Abstract—The Kaczmarz algorithm (KA) is a popular method for solving a system of linear equations. In this note we derive a new exponential convergence result for the KA. The key allowing us to establish the new result is to rewrite the KA in such a way that its solution path can be interpreted as the output from a particular dynamical system. The asymptotic stability results of the corresponding dynamical system can then be leveraged to prove exponential convergence of the KA. The new bound is also compared to existing bounds.

Index Terms—Kaczmarz algorithm, Stability analysis, Cyclic algorithm.

I. PROBLEM STATEMENT

In this note, we discuss the convergence properties of the Kaczmarz algorithm (KA) [1]. Since its introduction, the KA has been applied in many different fields and many new developments are reported. For applications and recent advances with respect to the KA, see for example [2]-[13].

Let us first briefly introduce the KA. It is used to find the vectors of the matrix

\[ A \]

converge. This is especially true when several consecutive row elements of the matrix

\[ A \]

are zero. In order to overcome this drawback, the Randomized Kaczmarz Algorithm (RKA) algorithm was introduced in [4].

The key of the RKA is that, instead of performing the hyperplane projections cyclically in a deterministic order, the projections are performed in a random order. More specifically, at time \( k \), select a hyperplane \( H_k \) to project with probability \( \| a_i \|^2 / \| A \|^2_p \), for \( p = 1, \ldots, m \). Note that \( \| \cdot \|_F \) is used to denote the Frobenius norm of a matrix. Intuitively speaking, the involved randomization is performing a kind of "preconditioning" to the original matrix equations [6]. The result is that a faster exponential convergence rate can be achieved, which is established in [4].

The specific and predefined ordering of the projections in the KA makes it challenging to obtain a theoretical convergence rate of the method. In [16], the authors build up the convergence rate of the KA by exploiting the Meany inequality [17], which works for the case \( \lambda = 1 \) in [2]. In [18], the authors also established a convergence rate for the KA, for a wider range of \( \lambda \in (0, 2) \). In Section III we will compare these results in more detail.

In this note, we present a different way to characterize the convergence property of the KA in [2] valid for all \( \lambda \) in the interval \( 0 < \lambda \leq 1 \). The key idea underlying our approach is to interpret the solution path of the KA as the output of a particular dynamical system. By studying the asymptotic stability of that dynamical system we obtain a new exponential convergence results for the KA.

The note will be organized as follows. In the subsequent section we make use of the sub-sequence \( \{x_{jm} - x_j\}_{j=0}^\infty \) to enable the derivation of the new exponential convergence result. In Section III we discuss its connections and differences to existing results. Finally, conclusions and some ideas for future work are provided in Section IV.

II. THE NEW CONVERGENCE RESULT

First, let us introduce the matrix \( B \in \mathbb{R}^{m \times n} \), for which the \( i \)-th row \( b_i^T \) is defined as \( b_i \triangleq \frac{a_i}{\| a_i \|^2} \), \( i = 1, \ldots, m \). Furthermore, let \( P_i \triangleq b_i b_i^T \) for \( i = 1, 2, \ldots, m \) and let \( \theta_k \triangleq x_k - x \) for \( k \geq 0 \). Using this new notation allows us to rewrite (2) according to

\[ \theta_{k+1} = (I - \lambda P_{i(k)}) \theta_k, \]

which can be interpreted as a discrete-time linear dynamical system. Hence, the mathematical equivalence of (2) and (3) allows us to study the KA as a linear dynamical system. Our
analysis and the proof techniques we employ are heavily inspired by the stability analysis of linear time varying systems, see e.g. [14], [15].

In what follows we will focus on analyzing the convergence rate of the sub-sequence \( \{\|v_{jm}\|^2\}_{j=0}^n \).

Given the fact that \( i(k) = \text{mod}(k, m) + 1 \), we have

\[
\theta_{(j+1)m} = \left( \prod_{i=1}^{m} (I - \lambda P_i) \right) \theta_{jm} \triangleq M_m \theta_{jm}.
\]

The following theorem gives the main result of the letter, which provides an upper bound on the spectral norm of \( M_m \).

**Theorem 1:** Let \( \rho \triangleq \|M_m\|_2 \), then it holds that

\[
\rho^2 \leq \rho_1 \triangleq 1 - \frac{\lambda}{(2 + \lambda^2 m^2)\|B^T\|_2^2}.
\]  

**Proof:** Let \( v_0 \in \mathbb{R}^n \) be a unit vector satisfying \( M_m v_0 = \rho v_0 \) and let \( v_i = (I - \lambda P_i) v_{i-1} \) for \( i = 1, \cdots, m \). It follows that \( v_m = M_m v_0 \) and \( \|v_m\|^2 = \rho^2 \).

Since \( 0 < \lambda \leq 1 \), it holds that \( 2\lambda - \lambda^2 \geq \lambda \). Also notice that since \( P_i^2 = P_i \), we have that

\[
(I - \lambda P_i)^2 = I - (2\lambda + \lambda^2)P_i \leq I - \lambda P_i,
\]

which holds for \( i = 1, \cdots, m \). Hence,

\[
\|v_i\|^2 = v_{i-1}^T (I - \lambda P_i)^2 v_{i-1} \\
\leq v_{i-1}^T (I - \lambda P_i) v_{i-1} \\
= \|v_{i-1}\|^2 - \lambda \|P_i v_{i-1}\|^2,
\]

which in turn implies that

\[
\lambda \sum_{i=1}^{m} \|P_i v_{i-1}\|^2 \leq \|v_0\|^2 - \|v_m\|^2 = 1 - \rho^2.
\]  

(5)

Also, for any \( i \in \{1, \cdots, m\} \), we have that

\[
\|v_i - v_0\| = \left\| \sum_{k=1}^{i} (v_k - v_{k-1}) \right\| = \lambda \left\| \sum_{k=1}^{i} P_k v_{k-1} \right\| \\
\leq \lambda \sum_{k=1}^{i} \|P_k v_{k-1}\| \leq \lambda \sqrt{i} \left( \sum_{k=1}^{i} \|P_k v_{k-1}\|^2 \right) \\\n\leq \sqrt{\lambda i} \left( \lambda \sum_{k=1}^{m} \|P_k v_{k-1}\|^2 \right)
\]

(6)

Together with (5), we get

\[
\|v_i - v_0\|^2 \leq \lambda i (1 - \rho^2).
\]

Meanwhile, we have that

\[
\lambda v_0^T B^T B v_0 = \lambda \sum_{k=1}^{m} v_k^T P_k v_0 = \lambda \sum_{k=1}^{m} \|P_k v_0\|^2 \\
= \lambda \sum_{k=1}^{m} \|P_k (v_{k-1} + (v_0 - v_{k-1}))\|^2 \\
\leq 2\lambda \sum_{k=1}^{m} \|P_k v_{k-1}\|^2 + 2\lambda \sum_{k=1}^{m} \|P_k (v_{k-1} - v_0)\|^2 \\
\leq 2\lambda \sum_{k=1}^{m} \|P_k v_{k-1}\|^2 + 2\lambda \sum_{k=1}^{m} \|v_{k-1} - v_0\|^2.
\]

Together with (5) and (6), we have that

\[
\lambda v_0^T B^T B v_0 \leq 2(1 - \rho^2) + 2\lambda \sum_{k=1}^{m} \|v_{k-1} - v_0\|^2.
\]

or equivalently

\[
\lambda v_0^T B^T B v_0 \leq (1 - \rho^2) \left( 2 + \lambda^2 m(m-1) \right),
\]  

(7)

thus

\[
\rho^2 \leq 1 - \frac{\lambda}{2 + \lambda^2 m(m-1)}.
\]

(8)

Finally, notice that \( m(m-1) \leq m^2 \) holds for any natural number \( m \), which concludes the proof.

**Remark 1:** An application of the Cauchy-Schwartz inequality shows that the right hand side of (4) can be optimized when \( \lambda = \sqrt{\frac{2}{m}} \).

**Remark 2:** Notice that the bound in (4) does not depend on the specific ordering for implementing the projections. The convergence rate of the KA should indeed depend on the ordering used in implementing the projections. In this sense, the bound in (4) gives a worst case bound, i.e. it bounds the slowest convergence speed among all the possible orderings for the projections.

The following corollary characterizes the convergence of the KA under \( \lambda = 1 \), which will be used in the subsequent section to enable comparison to the result given in [16] and the convergence rate of the RKA given in [4].

**Corollary 1:** For the KA with \( \lambda = 1 \) in (2), if \( m \geq n \geq 2 \), we have that

\[
\rho^2 \leq 1 - \frac{1}{m^2}\|B\|_2^2.
\]

(9)

**Proof:** We note that for \( m \geq 2 \) we have

\[
2 + m(m-1) \leq m^2,
\]

which when inserted into (8) proves (9).

Next, we will derive an improvement over the bound (4), enabled by partitioning the matrix \( A \) into non-overlapping sub-matrices. Let \( q = \lceil \frac{m}{2} \rceil + 1 \), where \( \lceil x \rceil \) denotes the smallest
number which is greater or equal to \(x\). Define the following
sets as \(T_i = \{(i-1)n+1, \cdots, in\}\), for \(i = 1, \cdots, q-1\) and \(T_q = \{(q-1)n+1, \cdots, m\}\). Further, for \(i = 1, \cdots, q\), define \(B_i\) as the sub-matrix of \(B\) with the rows indexed by the set \(T_i\), and \(N_i = \prod_{j \in T_i} (I - \lambda P_j)\).

**Corollary 2:** Based on the previous definitions, and further assume that all the sub-matrices \(B_i\) for \(i = 1, \cdots, q\) are of rank \(n\), then we have that

\[
\rho^2 \leq \rho_2 \triangleq \prod_{i=1}^{q} \left(1 - \frac{\lambda}{(2 + \lambda^2 n(n-1)) \|B_i\|_F^2}\right) \tag{10}
\]

**Proof:** Notice that since

\[M_m = N_q N_{q-1} \cdots N_2 N_1,
\]

we have that

\[
\rho^2 = \|M_m\|_F^2 \leq \prod_{i=1}^{q} \|N_i\|_F^2. \tag{11}
\]

For each \(N_i\), the spectral norm can be bounded analogously to what was done in Theorem 1, resulting in

\[
\|N_i\|_F^2 \leq 1 - \frac{\lambda}{(2 + \lambda^2 n(n-1)) \|B_i\|_F^2}.
\]

Finally, inserting this inequality into (11) concludes the proof.

**III. DISCUSSION AND NUMERICAL ILLUSTRATION**

In Section III-A and III-B we compare our new bound with the bounds provided by the Meany inequality [16, 17] and the RKA, respectively. In Section III-B we also provide a numerical illustration. Finally, Section III-C is devoted to a comparison with the bound provided in [18].

**A. Comparison with the bound given by Meany inequality**

In order enable a comparison to the results given by the Meany inequality [16, 17] we have to assume that \(m = n\) and \(\lambda = 1\).

Denote the singular values of \(B\) as \(\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n\), then the bound given by the Meany inequality can be written as

\[
\rho^2 \leq 1 - \prod_{i=1}^{n} \sigma_i^2, \quad \text{and the bound given in (9) can be written as} \quad \rho^2 \leq 1 - \frac{1}{n^2}.
\]

So, when

\[
\frac{\sigma_n^2}{n^2} \geq \prod_{i=1}^{n} \sigma_i^2 \quad \text{i.e.} \quad \prod_{i=1}^{n} \sigma_i^2 \leq \frac{1}{n^2}, \tag{12}
\]

holds, the bound in (9) is tighter. In the following lemma, we derive a sufficient condition, under which the inequality (12) holds.

**Lemma 1:** If \(\sigma_n^2 \leq \frac{(n-2)^{n-2}}{n^n}\) holds, the inequality in (12) is satisfied.

**Proof:** Notice that

\[
\prod_{i=1}^{n-1} \sigma_i^2 = \left(\prod_{i=1}^{n-2} \sigma_i^2\right) \sigma_n^2 \leq \left(\frac{\sum_{i=1}^{n-2} \sigma_i^2}{n-2}\right) \sigma_n^2 \leq \left(\frac{n}{n-2}\right) \sigma_n^2 = \sigma_n^2. \tag{13}
\]

The inequality (13) holds since

\[
\sum_{i=1}^{n-2} \sigma_i^2 \leq \sum_{i=1}^{n} \sigma_i^2 = \|B\|_F^2 = n.
\]

Hence,

\[
\frac{n}{n-2} \sigma_n^2 \leq \frac{(n-2)^{n-2}}{n^n} \tag{14}
\]

holds, or equivalently if

\[
\sigma_n^2 \leq \frac{(n-2)^{n-2}}{n^n} \tag{15}
\]

holds, then (12) holds, which concludes the proof.

**Remark 3:** Notice that the right hand side of the inequality in Lemma 1 is in the order of \(\frac{1}{n^2}\) for large \(n\). Furthermore, we note that the bound provided by Theorem 1 depends explicitly on the number of rows, while the bound provided by the Meany inequality does not.

**B. Comparison with the bound given by the RKA**

Let us now compare our new results to the results available for the RKA. Given by [9], for the sequence \(\{\theta_{jm}\}_{j=0}^{\infty}\) generated by the RKA, it holds that

\[
\mathbb{E}\|\theta_{jm}\|^2 \leq \left(1 - \frac{1}{\|A\|_F^2 \|A\|_F^2}\right)^{jm} \|\theta_0\|^2, \tag{14}
\]

for \(j \geq 1\), where \(\mathbb{E}\) denotes the expectation operator with respect to the random operations up to index \(jm\).

In order to compare (14) and (9), we make the assumption that \(A\) is a matrix with each row normalized, i.e. \(A = B\). It follows that \(\|B\|_F^2 = m \leq m^2\), and

\[
1 - \frac{1}{m^2} \|B\|_F^2 \geq 1 - \frac{1}{\|B\|_F^2 \|B\|_F^2}. \tag{15}
\]

Furthermore, since \(\|B\|_F^2 \|B\|_F^2 \geq 1\), we have that

\[
1 - \frac{1}{\|B\|_F^2 \|B\|_F^2} \geq \left(1 - \frac{1}{\|B\|_F^2 \|B\|_F^2}\right)^m, \tag{16}
\]

and combining (15) and (16), results in

\[
1 - \frac{1}{m^2} \|B\|_F^2 \geq 1 - \frac{1}{\|B\|_F^2 \|B\|_F^2}^m.
\]

The above inequality implies that the bound given by (9) is more conservative than the one given by the RKA, i.e. (14).

This is a reasonable result considering that the bound (9) is a worst case bound, as we remarked in the previous section.

Next, a numerical illustration is implemented to compare the bounds given by (9), (10) and (14). The setup is as follows. Generate \(A = \text{randn}(30, 3)\) and normalize each row to obtain \(B\), generate \(x = \text{randn}(3, 1)\) and compute \(y = Bx\). In the implementation of the RKA, we run 1000 realizations with the same initial value \(x_0\) to obtain the average performance. The results are reported in Fig. 1.

From the left panel in Fig. 1, we can see that the bound (14) for characterising the convergence of the RKA is closer to the real performance of the RKA, while the bounds given by (9) and (10) for bounding the convergence of the KA are further
away from the real performance of the KA. We remark again that the result is reasonable in the sense that the bounds (9) and (10) are bounding the worst case performance of the KA algorithm.

The right panel in Fig. 1 shows a zoomed illustration of the bound given by (9) and (10). We can observe that the bound given by (10) indeed improves upon (9). Note that the improvement is enabled by the partitioning of the rows of the matrix.

Note that neither of the bounds (18) and (17) can dominate the other for all $\lambda \in (0, 1]$, hence a better characterization of the convergence rate for a given $\lambda$ is achieved by combining both, which results in

$$\rho^2 \leq 1 - \max \left( \frac{(2 - \lambda) \lambda}{m \| B \|_2^2}, \left( \frac{1}{\| B \|_2^2} \right) \right).$$

(19)

C. Comparison with the bound given in [17]

We also assume that each row of the matrix $A$ is normalized, i.e. $A = B$. The result of the current work reads as

$$\rho^2 \leq 1 - \frac{\lambda}{(2 + \lambda^2 m^2) \| B \|_2^2},$$

(17)

where $\lambda \in (0, 1]$. The result given in [18] can be written as

$$\rho^2 \leq 1 - \frac{(2 - \lambda) \lambda}{m \| B \|_2^2},$$

(18)

where $\lambda \in (0, 2)$. Note that $m$ denotes the number of rows of the matrix $B$.

We will compare these two bounds from the following perspectives

1) An obvious difference is that the bound in (18) works for $\lambda \in (0, 2)$ while the bound in (17) only works for $\lambda \in (0, 1]$.  
2) A condition for the bound in (18) to hold is that $A$ has to be a square and invertible matrix. While the analysis and the bound in (17) of this work do not need such an assumption.  
3) The right hand side (RHS) of (18) is minimized when $\lambda = 1$, resulting in $\rho^2 \leq 1 - \frac{\lambda^2}{m \| B \|_2^2}$; The RHS of (17) is minimized when $\lambda = \frac{\sqrt{2}}{m}$, resulting in $\rho^2 \leq 1 - \frac{\lambda^2}{4m \| B \|_2^2}$. The calculation implies that when $m$ increases, both bounds decrease at a speed of $\frac{1}{m}$ when the best $\lambda$ are chosen.

Fig. 1. In the left panel, the curves with tags 'KA' and 'RKA' illustrate the real performance of the KA and the RKA. The curves with tags 'KA BD1', 'KA BD2' and 'RKA BD1' illustrate the bounds given by (9), (10) and (14), respectively. In the right panel, a zoomed illustration for the curves 'KA BD1' and 'KA BD2' in the left panel is given.

This note studies the convergence properties of the general KA from a dynamical system point of view, which gives a new bound for the exponential convergence of the algorithm.

IV. Conclusions

REFERENCES

[1] S. Kaczmarz, Angenaherte Auflosung von Systemen linearer Gleichungen, Bulletin International de l’Académie Polonaise des Sciences et des Lettres, 35, 355-357, 1937.  
[2] P. P. B. Eggermont, G. T. Herman, A. Lent, Iterative algorithms for large partitioned linear systems, with applications to image reconstruction, Linear algebra and its applications, 40, 37-67, 1981.  
[3] Y. Censor, Row-action methods for huge and sparse systems and their applications, SIAM review, 23(4), 444-466, 1981.  
[4] T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, Journal of Fourier Analysis and Applications, 15(2), 262-278, 2009.  
[5] F. Natterer, The Mathematics of Computerized Tomography, Wiley, New York, 1986.  
[6] T. Strohmer, R. Vershynin, Comments on the randomized Kaczmarz method, Journal of Fourier Analysis and Applications, 15(4), 437-440, 2009.  
[7] D. Needell, Randomized Kaczmarz solver for noisy linear systems, BIT Numerical Mathematics, 50(2), 395-403, 2010.  
[8] Y. Eldar, D. Needell, Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma, Numerical Algorithms, 58(2), 163-177, 2011.  
[9] X. Chen, A. Powell, Almost sure convergence for the Kaczmarz algorithm with random measurements, Journal of Fourier Analysis and Applications, 18(6), 1195-1214, 2012.  
[10] A. Zouzias, N. M. Freris, Randomized extended Kaczmarz for solving least-squares, SIAM Journal on Matrix Analysis and Applications, 2012.  
[11] G. Thoppe, V. Borkar, D. Manjunath, A stochastic Kaczmarz algorithm for network tomography, Automatica, 2013.  
[12] D. Needell, J. A. Tropp, Paved with Good Intentions: Analysis of a Randomized Block Kaczmarz Method, Linear Algebra and its Applications, 441, 199-221, 2014.  
[13] L. Dai, M. Soltanalian, and K. Pelckmans, On the Randomized Kaczmarz Algorithm, IEEE Signal Processing Letters, 21(3), 330-333, 2014.  
[14] L. Guo, Time-varying stochastic systems: stability, estimation and control, Jilin Science and Technology Press, Changchun, 1993.  
[15] L. Guo, L. Ljung, Exponential stability of general tracking algorithms, IEEE Transactions on Automatic Control, 40(8), 1376-1387, 1995.  
[16] A. Galantai, On the rate of convergence of the alternating projection method in finite dimensional spaces, Journal of mathematical analysis and applications, 310(1), 30-44, 2005.  
[17] R. K. Meany, A matrix inequality, SIAM Journal on Numerical Analysis, 6(1), 104-107, 1969.  
[18] J. Mandel, Convergence of the cyclical relaxation method for linear inequalities, Mathematical programming, 30(2), 218-228, 1984.