let $A_1, A_2 \subset [n]$ be arbitrary subsets of cardinality k_1, k_2 resp. with $k_1, k_2 \leq n/2$. Then*

$$\mathbb{P}\left[S^{(n)} \in \{A_1, A_1^c\}\right] \geq \mathbb{P}\left[S^{(n)} \in \{A_2, A_2^c\}\right]$$

if and only if

$$k_1 \leq k_2.$$

Proof. By the definition of the Bernoulli random set $S^{(n)}$, it follows for $j = 1,2$ that

$$\mathbb{P}\left[S \in \{A_j, A_j^c\}\right] = p^{k_j} q^{n-k_j} + p^{n-k_j} q^{k_j}; \tag{4.2}$$

Without loss of generality, we may assume that $p > q$. Now consider the difference of the respective probabilities,

$$\begin{aligned} \mathbb{P}\left[S \in \{A_1, A_1^c\}\right] &- \mathbb{P}\left[S \in \{A_2, A_2^c\}\right] \\ &= p^{k_1} q^{n-k_1} + q^{k_1} p^{n-k_1} - p^{k_2} q^{n-k_2} - q^{k_2} p^{n-k_2} \\ &= \left(p^{k_2} q^{n-k_2}\right)\left(1 - \left(\tfrac{p}{q}\right)^{n-k_1-k_2}\right)\left(\left(\tfrac{p}{q}\right)^{k_1-k_2} - 1\right). \end{aligned} \tag{4.3}$$

The first factor on the right hand side of (4.3) is strictly positive. If $k_1 \geq k_2$, then both the second and the third factor in (4.3) are non-positive. If $k_2 < k_1$, the second factor is negative (this follows from $k_1 \leq n/2$) and the third factor is positive. Consequently, the left hand side of (4.3) is non-negative if and only if $k_1 \leq k_2$. This completes the proof. □

To translate this into an upper bound on $P_{\pm,2,n}(p)$, we need the following lemma.

Lemma 4.7. *Let $n \geq 2$, $\mathcal{A} \subset 2^n$ be a symmetric Sperner-2 family and denote by $\mathcal{L}_j \subset 2^n$, $j \in [n]$, the family of all sets $A \in 2^n$ with $|A| = j$. Then*

$$\mathcal{N} := \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_{n-1} \cup \mathcal{L}_n$$

is not a subfamily of $\mathcal{A}^\xi$ for any sign pattern $\xi \in \{\pm 1\}^n$.

Proof. Suppose that $\mathcal{N}$ is a subfamily of $\mathcal{A}^{\xi}$ for an arbitrary $\xi \in \{\pm 1\}^n$, which, since $\mathcal{A}$ is a symmetric Sperner-2 family, implies that there exists a subfamily $\mathcal{B} \subset \mathcal{A}^{\xi}$ with $\mathcal{B} \cap \mathcal{B}^{-1} = \emptyset$ such that

$$\mathcal{N} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_{n-1} \cup \mathcal{L}_n = \mathcal{B} \cup \mathcal{B}^{-1}. \tag{4.4}$$

Using the identity $(\mathcal{A}^{\xi})^{\xi} = \mathcal{A}$, we can conclude that $\mathcal{B}^{\xi}$ and thus $\mathcal{B}^{-\xi}$ are Sperner families. We will now consider the two occurring cases in more detail. In the first case, $\mathcal{B}$ or $\mathcal{B}^{-1}$ contains $\emptyset$ and at least one set of cardinality 1. If $\emptyset \in \mathcal{B}$ and no set of cardinality 1 is contained in $\mathcal{B}$, it follows that $\mathcal{L}_1 \subset \mathcal{B}^{-1}$, which also implies that $\mathcal{L}_1^{-1} = \mathcal{L}_{n-1} \subset \mathcal{B}$. By interchanging the roles of $\mathcal{B}$ and $\mathcal{B}^{-1}$, it follows that the remaining second case is the one where either $\mathcal{B}$ or $\mathcal{B}^{-1}$ is equal to $\mathcal{L}_0 \cup \mathcal{L}_{n-1}$. By symmetry, it is in both cases enough to only consider the family $\mathcal{B}$. In the first case, suppose that there exists an index $i \in [n]$ such that $\{\emptyset, \{i\}\} \subset \mathcal{B}$. For the symmetric difference of the two sets, Proposition 2.8 implies that

$$\emptyset^{\xi} \Delta \{i\}^{\xi} = \emptyset \Delta \{i\} = \{i\},$$

i.e., it either must hold that $\emptyset^{\xi} = \{i\}^{\xi} \uplus \{i\}$ or that $\{i\}^{\xi} = \emptyset^{\xi} \uplus \{i\}$. Any of the two cases contradicts the fact that $\mathcal{B}^{\xi}$ is a Sperner family, since $\{i\}^{\xi} \subset \emptyset^{\xi}$ or $\emptyset^{\xi} \subset \{i\}^{\xi}$. We can analogously handle the case where $\mathcal{B}^{-1}$ contains $\emptyset$ and one subset of cardinality 1. Next, suppose that $\mathcal{B} = \mathcal{L}_0 \cup \mathcal{L}_{n-1}$. Again by Proposition 2.8, it holds that for any set $B \in \mathcal{L}_{n-1}$

$$|\emptyset^{\xi} \Delta B^{\xi}| = |\emptyset \Delta B| = |B| = n - 1.$$

Consequently, $\mathcal{B}^{\xi}$ must contain $|\mathcal{L}_{n-1}| = n$ distinct sets B with $|\emptyset^{\xi} \Delta B| = n - 1$, as $(\cdot)^{\xi}$ is a bijection of $[n]$ onto itself. We claim that this already contradicts the assumption that $\mathcal{B}^{\xi}$ is a Sperner family. If $\emptyset^{\xi}$ is equal to $\emptyset$ or $[n]$, $\mathcal{B}^{\xi}$ with $|\mathcal{B}^{\xi}| \geq 2$ cannot be a Sperner family. So suppose that $|\emptyset^{\xi}| = k$ with $1 \leq k \leq n-1$. Note that $|\emptyset^{\xi} \Delta B^{\xi}| = n - 1$ either implies that B^{ξ} and $\emptyset^{\xi}$ are disjoint with $|B^{\xi}| = n - k - 1$, or that B^{ξ} is intersecting $\emptyset^{\xi}$ in a single element and it holds that $|B^{\xi}| = n - k + 1$. In 2^n, there exist $n-k$ sets B^{ξ} for which the first assumption is satisfied and k sets B^{ξ} for which the second one holds. Since $\mathcal{B}^{\xi}$ is a Sperner family, it cannot contain both a set B_1 of the first type and a set B_2 of the second type, as this would imply that $B_1 \subsetneq (\emptyset^{\xi})^c \subsetneq B_2$. Therefore, it contains at most $\max\{k, n-k\}$ subsets B^{ξ} with $|\emptyset^{\xi} \Delta B^{\xi}| = n - 1$. This yields the desired contradiction and completes the proof. □

We can now finally derive a strong upper bound on $P_{\pm,2,3}(p)$ for all $p \in (0,1)$ and small n using a greedy approach.

Lemma 4.8. *Let $S^{(n)}$ be a random BERNOULLI set with parameter p and let further $\mathcal{A} \subset 2^n$ be a symmetric SPERNER-2 family and let $\xi \in \{\pm 1\}^n$ an arbitrary sign pattern. Then*

$$\mathbb{P}\left[S \in \mathcal{A}^{\xi}\right] \leq \mathbb{P}\left[S \in \mathcal{B}\right], \tag{4.5}$$

where $\mathcal{B}$ is the subfamily of $2^n \setminus \{\{1\},\{1\}^c\}$ consisting of the $\binom{n}{\lfloor (n-1)/2 \rfloor}$ sets that are smallest and largest in cardinality.
In particular, for $n \in \{2,4,5,6\}$ and $p \in (0,1)$, it holds that

$$P_{\pm,2,n}(p) < P_{\pm,2,3}(p).$$

Proof. Let $\mathcal{A} \subset 2^n$ be a symmetric SPERNER-2 family. By Corollary 4.4, it follows that

$$|\mathcal{A}^{\xi}| \leq 2\binom{n}{\lfloor (n-1)/2 \rfloor}.$$

If $p = 1/2$, the assertion of the lemma follows from Lemma 4.5; we may therefore assume that $p \neq 1/2$. In addition to the cardinality constraint, Lemma 4.7 implies that, if $n \geq 2$, the family

$$\mathcal{N} = \mathcal{L}_0 \cup \mathcal{L}_1 \cup \mathcal{L}_{n-1} \cup \mathcal{L}_n$$

is not a subfamily of $\mathcal{A}^{\xi}$. Furthermore, since $\mathcal{A}$ is a symmetric SPERNER-2 family, it holds that $\mathcal{A} = \mathcal{A}^{-1}$. Using these three observations, we can now obtain the upper bound

$$\mathbb{P}\left[S \in \mathcal{A}^{\xi}\right] \leq \max\left\{\mathbb{P}\left[S \in \mathcal{B}\right] : \mathcal{B} \subset 2^n,\ \mathcal{B} = \mathcal{B}^{-1},\ |\mathcal{B}| \leq 2\binom{n}{\lfloor (n-1)/2 \rfloor},\ \mathcal{N} \not\subset \mathcal{B}\right\}. \tag{4.6}$$

Recall from Lemma 4.6 that for $p \neq 1/2$ and subsets $A_1, A_2 \subset [n]$ of cardinality k_1, k_2, resp., with $k_1, k_2 \leq n/2$,

$$\mathbb{P}\left[S \in \{A_1, A_1^c\}\right] \geq \mathbb{P}\left[S \in \{A_2, A_2^c\}\right]$$

holds if and only if

$$k_1 \leq k_2.$$

If we neglect the constraint $\mathcal{N} \not\subset \mathcal{B}$ on the right hand side of (4.6) and only consider the cardinality and symmetry constraints, we can therefore construct a maximizer $\mathcal{C}$ in a greedy manner by selecting the $\binom{n}{\lfloor (n-1)/2 \rfloor}$ sets of smallest cardinality and their complements. However, for $n \geq 5$, a family constructed in this way will always be a superfamily of $\mathcal{N}$ and will thus violate the subfamily constraint. Again by Lemma 4.6, the family of largest probability which is symmetric and satisfies both the cardinality and the subfamily constraint is the one where we replace one of the sets of cardinality 1 and its complement contained in $\mathcal{C}$ with the subset of smallest cardinality $k \leq n/2$ not yet con-

tained in $\mathcal{C}$ together with its complement. The resulting family is then, up to indexing, the family $\mathcal{B}$ as described in the first part of the lemma.
For the second part, recall from Remark 3.11 that

$$P_{\pm,2,2}(p) < P_{\pm,2,3}(p)$$

for all p, and from Lemma 3.10 that $P_{\pm,2,3}(p) = 1 - pq$. Now let $S^{(n)}, n \in \{4,5,6\}$ be Bernoulli random sets with parameter p. For the probability $P_{\pm,2,4}$ and the family $\mathcal{B}_4 = \{\emptyset, \{2\}, \{3\}, \{4\}\}$, (4.5) reads

$$P_{\pm,2,4}(p) \leq \mathbb{P}\left[S^{(4)} \in \mathcal{B}_4 \cup \mathcal{B}_4^{-1}\right] = p^4 + q^4 + 3(p^3q + pq^3) = 1 - pq(1 + 4pq) < 1 - pq,$$

where the second equality uses that $p + q = 1$. We even have equality, since

$$\begin{aligned} \mathcal{B}_4 \cup \mathcal{B}_4^{-1} &= \Big\{\emptyset, \{2\}, \{3\}, \{4\}\Big\} \cup \Big\{\{1,2,3,4\}, \{1,3,4\}, \{1,2,4\}, \{1,2,3\}\Big\} \\ &= \Big\{\emptyset, \{1,3,4\}, \{1,2,4\}, \{1,2,3\}\Big\} \cup \Big\{\{1,2,3,4\}, \{2\}, \{3\}, \{4\}\Big\} \\ &= \mathcal{C}_4 \cup \mathcal{C}_4^{-1}, \end{aligned}$$

where $\mathcal{C}_4 = \{\{1\}, \{2,3\}, \{2,4\}, \{3,4\}\}^{(-1,1,1,1)^T}$. Moving on to the case $n = 5$, inequality (4.5) yields that

$$P_{\pm,2,5}(p) \leq p^5 + q^5 + 4(p^4q + pq^4) + 5(p^3q^2 + q^3p^2) =: Q_5(p).$$

Bearing in mind that $q = 1 - p$, we can expand all terms in $P_{\pm,2,3}(p) - Q_5(p)$ and end up with a polynomial in p, namely

$$P_{\pm,2,3}(p) - Q_5(p) = 2p^4 - 4p^3 + 2p^2 = 2p^2(p-1)^2,$$

which is strictly positive for all $p \in (0,1)$. This implies for all $p \in (0,1)$ that

$$P_{\pm,2,5}(p) \leq Q_5(p) < P_{\pm,2,3}(p).$$

In the case where $n = 6$, we obtain from inequality (4.5) that

$$P_{\pm,2,6}(p) \leq p^6 + q^6 + 5(p^5q + pq^5) + 9(p^4q^2 + p^2q^4) =: Q_6(p).$$

Proceeding in the same way as in the previous case, we can find that

$$\begin{aligned} P_{\pm,2,3}(p) - Q_6(p) &= -10p^6 + 30p^5 - 28p^4 + 6p^3 + 2p^2 \\ &= -p^2(p-1)^2(p - {}^1\!/\!_{10}(5 - 3\sqrt{5}))(p - {}^1\!/\!_{10}(5 + 3\sqrt{5})). \end{aligned}$$

Since ${}^1\!/\!_{10}(5 - 3\sqrt{5}) < 0$ and ${}^1\!/\!_{10}(5 + 3\sqrt{5}) > 1$, $P_{\pm,2,3}(p) - Q_6(p)$ is strictly positive and we therefore have that $P_{\pm,2,6}(p) \leq Q_6(p) < P_{\pm,2,3}(p)$ for all $p \in (0,1)$. □

To prove Theorem 1.2, it remains to develop a strong bound on $P_{\pm,2,n}(p)$ for $n \geq 7$, where Lemma 4.8 fails to produce a bound which is uniformly smaller than $P_{\pm,2,3}(p)$, e.g., for $p = 3/4$.

5. Bounding $P_{k,n}$ and $P_{\pm,k,n}$ using the LYM-inequality

In this section, we will reduce the problem of bounding the quantities $P_{k,n}$ and $P_{\pm,k,n}$, which are defined in Lemma 3.4 in terms of modulated SPERNER-k families, to a similar problem involving only standard SPERNER-k families. This will put us in a position to use the LYM-inequality and generalizations thereof to bound $P_{k,n}$ and $P_{\pm,k,n}$. The LYM-inequality was independently proven by BOLLOBÁS [Bol65], LUBELL[Lub66], MESHALKIN[Meš63] and YAMAMOTO[Yam54].

Theorem 5.1 (LYM-Inequality [Bol65, Lub66, Meš63, Yam54]). *Let $\mathcal{A} \subset 2^n$ be a SPERNER-1 family and denote by $\mathcal{A}_k \subset 2^n$ the family of all $A \in \mathcal{A}$ with $|A| = k$. Then*

$$\sum_{k=0}^{n} \frac{|\mathcal{A}_k|}{\binom{n}{k}} \leq 1. \tag{5.1}$$

In the general case, an analogous inequality reads as follows:

Theorem 5.2 ([EFK05]). *Let $\mathcal{A}$ be a SPERNER-k family. Then*

$$\sum_{i=0}^{n} \frac{|\mathcal{A}_i|}{\binom{n}{i}} \leq k.$$

Equality holds only if $\mathcal{A} = \{A \subset [n] \ : \ |A| \in K\}$ for some $K \subset [n]$ with $|K| = k$.

In order to be able to apply Theorem 5.2 for modulated SPERNER-k families, we introduce the following notation.

Definition 5.3. For any $\mathcal{A} \subset 2^n$ and any disjoint $I, J \subset [n]$, let $\mathcal{A}_{I,J} \subset 2^J$ be the family defined by

$$\mathcal{A}_{I,J} = \{A \subset J | A \dot\cup I \in \mathcal{A}\} .$$

Similar to Definition 2.15, we now have the following definition.

Definition 5.4. Let the family $\mathcal{B} \sqcup I \subset 2^n$ for $I \subset [n]$ and $\mathcal{B} \subset 2^n$ be given by

$$\mathcal{B} \sqcup I = \{B \cup I | B \in \mathcal{B}\} .$$

When we want to stress that $\mathcal{B} \sqcap I$ contains only the empty set, we write $\mathcal{B} \dot\sqcup I$ instead of $\mathcal{B} \sqcup I$.

Remark 5.5. As for $\mathcal{A}, \mathcal{B} \subset 2^n$ and disjoint $I, J \subset [n]$, it holds that

$$\begin{aligned}(\mathcal{A} \cup \mathcal{B})_{I,J} &= \{A \subset J | A \dot\cup I \in \mathcal{A} \cup \mathcal{B}\} \\ &= \{A \subset J | A \dot\cup I \in \mathcal{A}\} \cup \{A \subset J | A \dot\cup I \in \mathcal{B}\} \\ &= \mathcal{A}_{I,J} \cup \mathcal{B}_{I,J},\end{aligned}$$

we say that $(\cdot)_{I,J}$ is *union compatible*. We also have

$$\begin{aligned}(\mathcal{A}^{-1})_{I,J} &= \left\{A \subset J | A \dot\cup I \in \mathcal{A}^{-1}\right\} \\ &= \{A \subset J | A^c \dot\cup J^c \setminus I \in \mathcal{A}\} \\ &= (A_{J^c \setminus I, J})^{-1}.\end{aligned}$$

In the following, for arbitrary $\xi \in \{\pm 1\}^n$ and arbitrary $J \subset [n]$, we will denote by $\xi_J \in \{\pm 1\}^J$ the restriction of ξ to the coordinates indexed by J. This allows us to state the following lemma:

Lemma 5.6. *For $\mathcal{A} \subset 2^n$, $\xi \in \{\pm 1\}^n$ and any disjoint $J \subset [n]$, it holds that*

$$\mathcal{A}^{\xi} = \dot\bigcup_{I \subset J^c} (\mathcal{A}_{I,J})^{\xi_J} \dot\sqcup I^{\xi_{J^c}} .$$

Furthermore, if $\mathcal{A}^{\xi} \subset 2^n$ is a SPERNER-k family, then $(\mathcal{A}_{I,J})^{\xi_J} \subset 2^J$ is a SPERNER-k family for all $I \subset J^c$.

Proof. If every entry of ξ is equal to 1, the first part of Lemma 5.6 directly follows from Definition 5.3. Now let $\xi \in \{\pm 1\}^n$ be arbitrary. Recall that $(\cdot)^\xi$ is union compatible. This allows us to write

$$\mathcal{A}^\xi = \dot{\bigcup_{I \subset J^c}} \left(\mathcal{A}_{I,J} \sqcup I\right)^\xi ,$$

which is also a disjoint union, since $(\cdot)^\xi$ is a bijection of 2^n onto itself. As $\mathcal{A}_{I,J} \subset 2^J$ and $I \subset J^c$, we can rewrite this as

$$\mathcal{A}^\xi = \dot{\bigcup_{I \subset J^c}} (\mathcal{A}_{I,J})^{\xi_J} \sqcup I^{\xi_{J^c}} .$$

Now let $\mathcal{A}^\xi \subset 2^n$ be a SPERNER-k family and suppose that $\mathcal{A}_{I,J}^{\xi_J}$ is not a SPERNER-k family for some disjoint $I, J \subset [n]$, meaning that there exist $k+1$ sets $B_1, B_2, \ldots, B_{k+1} \in \mathcal{A}_{I,J}$ such that

$$B_1^{\xi_J} \subsetneq B_2^{\xi_J} \subsetneq \cdots \subsetneq B_{k+1}^{\xi_J}, \tag{5.2}$$

and therefore also

$$\left(B_1^{\xi_J} \mathbin{\dot\cup} I^{\xi_{J^c}}\right) \subsetneq \left(B_2^{\xi_J} \mathbin{\dot\cup} I^{\xi_{J^c}}\right) \subsetneq \cdots \subsetneq \left(B_{k+1}^{\xi_J} \mathbin{\dot\cup} I^{\xi_{J^c}}\right). \tag{5.3}$$

By construction of $\mathcal{A}_{I,J}$, there must exist $k+1$ sets $A_1, A_2, \ldots, A_{k+1} \in \mathcal{A}$ such that

$$A_j = B_j \mathbin{\dot\cup} I, \qquad j \in [k+1],$$

which translates to

$$A_j^\xi = B_j^{\xi_J} \mathbin{\dot\cup} I^{\xi_{J^c}}, \qquad j \in [k+1].$$

Together with the chain of inclusions (5.3), this now contradicts the assumption that $\mathcal{A}^\xi$ is a SPERNER-k family and completes the proof. □

The following theorem now reduces probability estimates for a BERNOULLI random set $S^{(n)}$ to estimates of BERNOULLI random sets $S^{(J)}$ for $J \subset [n]$. While the result for SPERNER-k families is straight-forward, the corresponding estimate for symmetric SPERNER-k families is more involved.

Theorem 5.7. *Let $\mathcal{A} \subset 2^n$ be a SPERNER-k family, $\xi \in \{\pm 1\}^n$ be an arbitrary sign pattern and $J \subset [n]$ be an arbitrary index set. Then, for the BERNOULLI random sets $S^{(n)}$ and $S^{(J)}$ with parameter p, it holds that*

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^\xi\right] \le \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B}\ \text{SPERNER-}k}} \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J}\right].$$

If $\mathcal{A}$ is symmetric, it holds that

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] \leq \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B}\ \textsc{Sperner}\text{-}\lceil k/2 \rceil}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-\xi_J}\right]\right). \tag{5.4}$$

Proof. Let $\mathcal{A} \subset 2^n$ and $\xi \in \{\pm 1\}^n$ be arbitrary. By Lemma 5.6, we can represent $\mathcal{A}^{\xi}$ by the disjoint union

$$\mathcal{A}^{\xi} = \dot{\bigcup_{I \subset J^c}} (\mathcal{A}_{I,J}^{\xi_J}) \sqcup I^{\xi_{J^c}}.$$

This allows us to write

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = \sum_{A \in \mathcal{A}} \mathbb{P}\left[S^{(n)} = A^{\xi}\right] = \sum_{I \subset J^c} \sum_{A \in \mathcal{A}_{I,J}} \mathbb{P}\left[S^{(n)} = A^{\xi_J} \dot{\cup} I^{\xi_{J^c}}\right]. \tag{5.5}$$

Since $A \in \mathcal{A}_{I,J}^{\xi_J} \subset 2^J$, it holds that for any subset $I \subset J^c$

$$(A^{\xi_J} \dot{\cup} I^{\xi_{J^c}})^c = (J \setminus A^{\xi_J}) \dot{\cup} (J^c \setminus I^{\xi_{J^c}}).$$

We can now rewrite the terms on the right-hand side of (5.5) as

$$\begin{aligned}
\mathbb{P}\left[S^{(n)} = A^{\xi_J} \dot{\cup} I^{\xi_{J^c}}\right] &= p^{|A^{\xi_J}|+|I^{\xi_{J^c}}|} q^{n-(|A^{\xi_J}|+|I^{\xi_{J^c}}|)} \\
&= p^{|A^{\xi_J}|} q^{|J|-|A^{\xi_J}|} p^{|I^{\xi_{J^c}}|} q^{|J^c|-|I^{\xi_{J^c}}|} \\
&= \mathbb{P}\left[S^{(J)} = A^{\xi_J}\right] \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right].
\end{aligned}$$

In (5.5), we therefore get

$$\begin{aligned}
\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] &= \sum_{I \subset J^c} \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \sum_{A \in \mathcal{A}_{I,J}} \mathbb{P}\left[S^{(J)} = A^{\xi_J}\right] \\
&= \sum_{I \subset J^c} \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \mathbb{P}\left[S^{(J)} \in \mathcal{A}_{I,J}^{\xi_J}\right].
\end{aligned} \tag{5.6}$$

Suppose now that $\mathcal{A} \subset 2^n$ is a Sperner-k family. Then Lemma 5.6 with $\xi = 1$ implies that $\mathcal{A}_{I,J} \subset 2^J$ also is a Sperner-k family. With (5.6), it follows that

$$\begin{aligned}
\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] &\leq \left(\max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B}\ \text{Sperner-}k}} \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J}\right]\right) \sum_{I \subset J^c} \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \\
&= \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B}\ \textsc{Sperner}\text{-}k}} \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J}\right],
\end{aligned}$$

since $(\cdot)^{\xi_{J^c}}$ is a bijection of 2^{J^c} onto itself, and thus

$$\sum_{I \subset J^c} \mathbb{P}\left[S^{J^c} = I^{\xi_{J^c}}\right] = 1.$$

For the second statement, we may assume that $J \neq \emptyset$. Let $d \in J$ be arbitrary and assume that $\mathcal{A} \subset 2^n$ is a symmetric SPERNER-k family. If k is even, it is of the form

$$\mathcal{A} = \bigcup_{i=1}^{k/2} \mathcal{A}_i \cup \mathcal{A}_i^{-1},$$

where $\mathcal{A}_i \subset 2^n$ is a SPERNER family for all $i \in [k/2]$. If k is odd, $\mathcal{A}$ is of the form

$$\mathcal{A} = \mathcal{A}_0 \cup \bigcup_{i=1}^{\lfloor k/2 \rfloor} \mathcal{A}_i \cup \mathcal{A}_i^{-1},$$

where the family $\mathcal{A}_i \subset 2^n$ is a SPERNER family for all $1 \leq i \leq \lfloor k/2 \rfloor$ and it additionally holds that $\mathcal{A}_0 = \mathcal{A}_0^{-1}$. Let $\mathcal{A}_{\lceil k/2 \rceil} \subset 2^n$ be the family of all sets contained in $\mathcal{A}_0$ which do not contain the index $d \in J$. Note that as a subfamily of a SPERNER family, $\mathcal{A}_{\lceil k/2 \rceil}$ must also be a SPERNER family. In this way, $\mathcal{A}_0 = \mathcal{A}_0^{-1}$ implies that $\mathcal{A}_{\lceil k/2 \rceil} \cup \mathcal{A}_{\lceil k/2 \rceil}^{-1} = \mathcal{A}_0$. It follows that if k is even or odd, we can write the symmetric SPERNER-k family $\mathcal{A}$ as

$$\mathcal{A} = \bigcup_{i=1}^{\lceil k/2 \rceil} \mathcal{A}_i \cup \mathcal{A}_i^{-1}, \tag{5.7}$$

where $\mathcal{A}_i \subset 2^n$ is a SPERNER family for all $1 \leq i \leq \lceil k/2 \rceil$ and, if k is odd, it additionally holds that $\mathcal{A}_{\lceil k/2 \rceil}$ contains no sets $A \subset [n]$ with $d \in A$. By possibly removing sets from some of the families $\mathcal{A}_i$, we may assume that all $\mathcal{A}_i, \mathcal{A}_i^{-1}$, $1 \leq i \leq \lceil k/2 \rceil$ are pairwise disjoint. Let $\mathcal{G} \subset 2^{J^c}$ be a family with

$$\mathcal{G} \mathbin{\dot\cup} \mathcal{G}^{-1} = 2^{J^c}.$$

Such a family always exists. For each $I \in \mathcal{G}$, let $p_{I,J}$ be the probability

$$\begin{aligned} p_{I,J} = \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \mathbb{P}\left[S^{(J)} \in \mathcal{A}_{I,J}^{\xi_J}\right] \\ + \mathbb{P}\left[S^{(J^c)} = J^c \setminus I^{\xi_{J^c}}\right] \mathbb{P}\left[S^{(J)} \in \mathcal{A}_{J^c \setminus I,J}^{\xi_J}\right]; \end{aligned} \tag{5.8}$$

we can now rewrite (5.6) as

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi}\right] = \sum_{I \in \mathcal{G}} p_{I,J}. \tag{5.9}$$

By Remark 5.5, it holds that for $I \subset J^c$, $(\cdot)$ is union compatible and one has $(\mathcal{A}^{-1})_{I,J} =$

$(\mathcal{A}_{J^c\setminus I,J})^{-1}$. Together with identity (5.7), it follows that

$$\mathcal{A}_{I,J} = \bigcup_{i=1}^{\lceil k/2\rceil} (\mathcal{A}_i)_{I,J} \cup ((\mathcal{A}_i)_{J^c\setminus I,J})^{-1}.$$

Now let $I \in \mathcal{J}$ be fixed. The families $\mathcal{A}_{I,J} \subset 2^J$ and $\mathcal{A}_{J^c\setminus I,J} \subset 2^J$ can now be written as

$$\mathcal{A}_{I,J} = \bigcup_{i=1}^{\lceil k/2\rceil} \left(\mathcal{B}_{i,1} \cup \mathcal{B}_{i,2}^{-1}\right) \quad \text{and} \quad \mathcal{A}_{J^c\setminus I,J} = \bigcup_{i=1}^{\lceil k/2\rceil} \left(\mathcal{B}_{i,2} \cup \mathcal{B}_{i,1}^{-1}\right). \tag{5.10}$$

where for each $i \subset \left[\lceil k/2\rceil\right]$, we set

$$\mathcal{B}_{i,1} = (\mathcal{A}_i)_{I,J} \quad \text{and} \quad \mathcal{B}_{i,2} = (\mathcal{A}_i)_{J^c\setminus I,J}.$$

Since $\mathcal{A}_i \subset 2^n$ is a Sperner family for each $1 \leq i \leq \lceil k/2\rceil$, Lemma 5.6 implies that for each i, the families $\mathcal{B}_{i,1}, \mathcal{B}_{i,2} \subset 2^J$ and therefore also $\mathcal{B}_{i,1}^{-1}, \mathcal{B}_{i,2}^{-1} \subset 2^J$ are Sperner families. Bearing in mind that $\mathcal{A}_i, \mathcal{A}_i^{-1}$, $1 \leq i \leq \lceil k/2\rceil$ are pairwise disjoint and that $(\cdot)^\nu$ is a bijection of 2^J onto itself for any sign pattern $\nu \in \{\pm 1\}^J$, it must hold that $B_{i,1}^{\xi_J}, B_{i,2}^{-\xi_J} \subset 2^J$, $1 \leq i \leq \lceil k/2\rceil$ are also pairwise disjoint. Analogously, the same must also hold for $B_{i,2}^{\xi_J}, B_{i,1}^{-\xi_J} \subset 2^J$, $1 \leq i \leq \lceil k/2\rceil$. Because of this and since $(\cdot)^{\xi_J}$ is union compatible, it follows with (5.10) and (5.11) for $p_{I,J}$, $I \in \mathcal{J}$ defined in (5.8) and

$$a_{I,J} := \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \quad \text{and} \quad b_{I,J} := \mathbb{P}\left[S^{(J^c)} = J^c \setminus I^{\xi_{J^c}}\right], \tag{5.11}$$

that

$$\begin{aligned} p_{I,J} &= \mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}}\right] \mathbb{P}\left[S^{(J)} \in \mathcal{A}_{I,J}^{\xi_J}\right] \\ &\quad + \mathbb{P}\left[S^{(J^c)} = J^c \setminus I^{\xi_{J^c}}\right] \mathbb{P}\left[S^{(J)} \in \mathcal{A}_{J^c\setminus I,J}^{\xi_J}\right] \\ &= a_{I,J}\, \mathbb{P}\left[S^{(J)} \in \dot{\bigcup_{i=1}^{\lceil k/2\rceil}} \left(\mathcal{B}_{i,1}^{\xi_J} \mathbin{\dot\cup} \mathcal{B}_{i,2}^{-\xi_J}\right)\right] + b_{I,J}\, \mathbb{P}\left[S^{(J)} \in \dot{\bigcup_{i=1}^{\lceil k/2\rceil}} \left(\mathcal{B}_{i,2}^{\xi_J} \mathbin{\dot\cup} \mathcal{B}_{i,1}^{-\xi_J}\right)\right] \\ &= a_{I,J} \sum_{i=1}^{\lceil k/2\rceil} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{-\xi_J}\right]\right) \\ &\quad + b_{I,J} \sum_{i=1}^{\lceil k/2\rceil} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{-\xi_J}\right]\right) \\ &= a_{I,J} \sum_{i=1}^{\lceil k/2\rceil} (B_{i,1} + \bar{B}_{i,2}) + b_{I,J} \sum_{i=1}^{\lceil k/2\rceil} (B_{i,2} + \bar{B}_{i,1}), \end{aligned} \tag{5.12}$$

where for each $i \in \lceil k/2 \rceil$, we set

$$\begin{aligned} B_{i,1} &:= \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{\xi_J}\right], & \bar{B}_{i,1} &:= \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{-\xi_J}\right], \\ B_{i,2} &:= \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{\xi_J}\right], & \bar{B}_{i,2} &:= \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{-\xi_J}\right]. \end{aligned} \tag{5.13}$$

We now aim to find an upper bound on $p_{I,J}$ which takes the structure of the appearing families into account. To this end, we claim that there exists a partition $P \cup Q = \left[\lceil k/2 \rceil\right]$ such that

$$p_{I,J} \le (a_{I,J} + b_{I,J})\left(\sum_{i \in P}(B_{i,1} + \bar{B}_{i,1}) + \sum_{i \in Q}(B_{i,2} + \bar{B}_{i,2})\right). \tag{5.14}$$

Indeed, let P be the set of all indices $i \in \left[\lceil k/2 \rceil\right]$ with

$$\operatorname*{argmax}_{j \in \{1,2\}} \left(a_{I,J}\bar{B}_{i,j} + b_{I,J}B_{i,j}\right) = 1,$$

and let Q be the set of all indices $i \in \left[\lceil k/2 \rceil\right]$ with

$$\operatorname*{argmax}_{j \in \{1,2\}} \left(a_{I,J}\bar{B}_{i,j} + b_{I,J}B_{i,j}\right) = 2.$$

In this way, it follows for arbitrary $i \in P$ that

$$\begin{aligned} a_{I,J}&(B_{i,1} + \bar{B}_{i,2}) + b_{I,J}(\bar{B}_{i,1} + B_{i,2}) - (a_{I,J} + b_{I,J})(B_{i,1} + \bar{B}_{i,1}) \\ &= a_{I,J}(\bar{B}_{i,2} - \bar{B}_{i,1}) + b_{I,J}(B_{i,2} - B_{i,1}) \\ &= (a_{I,J}\bar{B}_{i,2} + b_{I,J}B_{i,2}) - (a_{I,J}\bar{B}_{i,1} + b_{I,J}B_{i,1}) \le 0, \end{aligned}$$

which after rearranging reads

$$a_{I,J}(B_{i,1} + \bar{B}_{i,2}) + b_{I,J}(\bar{B}_{i,1} + B_{i,2}) \le (a_{I,J} + b_{I,J})(B_{i,1} + \bar{B}_{i,1}).$$

Analogously, it follows for each $i \in Q$ that

$$a_{I,J}(B_{i,1} + \bar{B}_{i,2}) + b_{I,J}(\bar{B}_{i,1} + B_{i,2}) \le (a_{I,J} + b_{I,J})(B_{i,2} + \bar{B}_{i,2}).$$

Together, these inequalities imply (5.14), or equivalently

$$\begin{aligned} p_{I,J} \le (a_{I,J} + b_{I,J})\Bigg(&\sum_{i \in P}\left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,1}^{-\xi_J}\right]\right) \\ &+ \sum_{i \in Q}\left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}_{i,2}^{-\xi_J}\right]\right)\Bigg). \end{aligned} \tag{5.15}$$

where we inserted the definition (5.13) in the last step. Now define the family

$$\mathcal{C} = \bigcup_{i \in P} \mathcal{B}_{i,1} \cup \bigcup_{i \in Q} \mathcal{B}_{i,2}^{-1} \subset 2^J,$$

which is a Sperner-$\lceil k/2 \rceil$ family by Lemma 2.3. Using once again that $B^{\xi_J}_{i,1}, B^{-\xi_J}_{i,2} \subset 2^J$, $1 \leq i \leq \lceil k/2 \rceil$ and also $B^{\xi_J}_{i,2}, B^{-\xi_J}_{i,1} \subset 2^J$, $1 \leq i \leq \lceil k/2 \rceil$ are pairwise disjoint and the fact that $(\cdot)^{-\xi_J}$ is union compatible, inequality (5.15) can be rewritten to

$$\begin{aligned} p_{I,J} &\leq (a_{I,J} + b_{I,J}) \left(\mathbb{P}\left[S^{(J)} \in \mathcal{C}^{\xi_J} \right] + \mathbb{P}\left[S^{(J)} \in \mathcal{C}^{-\xi_J} \right] \right) \\ &\leq (a_{I,J} + b_{I,J}) \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B} \text{ Sperner-}\lceil k/2 \rceil}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J} \right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-\xi_J} \right] \right). \end{aligned}$$

Combining this inequality with (5.9), we obtain

$$\begin{aligned} \mathbb{P}\left[S^{(n)} \in \mathcal{A}^{\xi} \right] &= \sum_{I \in \mathcal{G}} p_{I,J} \\ &\leq \sum_{I \in \mathcal{G}} (a_{I,J} + b_{I,J}) \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B} \text{ Sperner-}\lceil k/2 \rceil}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J} \right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-\xi_J} \right] \right) \\ &= \max_{\substack{\mathcal{B} \subset 2^J \\ \mathcal{B} \text{ Sperner-}\lceil k/2 \rceil}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J} \right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-\xi_J} \right] \right), \end{aligned}$$

since

$$\begin{aligned} \sum_{I \in \mathcal{G}} (a_{I,J} + b_{I,J}) &= \sum_{I \in \mathcal{G}} \left(\mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}} \right] + \mathbb{P}\left[S^{(J^c)} = J^c \setminus I^{\xi_{J^c}} \right] \right) \\ &= \sum_{I \in \mathcal{G}} \left(\mathbb{P}\left[S^{(J^c)} = I^{\xi_{J^c}} \right] + \mathbb{P}\left[S^{(J^c)} = (J^c \setminus I)^{\xi_{J^c}} \right] \right) = 1. \end{aligned} \tag{5.16}$$

The first equality in (5.16) is implied by

$$J^c \setminus I^{\xi_{J^c}} = I^{-\xi_{J^x}} = (J^c \setminus I)^{\xi_{J^c}};$$

the last equality follows by construction of the family $\mathcal{G}$ and the fact that $(\cdot)^{\xi_{J^c}}$ is a bijection of 2^{J^c} onto itself. This completes the proof. □

Remark 5.8. As a consequence of the construction made in (5.7), in order to come up with a smaller upper bound in the case where k is odd, we may additionally require in the maximum in (5.14) that one of the Sperner families $\mathcal{B}_\ell$ in the decomposition of the Sperner-$\lceil k/2 \rceil$ family $\mathcal{B}$, does not contain any complementing sets.

Theorem 5.7 now puts us in a position where we are able to apply the LYM-inequality in order to bound $P_{\pm,2,n}(p)$ and $P_{1,n}(p)$.

Corollary 5.9. *Let $\mathcal{A} \subset 2^n$ be a Sperner family, $\xi \in \{\pm1\}^n$ be a sign pattern, $J(\xi) = \{j|\xi_j = 1\}$ and $\bar{n}$ be an arbitrary integer with $0 \le \bar{n} \le \max\{|J(\xi)|, |J(-\xi)|\}$. Then for the Bernoulli random set $S^{(n)}$ with parameter p, it holds that*

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^\xi\right] \le \max_{0\le k\le \bar{n}} \binom{\bar{n}}{k} p^k q^{\bar{n}-k}. \tag{5.17}$$

If the family $\mathcal{A} \subset 2^n$ is a symmetric Sperner-2 family, we have

$$\mathbb{P}\left[S^{(n)} \in \mathcal{A}^\xi\right] \le \max_{0\le k\le \bar{n}} \binom{\bar{n}}{k} \left(p^k q^{\bar{n}-k} + p^{\bar{n}-k} q^k\right). \tag{5.18}$$

Consequently, for $0 \le \bar{n} \le \lceil n/2 \rceil$ and any $p \in (0,1)$, it holds that

$$P_{1,n}(p) \le \max_{0\le k\le \bar{n}} \binom{\bar{n}}{k} p^k q^{\bar{n}-k}$$

and

$$P_{\pm,2,n}(p) \le \max_{0\le k\le \bar{n}} \binom{\bar{n}}{k} \left(p^k q^{\bar{n}-k} + p^{\bar{n}-k} q^k\right).$$

Proof. We will only prove (5.18), since the proof of (5.17) can be done analogously. Let $\mathcal{A} \subset 2^n$ be an arbitrary symmetric Sperner-2 family and $\xi \in \{\pm1\}^n$ be arbitrary. Without loss of generality we may assume that ξ has more positive than negative entries and therefore $\max\{|J(\xi)|, |J(-\xi)|\} = |J(\xi)|$. Otherwise, we can rewrite $\mathcal{A}^\xi = (\mathcal{A}^{-1})^{-\xi}$ and note that the property of $\mathcal{A}$ being a Sperner-k or symmetric Sperner-k family is invariant under $(\cdot)^{-1}$. Applying Theorem 5.7 with $J \subset J(\xi)$ and $|J| = \bar{n}$ implies that

$$\begin{aligned} P\left[S^{(n)} \in \mathcal{A}^\xi\right] &\le \max_{\substack{\mathcal{B}\subset 2^J \\ \mathcal{B}\text{ Sperner}}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}^{\xi_J}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-\xi_J}\right]\right) \\ &= \max_{\substack{\mathcal{B}\subset 2^J \\ \mathcal{B}\text{ Sperner}}} \left(\mathbb{P}\left[S^{(J)} \in \mathcal{B}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-1}\right]\right), \end{aligned} \tag{5.19}$$

since $\xi_J \in \{\pm1\}^J$ is the constant 1 vector as $J \subset J(\xi)$. For a Sperner family $\mathcal{C} \subset 2^J$ and for $0 \le \ell \le n$, denote by $\mathcal{C}_\ell \subset \mathcal{C}$ the family of all $C \in \mathcal{C}$ of cardinality ℓ, and note that $(C^{-1})_\ell = (C_{\bar{n}-\ell})^{-1}$. It follows that

$$\mathbb{P}\left[S^{(J)} \in \mathcal{B}\right] + \mathbb{P}\left[S^{(J)} \in \mathcal{B}^{-1}\right] = \sum_{\ell=0}^{\bar{n}} |\mathcal{B}_\ell| p^\ell q^{\bar{n}-\ell} + \sum_{\ell=0}^{\bar{n}} |(\mathcal{B}^{-1})_\ell| p^\ell q^{\bar{n}-\ell}$$

$$\begin{aligned}
&= \sum_{\ell=0}^{\bar{n}} |\mathcal{B}_\ell| p^\ell q^{\bar{n}-\ell} + \sum_{\ell=0}^{\bar{n}} |\mathcal{B}_\ell| p^{\bar{n}-\ell} q^\ell = \sum_{\ell=0}^{\bar{n}} |\mathcal{B}_\ell| \left(p^\ell q^{\bar{n}-\ell} + p^{\bar{n}-\ell} q^\ell\right) \\
&= \sum_{\ell=0}^{\bar{n}} \frac{|\mathcal{B}_\ell|}{\binom{\bar{n}}{\ell}} \binom{\bar{n}}{\ell} \left(p^\ell q^{\bar{n}-\ell} + p^{\bar{n}-\ell} q^\ell\right) \\
&\leq \max_{0 \leq \ell \leq \bar{n}} \binom{\bar{n}}{\ell} \left(p^\ell q^{\bar{n}-\ell} + p^{\bar{n}-\ell} q^\ell\right),
\end{aligned}$$

where the last step follows from the LYM-inequality (Theorem 5.1). With (5.19) this yields (5.18). Since for arbitrary $\xi \in \{\pm 1\}^n$, we have $\max\{|J(\xi)|, |J(-\xi)|\} = |J(\xi)| \geq \lceil n/2 \rceil$ and (5.18) and (5.17) are independent of ξ, the last two claims of the corollary follow from the definition of $P_{k,n}$ and $P_{\pm,k,n}$ in Lemma 3.4. □

Next, we will prove Corollary 1.6 and Theorem 1.7.

Proof of Corollary 1.6 and Theorem 1.7. With Lemma 3.4, the assertions directly follow from Corollary 5.9. □

With Corollary 5.9, we can now bound $P_{\pm,2,n}(p)$ for $n \geq 7$.

Corollary 5.10. *Let $n \geq 7$ and $p \in (0,1)$, $p \neq 1/2$. Then*

$$P_{\pm,2,n}(p) < P_{\pm,2,3}(p). \tag{5.20}$$

Furthermore, for $n \geq 15$ and $p \in (0,1)$, one has

$$P_{\pm,2,n}(p) < \left(P_{\pm,2,3}(p)\right)^2. \tag{5.21}$$

Proof. For $n \geq 7$, we set $\bar{n} = 4 \leq \lceil n/2 \rceil$. Then, Corollary 5.9 implies that

$$P_{\pm,2,n}(p) \leq \max_{0 \leq k \leq 4} \binom{4}{k} \left(p^k q^{4-k} + p^{4-k} q^k\right) = \max_{0 \leq k \leq 2} Q_{4,k}(p), \tag{5.22}$$

where

$$Q_{4,k}(p) = \binom{4}{k} \left(p^k q^{4-k} + p^{4-k} q^k\right),$$

and the last inequality is by symmetry. Note that $P_{\pm,2,3}(p) = 1 - pq = p^2 + q^2 + pq$ by Lemma 3.10, it clearly holds that $Q_{4,0}(p) = p^4 + q^4 < p^2 + q^2 < P_{\pm,2,3}(p)$ for all $p \in (0,1)$. Next, we note that

$$Q_{4,1}(p) = 4(p^3 q + p q^3) = 1 - (p^4 + q^4 + 6p^2 q^2),$$

and hence

$$\begin{aligned} P_{\pm,2,3}(p) - Q_{4,1}(p) &= p^4 + q^4 + 6p^2q^2 - pq \\ &= p^4 + q^4 + 6p^2q^2 - pq(p+q)^2 \\ &= p^2(p + {}^1\!/\!{}_2 q)^2 + q^2(q + {}^1\!/\!{}_2 p)^2 + {}^7\!/\!{}_2 p^2q^2 > 0. \end{aligned}$$

For $k = 2$, it holds that

$$\begin{aligned} P_{\pm,2,3}(p) - Q_{4,2}(p) &= 1 - pq - 12p^2q^2 = -12p^4 + 24p^3 - 11p^2 - p + 1 \\ &= -12(p - {}^1\!/\!{}_2)^2 \left(p - {}^1\!/\!{}_6\left(3 - \sqrt{21}\right)\right)\left(p - {}^1\!/\!{}_6\left(3 + \sqrt{21}\right)\right). \end{aligned}$$

Since ${}^1\!/\!{}_6\left(3 - \sqrt{21}\right) < 0$ and ${}^1\!/\!{}_6\left(3 + \sqrt{21}\right) > 1$, it follows that $Q_{4,2}(p) < P_{\pm,2,3}(p)$ for all $p \in (0,1) \setminus \{{}^1\!/\!{}_2\}$. By (5.22), this proves the first part of the corollary. For the second part, we also aim to use Corollary 5.9. Since (5.18) is invariant under interchanging p and q, we may assume that $0 < p \leq {}^1\!/\!{}_2$. In the case of $n \geq 15$, (5.18) implies that

$$P_{\pm,2,n}(p) \leq \max_{0 \leq k \leq 8} \binom{8}{k}\left(p^kq^{8-k} + p^{8-k}q^k\right) = \max_{0 \leq k \leq 4} Q_{8,k}(p),$$

where

$$Q_{8,k}(p) := \binom{8}{k}\left(p^kq^{8-k} + p^{8-k}q^k\right). \tag{5.23}$$

We will now show that (5.23) is always smaller than $(P_{\pm,2,3}(p))^2$. First, consider $Q_{8,0}(p)$. For any $p \in (0,1)$, it holds that

$$Q_{8,0}(p) = p^8 + q^8 < p^4 + q^4 < \left(p^2 + q^2 + pq\right)^2 = (1 - pq)^2 = (P_{\pm,2,3}(p))^2.$$

For $Q_{8,k}$ with $k \geq 1$, consider the derivative

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p} Q_{8,k}(p) &= \binom{8}{k}\left(kp^{k-1}q^{8-k} - (8-k)p^kq^{8-k-1} + (8-k)p^{8-k-1}q^k - kp^{8-k}q^{k-1}\right) \\ &= \binom{8}{k}\left(p^{k-1}q^{8-k-1}(k(1-p) - (8-k)p) + p^{8-k-1}q^{k-1}((8-k)(1-p) - kp)\right) \\ &= \binom{8}{k}\left(p^{k-1}q^{8-k-1}(k - 8p) + p^{8-k-1}q^{k-1}((8-k) - 8p)\right), \end{aligned}$$

which has the same zeros and the same sign as the function

$$f(p) = (k - 8p) + ({}^p\!/\!{}_q)^{8-2k}\left((8-k) - 8p\right). \tag{5.24}$$

A dircet calculation shows that, for $k = 3, 4$, (5.24) only vanishes for $p = {}^1\!/\!{}_2$, where the sign changes from positive to negative, implying that both and $Q_{8,3}$ and $Q_{8,4}$ attain their maximum at $p = 1/2$. As $P_{\pm,2,3} = 1 - pq$ also attains its minimum at $1/2$, the estimate

$$Q_{8,3}(1/2) = 58/128 < Q_{8,4}(1/2) = 70/128 < 9/16 = \left(P_{\pm,2,3}(1/2)\right)^2$$

shows the result. For $k = 1, 2$, note that (5.24) is positive for $p \leq k/8$ and, for $\varepsilon > 0$, setting $p = \frac{k+\varepsilon}{8}$ and therefore $q = \frac{8-k-\varepsilon}{8}$, (5.24) is negative if and only if

$$\varepsilon > (8 - 2k - \varepsilon)\left(\frac{k+\varepsilon}{8-k-\varepsilon}\right)^{8-2k}. \tag{5.25}$$

For any ε that satisfies (5.25), the maximum of $Q_{8,k}$ is therefore attained in the interval $(\frac{k}{8}, \frac{k+\varepsilon}{8})$, implying that

$$Q_{8,k} \leq \binom{8}{k}\left(\left(\frac{k+\varepsilon}{8}\right)^k\left(\frac{8-k}{8}\right)^{8-k} + \left(\frac{8-k}{8}\right)^k\left(\frac{k+\varepsilon}{8}\right)^{8-k}\right). \tag{5.26}$$

Choosing $\varepsilon = 0.06$, which is a valid choice in (5.25) for $k = 1, 2$, inequality (5.26) now implies that $Q_{8,1}(p) < 0.42$ and $Q_{8,2}(p) < 0.34$ for all $p \in (0, 1)$; both upper bounds are clearly smaller than $\left(P_{\pm,2,3}(p)\right)^2 \geq 9/16$. This completes the proof. □

We have now everything at our disposal to prove Theorem 1.2 for $p \neq 1/2$.

Proof of Theorem 1.2 for $p \neq 1/2$. As in the case of $p = 1/2$, we aim to use Lemma 3.12 and Lemma 3.13. Let $M \sim \mathcal{E}^{(N,n)}$ be a Bernoulli random matrix with parameter $p \neq 1/2$, and note that $V = \{\pm 1\}$ is symmetric. Further, denote by P_1 the probability that there exists a vector $x \in \mathbb{R}^n$ with $\|x\|_0 = s$, $2 \leq s \leq 14$ such that $Mx \in \{\pm 1\}^n$ and denote by P_2 the probability that there exists a vector $x \in \mathbb{R}^n$ with $15 \leq \|x\|_0 \leq n$ such that $Mx \in \{\pm 1\}^n$. Lemma 4.8 and the first part of Corollary 5.10 now imply that

$$3 = \operatorname*{argmax}_{2 \leq s \leq 14} P_{\pm,2,s}(p)$$

is the unique minimizer. Furthermore, it holds by Lemma 3.10 that $|\mathfrak{F}_{\pm,2,3}| = 4$ and $P_{\pm,2,3}(p) = 1 - p(1-p)$. Lemma 3.12 therefore implies that

$$P_1 \leq 4\binom{n}{3}(1 - p(1-p))^N + o\left((1 - p(1-p))^N\right).$$

Applying now Lemma 3.13 for $\bar{Q} = P_{\pm,2,3} = 1 - pq$, we get

$$P_2 = o\left((1 - p(1-p))^N\right).$$

Here, condition (3.12) follows as $1 - pq = p^2 + q^2 + pq$ and (3.13) has been established in the second part of Corollary 5.10.

This completes the proof. □

We can now also prove Theorem 1.8.

Proof of Theorem 1.8. Proceeding in the same way as in the proof of Lemma 2.13 in (2.8), the assertion of the theorem directly follows from Theorem 1.2. □

Chapter II.

Ultrasonic Nondestructive Testing with Random Measurements

6. Introduction

Nondestructive testing (NDT) aims at discovering defects in materials such as metal or concrete, without damaging them [McM82]. It is usually performed directly after production of said material, in order to assure the demanding quality measures. There are several methods of nondestructive testing, such as visual inspection [ANMM93], radiography [Hal12], and electrical and magnetic testing such as Eddy current, see, e.g., [Bli12]. Here, we will focus on nondestructive testing using ultrasound [LMK12]. Ultrasonic nondestructive testing is a widely applied method for identifying defects in metals such as steel or aluminum [KK90]. An important application for instance is the inspection of weld seams [JC10], which are edges between two pieces of metal joined together via a welding process. Especially in the case of steel pipes, where a metal plate is bent into a cylinder and connected through welding, even small defects can lead to a reduced lifespan. It is therefore necessary to reliably detect common defects such as cracks, pores, and slag inclusions. To this end, the specimen gets insonified using an ultrasonic pulse emitted by a transducer, and the scattered ultrasonic signal then is recorded at another ultrasonic transducer [LMK12]. Performing several measurements placing the transducers on different locations then allows to identify defects in the material, using for instance the Time-Of-Flight Diffraction Method (TOFD), or the Synthetic Aperture Focusing Technique (SAFT), see, e.g., [SRD$^+$12].

In recent years, phased array probes, where several ultrasonic transducers are built into one physical component, became very popular in ultrasonic nondestructive testing. Phased array probes, in contrast to single element probes, can steer the ultrasonic pulse and hence focus it to different regions of the specimen, see, e.g., [Tho96]. Another approach of nondestructive testing using phased array probes is the Total Focusing Method (TFM) [HDW04]. Without steering the pulse, the specimen here is sequentially insonified by each of the individual ultrasonic transducers while the scattered ultrasonic signal is recorded at every transducer. This data acquisition method is also known as full-matrix-capture (FMC). The ultrasonic data is then algorithmically processed using the Synthetic Aperture Focusing Technique (cf.Section 8). TFM allows to reliably detect many types of defects, see, e.g., [JC10], and is therefore often described as the "gold standard" of ultrasonic nondestructive testing [Tho96]. One shortcoming of TFM compared to other methods is the relatively high amount of time needed for data acquisition. In many industrial settings, however, nondestructive testing usually requires extensive preparations during which other operations possibly need to shut down, see, e.g., [Caw01]. Therefore, there is a continuing effort to accelerate the data acquisition process. In order to achieve this goal in the context of TFM, we will propose to superpose ultrasonic measurements. To this end, each transducer of the phased array probe will insonify the specimen at a time chosen in an interval, which is significantly shorter than the time required for acquiring a full-matrix-capture. In this way, we can acquire ultrasonic data similar to a full-matrix-capture, but we also have to deal with overlapping measurements. By choosing the individual insonification times independent and uniformly distributed, and using an iterative version of SAFT, we will be able to diminish the effect of overlapping measurements. We will show that, in this way, under certain requirements on the sparsity of the defect and with high probability, one can efficiently use more ultrasonic data for defect identification as with a partial full-matrix-capture, acquired in the same amount of time. Note that the method of acquiring only a partial full-matrix-capture in order to reduce the measurement time, also has been considered by Schmitte et al. in [SNCO16].

Organization of the Chapter

In Section 7, starting from an analogous problem for point-like defects, we will first develop a simplified model of the signals acquired in ultrasonic nondestructive testing. This will give us the foundation for Section 8, where we discuss the *Synthetic Aperture Focusing Technique* (SAFT). For both basic and superposed measurements, we will analyze the defect images computed via SAFT in terms of the defect location in the case of point scatterers. In Section 9, we will develop an iterative version of the SAFT algorithm, which gives a significant improvement over the traditional SAFT algorithm for sparse defects and superposed measurements. Numerical results will be presented in Section 10.

Notation

Throughout this chapter, $\mathbb{R}_+$ will denote the positive real axis including 0 and for an integer n, $[n]$ will denote the set of integers from 1 to n. For any $D \subset \mathbb{R}^3$, we denote by $B(D, \mathbb{R}_+)$ the set of bounded functions from D to $\mathbb{R}_+$. Furthermore, let L_1 be the set of all functions f from $\mathbb{R}$ to $\mathbb{C}$ with

$$\|f\|_1 := \int_{\mathbb{R}} |f(t)| \, \mathrm{d}\, t < \infty.$$

We say that $f \in L_1$ has *bounded support* if there exists $T \geq 0$ such that $|f(t)| = 0$ for $t \geq |T|$. For any time-domain signal $u \in L_1$, $\hat{u}$ will denote the Fourier transform of u given by

$$\mathcal{F}\{u\}(\omega) = \hat{u}(\omega) = \int\limits_{\mathbb{R}} u(t) \exp(-i\omega t) \,\mathrm{d}\, t,$$

and we denote the corresponding inverse Fourier transform by $\mathcal{F}^{-1}\{\cdot\}$. If $\hat{u} \in L_1$, we have $\mathcal{F}^{-1}\{\hat{u}\} = u$, see, e.g., [Pin09]. For two functions $f, g \in L_1$, the convolution $f * g$ is given by

$$(f * g)(t) = \int\limits_{\mathbb{R}} f(\tau) g(t - \tau) \,\mathrm{d}\, \tau,$$

and it holds that $(f * g) \in L_1$, see, e.g., [Pin09]. For a closed subset $D \subset \mathbb{R}^3$, its boundary will be denoted by $\partial D \subset \mathbb{R}^3$. If $D_1, \ldots, D_s \subset \mathbb{R}^3$ are disjoint sets, we denote their union by

$$\dot{\bigcup_{j \in [s]}} D_j.$$

For any $r > 0$, $B_r(y)$ is the closed euclidean ball of radius r centered at y, i.e.,

$$B_r(y) = \left\{ x \in \mathbb{R}^3 \middle| \|x - y\|_2 \leq r \right\}.$$

For a subset $Y \subset \mathbb{R}^3$, we will also use the notation

$$B_r(Y) = \bigcup_{y \in Y} B_r(Y).$$

7. Model

In this section, we derive a model for scattered ultrasonic data arising after insonifying a specimen with an ultrasonic pulse. The key goal of this model is to capture the dependence of the observations on the location of the defect. In ultrasonic nondestructive testing, the specimen usually is insonified with a new pulse not before the scattered ultrasonic wave was recorded at each of the receiving transducer elements. We will refer to this data acquisition process as *basic measurements*, and derive a model for the corresponding ultrasonic data. Later, we will also discuss *superposed measurements*, where the specimen gets insonified by a transducer even before all data of the previous insonification was collected at the corresponding transducer. Here, the measured ultrasonic data is a superposition of basic measurements. The advantage of acquiring superposed measurements instead of basic measurements is a reduced measurement time. Basic ultrasonic measurements are characterized by the following definition.

Definition 7.1. Let $q_1, \ldots, q_m \in \mathbb{R}^3$ be the positions of the ultrasonic transducers and $p \in L_1$ be a fixed pulse. For each $i, j \in [m]$, a *basic measurement* $u_{i,j} \in L_1$ is the ultrasonic signal recorded at the transducer located at position q_j, after the specimen was insonified by the pulse p emitted by a transducer located at q_i.
Let $\mathcal{I} \subset [m]^2$ be the set containing all pairs (i, j) where a basic measurement $u_{i,j} \in L_1$ was measured. If

$$\mathcal{I} = [m]^2 =: \mathcal{I}_{\mathrm{FMC}},$$

we say that a *full-matrix-capture* was acquired.

To express basic measurements in terms of the scattering defect, we will consider two models. In a simpler first model, we will assume that the defect consists of a finite number of point scatterers. Subsequently, we expand this model to extended scatterers. For both cases, we will assume that the medium is homogeneous and isotropic. It therefore holds that the speed of sound c is constant. We will also make the simplifying assumption that the scattering properties of the defect $D \subset \mathbb{R}^3$ does not depend on the wavelength ω of the incident ultrasonic wave.

Basic Measurements of Point Scatterers

We will now study the scattering problem for point scatterers. For simplicity, we will neglect multiple scattering. This assumption will allow us to derive a linear model for the basic measurements $u_{i,j}, i, j \in [m]$. To this end, suppose that the specimen gets insonified by a time-harmonic spherical wave of frequency $\omega \in \mathbb{R}$, emitted by a transducer located at $q \in \mathbb{R}^3$. The corresponding ultrasonic wave is then given by the three dimensional free-space Green's function

$$\hat{G}(\omega, q, x) = \frac{1}{(4\pi)\|x - q\|_2} \exp\left(\mathrm{i}\tfrac{\omega}{c}\|x - q\|_2\right),$$

where c denotes the speed of sound, see for instance [Eva10]. Suppose now that located at $y_1, \ldots, y_s \in \mathbb{R}^3$, there are s point scatterers with scattering magnitudes $a(y_k) \in \mathbb{R}_+$, similar to the models used in [Bos13, FS12, AS13]. The spherical wave of frequency ω emitted at location q hits each of the s scatterers, which then acts as secondary source and also emits a time harmonic spherical wave of the same frequency ω. Since we neglect multiple scattering, the resulting scattered wave $u^s(\omega, x)$ at position $x \in \mathbb{R}^3$ is the superposition of the echos of the s point scatterers to the spherical wave. For any $x \in \mathbb{R}^3 \setminus \{y_1, \ldots, y_s\}$, the frequency domain signal $\hat{u}^s(\omega, x)$ is now given by

$$\begin{aligned}\hat{u}^s(\omega, x) &= \sum_{k=1}^{s} a(y_k)\hat{G}(\omega, q, y_k)\hat{G}(\omega, y_k, x) \\ &= \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi)^2 \|q - y_k\|_2 \|y_k - x\|_2} \exp\left(\mathrm{i}\tfrac{\omega}{c}(\|q - y_k\|_2 + \|y_k - x\|_2)\right).\end{aligned} \tag{7.1}$$

Now let the specimen be insonifed by a superposition of time-harmonic monochromatic spherical waves, emitted from a transducer at the location $q_i \in \mathbb{R}^3$. Suppose that, for a function $\hat{p} \in L_1$, each frequency $\omega \in \mathbb{R}$ gets emitted with phase and magnitude $\hat{p}(\omega)$. By linearity and (7.1), the basic measurement $\hat{u}_{i,j}$ recorded at the transducer at q_j, after the pulse p was emitted from q_i, is given by

$$\hat{u}_{i,j}(\omega) = \hat{p}(\omega)\hat{u}^s(\omega, q_j) = \hat{p}(\omega) \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} \exp\left(\mathrm{i}\omega(t_{i,j}(y_k))\right), \tag{7.2}$$

where the functions t_i and $t_{i,j}$ are, for $i, j \in [m]$ and $x \in \mathbb{R}^3$, defined as

$$t_i(x) := \tfrac{\|q_i - x\|_2}{c}, \qquad t_{i,j}(x) = t_i(x) + t_j(x). \tag{7.3}$$

Usually, $t_{i,j}(x) \in \mathbb{R}_+$ is referred to as *time-of-flight* of x with respect to i, j, see, e.g., [SRD⁺12]. With the inverse Fourier transform, we obtain

$$\begin{aligned}u_{i,j}(t) &= \mathcal{F}^{-1}\left\{\hat{p}(\omega) \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} \exp\left(\mathrm{i}\omega t_{i,j}(y_k)\right)\right\}(t) \\ &= \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} \mathcal{F}^{-1}\left\{\hat{p}(\omega) \exp\left(\mathrm{i}\omega t_{i,j}(y_k)\right)\right\}(t) \\ &= \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} \mathcal{F}^{-1}\{\hat{p}(\omega)\} * \mathcal{F}^{-1}\left\{\exp\left(\mathrm{i}\omega t_{i,j}(y_k)\right)\right\}(t) \\ &= \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} \left(p * \delta\left(\cdot + t_{i,j}(y_k)\right)\right)(t) \\ &= \sum_{k=1}^{s} \frac{a(y_k)}{(4\pi c)^2 t_i(y_k) t_j(y_k)} p\left(t - t_{i,j}(y_k)\right),\end{aligned} \tag{7.4}$$

where δ is the Dirac delta distribution. Here, $u_{i,j}$ is also absolute integrable for all $i, j \in [m]$, as $p \in L_1$. Similarly, if $p \in L_1$ is compactly supported, then $u_{i,j}$ must be compactly supported for arbitrary $i, j \in [m]$ as $t_{i,j} < \infty$. With the considerations above, we formulate the following model assumption.

Model Assumption 7.2. *Let $y_1, \dots, y_s \in \mathbb{R}^3$ be the locations of point scatterers. Let further $q_i \in \mathbb{R}^3 \setminus \{y_1, \dots, y_s\}$, $i \in [m]$ be arbitrary transducer locations. Let $p \in L_1$ be the pulse used to insonify the specimen. Then, the basic measurement $u_{i,j} \in L_1$ is for arbitrary $i, j \in [m]$ given by*

$$u_{i,j}(t) = \sum_{k=1}^{s} a_{i,j}(y_k) p(t - (t_{i,j}(y_k))),$$

where $a_{i,j}(y_k) \in \mathbb{R}_+$ for $k \in [s]$.

In Model Assumption 7.2, we neglect the explicit dependence of the coefficients $a_{i,j}(y_k)$ on the scattering magnitudes and the time-of-flight $t_i(x_k)$ and $t_j(x_k)$; the scattering problem for point scatters only serves as a toy model for the corresponding problem involving extended scatterers. Here, it is considerably more involved to compute the analogous density $a_{i,j} \in B(\partial D, \mathbb{R}_+)$, $i, j \in [m]$, where $D \subset \mathbb{R}^3$ is an extended defect $D \subset \mathbb{R}^3$. Extended scatterers are subject of the following section.

Basic Measurements of Extended Scatterers

Before we start with our considerations, we will first give the following definition.

Definition 7.3 (Extended Scatterers). A subset $D \subset \mathbb{R}^3$ is an *extended scatterer*, if it is closed, bounded and its complement D^c is connected. Furthermore, we say that D is (s, r)*-sparse*, if there exists a set of points $Y \subset \mathbb{R}^3$ with $|Y| \leq s$ such that

$$D \subset B_r(Y).$$

Note that by the Heine-Borel theorem, every extended scatterer $D \subset \mathbb{R}^3$ is also compact. For every $r > 0$, extended scatterers are therefore always (s, r)-sparse for suitably chosen $s \geq 0$. Now, let $D \subset \mathbb{R}^3$ be an extended scatterer, which can be written as

$$D = \dot{\bigcup_{k \in [s]}} D_k; \tag{7.5}$$

D_k, $k \in [s]$ are connected, but $D_{k_1} \cup D_{k_1}$ are not connected for $k_1 \neq k_2 \in [s]$. For simplicity, we will also assume that no ultrasonic wave can penetrate the defect D, and only the boundary ∂D of the defect has an impact on the measured ultrasonic signal. Analogous to (7.5), we can write

$$\partial D = \dot{\bigcup_{k \in [s]}} \partial D_k.$$

As before, let $q_1, \dots, q_m \in \mathbb{R}^3 \setminus D$, be the positions of the m transducers. While, for $k \in [s]$, we neglect multiple scattering by any point $x \in \partial D \setminus \partial D_k$, after the ultrasonic wave was scattered by any point $y \in \partial D_k$, we cannot neglect multiple scattering caused by points $y \in \partial D_k$ within the same defect ∂D_k. Here, the geometry of the defect boundary ∂D_k has immense impact on the magnitude of the scattered wave, see, e.g., [LMK12, KK90, Bos13]. In order to capture these dependencies without making restrictive model assumptions, we will assume that the scattering magnitude at a given point $y \in \partial D$ is not only a function of the location of the scatterer, but also depends on the locations $q_i, q_j \in \mathbb{R}^3$ of the corresponding transducers. To be more precise, we assume that for arbitrary $i, j \in [m]$, the scattering magnitudes are given by a bounded function $a_{i,j} \in B(\partial D, \mathbb{R}_+)$. Proceeding similarly as in the case of point scatterers in (7.4), it now follows for pulse functions $p \in L_1$ with FOURIER transform $\hat{p} \in L_1$ and $t_i(y), t_{i,j}(y)$ as in (7.3) that

$$\begin{aligned}
u_{i,j}(t) &= \mathcal{F}^{-1}\left\{\hat{p}(\omega)\int\limits_{\partial D} \frac{a_{i,j}(y)}{(4\pi c)^2 t_i(y)t_j(y)} \exp\left(\mathrm{i}\omega t_{i,j}(y)\right) \mathrm{d}\,y\right\} \\
&= \int\limits_{\partial D} \frac{a_{i,j}(y)}{(4\pi c)^2 t_i(y)t_j(y)} \mathcal{F}^{-1}\left\{p(\omega)\exp\left(\mathrm{i}\omega t_{i,j}(y)\right)\right\} \mathrm{d}\,y \\
&= \int\limits_{\partial D} \frac{a_{i,j}(y)}{(4\pi c)^2 t_i(y)t_j(y)} \left(p * \delta(\cdot + t_{i,j}(y))\right)(t) \,\mathrm{d}\,y \qquad (7.6) \\
&= \int\limits_{\partial D} \frac{a_{i,j}(y)}{(4\pi c)^2 t_i(y)t_j(y)} \left(p(\cdot + t_{i,j}(y))\right)(t) \,\mathrm{d}\,y \\
&= \int\limits_{\partial D} \tilde{a}_{i,j}(y) p(t - t_{i,j}(y)) \,\mathrm{d}\,y,
\end{aligned}$$

where we set

$$\tilde{a}_{i,j}(y) = \frac{a_{i,j}(y)}{(4\pi c)^2 t_i(y)t_j(y)} \in B(\partial D, \mathbb{R}_+).$$

Next, we will show that $u_{i,j} \in L_1$. Since $\tilde{a}_{i,j} \in B(\partial D, \mathbb{R}_+)$, there exists $M < \infty$ such that for arbitrary $i, j \in [m]$ and $y \in \partial D$, we have $|\tilde{a}_{i,j}(y)| \leq M$. Therefore

$$\begin{aligned}
\|u_{i,j}\|_1 &= \int\limits_{\mathbb{R}} \left|u_{i,j}(t)\right| \mathrm{d}\,t = \int\limits_{\mathbb{R}} \left|\int\limits_{\partial D} \tilde{a}_{i,j}(y) p(t - t_{i,j}(y)) \,\mathrm{d}\,y\right| \mathrm{d}\,t \\
&\leq M \int\limits_{\mathbb{R}} \int\limits_{\partial D} |p(t - t_{i,j}(y))| \,\mathrm{d}\,y \,\mathrm{d}\,t \\
&= M \int\limits_{\partial D} \int\limits_{\mathbb{R}} |p(t - t_{i,j}(y))| \,\mathrm{d}\,t \,\mathrm{d}\,y
\end{aligned}$$

$$\leq M \int_{\partial D} \|p\|_1 \, \mathrm{d}y < \infty,$$

since $p \in L_1$ and D is bounded. Also, if $p \in L_1$ is compactly supported, $u_{i,j}$ is also compactly supported for arbitrary $i, j \in [m]$. We can summarize the above considerations in the following model assumption.

Model Assumption 7.4. *Let $D \subset \mathbb{R}^3$ be an extended scatterer with boundary ∂D. Let further $q_i \in \mathbb{R}^3 \setminus D$, $i \in [m]$ be arbitrary transducer locations and $p \in L_1$ be the pulse used to insonify the specimen. Then, the basic measurement $u_{i,j}$ is, for arbitrary $i, j \in [m]$, given by*

$$u_{i,j}(t) = \int_{\partial D} a_{i,j}(y) p(t - t_{i,j}(y)), \tag{7.7}$$

where $a_{i,j} \in B(\partial D, \mathbb{R}_+)$.

Superposed Measurements

With the emerging availability of phased array probes, full-matrix-capture data aquisition (Definition 7.1) became very popular, see, e.g., [JC10]. Using phased array probes, for each $i \in [m]$, the m transducers allow to acquire all basic ultrasonic signals $u_{i,j} \in L_1$, $j \in [m]$ at the same time. Suppose that $p \in L_1$ is compactly supported. Then, by Model Assumption 7.2 and Model Assumption 7.4, the maximal time needed to collect the scattered ultrasonic signals at each of the m receiver is bounded and we will denote it by T. By passing through $i \in [m]$, a full-matrix-capture thus needs a total time of at most mT. To further reduce measurements time, we will superpose basic measurements. By insonifying the specimen with time-shifted versions of the same pulse emitted from different transducers, we are able to acquire ultrasonic data comparable to a full-matrix-capture in a considerably shorter period of time.

Definition 7.5. Let $q_1, \ldots, q_m \in \mathbb{R}^3$ denote the positions of ultrasonic transducers, $T_1, \ldots T_m \in \mathbb{R}$ be arbitrary shot times, and fix a pulse $p \in L_1$. Let u_j be the ultrasonic signal recorded by a transducer located at q_j after the specimen gets simultaneously insonified by each transducer located at q_i, $i \in [m]$ with respective pulses $p(\cdot - T_i) \in L_1$. Then, for $(i, j) \in \mathcal{I}_{\mathrm{FMC}} = [m]^2$, the function

$$\tilde{u}_{i,j}(t) = u_j(t + T_i)$$

is called *superposed measurement.*

Bearing in mind that superposed measurements are sums of basic measurements using time-shifted pulses, and that our derived models are linear, it is straightforward to apply Model Assumptions 7.2, 7.4 to the case of superposed measurements.

Lemma 7.6. *Let $D \subset \mathbb{R}^3$ be either a set of point scatterers or an extended scatterer, $q_i \in \mathbb{R}^3 \setminus D, i \in [m]$ be arbitrary transducer locations and $p \in L_1$ be a pulse. Let further the specimen be simultaneously insonified by each transducer located at q_i, $i \in [m]$ with respective pulses $p(\cdot - T_i) \in L_1$ and shot times $T_i \in \mathbb{R}$, $i \in [m]$. Then, for $(i,j) \in \mathcal{I}_{FMC} = [m]^2$, the superposed ultrasonic measurement $\tilde{u}_{i,j} \in L_1$ is given by*

$$\tilde{u}_{i,j}(t) = u_{i,j}(t) + \sum_{\substack{i' \in [m] \\ i' \neq i}} u_{i',j}(t + T_i - T_{i'}),$$

where for each $i' \in [m]$, $u_{i',j} \in L_1$ is the basic measurement as given in Model Assumption 7.2 or 7.4.

Proof. We will only prove the lemma in the point scatterer case, since the extended scatterer case is similar. To this end, let $y_1, \dots, y_s$ be the locations of the point scatterers. Then, by Model Assumption 7.2 for fixed $i \in [m]$, the basic measurement $u_{i,j}$ is for arbitrary $i, j \in [m]$ given by

$$u_{i,j}(t) = \sum_{k=1}^{s} a_{i,j}(y_k) p(t - t_{i,j}(y_k) - T_i),$$

where $a_{i,j}(y_k) \in \mathbb{R}_+$ for $k \in [s]$. By linearity, it holds for u_j as in Definition 7.5 that

$$u_j(t) = \sum_{i' \in [m]} \sum_{k=1}^{s} a_{i,j}(y_k) p(t - t_{i,j}(y_k) - T_{i'}).$$

Since this implies

$$\tilde{u}_{i,j}(t) = u_j(t + T_i) = \sum_{i' \in [m]} \sum_{k=1}^{s} a_{i,j}(y_k) p(t - t_{i,j}(y_k) + T_i - T_{i'}),$$

the proof is now complete. □

Let T and m as in the considerations which led to Definition 7.5. With $T_i = (i-1)T$, $i \in [m]$, basic measurements can be embedded into the framework of superposed measurements. Indeed, by Lemma 7.6, we have for arbitrary $(i,j) \in \mathcal{I}_{\mathrm{FMC}} = [m]^2$,

$$\tilde{u}_{i,j}(t) = u_{i,j}(t) + \sum_{\substack{i' \in [m] \\ i' \neq i}} u_{i',j}(t + (i - i')T). \tag{7.8}$$

By definition of T, the sum over $i' \neq i$ in (7.8) vanishes for $t \in [0,T)$; and it holds that

$$\tilde{u}_{i,j}(t) = u_{i,j}(t).$$

Since the goal of superposed measurements is to diminish the time needed for full-matrix-capture data acquisition, we will choose the shot times T_i, $i \in [m]$ in an interval $[0,S]$ with $S < (|m| - 1)T$. Doing so, the data aquisition only takes an amount of time strictly less than mT. In this case, however, the sum over $i' \neq i$ in (7.8) does not necessarily vanish; by the pigeonhole principle, there always exist indices $i, i' \in [m]$ such that $|T_i - T_{i'}| < T$. Exploiting the structure of the basic measurements $u_{i,j}$ and the dependence on the defect, we will nevertheless be able to reduce the amount of noise caused by overlapping measurements. This will be achieved by a modification of the Synthetic Aperture Focusing Technique, a standard defect imaging method in ultrasonic nondestructive testing. We will also discuss different methods of choosing the shot times $T_i \in [0,S]$, $i \in [m]$ in Section 8. There, we will see that choosing T_i, $i \in [m]$ independent and uniformly distributed in $[0,S]$, in general leads to good defect reconstructions.

8. Synthetic Aperture Focusing Technique

We will now introduce the Synthetic Aperture Focusing Technique (SAFT), which is a widely used defect imaging algorithm in ultrasonic nondestructive testing. In related fields such as radar and sonar, similar methods are known as Synthetic Aperture Radar (SAR), and Synthetic Aperture Sonar (SAS) [Hov80, Han11].

SAFT for Basic Measurements

SAFT uses the intuitive but heuristic approach of backprojecting the measured ultrasonic signals to all possible sources according to the time-of-flight, see, e.g., [Sey82]. For basic measurements as introduced in Section 7, the Synthetic Aperture Focusing Technique is given as follows.

Definition 8.1. Let $q_1, \ldots, q_m \in \mathbb{R}^3$ be arbitrary transducer locations, and for $\mathcal{I} \subset [m]^2$, let $u_{i,j} \in L_1$, $(i,j) \in \mathcal{I}$ be basic measurements. The *SAFT backprojection for basic measurements* $R_{\mathcal{I}}(x)$ for arbitrary $x \in \mathbb{R}^3$, is then given by

$$R_{\mathcal{I}}(x) = \sum_{(i,j)\in\mathcal{I}} u_{i,j}(t_{i,j}(x)). \tag{8.1}$$

SAFT, as it is formulated here, is a basic version of a whole class of algorithms. By using additional apodization weights in (8.1), it can for instance be adapted to characteristics of the ultrasonic probe, see, e.g., [SKF⁺12, HDW08]. Using full-matrix-capture measurements $\mathcal{I} = \mathcal{I}_{\text{FMC}} = [m]^2$ in (8.1), acquired by a phased array probe, is usually called the *Total Focusing Method* (TFM) [HDW04, JC10]. We have to point out that the SAFT backprojection is not a mathematically rigorous solution to the inverse problem for the assumed forward model. It nevertheless is a widely used algorithm for imaging defects in materials [SRD⁺12, Caw01]. Unlike more rigorous solutions of an corresponding inverse problem, like for instance the the wavenumber algorithm [HDW08], an important advantage of SAFT is its flexibility in terms of the arrangement of the transducers. Furthermore, the SAFT backprojection only relies on the locations of the transducers, the time-of-flight $t_{i,j}(x)$ of a point $x \in \mathbb{R}$ and the measured data. Therefore, it can be adopted to more realistic scenarios where the the speed of sound is not constant. This is usually the case, as the transducers often are contained in a coupling fluid such as water, see, e.g., [Caw01]. Additionally, the specimen itself can be anisotropic with varying speed of sound in different directions [SRD⁺12]. In both scenarios, it is possible to achieve good defect reconstructions by adjusting the time-of-flight $t_{i,j}(x)$ in (8.1) accordingly. Adjusted time-of-flights can for instance be computed via a fast marching method (FMM) based on Fermat's principle, see, e.g., [Set99]. In order to digitally process the ultrasonic signals $u_{i,j}$, they are sampled at a high sampling rate and discretized using an analog-to-digital converter. Therefore, only equidistant discrete samples of the ultrasonic signals are available for the SAFT backprojection. To account for this, one rounds the time-of-flight $t_{i,j}(x)$ appearing in (8.1) to the closest time $t \in \mathbb{R}$ where the sampled ultrasonic signal $u_{i,j}(t)$ is available [LMK12]. For imaging reasons, a discrete Hilbert transform often is applied to the discretized ultrasonic signal; the SAFT backprojection is then computed using the signals $\bar{u}_{i,j} + \mathrm{i}\mathcal{H}\{\bar{u}_{i,j}\}$, $(i,j) \in \mathcal{I}$, where $\bar{u}_{i,j}$ is the discretized version of the signal $u_{i,j}$ and $\mathcal{H}\{\bar{u}_{i,j}\}$ denotes the the *discrete Hilbert transform* of $\bar{u}_{i,j}$, see, e.g., [LMK12].

With the derived model for point scatterers, we now aim to illustrate the heuristics behind the SAFT backprojection. A more detailed description can be found,e.g., in [LMK12]. For this purpose, we will use a *raised cosine* $p_{N,\omega_0} \in L_1$ as a model for the pulse p. The raised cosine, especially in the case $N = 2$, is a widely used pulse model in ultrasonic nondestructive testing, see, e.g., [LMK12, Spi01].

Definition 8.2 (Raised cosine [LMK12]). For arbitrary frequency $\omega_0 \in \mathbb{R}_+$ and positive integer N, the *raised cosine* $p_{N,\omega_0} \in L_1$ is defined by

$$p_{N,\omega_0}(t) = \begin{cases} \frac{1}{2}\left(1 + \cos\frac{\omega_0}{N}t\right)\cos\omega_0 t & \text{if } -\frac{N\pi}{\omega_0} \le t \le \frac{N\pi}{\omega_0}, \\ 0 & \text{else.} \end{cases} \tag{8.2}$$

Let us now point out some properties of the raised cosine. Because of the window function $\left(1 + \cos\frac{\omega_0}{N}t\right)$, it holds that the absolute value $|p_{N,\omega_0}(t)|$ attains its maximum for $t = 0$. The window function is monotonically decreasing with $|t|$ until $|t| \ge {}^{(N\pi)}/_{\omega_0}$, where it holds that $p_{N,\omega_0}(t) = 0$. Obviously, p_{N,ω_0} is compactly supported. Furthermore, due to the factor $\cos\omega_0 t$, it is oscillating with frequency ω_0.

Suppose now that, located at $y \in \mathbb{R}^3$, there is a single point scatterer. Furthermore, let $q_1, \ldots, q_m \in \mathbb{R}^3$ denote the locations of the ultrasonic transducers. For a positive integer N and a frequency $\omega_0 \in \mathbb{R}_+$, let the specimen be insonified by a raised cosine $p_{N,\omega_0} \in L_1$, and let $\mathcal{I} \subset [m]^2$ be the set of acquired basic measurements. Then, by Model Assumption 7.2, the basic measurements are given by

$$u_{i,j}(t) = a_{i,j}(y)p_{N,\omega_0}(t - t_{i,j}(y)) \tag{8.3}$$

for all $(i,j) \in \mathcal{I}$ and $t \in \mathbb{R}$. For the SAFT backprojection, as defined in Definition 8.1, we have

$$|R_{\mathcal{I}}(y)| = \left|\sum_{(i,j)\in\mathcal{I}} u_{i,j}(t_{i,j}(y))\right| = \left|\sum_{(i,j)\in\mathcal{I}} a_{i,j}(y)p_{N,\omega_0}(0)\right| = \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y), \tag{8.4}$$

where y is again the location of the scatterer. On the other hand, for arbitrary $x \in \mathbb{R}^3$, we have

$$\begin{aligned} |R_{\mathcal{I}}(x)| &= \left|\sum_{(i,j)\in\mathcal{I}} u_{i,j}(t_{i,j}(x))\right| = \left|\sum_{(i,j)\in\mathcal{I}} a_{i,j}(y)p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y))\right| \\ &\le \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y)\left|p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y))\right| \\ &\le \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y)|p_{N,\omega_0}(0)| = |R_{\mathcal{I}}(y)|, \end{aligned} \tag{8.5}$$

since $|p_{N,\omega_0}(t)|$ attains is maximum at $t = 0$. It follows for a single point scatterer, that the absolute value of the SAFT backprojection attains its maximum exactly at the location of the scatterer. Depending on $x \in \mathbb{R}^3, x \neq y$, $|R_{\mathcal{I}}(x)|$ may be much smaller than $|R_{\mathcal{I}}(y)|$; this is due to two reasons. First, since $p_{N,\omega_0}(t) = 0$ for $|t| \ge {}^{N\pi}/_{\omega_0}$, some of the terms in (8.5) will be zero. In addition, the oscillating nature of $p_{N,\omega_0}(t)$ for $|t| < {}^{N\pi}/_{\omega_0}$ leads to destructive interference. For points $x \in \mathbb{R}^3$ in a close neighborhood of the scatterer y, however, the modulus of the SAFT backprojection at x will be comparable to the SAFT backprojection at y. The following two lemmas will allow us to capture this phenomenon. For a more detailed resolution analysis, we refer to [Tho84].

Lemma 8.3. *Let $\omega_0 \in \mathbb{R}_+$ be a frequency and N a positive integer. Then*

$$3/4 \leq p_{N,\omega_0}(t) \leq 1,$$

provided

$$|t| \leq 1/(\sqrt{3}\omega_0). \tag{8.6}$$

Proof. The upper bound is obvious. For the lower bound, expanding $\cos(x)$ in a power series, see, e.g., [BHL+12], it follows that $\cos t \geq 1 - \frac{1}{2}t^2$. For t as in (8.6), we therefore get

$$\begin{aligned} p_{N,\omega_0}(t) &= \tfrac{1}{2}\left(1 + \cos \tfrac{\omega_0}{N} t\right) \cos \omega_0 t \\ &\geq \tfrac{1}{2}(2 - \tfrac{1}{2}(\tfrac{\omega_0 t}{N})^2)(1 - \tfrac{1}{2}(\omega_0 t)^2) \\ &\geq \tfrac{1}{2}(2 - \tfrac{1}{2}(\omega_0 t)^2)(1 - \tfrac{1}{2}(\omega_0 t)^2) \\ &\geq 1 - \tfrac{3}{4}(\omega_0 t)^2 \geq \tfrac{3}{4}. \end{aligned} \tag{8.7}$$

This completes the proof. □

Lemma 8.4. *For arbitrary transducer locations $q_1, \ldots, q_m \in \mathbb{R}^3$ and arbitrary $x, y, \in \mathbb{R}^3$, it holds that*

$$|t_{i,j}(x) - t_{i,j}(y)| \leq \tfrac{2}{c}\|x - y\|_2,$$

where c is the speed of sound.

Proof. By triangle and reverse triangle inequality, we have

$$\begin{aligned} |t_{i,j}(x) - t_{i,j}(y)| &= \tfrac{1}{c}\big|\|q_i - x\|_2 + \|q_j - x\|_2 - \|q_i - y\|_2 - \|q_j - y\|_2\big| \\ &= \tfrac{1}{c}\big|\|(q_i - y) + (y - x)\|_2 + \|(q_j - y) + (y - x)\|_2 - \|q_i - y\|_2 - \|q_j - y\|_2\big| \\ &\leq \tfrac{2}{c}\|y - x\|_2. \end{aligned}$$

□

With the same setting which led to (8.3) and $x \in \mathbb{R}^3$ with $\|x-y\|_2 \leq c/(2\sqrt{3}\omega_0)$, Lemma 8.4 now implies for arbitrary $(i,j) \in \mathcal{I}$ that

$$|t_{i,j}(x) - t_{i,j}(y)| \leq \tfrac{2}{c}\|x - y\|_2 \leq 1/(\sqrt{3}\omega_0).$$

With Lemma 8.3, it therefore follows that

$$\begin{aligned} R_{\mathcal{I}}(x) &= \sum_{(i,j)\in\mathcal{I}} u_{i,j}(t_{i,j}(x)) \\ &= \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y) p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y)) \\ &\geq \tfrac{3}{4} \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y) \\ &= \tfrac{3}{4} R_{\mathcal{I}}(y). \end{aligned} \tag{8.8}$$

Hence, we can only expect to resolve a scatterer using SAFT up to length scales of the order c/ω_0.

We will now consider the case of multiple point scatterers instead of just one. If $y_1, \ldots, y_s \in \mathbb{R}^3$ are the locations of s point scatterers, Model Assumption 7.2 implies for arbitrary $(i,j) \in \mathcal{I}$ and arbitrary $x \in \mathbb{R}^3$ that

$$|R_{\mathcal{I}}(x)| = \left| \sum_{(i,j)\in\mathcal{I}} \sum_{k\in[s]} a_{i,j}(y_k) p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y)) \right|. \tag{8.9}$$

Now let $\ell \in [s]$ be arbitrary. At the location of the point scatterer y_ℓ, we get in (8.9)

$$|R_{\mathcal{I}}(y_l)| = \left| \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y_l) \right. \tag{8.10}$$

$$\left. + \sum_{(i,j)\in\mathcal{I}} \sum_{\substack{k\in[s] \\ k\neq l}} a_{i,j}(y_k) p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y)) \right|. \tag{8.11}$$

In contrast to the single scatterer case, it is not obvious that y_l, $l \in [s]$ are local maxima of $|R_{\mathcal{I}}(x)|$. This is caused by the additional sum involving the remaining point scatterer in (8.11). Countless results in the ultrasonic nondestructive testing literature, for appropriately chosen $q_1, \ldots, q_m \in \mathbb{R}^3$, show that one is nevertheless able to identify the scatterers, up to certain resolution limitations as described above, as local maxima of $|R_{\mathcal{I}}(x)|$. For this reason, we will use the performance of SAFT as a benchmark for the performance analysis of our modified approach.

SAFT for Superposed Measurements

The flexibility of the SAFT algorithm now allows us to easily adopt the SAFT backprojection to the case of superposed measurements.

Definition 8.5. Let $q_1, \ldots, q_m \in \mathbb{R}^3$ be arbitrary transducer locations and $\mathcal{I} \subset [m]^2$. Let further $T_i \in \mathbb{R}$, $i \in [m]$ be arbitrary shot times and $\tilde{u}_{i,j}$, $(i,j) \in \mathcal{I}$ be the corresponding superposed measurements. The *SAFT backprojection for superposed measurements* $\tilde{R}_{\mathcal{I}}(x)$ for arbitrary $x \in \mathbb{R}^3$, is then given by

$$\tilde{R}_{\mathcal{I}}(x) = \sum_{(i,j)\in\mathcal{I}} \tilde{u}_{i,j}(t_{i,j}(x)). \tag{8.12}$$

We can now directly related the SAFT backprojection using superposed measurements to the SAFT backprojection using the corresponding basic measurements.

Lemma 8.6. *Let $D \subset \mathbb{R}^3$ be either a set of point scatterers or an extended scatterer. Let further $q_i \in \mathbb{R}^3 \setminus D$, $i \in [m]$ be arbitrary transducer locations, $\mathcal{I} = \mathcal{I}_{FMC} = [m]^2$, and $p \in L_1$ be the pulse used to simultaneously insonify the specimen with corresponding shot times $T_i \in \mathbb{R}$, $i \in [m]$. Then,*

$$\tilde{R}_{\mathcal{I}}(x) = R_{\mathcal{I}}(x) + \sum_{(i,j)\in\mathcal{I}} \sum_{\substack{i'\in[m]\\ i'\neq i}} u_{i',j}(t_{i,j}(x) + T_i - T_{i'}), \tag{8.13}$$

where $u_{i',j} \in L_1$, $(i,j) \in \mathcal{I}$ are the basic measurement as given in Model Assumption 7.2 or 7.4; $R_{\mathcal{I}}(x)$ is the SAFT backprojection for basic measurements as in Definition 8.1.

Proof. The result of the lemma directly follows from Lemma 7.6, as

$$\begin{aligned}
\tilde{R}_{\mathcal{I}}(x) &= \sum_{(i,j)\in\mathcal{I}} \tilde{u}_{i,j}(t_{i,j}(x)) \\
&= \sum_{(i,j)\in\mathcal{I}} \left(u_{i,j}(t_{i,j}(x)) + \sum_{\substack{i'\in[m]\\ i'\neq i}} u_{i,j}(t_{i,j}(x) + T_i - T_{i'}) \right) \\
&= R_{\mathcal{I}}(x) + \sum_{(i,j)\in\mathcal{I}} \sum_{\substack{i'\in[m]\\ i'\neq i}} u_{i,j}(t_{i,j}(x) + T_i - T_{i'}). \qquad \square
\end{aligned}$$

We refer to the sum on the right hand side of (8.13) as *superposition noise*, which we will analyze in terms of the defect location and choice of shot times $T_i \in \mathbb{R}$, $i \in [m]$. The following lemma illustrates, why the superposition noise caused by superposing measurements using equidistant shot times can result in a highly ambiguous SAFT backprojection.

Lemma 8.7. *Let the transducers be arranged in an linear array, i.e., $q_i = ((i-1)a, 0, 0)^T$ for some $a > 0$ and $i \in [m]$. Let further $\mathcal{I} = \mathcal{I}_{FMC} = [m]^2$, $\omega_0 \in \mathbb{R}_+$ be an arbitrary frequency, N be a positive integer and, for $\varepsilon \le c/(4\sqrt{3}\omega_0)$,*

$$y_\varepsilon = (0, \varepsilon, 0)^T \qquad \text{and} \qquad \tilde{y}_\varepsilon = (-1/2(cT+a), \varepsilon, 0)^T \,. \tag{8.14}$$

If, for a single point scatterer located at y_ε and $T \ge N\pi/\omega_0 + (4\varepsilon+a)/c$, the specimen gets insonified by a raised cosine $p_{N,\omega_0} \in L_1$ and equidistant shot times

$$T_i = (i-1)T,$$

it follows that

$$\tilde{R}_{\mathcal{I}}(y_\varepsilon) = \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y_\varepsilon), \tag{8.15}$$

and

$$\tilde{R}_{\mathcal{I}}(\tilde{y}_\varepsilon) \ge 3/4 \sum_{\substack{(i,j)\in\mathcal{I}\\ i\neq 1}} a_{i+1,j}(y_\varepsilon), \tag{8.16}$$

where $a_{i,j}(y_\varepsilon) \in \mathbb{R}_+$, $i,j \in [m]$ are the scattering coefficients of y_ε as in Model Assumption 7.2.

Proof. By Lemma 8.6, we have for arbitrary $x \in \mathbb{R}^3$

$$\begin{aligned} \tilde{R}_{\mathcal{I}}(x) &= R_{\mathcal{I}}(x) + \sum_{(i,j)\in\mathcal{I}} \sum_{\substack{i'\in[m]\\ i'\neq i}} u_{i',j}(t_{i,j}(x) + T_i - T_{i'}) \\ &= R_{\mathcal{I}}(x) + \sum_{(i,j)\in\mathcal{I}} \sum_{\substack{i'\in[m]\\ i'\neq i}} a_{i',j}(y_\varepsilon) p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}), \end{aligned} \tag{8.17}$$

with $a_{i,j}(y_\varepsilon) \in \mathbb{R}_+$, $(i,j) \in \mathcal{I}$ as in Model Assumption 7.2. By Definition 8.1, it further holds that

$$R_{\mathcal{I}}(x) = \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y_\varepsilon) p_{N,\omega_0}(t_{i,j}(x) - t_{i,j}(y_\varepsilon)). \tag{8.18}$$

For $x = y_\varepsilon$ in (8.18), we get

$$R_{\mathcal{I}}(y_\varepsilon) = \sum_{(i,j)\in\mathcal{I}} a_{i,j}(y_\varepsilon) p_{N,\omega_0}(0) = \sum_{(i,j)\in i} a_{i,j}(y_\varepsilon).$$

To prove (8.15), it is therefore enough to show that the sum on the right hand side of (8.17) vanishes. To this end, observe that for arbitrary $i, i', j \in [m]$ with $i' \neq i$, we have

$$\begin{aligned} |t_{i,j}(y_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}| &\geq \Big| |t_{i,j}(y_\varepsilon) - t_{i,j}(y_\varepsilon) + (i - i')T| - |t_{i,j}(y_\varepsilon) - t_{i',j}(y_\varepsilon)| \Big| \\ &= \Big| |(i - i')|T - |t_{i,j}(y_\varepsilon) - t_{i',j}(y_\varepsilon)| \Big| . \end{aligned} \tag{8.19}$$

Analogously to (8.14), we define

$$y_0 = (0, 0, 0)^T \qquad \text{and} \qquad \tilde{y}_0 = (-1/2(cT + a), 0, 0)^T .$$

Since $\|y_\varepsilon - y_0\|_2 = \varepsilon$, Lemma 8.4 now implies that

$$\begin{aligned} |t_{i,j}(y_\varepsilon) - t_{i',j}(y_\varepsilon)| &\leq |t_{i,j}(y_0) - t_{i',j}(y_0)| + |t_{i,j}(y_\varepsilon) - t_{i,j}(y_0)| + |t_{i',j}(y_0) - t_{i',j}(y_\varepsilon)| \\ &\leq |\tfrac{a}{c}(i + j) - \tfrac{a}{c}(i' + j)| + 4\varepsilon/c \qquad (8.20) \\ &\leq \tfrac{a}{c}|i - i'| + 4\varepsilon/c. \end{aligned}$$

With (8.20) and the assumption of the lemma, we now can bound (8.19) as follows

$$\begin{aligned} |i - i'|T - |t_{i,j}(y_\varepsilon) - t_{i'.j}(y_\varepsilon)| &\geq |i - i'|T - \left(\tfrac{a}{c}|i - i'| + 4\varepsilon/c\right) \\ &= |i - i'|(T - \tfrac{a}{c}) - 4\varepsilon/c \\ &\geq (T - \tfrac{a}{c}) - 4\varepsilon/c \geq \tfrac{N\pi}{\omega_0}. \end{aligned}$$

Since $p_{N,\omega_0}(t) = 0$ for $|t| \geq \frac{N\pi}{\omega_0}$, this now implyies (8.15). For (8.16), observe that for arbitrary $i, j, i' \in [m]$, we have

$$\begin{aligned} &t_{i,j}(\tilde{y}_0) - t_{i',j}(y_0) + T_i - T_{i'} \\ &=\Big((T + a/c) + (i - 1)\tfrac{a}{c} + (j - 1)\tfrac{a}{c}\Big) - \Big((i' - 1)\tfrac{a}{c} + (j - 1)\tfrac{a}{c}\Big) + (i - i')T \qquad (8.21) \\ &=(T + a/c)(i - i' + 1). \end{aligned}$$

For $i, j \in [m]$, $i \neq 1$, $i' = i - 1$, it therefore follows with Lemma 8.4 and $\varepsilon \leq c/(4\sqrt{3}\omega_0)$ that

$$\begin{aligned} &|t_{i,j}(\tilde{y}_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}| \\ &\leq |t_{i,j}(\tilde{y}_0) - t_{i',j}(y_0) + T_i - T_{i'}| + |t_{i,j}(\tilde{y}_\varepsilon) - t_{i,j}(\tilde{y}_0)| + |t_{i',j}(\tilde{y}_0) - t_{i',j}(\tilde{y}_\varepsilon)| \qquad (8.22) \\ &\leq (4\varepsilon)/c \leq 1/(\sqrt{3}\omega_0). \end{aligned}$$

By Lemma 8.3, we therefore also have

$$p_{N,\omega_0}(t_{i,j}(\tilde{y}_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}) \geq 3/4.$$

With (8.17) and (8.18) for $x = \tilde{y}_\varepsilon$, inequality (8.16) now follows by observing that, for arbitrary $i, j, i' \in [m]$ with $i' \neq i - 1$, we have $|t_{i,j}(\tilde{y}_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}| \geq (N\pi)/\omega_0$.

Indeed, the reverse triangle inequality and (8.21) now yield

$$\begin{aligned}
&|t_{i,j}(\tilde{y}_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}| \\
&\quad \geq \Big| |t_{i,j}(\tilde{y}_0) - t_{i',j}(y_0) + T_i - T_{i'}| - |t_{i,j}(\tilde{y}_\varepsilon) - t_{i,j}(\tilde{y}_0) + t_{i',j}(\tilde{y}_0) - t_{i',j}(\tilde{y}_\varepsilon)| \Big| \qquad (8.23) \\
&\quad = \Big| |(T + a/c)(i - i' + 1)| - |t_{i,j}(\tilde{y}_\varepsilon) - t_{i,j}(\tilde{y}_0) + t_{i',j}(\tilde{y}_0) - t_{i',j}(\tilde{y}_\varepsilon)| \Big| .
\end{aligned}$$

Since by the assumption of the lemma

$$|(T + a/c)(i - i' + 1)| \geq (T + a/c) \geq T,$$

and

$$|t_{i,j}(\tilde{y}_\varepsilon) - t_{i,j}(\tilde{y}_0) + t_{i',j}(\tilde{y}_0) - t_{i',j}(\tilde{y}_\varepsilon)| \leq |t_{i,j}(\tilde{y}_\varepsilon) - t_{i,j}(\tilde{y}_0)| + |t_{i',j}(\tilde{y}_0) - t_{i',j}(\tilde{y}_\varepsilon)| \leq (4\varepsilon)/c,$$

it follows in (8.23) that

$$|t_{i,j}(\tilde{y}_\varepsilon) - t_{i',j}(y_\varepsilon) + T_i - T_{i'}| \geq T - (4\varepsilon)/c \geq (N\pi)/\omega_0. \qquad (8.24)$$

This completes the proof. □

Definition 8.8. Let $q_1, \ldots, q_m \in \mathbb{R}^3$ be arbitrary transducer locations, $\mathcal{J} \subset [m]^2$ be a set of measurements and $T_i \in \mathbb{R}$, $i \in [m]$ be arbitrary shot times. For arbitrary $x, y \in \mathbb{R}^3$ and arbitrary $\tau \geq 0$, define

$$\tilde{\mathcal{J}}(x, y; \tau) = \left\{ (i,j) \in \mathcal{J} \middle| \exists i' \in [m],\ i' \neq i :\ |t_{i,j}(x) - t_{i',j}(y) + T_{i'} - T_i| \leq \tau \right\}.$$

For arbitrary $X, Y \subset \mathbb{R}^3$, we set

$$\tilde{\mathcal{J}}(X, Y; \tau) = \bigcup_{x \in X} \bigcup_{y \in Y} \tilde{\mathcal{J}}(x, y; \tau).$$

Furthermore, define

$$\tilde{\mathcal{J}}^c(x, y; \tau) = \mathcal{J} \setminus \tilde{\mathcal{J}}(x, y; \tau),$$

and

$$\tilde{\mathcal{J}}^c(X, Y; \tau) = \mathcal{J} \setminus \tilde{\mathcal{J}}(X, Y; \tau).$$

With Definition 8.8 at hand, we will now analyze the superposition noise in the case of several point scatterers. To this end, suppose that, with m ultrasonic transducers located at $q_1, \ldots, q_m \in \mathbb{R}^3$, the specimen gets insonified by all m transducers with a raised cosine $p_{N,\omega_0}(\cdot - T_i)$ for arbitrary positive integer N, frequency $\omega_0 \in \mathbb{R}_+$ and shot times $T_i \in \mathbb{R}$, $i \in [m]$. Further, suppose that the defect consists of s point scatterers located at $Y = \{y_1, \dots, y_s\}$. Applying Lemma 8.6 together with Model Assumption 7.2, we now get

$$\begin{aligned} |\tilde{R}_{\mathcal{I}}(x)| =& \left| R_{\mathcal{I}}(x) + \sum_{\substack{i' \in [m] \\ i' \neq i}} \tilde{u}_{i,j}(t_{i,j}(x)) \right| \\ =& \left| R_{\mathcal{I}}(x) + \sum_{(i,j) \in \mathcal{I}} \sum_{\substack{i' \in [m] \\ i' \neq i}} \sum_{k \in [s]} a_{i',j}(y_k) p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y_k) + T_i - T_{i'}) \right|, \end{aligned} \tag{8.25}$$

with $a_{i,j}(y_k) \in \mathbb{R}_+$ for $k \in [s]$ and $i,j \in [m]$. The sum on the right hand side of (8.25) is the superposition noise that we have already encountered in Lemma 8.6. With $\tau = {}^{(N\pi)}/_{\omega_0}$ in Definition 8.8, such that $p_{N,\omega_0}(t) = 0$ for $|t| \geq \tau$, we now have for all $(i,j) \in \tilde{\mathcal{I}}^c(x, Y; {}^{N\pi}/_{\omega_0})$, $i' \in [m]$ with $i' \neq i$, and $y \in Y$ that

$$p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y) + T_i - T_{i'}) = 0.$$

We can therefore rewrite (8.25) to

$$|\tilde{R}_{\mathcal{I}}(x)| = \left| R_{\mathcal{I}}(x) + \sum_{(i,j) \in \tilde{\mathcal{I}}(x, Y; {}^{N\pi}/_{\omega_0})} \sum_{\substack{i' \in [m] \\ i' \neq i}} \sum_{k \in [s]} a_{i',j}(y_k) p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y_k) + T_i - T_{i'}) \right|.$$

As we already have seen in Lemma 8.7, even in the case of a single point scatterer, the superposition noise can lead to ambiguities and artifacts. Here, most of the terms corresponding to the superposition noise do not vanish. Indeed, by (8.22) and (8.24), for $\tau \geq {}^{1}/_{(\sqrt{3})\omega_0}$, it holds that $|\mathcal{I}(\tilde{y}_\varepsilon, y_\varepsilon; \tau)| \geq m(m-1)$. If the defect is sparse, and we choose the shot times T_i, $i \in [m]$ independent and uniformly distributed, then the set of measurements $\tilde{\mathcal{I}}(x, Y, \tau) \subset \mathcal{I}$ responsible for the superposition noise at x, is small in cardinality with high probability, as we will see in the following theorem.

Theorem 8.9. *Let $x, q_1, \dots, q_m \in \mathbb{R}^3$, $Y \subset \mathbb{R}^3$ with $|Y| \leq s$, and $\tau > 0$ be arbitrary. Let further T_i, $i \in [m]$ be independent and uniformly distributed on an interval I of length mL for some $L > 0$ and $\mathcal{I} = \mathcal{I}_{FMC} = [m]^2$. Then, for $\varepsilon, \delta > 0$, it holds that*

$$\mathbb{P}\left[\frac{|\tilde{\mathcal{I}}(x, Y; \tau)|}{m^2} \geq \delta \right] \leq \varepsilon,$$

provided

$$s \leq \frac{\varepsilon \delta L}{2\tau}.$$

The proof of Theorem 8.9 makes use of Markov's inequality, see, e.g., [Kle13].

Theorem 8.10 (Markov's inequality). *Let X be a non-negative random variable and $t > 0$. Then*

$$\mathbb{P}\left[X \geq t\right] \leq \frac{\mathbb{E}(x)}{t}.$$

Proof of Theorem 8.9. For arbitrary $\mathcal{G} \subset [m]^2$ and $(i,j) \in [m]^2$, let $1_{\mathcal{G}}$ be the characteristic function of $\mathcal{G}$, defined by

$$1_{\mathcal{G}}(i,j) = \begin{cases} 1 & (i,j) \in \mathcal{G}, \\ 0 & \text{else.} \end{cases}$$

We now aim to use Markov's inequality, and write

$$\mathbb{E}\left[|\tilde{\mathcal{G}}(x,Y;\tau)|\right] = \mathbb{E}\left[\sum_{(i,j)\in[m]^2} 1_{\tilde{\mathcal{G}}(x,Y;\tau)}(i,j)\right] = \sum_{(i,j)\in[m]^2} P_{i,j}, \tag{8.26}$$

where we set

$$P_{i,j} := \mathbb{P}[(i,j) \in \tilde{\mathcal{G}}(x,Y;\tau)] = \mathbb{E}[1_{\tilde{\mathcal{G}}(x,Y;\tau)}(i,j)].$$

Conditioning on all the $T_{i'}$, $i' \neq i$, we now have for arbitrary $i,j \in [m]$,

$$\begin{aligned} P_{i,j} &= \mathbb{P}\left[\exists y \in Y, \exists i' \in [m], i' \neq i : |t_{i,j}(x) - t_{i',j}(y) + T_i - T_{i'}| \leq \tau\right] \qquad (8.27) \\ &= \mathbb{E}\left[\mathbb{P}\left[\exists y \in Y, \exists i' \in [m], i' \neq i : |t_{i,j}(x) - t_{i',j}(y) + T_i - T_{i'}| \leq \tau \,\middle|\, T_{i'},\ i' \neq i\right]\right]. \end{aligned}$$

Now observe that for $i \in [m]$

$$\begin{aligned} &\mathbb{P}\left[\exists y \in Y, \exists i' \in [m], i' \neq i \ : \ |t_{i,j}(x) - t_{i',j}(y) + T_i - T_{i'}| \leq \tau \,\middle|\, T_{i'},\ i' \neq i\right] \\ &\leq \sum_{y\in Y} \sum_{\substack{i'\in[m] \\ i'\neq i}} \mathbb{P}\left[|t_{i,j}(x) - t_{i',j}(y) + T_i - T_{i'}| \leq \tau \,\middle|\, T_{i'},\ i' \neq i\right] \qquad (8.28) \\ &\leq {}^{(2\tau s)}/_{L}, \end{aligned}$$

as T_i is uniformly distributed on I with length mL and $|Y| \leq s$. Consequently, (8.28) and (8.27) imply

$$\mathbb{E}\left[|\tilde{\mathcal{G}}(x,Y;\tau)|\right] \leq \frac{2\tau s m^2}{L}.$$

Thus Markov's inequality now yields

$$\mathbb{P}\left[\frac{|\tilde{\mathcal{I}}(x,Y;\tau)|}{m^2} \geq \delta\right] \leq \frac{\mathbb{E}\left[|\tilde{\mathcal{I}}(x,Y;\tau)|\right]}{\delta m^2} \leq \frac{2\tau s}{\delta L},$$

which is bounded by ε provided

$$s \leq \frac{\varepsilon \delta L}{2\tau}.$$

This completes the proof. □

The observation of Theorem 8.9 will be the key ingredient for reducing the superposition noise via an iterative SAFT algorithm, which we will present in the following section.

9. Iterative Synthetic Aperture Focusing Technique

In order to reduce the superposition noise, we will now develop an iterative SAFT algorithm for superposed measurements. Since the defects occurring in applications are extended scatterers, we will restrict our analysis to this case. The following lemma will allow us to transfer the results of Section 8 to the case of extended scatterers.

Lemma 9.1. *Let $q_1, \ldots, q_m \in \mathbb{R}^3$ be arbitrary transducer locations and $T_i \in \mathbb{R}$, $i \in [m]$ be arbitrary shot times. For an arbitrary set of measurements $\mathcal{I} \subset [m]^2$, $X, Y \subset \mathbb{R}^3$, and $r_1, r_2, \tau \geq 0$, we have*

$$\tilde{\mathcal{I}}(B_{r_1}(X), B_{r_2}(Y); \tau) \subset \tilde{\mathcal{I}}(X, Y; \tau + {}^{2(r_1+r_2)}/_{c}).$$

In particular, for arbitrary $x \in \mathbb{R}^3$, we have

$$\tilde{\mathcal{I}}(x, B_{r_2}(Y); \tau) \subset \tilde{\mathcal{I}}(x, Y; \tau + {}^{(2r_2)}/_{c}).$$

Proof. Let $y \in Y$, $x \in X$ be arbitrary and $(i,j) \in \tilde{\mathcal{I}}(B_{r_1}(x), B_{r_2}(y); \tau)$. Then, there exist $x' \in B_{r_1}(x)$ and $y' \in B_{r_2}(y)$ such that $(i,j) \in \tilde{\mathcal{I}}(x', y'; \tau)$. This, in turn, implies that there exists $i' \in [m]$, $i' \neq i$ with $|t_{i,j}(x') - t_{i',j}(y') + T_{i'} - T_i| \leq \tau$. With Lemma 8.4, we get

$$\begin{aligned}|t_{i,j}(x) - t_{i',j}(y) + T_{i'} - T_i| \\ &\leq |t_{i,j}(x') - t_{i',j}(y') + T_{i'} - T_i| + |t_{i',j}(x) - t_{i',j}(x')| + |t_{i',j}(y') - t_{i',j}(y)| \\ &\leq \tau + |t_{i',j}(x) - t_{i',j}(x')| + |t_{i',j}(y') - t_{i',j}(y)| \\ &\leq \tau + {}^{2(r_1+r_2)}/_{c}.\end{aligned} \tag{9.1}$$

Bearing in mind that $(i,j) \in \tilde{\mathcal{I}}(B_{r_1}(x), B_{r_2}(y); \tau)$ was arbitrary, we now have

$$\tilde{\mathcal{I}}(B_{r_1}(x), B_{r_2}(y); \tau) \subset \tilde{\mathcal{I}}(x, y; \tau + {}^{2(r_1+r_2)}/_{c}). \tag{9.2}$$

Applying now (9.2) to all $x \in X$, $y \in Y$, we finally get

$$\begin{aligned}\tilde{\mathcal{I}}(B_{r_1}(x), B_{r_2}(Y); \tau) &= \bigcup_{x \in X} \bigcup_{y \in Y} \tilde{\mathcal{I}}(B_{r_1}(x), B_{r_2}(y); \tau) \\ &\subset \bigcup_{x \in X} \bigcup_{y \in Y} \tilde{\mathcal{I}}(x, y; \tau + {}^{2(r_1+r_2)}/_{c}) \\ &= \tilde{\mathcal{I}}(X, Y; \tau + {}^{2(r_1+r_2)}/_{c}).\end{aligned} \tag{9.3}$$

This establishes the first part of the lemma. The second part directly follows by setting $X = \{x\}$ and $r_1 = 0$. □

Let now $q_1, \ldots, q_m \in \mathbb{R}^3$ be arbitrary transducer locations, $\mathcal{I} = [m]^2$ be the set of measurements, and p_{N,ω_0} be a raised cosine for a frequency $\omega_0 \in \mathbb{R}$ and positive integer N. Let the shot times T_i, $i \in [m]$ be independent copies of a uniform random variable on an interval $I \subset \mathbb{R}$ to be chosen below and $D \subset \mathbb{R}^3$ be an extended scatterer. Model Assumption 7.4 together with Lemma 8.6 now implies that there exist bounded functions $a_{i,j} \in B(\partial D, \mathbb{R}_+)$, $(i,j) \in [m]^2$, such that for arbitrary $x \in \mathbb{R}^3$, we have

$$|\tilde{R}_{\mathcal{I}}(x)| = \Bigg| R_{\mathcal{I}}(x) \tag{9.4}$$

$$+ \sum_{(i,j) \in \mathcal{I}} \sum_{\substack{i' \in [m] \\ i \neq i'}} \int_{\partial D} a_{i',j}(y) p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y) + T_{i'} - T_i) \, \mathrm{d}\, y \Bigg|. \tag{9.5}$$

Consider the superposition noise in (9.5). With $\tau = {}^{(N\pi)}/_{\omega_0}$ in Definition 8.8, for all $(i,j) \in \tilde{\mathcal{I}}^c(x, \partial D; \tau)$ and arbitrary $i' \in [m]$, $i' \neq i$, we have $p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y) + T_{i'} - T_i) = 0$. Therefore, only the terms with $(i,j) \in \tilde{\mathcal{I}}(x, \partial D; {}^{(N\pi)}/_{\omega_0})$ contribute to the superposition noise in (9.5). We now propose an iterative SAFT algorithm, which adjusts the index set $\mathcal{I}$ used in SAFT at all points, where we aim to compute the SAFT backprojection. It is geared to reduce (9.5) by removing ultrasonic measurements indexed by $\tilde{\mathcal{J}}(x,S;\tau)$ for some $\tau > 0$, where $S \subset \mathbb{R}^3$ is a defect identified by SAFT in a greedy manner.

Algorithm 1: Iterative SAFT

```
Data: Superposed measurements $\tilde{u}_{i,j}$, $(i,j) \in \mathcal{J} = [m]^2$, shot times $T_i \in \mathbb{R}$, $i \in [m]$.
Input: Discretization of specimen $X \subset \mathbb{R}^3$,
       Radius $r > 0$,
       Time parameter $\tau > 0$,
       Maximum number of iterations $M$,
       Threshold parameter $\theta \in [0,1]$.
Variables: Maximum defect implication: $y_k \in X$,
           Identified defect: $S_k \in X$,
           Removed measurements for $x \in X$: $\mathcal{J}_k(x)$.
Result: Iterative SAFT backprojection $(\tilde{R}^{(M)}(x))_{x \in X}$.
1  Initialization: $\forall x \in X : \mathcal{J}_0(x) = \emptyset$, $\tilde{R}^{(0)}(x) = 0$ ; $S_0 = \emptyset$.
2  begin
3      for $k \in [M]$ do
4          for $x \in S_{k-1}$ do
5              $\tilde{R}^{(k)}(x) = \tilde{R}^{(k-1)}(x)$
6          for $x \in X \setminus S_{k-1}$ do
7              $\tilde{R}^{(k)}(x) = |\tilde{R}_{\mathcal{J} \setminus \mathcal{J}_{k-1}(x)}(x)|$
8          if $k<M$ then
9              $y_k = \text{argmax}_{x \in X \setminus S_{k-1}} |\tilde{R}^{(k)}(x)|$
10             $S_k = S_{k-1} \cup \left\{ x \in B_r(y_k) \cap X \middle| |\tilde{R}^{(k)}(x)| \geq \theta |\tilde{R}^{(k)}(y_k)| \right\}$
11             for $x \in X \setminus S_k$ do
12                 $\mathcal{J}_k(x) = \tilde{\mathcal{J}}(x, S_k; \tau)$
```

We will now analyze the iterative SAFT backprojection in terms of the defect D. Let $x \in \mathbb{R}^3$ be arbitrary and let $k \in [M]$ be the largest k' with $x \notin S_{k-1}$. It follows that

$$\left|\tilde{R}^{(M)}(x)\right| = \left|\tilde{R}^{(k)}(x)\right| = \left|\tilde{R}_{\mathcal{J} \setminus \mathcal{J}_{k-1}(x)}(x)\right| = \left|\tilde{R}_{\mathcal{J}^c(x,S_{k-1};\tau)}(x)\right|$$

$$= \Bigg| R_{\tilde{\mathcal{J}}^c(x,S_{k-1};\tau)}(x) \tag{9.6}$$

$$+ \sum_{(i,j) \in \tilde{\mathcal{J}}^c(x,S_{k-1};\tau)} \sum_{\substack{i' \in [m] \\ i \neq i'}} \int_{\partial D} a_{i',j}(y) p_{N,\omega_0}(t_{i,j}(x) - t_{i',j}(y) + T_{i'} - T_i) \, \mathrm{d}\, y \Bigg|. \tag{9.7}$$

In contrast to the superposition noise in (9.5), we now have (9.7), and (9.6) instead of the SAFT backprojection $R_{\mathcal{J}}(x)$. The greedy step in line 9 and line 10 of iteration k was aiming at identifying possible defects as S_{k-1}. It makes use of the heuristics behind the SAFT backprojection, which was discussed in Section 8. Now let $\tau = {}^{(N\pi)}/_{\omega_0}$, and suppose that, after $(M-1)$ iterations of Algorithm 1, we have $S_{M-1} \supset D$. This would in particular imply that D is $(M-1, r)$-sparse. Furthermore, with the definition of $\tilde{\mathcal{I}}(x, S_{k-1}; \tau)$ and since $S_{k-1} \subset S_{M-1}$, the superposition noise in (9.7) would vanish for all $x \in \mathbb{R}^3 \setminus S_{M-1}$.

For sparse defects, the idea of using superposed measurements for data aquisition and iterative SAFT for imaging rather than acquiring basic measurements and using SAFT, now is the following: Suppose that collecting basic measurements $u_{i,j} \in L_1$, $(i,j) \in \mathcal{I} = [m]^2$ takes an amount of time of mT, where T is the amount of time needed to measure all basic measurements $u_{i,j}$, $j \in [m]$ for $i \in [m]$ in parallel, using for instance a phased array probe. Let $\ell \in [m]$, $\ell \neq 1$ be arbitrary. Performing superposed measurements with shot times T_i, $i \in [m]$ chosen independent and uniformly distributed in $[0, (m-\ell)T]$, now allows us to acquire the same number of superposed measurements $\tilde{u}_{i,j}$, $(i,j) \in \mathcal{I} = [m]^2$ in a shorter period of time of at most $(m-\ell+1)T < mT$. In the same amount of time, we only can acquire $(m-\ell+1)m$ basic measurements by sequentially insonifying the specimen with $(m-\ell+1) \leq m$ ultrasonic transducers. While we can measure more data using superposed measurements, this also comes with the cost of superposition noise, which may cause ambiguities and artifacts as discussed in Section 8. To remove these artifacts, we use the iterative SAFT method instead of SAFT. For $x \in \mathbb{R}^3$ and $S_{k-1} \subset \mathbb{R}^3$ as above, after the superposition noise is removed, we only compute the SAFT backprojection with respect to $\tilde{\mathcal{I}}^c(x, S_{k-1}; \tau)$ in (9.6). This corresponds to effectively more information as used in SAFT with basic measurements acquired in the same amount of time, provided

$$|\tilde{\mathcal{I}}^c(x, S_{M-1}; \tau)| > (m-\ell+1)m.$$

Under certain requirements, this condition is satisfied for all $x \in X$ with high probability.

Theorem 9.2. *In Algorithm 1, let for arbitrary $\ell \in [m]$, $\ell \neq 1$ and $T > 0$, the shot times T_i, $i \in [m]$ be chosen independent and uniformly distributed in $[0, (m-\ell)T]$. Suppose that there exist $z \in \mathbb{R}^3$ and $R > 0$, such that $X \subset B_R(z)$. Then with probability at least $(1-\varepsilon)$, it holds that for all $x \in X$ and $[M] \ni k = \max\{k' | x \notin S_{k'-1}\}$*

$$|\mathcal{I} \setminus \mathcal{I}_k(x)| = |\tilde{\mathcal{I}}^c(x, S_{k-1}; \tau)| > (m-\ell+1)m, \tag{9.8}$$

provided

$$M \leq \frac{(\ell-1)(m-\ell)}{m} \frac{\varepsilon T}{2(\tau + {}^{2(R+r)}/_{c})}. \tag{9.9}$$

As S_{M-1} in Algorithm 1 is $(M-1,r)$-sparse, (9.9) also is a bound on the maximal sparsity of the identified defect.

Proof of Theorem 9.2. Let $x \in X$ be arbitrary and $k \in [M]$ be the largest k' with $x \notin S_{k'-1}$. Set $Y = \{y_1, \dots, y_{M-1}\} \subset \mathbb{R}^3$ with y_k, $k \in [M-1]$ as in line 9 of the algorithm. Since $S_{k-1} \subset B_r(Y)$, it now follows with Lemma 9.1 that

$$\begin{aligned} |\mathcal{G}_k(x)| &= |\tilde{\mathcal{G}}(x, S_{k-1}; \tau)| \leq |\tilde{\mathcal{G}}(B_R(z), B_r(Y); \tau)| \\ &\leq |\tilde{\mathcal{G}}(z, Y; \tau + {}^{2(R+r)}/_{c}). \end{aligned} \tag{9.10}$$

Applying Theorem 8.9 for $\varepsilon > 0$ and $\delta = \delta_l := {}^{(\ell-1)}/_{m}$, while bearing in mind that $|Y| \leq M$, we have with probability at most ε that

$$|\tilde{\mathcal{G}}(z, Y; \tau + {}^{2(R+r)}/_{c})| \geq \delta_l m^2 = (\ell - 1)m, \tag{9.11}$$

provided M satisfies (9.9). Inequality (9.11) together with (9.10) now implies

$$|\mathcal{G} \setminus \mathcal{G}_k(x)| = m^2 - |\mathcal{G}_k(x)| > (m - \ell + 1)m,$$

which completes the proof. □

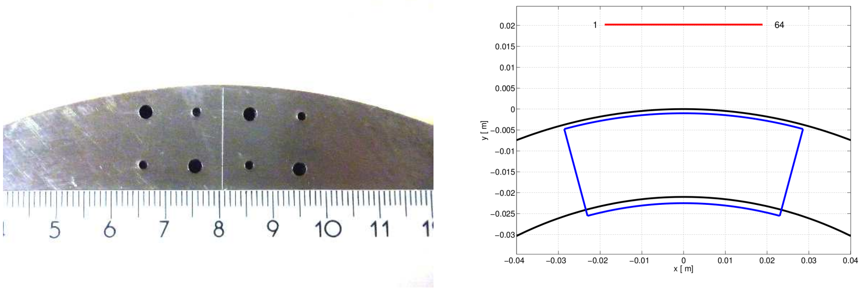
10. Numerical Results

The ultrasonic data for our experiments was provided by *Salzgitter Mannesmann* Forschung GmbH in Duisburg, Germany. It was also used in [SNCO16]. For our experiments, we consider a steel pipe of radius 223mm and a thickness of 21mm; 8 drilled holes of 1mm and 2mm diameter serve as model defects, see Figure 10.1(a). The full-matrix-capture basic measurements were acquired using a 64 element 5MHz phased array probe with a distance between consecutive transducers (pitch) of 0.6mm. As depicted in Figure 10.1(b), it is placed in a distance of 20mm above the specimen.

The corresponding time-of-flights were derived via Fermat's principle, see [SNCO16]. Each basic measurement was measured in 3.2×10^{-5}s with a sampling rate of 10^8Hz. From these, we simulated the corresponding superposed measurements using Lemma 7.6. In three experiments, we choose the shot times T_i, $i \in [64]$ independent and uniformly distributed in $[0, (m - \ell_k)T]$, for $T = 3.2 \times 10^{-5}$s and

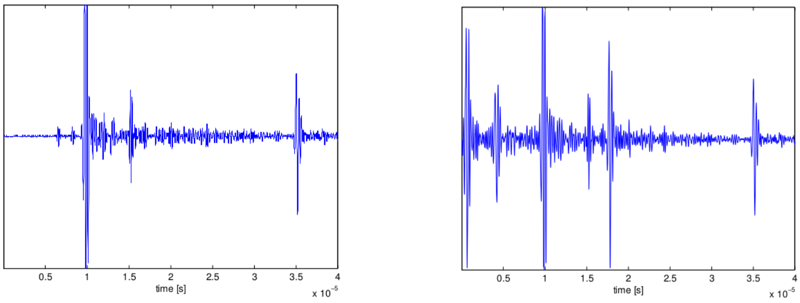
$$\ell_1 = 13, \qquad \ell_2 = 23, \qquad \ell_3 = 33.$$

In Figure 10.2, we exemplary show the basic measurement $u_{1,32}$ and the corresponding superposed measurements $\tilde{u}_{1,32}$ in the case of $\ell_1 = 13$. Here, one can see that many of the distinctive features in $u_{1,32}$ are also visible in $\tilde{u}_{1,32}$.



(a) model defect: 8 side-drilled holes **(b)** schematic experiment setup

Figure 10.1.: experiment setup



(a) $u_{1,32}$ **(b)** $\tilde{u}_{1,32}$ for $\ell_1 = 13$

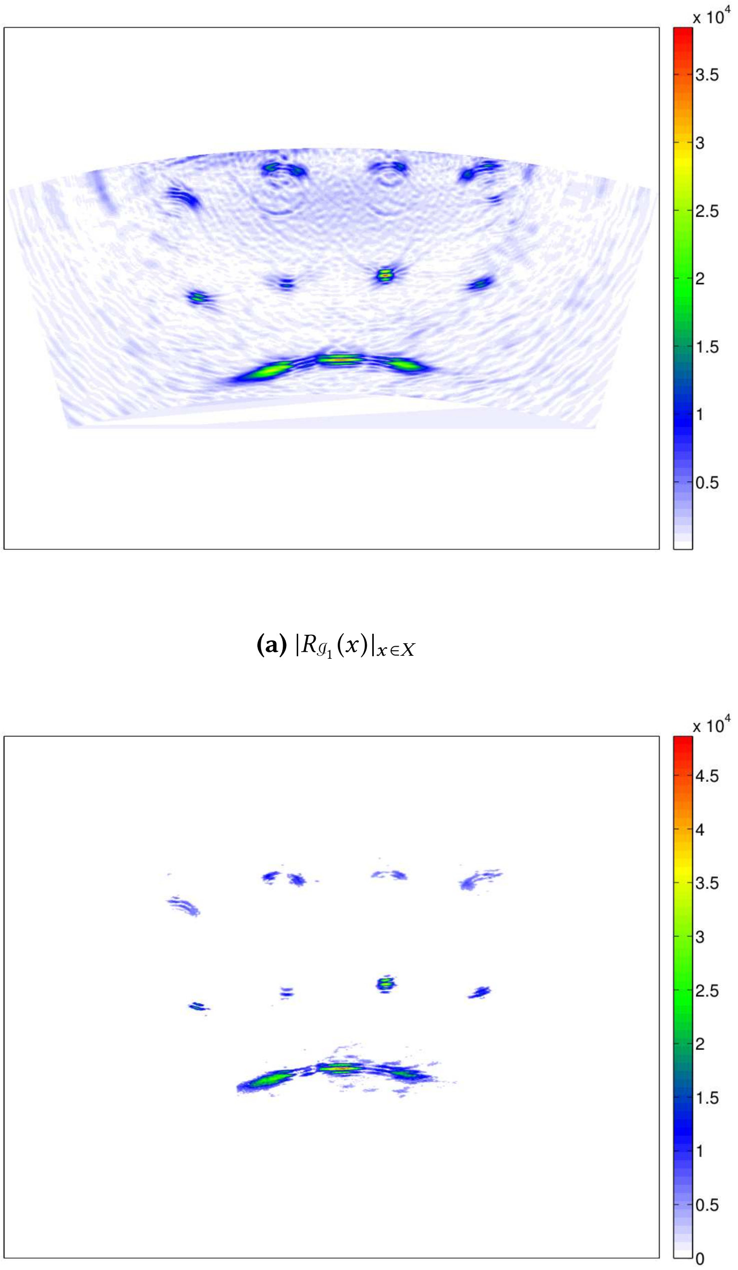
Figure 10.2.: corresponding basic (a) and superposed measurements (b)

We use all superposed measurements indexed by $\mathcal{I} = \mathcal{I}_{\mathrm{FMC}} = [m]^2$ in the corresponding iterative SAFT backprojection. As a benchmark, we compute the SAFT backprojection with $\mathcal{I} = I_k \times [m]$, where $I_k \subset [m]$ is evenly chosen with

$$|I_1| = 52, \quad |I_2| = 42, \quad |I_3| = 32.$$

In this way, we compare the performance of both methods with respect to ultrasonic data, which was acquired in the same amount of time. As it is common practice in nondestructive ultrasonic testing, we compute the discrete Hilbert transform of the basic and superposed measurements and perform the algorithms using the respective signals, see [LMK12] and the discussion of SAFT in Section 8. In all cases, we use a 300×375 pixel polar grid $X \subset \mathbb{R}^3$ to discretize the region of interest. It is bounded by the blue frame in Figure 10.1(b). For the iterative SAFT backprojection, we choose the parameters $r = 2.5\mathrm{mm}$, $\tau = 0.1\mathrm{ns}$, and $\theta = 0.4$. As the number M of iterations, we choose 20. In figures 10.3, 10.4, and 10.5, we compare the SAFT backprojection with the corresponding iterative SAFT backprojection restricted to the found defect S_{19} for ℓ_k, $k = 1, 2, 3$. The 8 drilled holes as well as the backwall are recognizable in the SAFT backprojections in Figures 10.3(a), 10.4(a), 10.5(a). As suggested by the colorbars, the pixel values are nearly proportional to the number of basic measurements, and therefore also to the amount of measurement time.

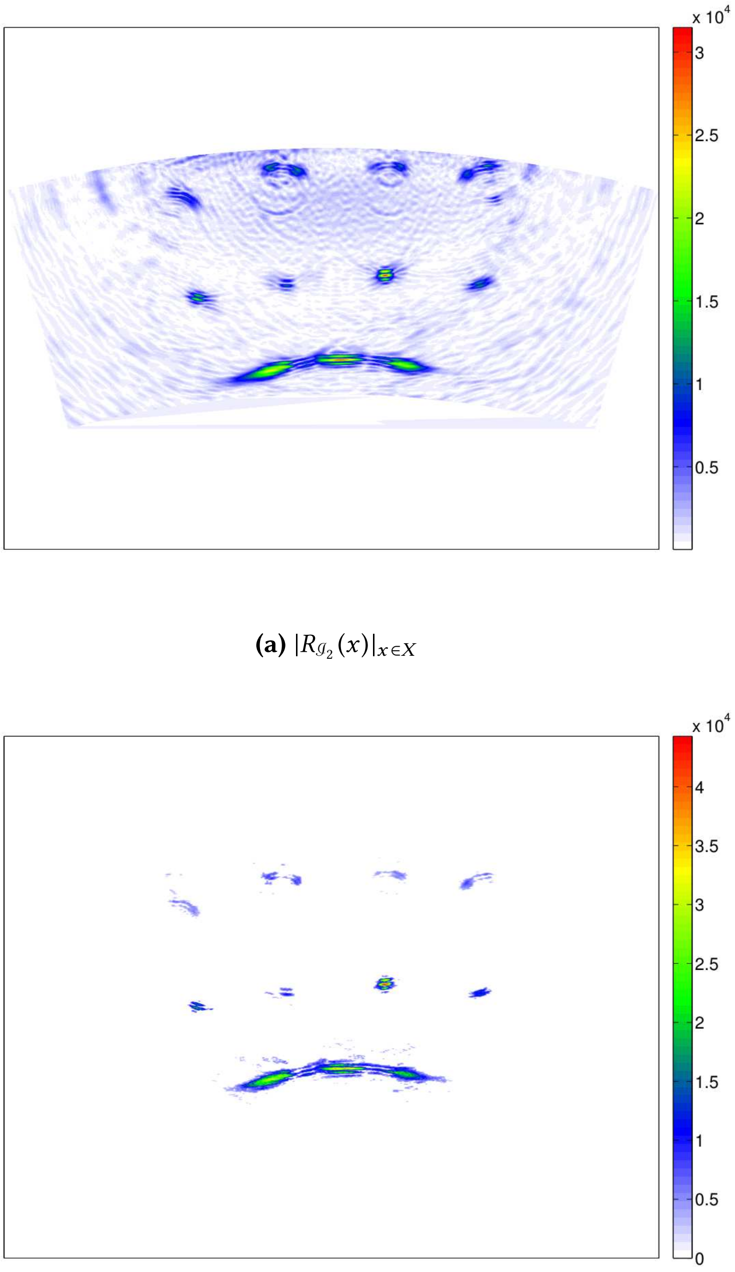
In Figures 10.3(b), 10.4(b), 10.5(b), we can see the iterative SAFT backprojections restricted to the identified defect S_{19}. Except for the last case, where we missed the two holes in the upper right, the 8 drilled holes and the backwall are recognizable in the three images. In all three cases, the maximum pixel value is almost the same; comparable high values are achieved at the backwall and the third hole in the second row; these are the defects found in the first iterations of the iterative SAFT algorithm. All remaining holes have smaller pixel values, as they are detected in later iterations and therefore using less measurements. Next, we consider Figures 10.6, 10.7, and 10.8. Comparing the first iteration with the last iteration of the iterative SAFT algorithm in the respective cases, the reconstructions after 20 iterations are less noisy. In order to directly compare the reconstructions of both methods, we investigate their differences in Figures 10.9, 10.10, 10.11. Consider 10.9(a), 10.10(a), and 10.11(a), where the difference between SAFT and iterative SAFT after one iteration is computed. Here, we can recognize the 8 holes and the backwall in red, indicating higher pixel values in the latter method. We also see the additional superposition noise in red. In Figures 10.9(b),10.10(b), and 10.11(b), we can see that much of the superposition noise is removed after 20 iterations. This comes with the cost of lower pixel values at the defects discovered in later iterations. But for sparse defects such as one drilled hole, which can be identified in a small number of iterations, iterative SAFT indeed gives better defect reconstructions, as suggested by third hole in the second row; this matches the theoretical considerations of Theorem 9.2.



(a) $|R_{\mathcal{g}_1}(x)|_{x\in X}$

(b) $\left|\tilde{R}^{(20)}_{\mathcal{g}}(x)\right|_{x\in S_{19}}$

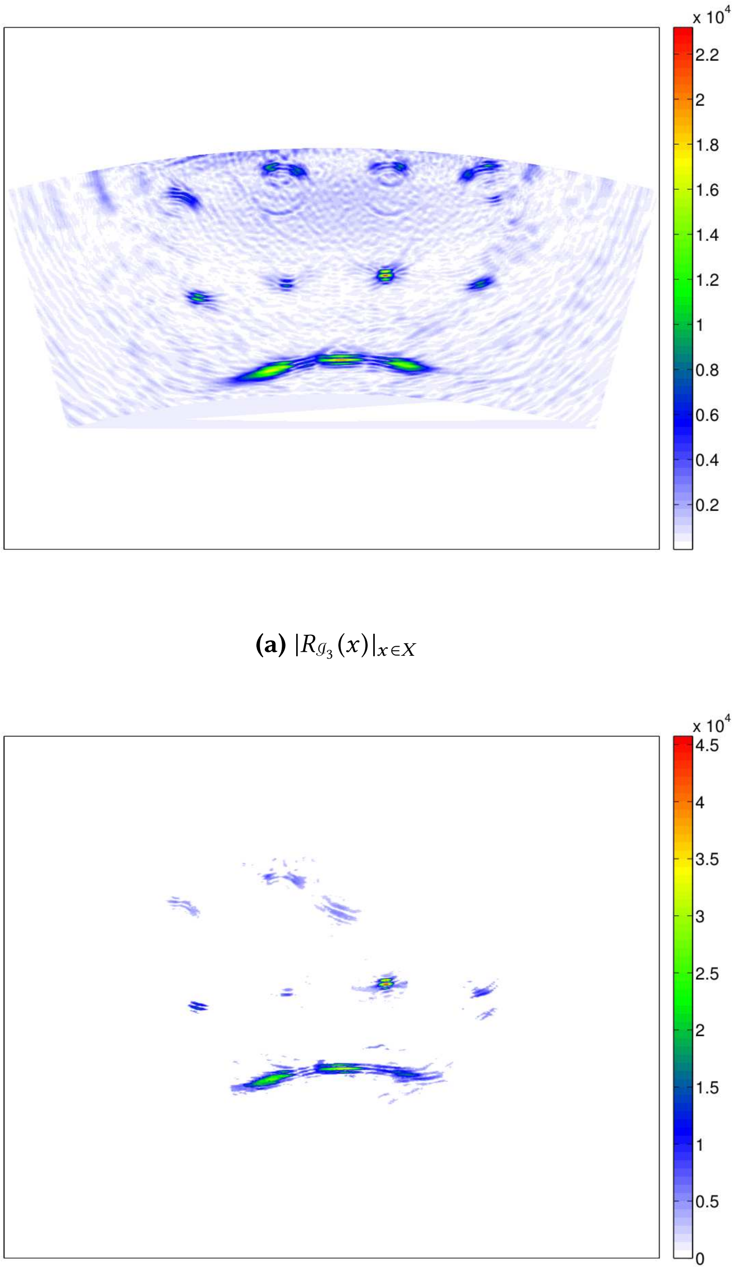
Figure 10.3.: SAFT with $|I_1| = 52$ (a) and corresponding iterative SAFT restricted to the found defect S_{19} for $\ell_1 = 13$ (b)



(a) $|R_{\mathcal{J}_2}(x)|_{x\in X}$

(b) $\left|\tilde{R}_{\mathcal{J}}^{(20)}(x)\right|_{x\in S_{19}}$

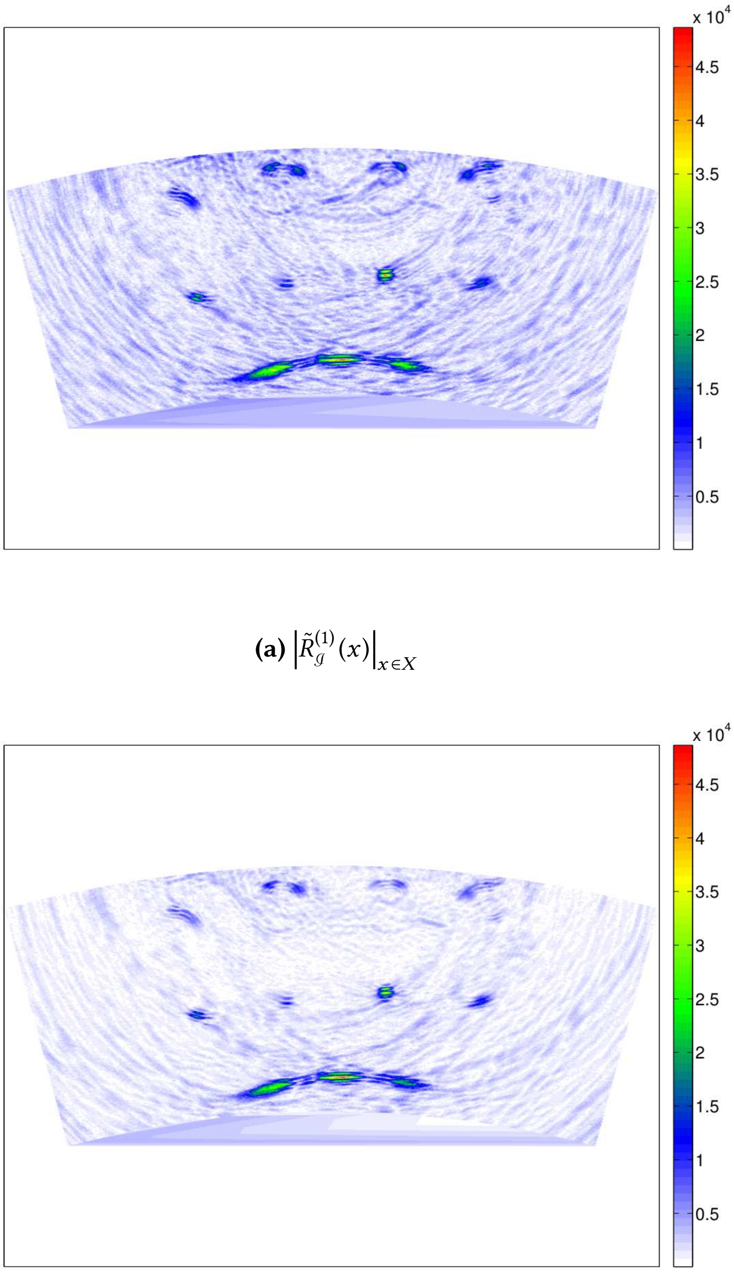
Figure 10.4.: SAFT with $|I_2| = 42$ (a) and corresponding iterative SAFT restricted to the found defect S_{19} for $\ell_2 = 23$ (b)



(a) $|R_{\mathscr{g}_3}(x)|_{x\in X}$

(b) $\left|\tilde{R}_{\mathscr{g}}^{(20)}(x)\right|_{x\in S_{19}}$

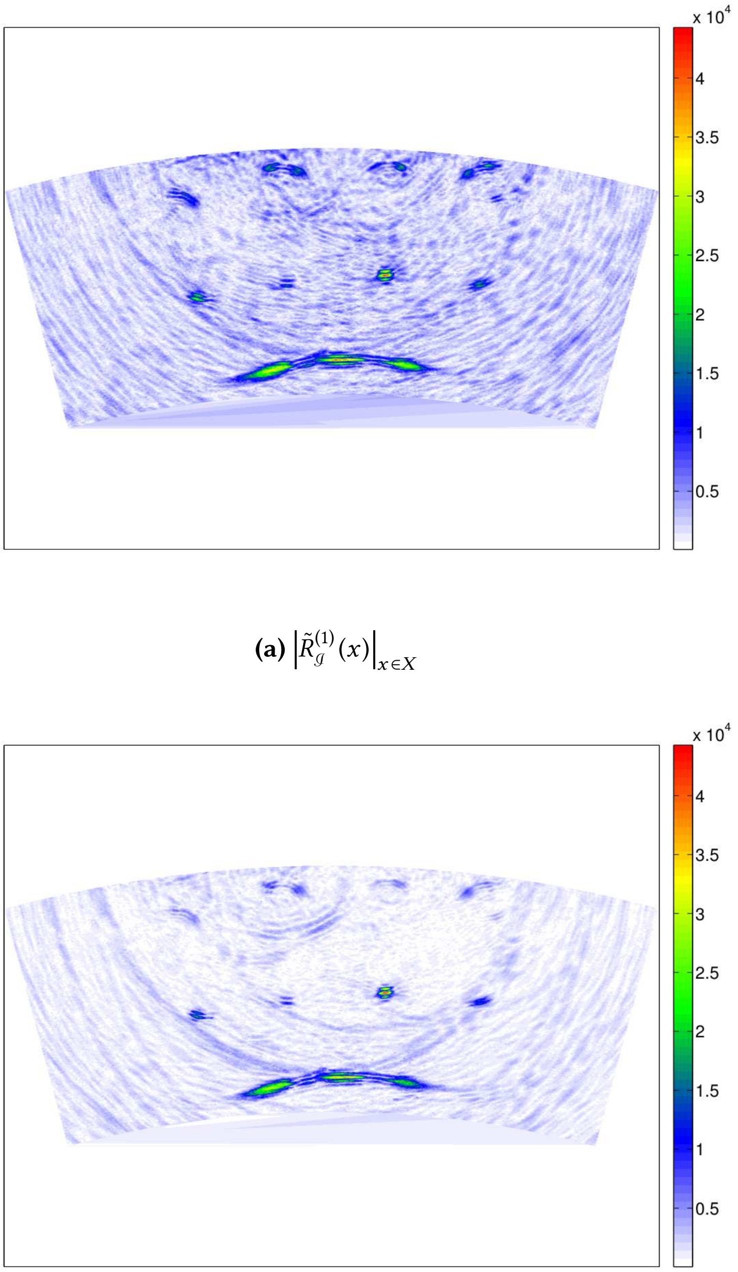
Figure 10.5.: SAFT with $|I_3| = 32$ (a) and corresponding iterative SAFT restricted to the found defect S_{19} for $\ell_3 = 33$ (b)



(a) $\left|\tilde{R}_{\mathcal{g}}^{(1)}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}_{\mathcal{g}}^{(20)}(x)\right|_{x\in X}$

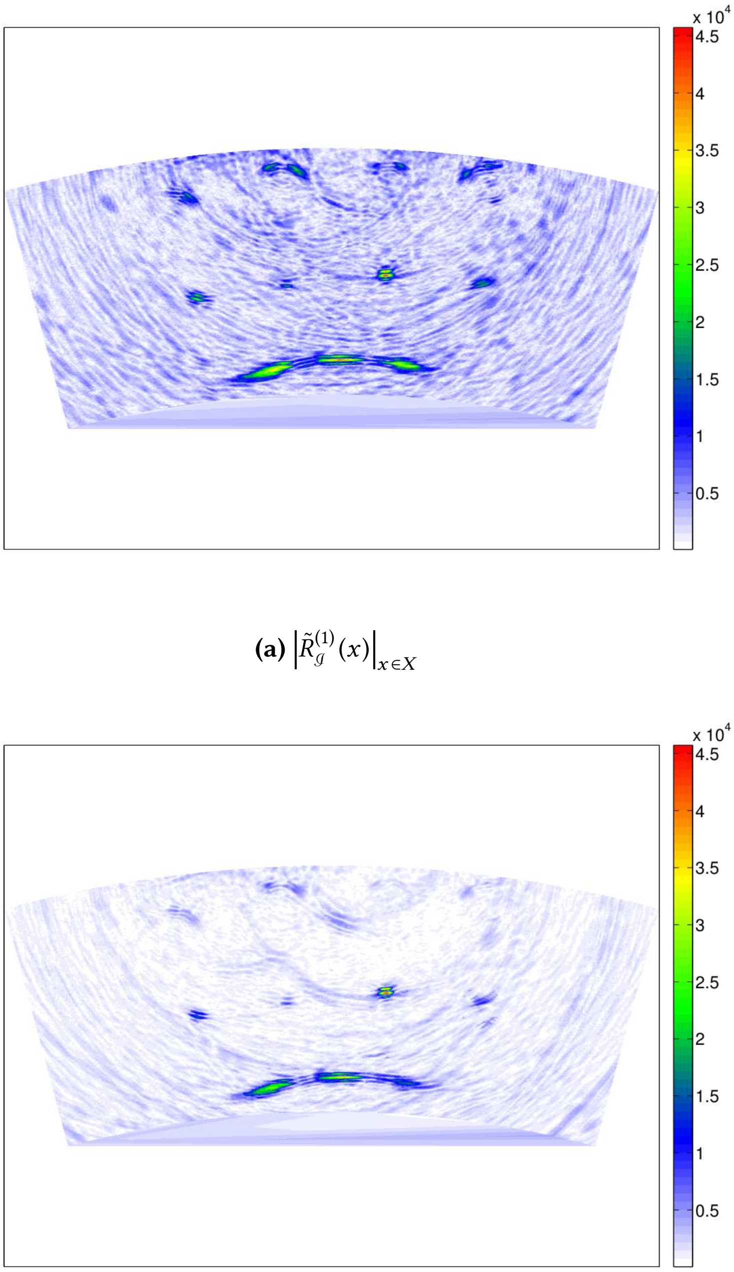
Figure 10.6.: iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_1 = 13$



(a) $\left|\tilde{R}_g^{(1)}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}_g^{(20)}(x)\right|_{x\in X}$

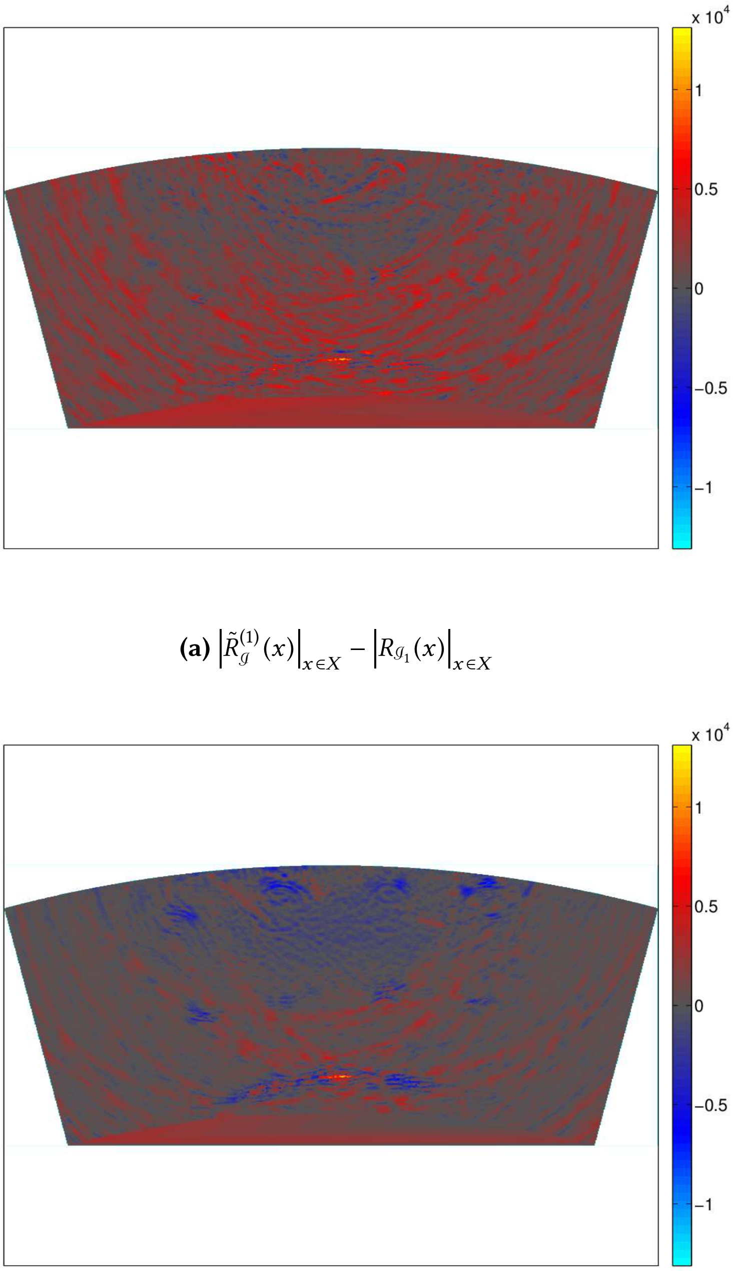
Figure 10.7.: iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_2 = 23$



(a) $\left|\tilde{R}^{(1)}_{\mathcal{G}}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}^{(20)}_{\mathcal{G}}(x)\right|_{x\in X}$

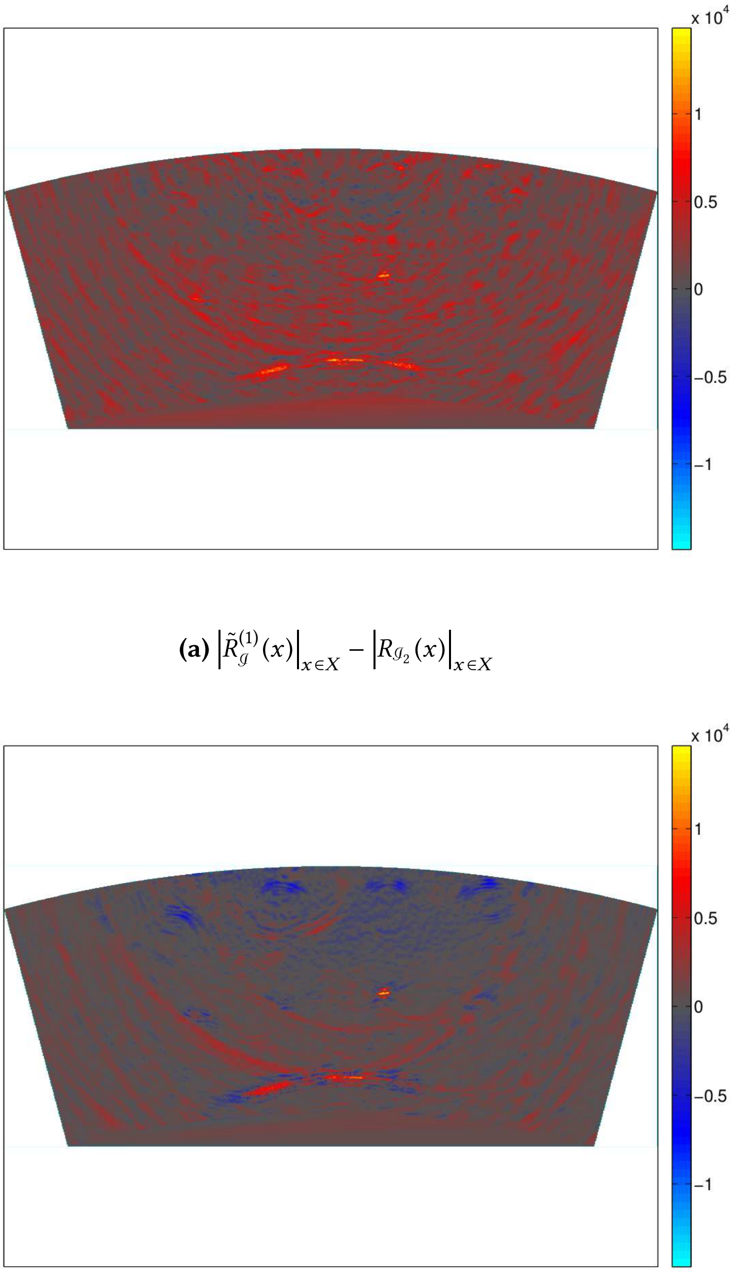
Figure 10.8.: iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_3 = 33$



(a) $\left|\tilde{R}^{(1)}_{\mathcal{G}}(x)\right|_{x\in X} - \left|R_{\mathcal{G}_1}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}^{(20)}_{\mathcal{G}}(x)\right|_{x\in X} - \left|R_{\mathcal{G}_1}(x)\right|_{x\in X}$

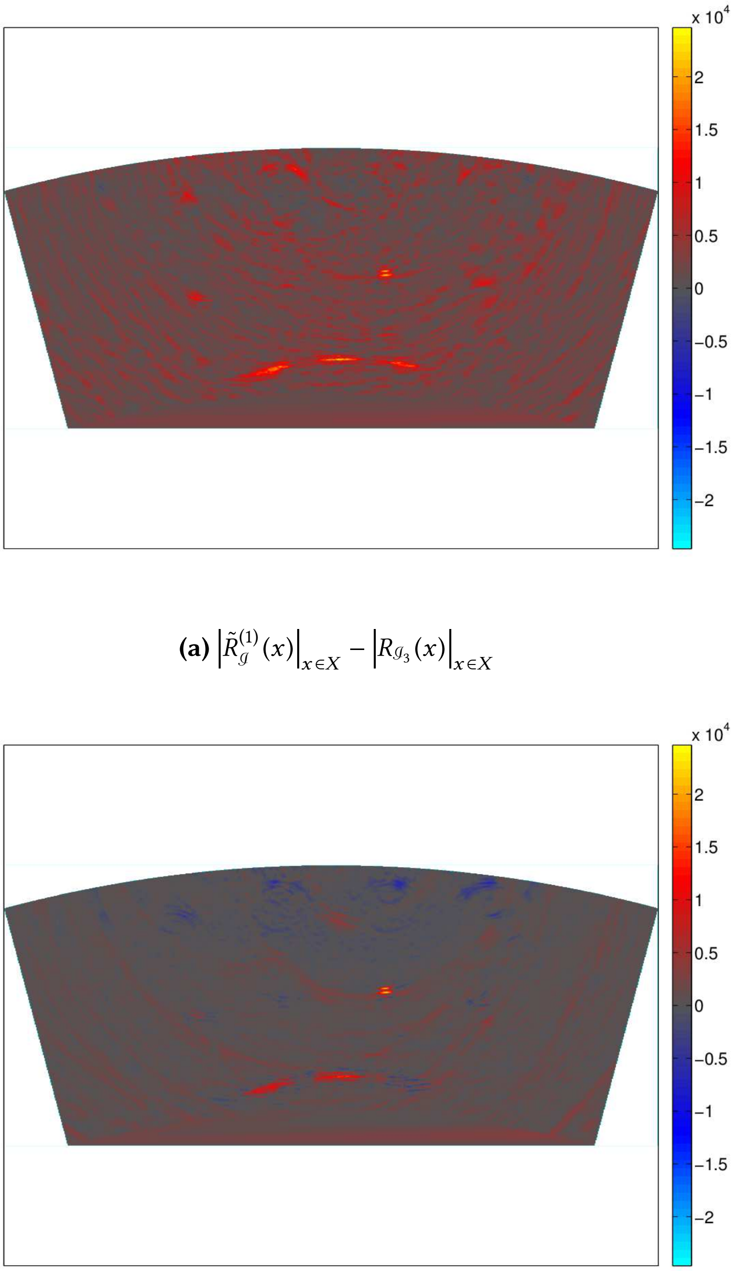
Figure 10.9.: difference of SAFT with $|I_1| = 52$ and iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_1 = 13$



(a) $\left|\tilde{R}_{\mathcal{g}}^{(1)}(x)\right|_{x\in X} - \left|R_{\mathcal{g}_2}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}_{\mathcal{g}}^{(20)}(x)\right|_{x\in X} - \left|R_{\mathcal{g}_2}(x)\right|_{x\in X}$

Figure 10.10.: difference of SAFT with $|I_2| = 42$ and iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_2 = 23$



(a) $\left|\tilde{R}^{(1)}_{\mathcal{G}}(x)\right|_{x\in X} - \left|R_{\mathcal{G}_3}(x)\right|_{x\in X}$

(b) $\left|\tilde{R}^{(20)}_{\mathcal{G}}(x)\right|_{x\in X} - \left|R_{\mathcal{G}_3}(x)\right|_{x\in X}$

Figure 10.11.: difference of SAFT with $|I_3| = 32$ and iterative SAFT after 1 (a) and 20 (b) iterations for $\ell_3 = 23$

Bibliography

[ANMM93] Allgaier, M. W. ; Ness, S. ; McIntire, P. ; Moore, P. O.: *Visual and optical testing.* American Society for Nondestructive Testing, 1993

[AS13] Azimipanah, A. ; ShahbazPanahi, S.: Experimental results of compressive sensing based imaging in ultrasonic non-destructive testing. In: *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013 IEEE 5th International Workshop on* IEEE, 2013, pp. 336–339

[BHL+12] Bronstein, I. N. ; Hromkovic, J. ; Luderer, B. ; Schwarz, H.-R. ; Blath, J. ; Schied, A. ; Dempe, S. ; Wanka, G. ; Gottwald, S. ; Zeidler, E. et al.: *Taschenbuch der Mathematik.* Vol. 1. Springer-Verlag, 2012

[BKG+05] Banerjee, A. ; Krumpelman, C. ; Ghosh, J. ; Basu, S. ; Mooney, R. J.: Model-based Overlapping Clustering. In: *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining.* New York, NY, USA : ACM, 2005 (KDD '05), pp. 532–537

[Bli12] Blitz, J.: *Electrical and magnetic methods of non-destructive testing.* Vol. 3. Springer Science & Business Media, 2012

[Bol65] Bollobás, B.: On generalized graphs. In: *Acta Math. Acad. Sci. Hungar* 16 (1965), pp. 447–452

[Bos13] Bossmann, F.: *Model based image pattern recognition in ultrasonic non-destructive testing*, University of Göttingen, PhD thesis, 2013

[BVW10] Bourgain, J. ; Vu, V. H. ; Wood, P. M.: On the singularity probability of discrete random matrices. In: *Journal of Functional Analysis* 258 (2010), No. 2, pp. 559–603

[Caw01] Cawley, P.: Non-destructive testing—current capabilities and future directions. In: *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials Design and Applications* 215 (2001), No. 4, pp. 213–223

[CV08] Costello, K. P. ; Vu, V. H.: The rank of random graphs. In: *Random Structures Algorithms* 33 (2008), No. 3, pp. 269–285

[EFK05] Erdős, P. L. ; Füredi, Z. ; Katona, G.: Two-part and k-Sperner families: new proofs using permutations. In: *SIAM Journal on Discrete Mathematics* 19 (2005), No. 2, pp. 489–500

[Erd45] Erdős, P.: On a lemma of Littlewood and Offord. In: *Bull. Amer. Math. Soc.* 51 (1945), pp. 898–902

[Eva10] Evans, L. C.: *Graduate Studies in Mathematics*. Vol. 19: *Partial differential equations*. Second. American Mathematical Society, Providence, RI, 2010

[FS12] Foroozan, F. ; ShahbazPanahi, S.: MUSIC-based array imaging in multi-modal ultrasonic non-destructive testing. In: *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th* IEEE, 2012, pp. 529–532

[HAK+12] Houseman, E. A. ; Accomando, W. P. ; Koestler, D. C. ; Christensen, B C. ; Marsit, C. J. ; Nelson, H. H. ; Wiencke, J.K. ; Kelsey, K. T.: DNA methylation arrays as surrogate measures of cell mixture distribution. In: *BMC Bioinformatics* 13 (2012), No. 1, pp. 1–16

[Hal12] Halmshaw, R.: *Industrial radiology: theory and practice*. Vol. 1. Springer Science & Business Media, 2012

[Han11] Hansen, R. E.: *Introduction to synthetic aperture sonar*. INTECH Open Access Publisher, 2011

[HDW04] Holmes, C. ; Drinkwater, B. ; Wilcox, P.: The post-processing of ultrasonic array data using the total focusing method. In: *Insight-Non-Destructive Testing and Condition Monitoring* 46 (2004), No. 11, pp. 677–680

[HDW08] Hunter, A. J. ; Drinkwater, B. W. ; Wilcox, P. D.: The wavenumber algorithm for full-matrix imaging using an ultrasonic array. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 55 (2008), November, No. 11, pp. 2450–2462

[Hov80] Hovanessian, S. A.: Introduction to synthetic array and imaging radars. In: *Dedham, Mass., Artech House, Inc., 1980. 173 p.* 1 (1980)

[JC10] Jobst, M. ; Connolly, G. D.: Demonstration of the application of the total focusing method to the inspection of steel welds. In: *10th European Conference on Non-Destructive Testing*, 2010

[KB08] Kabán, A. ; Bingham, E.: Factorisation and denoising of 0–1 data: a variational approach. In: *Neurocomputing* 71 (2008), No. 10, pp. 2291–2308

[KK90] Krautkrämer, J. ; Krautkrämer, H.: *Ultrasonic Testing of Materials*. 4. Springer, 1990

[KKS95] Kahn, J. ; Komlós, J. ; Szemerédi, E.: On the probability that a random ±1 matrix is singular. In: *J. Amer. Math. Soc.* 8 (1995), pp. 223–240

[Kle13] Klenke, Achim: *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013

[KS87] Kanter, I. ; Sompolinsky, H.: Associative recall of memory without errors. In: *Phys. Rev. A* 35 (1987), Jan, pp. 380–392

[LBY+03] Liao, J. C. ; Boscolo, R. ; Yang, Y-L ; Tran, L. M. ; Sabatti, C. ; Roychowdhury, V. P.: Network component analysis: Reconstruction of regulatory signals in biological systems. In: *Proceedings of the National Academy of Sciences* 100 (2003), Dezember, No. 26, pp. 15522–15527

[LL70] Levine, E. ; Lubell, D.: Sperner collections on sets of real variables. In: *Ann. New York Acad. Sci.* 175 (1970), pp. 272–276

[LMK12] Langenberg, K.-J. ; Marklein, M. ; K., Mayer: *Ultrasonic Nondestructive Testing of Materials: Theoretical Foundations.* CRC Press, 2012

[LO43] Littlewood, J. E. ; Offord, A. C.: On the number of real roots of a random algebraic equation. III. In: *Rec. Math. [Mat. Sbornik] N.S.* 12(54) (1943), pp. 277–286

[LS99] Lee, D. D. ; Seung, H. S.: Learning the parts of objects by non-negative matrix factorization. In: *Nature* 401 (1999), No. 6755, pp. 788–791

[Lub66] Lubell, D.: A short proof of Sperner's lemma. In: *J. Combinatorial Theory* 1 (1966), pp. 299

[McM82] McMaster, R. C.: Nondestructive testing handbook. Volume 1-Leak testing. (1982)

[Meš63] Mešalkin, L. D.: A generalization of Sperner's theorem on the number of subsets of a finite set. In: *Teor. Verojatnost. i Primenen* 8 (1963), pp. 219–220

[MGNR06] Meeds, E. ; Ghahramani, Z. ; Neal, R. M. ; Roweis, S. T.: Modeling dyadic data with binary latent factors. In: *Advances in neural information processing systems*, 2006, pp. 977–984

[Mil68] Milner, E. C.: A combinatorial theorem on systems of sets. In: *J. London Math. Soc.* 43 (1968), pp. 204–206

[MMG$^+$08] Miettinen, P. ; Mielikäinen, T. ; Gionis, A. ; Das, G. ; Mannila, H.: The discrete basis problem. In: *IEEE Transactions on Knowledge and Data Engineering* 20 (2008), No. 10, pp. 1348–1362

[Odl88] Odlyzko, A.: On subspaces spanned by random selections of±1 vectors. In: *journal of combinatorial theory, Series A* 47 (1988), No. 1, pp. 124–133

[Pin09] Pinsky, M. A.: *Graduate Studies in Mathematics.* Vol. 102: *Introduction to Fourier analysis and wavelets.* American Mathematical Society, Providence, RI, 2009

[PT94] Paatero, P. ; Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. In: *Environmetrics* 5 (1994), No. 2, pp. 111–126

[SBK03] Segal, E. ; Battle, A.J. ; Koller, D.: Decomposing gene expression into cellular processes. In: *Proc. Pacific Symposium on Biocomputing (PSB)*, World Scientific, January 2003, pp. 89–100

[Set99] Sethian, J. A.: *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science.* Vol. 3. Cambridge university press, 1999

[Sey82] Seydel, J: Ultrasonic synthetic-aperture focusing techniques in NDT. In: *Research techniques in nondestructive testing.* 6 (1982), pp. 1–47

[SHL13] Slawski, Martin ; Hein, Matthias ; Lutsik, Pavlo: Matrix factorization with binary components. In: Burges, C. J. (Ed.) ; Bottou, L. (Ed.) ; Ghahramani, Z. (Ed.) ; Weinberger, K. Q. (Ed.): *NIPS*, 2013, pp. 3210–3218

[SKF^{+}12] Scharrer, T. ; Koch, A. ; Fendt, K. T. ; Rupitsch, S. J. ; Sutor, A. ; Ermert, H. ; Lerch, R.: Ultrasonic defect detection in multi-material, axis-symmetric devices with an improved synthetic aperture focusing technique (SAFT). In: *2012 IEEE International Ultrasonics Symposium* IEEE, 2012, pp. 1039–1042

[SNCO16] Schmitte, T. ; Nemitz, O. ; Chichkov, N. ; Orth, T.: Application of the Total Focusing Method for Improved Defect Characterization in the Production of Steel Tubes, Pipes and Plates. In: *World Conference on Non-Destructive Testing (WCNDT)*, NDT.net, 2016

[Spe28] Sperner, E.: Ein Satz über Untermengen einer endlichen Menge. In: *Math. Z.* 27 (1928), pp. 544–548

[Spi01] Spies, M.: Semi-analytical elastic wave-field modeling applied to arbitrarily oriented orthotropic media. In: *The Journal of the Acoustical Society of America* 110 (2001), No. 1, pp. 68–79

[SRD^{+}12] Spies, M. ; Rieder, H. ; Dillhöfer, A. ; Schmitz, V. ; Müller, W.: Synthetic aperture focusing and time-of-flight diffraction ultrasonic imaging—past and present. In: *Journal of Nondestructive Evaluation* 31 (2012), No. 4, pp. 310–323

[SSU03] Schein, A. I. ; Saul, L. K. ; Ungar, L. H.: A Generalized Linear Model for Principal Component Analysis of Binary Data. In: *AISTATS* Vol. 3, 2003, pp. 10

[TCX12] Tu, S. ; Chen, R. ; Xu, L.: Transcription Network Analysis by A Sparse Binary Factor Analysis Algorithm. In: *J. Integrative Bioinformatics* 9 (2012), No. 2

[Tho84] Thomson, R. N.: Transverse and longitudinal resolution of the synthetic aperture focusing technique. In: *Ultrasonics* 22 (1984), No. 1, pp. 9–15

[Tho96] Thomenius, K. E.: Evolution of ultrasound beamformers. In: *Ultrasonics Symposium, 1996. Proceedings., 1996 IEEE* Vol. 2 IEEE, 1996, pp. 1615–1622

[TV07] Tao, T. ; Vu, V. H.: On the singularity probability of random Bernoulli matrices. In: *J. Amer. Math. Soc.* 20 (2007), No. 3, pp. 603–628

[Vee97] Veen, A. J. d.: Analytical method for blind binary signal separation. In: *IEEE Transactions on Signal Processing* 45 (1997), Apr, No. 4, pp. 1078–1082

[Yam54] Yamamoto, K.: Logarithmic order of free distributive lattice. In: *J. Math. Soc. Japan* 6 (1954), pp. 343–353

[ZLDZ07] Zhang, Z. ; Li, T. ; Ding, C. ; Zhang, X.: Binary matrix factorization with applications. In: *Seventh IEEE International Conference on Data Mining (ICDM 2007)* IEEE, 2007, pp. 391–400

Curriculum vitae

David James, M.Sc.

Personal Details

Date of birth	20 June 1986
Place of birth	Fulda, Germany
Birth name	Aschenbrücker
Nationality	German

Academic Education

Since 06/2013	**Doctoral study programme: Mathematical Sciences** at the Georg-August-Universität Göttingen Advisor: Prof. Dr. F. Krahmer
04/2010–05/2013	**Masters study programme: Mathematics** at the Rheinische Friedrich-Wilhelms-Universität Bonn Master thesis: *Sparse Recovery with Random Convolutions* Advisor of master thesis: Prof. Dr. H. Rauhut Degree: **Master of Science**
09/2007–04/2010	**Bachelor study programme: Mathematics** at the Philipps-Universität Marburg Bachelor thesis: *Elliptische Kurven, endliche Körper und eine Anwendung auf die Verschlüsselung in WLAN-basierter Lokalisation* Advisor of bachelor thesis: Prof. Dr. T. Bauer Degree: **Bachelor of Science**
04/2007–09/2007	**Bachelor study programme: Mathematics** at the Johann Wolfgang Goethe-Universität Frankfurt am Main

Publications

James, David ; Rauhut, Holger: Nonuniform Sparse Recovery with Random Convolutions. In: *Sampling Theory and Applications (SampTA)* (2015), No. 1, pp. 34–38.

08/2014–09/2014 **parental leave**