CrossMark

Compressed Sensing for Quaternionic Signals

Narciso Gomes[2] · **Stefan Hartmann**[1] ·
Uwe Kähler[1]

Abstract We study the problem of compressed sensing for quaternionic Fourier matrices as arising in color representation of images. We will show that such matrices are allowing a sparse reconstruction by means of an l_1-minimization with high probability. Examples of sparse sampling of color images are provided.

1 Introduction

In the last decade a new paradigm has taken hold in signal and image processing: compressed sensing. The possibility of reconstructing a signal by only a few measurements under the condition that the representation in a given basis or frame is sparse has allowed to look at new methods and algorithms. Although sparsity constraints are directly connected only with non-convex optimization the uniqueness property shown by Candes et al. [10] allows the application of simple convex algorithms, such as linear programming. In parallel, during the last 15 years quaternion-valued functions have been used to represent color images, in particular RGB images [11]. Hereby, repre-

✉ Uwe Kähler
ukaehler@ua.pt

1 Universidade de Aveiro, Aveiro, Portugal

2 University of Cabo Verde, Praia, Cabo Verde

Birkhäuser

sentations using the discrete and continuous quaternionic Fourier transforms play a particular important role [4]. In this paper we will show that it is possible to combine both approaches, i.e., to use sparse sampling methods in the quaternionic representation of color images. This is a priori not so evident due to the non-commutative structure of the quaternions. For instance, it is not clear that quaternionic sampling matrices will fulfil the RIP condition as the traditional condition for compressed sensing. Therefore, we intend to go back to the origins of compressed sensing and follow the original approach by Rauhut [28] to show that quaternionic color images allow sparse reconstruction by means of an l_1-minimization with high probability. As the principal example of a quaternionic Fourier sampling matrix we base ourselves on the discrete version of the classic quaternionic Fourier transform by Sommen [31]. The reason is twofold. First of all, it represents all the important difficulties coming from the quaternionic structure so that it can easily adapted to other cases. Second, it also represents the quaternionic Fourier transform which is most used in applications, this being the main reason why it has been re-invented so many times. We will show that it is possible to follow Rauhut's ideas despite the fact that in the quaternionic case non-commutativity of the multiplication does not allow a direct application. Furthermore, we will give some examples of sparse sampling of images given as quaternionic signals. It shows that the practical application of these ideas in the field of hypercomplex analysis is possible. As a final remark we would like to point out that the ideas in the present paper can be also applied to the wider area of interpolation of quaternionic valued functions.

2 Preliminaries

In this section we recall some basic notions about quaternions and quaternionic exponentials which will be needed in the sequel.

The algebra of quaternions $\mathbb{H}$ is a four-dimensional real associative division algebra with unit 1 spanned by the elements $\{\mathbf{I}, \mathbf{J}, \mathbf{K}\}$ endowed with the relations

$$\mathbf{I}^2 = \mathbf{J}^2 = \mathbf{K}^2 = \mathbf{IJK} = -1, \quad \mathbf{IJ} = -\mathbf{JI} = \mathbf{K}.$$

This algebra is non-commutative. The real and imaginary parts of a given quaternion

$$q = x_0 1 + x_1 \mathbf{I} + x_2 \mathbf{J} + x_3 \mathbf{K}$$

are defined as $\mathrm{Re}(q) = q_0 := x_0$, and $\mathrm{Im}(q) := x_1 \mathbf{I} + x_2 \mathbf{J} + x_3 \mathbf{K}$. Therefore, we have the natural embeddings of the real numbers and of $\mathbb{R}^3$ into quaternions given by

$$x_0 \in \mathbb{R} \to x_0 1 \in \mathbb{H} \quad \text{and} \quad (x_1, x_2, x_3) \in \mathbb{R}^3 \to x_1 \mathbf{I} + x_2 \mathbf{J} + x_3 \mathbf{K} \in \mathbb{H}.$$

This leads to the identifications

$$\mathbb{H} \equiv \mathbb{R}^4, \quad \mathrm{Im}\,\mathbb{H} \equiv \mathbb{R}^3, \quad \mathrm{Re}\,\mathbb{H} \equiv \mathbb{R},$$

where $\mathrm{Im}\,\mathbb{H}$ is the three dimensional space of pure imaginary quaternions, and hence, $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}^3$.

The *conjugation* on $\mathbb{H}$ is an automorphism of $\mathbb{H}$ onto itself given by

$$q = x_0 + \operatorname{Im} q \quad \rightarrow \quad \overline{q} = x_0 - \operatorname{Im} q,$$

together with the involution property

$$\overline{qp} = \overline{p}\,\overline{q}, \quad \forall p, q \in \mathbb{H}.$$

A purely imaginary quaternion with absolute value 1 is called an imaginary unit. We denote the set of all imaginary units by $\mathbb{S}^2$, that is,

$$\mathbb{S}^2 = \left\{ x_1\mathbf{I} + x_2\mathbf{J} + x_3\mathbf{K} \in \mathbb{H} : \sum_{i=1}^{3} x_i^2 = 1 \right\}.$$

2.1 Inner and Outer Products

Given two quaternions $q, p \in \mathbb{H}$ its quaternionic multiplication can be expressed in terms of the usual scalar and vector products on $\operatorname{Im} \mathbb{H} \sim \mathbb{R}^3$ by

$$q\overline{p} = (q_0 + \operatorname{Im} q)(p_0 - \operatorname{Im} p) = \underbrace{q_0 p_0 + \operatorname{Im} q \cdot \operatorname{Im} p}_{\in \mathbb{R}}$$
$$+ \underbrace{(-q_0 \operatorname{Im} p + p_0 \operatorname{Im} q - \operatorname{Im} q \times \operatorname{Im} p)}_{\in \operatorname{Im} \mathbb{H}}.$$

Moreover, we have

$$\langle q, p \rangle = \operatorname{Re}(q\overline{p}),$$

where $\langle \cdot, \cdot \rangle$ corresponds to the Euclidean scalar product defined on $\mathbb{H} \sim \mathbb{R}^4$, and the induced norm

$$\|q\|^2 = \operatorname{Re}(q\overline{q}) = \langle q, q \rangle$$

which satisfies the product rule $\|qp\| = \|q\|\,\|p\|$. Also, for every non-zero quaternion q it holds

$$q^{-1} = \frac{\overline{q}}{\|q\|^2}.$$

Furthermore, we can introduce the signum of a quaternion q defined via

$$\operatorname{sgn} q = \begin{cases} \frac{q}{\|q\|}, & q \neq 0 \\ 0, & q = 0. \end{cases}$$

Note that $\operatorname{sgn} q \in \mathbb{H}$. For $\operatorname{sgn} q$ we have the following properties.

Lemma 2.1 *Let $q = q_0 + q_1\mathbf{I} + q_2\mathbf{J} + q_3\mathbf{IJ}$ and $p = p_0 + p_1\mathbf{I} + p_2\mathbf{J} + p_3\mathbf{IJ}$ be two quaternions. It holds*

1. $\|\operatorname{sgn} p\| = 1,\ p \neq 0$;
2. $\overline{\operatorname{sgn} p}\, \operatorname{sgn} p = 1,\ p \neq 0$;
3. $p\, \overline{\operatorname{sgn} p} = \|p\|$;
4. $\|q\|\, \|\operatorname{sgn} p\| = \|q \operatorname{sgn} p\|$.

The proof is rather straightforward. For details we refer to [19].

For an invertible quaternion we have the following representation.

Theorem 2.2 *Let q be an invertible quaternion. Then,*

$$q = \|q\| \left[\cos(\theta) + \omega(q)\sin(\theta)\right],$$

where $\omega(q) = \frac{\operatorname{Im} q}{\|\operatorname{Im} q\|} \in \mathbb{S}^2 \subset \operatorname{Im} \mathbb{H}$ and $\theta \in [0, \pi]$ is such that $\cos\theta = \frac{q_0}{\|q\|}$ and $\sin\theta = \frac{\|\operatorname{Im} q\|}{\|q\|}$.

This leads to the following definition of a quaternionic exponential (see, e.g. [20]):

Definition 2.3 Given a purely imaginary quaternion $\omega \in \mathbb{S}^2 \subset \operatorname{Im} \mathbb{H}$, we define the quaternionic exponential $e^{\omega\theta}$ as

$$\theta \in \mathbb{R} \mapsto e^{\omega\theta} := \cos(\theta) + \omega\sin(\theta) \in \mathbb{H}. \tag{2.1}$$

We list some of its properties.

- $\|e^{\omega\theta}\| = 1$;
- $\|e^{\omega\theta}\| \leq e^{|\theta|}$;
- given $\omega_1, \omega_2 \in \mathbb{S}^2$, we have

$$e^{\omega_1\theta_1} e^{\omega_2\theta_2} = e^{\omega_1\theta_1 + \omega_2\theta_2}, \quad \forall \theta_1, \theta_2 \in [0, 2\pi],$$

only if $\omega_1\omega_2 = \omega_2\omega_1$.

The last property shows the non-commutative character of quaternionic exponentials and presents one of the major problems in the application of the classic approach by Rauhut [28].

3 Sparse Sampling of Quaternionic Signals

3.1 The Setting

While there is quite a variety of possibilities to consider concrete quaternionic Fourier transforms [29] we will consider here decompositions with respect to the atoms $\Psi_{k,l}(x, y) = e^{\mathbf{I}kx} e^{\mathbf{J}ly}$, $(k, l) \in \mathbb{Z}^2$. On the one side this corresponds to the original definition of a quaternionic Fourier transform by Sommen in 1982 [31] which has been reinvented several times since then. On the other side this quaternionic Fourier

transform is also the principal Fourier transform to be used in applications in image processing (c.f. [6]). Furthermore, this example can also be easily adapted to the other cases. Let us denote by $\prod_\rho$ the space of all quaternionic trigonometric polynomials of maximal order $\rho \in \mathbb{N}_0$ in two real variables. The dimension of $\prod_\rho$ is $d := (2\rho + 1)^2$. Therefore, an element $q \in \prod_\rho$ is of the form

$$q(x, y) = \sum_{(k,l)\in[-\rho,\rho]^2\cap\mathbb{Z}^2} e^{\mathbf{I}kx} e^{\mathbf{J}ly} c_{k,l}, \quad (x, y) \in [0, 2\pi]^2, \ c_{k,l} \in \mathbb{H}. \quad (3.1)$$

We want to determine the coefficients $c_{k,l}$ from a few given samples. To this end we consider the sequence of coefficients $c := (c_{k,l})_{k,l}$ as a vector (strictly speaking a tensor) such that c is supported on a set T which has a much smaller cardinality than the dimension of $\prod_\rho$. That is to say, the finite combination in (3.1) is "sparse". However, we do not require any information on T except its maximum size. Therefore, we introduce the set (not a linear space) $\prod_\rho(M) \subset \prod_\rho$ of all polynomials of type (3.1) such that the sequence of their coefficients c has support on a set $T \subset [-\rho, \rho]^2 \cap \mathbb{Z}^2$ with $|T| \leq M$, i.e., $q \in \prod_\rho(M)$ is of the form

$$q(x, y) = \sum_{(k,l)\in T\subset[-\rho,\rho]^2\cap\mathbb{Z}^2} e^{\mathbf{I}kx} e^{\mathbf{J}ly} c_{k,l}.$$

Furthermore, we consider a given sampling set $X := \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, x_N)\}$ as a set of independent random variables having uniform distribution on $[0, 2\pi]^2$. Thus, the main objective is to reconstruct $q \in \prod_\rho(M)$ from the samples $f(x_i, y_i)$ at those N randomly chosen points.

The standard way of determining the coefficients $c_{k,l}$ would be by applying a greedy algorithm, such as ℓ_0-minimization or matching pursuit. It basically corresponds to searching for the atom with the largest coefficient in each step [36]. Unfortunately, such an algorithm is NP-hard with exponentially growing computational costs. Therefore, we would like to use a different algorithm, such as basis pursuit. This consists in the following non-linear method of reconstructing $q \in \prod_\rho(M)$ from its sampled values $(q(x_1, y_1), \ldots, q(x_N, y_N))$. We minimize the ℓ_1-norm of the Fourier coefficients $c_{k,l}$,

$$||(c_{k,l})||_1 := \sum_{(k,l)\in[-\rho,\rho]^2\cap\mathbb{Z}^2} |c_{k,l}|,$$

under the constraint that the corresponding trigonometric polynomial matches q on the sampling points. In other words we solve the problem

$$||(c_{k,l})||_1 \to \min \quad \text{s.t.} \quad \sum_{(k,l)\in[-\rho,\rho]^2\cap\mathbb{Z}^2} e^{\mathbf{I}kx_i} e^{\mathbf{J}\tilde{k}y_i} c_{k,l} = q(x_i, y_i), \quad i = 1, \ldots, N.$$

$$(3.2)$$

We can directly point out that *basis pursuit* [12] can be performed with efficient convex optimization techniques [5] via linear programming. This makes it a much faster

algorithm than the simple ℓ_0-minimization. But, of course, the principal question is if we indeed can replace the ℓ_0-minimization by an ℓ_1-minimization.

3.2 Equivalence Between ℓ_0-minimization and Basis Pursuit

We would like to use *basis pursuit* [12] to reconstruct f completely by determining all coefficients $c_{k,l}$, $(k, l) \in [-\rho, \rho]^2 \cap \mathbb{Z}^2$. Unfortunately, a condition like the restrictive isometry property (RIP [1]) which is valid in all cases (including the worst) is too restrictive in our case. Therefore, we are going to give a probabilistic answer. The theorems below are analogues to the theorems given in Candes et al. [10] and in Rauhut [28].

Theorem 3.1 *Assume $f \in \prod_\rho(M)$ with some sparsity $M \in \mathbb{N}$. Let $(x_1, y_1), \ldots,$ $(x_N, y_N) \in [0, 2\pi]^2$ be independent random variables having uniform distribution on $[0, 2\pi]^2$. Choose $n \in \mathbb{N}$, $\beta > 0$, $\kappa > 0$ and $K_1, \ldots, K_n \in \mathbb{N}$ such that*

$$a := \sum_{m=1}^{n} \beta^{n/K_m} < 1 \quad \text{and} \quad \frac{\kappa}{1 - \kappa} \leq \frac{1 - a}{1 + a} M^{-3/2}. \tag{3.3}$$

Set $\theta := N/M$. Then with probability at least

$$1 - \left(d\beta^{-2n} \sum_{m=1}^{n} G_{2mK_m}(\theta) + \kappa^{-2} M G_{2n}(\theta) \right), \tag{3.4}$$

where $G_n(\theta) = \theta^{-n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} S_2(n, k)\theta^k$ and $S_2(n, k)$ denote the Stirling numbers of the second kind, f can be reconstructed exactly from its sampled values $f(x_1, y_1), \ldots, f(x_N, y_N)$ by solving the ℓ_1-minimization problem

$$\min \left\| (c_{k,l}) \right\|_1 := \min \sum_{(k,l) \in [-\rho, \rho]^2 \cap \mathbb{Z}^2} |c_{k,l}|,$$

$$s.t. \quad f(x_i, y_i) := \sum_{(k,l) \in [-\rho, \rho]^2 \cap \mathbb{Z}^2} e^{\mathbf{I}kx_i} e^{\mathbf{J}ly_i} c_{k,l}, \quad i = 1, \ldots, N. \tag{3.5}$$

While the above theorem provides exact constants we can give a version of the theorem which is somewhat easier to apply.

Theorem 3.2 *There exists an absolute constant $C > 0$ such that the following is true. Assume $f \in \prod_\rho(M)$ for some sparsity $M \in \mathbb{N}$. Let $(x_1, y_1), \ldots, (x_N, y_N) \in [0, 2\pi]^2$ be independent random variables having the uniform distribution on $[0, 2\pi]^2$. If for some $\epsilon > 0$ it holds*

$$N \geq CM \log(d/\epsilon) \tag{3.6}$$

then with probability at least $1 - \epsilon$ the function f can be recovered from its sampled values $f(x_i, y_i)$, $i = 1, \ldots, N$, by solving the minimization problem (3.5).

For our third theorem about reconstructing a sparse trigonometric polynomial from random samples we need some additional explanations. For convenience, we will use the notation $\prod_T$ as the set of all trigonometric polynomials whose coefficients are supported on T and we model the set $T \subset [-\rho, \rho]^2 \cap \mathbb{Z}^2$ of non-vanishing Fourier coefficients as random. Here we will treat the "generic" case and not arbitrary sparse polynomials. We expect to get better estimates for the probability of exact reconstruction.

Consider $0 < \tau < 1$. The probability that an index $(k, l) \in [-\rho, \rho]^2 \cap \mathbb{Z}^2$ belongs to T is assumed to be

$$\mathbb{P}((k, l) \in T) = \tau \tag{3.7}$$

independently of each (k, l). The choice of the sampling set X is stochastically independent. Furthermore, we also assume the same for the choice of T. So the length of T, i.e., $|T|$, has a binomial distribution and the expected size of $|T|$ is, obviously, $\mathbb{E}|T| = \tau d = \tau(2\rho + 1)^2$.

We will define now two auxiliary notations to announce the third theorem. For $n \in \mathbb{N}$ we define

$$W(n, N, \mathbb{E}|T|, d) := N^{-2n} \sum_{t=1}^{\min\{n, N\}} \frac{N!}{(N-t)!} \sum_{s=2}^{2n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} Q(2n, t, s, R) d^{-R} \tag{3.8}$$

and for $K, m \in \mathbb{N}$ we define

$$Z(K, m, N, \mathbb{E}|T|, d)$$
$$:= N^{-2Km} \sum_{t=1}^{\min\{Km, N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2Km} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}} Q^*(2K, m, t, s, R) d^{-R}. \tag{3.9}$$

By using Basis Pursuit our theorem is given as follows.

Theorem 3.3 *Let $(x_1, y_1), \ldots, (x_N, y_N) \in [0, 2\pi]^2$ be independent random variables having the uniform distribution on $[0, 2\pi]^2$. Further assume that $T := T_1 \times T_2$ (which is an independent set of $(x_{i_1}, y_{i_1}), \ldots, (x_{i_N}, y_{i_N})$) is a random subset of modeled by $\mathbb{P}((k, l) \in T) = \tau$ such that $\mathbb{E}|T| = \tau d \geq 1$. Choose $n \in \mathbb{N}$, $\alpha, \beta > 0$ and $K_1, \ldots, K_n \in \mathbb{N}$ such that*

$$a := \sum_{m=1}^{n} \beta^{n/K_m} < 1 \quad and \quad \frac{k}{1-k} \leq \frac{1-a}{1+a}((\alpha + 1)\mathbb{E}|T|)^{-3/2}. \tag{3.10}$$

Then with probability at least

$$1 - \left(\kappa^{-2} W(n, N, \mathbb{E}|T|, d) + \beta^{-2n} d \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, d) + \exp\left(-\frac{3\alpha^2}{6 + 2\alpha} \mathbb{E}|T|\right) \right) \tag{3.11}$$

any $f \in \prod_T \subset \prod_\rho(|T|)$ *can be reconstructed exactly from its sample values* $f(x_1, y_1), \dots, f(x_N, y_N)$ *by solving the minimization problem* (3.5)*.*

In the next section we will give the proof of these theorems.

4 Proof of the Main Theorems

First of all, we need to introduce some auxiliary notations. We abbreviate $\mathbb{Z}_\rho^2 = [-\rho, \rho]^2 \cap \mathbb{Z}^2$ and denote by $\ell_2(\mathbb{Z}_\rho^2), \ell_2(T), \ell_2(X)$ the ℓ_2-spaces of sequences indexed by our discrete sets $\mathbb{Z}_\rho^2$, T, and X, respectively. We also need the sampling operator $\mathcal{F}_X : \ell_2(\mathbb{Z}_\rho^2) \to \ell_2(X)$ given by

$$\left(\mathcal{F}_X c_{k,l}\right)(x, y) := \sum_{(k,l)\in\mathbb{Z}_\rho^2} \left(e^{\mathbf{I}kx} e^{\mathbf{J}ly} c_{k,l}\right), \quad (x, y) \in X.$$

By $\mathcal{F}_{TX}$ we denote its restriction to sequences with support only on T. Hence, $\mathcal{F}_{TX}$ is an operator acting from $\ell_2(T)$ in $\ell_2(X)$. Also, we have to consider their adjoint operators, $\mathcal{F}_X^* : \ell_2(X) \to \ell_2(\mathbb{Z}_\rho^2)$ and $\mathcal{F}_{TX}^* : \ell_2(X) \to \ell_2(T)$.

The next lemma is the fundamental lemma on which our subsequent investigations will be based.

Lemma 4.1 *Let $c \in \ell_2(\mathbb{Z}_\rho^2)$ and $T := \operatorname{supp} c$. Furthermore, let us assume that $\mathcal{F}_{TX} : \ell_2(T) \to \ell_2(X)$ is injective and suppose that there exists a $P \in \ell_2(\mathbb{Z}_\rho^2)$ with the following properties:*

(i) $P_{k,l} = \operatorname{sgn} \left(c_{k,l}\right)_{k,l}$ *for all $(k, l) \in T$,*
(ii) $\left|P_{k,l}\right| < 1$ *for all $(k, l) \notin T$,*
(iii) *there exists a $\lambda \in \ell_2(X)$ such that $P = \mathcal{F}_X^*\lambda$.*

Then c is the unique minimizer to the problem (3.5)*.*

This lemma is the fundamental lemma for [10,28]. Since unlike these cases we are working here with quaternion-valued vectors, i.e., a non-commutative structure, we are going to present a proof for our case.

Proof Let us assume $X \neq \emptyset$ and $c \neq 0$ to exclude the trivial cases. Furthermore, let us suppose that the vector P exist. Let r be any vector different to c with $\mathcal{F}_X r = \mathcal{F}_X c$. Consider $q := r - c$, then $\mathcal{F}_X q$ vanishes on X. This means that for $r_{k,l}, (k, l) \in T$, we have the following estimate

$$\begin{aligned}
|r_{k,l}| = |c_{k,l} + q_{k,l}| &= |(c_{k,l} + q_{k,l})\overline{\operatorname{sgn} c_{k,l}} \operatorname{sgn} c_{k,l}| \\
&= \left|\left(c_{k,l} \,\overline{\operatorname{sgn} c_{k,l}} + q_{k,l} \,\overline{\operatorname{sgn} c_{k,l}}\right) \operatorname{sgn} c_{k,l}\right| \\
&= |\,|c_{k,l}| + q_{k,l} \,\overline{\operatorname{sgn} c_{k,l}}||\overline{\operatorname{sgn} c_{k,l}}| \\
&= |\,|c_{k,l}| + q_{k,l} \,\overline{\operatorname{sgn} c_{k,l}}| \geq |c_{k,l}| + \operatorname{Re}\left(q_{k,l} \,\overline{\operatorname{sgn} c_{k,l}}\right) \\
&= |c_{k,l}| + \operatorname{Re}\left(q_{k,l} \,\overline{P_{k,l}}\right).
\end{aligned}$$

Thus, for any $(k, l) \in T$ we have $|c_{k,l}| + \mathrm{Re}\left(q_{k,l}\,\overline{P_{k,l}}\right) \leq |r_{k,l}|$. Otherwise, for $(k, l) \notin T$ we have $\mathrm{Re}\left(q_{k,l}\,\overline{P_{k,l}}\right) \leq |q_{k,l}| = |r_{k,l}|$ since $|P_{k,l}| < 1$. Thus

$$||r||_1 \geq ||p||_1 + \sum_{(k,l)\in[-\rho,\rho]^2\cap\mathbb{Z}^2} \mathrm{Re}\left(q_{k,l}\,\overline{P_{k,l}}\right).$$

Now, from condition (iii) we get

$$\sum_{(k,l)\in\mathbb{Z}^2_\rho} \mathrm{Re}\left(q_{k,l}\,\overline{P_{k,l}}\right) = \mathrm{Re}\left(\sum_{(k,l)\in\mathbb{Z}^2_\rho} q_{k,l}\,\overline{\left(\mathcal{F}_X^*\lambda\right)_{k,l}}\right)$$

$$= \mathrm{Re}\left(\sum_{i=1}^{N}(\mathcal{F}_X q)(x_i, y_i)\overline{\lambda(x_i, y_i)}\right) = 0 \qquad (4.1)$$

whereas $\mathcal{F}_X q$ vanishes. Thus, $\mathcal{F}_X P$ is supported on X and $\mathcal{F}_X q$ vanishes on X. Therefore, $||r||_{\ell_1} \geq ||c||_{\ell_1}$. The equality holds when $||q_{k,l}|| = \mathrm{Re}\left(q_{k,l}\,\overline{P_{k,l}}\right)$ for all $(k, l) \notin T$. Since $||P_{k,l}|| < 1$, this forces q to vanish outside of T. Taking in account the injectivity of $\mathcal{F}_{TX}$ we have that since $\mathcal{F}_X q$ vanishes on X, q vanishes identically and we get $r = c$. Thus, this shows that c is unique minimizer $c^\sharp$ to the problem (3.5).

□

For the invertibility we can state the following obvious lemma.

Lemma 4.2 *If $N \geq |T|$ then $\mathcal{F}_{TX}$ is injective almost surely.*

We need to show now that with high probability there exists a P with the required properties so that we can apply Lemma 4.1. To show this we proceed as in [28] but with the necessary modifications for the quaternionic case.

We introduce the restriction operator $R_T : \ell_2(\mathbb{Z}^2_\rho) \to \ell_2(T)$, $R_T c_{k,l} = c_{k,l}$ for $(k, l) \in T$. Its adjoint $R_T^* = E_T : \ell_2(T) \to \ell_2(\mathbb{Z}^2_\rho)$ is the operator that extends a vector outside T by zero, i.e., $(E_T d)_{k,l} = d_{k,l}$ for $(k, l) \in T$ and $(E_T d)_{k,l} = 0$ otherwise.

Now, for the moment let us assume that $\mathcal{F}_{TX}^*\mathcal{F}_{TX} : \ell_2(T) \to \ell_2(T)$ is invertible. This is true almost surely under the condition of Lemma 4.2 since $\mathcal{F}_{TX}$ is then injective. In this case we can represent P explicitly in the form

$$P := \mathcal{F}_X^*\mathcal{F}_{TX}(\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}R_T\,\mathrm{sgn}(c),$$

It is easy to check that P fulfills property (i) and property (iii) in Lemma 4.1 with

$$\lambda := \mathcal{F}_{TX}(\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}R_T\,\mathrm{sgn}\left(c_{k,l}\right)_{k,l} \in \ell_2(X).$$

This means that we have to show that P fulfills (ii) of Lemma 4.1 with high probability.

To this end we represent P in terms of the following operators

$$H := \ell_2(T) \to \ell_2(\mathbb{Z}^2_\rho), \quad H_0 := \ell_2(T) \to \ell_2(T),$$

where

$$H := NE_T - \mathcal{F}_{TX}^* \mathcal{F}_{TX} \quad H_0 := NI_T - \mathcal{F}_{TX}^* \mathcal{F}_{TX},$$

and I_T denotes the identity operator on $\ell_2(T)$. Furthermore, the matrix H_0 is self-adjoint, and H acts on c as

$$(Hc)_{k,l} = -\sum_{i=1}^{N} \sum_{\substack{(\tilde{k}, \tilde{l}) \in T \\ (\tilde{k}, \tilde{l}) \neq (k,l)}} e^{-\mathbf{J}ly_i} e^{\mathbf{I}(\tilde{k}-k)x_i} e^{\mathbf{J}\tilde{l}y_i} c_{\tilde{k}, \tilde{l}}, \tag{4.2}$$

By using H and H_0 we can rewrite P as

$$P = \left[(NE_T - H)(NI_T - H_0)^{-1} \right] R_T \operatorname{sgn}(c). \tag{4.3}$$

After joining the two terms into the last representation of P the result (4.3) looks very similar to the one of Rauhut [28], so we just adapt his approach below. Taking into account that we are interested in property (ii) from Lemma 4.1 we only need the values of P on $T^c = \mathbb{Z}_\rho^2 \backslash T$. Since $R_{T^c} E_T = 0$ we can write these components of P as

$$P_{k,\ell} = \left(-N^{-1} R_{T^c} H \left(I_T - N^{-1} H_0 \right)^{-1} R_T \operatorname{sgn}(c) \right)_{k,\ell} \quad \text{for all} \quad (k, \ell) \in T^c. \tag{4.4}$$

Let us take a closer look at the operator $\left(I_T - N^{-1} H_0 \right)^{-1}$. We would like to apply the von Neumann series, but that would require $\|N^{-1} H_0\| < 1$ which we cannot ensure. Therefore, let us assume that we can ensure $\|(N^{-1} H_0)^n\| < 1$ for some $n \in \mathbb{N}$. Then we get

$$\left(I_T - \left(N^{-1} H_0 \right)^n \right)^{-1} = I_T + A_n$$

with

$$A_n := \sum_{r=1}^{\infty} \left(N^{-1} H_0 \right)^{rn} \tag{4.5}$$

and, consequently,

$$\left(I_T - N^{-1} H_0 \right)^{-1} = (I_T + A_n) \sum_{m=0}^{n-1} \left(N^{-1} H_0 \right)^m.$$

This means that on the complement of T we have

$$R_{T^c} P = -N^{-1} H (I_T + A_n) \left(\sum_{m=0}^{n-1} \left(N^{-1} H_0 \right)^m \right) R_T \operatorname{sgn}(c) = - \left(P^{(1)} + P^{(2)} \right),$$

where

$$P^{(1)} = S_n \text{sgn}(c) \quad \text{and} \quad P^{(2)} = \frac{1}{N} H A_n R_T (I + S_{n-1}) \text{sgn}(c),$$

with $S_n := \sum_{m=0}^{n-1} (N^{-1} H R_T)^m$.

Since we would like to know when property (ii) fails we need to estimate the probability $\mathbb{P}(\sup_{(k,l) \in T^c} |P_{k,l}| \geq 1)$ by estimating $P^{(1)}$ and $P^{(2)}$ individually. To make this splitting we consider two arbitrary numbers $a_1, a_2 > 0$ satisfying $a_1 + a_2 = 1$. Then we have

$$\mathbb{P}\big(\sup_{(k,l) \in T^c} |P_{k,l}| \geq 1\big) \leq \mathbb{P}\left(\{\sup_{(k,l) \in T^c} |P_{k,l}^{(1)}| \geq a_1\} \cup \{\sup_{(k,l) \in T^c} |P_{k,l}^{(2)}| \geq a_2\}\right).$$
(4.6)

Thus, the first term can be estimated by

$$\mathbb{P}\big(|P_{k,l}^{(1)}| \geq a_1\big) = \mathbb{P}\big(|(S_n \text{sgn}(c))_{k,l}| \geq a_1\big)$$
$$\leq \mathbb{P}\left(\sum_{m=1}^{n} |((N^{-1} H R_T)^m \text{sgn}(c))_{k,l}| \geq a_1\right) =: \mathbb{P}(E_{k,l}). \quad (4.7)$$

while for the second term we have

$$\sup_{(k,\ell) \in T^c} |P_{k,l}^{(2)}| \leq \left\| P^{(2)} \right\|_\infty$$
$$\leq \left\| N^{-1} H A_n \right\|_{\ell^\infty(T) \to \ell^\infty(\Lambda)} (1 + \| R_T S_{n-1} \text{sgn}(c) \|_{\ell^\infty(T)}), \quad (4.8)$$

where $\Lambda := \ell^\infty(\mathbb{Z}_\rho^2)$ denotes the space of sequences indexed by $\mathbb{Z}_\rho^2$ with the supremum norm.

Let us first analyze the term $\| R_T S_{n-1} \text{sgn}(c) \|_{\ell^\infty(T)}$. This term is the same as in (4.7) and we obtain

$$\mathbb{P}\left(|(S_{n-1} \text{sgn}(c))_{k,\ell}| \geq a_1\right) \leq \mathbb{P}\left(\sum_{m=1}^{n} |((N^{-1} H R_T)^m \text{sgn}(c))_{k,\ell}| \geq a_1\right) = \mathbb{P}(E_{k,\ell})$$

Now, let us analyze the norm of the operator in (4.8). By definition we know that $\|A_n\|_\infty = \sup_r \sum_s |(A_n)_{r,s}|$ seen as a matrix operator. We also know that

$$\left\| N^{-1} H A_n \right\|_\infty \leq \left\| N^{-1} H \right\|_\infty \| A_n \|_\infty. \quad (4.9)$$

with $A_n = \sum_{r=1}^{\infty} \left(N^{-1} H_0 \right)^{rn}$.

The term $\|A_n\|_\infty$ can be estimated using the Frobenius norm. For the Frobenius norm we have $\|A\|_F^2 := Tr(AA^*) = \sum_{r,s} |A_{r,s}|^2$, where $Tr(AA^*)$ denotes the trace of AA^*. Let us assume that

$$\left\|\left(N^{-1}H_0\right)^n\right\|_F \leq \kappa < 1. \tag{4.10}$$

From the definition (4.5) of A_n, it follows that

$$\|A_n\|_F = \left\|\sum_{r=1}^\infty \left(N^{-1}H_0\right)^{rn}\right\|_F \leq \sum_{r=1}^\infty \left\|(N^{-1}H_0)^n\right\|_F^r \leq \sum_{r=1}^\infty \kappa^r = \frac{\kappa}{1-\kappa}.$$

Furthermore, since A_n has $|T|$ columns by Cauchy–Schwarz inequality we get

$$\|A_n\|_\infty^2 \leq \sup_i \sum_j |(A_n)_{i,j}|^2 \leq |T| \, \|A_n\|_F^2. \tag{4.11}$$

Thus, from (4.10) and $\|S_{n-1}\text{sgn}(c)\|_\infty < a_1$ it follows

$$\sup_{(k,\ell)\in T^c} |P_{k,l}^{(2)}| \leq (1 + a_1)\frac{\kappa}{1-\kappa}\, |T|^{\frac{3}{2}}. \tag{4.12}$$

This leads to the condition

$$\frac{\kappa}{1-\kappa} \leq \frac{a_2}{1+a_1}\, |T|^{-\frac{3}{2}} \tag{4.13}$$

to ensure $\sup_{(k,l)\in T^c} |P_{k,l}^{(2)}| \leq a_2$ as intended.

Also it follows from (4.13) that $\kappa < 1$ and $|T| \geq 1$ (we can exclude the case $T = \emptyset$ since it corresponds to the trivial case of $f = 0$ where ℓ_1-minimization will obviously recover f).

Now, since in Theorem 3.1 $|T|$ is deterministic and in Theorem 3.3 $|T|$ is a random variable we have to proceed differently in each case. In the case where $|T|$ is deterministic we have

$$\mathbb{P}\big(\sup_{(k,\ell)\in T^c} |P_{k,\ell}| \geq 1\big) \leq \sum_{(k,\ell)\in[-\rho,\rho]^2\cap\mathbb{Z}^2} \mathbb{P}(E_{k,\ell}) + \mathbb{P}\left(\left\|(N^{-1}H_0)^n\right\|_F \geq \kappa\right). \tag{4.14}$$

where we use the condition in (4.13).

Consider the second case, where $|T|$ is random by Theorem 3.3. If we assume

$$|T| \leq (\alpha + 1)\mathbb{E}|T|$$

with $\alpha > 0$ and also assume

$$\frac{\kappa}{1-\kappa} \leq \frac{a_2}{1+a_1}((\alpha+1)\mathbb{E}|T|)^{-3/2} \tag{4.15}$$

then clearly (4.13) is satisfied and consequently

$$\sup_{t \in T^c} \left| P_{k,l}^{(2)} \right| \le a_2.$$

This means that we get from (4.6) the following estimate

$$\mathbb{P}\big(\sup_{(k,l) \in T^c} |P_{k,l}| \ge 1\big) \le \mathbb{P}\Bigg(\bigcup_{(k,l) \in T^c} \left\{ \left| P_{k,l}^{(1)} \right| \ge a_1 \right\} \cup \left\{ \|R_T \mathrm{sgn}(c)\|_\infty \ge a_1 \right\}$$

$$\cup \left\{ \left\| (N^{-1} H_0)^n \right\|_F \ge \kappa \right\} \cup \left\{ |T| \ge (\alpha + 1) \mathbb{E}\,|T| \right\} \Bigg)$$

$$\le \mathbb{P}\Bigg(\bigcup_{(k,l) \in [-\rho,\rho]^2 \cap \mathbb{Z}^2} E_{k,l} \cup \left\{ \left\| (N^{-1} H_0)^n \right\|_F \ge \kappa \right\}$$

$$\cup \{ |T| \ge (\alpha + 1) \mathbb{E}\,|T| \} \Bigg)$$

$$\le \sum_{(k,l) \in [-\rho,\rho]^2 \cap \mathbb{Z}^2} \mathbb{P}(E_{k,l}) + \mathbb{P}\left(\left\| (N^{-1} H_0)^n \right\|_F \ge \kappa \right)$$

$$+ \mathbb{P}\left(|T| \ge (\alpha + 1) \mathbb{E}\,|T| \right). \tag{4.16}$$

Note that $|T|$ is the sum of independent random variables. For the third term of (4.16), we obtain

$$\mathbb{P}\big(|T| \ge \alpha \mathbb{E}\,|T| + \mathbb{E}\,|T| \big) \le \exp\left(-3(\alpha \mathbb{E}\,|T|)^2 / (6 \mathbb{E}\,|T| + 2(\alpha \mathbb{E}\,|T|)) \right)$$

$$= \exp\left(-\frac{3\alpha^2}{6 + 2\alpha} \mathbb{E}\,|T| \right). \tag{4.17}$$

Therefore, in the next two sections we have to estimate the quantities $\mathbb{P}(E_{k,\ell})$ and $\mathbb{P}(\|(N^{-1} H_0)^n\|_F \ge \kappa)$, respectively.

4.1 Analysis of the Powers of H_0

The objective of this section is to estimate the second term in (4.16) and (4.14). To this end we need to estimate Frobenius norm of our random matrix H_0. For this propose we will estimate $\left\| H_0^n \right\|_F^2$ by means of Markov's inequality. We will start with some auxiliary lemmas which will be needed for the proof of Lemma 4.5 in which we will determine the expectation value towards the random sampling set $X = \{(x_1, y_1), \ldots, (x_N, y_N)\}$.

Lemma 4.3 *Consider $\theta_r \in \mathbb{R}$ and $\underline{a} \in \{-1, 1\}$. For*

$$\lambda(a_r, \theta_r) = \begin{cases} \cos \theta_r, & a_r = +1 \\ \sin \theta_r, & a_r = -1, \end{cases}$$

it holds

$$2^{n-1} \prod_{r=1}^{n} \lambda(a_r, \theta_r) = \sum_{\alpha_2, \ldots, \alpha_n \in \{-1,1\}} \alpha_2^{\delta_2} \ldots \alpha_n^{\delta_n} \lambda \left(\prod_{j=1}^{n} a_j, \theta_1 + \alpha_2 \theta_2 + \cdots + \alpha_n \theta_n \right)$$

such that $\delta_1 := 0$ and

$$\delta_n := \delta \left(\prod_{j=1}^{n-1} a_j, a_n \right) \tag{4.18}$$

with

$$\delta(a_m, a_{m+1}) = \begin{cases} \frac{|a_m| - a_m}{2}, & a_m \cdot a_{m+1} = +1 \\ \frac{|a_m| + a_m}{2}, & a_m \cdot a_{m+1} = -1. \end{cases} \tag{4.19}$$

Proof The proof is given by mathematical induction. In the case of $n = 1$ we obtain $\lambda(a_1, \theta_1) = \sum_{\alpha_1 \in \{-1,1\}} \lambda(a_1, \theta_1) = \lambda(+1, \theta_1) + \lambda(-1, \theta_1) = \cos \theta_1 + \mathbf{I} \sin \theta_1$. Now, let us assume that our equality holds for some natural number n. We have to show that it holds for $n + 1$. Here, we can state

$$2^n \prod_{r=1}^{n+1} \lambda(a_r, \theta_r) = 2^{n-1} \prod_{r=1}^{n} \lambda(a_r, \theta_r) \cdot \lambda(a_{n+1}, \theta_{n+1})$$

$$= \sum_{\alpha_2, \ldots, \alpha_n \in \{-1,1\}} \alpha_2^{\delta_2} \ldots \alpha_n^{\delta_n} \lambda$$

$$\times \left(\prod_{j=1}^{n} a_j, \theta_1 + \alpha_2 \theta_2 + \cdots + \alpha_n \theta_n \right) \cdot \lambda(a_{n+1}, \theta_{n+1})$$

$$= \sum_{\alpha_2, \ldots, \alpha_n, \alpha_{n+1} \in \{-1,1\}} \alpha_2^{\delta_2} \ldots \alpha_n^{\delta_n} \cdot \alpha_{n+1}^{\delta_{n+1}} \lambda$$

$$\times \left(\prod_{j=1}^{n+1} a_j, \theta_1 + \alpha_2 \theta_2 + \cdots + \alpha_n \theta_n + \alpha_{n+1} \theta_{n+1} \right).$$

$\square$

Lemma 4.4 *Let $n \in \mathbb{N}$. It holds*

$$\prod_{r=1}^{n} \left(e^{\mathbf{J} A_r} \cos \theta_r + e^{\mathbf{J} C_r} \sin \theta_r \mathbf{I} \right) = \sum_{\underline{a} \in \{-1,1\}^n} e^{\mathbf{J} \sum_{r=1}^{n} \tilde{\sigma}(a_r)} \mathbf{I}^{\frac{1 - \prod_{j=1}^{n} a_j}{2}} \prod_{r=1}^{n} \lambda(a_r, \theta_r), \tag{4.20}$$

where

$$\tilde{\sigma}(\underline{a}, r) = \begin{cases} (-1)^{\pi(r)} A_r, & a_r = +1 \\ (-1)^{\pi(r)} C_r, & a_r = -1, \end{cases} \qquad \lambda(a_r, \theta_r) = \begin{cases} \cos \theta_r, & a_r = +1 \\ \sin \theta_r, & a_r = -1, \end{cases}$$

such that $\pi(1) = 0$ *and* $\pi(n) = \sum_{j=1}^{n-1} \frac{|a_j| - a_j}{2}$ ($n > 0$) *which counts the number of* $\sin$'s in the vector $\underline{a} = (\underline{a}_1, \ldots, \underline{a}_n) \in \{-1, 1\}^n$.

Proof Again, the proof will be given by mathematical induction.

For the case $n = 1$ we have

$$e^{\mathbf{J} A_1} \cos \theta_1 + e^{\mathbf{J} C_1} \sin \theta_1 \mathbf{I} = \sum_{\underline{a} \in \{-1, 1\}} e^{\mathbf{J}\tilde{\sigma}(a_1)} \sigma(a_1) = e^{\mathbf{J}\tilde{\sigma}(+1)} \sigma(+1) + e^{\mathbf{J}\tilde{\sigma}(-1)} \sigma(-1),$$

where $\tilde{\sigma}(+1) = (-1)^{\pi(1)} A_1 = A_1$ and $\tilde{\sigma}(-1) = (-1)^{\pi(1)} C_1 = C_1$. Note that we have $\pi(1) = 0$ as initial condition. For the induction step, let $n \in \mathbb{N}$ be given and suppose (4.20) is true for n. Then

$$\prod_{r=1}^{n+1} \left(e^{\mathbf{J} A_r} \cos \theta_r + e^{\mathbf{J} C_r} \sin \theta_r \mathbf{I} \right)$$

$$= \prod_{r=1}^{n} \left(e^{\mathbf{J} A_r} \cos \theta_r + e^{\mathbf{J} C_r} \sin \theta_r \mathbf{I} \right) \left(e^{\mathbf{J} A_{n+1}} \cos \theta_{n+1} + e^{\mathbf{J} C_{n+1}} \sin \theta_{n+1} \mathbf{I} \right)$$

$$= \prod_{r=1}^{n} \sum_{\underline{a} \in \{-1, 1\}^n} e^{\mathbf{J} \sum_{j=1}^{n} \tilde{\sigma}(a_r)} \mathbf{I}^{\frac{1 - \prod_{j=1}^{n} a_j}{2}} \prod_{r=1}^{n} \lambda(a_r, \theta_r) \left(e^{\mathbf{J} A_{n+1}} \cos \theta_{n+1} + e^{\mathbf{J} C_{n+1}} \sin \theta_{n+1} \mathbf{I} \right)$$

$$= \prod_{r=1}^{n} \sum_{\underline{a} \in \{-1, 1\}^n} e^{\mathbf{J} \sum_{r=1}^{n} \tilde{\sigma}(a_r) + (-1)^{\pi(n+1)} A_{n+1}} \mathbf{I}^{\frac{1 - \prod_{j=1}^{n} a_j}{2}} \prod_{r=1}^{n} \lambda(a_r, \theta_r) \cos \theta_{n+1}$$

$$+ \prod_{r=1}^{n} \sum_{\underline{a} \in \{-1, 1\}^n} e^{\mathbf{J} \sum_{r=1}^{n} \tilde{\sigma}(a_r) + (-1)^{\pi(n+1)} C_{n+1}} \mathbf{I}^{\frac{1 - \prod_{j=1}^{n} a_j}{2}} \prod_{r=1}^{n} \lambda(a_r, \theta_r) \sin \theta_{n+1} \mathbf{I}$$

$$= \prod_{r=1}^{n+1} \sum_{\underline{a} \in \{-1, 1\}^{n+1}} e^{\mathbf{J} \sum_{r=1}^{n+1} \tilde{\sigma}(a_r)} \mathbf{I}^{\frac{1 - \prod_{j=1}^{n+1} a_j}{2}} \prod_{j=1}^{n+1} \lambda(a_r, \theta_r),$$

such that $\underline{a} = (\underline{a}, a_{n+1}) \in \{-1, 1\}^{n+1}$ and $a_{n+1} = \pm 1$. $\qquad \square$

Using the above lemmas we can now estimate the expectation value of the Frobenius norm of our random matrix H_0^n. The need of these lemmas stems from the fact that we cannot use a direct approach via exponentials due to the non-commutativity of quaternions.

Lemma 4.5 *It holds*

$$\mathbb{E}_X \left[\| H_0^n \|_F^2 \right] = \sum_{t=1}^{\min\{n, N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A} \in P(2n, t)} \sum_{\substack{(k_1, l_1), \ldots, (k_{2n}, l_{2n}) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}}$$

$$\times \prod_{A \subset \mathcal{A}} \sum_{\substack{\underline{a} \in \{-1, 1\}^r \\ a_1 \dots a_r = +1}} \delta \left(\sum_{r \in A} \tilde{\varrho}(a_r) \right) 2^{1-|A|} \sum_{\alpha_2, \dots, \alpha_n \in \{-1, 1\}} \alpha_2^{\delta_2} \dots \alpha_{|A|}^{\delta_{|A|}} \delta \left(\theta_1 + \sum_{r \in A} \alpha_s \theta_s \right),$$

where $\delta(n)$ denotes the Kronecker δ_{0n} and $(k_{2n+1}, l_{2n+1}) = (k_1, l_1)$ and

$$\tilde{\varrho}(\underline{a}, l_r) = \begin{cases} (-1)^{\pi(r)}(l_{r+1} - l_r), & a_r = +1 \\ -(-1)^{\pi(r)}(l_{r+1} + l_r), & a_r = -1, \end{cases} \quad \theta_s = (k_{s+1} - k_s)$$

such that $\pi(|A|) = \sum_{j=1}^{|A|-1} \frac{|a_j| - a_j}{2}$.

Proof Since H_0^n is self adjoint we have from the definition $\mathbb{E}_X[\|H_0^n\|_F^2] = \mathbb{E}_X[\mathrm{tr}\, H_0^{2n}]$. Consider

$$H_0[(k_1, l_1), (k_2, l_2)] = (1 - \delta_{k_1, k_2} \delta_{l_1, l_2}) \sum_{i_1 = 1}^{N} \left[e^{-\mathbf{J} l_1 y_{i_1}} e^{\mathbf{I}(k_2 - k_1) x_{i_1}} e^{\mathbf{J} l_2 y_{i_1}} \right],$$

$$(k_1, l_1), (k_2, l_2) \in T.$$

Algebraic operations with quaternions constitute a big handicap because quaternions do not commute. Thus, we have to find a new strategy. What plays to our advantage are the anti-commutativity of the basis elements, that is, $\mathbf{IJ} = -\mathbf{JI}$. Moreover, this means that $\mathbf{I}e^{\mathbf{J}y} = e^{-\mathbf{J}y}\mathbf{I}$. Thus, we can simplify our expression as

$$e^{-\mathbf{J} l_r y_{i_r}} e^{\mathbf{I}(k_{r+1} - k_r) x_{i_r}} e^{\mathbf{J} l_{r+1} y_{i_r}} = e^{\mathbf{J}(l_{r+1} - l_r) y_{i_r}} \cos\left([k_{r+1} - k_r] x_{i_r}\right)$$
$$+ e^{-\mathbf{J}(l_{r+1} + l_r) y_{i_r}} \sin\left([k_{r+1} - k_r] x_{i_r}\right) \mathbf{I}. \quad (4.21)$$

Furthermore, we can write the $2n$-th power of the matrix H_0 as follows

$$H_0^{2n}\left[(k_1, l_1), (k_{2n+1}, l_{k_{2n+1}})\right]$$
$$= \sum_{\substack{(k_2, l_2), \dots, (k_{2n}, l_{2n}) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}} H_0[(k_1, l_1), (k_2, l_2)] \dots H_0\left[(k_{2n}, l_{2n})(k_{2n+1}, l_{k_{2n+1}})\right]$$
$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_{2n} = 1}}^{N} \sum_{\substack{(k_2, l_2), \dots, (k_{2n}, l_{2n}) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), j = 1, \dots, 2n}} \left[\prod_{r=1}^{2n} e^{-\mathbf{J} l_r y_{i_r}} e^{\mathbf{I}(k_{r+1} - k_r) x_{i_r}} e^{\mathbf{J} l_{r+1} y_{i_r}} \right].$$

Since we need the diagonal element of the $2n$-th power of our matrix to calculate the norm $\|H_0^n\|_F^2 = \mathrm{Tr}\, H_0^{2n}$ we require $(k_{2n+1}, l_{k_{2n+1}}) = (k_1, l_1)$. Therefore, using the linearity of the expectation value it follows

$$\mathbb{E}_X\left[\mathrm{Tr}H_0^{2n}\right]$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_{2n} = 1}}^{N} \sum_{\substack{(k_1, \ell_1), \ldots, (k_{2n}, l_{2n}) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}} \mathbb{E}_X\left[\prod_{r=1}^{2n} e^{-\mathbf{J}l_r y_{i_r}} e^{\mathbf{I}(k_{r+1}-k_r)x_{i_r}} e^{\mathbf{J}l_{r+1} y_{i_r}}\right].$$

Furthermore, some indices i_r might be the same which means that we cannot use directly the product rule for the expectation value since it is only valid for independent random variables. This is where we consider a set partition. We associate a partition $\mathcal{A} = (A_1, A_2, \ldots, A_t)$ of $\{1, \ldots, 2n\}$ to a certain vector $i_1, \ldots, i_{2n}$ such that $i_r = i_{r'}$ if and only if r and r' are contained in the same set $A_i \in \mathcal{A}$. This is allows us to unambiguously write i_A instead of i_r if $r \in A$. Now, the independence of the variables (x_A, y_A) yields

$$\mathbb{E}_X\left[\prod_{r=1}^{2n} \exp\left(-\mathbf{J}l_r y_{i_r}\right) \exp\left(\mathbf{I}(k_{r+1} - k_r)x_{i_r}\right) \exp\left(\mathbf{J}l_{r+1} y_{i_r}\right)\right]$$

$$= \mathbb{E}_X\left[\prod_{A \in \mathcal{A}} \prod_{r \in A} \exp\left(-\mathbf{J}l_r y_A\right) \exp\left(\mathbf{I}(k_{r+1} - k_r)x_A\right) \exp\left(\mathbf{J}l_{r+1} y_A\right)\right]$$

$$= \prod_{A \in \mathcal{A}} \mathbb{E}_X\left[\prod_{r \in A} \exp\left(-\mathbf{J}l_r y_A\right) \exp\left(\mathbf{I}(k_{r+1} - k_r)x_A\right) \exp\left(\mathbf{J}l_{r+1} y_A\right)\right]. \quad (4.22)$$

Let us introduce now

$$\tilde{\sigma}(\underline{a}, l_r) = \begin{cases} (-1)^{\pi(r)}(l_{r+1} - l_r)y_A, & a_r = +1 \\ (-1)^{\pi(r)}(-l_{r+1} - l_r)y_A, & a_r = -1, \end{cases}$$

and

$$\lambda(a_r, \theta_r) = \begin{cases} \cos(k_{r+1} - k_r)x_A, & a_r = +1 \\ \sin(k_{r+1} - k_r)x_A, & a_r = -1 \end{cases},$$

and by applying Lemma 4.4 we get for the expectation value in (4.22)

$$\mathbb{E}_X\left(\sum_{\underline{a}\in\{-1,1\}^r} e^{\mathbf{J}\sum_{r\in A}\tilde{\sigma}(a_r)} \prod_{r\in A} \lambda(a_r, \theta_r)\right)$$

$$= \mathbb{E}_X\left(\sum_{\substack{\underline{a} \in \{-1, 1\}^r \\ a_1 \ldots a_r = +1}} e^{\mathbf{J}\sum_{r\in A}\tilde{\sigma}(a_r)} \prod_{r\in A} \lambda(a_r, \theta_r)) + \sum_{\substack{\underline{a} \in \{-1, 1\}^r \\ a_1 \ldots a_r = -1}} e^{\mathbf{J}\sum_{r\in A}\tilde{\sigma}(a_r)}\mathbf{I} \prod_{r\in A} \lambda(a_r, \theta_r)\right)$$

$$
= \mathbb{E}_X \left(\sum_{\substack{\underline{a} \, \in \, \{-1, \, 1\}^r \\ a_1 \ldots a_r \, = \, +1}} e^{\mathbf{J} \sum_{r \in A} \tilde{\sigma}(a_r)} \prod_{r \in A} \lambda(a_r, \theta_r) \right)
$$

$$
+ \, \mathbb{E}_X \left(\sum_{\substack{\underline{a} \, \in \, \{-1, \, 1\}^r \\ a_1 \ldots a_r \, = \, -1}} e^{\mathbf{J} \sum_{r \in A} \tilde{\sigma}(a_r)} \mathbf{I} \prod_{r \in A} \lambda(a_r, \theta_r) \right). \tag{4.23}
$$

The last line in (4.23) can be analyzed as follows: when $a_1 \ldots a_r = +1$ then the term is independent of $\mathbf{I}$ and when $a_1 \ldots a_r = -1$ the term depends on $\mathbf{I}$. This can be easily seen as in the first case the number of $a_j = -1$ which corresponds to the terms containing $\mathbf{I}$ being even. For this reason the basis element $\mathbf{I}$ vanishes because $\mathbf{I}^{2r} = \pm 1$, with $r \in \mathbb{N}$. For the second term in (4.23) the number of terms with $a_j = -1$ is odd, matching the number of sin-functions in the product. That means the basis element $\mathbf{I}$ never vanishes for the second term since $\mathbf{I}^{2r-1} = \pm \mathbf{I}$, with $r \in \mathbb{N}$.

We emphasized here that the expectation value of the second term vanishes because it is given in terms of sin-functions which vanish when x_{s_A} is integrated over $[0, 2\pi]$. Therefore, only the first term is surviving. In turn, the first term has 2^{n-1} sub-terms. Thus, we can state

$$
\mathbb{E}_X \left[\sum_{\substack{\underline{a} \, \in \, \{-1, \, 1\}^r \\ a_1 \ldots a_r \, = \, +1}} e^{\mathbf{J} \sum_{r \in A} \tilde{\sigma}(a_r)} \prod_{r \in A} \lambda(a_r, \theta_r) \right]
$$

$$
= \sum_{\substack{\underline{a} \, \in \, \{-1, \, 1\}^r \\ a_1 \ldots a_r \, = \, -1}} \int_{[0, 2\pi]} e^{\mathbf{J} \sum_{r \in A} \tilde{\sigma}(a_r)} dy \int_{[0, 2\pi]} \prod_{r \in A} \lambda(a_r, \theta_r) dx
$$

where

$$
\tilde{\sigma}(a_r) = \begin{cases} (-1)^{\pi(r)} (l_{r+1} - l_r) y_r, & a_r = +1 \\ -(-1)^{\pi(r)} (l_{r+1} + l_r) y_r, & a_r = -1, \end{cases}
$$

and

$$
\lambda(a_r, \theta_r) = \begin{cases} \cos[(k_{r+1} - k_r) x_r], & a_r = +1 \\ \sin[(k_{r+1} - k_r) x_r], & a_r = -1, \end{cases}
$$

such that $\pi(n) = \sum_{j=1}^{n-1} \frac{1-a_j}{2}$. The last quantity $\pi(n)$ does nothing else then counting the number of sin-functions. By invoking Lemma 4.3 for the product altogether we obtain

$$\mathbb{E}_X \left[\exp \left(\mathbf{J} \sum_{r \in A} \tilde{\sigma}(\underline{a}, l_r) \right) \prod_{s \in A} \lambda(a_r, \theta_r) \right]$$

$$= \sum_{\substack{\underline{a} \in \{-1, 1\}^r \\ a_1 \dots a_r = +1}} \delta \left(\sum_{r \in A} \tilde{\varrho}(a_r) \right) 2^{1-|A|} \sum_{\alpha_2, \dots, \alpha_n \in \{-1, 1\}} \alpha_2^{\delta_2} \dots \alpha_{|A|}^{\delta_{|A|}} \delta \left(\theta_1 + \sum_{r \in A} \alpha_r \theta_r \right),$$

such that

$$\tilde{\varrho}(\underline{a}, l_r) = \begin{cases} (-1)^{\pi(r)} (l_{r+1} - l_r), & a_r = +1 \\ -(-1)^{\pi(r)} (l_{r+1} + l_r), & a_r = -1. \end{cases} \tag{4.24}$$

Furthermore, the condition $|A| \geq 2$ for all $A \in \mathcal{A}$ should be satisfied, i.e., we consider partitions $P(2n, t)$ with $t \geq 2$. Moreover, the number of vectors $(A_1, A_2, \dots, A_t) \in \{1, \dots, N\}^t$ with different entries is exactly $N \dots (N - t + 1) = N!/(N - t)!$ if $N \geq t$ and 0 if $N \leq t$. $\qquad\square$

For later reference let us introduce the following notation

$$C_{\mathbb{H}}(\mathcal{A}, T) := \sum_{\substack{(k_1, l_1), \dots, (k_{2n}, l_{2n}) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}} \sum_{\substack{\underline{a} \in \{-1, 1\}^r \\ a_1 \dots a_r = +1}} \prod_{A \in \mathcal{A}}$$

$$\times \delta \left(\sum_{s \in A} \tilde{\varrho}(a_s) \right) 2^{1-|A|} \sum_{\alpha_2, \dots, \alpha_n \in \{-1, 1\}} \alpha_2^{\delta_2} \dots \alpha_{|A|}^{\delta_{|A|}} \delta \left(\theta_1 + \sum_{r \in A} \alpha_r \theta_r \right),$$

$$\tag{4.25}$$

with

$$\tilde{\varrho}(\underline{a}, l_s) = \begin{cases} \beta_s (l_{s+1} - l_s), & a_s = +1 \\ -\beta_s (l_{s+1} + l_s), & a_s = -1. \end{cases}$$

and $\beta_s = (-1)^{\pi(s)}$ where $\pi(s) = \sum_{j=1}^{n-1} \frac{|a_j| - a_j}{2}$.

4.2 Analysis of $\mathbb{P}(E_{k,l})$

While we have the estimate for the Frobenius norm of H_0^n from the previous section we still need to study the probability $\mathbb{P}(E_{k,l})$, i.e., the first term in (4.16) and (4.14). Similar to [28] consider the numbers $\beta_m > 0$, $m = 1, \dots, n$, such that

$$\sum_{m=1}^{n} \beta_m = a_1$$

and $K_m \in \mathbb{N}$, $m = 1, \dots, n$, some natural numbers. Let $(k, l) \in Z_\rho^2$. By applying Markov's inequality it follows

$$\mathbb{P}(E_{k,l}) = \mathbb{P}\left(\sum_{m=1}^{n}|((N^{-1}HR_T)^m\,\mathrm{sgn}(c))_{k,l}| \geq a_1\right)$$

$$\leq \sum_{m=1}^{n}\mathbb{P}\left(N^{-m}|((HR_T)^m\,\mathrm{sgn}(c))_{k,l}| \geq \beta_m\right)$$

$$= \sum_{m=1}^{n}\mathbb{P}\left(N^{-2mK_m}|((HR_T)^m\,\mathrm{sgn}(c))_{k,l}|^{2K_m} \geq \beta_m^{2K_m}\right)$$

$$\leq \sum_{m=1}^{n}\mathbb{E}_X\left(|((HR_T)^m\,\mathrm{sgn}(c))_{k,l}|^{2K_m}\right)N^{-2mK_m}\beta_m^{-2K_m}. \qquad (4.26)$$

Let us consider $\beta_m = \beta^{n/K_m}$, i.e., $\beta_m^{-2K_m} = \beta_m^{-2n}$. Now, from (4.26) it follows

$$\mathbb{P}(E_{k,l}) \leq \beta_m^{-2n}\sum_{m=1}^{n}\mathbb{E}_X\left(|((HR_T)^m\,\mathrm{sgn}(c))_{k,l}|^{2K_m}\right)N^{-2mK_m} \qquad (4.27)$$

while the condition $a_1 = \sum_{m=1}^{n}\beta_m$ can be written as

$$a_1 = a = \sum_{m=1}^{n}\beta^{n/K_m} < 1.$$

The following lemma attends the expectation value appearing in (4.27). While Lemma 4.6 is similar to Lemma 4.5, it requires much more work and, unfortunately, lengthy calculations. That is why we need to refer to Appendix B in the Ph.D. thesis [19] for the lengthy calculations which form part of this proof.

Lemma 4.6 *For a sequence* $c = (c_{k,l}) \in \ell_2(\mathbb{Z}_\rho^2)$ *with* $\mathrm{supp}\,c = T$, *it holds*

$$\mathbb{E}_X\left[\left|\left((HR_T)^m\,\mathrm{sgn}(c)\right)_{k,l}\right|^{2K}\right]$$

$$\leq \sum_{t=1}^{\min\{Km,N\}}\frac{N!}{(N-t)!}\sum_{\mathcal{A}\in P(2Km,t)}\sum_{\substack{(k_1^{(1)},l_1^{(1)}),\,\ldots,\,(k_m^{(1)},l_m^{(1)})\,\in\,T \\ \vdots \\ (k_1^{(K)},l_1^{(K)}),\,\ldots,\,(k_m^{(K)},l_m^{(K)})\,\in\,T \\ (k_j^{(p)},l_j^{(p)})\neq(k_{j+1}^{(p)},l_{j+1}^{(p)})}}2^{3-m}$$

$$\times \sum_{(r_1,\ldots,r_K)\subset V(\{1,\ldots,16\},K)}\sum_{\substack{\underline{a}\,\in\,\{-1,1\}^m \\ a_1^{(1)}\ldots a_m^{(1)}=\pm1 \\ \vdots \\ a_1^{(K)}\ldots a_m^{(K)}=\pm1}}\sum_{\substack{\alpha_2^{(1)},\,\ldots,\,\alpha_m^{(1)}\,\in\,\{-1,1\} \\ \vdots \\ \alpha_2^{(K)},\,\ldots,\,\alpha_m^{(K)}\,\in\,\{-1,1\}}}$$

$$\times \prod_{A \in \mathcal{A}} \delta \left(\sum_{(r,p) \in A} \alpha_r^{(p)} \theta_r^{(p)} \right) \delta \left(\sum_{(s,p) \in A} \beta_s^{(p)} \phi_s^{(p)} \right),$$

where $\theta_r^{(p)} = (k_{r+1}^{(p)} - k_r^{(p)})$ and $\phi_s^{(p)} = \begin{cases} (l_{s+1}^{(p)} - l_s^{(p)}), & a_s = +1 \\ -(l_{s+1}^{(p)} + l_s^{(p)}), & a_s = -1, \end{cases}$ with $\alpha_r^{(p)} \in$ $\{-1, 1\}$ and $\beta_s^{(p)} = (-1)^{\pi(s)}$ such that $\pi(s) = \sum_{j=1}^{s-1} \frac{|a_j| - a_j}{2}$ counts the number of sin's.

Proof We want to compute the expectation value $\mathbb{E}_X$ of the Kth-power of the modulus, i.e., the expectation value of the quantity $|((HR_T)^m \operatorname{sgn}(c_{k,l}))_{k,l}|^{2K}$. But there exist a major problem due the non-commutativity in the quaternionic setting. While in the previous proof we could still use the anti-commutativity of the basis elements **I** and **J** we now have to take into account quaternion-valued coefficients. That means we have to compute first $(HR_T)^m$ and then multiply it by the coefficient vector from the signal. This we are going to do in the way that we write the previous quantity as $(|((HR_T)^m \sigma)_{k,l}|^2)^K = (Q_0^2 + Q_1^2 + Q_2^2 + Q_3^2)^K$. Of course, breaking the power of a quaternion in several small pieces and joining afterwards usually involves lengthy calculations. We refer for the calculation of the large formulae to Appendix B in the Ph.D. thesis [19]. Thus, we get

$$\big((HR_T)\,[(k_1, l_1), (k_{m+1}, l_{m+1})]\big)^m$$

$$= \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m)\, \in\, T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}),\ j = 1, \ldots, m}} HR_T\,[(k_1, l_1), (k_2, l_2)] \ldots HR_T\,[(k_m, l_m)\,(k_{m+1}, l_{m+1})]$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m)\, \in\, T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}),\ j = 1, \ldots, m}} \left[\prod_{r=1}^{m} e^{-\mathbf{J} l_r\, y_{i_r}} e^{\mathbf{I}(k_{r+1} - k_r) x_{i_r}} e^{\mathbf{J} l_{r+1}\, y_{i_r}} \right]$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m)\, \in\, T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}),\ j = 1, \ldots, m}} \sum_{\underline{a}\, \in\, \{-1, 1\}^m} e^{\mathbf{J} \sum_{r=1}^{m} \beta_r \phi_r} \prod_{s=1}^{m} \lambda(a_s, \theta_s)$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m)\, \in\, T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}),\ j = 1, \ldots, m}} \sum_{\substack{\underline{a}\, \in\, \{-1, 1\}^m \\ a_1 \ldots a_m = +1}} 2^{1-m}$$

$$\times e^{\mathbf{J} \sum_{r=1}^{m} \beta_r \phi_r} \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_m^{\delta_m} \lambda \left(\prod_{j=1}^{m} a_j, \theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right)$$

$$+ \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), \; j = 1, \ldots, m}} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1 \ldots a_m = -1}} 2^{1-m}$$

$$\times e^{\mathbf{J} \sum_{r=1}^{m} \beta_r \phi_r} \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_m^{\delta_m} \lambda \left(\prod_{j=1}^{m} a_j, \theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right).$$

In accordance with the proof of our previous Lemma 4.3, we have

$$\left((H R_T) \left[(k_1, l_1), (k_{m+1}, l_{m+1}) \right] \right)^m$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), \; j = 1, \ldots, m}} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1 \ldots a_m = +1}} 2^{1-m}$$

$$\times e^{\mathbf{J} \sum_{r=1}^{m} \beta_r \phi_r} \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_m^{\delta_m} \cos \left(\theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right)$$

$$+ \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), \; j = 1, \ldots, m}} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1 \ldots a_m = -1}} 2^{1-m}$$

$$\times e^{\mathbf{J} \sum_{r=1}^{m} \beta_r \phi_r} \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_m^{\delta_m} \sin \left(\theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right) \mathbf{I}$$

$$= \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), \; j = 1, \ldots, m}} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1 \ldots a_m = +1}} 2^{1-m}$$

$$\times \left(\cos \left(\sum_{r=1}^{m} \beta_r \phi_r \right) + \mathbf{J} \sin \left(\sum_{r=1}^{m} \beta_r \phi_r \right) \right) \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_m^{\delta_m} \cos \left(\theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right)$$

$$+ \sum_{\substack{i_1 = 1 \\ \vdots \\ i_m = 1}}^{N} \sum_{\substack{(k_2, l_2), \ldots, (k_m, l_m) \in T \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1}), \; j = 1, \ldots, m}} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1 \ldots a_m = -1}} 2^{1-m}$$

$$\left(\cos \left(\sum_{r=1}^{m} \beta_r \phi_r \right) + \mathbf{J} \sin \left(\sum_{r=1}^{m} \beta_r \phi_r \right) \right) \sum_{\alpha_2, \ldots, \alpha_m \in \{-1, 1\}} \alpha_2^{\delta_2} \ldots \alpha_n^{\delta_m} \sin \left(\theta_1 + \sum_{r=2}^{m} \alpha_r \theta_r \right) \mathbf{I}$$

$$
= \sum_{\substack{i_1=1 \\ \vdots \\ i_m=1}} \sum_{\substack{(k_2,l_2),\ldots,(k_m,l_m)\in T \\ (k_j,l_j)\neq(k_{j+1},l_{j+1}),\ j=1,\ldots,m}} \sum_{\substack{\underline{a}\in\{-1,1\}^m \\ a_1\ldots a_m=+1}} 2^{1-m}
$$

$$
\times \cos\left(\sum_{r=1}^m \beta_r\phi_r\right) \sum_{\alpha_2,\ldots,\alpha_m\in\{-1,1\}} \alpha_2^{\delta_2}\ldots\alpha_m^{\delta_m}\cos\left(\theta_1+\sum_{r=2}^m \alpha_r\theta_r\right)
$$

$$
+ \sum_{\substack{i_1=1 \\ \vdots \\ i_m=1}} \sum_{\substack{(k_2,l_2),\ldots,(k_m,l_m)\in T \\ (k_j,l_j)\neq(k_{j+1},l_{j+1}),\ j=1,\ldots,m}} \sum_{\substack{\underline{a}\in\{-1,1\}^m \\ a_1\ldots a_m=+1}} 2^{1-m}
$$

$$
\times \sin\left(\sum_{r=1}^m \beta_r\phi_r\right) \sum_{\alpha_2,\ldots,\alpha_m\in\{-1,1\}} \alpha_2^{\delta_2}\ldots\alpha_m^{\delta_m}\cos\left(\theta_1+\sum_{r=2}^m \alpha_r\theta_r\right)\mathbf{J}
$$

$$
+ \sum_{\substack{i_1=1 \\ \vdots \\ i_m=1}} \sum_{\substack{(k_2,l_2),\ldots,(k_m,l_m)\in T \\ (k_j,l_j)\neq(k_{j+1},l_{j+1}),\ j=1,\ldots,m}} \sum_{\substack{\underline{a}\in\{-1,1\}^m \\ a_1\ldots a_m=-1}} 2^{1-m}
$$

$$
\times \cos\left(\sum_{r=1}^m \beta_r\phi_r\right) \sum_{\alpha_2,\ldots,\alpha_m\in\{-1,1\}} \alpha_2^{\delta_2}\ldots\alpha_m^{\delta_m}\sin\left(\theta_1+\sum_{r=2}^m \alpha_r\theta_r\right)\mathbf{I}
$$

$$
+ \sum_{\substack{i_1=1 \\ \vdots \\ i_m=1}} \sum_{\substack{(k_2,l_2),\ldots,(k_m,l_m)\in T \\ (k_j,l_j)\neq(k_{j+1},l_{j+1}),\ j=1,\ldots,m}} \sum_{\substack{\underline{a}\in\{-1,1\}^m \\ a_1\ldots a_m=-1}} 2^{1-m}
$$

$$
\times \sin\left(\sum_{r=1}^m \beta_r\phi_r\right) \sum_{\alpha_2,\ldots,\alpha_m\in\{-1,1\}} \alpha_2^{\delta_2}\ldots\alpha_m^{\delta_m}\sin\left(\theta_1+\sum_{r=2}^m \alpha_r\theta_r\right)\mathbf{JI}
$$

$$
= P_0 + \mathbf{I}P_1 + \mathbf{J}P_2 - \mathbf{K}P_3.
$$

Let us consider $\sigma_m = \sigma_m(k_1,l_1) = \sigma_{0m}(k_1,l_1) + \mathbf{I}\sigma_{1m}(k_1,l_1) + \mathbf{J}\sigma_{2m}(k_1,l_1) + \mathbf{K}\sigma_{3m}(k_1,l_1)$. We are interested to compute

$$
\sum_{(k_1,l_1)\in T} [P_0 + \mathbf{I}P_1 + \mathbf{J}P_2 - \mathbf{K}P_3]
$$
$$
\times [\sigma_{0m}(k_1,l_1) + \mathbf{I}\sigma_{1m}(k_1,l_1) + \mathbf{J}\sigma_{2m}(k_1,l_1) + \mathbf{K}\sigma_{3m}(k_1,l_1)]
$$
$$
= \sum_{(k_1,l_1)\in T} \Big[(P_0\sigma_{0m}(k_1,l_1) - P_1\sigma_{1m}(k_1,l_1) - P_2\sigma_{2m}(k_1,l_1) + P_3\sigma_{3m}(k_1,l_1))
$$

$$+ \left(P_0\sigma_{1m}(k_1, l_1) + P_1\sigma_{0m}(k_1, l_1) + P_2\sigma_{\rho 3} + P_3\sigma_{2m}(k_1, l_1) \right) \mathbf{I}$$
$$+ \left(P_0\sigma_{2m}(k_1, l_1) - P_1\sigma_{3m}(k_1, l_1) + P_2\sigma_{0m}(k_1, l_1) - P_3\sigma_{1m}(k_1, l_1) \right) \mathbf{J}$$
$$+ \left. (P_0\sigma_{03m}(k_1, l_1) + P_1\sigma_{2m}(k_1, l_1) - P_2\sigma_{1m}(k_1, l_1) - P_3\sigma_{0m}(k_1, l_1)) \mathbf{K} \right]$$

$$= \sum_{(k_1, l_1) \in T} (P_0\sigma_{0m}(k_1, l_1) - P_1\sigma_{1m}(k_1, l_1) - P_2\sigma_{2m}(k_1, l_1) + P_3\sigma_{3m}(k_1, l_1))$$

$$+ \sum_{(k_1, l_1) \in T} \left(P_0\sigma_{1m}(k_1, l_1) + P_1\sigma_{0m}(k_1, l_1) + P_2\sigma_{\rho 3} + P_3\sigma_{2m}(k_1, l_1) \right) \mathbf{I}$$

$$+ \sum_{(k_1, l_1) \in T} (P_0\sigma_{2m}(k_1, l_1) - P_1\sigma_{3m}(k_1, l_1) + P_2\sigma_{0m}(k_1, l_1) - P_3\sigma_{1m}(k_1, l_1)) \mathbf{J}$$

$$+ \sum_{(k_1, l_1) \in T} (P_0\sigma_{03m}(k_1, l_1) + P_1\sigma_{2m}(k_1, l_1) - P_2\sigma_{1m}(k_1, l_1) - P_3\sigma_{0m}(k_1, l_1)) \mathbf{K}$$

$$= Q_0 + Q_1\mathbf{I} + Q_2\mathbf{J} + Q_3\mathbf{K}.$$

Let us consider the expectation value applied to the sum. As in the proof of Lemma 4.5 we have to take into account that some of the indices $i_r^{(p)}$ might coincide. This means that we have to use again set partitions. Let $(i_r^{(p)})_{r=1,\dots,m}^{p=1,\dots,2K} \subset \{1, \dots, N\}^{2Km}$ be a vector of indices and let $\mathcal{A} = (A_1, A_2, \dots, A_t)$, $A_i \subset \{1, \dots, m\} \times \{1, \dots, 2K\}$ be a corresponding partition such that (r, p) and (r', p') are contained in the same block if and only if $i_r^{(p)} = i_{r'}^{(p')}$. For some $A \in \mathcal{A}$ we may unambiguously write i_A instead of $i_r^{(p)}$ if $(r, p) \in A$. Once again, if $A \subset \mathcal{A}$ contains only one element then the last expression vanishes due to the condition $(k_r^{(p)}, l_r^{(p)}) \neq (k_{r-1}^{(p)}, l_{r-1}^{(p)})$. Thus, we only need to consider partitions $\mathcal{A} \in P(2Km, t)$ with $t > 1$. Now we are able to rewrite the inequality as

$$\mathbb{E}_X \left[\left| \left((HR_T)^m \operatorname{sgn}(c_{k,l}) \right)_{k,l} \right|^{2K} \right]$$

$$\leq \sum_{t=1}^{\min\{Km, N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A} \in P(2Km, t)} \sum_{\substack{(k_1^{(1)}, l_1^{(1)}), \dots, (k_m^{(1)}, l_m^{(1)}) \in T \\ \vdots \\ (k_1^{(K)}, l_1^{(K)}), \dots, (k_m^{(K)}, l_m^{(K)}) \in T \\ (k_j^{(p)}, l_j^{(p)}) \neq (k_{j+1}^{(p)}, l_{j+1}^{(p)})}} 2^{3-m}$$

$$\times \sum_{(r_1, \dots, r_K) \subset V(\{1, \dots, 16\}, K)} \sum_{\substack{\underline{a} \in \{-1, 1\}^m \\ a_1^{(1)} \dots a_m^{(1)} = \pm 1 \\ \vdots \\ a_1^{(K)} \dots a_m^{(K)} = \pm 1}} \sum_{\substack{\alpha_2^{(1)}, \dots, \alpha_m^{(1)} \in \{-1, 1\} \\ \vdots \\ \alpha_2^{(K)}, \dots, \alpha_m^{(K)} \in \{-1, 1\}}} \prod_{A \in \mathcal{A}}$$

$$\times \prod_{A \in \mathcal{A}} \delta \left(\sum_{(r,p) \in A} \alpha_r^{(p)} \theta_r^{(p)} \right) \delta \left(\sum_{(s,p) \in A} \beta_s^{(p)} \phi_s^{(p)} \right),$$

where $\theta_r^{(p)} = (k_{r+1}^{(p)} - k_r^{(p)})$ and $\phi_s^{(p)} = \begin{cases} (l_{s+1}^{(p)} - l_s^{(p)}), & a_s = +1 \\ -(l_{s+1}^{(p)} + l_s^{(p)}), & a_s = -1, \end{cases}$ which $\alpha_r^{(p)} \in$

$\{-1, 1\}$ and $\beta_s^{(p)} = (-1)^{\pi(s)}$ with $\pi(s) = \sum_{j=1}^{s-1} \frac{|a_j| - a_j}{2}$. $\qquad \square$

To simplify our work in the next section we abbreviate

$$B_{\mathbb{H}}(\mathcal{A}, T) = \sum_{\substack{(k_1^{(1)}, l_1^{(1)}), \dots, (k_m^{(1)}, l_m^{(1)}) \in T \\ \vdots \\ (k_1^{(K)}, l_1^{(K)}), \dots, (k_m^{(K)}, l_m^{(K)}) \in T \\ (k_j^{(p)}, l_j^{(p)}) \neq (k_{j+1}^{(p)}, l_{j+1}^{(p)})}} 2^{3-m}$$

$$\times \sum_{(r_1, \dots, r_K) \subset V(\{1, \dots, 16\}, K)} \sum_{\substack{a \in \{-1, 1\}^m \\ a_1^{(1)} \dots a_m^{(1)} = \pm 1 \\ \vdots \\ a_1^{(K)} \dots a_m^{(K)} = \pm 1}} \sum_{\substack{\alpha_2^{(1)}, \dots, \alpha_m^{(1)} \in \{-1, 1\} \\ \vdots \\ \alpha_2^{(K)}, \dots, \alpha_m^{(K)} \in \{-1, 1\}}}$$

$$\times \prod_{A \in \mathcal{A}} \delta \left(\sum_{(r,p) \in A} \alpha_r^{(p)} \theta_r^{(p)} \right) \delta \left(\sum_{(s,p) \in A} \beta_s^{(p)} \phi_s^{(p)} \right). \tag{4.28}$$

4.3 Proof of Theorem 3.1

Using our lemmas from the previous section, in particular Lemma 4.5, we will give now the proof of Theorem 3.1. The value of the quantity $C_{\mathbb{H}}(\mathcal{A}, T)$ as defined in (4.25) depends on T and $\mathcal{A} \in P(2n, t)$. Here, the indices $((k_1, l_1), \dots, (k_{2n}, l_{2n})) \in T^{2n}$ are conditioned by the $|\mathcal{A}| = t$ linear constraints $\sum_{r \in A} (k_{r+1} - k_r) = 0$ and $\sum_{s \in A} (l_{s+1} \pm l_s) = 0$ for all $A \in \mathcal{A}$. These constraints are independent except for $\sum_{r=1}^{2n} (k_{r+1} - k_r) = 0$ and $\sum_{s=1}^{2n} (l_{s+1} \pm l_s) = 0$. Thus, from (4.25) we can estimate

$$C_{\mathbb{H}}(\mathcal{A}, T) \leq |T|^{2n-t+1} \leq M^{2n-t+1}. \tag{4.29}$$

By Lemma 4.5 we obtain (since in Theorem 3.1 T is not random which means $\mathbb{E} = \mathbb{E}_X$)

$$\mathbb{E} \left[\|H_0^n\|_F^2 \right] \leq \sum_{t=1}^{\min\{n, N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A} \in P(2n, t)} |T|^{2n-t+1} \leq M^{2n+1} \sum_{t=1}^{n} \left(\frac{N}{M} \right)^t S_2(2n, t),$$

where $S_2(n, t) = |P(2n, t)|$ are the associated Stirling numbers of the second kind. Let us put $\theta = \frac{N}{M}$. By Markov's inequality it follows

$$\mathbb{P}\left(\left\|\left(N^{-1}H_0\right)^n\right\|_F \geq \kappa\right) = \mathbb{P}\left(\left\|H_0^n\right\|_F^2 \geq N^{2n}\kappa^2\right) \leq N^{-2n}\kappa^{-2}\,\mathbb{E}\left[\left\|H_0^n\right\|_F^2\right]$$
$$\leq \kappa^{-2}\,M\,\theta^{-2n}F_{2n}(\theta) = \kappa^{-2}\,M\,G_{2n}(\theta).$$

Let us point out that $\kappa < 1$. This implies $\left\|(N^{-1}H_0)^n\right\|_F \leq \kappa$ and, therefore, $(I_T - (N^{-1}H_0)^n)$ and also $[\mathcal{F}_{TX}^*\mathcal{F}_{TX}] = N(I_T - N^{-1}H_0)$ are invertible. In particular, $\mathcal{F}_{TX}$ is injective.

Let us now take a look at $P(E_{k,l})$. By Lemma 4.6 we need to estimate $B_{\mathbb{H}}(\mathcal{A}, T)$, which represents the number of elements $(k_j^{(p)}, l_j^{(p)}) \in T^{2Km}$ that satisfy $\sum_{(r,p)}(k_r^{(p)} - k_{r-1}^{(p)}) = 0$ and $\sum_{(s,p)}(k_{s+1}^{(p)} \pm k_s^{(p)}) = 0$ for all $A \in \mathcal{A}$ with $\mathcal{A} \in P(2Km, t)$. Since we have t independent linear constraints $B_{\mathbb{H}}(\mathcal{A}, T)$ is bounded from above by $|T|^{2Km-t} \leq M^{2Km-t}$. Thus, by putting again $\theta = \frac{N}{M}$, we get

$$\mathbb{E}_X\left[\left|\left((HR_T)^m\,\mathrm{sgn}\,c\right)_{k\tilde{k}}\right|^{2K}\right] \leq \sum_{t=1}^{Km} N^t S_2(2Km, t)M^{2Km-t} = M^{2Km}F_{2Km}(\theta),$$

and, therefore, we get

$$P(E_{k,l}) = \beta^{-2n}\sum_{m=1}^{n}\theta_{2mK_m}F_{2mK_m}(\theta) = \beta^{-2n}\sum_{m=1}^{n}G_{2mK_m}(\theta).$$

Let us denote by $\mathbb{P}(\text{failure})$ the probability that exact reconstruction of f by means of ℓ_1-minimization fails.

By Lemma 4.1 and the above estimates it follows

$$\mathbb{P}(\text{failure}) \leq \mathbb{P}\left(\{\mathcal{F}_{TX}\{\text{is not injective}\} \cup \{\sup_{(k,l)\in T^c}|P_{k,l}| \geq 1\}\right)$$
$$\leq \sum_{(k,l)\in\mathbb{Z}_\rho^2}\mathbb{P}(E_{k,l}) + \mathbb{P}\left(\left\|\left(N^{-1}H_0\right)^n\right\| \geq \kappa\right)$$
$$\leq (2\rho + 1)^2\beta^{-2n}\sum_{m=1}^{n}G_{2mK_m}(\theta) + \kappa^{-2}\,M\,G_{2n}(\theta)$$

under the conditions

$$a_1 = a = \sum_{m=1}^{n}\beta^{n/K_m} < 1, \quad a_2 + a_1 = 1, \quad \text{i.e., } a_2 = 1 - a,$$
$$\frac{\kappa}{1 - \kappa} \leq \frac{a_2}{1 + a_1}M^{-3/2} = \frac{1 - a}{1 + a}M^{-3/2}.$$

$\square$

Let us remark that, given n, a possible choice for K_m is $K_m \approx m/n$, $m = 1, \ldots, n$, by rounding m/n to the nearest integer. Then β is chosen quite close to the maximal value such that $a = \sum_{m=1}^{n} \beta^{n/K_m} < 1$. By our choice of K_m we approximately have

$$\sum_{m=1}^{n} \beta^{n/K_m} \approx \sum_{m=1}^{n} \beta^m \approx \frac{\beta}{1 - \beta}.$$

Thus, the optimal β will always be close to $1/2$.

4.4 Proof of Theorem 3.2

The proof of the Theorem 3.2 is practically the same as the proof of Theorem 2.1 in [28], because it depends only on set partitions without the algebraic structure of quaternions entering the picture. For more details we refer to [19].

4.5 Proof of Theorem 3.3

To prove our theorem we will need some modifications of Lemmas 4.5 and 4.6, in particular for the quantities $C_{\mathbb{H}}(\mathcal{A}, T)$ and $B_{\mathbb{H}}(\mathcal{C}, T)$ defined in (4.25) and (4.28), respectively. Hereby, we have to consider T as a random set modeled by (3.7), similarly as [28]. We will start first with (4.25) as follows.

Lemma 4.7 *For $\mathcal{A} \in P(2n, t)$ we have*

$$\mathbb{E}\left[C_{\mathbb{H}}(\mathcal{A}, T)\right]$$
$$\leq \sum_{s=2}^{n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{-2R} \#\{\mathcal{B} \in U(2n, s), \text{ rank } M(\mathcal{A}, \mathcal{B}) = R\},$$

where $M = M(\mathcal{A}, \mathcal{B})$ denote the $t \times s$ matrix whose entries are given by

$$M_{i,j} := |A_i \cap B_j| - |(A_i + 1) \cap B_j|, \quad 1 \leq i \leq t, \quad 1 \leq j \leq s.$$

Proof Using the linearity of the expectation value we get

$$\mathbb{E}\left[C_{\mathbb{H}}(\mathcal{C}, T)\right] = \mathbb{E}\left[\sum_{\substack{(k_1, l_1), \ldots, (k_{2n}, l_{2n}) \in \mathbb{Z}_\rho^2 \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}} \prod_{j=1}^{2n} \mathbb{1}_{\{(k_j, l_j) \in T\}} \right.$$
$$\left. \times \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A}(k_{r+1} - k_r)\right) \delta\left(\sum_{r \in A}(l_{r+1} \pm l_r)\right)\right]$$

$$= \sum_{\substack{(k_1, l_1), \ldots, (k_{2n}, l_{2n}) \in \mathbb{Z}_\rho^2 \\ (k_j, l_j) \neq (k_{j+1}, l_{j+1})}} \mathbb{E}\left[\prod_{j=1}^{2n} \mathbb{1}_{\{(k_j, l_j) \in T\}}\right]$$

$$\times \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A}(k_{r+1} - k_r)\right) \delta\left(\sum_{r \in A}(l_{r+1} \pm l_r)\right).$$

Hereby, $\mathbb{1}_{\{(k,l) \in T\}}$ denotes the characteristic function of the set T, i.e., it is 1 if and only if $(k, l) \in T$. The expression $\mathbb{E}\left[\prod_{j=1}^{2n} \mathbb{1}_{\{(k_j, l_j) \in T\}}\right]$ depends on on the number of (k_j, l_j) belonging to T. In this way, we can play again with partitions. If $((k_1, l_1), \ldots, (k_{2n}, l_{2n})) \in (\mathbb{Z}_\rho^2)^{2n}$ is a vector satisfying $(k_j, l_j) \neq (k_{j+1}, l_{j+1})$ then we can associate to it a partition $\mathcal{B} = (B_1, \ldots, B_s)$ of $\{1, \ldots, 2n\}$ such that j and j' are in the same set B_i if and only if $(k_j, l_j) = (k_{j'}, l_{j'})$. Obviously, j and $j + 1$ have to be contained in different blocks for all j due to the condition $(k_j, l_j) \neq (k_{j+1}, l_{j+1})$ (once again we use the convention that $2n + 1$ is identified with 1). In other words $\mathcal{B}$ has no adjacencies, i.e., $\mathcal{B} \in U(2n, s)$. Now if $\mathcal{B}$ has $|\mathcal{B}| = s$ blocks we use the probability model for T given by $\mathbb{P}((k, l) \in T) = \tau$, $0 < \tau < 1$, (see (3.7)) and stochastic independence to get

$$\mathbb{E}\left[\prod_{i=1}^{2n} \mathbb{1}_{\{(k_i, l_i) \in T\}}\right] = \mathbb{E}\left[\prod_{j=1}^{s} \mathbb{1}_{\{(k,l)_{B_j} \in T\}}\right] = \prod_{j=1}^{s} \mathbb{E}\left[\mathbb{1}_{\{(k,l)_{B_j} \in T\}}\right] = \tau^s, \quad (4.30)$$

where (unambiguously) $(k, l)_{B_j} = (k, l)_i$ if $i \in B_j$. By introducing the notation $\sigma_\mathcal{B}(r) = j$ if and only if $r \in B_j \in \mathcal{B}$ we can write

$$\mathbb{E}\left[C_\mathbb{H}(\mathcal{C}, T)\right]$$

$$= \sum_{s=2}^{n} \tau^s \sum_{\mathcal{B} \in U(2n, s)} \sum_{\substack{(k_1, l_1), \ldots, (k_{2n}, l_{2n}) \in \mathbb{Z}_\rho^2 \\ (k_i, l_i) \text{ p.w. different}}} \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A}(k_{\sigma_\mathcal{B}(r+1)} - k_{\sigma_\mathcal{B}(r)})\right).$$

Clearly, we have $\prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A}(k_{\sigma_\mathcal{B}(r+1)} - k_{\sigma_\mathcal{B}(r)})\right) = 1$ and $\prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A}(l_{\sigma_\mathcal{B}(r+1)} - l_{\sigma_\mathcal{B}(r)})\right) = 1$ if and only if, respectively,

$$\sum_{r \in A}(k_{\sigma_\mathcal{B}(r+1)} - k_{\sigma_\mathcal{B}(r)}) = 0, \quad \text{for all } A \in \mathcal{A} \tag{4.31}$$

and

$$\sum_{r \in A}(l_{\sigma_\mathcal{B}(r+1)} \pm l_{\sigma_\mathcal{B}(r)}) = 0, \quad \text{for all } A \in \mathcal{A}. \tag{4.32}$$

Otherwise, the terms are 0. For $j \in \{1, \ldots, s\}$ the term (k_j, l_j) appears $|A_i \cap B_j|$ times in the terms $(k_{\sigma_\mathcal{B}(r)}, l_{\sigma_\mathcal{B}(r)})$ when r runs through $A_i \in \mathcal{A}$. Let $M = M(\mathcal{A}, \mathcal{B})$ denote the $t \times s$ matrix whose entries are given by

$$M_{i,j} := |A_i \cap B_j| - |(A_i + 1) \cap B_j|, \quad 1 \le i \le t, 1 \le j \le s.$$

Then (4.31) and (4.32) are satisfied if and only if $((k_1, l_1), \ldots, (k_s, l_s)) \in (\mathbb{Z}_\rho^2)^s$ belongs to the kernel of $M(\mathcal{A}, \mathcal{B})$. Thus, if the rank of $M(\mathcal{A}, \mathcal{B})$ equals R then the number of vectors $(k_1, l_1), \ldots, (k_s, l_s) \in (\mathbb{Z}_\rho^2)^s$ for which (4.31) and (4.32) are satisfied can be bounded by $((2\rho + 1)^2)^{s-R}$. Here we even neglected the condition that $(k_1, l_1), \ldots, (k_s, l_s)$ should be pairwise different. So finally we obtain

$$\mathbb{E}\left[C_{\mathbb{H}}(\mathcal{C}, T)\right] \le \sum_{s=2}^{n} \tau^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{2(s-R)} \#\{\mathcal{B} \in U(n, s), \text{ rank } M(\mathcal{A}, \mathcal{B}) = R\}$$

$$= \sum_{s=2}^{n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{-2R} \#\{\mathcal{B} \in U(n, s), \text{rank } M(\mathcal{A}, \mathcal{B}) = R\},$$

where we substituted $\mathbb{E}|T| = \tau D$. $\qquad\square$

Since $E = E_X E_T$ by Fubini's theorem and stochastic independence of T and X the previous result yields together with Lemma 4.5

$$\mathbb{E}\left[\left\|H_0^n\right\|_F^2\right] \le \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A} \in P(2n,t)} \sum_{s=2}^{n} (\mathbb{E}|T|)^s$$

$$\times \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{-2R} \#\{\mathcal{B} \in U(n, s), \text{ rank } M(\mathcal{A}, \mathcal{B}) = R\}$$

$$\le \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{s=2}^{n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{-2R} Q(2n, t, s, R)$$

$$= N^{2n} W(n, N, \mathbb{E}|T|, D).$$

Hereby, $Q(2n, t, s, R) := \#\{(\mathcal{A}, \mathcal{B}) : \mathcal{A} \in P(n, t), \mathcal{B} \in U(n, s), \text{rank} M(\mathcal{A}, \mathcal{B}) = R\}$ and

$$W(n, N, \mathbb{E}|T|, \rho)$$
$$:= N^{-2n} \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{s=2}^{n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho + 1)^{-2R} Q(2n, t, s, R).$$

Applying Markov's inequality we obtain

$$\mathbb{P}\left(\left\|(N^{-1}H_0)^n\right\|_F \ge \kappa\right) \le N^{-2n} \kappa^{-2} \mathbb{E}\left[\left\|H_0^n\right\|_F^2\right] \le \kappa^{-2} W(n, N, \mathbb{E}|T|, \rho).$$

Obviously, like in the proof of Theorem 3.3 $\mathcal{F}_{TX}$ is injective if $\left\|(N^{-1}H_0)^n\right\|_F < 1$.

As in the case where T is not random we have to estimate $\mathbb{P}(E_{k,l})$. This has to be done like in Lemma 4.6, i.e., we need to estimate the expectation value of $B_{\mathbb{H}}(\mathcal{C}, T)$ defined in (4.28).

Lemma 4.8 *Suppose* $\mathcal{A} \in P(2Km, t)$ *then we have*

$$
\mathbb{E}\left[B_{\mathbb{H}}(\mathcal{C}, T)\right]
$$
$$
\leq \sum_{s=1}^{2Km} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} (2\rho+1)^{-2R} \#\{\mathcal{B} \in U^*(2K, m, s), \quad \text{rank } L(\mathcal{A}, \mathcal{B}) = R\},
$$

where similar to [28] $U^*(2K, m, s)$ *denotes all partitions of* $[K] \times [m]$ *such that* (r, p)
and $(r, p + 1)$ *are not contained in the same block.* $L = L(\mathcal{A}, \mathcal{B})$ *denote the* $t \times s$
matrix whose entries are given by

$$
L_{i,j} := \sum_{(p,u)\in A_i \cap B_j} (-1)^p - \sum_{(p,u)\in (A_i-1)\cap B_j} (-1)^p.
$$

Proof In the same way as in the proof of Lemma 4.7 we can consider

$$
\mathbb{E}\left[B_{\mathbb{H}}(\mathcal{C}, T)\right] = \sum_{\substack{(k_1,l_1),\ldots,(k_{2n},l_{2n}) \in \mathbb{Z}_\rho^2 \\ (k_j,l_j) \neq (k_{j+1},l_{j+1})}} \mathbb{E}\left[\prod_{(p,j)\in[2K]\times[m]} \mathbb{1}_{\{(k_j,l_j)\in T\}}\right]
$$
$$
\times \prod_{A\in\mathcal{A}} \delta\left(\sum_{(r,p)\in A} \alpha(k_{r+1}^{(p)} - k_r^{(p)})\right) \delta\left(\sum_{(r,p)\in A} \alpha(l_{r+1}^{(p)} - l_r^{(p)})\right).
$$

Hereby, $\mathbb{E}[\prod_{(p,j)\in[2K]\times[m]} \mathbb{1}_{\{(k_j^{(p)},l_j^{(p)})\in T\}}]$ depends only on the number of different
$(k_r^{(p)}, l_r^{(p)})$'s. Therefore, if $((k_1^{(1)},l_1^{(1)}),\ldots,(k_m^{(2K)},l_m^{(2K)})) \in (\mathbb{Z}_\rho^2)^{(2Km)}$ satisfies

$$
\left(k_j^{(p)}, l_j^{(p)}\right) \neq \left(k_{j+1}^{(p)}, l_{j+1}^{(p)}\right) \quad \text{for all } j \in [m], \quad p \in [2K] \tag{4.33}
$$

we can identify it with partition $\mathcal{B} = (B_1, \ldots, B_s)$ of $[2K]$ such that (p, j) and (p', j')
are belonging to the same block if and only if $k_j^{(p)} = k_{j'}^{(p')}$. This implies that $\mathcal{B}$ belongs
to $U^*(2K, m, s)$ since (p, j) and $(p, j + 1)$ cannot be in the same block due to the
condition (4.33). Now $\mathcal{B}$ has s sets which means that we have s different values of
$(k_j^{(p)}, l_j^{(p)})$ and

$$
\mathbb{E}\left[\prod_{(p,j)\in[2K]\times[m]} \mathbb{1}_{\{(k_j^{(p)},l_j^{(p)})\in T\}}\right] = \tau^s
$$

as in (4.30). Using again the notation $\sigma_\mathcal{B}(p, j) = i$ if $(p, j) \in B_i \in \mathcal{B}$ and $\sigma(p, 0) = 0$
we get

$$\mathbb{E}[B_{\mathbb{H}}(\mathcal{C}, T)] = \sum_{s=1}^{2n} \tau^s \sum_{\mathcal{B} \in U^*(2K, m, s)} \sum_{\substack{(k_1, l_1), \ldots, (k_{2n}, l_{2n}) \in \mathbb{Z}_\rho^2 \\ (k_i, l_i) \text{ p.w. different}}}$$

$$\times \prod_{A \in \mathcal{A}} \delta \left(\sum_{(p,j) \in A} \alpha(k_{\rho \mathcal{B}(p,j+1)} - k_{\rho \mathcal{B}(p,j)}) \right) \delta \left(\sum_{(p,j) \in A} \alpha(l_{\rho \mathcal{B}(p,j+1)} \pm l_{\rho \mathcal{B}(p,j)}) \right).$$

The term $\prod_{A \in \mathcal{A}} \delta \left(\sum_{(p,j) \in A} \alpha(l_{\rho \mathcal{B}(p,j+1)} - l_{\rho \mathcal{B}(p,j)}) \right) \delta \left(\sum_{(p,j) \in A} \alpha(l_{\rho \mathcal{B}(p,j+1)} \pm l_{\rho \mathcal{B}(p,j)}) \right)$ is non-zero if and only if

$$\sum_{(p,j) \in A} \alpha(k_{\rho \mathcal{B}(p,j+1)} - k_{\rho \mathcal{B}(p,j)}) = 0 \quad \text{for all } A \in \mathcal{A}$$

and

$$\sum_{(p,j) \in A} \alpha(l_{\rho \mathcal{B}(p,j+1)} \pm l_{\rho \mathcal{B}(p,j)}) = 0 \quad \text{for all } A \in \mathcal{A}.$$

Similar to [28] this leads to the estimate

$$\mathbb{E}[B(\mathcal{A}, T)] \leq \sum_{s=1}^{2Km} \tau^s \sum_{R=0}^{\min\{s,t\}} D^{s-R} \#\{\mathcal{B} \in U^*(2K, m, s), \text{ rank } L(\mathcal{A}, \mathcal{B}) = R\}.$$

Since $\mathbb{E}|T| = \tau D$ this proves the lemma. $\qquad \square$

Using our Lemma 4.6 the previous result gives us

$$\mathbb{E}\left[|((H R_T)^m \sigma)_{k,l}|^{2K} \right]$$
$$\leq \sum_{t=1}^{\min\{Km, N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2Km} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}} Q^*(2K, m, t, s, R) D^{-R}$$
$$= N^{2Km} Z(K, m, N, \mathbb{E}|T|, D)$$

where $Q^*(2K, m, t, s, R)$ are numbers defined by

$$Q^*(2K, m, t, s, R)$$
$$:= \#\{(\mathcal{A}, \mathcal{B}), \mathcal{A} \in P(2Km, t), \mathcal{B} \in U^*(2K, m, s), \text{rank } L(\mathcal{A}, \mathcal{B}) = R\}.$$

From (4.27) we get

$$\mathbb{P}(E_{k,l}) \leq \beta^{-2n} \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, D).$$

Finally, again let $\mathbb{P}(\text{failure})$ denote the probability that exact reconstruction of f fails. By Lemma 4.1, expressions (4.16) and (4.17), and using the fact $\{\mathcal{F}_{TX} \text{ is not injective}\} \subset \{\left\| (N^{-1}H_0)^n \right\|_F \geq \kappa\}$ we finally obtain

$$\mathbb{P}(\text{failure}) \leq \mathbb{P}\left(\{\mathcal{F}_{TX}\{\text{is not injective}\} \cup \{\sup_{(k,l) \in T^c} |P_{k,l}| \geq 1\} \right)$$

$$\leq \sum_{(k,l) \in \mathbb{Z}_\rho^2} \mathbb{P}(E_{k,l}) + \mathbb{P}\left(\left\| \left(N^{-1}H_0 \right)^n \right\| \geq \kappa \right) + \mathbb{P}(|T| \geq (\alpha+1)\mathbb{E}|T|)$$

$$\leq D\beta^{-2n} \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, D) + \kappa^{-2} W(n, N, \mathbb{E}|T|, D)$$

$$+ \exp\left(-\frac{3\alpha^2}{6+2\alpha} \mathbb{E}|T| \right)$$

under the conditions (see also (4.15))

$$a_1 = a = \sum_{m=1}^{n} \beta^{n/K_m} < 1, \quad a_2 + a_1 = 1, \text{ i.e., } a_2 = 1 - a,$$

$$\frac{\kappa}{1-\kappa} \leq \frac{a_2}{1+a_1}((\alpha+1)\mathbb{E}|T|)^{-3/2} = \frac{1-a}{1+a}((\alpha+1)\mathbb{E}|T|)^{-3/2}.$$

This proves Theorem 3.3.

5 Applications

Although there exists a large variety of applications of quaternionic signals, in this section we will restrict ourselves to the case of color-encoded images represented by quaternions, i.e., the R-,G-, and B-components of the image are represented by the **I**-, **J**-, **K**-components of a given quaternion, respectively. In what follows we present some numerical experiments. Namely, we reconstruct a given signal by means of the ℓ_1-minimization problem

$$\min \|x\|_1 = \sum_{s=1}^{N} |x_s|$$
$$\text{s.t.} \quad \Phi x = y, \tag{5.1}$$

where $\Phi \in \mathbb{H}^{M \times N}$ is our quaternionic sampling matrix, $y \in \mathbb{H}^{M \times 1}$ is the vector consisting of chosen random pixels from the given image under the assumption that the signal vector $x \in \mathbb{H}^{N \times 1}$ is k-sparse, i.e., at most k entries of x are non-zero. The support of x is the set of indexes corresponding to the non-zero entries, supp $x = \{s \in \{1, \ldots, N\} : x_s \neq 0\}$. We aim to illustrate that an effective signal reconstruction is possible even if using only a small amount of the signal information.

For the implementation of the ℓ_1-minimization algorithm we use the Matlab toolbox ℓ1-Magic [9].

Since a quaternion is defined as follows

$$q = R(q) + \mathbf{I}I(q) + \mathbf{J}J(q) + \mathbf{K}K(q)$$

where $R(q)$ is the real part of the quaternion and $I(q)$, $J(q)$ and $K(q)$ are its three imaginary components we have for vectors and matrices of quaternions the decomposition

$$\mathbf{v} = R(\mathbf{v}) + \mathbf{I}I(\mathbf{v}) + \mathbf{J}J(\mathbf{v}) + \mathbf{K}K(\mathbf{v})$$
$$\mathbf{M} = R(\mathbf{M}) + \mathbf{I}I(\mathbf{M}) + \mathbf{J}J(\mathbf{M}) + \mathbf{K}K(\mathbf{M}).$$

Since $\ell 1$-Magic works only with real-valued vectors we need to modify our quaternion-valued system. We rewrite the quaternionic multiplication $(\alpha + \mathbf{I}\beta + \mathbf{J}\gamma + \mathbf{K}\delta)(x + \mathbf{I}y + \mathbf{J}v + \mathbf{K}w) = a + \mathbf{I}b + \mathbf{J}c + \mathbf{K}d$ as a matrix-vector multiplication, i.e.,

$$\begin{pmatrix} \alpha & \beta & \gamma & \delta \\ -\beta & \alpha & -\delta & \gamma \\ -\gamma & \delta & \alpha & -\beta \\ -\delta & -\gamma & \beta & \alpha \end{pmatrix} \begin{pmatrix} x \\ y \\ v \\ w \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

This allows us to rewrite the quaternionic linear system (5.1) in the form $\tilde{y} = \mathcal{M}\tilde{x}$ with

$$\mathcal{M} = \begin{pmatrix} R(\Phi) & -I(\Phi) & J(\Phi) & -K(\Phi) \\ -I(\Phi) & R(\Phi) & -K(\Phi) & J(\Phi) \\ -J(\Phi) & K(\Phi) & R(\Phi) & -I(\Phi) \\ -K(\Phi) & -J(\Phi) & I(\Phi) & R(\Phi) \end{pmatrix},$$

Fig. 1 Lena—original image

Fig. 2 Lena image—reconstructed from 40,000 pixels
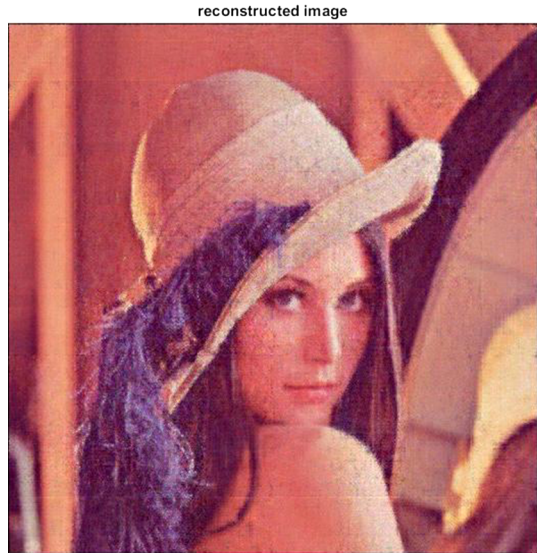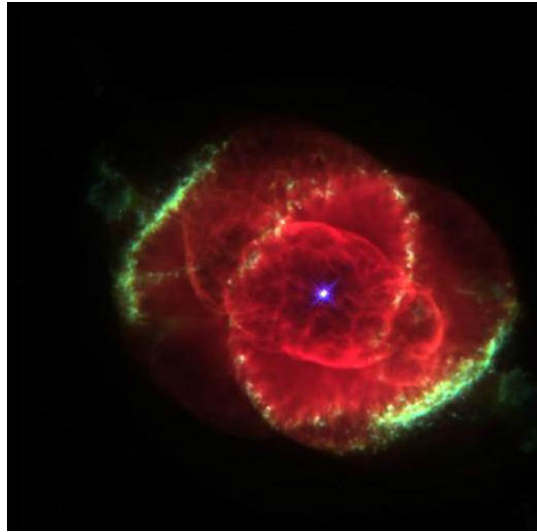

reconstructed image

Fig. 3 Galaxy—original image

$$\tilde{x} = \begin{pmatrix} R(x) \\ I(x) \\ J(x) \\ K(x) \end{pmatrix}, \quad \text{and} \quad \tilde{y} = \begin{pmatrix} R(y) \\ I(y) \\ J(y) \\ K(y) \end{pmatrix}.$$

In our implementation we use the sampling matrix in its explicit form. This leads to large requirements in terms of memory since we deal with a matrix of dimension $M \times N^2$. To be able to work with such large matrices, we used a strategy of dividing the image into 8×8-blocks followed by an individual reconstruction of each block. The

Fig. 4 Galaxy
image—reconstructed from
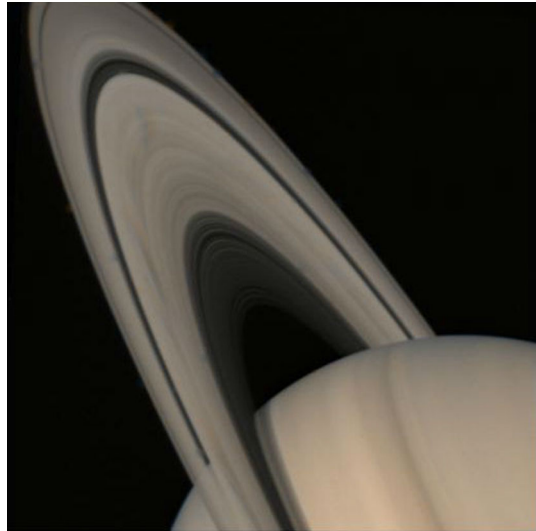40,000 pixels

reconstructed image
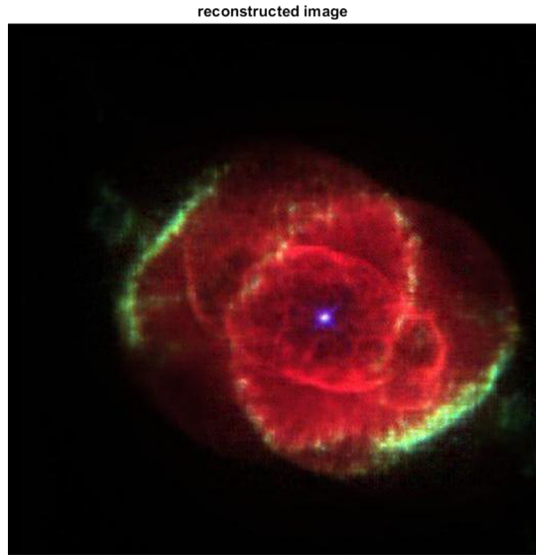
Fig. 5 Saturn—original image

sampling matrices for the blocks themselves are constructed by using DCT (Discrete Cosine Transform) and DST (Discrete Sine Transform). Since we study each block individually, we obviously get the additional problem of reassembling them. To this end we introduce reflexive boundary conditions on each block.

We performed the calculations on a computer with Intel(R) Core(TM) i7-4790U CPU 3.60 GHz, RAM 16 GB, Windows 8.1, OS 64-bit(win64) and running Matlab R2012b.

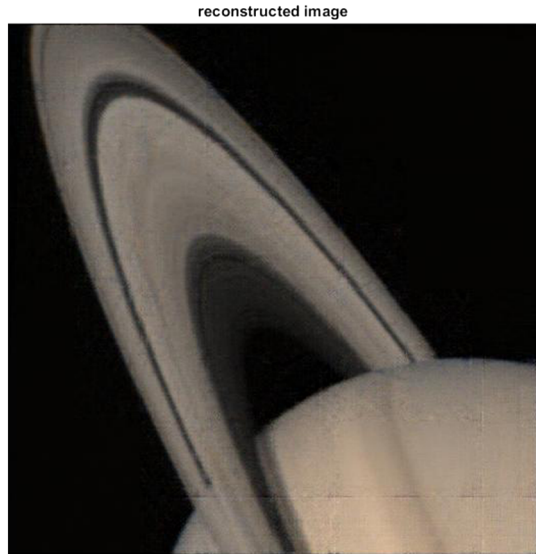**Fig. 6** Saturn
image—reconstructed from
40,000 pixels

Fig. 7 Chips—original image

For our examples we choose the following images: Lena (Fig. 1), Galaxy (Fig. 3), and the Saturn rings (Fig. 5), each with $N^2 = 262,144$ pixels (512×512). Lena image was chosen as one of the standard examples in image processing. Galaxy and Saturn have large black parts which of course can affect the reconstruction when by random choice mainly black pixels are chosen. Furthermore, Saturn is also easily visible as a gray-scale image. For the reconstruction we use 40,000 randomly chosen pixels ($M = 625$ samples in each block) which corresponds to $\approx 15.26\%$ of the total information. The reconstructed images can be seen in Figs. 2, 4, and 6, respectively.

Fig. 8 Reconstructed Chips
image using information of
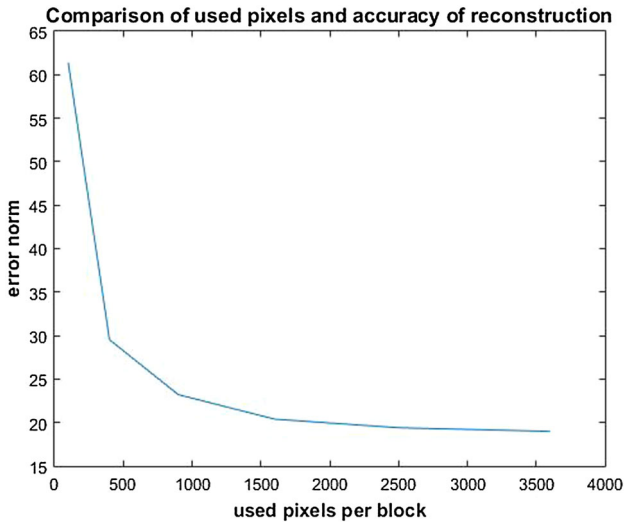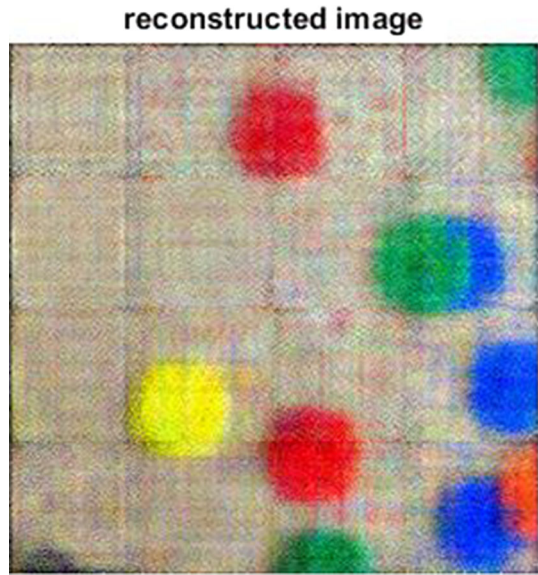10,000 pixels

reconstructed image

Fig. 9 Comparison between number of random samples and reconstruction error

Another particularly interesting example is the chips image (Fig. 7). This image
combines textures and sharp edges and represents a kind of worst case scenario. In
this example, the image has $N^2 = 65,536$ pixels (256×256). For the reconstruction
we use 10,000 pixels ($M = 625$ samples in each block). Since the overall size of
the image is smaller we just use a decomposition into 4×4 blocks. Under the same
conditions as above ($\approx 15.26\%$ of the original pixels where taken as random samples)
we obtain our reconstruction (see Fig. 8).

While the image is still blurred (as expected) it still contains all relevant details.

To show the error reduction we provide Fig. 9 which shows the relation between the number of used pixels and the error given by the l_2-norm of the matrix containing the difference in the pixels between the original chips image (Fig. 7) and the reconstructed image. The figure shows an exponential decay in the error. It also shows why we opted to use the values of 10,000 pixels in the reconstruction since it is an interesting compromise between numbers of pixels and quality of the reconstructed image.

References

1. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)
2. Bass, R.F., Gröchenig, K.: Random sampling of multivariate trigonometric polynomials. SIAM J. Math. Anal. **36**(3), 773–795 (2004)
3. Bas, P., Le Bihan, N., Chassery, J.M.: Color image watermarking using quaternion Fourier transform. In: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, pp. 521–524. ICASSP, Hong-Kong (2003)
4. Batard, T., Berthier, M., Saint-Jean, C.: Clifford-Fourier transform for color image processing. Geometric Algebra Computing, pp. 135–162. Springer, London (2010)
5. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
6. Bülow, T.: Hypercomplex spectral signal representations for the processing and analysis of images, Ph.D. Thesis, University of Kiel (1999)
7. Bülow, T., Sommer, G.: Hypercomplex signals a novel extension of the analytic signal to the multidimensional case. IEEE Trans. Signal Process. **49**(11), 2844–2852 (2001)
8. Candés, E.: A restricted isometry property and it implications for Compressed Sensing. Comptes Rendus Math. **346**, 589–592 (2008)
9. Candés, E., Romberg, J.: ℓ1-MAGIC: Recovery of Sparse Signals via Convex Programming (2005). http://www.acm.caltech.edu/l1magic
10. Candés, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)
11. Carré, P., Berthier, M.: Color Representation and Processes with Clifford Algebra. In: Advanced Color Image Processing and Analysis, pp. 147–179. Springer, New York (2013)
12. Chen, S., Donoho, D.L.: Basis Pursuit: Technical report, Department of Statistics, Stanford University (1994). http://statweb.stanford.edu/donoho/Reports/1994/asilomar
13. DeVore, R.A.: Nonlinear approximation. Acta Numer. **7**, 51–150 (1998)
14. Donoho, D.L.: Compressed Sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)
15. Dubey, V.R.: Quaternion Fourier Transform for Colour Images. Int. J. Comput. Sci. Inf. Technol. **5**(3), 4411–4416 (2014)
16. Ell, T.A.: Quaternion-fourier transforms for analysis of two-dimensional linear time-invariant partial differential systems. In: Proceeding of the 32nd Conference on Decision and Control, pp. 1830–1841. San Antonio, Texas (1993)
17. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Birkhäuser, Basel (2013)
18. Gohberg, I.: Continuous and Discrete Fourier Transforms, Extension Problems and Wiener–Hopf Equations, vol. Operator Theory: Advances and Applications. Birkhüuser, Basel (2012)
19. Gomes, N.R.: Compressed Sensing in Clifford Analysis, Ph.D. thesis, Universidade de Aveiro (2015)
20. Gürlebeck, K., Sprößig, W.: Quaternionic and Clifford calculus for physicists and engineers, Mathematical Methods in Practice. Wiley, Chichester (1997)
21. Hitzer, E.: Quaternion Fourier transform on quaternion fields and generalizations. Adv. Appl. Clifford Algebra **17**(3), 497–517 (2007)
22. Lebedev, V.I.: An Introduction to Functional Analysis in Computational Mathematics: An Introduction. Birkhäuser, Boston (2012)
23. Lee, S.: Image representation using 2D Gabor wavelets. IEEE Trans. Pattern Anal. Mach. Intell. **18**(10), 959–971 (1996)

24. Mallat, S., Zhang, Z.: Matching Pursuit With Time-Frequency Dictionaries. IEEE Trans. Signal Process. **41**(12), 3397–3415 (1993)
25. Mawardi, B., Hitzer, E., Hayashi, A., Ashino, R.: An Uncertainty Principle for Quaternion Fourier Transform. Comput. Math. Appl. **56**(9), 2398–2410 (2008)
26. Nguyen, H.T., Do, M.N.: Robust multichannel sampling. In: 15th Proc. IEEE Int. Conf. Image Process, pp. 653–656 (2008)
27. Pei, S.C., Ding, J.J., Chang, J.H.: Efficient implementation of quaternion Fourier transform, convolution, and correlation by 2-D complex FFT. IEEE Trans. Signal Process. **49**(11), 2783–2797 (2001)
28. Rauhut, H.: Random sampling of sparse trigonometric polynomials. Appl. Comput. Harmon. Anal. **22**(1), 16–42 (2007)
29. Sangwine, S.J.: The discrete quaternion fourier transform. In: Proc. 6th Int. Conf. Image Process. Appl. vol. 2, pp. 14–17 (1997)
30. Sangwine, S.J., Ell, T.A.: Hypercomplex Fourier transforms of color images. IEEE Trans. Signal Process. **16**(1), 22–35 (2007)
31. Sommen, F.: Hypercomplex Fourier and Laplace transforms I. Hypercomplex Fourier and Laplace transforms I. Ill. J. Math. **341**(9), 561566 (1982)
32. Sommen, F., Sprößig, W.: Introduction to Clifford analysis, Mathematical Methods in the Applied Sciences, vol. 25, pp. 16–18. John Wiley & Sons, Ltd., New York (2002)
33. Sudbery, A.: Quaternionic Analysis. Math. Proc. Camb. Philos. Soc. **85**, 199–225 (1979)
34. Teschke, G.: Multi-frame representations in linear inverse problems with mixed multiconstraints. Appl. Comput. Harmon. Anal. **22**(1), 43–60 (2007)
35. Todd, A., Ell, A., Le Bihan, N., Sangwine, S.J.: Quaternion Fourier Transforms for Signal and Image Processing. Wiley-ISTE, New York (2014)
36. Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theor. **50**(10), 2231–2242 (2004)