# Analysis of Signal Reconstruction Algorithms Based on Consistency Constraints

By

Chang-Hsin Lee

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematics

August 11, 2017

Nashville, Tennessee

Approved:

Alexander Powell, Ph.D.

Akram Aldroubi, Ph.D.

Brett Byram, Ph.D.

Doug Hardin, Ph.D.

Mike Neamtu, Ph.D.

*To Mom and Dad*

## ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Alexander Powell, for his guidance and support in my career. It was a privilege to work and to have many conversations and discussions with Alex, who had introduced me to areas of mathematics that I would not have known otherwise. I would also like to thank the remaining members of my Ph.D. committee, Dr. Akram Aldroubi, Dr. Brett Byram, Dr. Doug Hardin, and Dr. Mike Neamtu, for taking the time to review my dissertation.

It was fun to know the graduate students and staff in the mathematics department at Vanderbilt, who had made the department a welcoming environment and had been tremendously helpful in many different ways.

Additionally, I would like to thank all the members of the Church in Nashville, who encouraged and supported me in many ways from start to finish.

Finally, I would like to thank my family. I am eternally thankful to my parents for supporting my studies, and none of this work would have been possible without them. I am also thankful to my uncle, John Sun, my aunts, Hsiu-Er Lin and Dolly Chao, and their families for their help and support throughout my time in Nashville.

TABLE OF CONTENTS

CHAPTER 1

# INTRODUCTION

## 1.1 The reconstruction problem

A fundamental problem in signal processing called *signal reconstruction*, or *signal recovery*, is the determination of a signal $x \in \mathbb{R}^d$ from a sequence of samples $\{y_n\}$ obtained from $x$. Studies on the reconstruction problem have resulted in major breakthroughs in technology in the past century, and practical solutions to the problem are still essential in the advancement of fields such as image processing and speech recognition. In this work, we will consider the reconstruction problem in the framework that the samples $y_n \in \mathbb{R}$ are *linear measurements*, which are defined as inner products $y_n = \langle x, \varphi_n \rangle$ between $x$ and a set of measurement vectors $\{\varphi_n\} \subset \mathbb{R}^d$. In particular, our focus will be on algorithms that seek to either recover the true signal $x$ or produce an estimate $\widetilde{x} \in \mathbb{R}^d$ of $x$ using linear measurements. Many questions that are of both practical and theoretical interest then arise naturally under this setting:

- Which algorithms allow one to reconstruct $x$ perfectly, if possible?
- If one cannot exactly recover $x$ from $\{y_n\}$, then how precise can the estimate $\widetilde{x}$ be?
- How does noise or quantization of measurements affect the algorithms?
- Can one solve the problem if the measurements come in a streaming fashion?

In this chapter, we will quickly review some developments centered around the aforementioned questions in the field of digital signal processing.

## 1.2 Reconstruction from noiseless measurements

Suppose that the measurements are free from noise, then solving the reconstruction problem from linear measurements $y_n = \langle x, \varphi_n \rangle$ is equivalent to solving a system of linear equations in the form of

$$Ax = y, \tag{1.1}$$

where $x \in \mathbb{R}^d$, $y = (y_1, y_2, ..., y_N)^T \in \mathbb{R}^N$, and $A \in \mathbb{M}_{N,d}(\mathbb{R})$ is an $N$ by $d$ matrix whose $j$-th row is $\varphi_j \in \mathbb{R}^d$. The most common approach to produce an estimate $\widetilde{x}$ from (1.1) is by setting

$$\widetilde{x} = \arg\min_{z \in \mathbb{R}^d} \|Az - y\|_2. \tag{1.2}$$

In other words, a solution is obtained by minimizing a cost function $f(z) = \|Az - y\|_2$. It is well-known that the solution to (1.2) is $\widetilde{x} = A^\dagger y$, where $A^\dagger$ is the Moore-Penrose pseudoinverse of A. However, because most implementations of the pseudoinverse rely on calculating singular value decompositions, computing $\widetilde{x}$ from the pseudoinverse quickly becomes a rather expensive task to perform as the number of measurements increases. Therefore, practical solutions to (1.2) are often computed iteratively using a family of algorithms known as *gradient descent* [24].

The system of linear equations (1.1) is called an overdetermined system if $N > d$, and an underdetermined system if $N < d$. Other than gradient descent, there are other iterative algorithms that solve (1.1), which one may prefer over gradient descent for practical considerations such as better convergence rates or support for streaming data. For example, there is also a family of algorithms called *Projection onto Convex Sets* (POCS) methods, also known as *Alternating Projection* methods, that solves (1.1) by iteratively projecting the estimate onto the solution space of each equation in the system. POCS methods have found applications in computerized tomography [10], and here we will outline a particular method from the POCS family called the Kaczmarz method [14]. Given an arbitrary initial estimate $x_0 \in \mathbb{R}^d$, the Kaczmarz algorithm iteratively produces a new estimate $x_{j+1} \in \mathbb{R}^d$

from the previous estimate $x_j \in \mathbb{R}^d$ by

$$x_{j+1} = x_j + \frac{y_i - \langle \varphi_i, x_j \rangle}{\|\varphi_i\|^2} \varphi_i, \tag{1.3}$$

where $i = j \mod N+1$ and $\varphi_i$ is the $i$-th row of $A$. The estimates $\{x_n\}$ produced by the Kaczmarz method have long been known to converge to the least square solution (1.2), but bounds on the convergence rate remained unclear. Recently, a randomized version of the Kaczmarz algorithm was proposed in [26] that works as follows: instead of choosing $\varphi_i$ in a cyclic manner, each row $\varphi_i$ is sampled randomly with a probability proportional to its Euclidean norm $\|\varphi_i\|$. The randomized Kaczmarz algorithm with row sampling was shown to converge exponentially fast for consistent overdetermined systems in [26], and further analysis provided bounds on almost sure exponential convergence with random measurements in [1]. Moreover, in [16], the randomized Kaczmarz algorithm with row sampling is found to be a special case of stochastic gradient descent [21].

When one has an underdetermined system in (1.1), regularization can be used to find an estimate $\widetilde{x}$ in the solution space, but it is unlikely to recover the true signal $x$. However, much work has been done in the last two decades in the field of compressed sensing to find sparse solutions for underdetermined systems by exploiting structures of the matrix $A$ and by making use of L1-optimization techniques. Such solvers of underdetermined systems fall outside the scope of our discussion, and we refer the readers to [6] for a systematic overview of the theory of compressed sensing.

### 1.3 The noise model

When noise is present, (1.1) becomes

$$Ax = y + \varepsilon, \tag{1.4}$$

where $\varepsilon \in \mathbb{R}^N$ is a random vector. The presence of noise in (1.4) means one cannot solve for $x \in \mathbb{R}^d$ exactly and has to seek instead an estimate $\widetilde{x} \in \mathbb{R}^d$ of $x$. Besides introducing uncertainty to estimates, noise also affects the stability of algorithms. As a result, many algorithms in Section 1.2 no longer have a known guarantee for general $A$, and additional conditions are required. One common constraint on $A$ in a noisy situation is to require an upper bound on the condition number $\text{cond}(A)$. If $\{\varphi_n\}$ happens to form a *frame*, which we will define in the next section, then bounding the $\text{cond}(A)$ can also be thought of as bounding the ratio between optimal upper frame bound and lower frame bound as defined in (1.6).

A natural source of error is the round-off error inherited from the precision of measuring instruments. In the language of digital signal processing, the errors are the result of *quantization* and are often modeled by *dithering*. In this section, we will describe the idea of quantization and a procedure called subtractive dithering [22, 23] that allows one to model the quantization error as noises uniformly distributed on $[-\delta, \delta]$. Let $\mathscr{A} \subset \mathbb{R}$ be a finite set, which we call the *quantization alphabet*, and define the scalar quantizers $\mathscr{Q}$ associated with $\mathscr{A}$ by

$$\forall u \in \mathbb{R},\ \mathscr{Q}(u) = \mathscr{Q}_{\mathscr{A}}(u) = \text{argmin}_{q \in \mathscr{A}} |u - q|.$$

Let $K \in \mathbb{N}$ and define a $2K$-level *midrise* quantization alphabet with stepsize $2\delta$ by

$$A_K^{\delta} = \{-(2K-1)\delta, \ldots, -3\delta, -\delta, \delta, 3\delta, \ldots, (2K-1)\delta\}.$$

In the case of subtractive dithering, $\delta > 0$ is half of the quantization step size and is known *a priori*. Subtractive dithering then works as follows. Given an i.i.d. random noise sequence $\{w_n\}_{n=1}^N$ known as a dither, we use $A_K^{\delta}$ to quantize the dithered noise sequence $\langle x, \varphi_n \rangle + w_n$ to obtain a quantized sequence $\widetilde{q}_n = Q_{\delta}(\langle x, \varphi_n \rangle + w_n)$. The dithers are then subtracted from the quantized sequence to give $q_n := \widetilde{q}_n - w_n$. With appropriate dither $\{w_n\}_{n=1}^N$, the errors $\varepsilon_n = \langle x, \varphi_n \rangle - q_n$ are i.i.d. uniform random variables on $[-\delta, \delta]$. For this reason, we will adopt a uniform noise model throughout our work.

## 1.4 Frames

Generating a sequence of linear measurements $\{y_n\} \subset \mathbb{R}$ from a signal $x \in \mathbb{R}^d$ is equivalent to representing $x$ by $y = (y_1, \ldots, y_N)^T \in \mathbb{R}^N$ using a set of vectors $\{\varphi_n\}$ by setting $y_n = \langle x, \varphi_n \rangle$. The ability to represent a vector by a given collection of vectors is of central importance in linear algebra, for which the consideration often starts with *basis vectors* that provide a unique representation of vectors in a vector space. For simplicity, we will restrict our discussion to only vectors in Hilbert spaces in this section. A countable sequence $\{\varphi_n\} \subset \mathscr{H}$ is a basis for a Hilbert space $\mathscr{H}$ if, for all $x \in \mathscr{H}$, there exist a unique sequence of scalars $a_n = a_n(x)$ such that

$$x = \sum_n a_n \varphi_n, \tag{1.5}$$

where the series converges in the norm of $\mathscr{H}$. From the definition, if one has access to the measurement vectors $\varphi_n$ and corresponding $a_n$, then one can estimate $x$ by considering the partial sums $x_j = \sum_{n=1}^{j} a_n \varphi_n$. However, the requirement that (1.5) being a unique representation means every coefficient $a_n$ is necessary in reconstructing $x$. Therefore, rather than finding a unique representation with basis vectors, one may prefer using a set of vectors called *frames* that provides a redundant representation of the signal $x$. As our work in the following chapters focused on solving overdetermined systems, we will now briefly review the definition and basic properties of a frame. For an introduction to the theory of bases and frames, we refer the readers to [9].

**Definition 1.4.1.** *Let $\mathscr{H}$ be a Hilbert space and $V = \{\varphi_n\} \subset \mathscr{H}$ be a sequence in $\mathscr{H}$. We say $V$ is a frame for $\mathscr{H}$ if there exist numbers $0 < A \leq B < \infty$ such that for all $x \in \mathscr{H}$,*

$$A\|x\|^2 \leq \sum_n |\langle x, \varphi_n \rangle|^2 \leq B\|x\|^2. \tag{1.6}$$

*Moreover, $V$ is called a tight frame if $A = B$ in* (1.6).

The next definition helps to simplify symbols involved in reconstructing $x$ from $\varphi_n$.

**Definition 1.4.2.** *Given a frame $V = \{\varphi_n\} \subset \mathscr{H}$, the frame operator $S : H \to H$ is defined by*

$$Sx = \sum_n \langle x, \varphi_n \rangle \varphi_n, \quad x \in \mathscr{H}.$$

From Theorem 8.13 in [9], if we have a frame $V \subset \mathscr{H}$ with frame bounds $0 < A \leq B < \infty$, the associated frame operator S is a topological isomorphism of $H$ onto itself, and $\{\mathrm{S}^{-1}\varphi_n\}$ is also a frame for $\mathscr{H}$ with frame bounds $B^{-1}$ and $A^{-1}$. The frame $\{\mathrm{S}^{-1}\varphi_n\}$ is called the *canonical dual frame* for $\{\varphi_n\}$ that one can use to reconstruct $x$ using the next proposition.

**Proposition 1.4.3.** *Given a frame $V \subset \mathscr{H}$, then for each $x \in \mathscr{H}$,*

$$x = \sum_n \langle x, \widetilde{\varphi}_n \rangle \varphi_n = \sum_n \langle x, \varphi_n \rangle \widetilde{\varphi}_n, \tag{1.7}$$

*where both series converge unconditionally in the norm of $\mathscr{H}$.*

**Definition 1.4.4** (Linear reconstruction)**.** *Let $x \in \mathbb{R}^d$. Given a set of measurement vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ and linear measurements $\{\langle x, \varphi_n \rangle + \varepsilon_n\}_{n=1}^N$ of $x$ corrupted with a noise sequence $\{\varepsilon_n\}_{n=1}^N$, the linear construction estimate at the $j$-th step is defined by*

$$x_j = \sum_{n=1}^{j} (y_n + \varepsilon_n) \widetilde{\varphi}_n.$$

One can show that when $\varepsilon_n$ are i.i.d. uniform noise for $n = 1, \ldots, N$, then the mean squared error of linear reconstruction converges to 0 in $\mathscr{O}(N^{-1})$ when the measurement vectors $\{\varphi_n\}$ form a unit-norm tight frame [7].

## 1.5 Reconstruction based on consistency constraints

Linear reconstruction has an optimal mean error rate $\mathscr{O}(N^{-1})$ when the measurement vectors $\{\varphi_n\}$ form a unit-norm tight frame and the noises $\varepsilon_n$ are i.i.d. uniformly distributed, but the algorithm itself does not take advantage of the uniform noise model. If we know the $j$-th measurement $\langle x, \varphi_j \rangle$ is corrupted by a noise $\varepsilon_j$ uniformly distributed on $[-\delta, \delta]$, then the $j$-th estimate $x_j \in \mathbb{R}^d$ of an iterative algorithm can address this fact by enforcing *consistency constraints* that

$$\left|\langle x, \varphi_n \rangle - \langle x_j, \varphi_n \rangle + \varepsilon_n\right| \leq \delta, \; \forall n = 1, \ldots, j. \tag{1.8}$$

In other words, if we set the estimate error at $j$-th step to be $z_j = x - x_j$, then (1.8) ensures

$$\left|\langle z_j, \varphi_n \rangle + \varepsilon_n\right| \leq \delta, \; \forall n = 1, \ldots, j,$$

which says that the magnitude of error in a direction defined by $\varphi_n$ should not be greater than the upper bound of quantization error $\delta$. In the subsequent chapters, we analyze two algorithms that are developed on the idea of enforcing consistent constraints: consistent reconstruction and the Rangan-Goyal algorithm. In particular, we will investigate their expected convergence behaviors with random measurement vectors. The bottom line is that, with i.i.d. random measurement vectors under reasonable conditions, the $p$-th error moment of both algorithms converges to 0 in $\mathscr{O}(N^{-p})$. We present the algorithms here.

**Definition 1.5.1** (Consistent reconstruction)**.** *Let $x \in \mathbb{R}^d$. Given a set of measurement vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ and linear measurements $\{\langle x, \varphi_n \rangle + \varepsilon_n\}_{n=1}^N$ of $x$ that are corrupted with a uniform distributed noise sequence $\{\varepsilon_n\}_{n=1}^N \subset [-\delta, \delta]$, the consistent reconstruction estimate $\widetilde{x} \in \mathbb{R}^d$ for $x$ is produced by selecting an arbitrary solution to the linear feasibility problem*

$$|\langle \widetilde{x}, \varphi_n \rangle - \langle x, \varphi_n \rangle - \varepsilon_n| \leq \delta, \;\; \forall n = 1, \ldots, N. \tag{1.9}$$

In consistent reconstruction, because of the freedom to choose any solution from the convex region defined by (1.9) in $\mathbb{R}^d$, the analysis in Chapter 2 concerns the worst case error in (1.9).

**Definition 1.5.2** (Rangan-Goyal algorithm)**.** *Let $x \in \mathbb{R}^d$. Given a set of measurement vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$, linear measurements $\{\langle x, \varphi_n \rangle + \varepsilon_n\}_{n=1}^N$ of $x$ that are corrupted with a uniform distributed noise sequence $\{\varepsilon_n\}_{n=1}^N \subset [-\delta, \delta]$, and an initial estimate $x_0 \in \mathbb{R}^d$, the Rangan-Goyal algorithm produces the $j$-th estimate iteratively by setting*

$$x_j := x_{j-1} + \frac{\varphi_j}{\|\varphi_j\|^2} T_\delta \left( \langle x, \varphi_j \rangle - \langle x_{j-1}, \varphi_j \rangle + \varepsilon_j \right),$$

*where $T_\delta : \mathbb{R} \to \mathbb{R}$ is a soft-thresholding function defined by*

$$T_\delta(y) = \begin{cases} y - \delta, & \text{if } y > \delta, \\ 0, & \text{if } |y| \leq \delta, \\ y + \delta, & \text{if } y < -\delta. \end{cases} \tag{1.10}$$

The Rangan-Goyal algorithm does not actually force each estimate $x_j$ to satisfy the consistency constraints (1.8). Instead the algorithm only updates the estimate $x_j$ if the previous estimate $x_{j-1}$ breaks the $j$-th consistency constraint.

# CHAPTER 2

# CONSISTENT RECONSTRUCTION

## 2.1 Overview of algorithm

Consistent reconstruction is a method for estimating a signal $x \in \mathbb{R}^d$ from a collection of linear measurements that have been corrupted by uniform noise or, more generally, bounded noise. Estimation with uniform noise arises naturally in quantization problems in signal processing, especially in connection with dithering and the uniform noise model [13, 20]. Consistent reconstruction has been used as a signal recovery method for memoryless scalar quantization [3, 2, 8, 20, 28], Sigma-Delta quantization [27], and compressed sensing [11, 12, 17]. See [19] for background and motivation on consistent reconstruction and estimation with uniform noise.

Let $x \in \mathbb{R}^d$ be an unknown signal and let $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ be a given spanning set for $\mathbb{R}^d$ that is used to make linear measurements $\langle x, \varphi_n \rangle$ of $x$. We consider the problem of recovering an estimate for $x$ from noisy measurements

$$q_n = \langle x, \varphi_n \rangle + \varepsilon_n, \quad 1 \leq n \leq N, \tag{2.1}$$

where $\{\varepsilon_n\}_{n=1}^N$ are random variables in $\mathbb{R}$. For the setting of this chapter, we suppose the noise level $\delta > 0$ is fixed and known, the noise model is that $\{\varepsilon_n\}_{n=1}^N$ are independent uniform random variables on $[-\delta, \delta]$ but are unknown, the collection $\{\varphi_n\}_{n=1}^N$ is known but randomly generated, and the true signal $x$ is unknown. We focus on the situation when $\{\varphi_n\}_{n=1}^N$ are independent versions of a random vector $\varphi \in \mathbb{R}^d$ whose distribution we refer to as the sampling distribution.

Consistent reconstruction seeks an estimate $\widetilde{x}$ for the unknown signal $x$ that is consistent with the knowledge that the noise is bounded in $[-\delta, \delta]$. Specifically, consistent reconstruction produces an estimate $\widetilde{x} \in \mathbb{R}^d$ for $x$ by selecting any solution of the linear feasibility problem

$$|\langle \widetilde{x}, \varphi_n \rangle - q_n| \leq \delta, \quad 1 \leq n \leq N. \tag{2.2}$$

There are generally infinitely many solutions to this feasibility problem. In this chapter, we mainly focus on the worst case error associated to consistent reconstruction.

### 2.1.1 Worst case error

To describe the worst case error of consistent reconstruction, note that if $\widetilde{x}$ is any solution to (2.2), then the error $(\widetilde{x} - x)$ lies in each of the closed convex sets

$$E_n = \left\{ u \in \mathbb{R}^d : |\langle u, \varphi_n \rangle - \varepsilon_n| \leq \delta \right\}. \tag{2.3}$$

whose intersection forms the following error polytope

$$P_N = \bigcap_{n=1}^{N} E_n, \tag{2.4}$$

that is the set of all possible errors associated to consistent reconstruction (2.2). The worst case error $W_N$ associated to consistent reconstruction is thus defined by

$$W_N = \max \{ \|u\| : u \in P_N \}, \tag{2.5}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$.

### 2.1.2 Background

The main results in [19] proved error bounds for the expected worst case error squared $\mathbb{E}[(W_N)^2]$ of consistent reconstruction when the sampling vectors $\{\varphi_n\}_{n=1}^N$ are drawn at random from a suitable probability distribution on the unit sphere $\mathbb{S}^{d-1}$.

The work in [19] considered sampling vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{S}^{d-1}$ that are independently drawn instances of a unit-norm random vector $\varphi$ that satisfies the following admissibility condition:

$$\exists \alpha \geq 1, \exists 0 < s \leq 1, \ \ \forall 0 \leq t \leq 1, \ \ \forall x \in \mathbb{S}^{d-1}, \quad \Pr[|\langle x, \varphi \rangle| \leq t] \leq \alpha t^s. \tag{2.6}$$

See Section 5 of [19] for further discussion of the admissibility condition (2.6). For example, if $\varphi$ is uniformly distributed on $\mathbb{S}^{d-1}$ then $\varphi$ satisfies (2.6) with $s = 1$ and $\alpha = \frac{2\Gamma(\frac{d}{2})}{\sqrt{\pi}\Gamma(\frac{d-1}{2})}$. On the other hand, if $\varphi$ has a point mass then $\varphi$ does not satisfy (2.6).

Suppose that $\{\varphi_n\}_{n=1}^N \subset \mathbb{S}^{d-1}$ are independently drawn at random according to a distribution that satisfies the admissibility condition (2.6). Theorem 5.5 and Corollary 5.6 in [19] prove that there exist absolute constants $c_1, c_2 > 0$ such that if

$$N \geq c_2 d \ln(32(2\alpha)^{1/s}),$$

then the expected worst case error squared for consistent reconstruction satisfies

$$\mathbb{E}[(W_N)^2] \leq \frac{c_1 \delta^2 d^2 (2\alpha)^{1/s} \ln^2(16(2\alpha)^{1/s})}{(N+1)(N+2)}.$$

Moreover, in the special case when $\{\varphi_N\}_{n=1}^N$ are drawn independently at random according to the uniform distribution on $\mathbb{S}^{d-1}$, Theorem 6.1 and Corollary 6.2 in [19] proved a refined error bound with a constant that has cubic dependence on the the dimension

$$\mathbb{E}[(W_N)^2] \leq \frac{c\delta^2 d^3}{N^2}.$$

For perspective, it is known that mean squared error rates of order $1/N^2$ are generally optimal for estimation with uniform noise, see [20].

### 2.1.3 Overview and main results

The error bounds for consistent reconstruction in [19] only considered the mean squared error $\mathbb{E}[(W_N)^2]$ and only considered the admissibility condition (2.6) in the setting of unit-norm random vectors (for example, this excludes the case of Gaussian random vectors). The main contributions of this chapter are twofold:

1. We prove bounds on general error moments $\mathbb{E}[(W_N)^p]$ for consistent reconstruction. Our main results show that the error decreases like $\mathbb{E}[(W_N)^p] \lesssim 1/N^p$, as the number of measurements $N$ increases.

2. We establish a general admissibility condition on the sampling distribution that does not require $\varphi$ to be unit-norm.

In Section 2.2, we prove our first main result, Theorem 2.2.5, which gives upper bounds on $\mathbb{E}[(W_N)^p]$ for unit-norm sampling distributions. Section 2.3 builds on Theorem 2.2.5 and proves our second main result, Theorem 2.3.4, for general sampling distributions that need not be unit-norm.

## 2.2 Error moments for consistent reconstruction: unit-norm distributions

In this section we prove our first main result, Theorem 2.2.5. Theorem 2.2.5 extends Theorem 5.5 in [19] to the setting of general error moments $\mathbb{E}[(W_N)^p]$. In this section, we assume that the sampling vectors $\{\varphi_n\}_{n=1}^N$ are unit-norm and satisfy the admissibility condition (2.6). We shall later remove the unit-norm requirement from the admissibility condition in Section 2.3.

### 2.2.1 Consistent reconstruction and coverage problems

We begin by recalling a useful connection between consistent reconstruction and a problem on covering the sphere by random sets.

**Definition 2.2.1.** *Let* $\{\varphi_n\}_{n=1}^N$ *be a set of unit-norm vectors and let* $\{\varepsilon_n\}_{n=1}^N \subset [-\delta,\delta]$. *For each* $\lambda > 0$, *define*

$$B_n(\lambda) = B(\varphi_n, \varepsilon_n, \lambda) = \left\{ u \in \mathbb{S}^{d-1} : \langle u, \varphi_n \rangle > \frac{\varepsilon_n + \delta}{\lambda} \text{ or } \langle u, \varphi_n \rangle < \frac{\varepsilon_n - \delta}{\lambda} \right\}$$
$$= \left\{ u \in \mathbb{S}^{d-1} : |\lambda \langle u, \varphi_n \rangle - \varepsilon_n| > \delta \right\}. \tag{2.7}$$

In our setting, the sets $B_n(\lambda)$ are random subsets of $\mathbb{S}^{d-1}$ because $\{\varphi_n\}_{n=1}^N$ and $\{\varepsilon_n\}_{n=1}^N$ are random.

Note that each $B_n(\lambda)$ can be expressed as a union of two (possibly empty) antipodal open spherical caps of different sizes

$$B_n(\lambda) = \operatorname{Cap}\left(\varphi_n, \theta_n^+\right) \cup \operatorname{Cap}\left(-\varphi_n, \theta_n^-\right), \tag{2.8}$$

where the angular radii $\theta_n^+$ and $\theta_n^-$ are given by

$$\theta_n^+ = \begin{cases} \arccos\left(\frac{\delta+\varepsilon_n}{\lambda}\right), & \text{if } \delta + \varepsilon_n < \lambda, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\theta_n^- = \begin{cases} \arccos\left(\frac{\delta-\varepsilon_n}{\lambda}\right), & \text{if } \delta - \varepsilon_n < \lambda, \\ 0, & \text{otherwise.} \end{cases}$$

The following lemma shows a connection between consistent reconstruction and the problem of covering the unit sphere by the random sets $B_n(\lambda)$, see Lemma 4.1 in [19].

**Lemma 2.2.2.** *For all $\lambda > 0$, the worst case error satisfies*

$$\Pr\left[W_N > \lambda\right] \leq \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right]. \tag{2.9}$$

The following lemmas collect upper bounds on $\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right]$ that are spread out over various parts of [19].

**Lemma 2.2.3.** *If $\lambda \geq 4\delta$ then*

$$\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right] \leq 4^{d-1}(4^s\alpha)^N \left(\frac{\delta}{\lambda}\right)^{sN-d+1}. \tag{2.10}$$

Lemma 2.2.3 was shown in equation (5.9) in [19].

**Lemma 2.2.4.** *If $0 \leq \lambda \leq 4(2\alpha)^{1/s}\delta$ then*

$$\begin{aligned} &\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right] \\ &\leq \sum_{k=0}^{N} q(k,d-1,\alpha,s)\binom{N}{k}\left(1-\frac{\lambda}{4\delta(2\alpha)^{1/s}}\right)^{N-k}\left(\frac{\lambda}{4\delta(2\alpha)^{1/s}}\right)^{k}, \end{aligned} \tag{2.11}$$

*where $q(k,d-1,\alpha,s)$ satisfies*

$$q(k,d-1,\alpha,s) \leq 1, \tag{2.12}$$

*and*

$$k \geq \frac{2d\ln(16(2\alpha)^{1/s})}{\ln(4/3)} \quad \Longrightarrow \quad q(k,d-1,\alpha,s) \leq \left(\frac{3}{4}\right)^{k/2}. \tag{2.13}$$

The bound (2.11) appears in (5.12) in [19]. The bound (2.12) follows from (5.11) in [19], and the bound (2.13) appears in Step VI in the proof of Theorem 5.5 in [19].

### 2.2.2 Error moment bounds

We now prove our first main result that provides error moment bounds for consistent reconstruction.

**Theorem 2.2.5.** *Suppose that $\{\varphi_n\}_{n=1}^N \subset \mathbb{S}^{d-1}$ are independently drawn at random according to a distribution that satisfies the admissibility condition* (2.6) *with parameters $\alpha \geq 1$ and $0 < s \leq 1$. If $p \in \mathbb{N}$ and $N \geq (d+p)/s$, then the pth error moment for consistent reconstruction satisfies*

$$\mathbb{E}[(W_N)^p] \leq C' \delta^p \left( \prod_{j=1}^{p} (N+j) \right)^{-1} + C'' \delta^p \left(\frac{1}{2}\right)^N, \tag{2.14}$$

*where*

$$C' = C'_{p,\alpha,s} = 2p(4(2\alpha)^{1/s})^p \left( \frac{2d \ln(16(2\alpha)^{1/s})}{\ln(4/3)} + p \right)^p \left( \sum_{k=1}^{\infty} (k+1)^{p-1} (3/4)^{k/2} \right),$$

*and*

$$C'' = C''_{p,\alpha,s,d} = 2p(32(2\alpha)^{1/s})^{p+d-1}.$$

*Proof.* We proceed by directly building on the proof of Theorem 5.5 in [19].

*Step 1.* We need to compute

$$\mathbb{E}[(W_N)^p] = p \int_0^{\infty} \lambda^{p-1} \Pr[W_N > \lambda] d\lambda. \tag{2.15}$$

By Lemma 2.2.2, we have

$$\mathbb{E}[(W_N)^p] \leq p \int_0^{\infty} \lambda^{p-1} \Pr\left[ \mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda) \right] d\lambda. \tag{2.16}$$

Thus, it suffices to bound the integral on right side of (2.16).

*Step 2.* We shall bound the integral in (2.16) by breaking it up into three separate integrals. We begin by estimating the integral in the range $0 \le \lambda \le 4\delta(2\alpha)^{1/s}$.

Using (2.11) and a change of variables gives

$$
\begin{aligned}
& p\int_0^{4\delta(2\alpha)^{1/s}} \lambda^{p-1}\,\mathrm{Pr}\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right] d\lambda \\
& \le p\sum_{k=0}^{N} q(k,d-1,\alpha,s)\binom{N}{k}\int_0^{4\delta(2\alpha)^{1/s}} \lambda^{p-1}\left(1-\frac{\lambda}{4\delta(2\alpha)^{1/s}}\right)^{N-k}\left(\frac{\lambda}{4\delta(2\alpha)^{1/s}}\right)^{k} d\lambda \\
& = p\sum_{k=0}^{N} q(k,d-1,\alpha,s)\binom{N}{k}\left(4\delta(2\alpha)^{1/s}\right)^{p}\int_0^1 v^{k+p-1}(1-v)^{N-k}dv \\
& = p\left(4\delta(2\alpha)^{1/s}\right)^{p}\sum_{k=0}^{N} q(k,d-1,\alpha,s)\binom{N}{k}\frac{(N-k)!(k+p-1)!}{(N+p)!} \\
& = p\left(4\delta(2\alpha)^{1/s}\right)^{p}\left(\prod_{j=1}^{p}(N+j)\right)^{-1}\left[\sum_{k=0}^{N}\frac{(k+p-1)!}{k!}\,q(k,d-1,\alpha,s)\right].
\end{aligned}
\tag{2.17}
$$

Here, we used the property of the beta function that

$$
\int_0^1 v^{k+p-1}(1-v)^{N-k}dv = \frac{(N-k)!(k+p-1)!}{(N+p)!}. \tag{2.18}
$$

It remains to bound the sum $\sum_{k=0}^{N}\frac{(k+p-1)!}{k!}q(k,d-1,\alpha,s)$ in (2.17). We will bound this sum by breaking it up into two separate sums, in an analogous manner to Step VI in the proof of Theorem 5.5 in [19]. Let

$$
K = \left\lfloor \frac{2d\ln(16(2\alpha)^{1/s})}{\ln(4/3)} \right\rfloor. \tag{2.19}
$$

Since $q(k,d-1,\alpha,s) \le 1$, we have

$$
\sum_{k=0}^{K}\frac{(k+p-1)!}{k!}q(k,d-1,\alpha,s) \le \sum_{k=0}^{K}(K+p-1)^{p-1} \le (K+p)^{p}. \tag{2.20}
$$

Using (2.13) we have

$$\begin{aligned}\sum_{k=K+1}^{N} \frac{(k+p-1)!}{k!} q(k,d-1,\alpha,s) &\leq \sum_{k=K+1}^{\infty} \frac{(k+p-1)!}{k!} \left(\frac{3}{4}\right)^{k/2} \\ &\leq \sum_{k=K+1}^{\infty} (k+p-1)^{p-1} \left(\frac{3}{4}\right)^{k/2} \\ &= \sum_{k=1}^{\infty} (k+K+p-1)^{p-1} \left(\frac{3}{4}\right)^{(k+K)/2} \\ &\leq (K+p)^{p-1} \sum_{k=0}^{\infty} (k+1)^{p-1} \left(\frac{3}{4}\right)^{k/2} \\ &= (K+p)^{p-1} S_p, \end{aligned} \tag{2.21}$$

where $S_p = \sum_{k=1}^{\infty}(k+1)^{p-1}(3/4)^{k/2}$ satisfies $1 < S_p < \infty$.

By (2.20) and (2.21) we have

$$\sum_{k=0}^{N} \frac{(k+p-1)!}{k!} q(k,d-1,\alpha,s) \leq (K+p)^p(1+S_p) \leq 2(K+p)^p S_p. \tag{2.22}$$

Combining (2.17) and (2.22) yields

$$\begin{aligned}&p\int_0^{4\delta(2\alpha)^{1/s}} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right] d\lambda \\ &\qquad\leq 2p(4\delta(2\alpha)^{1/s})^p (K+p)^p S_p \left(\prod_{j=1}^{p}(N+j)\right)^{-1}. \end{aligned} \tag{2.23}$$

*Step 3.* Next, we bound the integral (2.16) in the range $4\delta(2\alpha)^{1/s} \leq \lambda \leq 8\delta(2\alpha)^{1/s}$. By Lemma 2.2.3 we know that in this range of $\lambda$,

$$\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B_n(\lambda)\right] \leq (16(2\alpha)^{1/s})^{d-1}\left(\frac{1}{2}\right)^{N}.$$

Thus

$$
\begin{aligned}
p\int_{4\delta(2\alpha)^{1/s}}^{8\delta(2\alpha)^{1/s}} \lambda^{p-1}\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^{N}B_n(\lambda)\right]d\lambda \\
&\le p(16(2\alpha)^{1/s})^{d-1}\left(\frac{1}{2}\right)^{N}\int_{4\delta(2\alpha)^{1/s}}^{8\delta(2\alpha)^{1/s}}\lambda^{p-1}d\lambda \\
&\le \delta^{p}(16(2\alpha)^{1/s})^{d+p-1}\left(\frac{1}{2}\right)^{N}. \qquad (2.24)
\end{aligned}
$$

*Step 4.* We next bound the integral (2.16) in the range $\lambda \ge 8\delta(2\alpha)^{1/s}$. By Lemma 2.2.3 we know that in this range of $\lambda$,

$$
\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^{N}B_n(\lambda)\right] \le 4^{d-1}(4^s\alpha)^N\left(\frac{\delta}{\lambda}\right)^{sN-d+1}.
$$

It follows that when $N \ge (d+p)/s$,

$$
\begin{aligned}
p\int_{8\delta(2\alpha)^{1/s}}^{\infty} \lambda^{p-1}\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^{N}B_n(\lambda)\right]d\lambda \\
&\le p\cdot 4^{d-1}(4^s\alpha)^N\delta^{sN-d+1}\int_{8\delta(2\alpha)^{1/s}}^{\infty}\lambda^{p-sN+d-2}d\lambda \\
&= p\cdot 4^{d-1}(4^s\alpha)^N\delta^{sN-d+1}\left(\frac{(8\delta(2\alpha)^{1/s})^{p-sN+d-1}}{sN-p-d+1}\right) \\
&\le p\cdot\delta^{p}(32(2\alpha)^{1/s})^{p+d-1}\left(\frac{1}{2}\right)^{N}. \qquad (2.25)
\end{aligned}
$$

Combining (2.16), (2.23), (2.24) and (2.25) completes the proof. □

Theorem 2.2.5 yields the following corollary.

**Corollary 2.2.6.** *Suppose that $\{\varphi_n\}_{n=1}^N \subset \mathbb{S}^{d-1}$ are independently drawn at random according to a distribution that satisfies the admissibility condition* (2.6) *with parameters $\alpha \ge 1$ and $0 < s \le 1$. If $p \in \mathbb{N}$ and*

$$
N \ge \max\left\{\frac{2}{\ln 2}\left[\ln\left(\frac{C''}{C'}\right)+2p\ln\left(\frac{4p}{e\ln 2}\right)\right],\frac{d+p}{s}\right\}, \qquad (2.26)
$$

*then*

$$\mathbb{E}[(W_N)^p] \le 2C'\delta^p \left(\prod_{j=1}^{p}(N+j)\right)^{-1}, \tag{2.27}$$

*where $C', C''$ are as in Theorem 2.2.5.*

*Proof.* In view of Theorem 2.2.5, it suffices to show that if $N$ satisfies (2.26) then

$$C''\left(\frac{1}{2}\right)^N \le C'\left(\prod_{j=1}^{p}(N+j)\right)^{-1}.$$

Equivalently, it suffices to show

$$\ln\left(\frac{C''}{C'}\right) + \sum_{j=1}^{p}\ln(N+j) \le N\ln 2. \tag{2.28}$$

To begin, note that

$$\forall x > 0, \quad \ln(x) \le x - 1,$$

gives

$$\begin{aligned}\ln(N) &= \ln\left(\frac{N\ln 2}{4p}\right) + \ln\left(\frac{4p}{\ln 2}\right)\\ &\le \frac{N\ln 2}{4p} - 1 + \ln\left(\frac{4p}{\ln 2}\right)\\ &= \frac{N\ln 2}{4p} + \ln\left(\frac{4p}{e\ln 2}\right).\end{aligned} \tag{2.29}$$

Next, use (2.29) and $N \ge (d+p)/s \ge \max\{p, 2\}$ to obtain

$$\begin{aligned}\sum_{j=1}^{p}\ln(N+j) &= \sum_{j=1}^{p}\left[\ln(N) + \ln\left(1+\frac{j}{N}\right)\right]\\ &\le p\ln(N) + p\ln 2\\ &\le 2p\ln(N)\\ &\le \frac{N\ln 2}{2} + 2p\ln\left(\frac{4p}{e\ln 2}\right).\end{aligned} \tag{2.30}$$

In view of (2.30), to show (2.28) it suffices to have

$$\ln\left(\frac{C''}{C'}\right)+\frac{N\ln 2}{2}+2p\ln\left(\frac{4p}{e\ln 2}\right)\leq N\ln 2. \tag{2.31}$$

Since (2.31) holds by the assumption (2.26), this completes the proof. $\square$

We conclude this section with some perspective on the dimension dependence of the constant $C'$ in Theorem 2.2.5 and Corollary 2.2.6. We consider the special case when $\varphi$ is uniformly distributed on the unit-sphere $\mathbb{S}^{d-1}$ with $d \geq 3$. In this case, one may take $s = 1$ and $\alpha = \frac{2\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)}$ in (2.6), see Example 5.1 in [19], and the constant $C'$ is of order $\left(d^{\frac{3}{2}}\ln d\right)^p$. Here, the logarithmic factor $\ln d$ is an artifact of the general setting of Theorem 2.2.5. In particular, for $p = 2$ the refined analysis in Theorem 6.1 and Corollary 6.2 of [19] shows that the factor $\ln d$ can be removed when $\varphi$ is uniformly distributed on the unit-sphere $\mathbb{S}^{d-1}$. A similar analysis extends to moments with general values of $p \in \mathbb{N}$ and shows that the factor $\ln d$ can be replaced by an absolute constant that is independent of $d$.

## 2.3 Error moments for consistent reconstruction: general distributions

In Section 2.2 we proved bounds on the $p$th error moment for consistent reconstruction when the measurements are made using i.i.d. copies of a unit-norm random vector $\varphi \in \mathbb{S}^{d-1}$. In this section, we relax the unit-norm constraint to accommodate more general distributions.

### 2.3.1 General admissibility condition

**Definition 2.3.1.** *We shall say that a random vector $\varphi \in \mathbb{R}^d$ satisfies the general admissibility condition if the following conditions hold:*

- $\varphi = a\psi$*, where $a$ is a non-negative random variable, $\psi$ is a unit-norm random vector,* *and $a$ and $\psi$ are independent.*

- $\psi$ *satisfies the admissibility condition* (2.6).
- $\exists C > 0$ *such that*

$$\forall \lambda > 0, \quad \lambda \Pr[a\lambda \leq 1] \leq C. \tag{2.32}$$

- $r_a = \Pr[a > 1]$ *satisfies* $0 < r_a < 1$.

**Example 2.3.2.** *A sufficient condition for the small-ball inequality* (2.32) *to hold is when $a$ is an absolutely continuous random variable whose probability density function $f$ is in $L^\infty(\mathbb{R})$. In this case, for each $\lambda > 0$,*

$$\Pr[a\lambda \leq 1] = \Pr\left[a \leq \frac{1}{\lambda}\right] = \int_0^{1/\lambda} f(a)\, da \leq \frac{\|f\|_\infty}{\lambda}.$$

*This shows that a large class of probability distributions satisfy the conditions in Definition 2.3.1. For example, if $\varphi$ is a random vector whose entries are i.i.d zero mean Gaussian random variables, then $\varphi$ satisfies the conditions in Definition 2.3.1.*

In Definition 2.3.1, there would be no loss of generality if $a$ were scaled differently so that $0 < \Pr[a > T] < 1$ for some $T > 0$. In particular, suppose that $\varphi_n = a_n \psi_n$ with $0 < \Pr[a_n > T] < 1$, and $q_n = \langle x, \varphi_n \rangle + \varepsilon_n$ with $\varepsilon_n$ uniformly distributed on $[-\delta, \delta]$. Then $\widetilde{x} \in \mathbb{R}^d$ satisfies

$$|\langle \widetilde{x}, \varphi_n \rangle - q_n| \leq \delta \qquad \text{if and only if} \qquad |\langle \widetilde{x}, \varphi_n' \rangle - q_n'| \leq \delta',$$

where $\varphi_n' = \varphi_n / T = a_n' \psi_n$ and $a_n' = a_n / T$ and $q_n' = \langle x, \varphi_n' \rangle + \varepsilon_n'$, where $\varepsilon_n' = \varepsilon_n / T$ is uniformly distributed on $[-\delta', \delta']$ with $\delta' = \delta / T$.

2.3.2 Coverage problems revisited

Suppose that $\{\varphi_n\}_{n=1}^N$ are i.i.d. versions of a random vector $\varphi$ that satisfies the conditions of Definition 2.3.1. In particular, $\varphi_n = a_n\psi_n$, where $\{a_n\}_{n=1}^N$ i.i.d. versions of a random variable $a$, and $\{\psi_n\}_{n=1}^N$ are i.i.d. versions of a random vector $\psi$. Similar to Lemma 2.2.2, the worst case error $W_N$ for consistent reconstruction can be bounded by

$$\Pr[W_N > \lambda] \leq \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right], \tag{2.33}$$

where $B(\psi_n, \varepsilon_n, a_n\lambda)$ is defined using (2.7).

#### 2.3.2.1 Conditioning and a bound by caps with $a_n = 1$

The following lemma bounds (2.33) by coverage probabilities involving caps with $a_n = 1$.

**Lemma 2.3.3.** *Suppose $\{\varphi_n\}_{n=1}^N$, with $\varphi_n = a_n\psi_n$, are i.i.d. versions of a random vector $\varphi$ that satisfies the conditions of Definition 2.3.1. Then*

$$\begin{aligned}
&\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] \\
&\leq \sum_{j=1}^{N} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \operatorname{bino}(j,N,r) + (1-r)^N,
\end{aligned} \tag{2.34}$$

*where*

$$\operatorname{bino}(j,N,r) = \binom{N}{j} r^j (1-r)^{N-j},$$

*and $r = r_a = \Pr[a > 1]$ is as in Definition 2.3.1.*

*Proof.* Let $\mathscr{J}_{j,N}$ denote the event that exactly $j$ elements of $\{a_n\}_{n=1}^N$ satisfy $a_n > 1$. Since the $\{a_n\}_{n=1}^N$ are independent versions of the random variable $a$,

$$\begin{aligned}\Pr[\mathscr{J}_{j,N}] &= \binom{N}{j}(\Pr[a>1])^j(1-\Pr[a>1])^{N-j}\\ &= \binom{N}{j} r^j(1-r)^{N-j} = \mathrm{bino}(j,N,r).\end{aligned}$$

Thus,

$$\begin{aligned}&\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^N B(\psi_n,\varepsilon_n,a_n\lambda)\right]\\ &=\sum_{j=0}^N \Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^N B(\psi_n,\varepsilon_n,a_n\lambda)\,\middle|\,\mathscr{J}_{j,N}\right]\Pr\left[\mathscr{J}_{j,N}\right]\\ &=\sum_{j=0}^N \Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^N B(\psi_n,\varepsilon_n,a_n\lambda)\,\middle|\,\mathscr{J}_{j,N}\right]\mathrm{bino}(j,N,r). \qquad (2.35)\end{aligned}$$

By (2.7), when $a_n>1$ we have $B(\psi_n,\varepsilon_n,a_n\lambda)\supset B(\psi_n,\varepsilon_n,\lambda)$. Thus for $1\le j\le N$,

$$\begin{aligned}\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^N B(\psi_n,\varepsilon_n,a_n\lambda)\,\middle|\,\mathscr{J}_{j,N}\right] &\le \Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{\{n:a_n>1\}} B(\psi_n,\varepsilon_n,a_n\lambda)\,\middle|\,\mathscr{J}_{j,N}\right]\\ &\le \Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{\{n:a_n>1\}} B(\psi_n,\varepsilon_n,\lambda)\,\middle|\,\mathscr{J}_{j,N}\right]\\ &= \Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{n=1}^j B(\psi_n,\varepsilon_n,\lambda)\right], \qquad (2.36)\end{aligned}$$

where the last equality holds because $\{a_n\}_{n=1}^N$ are i.i.d. random variables that are independent of the i.i.d. random vectors $\{\psi_n\}_{n=1}^N$. For $j=0$, we use the trivial bound

$$\Pr\left[\mathbb{S}^{d-1}\not\subset\bigcup_{\{n:a_n>1\}} B(\psi_n,\varepsilon_n,\lambda)\,\middle|\,\mathscr{J}_{j,N}\right]\le 1.$$

Combining (2.35) and (2.36) completes the proof. □

To bound the binomial terms in Lemma 2.3.3 it will be useful to recall Hoeffding's inequality for Bernoulli random variables. If $0 < p < 1$ and $m \leq Np$, then

$$\sum_{j=0}^{m} \mathrm{bino}(j,N,p) \leq \exp\left(-2\,(Np-m)^2/N\right). \tag{2.37}$$

#### 2.3.2.2 Covering and discretization

A useful technique for bounding coverage probabilities such as (2.33) is to discretize the problem by discretizing the sphere $\mathbb{S}^{d-1}$ with an $\varepsilon$-net, see [5]. In this section, we briefly recall necessary aspects of this discretization method as used in [19].

Recall that a set $\mathscr{N}_\varepsilon \subset \mathbb{S}^{d-1}$ is a geodesic $\varepsilon$-net for $\mathbb{S}^{d-1}$ if

$$\forall x \in \mathbb{S}^{d-1},\ \exists z \in \mathscr{N}_\varepsilon, \quad \text{such that} \quad \arccos(\langle x,z\rangle) \leq \varepsilon.$$

For the remainder of this section, let $\mathscr{N}_\varepsilon$ be a geodesic $\varepsilon$- net of cardinality

$$\#(\mathscr{N}_\varepsilon) \leq \left(\frac{8}{\varepsilon}\right)^{d-1}.$$

It is well known that geodesic $\varepsilon$-nets of such cardinality exist, e.g., see Lemma 13.1.1 in [15] or Section 2.2 in [19].

Recalling (2.8), define the shrunken bi-cap $T_\varepsilon\left[B(\psi_n,\varepsilon_n,a_n\lambda)\right]$ by

$$T_\varepsilon\left[B(\psi_n,\varepsilon_n,a_n\lambda)\right] = \mathrm{Cap}\left(\psi_n, T_\varepsilon(\theta_n^+)\right) \cup \mathrm{Cap}\left(-\psi_n, T_\varepsilon(\theta_n^-)\right),$$

where

$$T_\varepsilon(\theta) = \begin{cases} \theta-\varepsilon, & \text{if } \theta \geq \varepsilon, \\ 0, & \text{if } 0 \leq \theta \leq \varepsilon. \end{cases}$$

Similar to equations (5.4) and (5.5) in [19], the coverage probability (2.33) can be discretized as follows

$$\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] \leq \Pr\left[\mathscr{N}_\varepsilon \not\subset \bigcup_{n=1}^{N} T_\varepsilon\left[B(\psi_n, \varepsilon_n, a_n\lambda)\right]\right]$$
$$\leq \left(\frac{8}{\varepsilon}\right)^{d-1} \left(\sup_{z\in\mathbb{S}^{d-1}} \Pr\left[z \notin T_\varepsilon\left[B(\psi_n, \varepsilon_n, a_n\lambda)\right]\right]\right)^N. \tag{2.38}$$

Similar to equation (5.6) in [19], one has that

$$B(\psi_n, \varepsilon_n, a_n\lambda) \supset \left\{u \in \mathbb{S}^{d-1} : |\langle u, \psi_n\rangle| > \frac{2\delta}{a_n\lambda}\right\}$$

and

$$T_\varepsilon\left[B(\psi_n, \varepsilon_n, a_n\lambda)\right] \supset \left\{u \in \mathbb{S}^{d-1} : |\langle u, \psi_n\rangle| > \frac{2\delta}{a_n\lambda} + \varepsilon\right\}.$$

This gives

$$\Pr\left[z \notin T_\varepsilon\left[B(\psi_n, \varepsilon_n, a_n\lambda)\right]\right] \leq \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon\right]. \tag{2.39}$$

### 2.3.3 Moment bounds for general distributions

We now state our next main theorem.

**Theorem 2.3.4.** *Suppose that $\{\varphi_n\}_{n=1}^N$ are i.i.d. versions of a random vector $\varphi$ that satisfies the conditions of Definition 2.3.1. Let $r = r_a = \Pr[a > 1]$ be as in Definition 2.3.1. If*

$$N \geq \frac{2(d+p)}{sr}, \tag{2.40}$$

*then the pth error moment for consistent reconstruction satisfies*

$$\mathbb{E}\left[(W_N)^p\right] \leq pC'\left(\frac{2\delta}{Nr}\right)^p + pC''\delta^p\left(\frac{1}{2}\right)^{Nr/2} + \delta^p\Lambda^p e^{-Nr^2/2} + \delta^p C'''\left(\frac{1}{2}\right)^N,$$

*where $C', C''$ are as in Theorem 2.2.5, $\Lambda$ is defined by* (2.42) *and* (2.57)*, and $C'''$ is defined by* (2.60) *and* (2.57).

*Proof.* As in Theorem 2.2.5 we shall use (2.15). In view of (2.33), we need to estimate

$$\mathbb{E}[(W_N)^p] \leq p \int_0^\infty \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] d\lambda. \tag{2.41}$$

*Step 1.* We begin by estimating the integral in (2.41) over the range $0 \leq \lambda \leq \Lambda\delta$, where

$$\Lambda = \max\{\Lambda_0, \Lambda_1\}, \quad \text{with} \quad \Lambda_0 = \frac{2^{s+3}C}{\alpha} \quad \text{and} \quad \Lambda_1 = 4\left(2K''\right)^{\frac{s+1}{s}}, \tag{2.42}$$

where $C, \alpha, s$ are the parameters in (2.6) and Definition 2.3.1, and $K''$ is defined in (2.57).

By Lemma 2.3.3 we have

$$\begin{aligned}
&p \int_0^{\Lambda\delta} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] d\lambda \\
&\leq p \int_0^{\Lambda\delta} \lambda^{p-1} \sum_{j=0}^{N} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \mathrm{bino}(j,N,r) d\lambda \\
&= p \int_0^{\Lambda\delta} \lambda^{p-1} \sum_{j=0}^{\lfloor Nr/2 \rfloor} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \mathrm{bino}(j,N,r) d\lambda &(2.43) \\
&\quad + p \int_0^{\Lambda\delta} \lambda^{p-1} \sum_{j=\lceil Nr/2 \rceil}^{N} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \mathrm{bino}(j,N,r) d\lambda. &(2.44)
\end{aligned}$$

Hoeffding's inequality and the trivial bound $\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \leq 1$ can be used to bound (2.43) as follows

$$\begin{aligned}
p \int_0^{\Lambda\delta} \lambda^{p-1} &\sum_{j=0}^{\lfloor Nr/2 \rfloor} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \mathrm{bino}(j,N,r) d\lambda \\
&\leq p \int_0^{\Lambda\delta} \lambda^{p-1} \left(\sum_{j=0}^{\lfloor Nr/2 \rfloor} \mathrm{bino}(j,N,r)\right) d\lambda \\
&\leq p \left(e^{-Nr^2/2}\right) \int_0^{\Lambda\delta} \lambda^{p-1} d\lambda \\
&= \delta^p \Lambda^p e^{-Nr^2/2}. &(2.45)
\end{aligned}$$

To bound the integral in (2.44), recall (2.40) and note that if $j$ satisfies $(d+p)/s \leq \lceil Nr/2 \rceil \leq j \leq N$, then the bounds on (2.16) obtained in the proof of Theorem 2.2.5 give that

$$\begin{aligned} p\int_0^{\Lambda\delta} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] d\lambda & \\ \leq p\int_0^{\infty} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] d\lambda & \\ \leq C'\delta^p \left(\prod_{l=1}^{p}(j+l)\right)^{-1} + C''\delta^p \left(\frac{1}{2}\right)^j & \\ \leq \frac{C'\delta^p}{j^p} + C''\delta^p \left(\frac{1}{2}\right)^j, & \end{aligned} \tag{2.46}$$

where $C'$ and $C''$ are as in Theorem 2.2.5.

Using (2.46), along with $\sum_{j=0}^{N} \text{bino}(j,N,r) = 1$, one may bound (2.44) as follows

$$\begin{aligned} p\sum_{j=\lceil Nr/2 \rceil}^{N} \int_0^{\Lambda\delta} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{j} B(\psi_n, \varepsilon_n, \lambda)\right] \text{bino}(j,N,r) d\lambda & \\ \leq p \sum_{j=\lceil Nr/2 \rceil}^{N} \text{bino}(j,N,r) \left[\frac{C'\delta^p}{j^p} + C''\delta^p \left(\frac{1}{2}\right)^j\right] & \\ \leq p\delta^p \left[\frac{2^p C'}{(Nr)^p} + C''\left(\frac{1}{2}\right)^{Nr/2}\right] \sum_{j=\lceil Nr/2 \rceil}^{N} \text{bino}(j,N,r) & \\ \leq p\delta^p \left[C'\left(\frac{2}{Nr}\right)^p + C''\left(\frac{1}{2}\right)^{Nr/2}\right]. & \end{aligned} \tag{2.47}$$

Applying the bounds (2.45) and (2.47) to (2.43) and (2.44) gives

$$\begin{aligned} p\int_0^{\Lambda\delta} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] d\lambda & \\ \leq \delta^p \Lambda^p e^{-Nr^2/2} + pC'\left(\frac{2\delta}{Nr}\right)^p + pC''\delta^p \left(\frac{1}{2}\right)^{Nr/2}. & \end{aligned} \tag{2.48}$$

*Step 2.* We next estimate the integral in (2.41) over the range $\lambda \geq \Lambda\delta$. By (2.38) and

(2.39) we have

$$\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] \leq \left(\frac{8}{\varepsilon}\right)^{d-1} \left(\sup_{z\in\mathbb{S}^{d-1}} \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon\right]\right)^{N}. \tag{2.49}$$

We therefore need to bound $\Pr[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon]$.

For the remainder of this step set

$$A = \left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}} \left(\frac{4\delta}{\lambda}\right)^{\frac{s}{s+1}} \quad \text{and} \quad \varepsilon = \frac{2\delta}{A\lambda} = \left(\frac{1}{2}\right)\left(\frac{4\delta C}{\lambda\alpha}\right)^{\frac{1}{s+1}}. \tag{2.50}$$

By (2.42), note that $\lambda \geq \Lambda\delta \geq \Lambda_0\,\delta$ implies that $0 < \varepsilon \leq 1/4$.

For any $z \in \mathbb{S}^{d-1}$ we have

$$\Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon\right] = \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon \,\middle|\, a_n > A\right] \Pr[a_n > A] \tag{2.51}$$

$$+ \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon \,\middle|\, a_n \leq A\right] \Pr[a_n \leq A]. \tag{2.52}$$

We now bound the terms appearing in (2.51). Recall that $\lambda \geq \Lambda\delta$ implies that $4\delta/(A\lambda) = 2\varepsilon \leq 1/2$. By our choice of $\varepsilon$ in (2.50), and using the admissibility assumption (2.6), for each $\lambda \geq \Lambda\delta$ one has

$$\begin{aligned}
&\Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{a_n\lambda} + \varepsilon \,\middle|\, a_n > A\right] \Pr[a_n > A] \\
&\leq \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{2\delta}{A\lambda} + \varepsilon \,\middle|\, a_n > A\right] \Pr[a_n > A] \\
&= \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{4\delta}{A\lambda} \,\middle|\, a_n > A\right] \Pr[a_n > A] \\
&\leq \Pr\left[|\langle z, \psi_n\rangle| \leq \frac{4\delta}{A\lambda}\right] \\
&\leq \alpha\left(\frac{4\delta}{A\lambda}\right)^{s}.
\end{aligned} \tag{2.53}$$

To bound (2.52), note that by (2.32) one has $\Pr[a_n \le A] \le CA$, and thus

$$\Pr\left[|\langle z, \psi_n\rangle| \le \frac{2\delta}{a_n\lambda} + \varepsilon \,\middle|\, a_n \le A\right] \Pr[a_n \le A] \le \Pr[a \le A] \le CA. \tag{2.54}$$

Using the bounds (2.53) and (2.54) in (2.51) and (2.52) gives

$$\Pr\left[|\langle z, \psi_n\rangle| \le \frac{2\delta}{a_n\lambda} + \varepsilon\right] \le \alpha \left(\frac{4\delta}{A\lambda}\right)^s + CA. \tag{2.55}$$

Since our choice of $A$ in (2.50) gives

$$\alpha \left(\frac{4\delta}{A\lambda}\right)^s = CA,$$

we have

$$\Pr\left[|\langle z, \psi_n\rangle| \le \frac{2\delta}{a_n\lambda} + \varepsilon\right] \le 2CA = 2C\left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}} \left(\frac{4\delta}{\lambda}\right)^{\frac{s}{s+1}}. \tag{2.56}$$

Thus, combining (2.49) and (2.56), gives

$$\begin{aligned}\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] &\le \left(\frac{8}{\varepsilon}\right)^{d-1} \left[2C\left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}} \left(\frac{4\delta}{\lambda}\right)^{\frac{s}{s+1}}\right]^N \\ &= \left(16\left(\frac{\alpha\lambda}{4\delta C}\right)^{\frac{1}{s+1}}\right)^{d-1} \left[2C\left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}} \left(\frac{4\delta}{\lambda}\right)^{\frac{s}{s+1}}\right]^N.\end{aligned}$$

To simplify notation, let

$$K' = \left(16\left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}}\right)^{d-1} \quad \text{and} \quad K'' = 2C\left(\frac{\alpha}{C}\right)^{\frac{1}{s+1}}, \tag{2.57}$$

so that

$$\Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] \le K' \left(\frac{\lambda}{4\delta}\right)^{\frac{d-1}{s+1}} \left[K'' \left(\frac{4\delta}{\lambda}\right)^{\frac{s}{s+1}}\right]^N$$

$$= K' (K'')^N \left(\frac{4\delta}{\lambda}\right)^{\left(\frac{sN-d+1}{s+1}\right)}. \tag{2.58}$$

Since $0 < s \le 1$ and $0 < r < 1$, note that (2.40) implies $\left(\frac{sN-d+1}{s+1} - p + 1\right) \ge 2$. By (2.58) we have

$$p\int_{\Lambda\delta}^{\infty} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] d\lambda$$

$$\begin{aligned}
&\le pK'(K'')^N \int_{\Lambda\delta}^{\infty} \lambda^{p-1} \left(\frac{4\delta}{\lambda}\right)^{\left(\frac{sN-d+1}{s+1}\right)} d\lambda \\
&= pK'(K'')^N (4\delta)^{p-1} \int_{\Lambda\delta}^{\infty} \left(\frac{4\delta}{\lambda}\right)^{\left(\frac{sN-d+1}{s+1} - p + 1\right)} d\lambda \\
&= pK'(K'')^N (4\delta)^{p} \int_{\Lambda/4}^{\infty} \left(\frac{1}{\lambda}\right)^{\left(\frac{sN-d+1}{s+1} - p + 1\right)} d\lambda \\
&= pK'(K'')^N (4\delta)^{p} \left(\frac{\Lambda}{4}\right)^{p - \frac{sN-d+1}{s+1}} \left(\frac{sN-d+1}{s+1} - p\right)^{-1}. \\
&\le pK'(K'')^N (4\delta)^{p} \left(\frac{\Lambda}{4}\right)^{p - \frac{sN-d+1}{s+1}} \\
&= pK'(4\delta)^{p} \left(\frac{\Lambda}{4}\right)^{p + \frac{d-1}{s+1}} \left[K'' \left(\frac{4}{\Lambda}\right)^{\frac{s}{s+1}}\right]^N.
\end{aligned}$$

Since (2.42) implies that $K'' \left(\frac{4}{\Lambda}\right)^{\frac{s}{s+1}} \le 1/2$, it follows that

$$p\int_{\Lambda\delta}^{\infty} \lambda^{p-1} \Pr\left[\mathbb{S}^{d-1} \not\subset \bigcup_{n=1}^{N} B(\psi_n, \varepsilon_n, a_n\lambda)\right] d\lambda \le \delta^p C''' \left(\frac{1}{2}\right)^N, \tag{2.59}$$

where

$$C''' = pK' 4^p \left(\frac{\Lambda}{4}\right)^{p + \frac{d-1}{s+1}}. \tag{2.60}$$

Combining (2.41), (2.48) and (2.59) completes the proof. □

Similar to Corollary 2.2.6, the following corollary of Theorem 2.3.4 shows that $\mathbb{E}[(W_N)^p]$ is at most of order $1/N^p$ when $N$ is sufficiently large.

**Corollary 2.3.5.** *Let* $\{\varphi_n\}_{n=1}^N$ *be as in Theorem 2.3.4. There exist constants* $C_1, C_2 > 0$ *such that*

$$\forall N \geq C_1, \quad \mathbb{E}\left[(W_N)^p\right] \leq \frac{C_2 \delta^p}{N^p}. \tag{2.61}$$

*The constants* $C_1, C_2$ *depend on* $\alpha, s, C, p, d$.

# CHAPTER 3

# THE RANGAN-GOYAL ALGORITHM

## 3.1 Preliminaries

Suppose that an unknown signal $x \in \mathbb{R}^d$ was measured with a sequence of measurement vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$. The sequence of linear measurements $\{\langle x, \varphi_n \rangle + \varepsilon_n\rangle\}_{n=1}^N$ are known to be corrupted by noises $\{\varepsilon_n\}_{n=1}^N \subset \mathbb{R}$. The question is, how well can one recover the signal $x$ from the sequence of linear measurements? Under the assumption that the noises $\varepsilon_n$ are i.i.d. random variables uniformly distributed on $[-\delta, \delta]$ for some $\delta > 0$, Rangan and Goyal (RG) proposed an iterative algorithm in [20] that estimates $x$ from $\{\langle x, \varphi_n \rangle + \varepsilon_n\rangle\}_{n=1}^N$ as follows. Initializing with arbitrary $x_0 \in \mathbb{R}^d$, at each step the RG algorithm updates the estimate $x_n$ by setting

$$x_n := x_{n-1} + \frac{\varphi_n}{\|\varphi_n\|^2} T_\delta\left(\langle x, \varphi_n \rangle - \langle x_{n-1}, \varphi_n \rangle + \varepsilon_n\right), \tag{3.1}$$

where $T_\delta : \mathbb{R} \to \mathbb{R}$ is a soft-thresholding function defined by

$$T_\delta(y) = \begin{cases} y - \delta, & \text{if } y > \delta, \\ 0, & \text{if } |y| \leq \delta, \\ y + \delta, & \text{if } y < -\delta. \end{cases} \tag{3.2}$$

The use of the soft-thresholding function $T_\delta$ is motivated by the consistency constraint: at each iteration, our estimate error $x - x_{n-1}$ should satisfy

$$|\langle x - x_{n-1}, \varphi_n \rangle| \leq \delta,$$

and if not, the RG algorithm shrinks $x - x_{n-1}$ in a direction opposite to the direction of $\varphi_n$

by an amount depending on $\delta$. In this chapter, we will investigate the convergence behavior of the RG algorithm. In particular, we are interested in general moments of the estimate error $z_n = x - x_n \in \mathbb{R}^d$ at the $n$-th iteration, which can be written as

$$z_n = z_{n-1} - \frac{\varphi_n}{\|\varphi_n\|^2} T_\delta\left(\langle z_{n-1}, \varphi_n\rangle + \varepsilon_n\right).$$

### 3.1.1 Background

When the measurement vectors $\{\varphi_n\}_{n=1}^N$ are taken from i.i.d distributions, the RG algorithm has been shown to converge almost surely in [20]. Moreover, the second error moment of the RG algorithm has been shown to converge to 0 at the rate $\mathscr{O}(N^{-2})$ if $\varphi_n$ are i.i.d. copies of a unit-norm random vector $\varphi$ which satisfies a set of reasonable assumptions [18]. The goal of this chapter is to extend the results in [18] by showing the $p$-th error moment of the RG algorithm converges for any $p \geq 2$ under reasonable assumptions on the set of random measurement vectors $\{\varphi_n\}_{n=1}^N$.

In this chapter, we expand the result in [18] by replacing the unit-norm condition on measurement vectors $\|\varphi\| = 1$ by $\|\varphi\| \leq A$ for some $A > 0$ and generalize the result on the convergence rate of second error moment to $p$-th error moment for all $p \geq 2$. To this end, we will impose a set of assumptions on the random vector $\varphi$ similar to a probabilistic frame condition as follows:

**Assumption 3.1.1.** *We require the random measurement vectors $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ and noises $\{\varepsilon_n\}_{n=1}^N \subset \mathbb{R}$ in (3.1) to satisfy the following conditions:*

- *$\{\varphi_n\}_{n=1}^N$ are i.i.d. copies of a single random vector $\varphi \in \mathbb{R}^d$.*
- *$\{\varepsilon_n\}_{n=1}^N$ are i.i.d. uniform random variables on $[-\delta, \delta]$.*
- *$\varphi_j$ and $\varepsilon_k$ are independent for all pairs onf indices $(j,k)$.*

- *Given the random vector $\varphi$, there exist constants $C_p, C_p' > 0$ for each positive integer $p$ such that*

$$C_p \|x\|^p \mathbb{E}\|\varphi\| \leq \mathbb{E}\left[\left|\left\langle x, \frac{\varphi}{\|\varphi\|}\right\rangle\right|^p \|\varphi\|\right] \leq C_p' \|x\|^p \mathbb{E}\|\varphi\|, \quad \textit{for all } x \in \mathbb{R}^d. \tag{3.3}$$

- *The constants in (3.3) satisfy $C_{4k-1} > \frac{4k}{4k+1} C'_{4k+1}$ for all $k \geq 1$.*

**Example 3.1.2.** *As an example of a set of measurement vectors that satisfies Assumption 3.1.1, consider the case when the measurement vectors $\{\varphi_n\}_{n=1}^N$ are i.i.d. copies of a single Gaussian random vector $\varphi$. That is, $\varphi = \left(\varphi^{(1)}, \varphi^{(2)}, ..., \varphi^{(d)}\right)$ where each $\varphi^{(i)} \sim \mathcal{N}(\mu, \sigma)$ for some fixed $\mu$ and $\sigma > 0$. Assume without loss of generality that $\mathbb{E}\|\varphi\| = 1$, we will show this setup satisfies the fourth and fifth conditions in Assumption 3.1.1. First, note that $\varphi$ being isotropic implies that $\varphi/\|\varphi\|$ and $\|\varphi\|$ are independent. Therefore $|\langle x, \varphi/\|\varphi\|\rangle|^p$ and $\|\varphi\|$ are also independent random variables. Hence,*

$$\begin{aligned}
&\mathbb{E}\left[\left|\left\langle x, \frac{\varphi}{\|\varphi\|}\right\rangle\right|^p \|\varphi\|\right] \\
&\quad = \mathbb{E}\left|\left\langle x, \frac{\varphi}{\|\varphi\|}\right\rangle\right|^p \cdot \mathbb{E}\|\varphi\| \\
&\quad = \|x\|^p \cdot \mathbb{E}\left|\left\langle \frac{x}{\|x\|}, \frac{\varphi}{\|\varphi\|}\right\rangle\right|^p.
\end{aligned}$$

*Let us define the random variable $Z := |\langle x/\|x\|, \varphi/\|\varphi\|\rangle|$. Since $\varphi/\|\varphi\|$ is a uniform random variable distributed on $\mathbb{S}^{d-1}$, $Z$ has the following probability density function $f_Z$ according to [25], where $\Gamma$ is the gamma function:*

$$f_Z(z) = \begin{cases} 2C_d(1-z^2)^{(d-3)/2}, & \textit{if } z \in [0,1], \\ 0, & \textit{otherwise}, \end{cases}$$

$$\textit{where } \; C_d = \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)}, \quad \textit{and } \; \Gamma(z) := \int_0^\infty t^{z-1}e^{-t}dt.$$

*Therefore, for all $p > 0$ and $d > 2$,*

$$\begin{aligned}
\mathbb{E}\left|\left\langle \frac{x}{\|x\|}, \frac{\varphi}{\|\varphi\|}\right\rangle\right|^p &= \mathbb{E}|Z|^p \\
&= 2C_d \int_0^1 z^p(1-z^2)^{(d-3)/2}dz \\
&= C_d \int_0^1 t^{(p-1)/2}(1-t)^{(d-3)/2}dt \\
&= C_d \cdot B\left(\frac{p+1}{2}, \frac{d-1}{2}\right) \\
&= \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)} \cdot \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{d-1}{2}\right)}{\Gamma\left(\frac{p+d}{2}\right)} \\
&= \frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{p+d}{2}\right)},
\end{aligned}$$

*where $B = B(x,y)$ is the beta function,*

$$B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}dt, \;\; \forall x,y > 0.$$

*If we set*

$$C_p = C_p' := \frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{p+d}{2}\right)},$$

*then from the property of gamma function that* $\Gamma(z+1) = z\Gamma(z)$, *one has*

$$\begin{aligned}\frac{C_{4k-1}}{C'_{4k+1}} &= \frac{\Gamma(2k)}{\Gamma(2k+1)} \cdot \frac{\Gamma\left(\frac{4k+d+1}{2}\right)}{\Gamma\left(\frac{4k+d-1}{2}\right)} \\ &= \frac{1}{2k} \cdot \frac{4k+d-1}{2}. \end{aligned} \tag{3.4}$$

*When combined with* $k, d \geq 1$, (3.4) *means*

$$\frac{C_{4k-1}}{C'_{4k+1}} \geq 1 > \frac{4k}{4k+1}.$$

*So both the fourth and the fifth conditions in Assumption 3.1.1 hold for i.i.d. Gaussian random vectors* $\{\varphi_n\}_{n=1}^N$ *when* $\mathbb{E}\|\varphi\| = 1$.

### 3.1.2 Main results

In the next section, Theorem 3.2.1 provides bounds on the $p$-th error moment of RG algorithm for all $p \geq 2$ when RG is initialized with a good condition on a set of random measurement vectors $\{\varphi_n\}_{n=1}^N$ that satisfies Assumption 3.1.1 and $\|\varphi\| \leq A$. Theorem 3.3.6 then extends the result to general initial condition. In both theorems, the error moments are shown to be decreasing like

$$\mathbb{E}[\|x - x_N\|^p] \lesssim \frac{1}{N^p},$$

as the number of measurements $N$ increases.

## 3.2 Error moments under good initial condition

The behavior of the RG algorithm depends the magnitude of the estimate error $z_j := x - x_j$. In this section, we restrict the problem to the initial condition that $\|z_0\| = \|x - x_0\| \leq 2\delta/A$, which we refer to as the "good" initial condition. To simplify notation, for conditional expectation of a random variable $X = X(Y,Z)$ on $Y$, we write $\mathbb{E}_Y[X]$ instead of

$\mathbb{E}[X|Z]$ with the implicit understanding that the expected value is conditioned on $Z$. The main result here is that the $p$-th error moment of RG algorithm converges to zero at the rate $\mathscr{O}(N^{-p})$ for any $p \geq 2$, shown in Corollary 3.2.3 of the main theorem, if we also require that $\|\varphi\| \leq A$. The main theorem bounds the even error moments of RG algorithm:

**Theorem 3.2.1.** *Suppose that $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ are i.i.d. versions of a single random vector $\varphi \in \mathbb{R}^d$ with $\|\varphi\| \leq A$ for some $A > 0$, that $\{\varepsilon_n\}_{n=1}^N$ are i.i.d. uniform random variables on $[-\delta, \delta]$, and that Assumption 3.1.1 holds. Given $x \in \mathbb{R}^d$ and noisy linear measurements $\{\langle x, \varphi_n\rangle + \varepsilon\}_{n=1}^N$, let $\{x_n\}_{n=0}^N \subset \mathbb{R}^d$ denote the RG estimate of $x$ with initial estimate $x_0 \in \mathbb{R}^d$ at the n-th step. If $\|x - x_0\| \leq 2\delta/A$, then for all $p \in \mathbb{N}$ such that $p+1 \leq n \leq N$, one has*

$$\mathbb{E}\|x - x_n\|^{2p} \leq \left[\frac{2p}{M_{p,\delta}(n+1)}\right]^{2p}.$$

*The constant $M_{p,\delta}$ is defined by*

$$M_{p,\delta} := \min\left\{\frac{p}{\delta\sqrt{p+1}}, -\frac{\lambda}{2\delta}\sum_{k=1}^{\lfloor p/2 \rfloor} A_{2k-1}\left(C_{4k-1} - \frac{4k}{4k+1}C'_{4k+1}\right)\right\} > 0, \tag{3.5}$$

*where $\lambda = \mathbb{E}\|\varphi\|$, the constants $C_k$ and $C'_k$ are defined in Assumption 3.1.1, and $A_k$ are defined by*

$$A_k = \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^{k-j}}{(k+j+1)}.$$

**Corollary 3.2.2.** *With the same assumptions on $\{\varphi_n\}_{n=1}^N$ and $\{\varepsilon_n\}_{n=1}^N$ and notations in Theorem 3.2.1, if $\|x - x_0\| > 2\delta/A$ but $\|x - x_{n_0}\| \leq 2\delta/A$ holds for some $n_0 \geq 1$ instead, then for all $n \leq N$ and $p \in \mathbb{N}$ that satisfies $p+1 \leq n - n_0$ we have*

$$\mathbb{E}\left[\|x - x_n\|^2 \,\middle|\, x - x_{n_0}\right] \leq \left[\frac{2p}{M_{p,\delta}((n-n_0)+1)}\right]^{2p}.$$

Other error moment bounds are then derived from the even error moment bounds:

**Corollary 3.2.3.** *Under the same assumptions and notations in Theorem 3.2.1, for all $p \in \mathbb{N}$ such that $p+1 \leq n \leq N$ one has*

$$\mathbb{E}\|z_n\|^{2p+q} \leq \left[\frac{2p}{M_{p,\delta}(n+1)}\right]^{2p+q}, \ \forall\, 0 < q < 2.$$

*If $\|x-x_0\| > 2\delta/A$ but $\|x-x_{n_0}\| \leq 2\delta/A$ holds for some $n_0 \geq 1$ instead, we also have for all $n \leq N$ and $p \geq 2$ that satisfies $p+1 \leq n-n_0$*

$$\mathbb{E}\left[\|x-x_n\|^{2p+q} \,\middle|\, x-x_{n_0}\right] \leq \left[\frac{2p}{M_{p,\delta}((n-n_0)+1)}\right]^{2p+q}.$$

*Proof.* Applying Holder's inequality, we obtain

$$\begin{aligned}
\mathbb{E}\|z_n\|^{2p+q} &= \mathbb{E}\left[\|z_n\|^{2p+q} \cdot 1\right] \\
&\leq \left(\mathbb{E}\|z_n\|^{(2p+q)\cdot\frac{2p}{2p+q}}\right)^{\frac{2p+q}{2p}} \cdot 1 \\
&= \left(\mathbb{E}\|z_n\|^{2p}\right)^{\frac{2p+q}{2p}} \\
&\leq \left[\frac{2p}{M_{p,\delta}(n+1)}\right]^{2p+q}.
\end{aligned}$$

□

The proof of Theorem 3.2.1 will rely on the following lemmas, the first of which from [20] states that the errors of RG estimates decrease monotonically:

**Lemma 3.2.4.** *Let $x \in \mathbb{R}^d$. With any measurements vectors $\{\varphi_n\}_{n=1}^N$, any uniformly noise sequence $\{\varepsilon_n\}_{n=1}^N \subset [-\delta,\delta]$, and any initial condition $x_0 \in \mathbb{R}^d$, the RG estimates $\{x_n\}_{n=1}^N$ satisfy*

$$\|x-x_n\| \leq \|x-x_{n-1}\|, \ \ \forall 1 \leq n \leq N.$$

The next lemma is a combinatorial result.

**Lemma 3.2.5.** *Let*

$$A_k := \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^{k-j}}{(k+j+1)},$$

*then* $A_1 = -2/3$*, and* $A_{k+1}/A_k = -(2k+2)/(2k+3)$ *for all* $k \in \mathbb{N}$.

*Proof.* Straightforward calculation shows that $A_1 = -2/3$. For $k > 1$, define

$$D_k(x) := \sum_{j=0}^{k} \binom{k}{j} \frac{x^{k+j+1}}{k+j+1}.$$

Note that $A_k = (-2)^{2k+1} \cdot D_k(-1/2)$. and $A_{k+1}/A_k = (-2)^2 \cdot D_{k+1}(-1/2)/D_k(-1/2)$. Therefore it suffices to show that

$$\frac{D_{k+1}(-1/2)}{D_k(-1/2)} = -\frac{k+1}{4k+6}.$$

Now $(D_k)'(x) = \sum_{j=0}^{k} \binom{k}{j} x^{k+j} = x^k(1+x)^k$ and $D_k(0) = 0$, so

$$\begin{aligned} D_k(-1/2) &= D_k(-1/2) - D_k(0) \\ &= \int_0^{-1/2} x^k(1+x)^k dx \\ &= (-1)^{k+1} \int_0^{1/2} t^k(1-t)^k dt \\ &= (-1)^{k+1} B(1/2; k+1, k+1), \end{aligned}$$

where $B(x;a,b) = B_x(a,b)$ is the incomplete Beta function as defined in [4]. By convention, let $B(a,b)$ denote the Beta function and $I_x(a,b)$ denote the regularized incomplete Beta function where $B(a,b) =: I_x(a,b)B_x(a,b)$. With the identity that $I_x(a,a) = \frac{1}{2} I_{4x(1-x)}(a, \frac{1}{2})$

[4, Eq. 8.17.6] and noting that $I_1(a,b) = 1$, we have

$$\begin{aligned}
\frac{D_{k+1}(-1/2)}{D_k(-1/2)} &= -\frac{B(1/2;k+2,k+2)}{B(1/2;k+1,k+1)} \\
&= -\frac{I_{1/2}(k+2,k+2)B(k+2,k+2)}{I_{1/2}(k+1,k+1)B(k+1,k+1)} \\
&= -\frac{B(k+2,k+2)}{B(k+1,k+1)} \\
&= -\frac{k+1}{(k+1)+(k+2)} \cdot \frac{B(k+1,k+2)}{B(k+1,k+1)} \\
&= -\frac{k+1}{(k+1)+(k+2)} \cdot \frac{1}{2} \\
&= -\frac{k+1}{4k+6},
\end{aligned}$$

as desired. Thus the proof is complete. □

The next lemma computes the error moment $\mathbb{E}_{\varepsilon_n}\|z_n\|^p$ conditioning on both $\varphi_n$ and $z_{n-1}$.

**Lemma 3.2.6.** *With $\varphi \in \mathbb{R}^d$ and $\delta > 0$ fixed, let $\varepsilon$ be a uniform random variable on $[-\delta,\delta]$. For $u \in \mathbb{R}$, define $F_\delta : \mathbb{R} \to \mathbb{R}$ by $F_\delta(u) = F_\delta(u,\varepsilon) := (1/\|\varphi\|^2)T_\delta(u+\varepsilon)[T_\delta(u+\varepsilon) - 2u]$ where $T_\delta$ is the soft-thresholding function in (3.2). If $|u| \leq 2\delta$, then for all $m > 0$,*

$$\mathbb{E}_\varepsilon [F_\delta(u,\varepsilon)^m] = \sum_{j=0}^{m} \binom{m}{j} \frac{(-2)^{m-j}}{\|\varphi\|^{2m}} \frac{|u|^{2m+1}}{2\delta(m+j+1)}.$$

*Proof.* First, we calculate the expected value of powers of $T(u+\varepsilon)$ where the uniform random variable $\varepsilon$ has a constant probability density $p(\varepsilon) = 1/2\delta$. Note that $T(u+\varepsilon)$ is nonzero when $u+\varepsilon-\delta > 0$ or $u+\varepsilon+\delta < 0$. Since $|\varepsilon| \leq \delta$, only $u+\varepsilon-\delta > 0$ can be true when $u > 0$. So for $0 < u \leq 2\delta$ and $m > 0$,

$$\begin{aligned}\mathbb{E}\left[T(u+\varepsilon)^m\right] &= \int_{-\delta}^{\delta} T\left(u+\varepsilon\right)^m p\left(\varepsilon\right) d\varepsilon \\ &= \int_{\delta-u}^{\delta} (u+\varepsilon-\delta)^m \frac{d\varepsilon}{2\delta} \\ &= \frac{u^{m+1}}{2\delta\left(m+1\right)}.\end{aligned}$$

On the other hand, only $u+\varepsilon+\delta < 0$ can be true when $-2\delta \leq u < 0$, so in this case

$$\begin{aligned}\mathbb{E}\left[T(u+\varepsilon)^m\right] &= \int_{-\delta}^{\delta} T\left(u+\varepsilon\right)^m dp\left(\varepsilon\right) \\ &= \int_{-\delta}^{-\delta-u} (u+\varepsilon+\delta)^m \frac{d\varepsilon}{2\delta} \\ &= -\frac{u^{m+1}}{2\delta\left(m+1\right)}.\end{aligned}$$

Hence for $|u| \leq 2\delta$ one has

$$\mathbb{E}_\varepsilon\left[T\left(u+\varepsilon\right)^m\right] = \begin{cases} \dfrac{u^{m+1}}{2\delta\left(m+1\right)}, & \text{if } u \geq 0, \\ -\dfrac{u^{m+1}}{2\delta\left(m+1\right)}, & \text{if } u < 0. \end{cases} \tag{3.6}$$

From (3.6) it follows that for $|u| \leq 2\delta$ and $m > 0$,

$$\begin{aligned}\mathbb{E}_\varepsilon\left[F_\delta(u,\varepsilon)^m\right] &= \mathbb{E}_\varepsilon\left[\frac{1}{\|\varphi\|^{2m}} T_\delta\left(u+\varepsilon\right)^m \left[T_\delta\left(u+\varepsilon\right) - 2u\right]^m\right] \\ &= \mathbb{E}_\varepsilon\left[\frac{1}{\|\varphi\|^{2m}} T_\delta\left(u+\varepsilon\right)^m \sum_{j=0}^{m} \binom{m}{j} T_\delta\left(u+\varepsilon\right)^j \left(-2u\right)^{m-j}\right] \\ &= \sum_{j=0}^{m} \binom{m}{j} \frac{\left(-2u\right)^{m-j}}{\|\varphi\|^{2m}} \mathbb{E}_\varepsilon\left[T_\delta\left(u+\varepsilon\right)^{m+j}\right] \\ &= \sum_{j=0}^{m} \binom{m}{j} \frac{\left(-2\right)^{m-j}}{\|\varphi\|^{2m}} \frac{|u|^{2m+1}}{2\delta\left(m+j+1\right)},\end{aligned}$$

and the proof is complete. □

**Lemma 3.2.7.** *For $p \geq 1$ and $C > 0$, define $f_p : \mathbb{R} \to \mathbb{R}$ by $f_p(x) = x - Cx^{(p+1)/p}$. Setting $K = (p/C)^p \geq 1$, then for all $N \geq p+1$ we have*

$$f_p(x) \leq \frac{K}{(N+1)^p}, \ \forall x \in \left[0, \frac{K}{N^p}\right].$$

*Proof.* First derivative test shows that $f$ is increasing between 0 and $K/(p+1)^p$. Hence $N \geq p+1$ implies

$$f_p\left(\left[0, \frac{K}{N^p}\right]\right) \subset \left[0, f_p\left(\frac{K}{N^p}\right)\right],$$

and it is sufficient to show that

$$f_p\left(\frac{K}{N^p}\right) = \frac{K}{N^p} - \frac{CK^{(p+1)/p}}{N^{p+1}} \leq \frac{K}{(N+1)^p},$$

which is equivalent to showing that

$$(N - CK^{1/p})(N+1)^p \leq N^{p+1}. \tag{3.7}$$

By rearranging terms in (3.7), the problem reduces to proving that

$$CK^{1/p}N^{p+1}(N+1)^p \geq N(N+1)^p,$$

which is guaranteed by the assumption that $CK^{1/p} \geq p \geq 1$. □

We are now ready to prove the main theorem in this section.

*Proof of Theorem.* Let us define $z_k = x - x_k \in \mathbb{R}^d$ to be the error of RG estimate at $k$-th iteration. Set $u_n = \langle z_{n-1}, \varphi_n \rangle$, then

$$\begin{aligned}
\|z_n\|^{2p} &= \left(\|z_{n-1}\|^2 + F_\delta(u_n, \varepsilon_n)\right)^p \\
&= \sum_{m=0}^{p} \binom{p}{m} \|z_{n-1}\|^{2(p-m)} \left[F_\delta(u_n, \varepsilon_n)\right]^m.
\end{aligned}$$

We will now compute the error moments conditioned on $\varepsilon_n$, $\varphi_n$, and $z_{n-1}$ successively. First, using Lemma 3.2.6 we have the error moment conditioned on $\varepsilon_n$:

$$\begin{aligned}\mathbb{E}_{\varepsilon_n}\|z_n\|^{2p} &= \sum_{m=0}^{p}\binom{p}{m}\|z_{n-1}\|^{2(p-m)}\mathbb{E}_{\varepsilon_n}\left[(F_\delta)^m(u,\varepsilon)\right] \\ &= \|z_{n-1}\|^{2p}+\sum_{m=1}^{p}\|z_{n-1}\|^{2(p-m)}\sum_{j=0}^{m}\binom{m}{j}\frac{(-2)^{m-j}}{\|e\|^{2m}}\frac{|u|^{2m+1}}{2\delta\,(m+j+1)}. \end{aligned} \tag{3.8}$$

Next, taking expected value on (3.8) with respect to $\varphi_n$ results in

$$\begin{aligned}\mathbb{E}_{\varepsilon_n,\varphi_n}\|z_n\|^{2p} &= \|z_{n-1}\|^{2p}+\frac{1}{2\delta}\sum_{m=1}^{p}\sum_{j=0}^{m}\binom{m}{j}\frac{(-2)^{m-j}}{m+j+1}\|z_{n-1}\|^{2(p-m)}\times\mathbb{E}_{\varphi_n}\left[\frac{|u|^{2m+1}}{\|\varphi_n\|^{2m}}\right] \\ &= \|z_{n-1}\|^{2p} \\ &\quad +\frac{1}{2\delta}\sum_{m=1}^{p}\left(\sum_{j=0}^{m}\binom{m}{j}\frac{(-2)^{m-j}}{m+j+1}\|z_{n-1}\|^{2(p-m)}\right) \\ &\quad \times\mathbb{E}_{\varphi_n}\left[\left|\left\langle z_{n-1},\frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^{2m+1}\|\varphi_n\|\right].\end{aligned}$$

Letting $\lambda := \mathbb{E}\|\varphi\|$, the fourth condition in Assumption 3.1.1 says

$$C_{2m+1}\lambda\|z_{n-1}\|^{2m+1}\le\mathbb{E}_{\varphi_n}\left[\left|\left\langle z_{n-1},\frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^{2m+1}\|\varphi_n\|\right]\le C'_{2m+1}\lambda\|z_{n-1}\|^{2m+1}.$$

So, by choosing $\tilde{C}_{2m+1}=C_{2m+1}$ if $m$ is odd and $\tilde{C}_{2m+1}=C'_{2m+1}$ if $m$ is even, one has

$$\begin{aligned}\mathbb{E}_{\varepsilon,\varphi_n}\|z_n\|^{2p} &\le \|z_{n-1}\|^{2p}+\frac{\lambda}{2\delta}\left(\sum_{m=1}^{p}\sum_{j=0}^{m}\binom{m}{j}\frac{(-2)^{m-j}}{m+j+1}\tilde{C}_{2m+1}\right)\|z_{n-1}\|^{2p+1} \\ &= \|z_{n-1}\|^{2p}+\frac{\lambda}{2\delta}\sum_{k=1}^{\lfloor p/2\rfloor}\left(A_{2k-1}C_{4k-1}+A_{2k}C'_{4k+1}\right)\|z_{n-1}\|^{2p+1} \\ &= \|z_{n-1}\|^{2p}+\frac{\lambda}{2\delta}\sum_{k=1}^{\lfloor p/2\rfloor}A_{2k-1}\left(C_{4k-1}-\frac{4k}{4k+1}C'_{4k+1}\right)\|z_{n-1}\|^{2p+1}.\end{aligned}$$

Because Assumption 3.1.1 requires $C_{4k-1}/C'_{4+1} > 4k/(4k+1)$ for all $k \geq 1$, if we set

$$M_{p,\delta} := \min\left\{\frac{pA}{\delta\sqrt{p+1}}, \frac{-\lambda}{2\delta}\sum_{k=1}^{\lfloor p/2 \rfloor} A_{2k-1}\left(C_{4k-1} - \frac{4k}{4k+1}C'_{4k+1}\right)\right\}, \tag{3.9}$$

then $M_{p,\delta} > 0$ and

$$\mathbb{E}_{\varepsilon,\varphi_n}\|z_n\|^{2p} \leq \|z_{n-1}\|^{2p} - M_{p,\delta}\|z_{n-1}\|^{2p+1}.$$

Finally, by taking the expected value with respect to $z_{n-1}$, it follows from Hölder's inequality that

$$\begin{aligned}\mathbb{E}\|z_n\|^{2p} &\leq \mathbb{E}_{z_{n-1}}\|z_{n-1}\|^{2p} - M_{p,\delta}\mathbb{E}_{z_{n-1}}\|z_{n-1}\|^{2p+1} \\ &\leq \mathbb{E}_{z_{n-1}}\|z_{n-1}\|^{2p} - M_{p,\delta}\left(\mathbb{E}_{z_{n-1}}\|z_{n-1}\|^{2p}\right)^{\frac{2p+1}{2p}}.\end{aligned}$$

At this point we set

$$K_{p,\sigma} := (2p/M_{p,\delta})^{2p}.$$

To apply Lemma 3.2.7, we need to make sure that $x = \mathbb{E}_{z_{n-1}}\|z_{n-1}\|^{2p} \leq (2p/M_{p,\delta})^{2p}/N^p$. Because $\|z_{n-1}\| \leq \|z_0\| \leq 2\delta/A$ implies $\mathbb{E}\|z_{n-1}\|^{2p} \leq (2\delta/A)^{2p}$, by bounding $1/N^p$ above with $1/(p+1)^p$ we obtain the necessary condition

$$M_{p,\delta} \leq \frac{pA}{\delta\sqrt{p+1}},$$

which is guaranteed by (3.9). So from Lemma 3.2.7 one has

$$\mathbb{E}\|z_n\|^{2p} \leq \frac{K_{p,\sigma}}{(n+1)^{2p}} = \left[\frac{2p}{M_{p,\delta}(n+1)}\right]^{2p}, \ \forall n \geq p+1.$$

Therefore the $2p$-th error moment of RG algorithm converges to zero in $\mathscr{O}\left(N^{-2p}\right)$ for all $p \in \mathbb{N}$ and the proof of theorem is complete.

□

## 3.3 General error moments

In this section we build up on Theorem 3.2.1 and provide bounds on the error moments under general initial condition where one may have $\|z_0\| > 2\delta/A$ while still under the assumption that $\|\varphi\| \leq A$. The idea is the following: because the error $z_j$ decreases monotonically, after many iterations it is unlikely to have $\|z_{n_0}\| > 2\delta/A$ for some $n_0 > 0$. In fact, as shown in Lemma 3.3.3, if $\|z_0\| > 2\delta/A$, then with positive probability, each iteration of RG reduces the estimate error by at least $\delta/A$. Hence, given initial condition $\|z_0\| > 0$, we can compute the expected number of steps needed for RG estimate error $z_j$ to fall into the range where we can apply Theorem 3.2.1. The first lemma provides a deterministic condition on the random variable $\varphi$.

**Lemma 3.3.1.** *Suppose that the random measurement vector $\varphi$ satisfies $\|\varphi\| \leq A$ and Assumption 3.1.1. Let*

$$B = \frac{1}{2}C_1\mathbb{E}\|\varphi\| \ \text{ and } \ \alpha = \left(\frac{B}{A}\right)^2,$$

*where $C_1$ is defined in Assumption 3.1.1. Then for all $x \in \mathbb{R}^d$ with $\|x\| > 2\delta/A$, we have*

$$\Pr\left[|\langle x, \varphi\rangle| \geq 2\delta\sqrt{\alpha} \ \text{ and } \ \|\varphi\| \geq B\right] \geq \frac{B}{A-B} > 0.$$

*Proof.* First note that $2B = C_1 \mathbb{E}\|\varphi\| \le A$ from Assumption 3.1.1. Also one has

$$\begin{aligned}
2B\|x\| &\le \mathbb{E}\,|\langle x,\varphi\rangle| \\
&= \mathbb{E}\left[|\langle x,\varphi\rangle| \,\middle|\, |\langle x,\varphi\rangle| > B\|x\|\right] \Pr\Big[\,|\langle x,\varphi\rangle| > B\|x\|\Big] \\
&+ \mathbb{E}\left[|\langle x,\varphi\rangle| \,\middle|\, |\langle x,\varphi\rangle| \le B\|x\|\right] \Pr\Big[\,|\langle x,\varphi\rangle| \le B\|x\|\Big] \\
&\le \|x\|A \cdot \Pr\Big[\,|\langle x,\varphi\rangle| > B\|x\|\Big] + B\|x\| \cdot \Pr\Big[|\langle x,\varphi\rangle| \le B\|x\|\Big] \\
&= \|x\|A \cdot \Pr\Big[\,|\langle x,\varphi\rangle| > B\|x\|\Big] + B\|x\|\left(1 - \Pr\Big[\,|\langle x,\varphi\rangle| > B\|x\|\Big]\right). \qquad (3.10)
\end{aligned}$$

Rearranging (3.10) to get

$$B\|x\| \le (A-B)\|x\|\Pr\Big[\,|\langle x,\varphi\rangle| > B\|x\|\Big],$$

and as a result

$$\Pr\big[\,|\langle x,\varphi\rangle| > B\|x\|\big] \ge \frac{B}{A-B}.$$

Note that $\|\varphi\| < B$ implies $|\langle x,\varphi\rangle| \le B\|x\|$, hence

$$\Pr\big[|\langle x,\varphi\rangle| > B\|x\| \text{ and } \|\varphi\| < B\big] = 0. \qquad (3.11)$$

Finally, because

$$B\|x\| \ge B\left(\frac{2\delta}{A}\right) = 2\delta\sqrt{\alpha},$$

from (3.11) we obtain

$$\begin{aligned}
&\Pr\big[|\langle x,\varphi\rangle| \ge 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \ge B\big] \\
&\quad \ge \Pr\big[\,|\langle x,\varphi\rangle| > B\|x\| \text{ and } \|\varphi\| \ge B\big] \\
&\quad = \Pr\big[\,|\langle x,\varphi\rangle| > B\|x\|\big] \\
&\quad \ge \frac{B}{A-B},
\end{aligned}$$

and the proof is complete. □

The next lemma provides a lower bound on the amount of estimate error reduced by RG at each step.

**Lemma 3.3.2.** *For fixed $\delta > 0$, let $\varepsilon$ be a random variable uniformly distributed on $[-\delta, \delta]$. Define $F_\delta : \mathbb{R} \to \mathbb{R}$ by*

$$F_\delta(u) = F_\delta(u, \varepsilon) := \frac{1}{\|e\|^2} \phi_\delta(u+\varepsilon) \left[\phi_\delta(u+\varepsilon) - 2u\right]. \tag{3.12}$$

*Fix $0 < \alpha < 1$, then for $u \in \mathbb{R}^d$,*

$$\Pr\left[F_\delta(u) < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] > \sqrt{3\alpha/4} > 0,$$

*for all $\lambda$ such that $-\alpha\delta^2 \leq B^2\lambda < 4\delta^2 - u^2$.*

*Proof.* Without loss of generality, one can assume that $u > 0$. Since $u > 0$, either $u > 2\delta$ or $2\delta\sqrt{\alpha} \leq u \leq 2\delta$ from the conditioning. So there are two possible cases depending on magnitude of $u$.

**Case 1. ($u > 2\delta$)** When $u > 2\delta$,

$$F_\delta(u, \varepsilon) = \frac{1}{\|e\|^2}(u + \varepsilon - \delta)\left[(\varepsilon - \delta) - u\right] = \frac{1}{\|e\|^2}\left[(\varepsilon - \delta)^2 - u^2\right].$$

Because $\|\varphi\| \geq B$, $\frac{1}{B^2}\left[(\varepsilon-\delta)^2-u^2\right] < \lambda \implies F_\delta(u,\varepsilon) < \lambda$ and hence

$$\begin{aligned}
&\Pr\left[F_\delta(u,\varepsilon) < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&\geq \Pr\left[\frac{1}{B^2}\left[(\varepsilon-\delta)^2-u^2\right] < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&= \Pr\left[(\varepsilon-\delta)^2-u^2 < B^2\lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&= \begin{cases} 0, & \text{if } B^2\lambda < -u^2, \\ \frac{\sqrt{u^2+B^2\lambda}}{2\delta}, & \text{if } -u^2 \leq B^2\lambda \leq 4\delta^2-u^2 \\ 1, & \text{if } B^2\lambda > 4\delta^2-u^2. \end{cases}
\end{aligned}$$

Since $u > 2\delta > 2\delta\sqrt{\alpha}$, we have

$$\begin{aligned}
&\Pr\left[F_\delta(u,\varepsilon) < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq A\right] \\
&\quad \geq \frac{\sqrt{u^2+B^2\lambda}}{2\delta} \geq \frac{\sqrt{4\alpha\delta^2-\alpha\delta^2}}{2\delta} = \sqrt{3\alpha/4} > 0,
\end{aligned}$$

for all $\lambda$ such that $-\alpha\delta^2 \leq B^2\lambda < 4\delta^2-u^2$.

**Case 2. ($2\delta\sqrt{\alpha} \leq u \leq 2\delta$)** In this case,

$$F_\delta(u,\varepsilon) = \begin{cases} 0, & \text{if } -\delta \leq \varepsilon \leq \delta-u, \\ (\varepsilon-\delta)^2-u^2, & \text{if } \delta-u < \varepsilon \leq \delta. \end{cases}$$

It follows that

$$
\begin{aligned}
&\Pr\left[F_\delta(u,\varepsilon) < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&= \begin{cases} 0, & \text{if } B^2\lambda < -u^2, \\ \frac{\sqrt{u^2+B^2\lambda}}{2\delta}, & \text{if } -u^2 \leq B^2\lambda \leq 0, \\ 1, & \text{if } B^2\lambda > 0. \end{cases}
\end{aligned}
$$

Since $u > 2\delta\sqrt{\alpha}$, we have similarly

$$
\begin{aligned}
&\Pr\left[F_\delta(u,\varepsilon) < \lambda \,\middle|\, |u| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&\geq \frac{\sqrt{u^2+B^2\lambda}}{2\delta} \geq \frac{\sqrt{4\alpha\delta^2-\alpha\delta^2}}{2\delta} = \sqrt{3\alpha/4} > 0,
\end{aligned}
$$

for $\lambda$ such that $-\alpha\delta^2 \leq B^2\lambda < 0$.

□

Following Lemma 3.3.1 and Lemma 3.3.2, the next lemma combines the result and provides a lower bound on the amount of error reduced whenever the input $x_j$ of RG is more than $2\delta/A$ away from the true $x$.

**Lemma 3.3.3.** *Fix $\delta > 0$. Suppose $\varepsilon$ is a uniform random variable on $[-\delta,\delta]$ and the random vector $e \in \mathbb{R}^d$ satisfies Assumption 3.1.1. Assume also that $\varphi$, $\varepsilon$ are independent. Let $F_\delta$ be defined as in (3.12), then with the constants A, B, and $\alpha$ defined in Lemma 3.3.1, for $z \in \mathbb{R}^d$ with $\|z\| > 2\delta/A$ we have*

$$
Pr\left[F_\delta\left(\langle z,\varphi\rangle,\varepsilon\right) < -\left(\frac{\delta}{A}\right)^2\right] > \frac{\sqrt{3}B^2}{2A(A-B)} =: C_{A,B} > 0.
$$

*Proof.* The result follows immediately from from Lemma 3.3.1 and 3.3.2 with $\alpha = (B/A)^2$

since

$$
\begin{aligned}
&\Pr\left[F_\delta\left(\langle z,\varphi\rangle,\varepsilon\right) < -\left(\frac{\delta}{A}\right)^2\right] \\
=&\Pr\left[F_\delta\left(\langle z,\varphi\rangle,\varepsilon\right) < \frac{-\alpha\delta^2}{B^2}\right] \\
\geq&\Pr\left[F_\delta\left(\langle z,\varphi\rangle,\varepsilon\right) < \frac{-\alpha\delta^2}{B^2}\,\middle|\, |\langle z,\varphi\rangle| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
&\cdot \Pr\left[|\langle z,\varphi\rangle| \geq 2\delta\sqrt{\alpha} \text{ and } \|\varphi\| \geq B\right] \\
>&\frac{B}{A-B}\sqrt{\frac{3\alpha}{4}} > 0.
\end{aligned}
$$

□

The next lemma gives a bound on the probability of having the estimate error $z_n$ staying outside the "good" region where $\|z_n\| \leq 2\delta/A$, after $n$ iterations of RG.

**Lemma 3.3.4.** *Under the same assumptions and notations in Lemma 3.3.3, if we set $M$ to be the smallest positive integer such that*

$$\|z_0\|^2 - M(\delta/A)^2 < (2\delta/A)^2,$$

*then for all $n > M-1$ one has*

$$Pr\left[\|z_n\| > \frac{2\delta}{A}\right] \leq an^b c^n,$$

*where*

$$a = M(1-C_{A,B})^{-M},\, b = (M-1),\, c = (1-C_{A,B}). \tag{3.13}$$

*Proof.* Let $\mathscr{I}$ be the event $\mathscr{I} = \left\{\|z_j\| > 2\delta/A,\ \forall j = 0,\ldots,n-1\right\}$. Because

$$\|z_n\|^2 = \|z_0\|^2 + \sum_{j=0}^{n-1} F_j(\langle z_{j-1},\varphi_j\rangle,\varepsilon_j),$$

we define $X$ to be the random variable

$$X = \text{Card}\left\{ j : F_j(\langle z_{j-1}, \varphi_j \rangle, \varepsilon_j) < -\left(\frac{\delta}{A}\right)^2,\ j = 0, 1, \ldots, n-1 \right\}.$$

$X$ can be viewed as the number of "good" RG iterations where we have $z_j < z_{j-1} - (\delta/A)$ and the RG estimate error decreases for more than $\delta/A$. From the definition of $X$ and $M$, we have

$$\Pr\left[X \geq M \,|\, \mathscr{I}\right] \leq \Pr\left[\|z_n\| \leq \frac{2\delta}{A} \,\middle|\, \mathscr{I}\right],$$

which implies

$$\Pr\left[\|z_n\| > \frac{2\delta}{A} \,\middle|\, \mathscr{I}\right] \leq \Pr\left[X < M \,|\, \mathscr{I}\right].$$

However, having $\|z_{j+1}\| \leq \|z_j\|$ for all $j = 0, 1, \ldots, n-1$ tells us that

$$\Pr\left[\|z_n\| > \frac{2\delta}{A} \,\middle|\, \mathscr{I}^c\right] = 0,$$

so

$$\Pr\left[\|z_n\| > \frac{2\delta}{A}\right] \leq \Pr\left[X < M \,|\, \mathscr{I}\right].$$

The next step is to bound $\Pr\left[X < M \,|\, \mathscr{I}\right]$, the probability of having less than $M$ good RG iterations. We can do so by bounding the probability of having exactly $k$ good iterations, where $k < M$ , as follows. For any $0 \leq k < M$, fix a set of indices

$$\{j_0 < j_1 < \ldots < j_{n-k}\} \subset \{1, \ldots, n\},$$

where $j_0, j_1, \ldots, j_{n-k}$ are indices of bad iterations during which the RG estimate error possibly decreases for less than or equal to $\delta/A$. Then, from the independency assumption on $\varphi_j$ and $\varepsilon_j$, we can apply Lemma 3.3.3 on the largest index $j_{n-k}$ to get a bound on the probability of having bad iterations at exactly the $j_0, \ldots, j_{n-k}$-th iterations of RG. Hence,

$$\begin{aligned}
&\Pr\left[F_{j_l}(\langle z_{j_l-1}, \varphi_{j_l}\rangle, \varepsilon_{j_l}) \geq -\left(\frac{\delta}{A}\right)^2, l = 1, \ldots, n-k \,\middle|\, \mathscr{I}\right] \\
&\quad = \Pr\left[F_{j_l}(\langle z_{j_l-1}, \varphi_{j_l}\rangle, \varepsilon_{j_l}) \geq -\left(\frac{\delta}{A}\right)^2, l = 1, \ldots, n-k-1 \,\middle|\, \mathscr{I}\right] \\
&\quad\quad \times \Pr\left[F_{j_{n-k}}(\langle z_{j_{n-k}-1}, \varphi_{j_{n-k}}\rangle, \varepsilon_{j_{n-k}}) \geq -\left(\frac{\delta}{A}\right)^2 \,\middle|\, \mathscr{I}\right] \\
&\quad \leq (1 - C_{A,B}) \cdot \Pr\left[F_{j_l}(\langle z_{j-1}, \varphi_{j_l}\rangle, \varepsilon_{j_l}) \geq -\left(\frac{\delta}{A}\right)^2, l = 1, \ldots, n-k-1 \,\middle|\, \mathscr{I}\right].
\end{aligned}$$

By repeated application of Lemma 3.3.3 on the largest remaining index $j_l$, we obtain an upper bound on the probability of having $k$ good iterations:

$$\Pr\left[F_{j_l}(\langle z_{j_l-1}, \varphi_{j_l}\rangle, \varepsilon_{j_l}) \geq -\left(\frac{\delta}{A}\right)^2, l = 1, \ldots, n-k \,\middle|\, \mathscr{I}\right] \leq (1 - C_{A,B})^{n-k}. \tag{3.14}$$

Therefore for any $k < M$,

$$\Pr[X = k \,|\, \mathscr{I}] \leq \binom{n}{k} (1 - C_{A,B})^{n-k}.$$

Finally, for $n > M-1$,

$$\begin{aligned}
\Pr\left[\|z_n\| > \frac{2\delta}{A}\right] &\leq \Pr\left[X < M \,|\, \mathscr{I}\right] \\
&= \sum_{k=0}^{M-1} \Pr\left[X = k \,|\, \mathscr{I}\right] \\
&\leq \sum_{k=0}^{M-1} \binom{n}{k} (1 - C_{A,B})^{n-k} \\
&\leq (1 - C_{A,B})^{n-M} \sum_{k=0}^{M-1} \binom{n}{k} \\
&\leq (1 - C_{A,B})^{n-M} \sum_{k=0}^{M-1} n^k \\
&\leq M n^{M-1} (1 - C_{A,B})^{n-M}.
\end{aligned}$$

□

**Lemma 3.3.5.** *Given* $b > 0$ *and* $0 < c < 1$, *then for each* $p \in \mathbb{N}$, *we have*

$$\forall n > 1, \qquad \sum_{j=1}^{n} \frac{j^b c^j}{(n-j)^{2p}} \leq \frac{D_p}{n^{2p}},$$

*where* $D_p$ *is defined in* (3.15).

*Proof.* Because $0 < c < 1$, given $p > 0$, the product $j^{b+2p} c^j$ is dominated by $c^j$. Hence we can find constants $\widetilde{D}_p = \widetilde{D}_p(b,c)$ such that $j^{b+2p} c^j \leq \widetilde{D}_p$ for all $j \in \mathbb{N}$. Therefore,

$$\begin{aligned}
\sum_{j=1}^{n} \frac{j^b c^j}{(n-j)^{2p}} &\leq \widetilde{D}_p \sum_{j=1}^{\lceil n/2 \rceil} \frac{1}{j^{2p}(n-j)^{2p}} + \widetilde{D}_p \sum_{j=\lfloor n/2 \rfloor}^{n} \frac{1}{j^{2p}(n-j)^{2p}} \\
&\leq 2\widetilde{D}_p \left(\frac{2}{n}\right)^{2p} \sum_{j=1}^{\infty} \frac{1}{j^{2p}} \leq \frac{D_p}{n^{2p}},
\end{aligned}$$

where

$$D_p = 2\widetilde{D}_p \cdot 2^{2p} \cdot \sum_{j=1}^{\infty} \frac{1}{j^{2p}}. \tag{3.15}$$

□

We are now ready to prove the theorem.

**Theorem 3.3.6.** *Fix $p \in \mathbb{N}$. Suppose that $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ are i.i.d. random vectors with $\|\varphi_n\| \leq A$, that $\{\varepsilon_n\}_{n=1}^N$ are i.i.d. uniform random variables on $[-\delta, \delta]$, and that Assumption 3.1.1 holds. Given $x \in \mathbb{R}^d$ and noisy measurements $\{\langle x, \varphi_n\rangle + \varepsilon_n\}_{n=1}^N$, let $\{x_n\}_{n=0}^N \in \mathbb{R}^d$ denote the RG estimate of $x$ with initial estimate $x_0 \in \mathbb{R}^d$. Denote the error of RG estimate at the n-th iteration by $z_n := x - x_n$. Then for all $N > p+1$, we have*

$$\mathbb{E}\|z_N\|^{2p} \leq \|z_0\|^{2p}(p+1)aN^b c^{N-p} + \left(\frac{2p}{M_{p,\delta}}\right)^{2p} \frac{D_p}{[N-(p+1)]^{2p}}, \tag{3.16}$$

*where $a, b, c$ are defined in* (3.13)*, $M_{p,\delta}$ is defined in* (3.5)*, and $D_p$ is defined in* (3.15)*.*

*Proof.* First, we condition $\mathbb{E}\|z_N\|^{2p}$ with the probability that the error $z_j$ falls into the applicable range of Theorem 3.2.1. The sum is split at $j = N-p$ because Theorem 3.2.1 does not apply to the first $p+1$ terms. Therefore,

$$\begin{aligned}\mathbb{E}\|z_N\|^{2p} = \mathbb{E}&\left[\|z_N\|^{2p} \,\middle|\, \|z_{N-1}\| > \frac{2\delta}{A}\right] \Pr\left[\|z_{N-1}\| > \frac{2\delta}{A}\right] \\ &+ \sum_{j=N-p}^{N-1} \left[\|z_N\|^{2p} \,\middle|\, \|z_j\| > \frac{2\delta}{A}, \|z_{j-1}\| \leq \frac{2\delta}{A}\right] \Pr\left[\|z_j\| > \frac{2\delta}{A}, \|z_{j-1}\| \leq \frac{2\delta}{A}\right] \\ &+ \sum_{j=1}^{N-p-1} \left[\|z_N\|^{2p} \,\middle|\, \|z_j\| > \frac{2\delta}{A}, \|z_{j-1}\| \leq \frac{2\delta}{A}\right] \Pr\left[\|z_j\| > \frac{2\delta}{A}, \|z_{j-1}\| \leq \frac{2\delta}{A}\right].\end{aligned} \tag{3.17}$$

Next, we bound the probability in the first two terms of (3.17) with Lemma 3.3.4 and monotonicity of $z_j$ that $\|z_{j+1}\| \leq \|z_j\|$. We also bound the third term in (3.17) by applying Theorem 3.2.1, Lemma 3.3.4, and Lemma 3.3.5 to obtain

$$\begin{aligned}\mathbb{E}\|z_N\|^{2p} &\leq \|z_0\|^{2p} \sum_{j=N-p}^{N} a j^b c^j + \sum_{j=1}^{N-p-1} \left[\frac{2p}{M_{p,\delta}((n-j)+1)}\right]^{2p} a j^b c^j \\ &\leq \|z_0\|^{2p}(p+1)aN^b c^{N-p} + \left(\frac{2p}{M_{p,\delta}}\right)^{2p} \frac{D_p}{[N-(p+1)]^{2p}}.\end{aligned}$$

□

Because the first term in (3.16) is dominated by $c^N$ with $0 < c < 1$, we have the following corollary.

**Corollary 3.3.7.** *With the assumptions and notations in Theorem 3.3.6, for each* $p \in \mathbb{N}$ *there exists a constant* $C > 0$ *such that for* $N > p+1$*,*

$$\mathbb{E}\|z_N\|^{2p} \leq \frac{C}{N^{2p}}.$$

BIBLIOGRAPHY


[1] X. Chen and A. M. Powell. Almost sure convergence of the Kaczmarz algorithm with random measurements. *J. Fourier Anal. Appl.*, 18(6):1195–1214, 2012.

[2] Z. Cvetković. Resilience properties of redundant expansions under additive noise and quantization. *IEEE Trans. Inform. Theory*, 49(3):644–656, 2003.

[3] Z. Cvetković and M. Vetterli. On simple oversampled A/D conversion in $L^2(\mathbb{R})$. *IEEE Trans. Inform. Theory*, 47(1):146–154, 2001.

[4] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.14 of 2016-12-21. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[5] L. Flatto and D. J. Newman. Random coverings. *Acta Math.*, 138(3–4):241–264, 1977.

[6] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[7] V. K. Goyal, J. Kovačević, and J. A. Kelner. Quantized frame expansions with erasures. *Appl. Comput. Harmon. Anal.*, 10(3):203–233, 2001.

[8] V. K. Goyal, M. Vetterli, and N. T. Thao. Quantized overcomplete expansions in $\mathbf{R}^N$: analysis, synthesis, and algorithms. *IEEE Trans. Inform. Theory*, 44(1):16–31, 1998.

[9] C. Heil. *A basis theory primer*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, expanded edition, 2011.

[10] G. T. Herman. *Fundamentals of computerized tomography*. Advances in Pattern Recognition. Springer, Dordrecht, second edition, 2009. Image reconstruction from projections.

[11] L. Jacques, D. K. Hammond, and J. M. Fadili. Dequantizing compressed sensing: when oversampling and non-Gaussian constraints combine. *IEEE Trans. Inform. Theory*, 57(1):559–571, 2011.

[12] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Trans. Inform. Theory*, 59(4):2082–2102, 2013.

[13] D. Jimenez, L. Wang, and Y. Wang. White noise hypothesis for uniform quantization errors. *SIAM J. Math. Anal.*, 38(6):2042–2056, 2007.

[14] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de lAcademie Polonaise des Sciences et des Lettres*, 35:355–357, 1937.

[15] J. Matoůsek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[16] D. Needell, N. Srebro, and R. Ward. Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. *Math. Program.*, 155(1-2, Ser. A):549–573, 2016.

[17] Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming. *Comm. Pure Appl. Math.*, 66(8):1275–1297, 2013.

[18] A. M. Powell. Mean squared error bounds for the Rangan-Goyal soft thresholding algorithm. *Appl. Comput. Harmon. Anal.*, 29(3):251–271, 2010.

[19] A. M. Powell and J. T. Whitehouse. Error bounds for consistent reconstruction: random polytopes and coverage processes. *Found. Comput. Math.*, 16(2):395–423, 2016.

[20] S. Rangan and V. K. Goyal. Recursive consistent estimation with bounded noise. *IEEE Trans. Inform. Theory*, 47(1):457–464, 2001.

[21] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[22] L. Roberts. Picture coding using pseudo-random noise. *IRE Transactions on Information Theory*, 8(2):145–154, February 1962.

[23] L. Schuchman. Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, 12(4):162–165, December 1964.

[24] J. R. Shewchuk et al. An introduction to the conjugate gradient method without the agonizing pain, 1994.

[25] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *J. Appl. Probab.*, 19(1):221–228, 1982.

[26] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.

[27] N. T. Thao and M. Vetterli. Deterministic analysis of oversampled a/d conversion and decoding improvement based on consistent estimates. *IEEE Transactions on Signal Processing*, 42(3):519–531, Mar 1994.

[28] N. T. Thao and M. Vetterli. Reduction of the mse in r-times oversampled a/d conversion o(1/r) to o(1/r2). *IEEE Transactions on Signal Processing*, 42(1):200–203, Jan 1994.