ALMOST SURE CONVERGENCE OF THE KACZMARZ ALGORITHM WITH RANDOM MEASUREMENTS

XUEMEI CHEN AND ALEXANDER M. POWELL

Abstract. The Kaczmarz algorithm is an iterative method for reconstructing a signal $x \in \mathbb{R}^d$ from an overcomplete collection of linear measurements $y_n = \langle x, \varphi_n \rangle$, $n \geq 1$. We prove quantitative bounds on the rate of almost sure exponential convergence in the Kaczmarz algorithm for suitable classes of random measurement vectors $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$. Refined convergence results are given for the special case when each φ_n has i.i.d. Gaussian entries and, more generally, when each $\varphi_n / \|\varphi_n\|$ is uniformly distributed on $\mathbb{S}^{d-1}$. This work on almost sure convergence complements the mean squared error analysis of Strohmer and Vershynin for randomized versions of the Kaczmarz algorithm.

1. Introduction

The Kaczmarz algorithm is a classical iterative method for solving an overdetermined consistent linear system $\Phi x = y$. The algorithm is based on the mechanism of projection onto convex sets and also falls into the class of row-action methods. Within the spectrum of linear solvers, some key features of the Kaczmarz algorithm are its scalability and its simplicity; a single inner product is the dominant computation in each step of the algorithm. This has made the Kaczmarz algorithm a good candidate for high dimensional problems.

To describe the Kaczmarz algorithm let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ be a given spanning set for $\mathbb{R}^d$. Suppose that $x \in \mathbb{R}^d$ is unknown but that one is given access to the linear measurements $y_n = \langle x, \varphi_n \rangle$, for $n \geq 1$. The Kaczmarz algorithm is used to approximately recover x from the linear measurements $\{y_n\}_{n=1}^\infty$.

The Kaczmarz algorithm starts with an arbitrary initial estimate $x_0 \in \mathbb{R}^d$ and produces approximate solutions $x_n \in \mathbb{R}^d$ by the following iteration:

$$\forall\, n \geq 1, \qquad x_n = x_{n-1} + \frac{y_n - \langle \varphi_n, x_{n-1} \rangle}{\|\varphi_n\|^2}\, \varphi_n. \tag{1.1}$$

Geometrically, this is an iterative projection algorithm that updates the estimate $x_{n-1} \in \mathbb{R}^d$ by orthogonally projecting it onto the affine hyperplane

$$H_n = \{u \in \mathbb{R}^d : \langle u, \varphi_n \rangle = y_n\}.$$

The initial convergence analysis for this algorithm in [7] focuses on finite dimensional spaces, but there are also subsequent extensions to infinite dimensional spaces, e.g., in [5, 8, 23].

The original work of Kaczmarz applies (1.1) to a fixed $N \times d$ system $\Phi x = y$ by iterating the algorithm on an infinite periodic extension of the system, see [7]. In particular, for $n > N$

Date: June 4, 2012.

Key words and phrases. Almost sure convergence, convergence rates, Kaczmarz algorithm.

one defines $\varphi_n = \varphi_{p(n)}$ and $y_n = y_{p(n)}$ where $p(n) \in \{1, \cdots, N\}$ is the unique integer such that $p(n) \equiv n$ modulo N. Kaczmarz showed that iteratively cycling through the system in this manner produces estimates x_n that are guaranteed to converge to x:

$$\lim_{n\to\infty} \|x - x_n\| = 0.$$

It is well known that the algorithm (1.1) produces monotonically improving approximations as the iteration number increases. Specifically, for any $x \in \mathbb{R}^d$ and $\{\varphi_n\}_{n=1}^N \subset \mathbb{R}^d$ and any initial estimate $x_0 \in \mathbb{R}^d$ the Kaczmarz algorithm satisfies

$$\|x - x_{n+1}\| \leq \|x - x_n\|. \tag{1.2}$$

This basic fact can, for example, be seen as a corollary to Proposition 3.1. However, it can be difficult to quantify the associated rates of convergence in (1.1). Geometric considerations imply that the specific rate at which the Kaczmarz algorithm converges depends strongly on the order in which measurements y_n are entered into the algorithm, and in certain circumstances the convergence can be quite slow. Motivated by this, Strohmer and Vershynin [20, 21] investigated a *randomized* version of the Kaczmarz algorithm where the new information (y_n, φ_n) processed at each step of the algorithm (1.1) is randomly selected from among the N measurements. They proved that this randomized approach achieves mean squared error with a rate that is quantifiable in terms of a particular matrix condition number $\kappa(\Phi)$ as

$$\mathbb{E}\|x - x_n\|^2 \leq (1 - \kappa(\Phi)^{-2})^n \ \|x - x_0\|^2. \tag{1.3}$$

The theoretical and numerical analysis of the randomized Kaczmarz algorithm in [21] shows that this method converges exponentially fast and has features that are competitive with (and sometimes superior to) standard approaches such as the conjugate gradient method.

In addition to the analysis of mean squared convergence rates, there is recent work that highlights other favorable properties of the Kaczmarz algorithm. The work in [12] shows that the algorithm is robust against noise in the measurements y_n. There is work in [4] on accelerating the convergence of the Kaczmarz algorithm in high dimensions with help of the Johnson-Lindenstrauss Lemma. The discussion in [1, 2, 22] addresses choices of randomization for the algorithm.

The Kaczmarz algorithm and its variants appear in a wide variety of settings. For example, it has been applied to computer tomography and image processing in [11, 17], and used for sparse signal recovery in compressed sensing in [19]. In signal processing, the closely related Rangan-Goyal algorithm is used for consistent reconstruction of quantized data, see [14, 13].

Overview and main results. The main aim of this article is to study the issue of *almost sure convergence rates* for the Kaczmarz algorithm with random measurement vectors $\{\varphi_n\}_{n=1}^\infty$. We prove that the Kaczmarz algorithm *almost surely* converges exponentially fast and we provide quantitative bounds on the associated convergence rate.

The paper is organized as follows. Section 2 provides definitions and background properties of the random measurement vectors $\{\varphi_n\}_{n=1}^\infty$. Section 3 gives basic formulas for the error $\|x - x_n\|$ in the Kaczmarz algorithm, and Section 4 gives basic bounds on the moments $\mathbb{E}\|x - x_n\|^{2s}$ with $s > 0$.

Our main results appear in Section 5 and Section 6. Our first main result, Theorem 5.3 in Section 5, provides sharp almost sure rates of exponential convergence for the Kaczmarz algorithm in the important case when the normalized measurement vectors $\varphi_n/\|\varphi_n\|$ are independent and uniformly distributed on $\mathbb{S}^{d-1}$ (for example, this applies to random vectors φ_n with i.i.d. Gaussian entries). Our next main results, Theorem 6.2 and Theorem 6.3 in Section 6, provide quantitive bounds on the rate of almost sure exponential convergence for general classes of random measurement vectors.

2. Random measurements

This section will discuss conditions on the random measurement vectors $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ which will be needed in our analysis of almost sure convergence in the Kaczmarz algorithm.

Suppose that the random measurement vectors $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ are used for the Kaczmarz algorithm (1.1). We always assume that each φ_n is almost surely nonzero, $\Pr[\,\|\varphi_n\| = 0\,] = 0$, to ensure that the Kaczmarz iteration (1.1) is well defined. Since most of our error analysis only involves the normalized random vectors $\varphi_n/\|\varphi_n\|$, the assumption that each φ_n is almost surely nonzero also guarantees that each $\varphi_n/\|\varphi_n\|$ is well defined.

Our general analysis of the Kaczmarz algorithm will require that the normalized random measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ be independent but not necessarily identically distributed. Since it is common in practice to make assumptions directly on the measurement vectors $\{\varphi_n\}_{n=1}^{\infty}$, it is useful to note that independence of the measurement vectors $\{\varphi_n\}_{n=1}^{\infty}$ is a strictly stronger assumption than independence of the normalized measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$. Our analysis will allow the possibility of non-independent $\{\varphi_n\}_{n=1}^{\infty}$, but will always require that $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ be independent.

Lemma 2.1. *If the random vectors $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ are independent and almost surely nonzero then the normalized random vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are also independent.*

As mentioned above, the converse of Lemma 2.1 is not true.

Example 2.2. Let θ_1, θ_2 be independent random variables that are uniform on $[0, 2\pi)$. Define the random vectors $\varphi_1 = (\cos\theta_1, \sin\theta_1)$ and φ_2 as follows:

$$\varphi_2 = \begin{cases} (\cos\theta_2, \sin\theta_2), & \text{if } 0 \le \theta_1 < \pi, \\ 2(\cos\theta_2, \sin\theta_2), & \text{if } \pi \le \theta_1 < 2\pi. \end{cases}$$

Then $\varphi_1/\|\varphi_1\|$ and $\varphi_2/\|\varphi_2\|$ are independent, but φ_1, φ_2 are not independent.

Our analysis of almost sure convergence will involve the following frame-type assumptions on the normalized random measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$.

Definition 2.3. Let $s > 0$ be fixed. The unit-norm random vector $u \in \mathbb{R}^d$ has the *Kaczmarz bound* $0 < \alpha < 1$ *of order* s if

$$\forall x \in \mathbb{S}^{d-1}, \qquad \left(\mathbb{E}\left(1 - |\langle x, u\rangle|^2\right)^s\right)^{1/s} \le \alpha. \tag{2.1}$$

If (2.1) holds with equality then we shall say that the Kaczmarz bound is *tight*.

Convergence rates in the Kaczmarz algorithm will depend on the specific value of the Kaczmarz bound $0 < \alpha < 1$. Qualitatively, if $u \in \mathbb{R}^d$ is a given random vector and $s > 0$ is fixed, note that (2.1) holds for some $0 < \alpha < 1$ if and only if u is not concentrated on a subspace of $\mathbb{R}^d$ with positive codimension.

In the special case when $s = 1$, Definition 2.3 reduces to the notion of probabilistic frame and deserves further mention.

Definition 2.4. The random vector $u \in \mathbb{R}^d$ has the *probabilistic lower frame bound* $\beta > 0$ if

$$\forall x \in \mathbb{R}^d, \quad \mathbb{E}|\langle x, u\rangle|^2 \geq \beta \|x\|^2. \tag{2.2}$$

The random vector $u \in \mathbb{R}^d$ is a *tight probabilistic frame* if (2.2) holds with equality

$$\forall x \in \mathbb{R}^d, \quad \mathbb{E}|\langle x, u\rangle|^2 = \beta \|x\|^2. \tag{2.3}$$

If $u \in \mathbb{S}^{d-1}$ is a unit-norm tight probabilistic frame we shall simply say that u is *isotropic*.

Thus, a Kaczmarz bound $0 < \alpha < 1$ of order $s = 1$ corresponds to a probabilistic frame bound $\beta = 1 - \alpha$. A condition similar to (2.2) was used for the analysis of the Rangan-Goyal algorithm in [14], cf. [13]. Random vectors satisfying the probabilistic tight frame condition (2.3) are fully characterized in [3] and it is shown that if u is isotropic then the constant β in (2.3) must satisfy

$$\beta = \beta_d = 1/d. \tag{2.4}$$

We refer the reader to [3] for further background on the interesting properties of probabilistic frames.

Example 2.5. If $u \in \mathbb{R}^d$ is uniformly distributed on $\mathbb{S}^{d-1}$ then u is isotropic.

Example 2.6. Let $\{f_n\}_{n=1}^N \subset \mathbb{S}^{d-1}$ be a deterministic unit-norm tight frame for $\mathbb{R}^d$, i.e.,

$$\forall x \in \mathbb{R}^d, \quad \|x\|^2 = \frac{d}{N} \sum_{n=1}^{N} |\langle x, f_n\rangle|^2.$$

If the discrete random vector $u \in \mathbb{R}^d$ is defined to be uniformly distributed on the set $\{f_n\}_{n=1}^N$, then u satisfies (2.3). For example, if $\{f_n\}_{n=1}^d \subset \mathbb{R}^d$ is an orthonormal basis for $\mathbb{R}^d$ and $u \in \mathbb{R}^d$ randomly selects an element of this basis then the random vector u satisfies (2.3).

Example 2.7. Let F be a full rank $N \times d$ matrix and let $\{f_n\}_{n=1}^N \subset \mathbb{R}^d$ be the rows of F. Let $u \in \mathbb{R}^d$ be the discrete random vector with the probability mass function

$$\forall\ 1 \leq k \leq N, \quad \Pr[u = f_k] = \|f_k\|^2 / \sum_{n=1}^{N} \|f_n\|^2.$$

It was shown in [21] that u has a probabilistic lower frame bound $\beta > 0$ that satisfies

$$\beta \geq \left(\frac{1}{\kappa(F)}\right)^2 = \frac{1}{\|F\|_{\mathrm{Fr}}^2 \|F^{-1}\|_2^2}. \tag{2.5}$$

For our analysis of almost sure convergence it will be useful to have a version of Definition 2.3 for the limiting case $s = 0$. The following standard lemma will be useful for this (for example, see page 71 of [15]).

Lemma 2.8. *Let η be a random variable such that $\mathbb{E}|\eta|^s < \infty$ for some $s > 0$. Then*

$$\inf_{s>0}(\mathbb{E}|\eta|^s)^{1/s} = \lim_{s\to 0}(\mathbb{E}|\eta|^s)^{1/s} = \exp(\mathbb{E}\log|\eta|).$$

Corollary 2.9. *If $u \in \mathbb{S}^d$ is a random unit-vector then*

$$\forall x \in \mathbb{S}^d, \quad \lim_{s\to 0}(\mathbb{E}(1-|\langle x,u\rangle|^2)^s)^{1/s} = \exp\left(\mathbb{E}[\log(1-|\langle x,u\rangle|^2)]\right). \tag{2.6}$$

In both Lemma 2.8 and Corollary 2.9 we interpret $\exp(-\infty) = 0$. Motivated by Corollary 2.9, the following definition will naturally arise in our analysis of almost sure convergence in the Kaczmarz algorithm.

Definition 2.10. The random unit-vector $u \in \mathbb{S}^{d-1}$ has a *logarithmic Kaczmarz bound* $0 < \rho < 1$ if

$$\forall x \in \mathbb{S}^{d-1}, \quad \exp\left(\mathbb{E}[\log(1-|\langle x,u\rangle|^2)]\right) \le \rho. \tag{2.7}$$

We say that $u \in \mathbb{S}^{d-1}$ has a *tight logarithmic Kaczmarz bound* ρ if (2.7) holds with equality.

For perspective, $\mathbb{E}[\log(1-|\langle x,u\rangle|^2)]$ in (2.7) can be expressed as a perturbation of the familiar logarithmic potential [16] by

$$\forall x \in \mathbb{S}^{d-1}, \quad \mathbb{E}[\log(1-|\langle x,u\rangle|^2)] = 2\mathbb{E}[\log\|x-u\|] + \mathbb{E}[\log(1-4^{-1}\|x-u\|^2)].$$

Note that for $x, u \in \mathbb{S}^{d-1}$, $L(x,u) = \log(1-|\langle x,u\rangle|^2)$ is singular at both $u = x$ and $u = -x$.

Random vectors $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ with the following properties will play an important role in Section 5. For convenience we collect these properties in the following definition.

Definition 2.11. We shall say that the random vectors $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ have the *normalized independence and uniformity (NIU) property* if each φ_n is almost surely nonzero and if the normalized vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent and uniformly distributed on $\mathbb{S}^{d-1}$.

Lemma 2.1 and Example 2.2 provide insight into the assumption in Definition 2.11 that $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ be independent. The following examples provide some insight into the condition that each $\varphi_n/\|\varphi_n\|$ is uniformly distributed on $\mathbb{S}^{d-1}$.

Example 2.12. Let $u \in \mathbb{R}^d$ be a uniform random vector on $\mathbb{S}^{d-1}$. We shall consider a random vector $\varphi \in \mathbb{R}^d$ to be *radial* if it is of the form $\varphi = ru$ where $r \in \mathbb{R}$ is a random variable that is independent of u. If the random vector $\varphi \in \mathbb{R}^d$ is radial and almost surely nonzero then $\varphi/\|\varphi\|$ is uniform on $\mathbb{S}^{d-1}$. For example, if $\varphi \in \mathbb{R}^d$ is a random Gaussian vector with i.i.d. $N(0,1)$ entries then $\varphi/\|\varphi\|$ is uniformly distributed on $\mathbb{S}^{d-1}$.

Example 2.13. Let θ be uniformly distributed on $[0, 2\pi)$. Define the random vector $\varphi \in \mathbb{R}^2$ by

$$\varphi = \begin{cases} (\cos\theta, \sin\theta), & \text{if } 0 \le \theta < \pi, \\ 2(\cos\theta, \sin\theta), & \text{if } \pi \le \theta < 2\pi. \end{cases}$$

Then $\varphi/\|\varphi\|$ is uniformly distributed on $\mathbb{S}^1$ but $\varphi \in \mathbb{R}^2$ is not radial.

3. Basic error formulas for the Kaczmarz algorithm

The following error formulas for the Kaczmarz algorithm will play an important role throughout this paper.

Proposition 3.1. *Suppose that $x \in \mathbb{R}^d$ and that the measurement vectors $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ are nonzero. Suppose that the measurements $y_n = \langle x, \varphi_n \rangle$, with $n \geq 1$, are used as input to the Kaczmarz algorithm with initial estimate $x_0 \in \mathbb{R}^d$.*

The error $z_n = x - x_n$ after the nth iteration of the Kaczmarz algorithm satisfies

$$\|z_n\|^2 = \|z_{n-1}\|^2 - \left| \left\langle z_{n-1}, \frac{\varphi_n}{\|\varphi_n\|} \right\rangle \right|^2 \tag{3.1}$$

and

$$\|z_n\|^2 = \|z_0\|^2 \prod_{k=1}^{n} \left(1 - \left| \left\langle \frac{z_{k-1}}{\|z_{k-1}\|}, \frac{\varphi_k}{\|\varphi_k\|} \right\rangle \right|^2 \right). \tag{3.2}$$

We adopt the convention that $z_{k-1}/\|z_{k-1}\| = 0$ is the zero vector when $\|z_{k-1}\| = 0$.

Proof. The defining iteration (1.1) can be written in terms of the error $z_n = x - x_n$ as

$$z_n = z_{n-1} - \left\langle z_{n-1}, \frac{\varphi_n}{\|\varphi_n\|} \right\rangle \frac{\varphi_n}{\|\varphi_n\|}.$$

Since φ_n is orthogonal to z_n, the equation (3.1) now follows:

$$\|z_n\|^2 = \|z_{n-1}\|^2 - \left| \left\langle z_{n-1}, \frac{\varphi_n}{\|\varphi_n\|} \right\rangle \right|^2 = \|z_{n-1}\|^2 \left(1 - \left| \left\langle \frac{z_{n-1}}{\|z_{n-1}\|}, \frac{\varphi_n}{\|\varphi_n\|} \right\rangle \right|^2 \right). \tag{3.3}$$

A repeated application of (3.3) gives that for all $0 \leq l \leq n-1$,

$$\|z_n\|^2 = \|z_l\|^2 \prod_{k=l+1}^{n} \left(1 - \left| \left\langle \frac{z_{k-1}}{\|z_{k-1}\|}, \frac{\varphi_k}{\|\varphi_k\|} \right\rangle \right|^2 \right). \tag{3.4}$$

When $l = 0$ this yields the formula (3.2). ☐

From Proposition 3.1, we see that the monotonicity of the Kaczmarz algorithm in (1.2) is an immediate corollary of (3.1). Consequently, if $z_l = 0$ for some $l \geq 1$ then $z_j = 0$ for all $j \geq l$. So, if $z_l = 0$, the convention that $z_k/\|z_k\| = 0$ for $k \geq l$ simply sets each term in the partial product in (3.4) to be one. While it is possible to have the desirable outcome of finite convergence to zero error $\|z_l\| = 0$, this will generally not be the case for continuous random measurements. For example, if the normalized measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are absolutely continuous with respect to the normalized surface measure on $\mathbb{S}^{d-1}$ then by (3.2) each error z_k is almost surely nonzero.

Corollary 3.2. *Suppose the measurement vectors $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ are random vectors such that each φ_n is almost surely nonzero. Additionally suppose that $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are independent and that each $\varphi_n/\|\varphi_n\|$ is absolutely continuous with respect to the uniform measure on $\mathbb{S}^{d-1}$. If the initial error $z_0 = x - x_0$ in the Kaczmarz algorithm is nonzero then for each $k \geq 1$, there holds* $\Pr[\|x - x_k\| = 0] = \Pr[z_k = 0] = 0$.

4. Moment bounds in the Kaczmarz algorithm

The following moment bound and its proof is motivated by the work in [21] on mean squared error.

Theorem 4.1. *Let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and such that $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent. Let $s>0$ be fixed and assume that each $\varphi_n/\|\varphi_n\|$ has the common Kaczmarz bound $\alpha>0$ of order s, as in (2.1).*

The error after the nth iteration of the Kaczmarz algorithm satisfies

$$\mathbb{E}\|x-x_n\|^{2s} \leq \alpha^{ns}\|x-x_0\|^{2s}. \tag{4.1}$$

If additionally the Kaczmarz bound α is tight then

$$\mathbb{E}\|x-x_n\|^{2s} = \alpha^{ns}\|x-x_0\|^{2s}. \tag{4.2}$$

Proof. Let $z_n = x - x_n$ and $u_n = \varphi_n/\|\varphi_n\|$. Note that $z_{n-1} = z_{n-1}(z_0, u_1, \cdots, u_{n-1})$ is a function of the deterministic initial error $z_0 \in \mathbb{R}^d$ and the independent random vectors $\{u_k\}_{k=1}^{n-1} \subset \mathbb{S}^{d-1}$. In particular, u_n and $z_{n-1}/\|z_{n-1}\|$ are independent. Let the measure μ_k denote the probability distribution of u_k.

Since u_n satisfies the Kaczmarz bound α of order s, it follows that if $\{\widetilde{u}_k\}_{k=1}^{n-1} \subset \mathbb{S}^{d-1}$ are fixed unit-vectors then

$$\int_{\mathbb{S}^{d-1}} \left(1 - \left|\left\langle \frac{z_{n-1}(z_0,\widetilde{u}_1,\cdots,\widetilde{u}_{n-1})}{\|z_{n-1}(z_0,\widetilde{u}_1,\cdots,\widetilde{u}_{n-1})\|}, u_n\right\rangle\right|^2\right)^s d\mu_n(u_n) \leq \alpha^s. \tag{4.3}$$

Since z_{n-1} and $u_n = \varphi_n/\|\varphi_n\|$ are independent, (4.3) implies that

$$\mathbb{E}\left[\left(1 - \left|\left\langle \frac{z_{n-1}}{\|z_{n-1}\|}, \frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^2\right)^s \,\middle|\, z_{n-1}\right] \leq \alpha^s. \tag{4.4}$$

The error representation (3.1) together with (4.4) and properties of conditional expectations and the fact that the random vectors $\{u_k\}_{k=1}^n$ are independent implies

$$\begin{aligned}
\mathbb{E}\|z_n\|^{2s} &= \mathbb{E}\left(\|z_{n-1}\|^{2s}\left(1 - \left|\left\langle \frac{z_{n-1}}{\|z_{n-1}\|}, \frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^2\right)^s\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[\|z_{n-1}\|^{2s}\left(1 - \left|\left\langle \frac{z_{n-1}}{\|z_{n-1}\|}, \frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^2\right)^s \,\middle|\, z_{n-1}\right]\right) \\
&= \mathbb{E}\left(\|z_{n-1}\|^{2s}\ \mathbb{E}\left[\left(1 - \left|\left\langle \frac{z_{n-1}}{\|z_{n-1}\|}, \frac{\varphi_n}{\|\varphi_n\|}\right\rangle\right|^2\right)^s \,\middle|\, z_{n-1}\right]\right) \\
&\leq \mathbb{E}\left(\|z_{n-1}\|^{2s}\alpha^s\right) \\
&= \alpha^s\ \mathbb{E}\|z_{n-1}\|^{2s}.
\end{aligned} \tag{4.5}$$

Iterating (4.5) yields (4.1). A similar computation shows that if each $\varphi_n/\|\varphi_n\|$ has a tight Kaczmarz bound α then (4.2) holds. ∎

Taking $s = 1$ in Theorem 4.1 gives the following mean squared error bound for the Kaczmarz algorithm. Corollary 4.2 is essentially the same as the mean squared error bounds in [21] but is expressed under a superficially more general model of randomization using probabilistic frames instead of the finite random vectors as in Example 2.7. Theorem 4.1 and Corollary 4.2 should both be viewed as reflections of the work in [21].

Corollary 4.2. *Let $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and such that $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are independent. If each $\varphi_n/\|\varphi_n\|$ has the common probabilistic lower frame bound $\beta > 0$ then the error after the nth iteration of the Kaczmarz algorithm satisfies*

$$\mathbb{E}\|x - x_n\|^2 \leq (1-\beta)^n \|x - x_0\|^2. \tag{4.6}$$

If additionally each $\varphi_n/\|\varphi_n\|$ is isotropic (2.3) *then*

$$\mathbb{E}\|x - x_n\|^2 = (1-d^{-1})^n \|x - x_0\|^2. \tag{4.7}$$

Similar to [21], Corollary 4.2 yields the following examples. Versions of these examples appear in [21] under a slightly different statement of randomization, so we include them here to illustrate analogs for randomization using probabilistic frames and for random measurements satisfying Definition 2.11.

Example 4.3. If $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ satisfy the properties of Definition 2.11 then each $\varphi_n/\|\varphi_n\|$ is isotropic with tight probabilistic frame bound $\beta = 1/d$. Thus the mean squared error of the Kaczmarz algorithm for measurements with the properties of Definition 2.11 satisfies

$$\mathbb{E}\|x - x_n\|^2 = (1-d^{-1})^n \|x - x_0\|^2.$$

Example 4.4 (Computational Complexity)**.** Let $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ be random vectors satisfying the properties of Definition 2.11. Given $\epsilon > 0$, let n_ϵ be the smallest number of iterations of the Kaczmarz algorithm needed to ensure ϵ-precise mean squared error

$$\mathbb{E}\|x - x_{n_\epsilon}\|^2 \leq \epsilon^2 \|x - x_0\|^2.$$

By (4.2), we seek the smallest integer n_ϵ such that $(1-\beta)^{n_\epsilon} \leq \epsilon^2$. Since $\beta = \beta_d = 1/d$, in high dimensions we have $\log(1-\beta) \approx -\beta = -1/d$ and

$$n_\epsilon = \left\lceil \frac{2\log\epsilon}{\log(1-d^{-1})} \right\rceil \approx 2d\,|\log\epsilon|. \tag{4.8}$$

By (4.8), $\mathcal{O}(d)$ iterations suffice to ensure ϵ-precise mean squared error. Moreover, since each iteration of the Kaczmarz algorithm requires $\mathcal{O}(d)$ elementary operations, ϵ-precision is achieved with an overall quadratic complexity of $\mathcal{O}(d^2)$ operations.

Example 4.5. Theorem 4.1 together with Example 2.7 recovers the mean squared error bound (1.3) from [21]. In particular, if the randomization from Example 2.7 is used to solve a given $N \times d$ system $\Phi x = y$ then the Kaczmarz bound α of order $s = 1$ satisfies $\alpha \leq 1 - [\kappa(\Phi)]^{-2}$ so that $\mathbb{E}\|x - x_n\|^2 \leq \alpha^n \|x - x_0\|^2 \leq (1 - [\kappa(\Phi)]^{-2})^n \|x - x_0\|^2$.

5. Almost sure convergence for uniform random measurements on $\mathbb{S}^{d-1}$

The product error representation in (3.2) will play an important role in our analysis of almost sure convergence in the Kaczmarz algorithm. It will be convenient to introduce the following notation for the individual random variables in the random product (3.2):

$$\xi_k = \left(1 - \left|\left\langle \frac{z_{k-1}}{\|z_{k-1}\|}, \frac{\varphi_k}{\|\varphi_k\|}\right\rangle\right|^2\right). \tag{5.1}$$

Since the first step of the Kaczmarz algorithm requires an initial estimate $x_0 \in \mathbb{R}^d$, each random variable ξ_k is implicitly parametrized by the initial error $z_0 = x - x_0 \in \mathbb{R}^d$. When needed, we emphasize this dependence by writing $\xi_k = \xi_k(z_0)$.

With the notation (5.1), the error in the Kaczmarz algorithm satisfies

$$\|x - x_n\|^2 = \|x - x_0\|^2 \left(\prod_{k=1}^{n} \xi_k\right). \tag{5.2}$$

In general, the random variables $\{\xi_k\}_{k=1}^\infty$ defined by (5.1) need not be independent, e.g., see Example 6.1. However, in the special case when the random measurements $\{\varphi_n\}_{n=1}^\infty$ satisfy the conditions of Definition 2.11, it will follow that the random variables $\{\xi_n\}_{n=1}^\infty$ are independent and identically distributed. This will have pleasant consequences for the subsequent error analysis.

The results of this section are stated under the general requirement that the normalized vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty \subset \mathbb{S}^{d-1}$ are independent uniformly distributed random vectors on $\mathbb{S}^{d-1}$, but it is worth noting that this includes the special case when the $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ are independent Gaussian random vectors with i.i.d. $N(0,1)$ entries, see Example 2.12. In particular, the results of this section immediately apply to give almost sure convergence rates for the Kaczmarz algorithm in the case of Gaussian measurement vectors.

Lemma 5.1. *Fix $z_0 \in \mathbb{R}^d$. Let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and such that the normalized random measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent and uniformly distributed on $\mathbb{S}^{d-1}$. Then the random variables $\{\xi_n\}_{n=1}^\infty$ defined by* (5.1) *are independent and identically distributed versions of the random variable*

$$\xi = 1 - |\langle e_1, u\rangle|^2, \tag{5.3}$$

where $e_1 = (1, 0, \cdots, 0) \in \mathbb{R}^d$ and $u \in \mathbb{R}^d$ is a uniform random vector on $\mathbb{S}^{d-1}$. The random variable ξ does not depend on z_0 but does depend on the dimension d.

Proof. Let $u_n = \varphi_n/\|\varphi_n\|$. The hypotheses on $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ mean that $\{u_n\}_{n=1}^\infty$ are independent random variables that are uniformly distributed on $\mathbb{S}^{d-1}$. Without loss of generality we assume that $z_0 \neq 0$. Moreover, as noted in the discussion following Proposition 3.1, since each u_n is absolutely continuous, we have that $\Pr[z_k = 0] = 0$ for all k.

Note that the random vector

$$z_{n-1} = z_{n-1}(z_0, u_1, \cdots, u_{n-1})$$

is a function of the nonrandom initial error z_0 and the independent random vectors $\{u_k\}_{k=1}^{n-1}$. Thus, z_{n-1} and u_n are independent random vectors. This independence along with the rotational symmetry of u_n now implies that if $e_1 = (1, 0, \cdots, 0)$ then ξ_n has the same distribution as the random variable $\left(1 - |\langle e_1, u_n\rangle|^2\right)$. This shows that the random variables $\{\xi_n\}_{n=1}^\infty$ are identically distributed.

It remains to show that the random variables $\{\xi_n\}_{n=1}^\infty$ are independent. Note that the random variable

$$\xi_n = \xi_n(z_0, u_1, \cdots, u_n)$$

is a function of the nonrandom initial error z_0 and the independent random vectors $\{u_k\}_{k=1}^n$. Given constants $\{\beta_n\}_{n=1}^\infty \subset \mathbb{R}$, for each $n \geq 1$ let

$$A_n = \{(u_1, \cdots, u_n) \in (\mathbb{S}^{d-1})^n : \xi_n(z_0, u_1, \cdots, u_n) \leq \beta_n\}$$

denote the event that $\xi_n \leq \beta_n$, and let

$$B_n = \bigcap_{k=1}^n A_k$$

denote the event that $\xi_k \leq \beta_k$ holds for all $1 \leq k \leq n$. Let χ_{A_n} and χ_{B_n} denote the indicator functions of the events A_n and B_n respectively. Let μ denote the normalized surface measure on $\mathbb{S}^{d-1}$.

If $\{\widetilde{u}_k\}_{k=1}^{n-1} \subset \mathbb{S}^{d-1}$ are fixed unit-vectors then the rotational symmetry of u_n implies

$$\begin{aligned}
\int_{\mathbb{S}^{d-1}} \chi_{A_n}(\widetilde{u}_1, \cdots, \widetilde{u}_{n-1}, u_n)\, d\mu(u_n) &= \Pr[\xi_n(z_0, \widetilde{u}_1, \cdots, \widetilde{u}_{n-1}, u_n) \leq \beta_n] \\
&= \Pr\left[1 - \left|\left\langle \frac{z_{n-1}(z_0, \widetilde{u}_1, \cdots, \widetilde{u}_{n-1})}{\|z_{n-1}(z_0, \widetilde{u}_1, \cdots, \widetilde{u}_{n-1})\|}, u_n \right\rangle\right|^2 \leq \beta_n\right] \\
&= \Pr[1 - |\langle e_1, u_n\rangle|^2 \leq \beta_n] \\
&= \Pr[\xi_n \leq \beta_n]. \qquad (5.4)
\end{aligned}$$

The independence of $\{u_n\}_{n=1}^\infty$ together with (5.4) allows one to compute as follows:

$$\begin{aligned}
&\Pr[\xi_1 \leq \beta_1, \cdots, \xi_n \leq \beta_n] = \Pr[B_n] = \mathbb{E}[\chi_{B_n}] = \mathbb{E}[\chi_{B_{n-1}}\ \chi_{A_n}] \\
&\quad = \int_{(\mathbb{S}^{d-1})^{n-1}} \chi_{B_{n-1}}(u_1, \cdots, u_{n-1}) \left(\int_{\mathbb{S}^{d-1}} \chi_{A_n}(u_1, \cdots, u_{n-1}, u_n) d\mu(u_n) \right) d\mu(u_1) \cdots d\mu(u_{n-1}) \\
&\quad = \Pr[\xi_n \leq \beta_n] \int_{(\mathbb{S}^{d-1})^{n-1}} \chi_{B_{n-1}}(u_1, \cdots, u_{n-1})\, d\mu(u_1) \cdots d\mu(u_{n-1}) \\
&\quad = \Pr[\xi_n \leq \beta_n]\ \Pr[B_{n-1}] \\
&\quad = \Pr[\xi_n \leq \beta_n]\ \Pr[\xi_1 \leq \beta_1, \cdots, \xi_{n-1} \leq \beta_{n-1}].
\end{aligned}$$

Iterating this argument yields that for any $n \geq 1$

$$\Pr[\xi_1 \leq \beta_1, \cdots, \xi_n \leq \beta_n] = \prod_{k=1}^n \Pr[\xi_k \leq \beta_k].$$

This implies that the random variables $\{\xi_n\}_{n=1}^\infty$ are independent.

□

Lemma 5.2. *Let $d \geq 2$ be an integer and let ξ be the random variable given by* (5.3). *Then*

$$\mathbb{E}(\log \xi) = \frac{\omega_{d-2}}{\omega_{d-1}} \int_0^\pi \sin^{d-2}\theta \, \log(\sin^2\theta) \, d\theta, \tag{5.5}$$

and

$$\mathbb{E}(\log \xi)^2 = \frac{\omega_{d-2}}{\omega_{d-1}} \int_0^\pi \sin^{d-2}\theta \, (\log(\sin^2\theta))^2 d\theta, \tag{5.6}$$

where $\omega_d = \dfrac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})}$ *is the surface area of* $\mathbb{S}^d$.

Proof. Let $e_1 = (1, 0, \cdots, 0) \in \mathbb{R}^d$ and let μ be the normalized surface measure on $\mathbb{S}^{d-1}$. By Lemma 5.1

$$\begin{aligned}
\mathbb{E}(\log \xi) &= \int_{\mathbb{S}^{d-1}} \log\left(1 - |\langle e_1, u\rangle|^2\right) d\mu(u) \\
&= \frac{\omega_{d-2}}{\omega_{d-1}} \int_{-1}^1 (\sqrt{1-s^2})^{d-3} \log(1-s^2) ds \\
&= \frac{\omega_{d-2}}{\omega_{d-1}} \int_0^\pi \sin^{d-2}\theta \, \log(\sin^2\theta) \, d\theta.
\end{aligned}$$

Similarly,

$$\mathbb{E}(\log \xi)^2 = \int_{\mathbb{S}^{d-1}} (\log(1 - |\langle e_1, u\rangle|^2))^2 d\mu(u) = \frac{\omega_{d-2}}{\omega_{d-1}} \int_0^\pi \sin^{d-2}\theta (\log(\sin^2\theta))^2 d\theta.$$

□

The independence of the random variables in $\{\xi_n\}_{n=1}^\infty$ in Lemma 5.1 will allow us to apply classical tools such as the Strong Law of Large Numbers, the Central Limit Theorem, and the Law of the Iterated Logarithm to our analysis of almost sure convergence properties of the Kaczmarz algorithm.

Theorem 5.3. *Let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and such that the normalized random measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent and uniformly distributed on $\mathbb{S}^{d-1}$. Let $R = \exp(-\mathbb{E}\log\xi)$ and $\sigma^2 = \mathbb{E}(\log\xi)^2 - (\mathbb{E}\log\xi)^2$ be as computed in Lemma 5.2. Then the error in the Kaczmarz algorithm satisfies*

$$\lim_{n\to\infty} \|x - x_n\|^{2/n} = R^{-1}, \qquad \textit{almost surely,} \tag{5.7}$$

and

$$\forall\ t \in \mathbb{R}, \quad \lim_{n\to\infty} \Pr\left[R^n \|x - x_n\|^2 \geq \|x - x_0\|^2 e^{t\sqrt{n\sigma^2}}\right] = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du, \tag{5.8}$$

and

$$\limsup_{n\to\infty} \left(R^n \|x - x_n\|^2\right)^{\frac{1}{\sqrt{2\sigma^2 n \log(\log n)}}} = e, \qquad \textit{almost surely.} \tag{5.9}$$

Note that the constants $R = R_d > 1$ and $\sigma = \sigma_d > 0$ only depend on the dimension d.

Proof. Let

$$S_n = \log\left(\prod_{k=1}^n \xi_k\right) = \sum_{k=1}^n \log(\xi_k). \tag{5.10}$$

By Lemma 5.1 the $\{\xi_k\}_{k=1}^\infty$ are independent versions of the random variable ξ given by (5.3). By Lemma 5.2, $\mathbb{E}(\log\xi) = \log(1/R)$ and $\mathrm{Var}(\log\xi) = \sigma^2$ are both finite.

Applying the Strong Law of Large Numbers to (5.10) yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \log\xi_k = \mathbb{E}(\log\xi) = \log(1/R), \quad \text{a.s.} \tag{5.11}$$

Taking the exponential of (5.11) gives

$$\lim_{n\to\infty}\left(\prod_{k=1}^n \xi_n\right)^{\frac{1}{n}} = \exp(\mathbb{E}(\log\xi)) = R^{-1}, \quad \text{a.s.} \tag{5.12}$$

Equation (5.7) now follows from (5.2) and (5.12).

Applying the Central Limit Theorem to (5.10) gives

$$\forall\, t\in\mathbb{R}, \qquad \lim_{n\to\infty} \Pr\left(\frac{\sum_{k=1}^n \log\xi_k - n\log(1/R)}{\sqrt{n\sigma^2}} \le t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-u^2/2}du. \tag{5.13}$$

Exponentiating and reorganizing (5.13) gives

$$\forall\, t\in\mathbb{R}, \qquad \lim_{n\to\infty} \Pr\left(\prod_{k=1}^n R\xi_k \ge e^{t\sqrt{n\sigma^2}}\right) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-u^2/2}du. \tag{5.14}$$

Equation (5.8) now follows from (5.2) and (5.14).

To prove (5.9), apply the Law of the Iterated Logarithm to $\log(R\xi_n)$. Since $\mathbb{E}(\log(R\xi)) = \mathbb{E}(\log\xi + \log R) = 0$ and $\mathrm{Var}(\log(R\xi)) = \mathbb{E}(\log\xi + \log R)^2 = \mathbb{E}(\log\xi - \mathbb{E}(\log\xi))^2 = \sigma^2$, there holds

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^n \log(R\xi_k)}{\sqrt{2\sigma^2 n\log(\log n)}} = 1, \quad \text{a.s.}$$

which yields

$$\limsup_{n\to\infty}\left(\prod_{k=1}^n R\xi_k\right)^{\frac{1}{\sqrt{2\sigma^2 n\log(\log n)}}} = e, \quad \text{a.s.}$$

This implies (5.9). □

For a different perspective on Theorem 5.3 we shall use following lemma.

Lemma 5.4. *Given $A > 0$ and a nonnegative sequence $\{a_n\}_{n=1}^\infty \subset \mathbb{R}$, the following two statements are equivalent:*

(*a*) $\lim_{n\to\infty} (a_n)^{1/n} = 1/A$

(*b*) $\forall\; 0 < r < A,\ \lim_{n\to\infty} r^n a_n = 0$ *and* $\forall\; A < r,\ \lim_{n\to\infty} r^n a_n = \infty.$

Proof. To prove that (a) implies (b), suppose $0 < r < A$. There exists c such that $1/A < c < 1/r$, and hence $(a_n)^{1/n} < c$ for all sufficiently large n. Therefore

$$\limsup_{n\to\infty} r^n a_n \leq \limsup_{n\to\infty} r^n c^n = 0.$$

A similar argument applies to the case when $r > A$.

To prove that (b) implies (a), proceed by contrapositive and suppose (a) does not hold. Then there exists $\epsilon > 0$ and a strictly increasing sequence $\{n_k\}_{k=1}^\infty$ of positive integers such that $|a_{n_k}^{1/n_k} - 1/A| > \epsilon$. Without loss of generality, consider the case when infinitely many $a_{n_k}^{1/n_k}$ satisfy $a_{n_k}^{1/n_k} - 1/A > \epsilon$. Picking $r = (1/A + \epsilon/2)^{-1} < A$ gives that

$$\limsup_{k\to\infty} r^{n_k} a_{n_k} \geq \limsup_{k\to\infty} r^{n_k}(1/A+\epsilon)^{n_k} = \infty,$$

which means that (b) does not hold. □

Thus, (5.7) in Theorem 5.3 can be stated as follows.

Corollary 5.5. *Let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and such that the normalized random measurement vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent and uniformly distributed on $\mathbb{S}^{d-1}$. Let $R > 1$ be the constant defined in Theorem 5.3.*

If $0 < r < R$ then

$$\lim_{n\to\infty} r^n \|x - x_n\|^2 = 0, \qquad \textit{almost surely.} \tag{5.15}$$

If $r > R$ then

$$\lim_{n\to\infty} r^n \|x - x_n\|^2 = \infty, \qquad \textit{almost surely.} \tag{5.16}$$

The boundary case $r = R$ in Corollary 5.5 is addressed by (5.8) and (5.9). For example, taking $t = 0$ in (5.8) of Theorem 5.3 shows that one does not have almost sure convergence of $R^n\|x - x_n\|^2$ to 0. Likewise, one does not have almost sure convergence of $R^n\|x - x_n\|^2$ to infinity either.

Example 5.6. To compare the almost sure convergence rates in Theorem 5.3 with the mean squared convergence rates in Corollary 4.2, let $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^2$ be independent random vectors that are uniformly distributed on $\mathbb{S}^1$. In dimension $d = 2$, we have that each φ_n is isotropic with $\beta = \beta_2 = 1/2$. Moreover, $\omega_1 = 2\pi$ and $\omega_0 = (2\sqrt{\pi})/\Gamma(1/2) = 2$, so that the constant R from Theorem 5.3 satisfies

$$R = \exp\left(-\frac{1}{\pi}\int_0^\pi \log\sin^2\theta d\theta\right) = 4. \tag{5.17}$$

The computation of the integral in (5.17) follows from the fact that the Lobachevsky function

$$L(t) = -\int_0^t \log|2\sin\theta| d\theta = -t\log 2 - \frac{1}{2}\int_0^t \log\sin^2\theta d\theta$$

is π-periodic, e.g., see the appendix in [10]. So, $L(\pi) = L(0) = 0$ and this implies (5.17).

By (4.2), the mean squared error satisfies

$$\forall\, n \geq 1, \qquad \mathbb{E}\|x - x_n\|^2 = (1/2)^n\|x - x_0\|^2. \tag{5.18}$$

By Corollary 5.5, we have the following almost sure convergence:

$$\forall\ 0 < r < 4, \qquad \lim_{n\to\infty} r^n \|x - x_n\|^2 = 0, \quad \text{almost surely.} \tag{5.19}$$

In particular, the mean squared error decreases at the rate $(1/2)^n$, whereas the squared error nearly decreases at the rate of $(1/4)^n$ in an almost sure sense.

Example 5.7 (Gaussian measurements)**.** If the measurement vectors $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^d$ are independent and have i.i.d. $N(0,1)$ entries then the results of Theorem 5.3 and Corollary 5.5 apply to give almost sure convergence rates for Gaussian measurement vectors. In particular, in dimension $d = 2$, the mean squared convergence result (5.18) and almost sure convergence result (5.19) both still hold for Gaussian vectors $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{R}^2$ and allow one to compare the mean squared convergence rate with the almost sure convergence rate.

6. Almost sure convergence for general random measurements

The results of Section 5 showed that if the measurement vectors $\{\varphi_n\}_{n=1}^\infty$ satisfy the conditions of Definition 2.11 then the random variables $\{\xi_n\}_{n=1}^\infty$ defined in (5.1) are independent and identically distributed, and moreover do not depend on the initial error z_0. This, in turn, made it possible to apply classical results on sums of i.i.d. random variables to the convergence analysis in Theorem 5.3.

For general measurement vectors $\{\varphi_n\}_{n=1}^\infty$ without the properties in Definition 2.11, it is possible for the random variables $\{\xi_n\}_{n=1}^\infty$ to be neither independent nor identically distributed (see Example 6.1 below), and it is not possible to directly apply the classical convergence results used for Theorem 5.3. In this section we address almost sure convergence of the Kaczmarz algorithm when a general collection of random measurements $\{\varphi_n\}_{n=1}^\infty$ is used.

Example 6.1. Let $\varphi \in \mathbb{R}^2$ be a discrete random vector that satisfies

$$\Pr[\varphi = (1,0)] = 2/3 \qquad \text{and} \qquad \Pr[\varphi = (0,1)] = 1/3.$$

Let φ_1, φ_2 be independent versions of φ. We consider the random variables $\xi_1(z_0), \xi_2(z_0)$ that arise in the first two iterations of the Kaczmarz algorithm when $x = (\sqrt{3}/2, 1/2)$, $x_0 = (0,0)$, and the initial error $z_0 = x - x_0$ satisfies $z_0 = (\sqrt{3}/2, 1/2)$.

A direct computation shows that ξ_1 satisfies

$$\Pr[\xi_1 = 1/4] = 2/3 \qquad \text{and} \qquad \Pr[\xi_1 = 3/4] = 1/3.$$

Similarly, by considering a tree of probabilities, ξ_2 can be shown to satisfy

$$\Pr[\xi_2 = 1] = 5/9, \qquad \text{and} \qquad \Pr[\xi_2 = 0] = 4/9.$$

Moreover, it can be shown that $\Pr[\xi_1 = 3/4, \xi_2 = 1] = 1/9$. Thus ξ_1, ξ_2 are neither independent nor identically distributed.

Theorem 6.2. *Let $\{\varphi_k\}_{k=1}^\infty \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero and for which $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent. Let $s > 0$ be fixed and suppose that each $\varphi_n/\|\varphi_n\|$ has* *the common Kaczmarz bound* $0 < \alpha < 1$ *of order* s*. Then there exists a random variable* X *satisfying* $\mathbb{E}|X|^s < \infty$ *such that*

$$\lim_{n\to\infty} (1/\alpha)^n \|x - x_n\|^2 = X, \qquad \textit{almost surely.} \tag{6.1}$$

Consequently,

$$\forall\ 0 < r < 1/\alpha, \qquad \lim_{n\to\infty} r^n \|x - x_n\|^2 = 0, \qquad \textit{almost surely.} \tag{6.2}$$

Proof. Let $Y_n = (1/\alpha)^{sn}\|x-x_n\|^{2s} = (1/\alpha)^{sn}\|z_n\|^{2s}$ and let $\mathcal{F}_n$ be the sigma algebra generated by the random vectors $\varphi_1/\|\varphi_1\|, \cdots, \varphi_n/\|\varphi_n\|$. It can be shown that Y_n is measureable with respect to $\mathcal{F}_n$. Similar computations as in the proof of Theorem 4.1 show that

$$\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[(1/\alpha)^{sn}\|z_n\|^{2s} \mid \mathcal{F}_{n-1}] \leq (1/\alpha)^{s(n-1)}\|z_{n-1}\|^{2s}(1/\alpha)^s\alpha^s = Y_{n-1}.$$

Thus $\{(Y_n, \mathcal{F}_n)\}_{n=1}^\infty$ is a supermartingale. Moreover, by Theorem 4.1, there holds

$$\forall n \geq 1, \quad \mathbb{E}[Y_n] \leq \|z_0\|^{2s}.$$

An application of Doob's martingale convergence theorem (for example, see Theorem 1 on page 508 of [18]) to the submartingale $\{(-Y_n, \mathcal{F}_n)\}_{n=1}^\infty$ shows that the limit

$$\lim_{n\to\infty} Y_n = Y, \qquad \text{exists almost surely,}$$

and the limit satisfies $\mathbb{E}|Y| < \infty$. Thus,

$$\lim_{n\to\infty} (1/\alpha)^{sn}\|x - x_n\|^{2s} = Y, \qquad \text{almost surely.} \tag{6.3}$$

Letting $X = Y^{1/s}$, and taking the $1/s$ power of (6.3), we obtain (6.1)

$$\lim_{n\to\infty} (1/\alpha)^n\|x - x_n\|^2 = X, \qquad \text{almost surely.}$$

This implies (6.2) and completes the proof. □

The martingale convergence theorem is a natural tool for analyzing the Kaczmarz algorithm and was previously used in [9] to discuss almost sure convergence of such algorithms (but with convergence rates only studied for mean squared error). The almost sure convergence rates established in Theorems 5.3 and 6.2 complement the mean squared error rates in [9, 21]. Martingale and Markov chain methods were also previously applied to the error analysis of closely related algorithms such as the Rangan-Goyal algorithm in [14] and the Gibbs sampler in [6]. In the present setting, it is possible to give a direct alternative proof of the bound (6.2) in Theorem 6.2 without appealing to martingale convergence in the following manner.

Alternative Proof of Equation (6.2). Let $\{\xi_k\}_{k=1}^\infty$ be as in (5.1). Fix $0 < r < 1/\alpha$ and let

$$P_n = r^n \prod_{k=1}^n \xi_k.$$

Recall that $P_n \geq 0$ and that $r^n\|x - x_n\|^2 = P_n\|x - x_0\|^2$. To prove (6.2), it suffices to show that

$$\forall \epsilon > 0, \quad \lim_{N\to\infty} \Pr\left(\bigcup_{n=N}^{\infty} \{P_n > \epsilon\}\right) = 0. \tag{6.4}$$

Let $\epsilon > 0$ be fixed. A union bound together with Chebyshev's inequality implies that

$$\Pr\left(\bigcup_{n=N}^{\infty} \{P_n > \epsilon\}\right) \leq \sum_{n=N}^{\infty} \Pr(P_n > \epsilon) \leq \sum_{n=N}^{\infty} \frac{\mathbb{E}(P_n^s)}{\epsilon^s}. \tag{6.5}$$

Theorem 4.1 shows that

$$\mathbb{E}(P_n^s) \leq (r\alpha)^{sn}. \tag{6.6}$$

Combining (6.5) and (6.6), it follows that

$$\Pr\left(\bigcup_{n=N}^{\infty} \{P_n > \epsilon\}\right) \leq \frac{1}{\epsilon^s} \sum_{n=N}^{\infty} (r\alpha)^{sn} \leq \frac{(r\alpha)^{sN}}{\epsilon^s(1 - r^s\alpha^s)}. \tag{6.7}$$

Since $0 < r\alpha < 1$, it follows that (6.4) holds. This completes the proof. □

The next result improves the conclusion of Theorem 6.2 by considering the limiting case when $s = 0$. Unlike Theorem 6.2, the following theorem assumes that the $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are identically distributed.

Theorem 6.3. *Let $\{\varphi_n\}_{n=1}^{\infty} \subset \mathbb{R}^d$ be random vectors that are almost surely nonzero. Assume that the normalized vectors $\{\varphi_n/\|\varphi_n\|\}_{n=1}^{\infty}$ are independent and identically distributed versions of a random vector $u \in \mathbb{S}^{d-1}$ and assume that u has the logarithmic Kaczmarz bound $0 < \rho < 1$. Then the error in the Kaczmarz algorithm satisfies*

$$\forall\ 0 < r < 1/\rho, \quad \lim_{n\to\infty} r^n\|x - x_n\|^2 = 0, \quad \textit{almost surely.}$$

Proof. Fix $0 < r < 1/\rho$ and take α such that $\rho < \alpha < 1/r$. By Corollary 2.9,

$$\forall x \in \mathbb{S}^{d-1}, \qquad \inf_{s>0} (\mathbb{E}(1 - |\langle x, u\rangle|^2)^s)^{1/s} \leq \rho.$$

So, for every $x \in \mathbb{S}^{d-1}$ there exists $s_x > 0$ such that

$$(\mathbb{E}(1 - |\langle x, u\rangle|^2)^{s_x})^{1/s_x} < \alpha.$$

It follows from the Lebesgue Dominated Convergence Theorem that

$$\forall x \in \mathbb{S}^{d-1}, \quad \lim_{\|y\|=1; y\to x} (\mathbb{E}(1 - |\langle y, u\rangle|^2)^{s_x})^{1/s_x} = (\mathbb{E}(1 - |\langle x, u\rangle|^2)^{s_x})^{1/s_x} < \alpha.$$

So, for every $x \in \mathbb{S}^{d-1}$, there exists an open neighborhood $U_x \subset \mathbb{S}^{d-1}$ of x such that

$$\forall y \in U_x, \quad (\mathbb{E}(1 - |\langle y, u\rangle|^2)^{s_x})^{1/s_x} < \alpha.$$

Since $\mathbb{S}^{d-1}$ is compact and $\mathbb{S}^{d-1} \subset \cup_{x\in\mathbb{S}^{d-1}} U_x$, there exists a finite subcover $\{U_{x_j}\}_{j=1}^J$ of $\{U_x\}_{x\in\mathbb{S}^{d-1}}$. Letting $s^* = \min\{s_{x_j}\}_{j=1}^J$ and using Lyapunov's inequality (for example, see page 193 of [18]), we obtain

$$\forall x \in \mathbb{S}^{d-1}, \quad (\mathbb{E}(1 - |\langle x, u\rangle|^2)^{s^*})^{1/s^*} < \alpha.$$

Since the $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent and identically distributed versions of the random vector u, each $\varphi_n/\|\varphi_n\|$ has the common Kaczmarz bound α of order $s^* > 0$. Since $r < 1/\alpha$ we conclude by Theorem 6.2 that $\lim_{n\to\infty} r^n \|x - x_n\|^2 = 0$ almost surely. □

Theorem 6.3 provides stronger error bounds than Theorem 6.2 since by Lemma 2.8 and Corollary 2.9, a logarithmic Kaczmarz bound ρ satisfies $\rho \leq \alpha = \alpha_s$ for each Kaczmarz bound α of order $s > 0$. In the special case when the $\{\varphi_n/\|\varphi_n\|\}_{n=1}^\infty$ are independent uniform random vectors on $\mathbb{S}^{d-1}$, Theorem 6.3 recovers the sharp bound (5.15) of Corollary 5.5. In particular, if $u = \varphi/\|\varphi\|$ is uniformly distributed on $\mathbb{S}^{d-1}$ then the logarithmic Kaczmarz bound ρ is tight and satisfies

$$\forall x \in \mathbb{S}^{d-1}, \quad \rho = \exp[\mathbb{E}\log(1 - |\langle x, u\rangle|^2)] = \exp(\mathbb{E}(\log \xi)) = 1/R,$$

where R and $\mathbb{E}(\log \xi)$ are as in Lemma 5.2 and Theorem 5.3.

Acknowledgments

A. Powell was supported in part by NSF DMS Grant 0811086, and gratefully acknowledges the Academia Sinica Institute of Mathematics (Taipei, Taiwan) for its hospitality and support. He especially thanks Professors Yunshyong Chou, Chii-Ruey Hwang, and Shuenn-Jyi Sheu for valuable discussions related to the material.

The authors thank the anonymous reviewers for their helpful comments and especially for pointing out the reference [9].

References

1. Editorial comment on the two following letters to the editor, concerning Kaczmarz algorithms. Journal of Fourier Analysis and Applications, 15 (2009), no. 4, 429-430.
2. Y. Censor, G. Herman, and M. Jiang. A note on the behavior of the randomized Kaczmarz algorithm of Strohmer and Vershynin. Journal of Fourier Analysis and Applications, 15 (2009), no. 4, 431-436.
3. M. Ehler and K. Okoudjou. Probabilistic frames: an overview. To appear in: Finite Frames: Theory and Applications, edited by P. Casazza and G. Kutyniok, Birkhäuser.
4. Y. Eldar and D. Needell. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma. Numerical Algorithms, 58 (2011), no. 2, 163–177.
5. R. Haller and R. Szwarc. Kaczmarz algorithm in Hilbert space. Studia Mathematica, 169 (2005), no.2, 123–132.
6. C.-R. Hwang and S.-J. Sheu. On the geometrical convergence of the Gibbs sampler in $\mathbf{R}^{d*}$. Journal of Multivariate Analysis, 66 (1998), 22–37.
7. S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. Bulletin International de l'Académie Polonaise de Sciences A, (1937), 355–357. English translation in: Approximate solution of systems of linear equations. International Journal of Control 57 (1993), 1269–1271.
8. S. Kwapień and J. Mycielski. On the Kaczmarz algorithm of approximation in infinite-dimensional spaces. Studia Mathematica, 148 (2001), no.1, 75–86.
9. D. Leventhal and A.S. Lewis. Randomized methods for linear constraints: convergence rates and conditioning. Mathematics of Operations Research, 35 (2010), no. 3, 641–654.
10. J. Milnor. Hyperbolic geometry: the first 150 years. Bulletin of the American Mathematical Society, 6 (1982), no. 1, 9–24.
11. F. Natterer. The Mathematics of Computerized Tomography. Wiley, New York, 1986.
12. D. Needell. Randomized Kaczmarz solver for noisy linear systems. BIT Numerical Mathematics, 50 (2010), no. 2, 395–403.

13. A.M. Powell. Mean squared error bounds for the Rangan-Goyal soft thresholding algorithm. Applied and Computational Harmonic Analysis, 29 (2010), no. 3, 251–271.
14. S. Rangan and V.K. Goyal. Recursive consistent estimation with bounded noise. IEEE Transactions on Information Theory, 47, no.1, (2001) 457–464.
15. W. Rudin. Real and Complex Analysis, 3rd Edition. McGraw-Hill, New York, 1987.
16. E. Saff and V. Totik. Logarithmic potentials with external fields. First edition. Springer, New York, 1997.
17. K.M. Sezan and H. Stark. Applications of convex projection theory to image recovery in tomography and related areas. In H. Stark, editor, Image Recovery: Theory and application, 415462. Academic Press, 1987.
18. A.N. Shiryaev. Probability, Second Edition. Springer, New York, 1989.
19. S. Sra and J. Tropp. Row-action methods for compressed sensing. Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 3, 868-871, Toulouse, May 2006.
20. T. Strohmer and R. Vershynin. A randomized solver for linear systems with exponential convergence. Approximation, randomization and combinatorial optimization, 499507, Lecture Notes in Computer Science, 4110, Springer, Berlin, 2006.
21. T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. Journal of Fourier Analysis and Applications, 15 (2009), no. 2, 262–278.
22. T. Strohmer and R. Vershynin. Comments on the randomized Kaczmarz method. Journal of Fourier Analysis and Applications, 15 (2009), no. 4, 437–440.
23. R. Szwarc. Kaczmarz algorithm in Hilbert space and tight frames. Applied and Computational Harmonic Analysis, 22 (2007), no. 3, 382–385.

Vanderbilt University, Department of Mathematics, Nashville, TN 37240, USA
E-mail address: xuemei.chen@vanderbilt.edu

Vanderbilt University, Department of Mathematics, Nashville, TN 37240, USA
E-mail address: alexander.m.powell@vanderbilt.edu