

# A Note on the Behavior of the Randomized Kaczmarz Algorithm of Strohmer and Vershynin

Yair Censor<sup>1</sup>, Gabor T. Herman<sup>2</sup> and Ming Jiang<sup>3</sup>

<sup>1</sup>Department of Mathematics,  
University of Haifa, Mt. Carmel, Haifa 31905, Israel  
(yair@math.haifa.ac.il)

<sup>2</sup>Department of Computer Science,  
The Graduate Center, City University of New York,  
365 Fifth Avenue, New York, NY 10016, USA  
(gabortherman@yahoo.com)

<sup>3</sup>LMAM, School of Mathematical Sciences, Peking University,  
5 Yi He Yuan Street, Beijing 100871, P.R. China  
(ming-jiang@pku.edu.cn)

August 10, 2008. Revised: January 8, 2009.

## Abstract

In a recent paper by T. Strohmer and R. Vershynin [“A Randomized Kaczmarz Algorithm with Exponential Convergence”, *Journal of Fourier Analysis and Applications*, published online on April 25, 2008] a “randomized Kaczmarz algorithm” is proposed for solving systems of linear equations  $\{\langle a^i, x \rangle = b_i\}_{i=1}^m$ . In that algorithm the next equation to be used in an iterative Kaczmarz process is selected with a

probability proportional to  $\|a^i\|^2$ . The paper illustrates the superiority of this selection method for the reconstruction of a bandlimited function from its nonuniformly spaced sampling values.

In this note we point out that the reported success of the algorithm of Strohmer and Vershynin in their numerical simulation depends on the specific choices that are made in translating the underlying problem, whose geometrical nature is “find a common point of a set of hyperplanes”, into a system of algebraic equations. If this translation is carefully done, as in the numerical simulation provided by Strohmer and Vershynin for the reconstruction of a bandlimited function from its nonuniformly spaced sampling values, then indeed good performance may result. However, there will always be legitimate algebraic representations of the underlying problem (so that the set of solutions of the system of algebraic equations is exactly the set of points in the intersection of the hyperplanes), for which the selection method of Strohmer and Vershynin will perform in an inferior manner.

**Key words.** Kaczmarz algorithm, projection method, rate of convergence.

**AMS 2000 subject classifications.** 15A06, 15A29, 65F10.

## 1 Introduction

Kaczmarz’s algorithm [21] is a sequential projection method for the solution of linear systems of equations of the form  $Ax = b$ . It was rediscovered in the field of image reconstruction from projections in [13] where it is called the Algebraic Reconstruction Technique (ART), see, e.g., [5, 7, 17], and it is obtained also by applying to the hyperplanes described by the linear system the method of successive projections onto convex sets. The latter is called in the literature POCS (for projections onto convex sets), see, e.g., [25], and was originally published by Bregman [3] and further studied by Gubin, Polyak and Raik [15], see, e.g., Bauschke and Borwein [1] and Combettes [9] for reviews. It belongs also to the class of *row-action methods* described in [6], see also [7]. The literature on Kaczmarz’s algorithm is vast and ranges from beautiful mathematical theoretical results to many successful applications in important real-world problems. Kaczmarz’s method is also a special case of several other algorithmic structures, see, e.g., [4, 8, 12, 20].

In a recent paper [24] (see also [23]) Strohmer and Vershynin propose a

randomized Kaczmarz’s algorithm and claim that “it outperforms all previously known methods on general extremely overdetermined systems.” In this note we point out that the reported success of the algorithm of [24] in their numerical simulation depends on the specific choices that are made in translating the underlying problem, whose geometrical nature is “find a common point of a set of hyperplanes”, into a system of algebraic equations. If this translation is carefully done, as in the numerical simulation provided by [24] for the reconstruction of a bandlimited function from its nonuniformly spaced sampling values, then indeed good performance may result. However, there will always be legitimate algebraic representations of the underlying problem (so that the set of solutions of the system of algebraic equations is exactly the set of points in the intersection of the hyperplanes), for which the selection method of [24] will perform in an inferior manner.

## 2 The randomized Kaczmarz’s algorithm

Each equation of the  $m \times n$  system  $Ax = b$  gives rise to a hyperplane

$$H_i = \{x \in R^n \mid \langle a^i, x \rangle = b_i\} \quad (1)$$

in the  $n$ -dimensional Euclidean space  $R^n$ , where  $a^i \in R^n$  and  $b_i \in R$  are the  $i$ -th row of the matrix  $A$  and the  $i$ -th component of the right-hand side vector  $b \in R^m$  of the linear system, respectively. The Euclidean inner product is  $\langle a^i, x \rangle := \sum_{j=1}^n a_j^i x_j$  where  $a^i = (a_j^i)_{j=1}^n$  and  $x = (x_j)_{j=1}^n$  and  $\|a^i\|$  denotes the Euclidean norm. The randomized Kaczmarz’s algorithm of Strohmer and Vershynin is as follows.

**Algorithm 1** [24, Algorithm 1]

*Initialization:*  $x^0 \in R^n$  is arbitrary.

*Iterative Step:* Given the current iterate  $x^k$ , calculate the next iterate  $x^{k+1}$  by

$$x^{k+1} = x^k + \frac{b_{r(i)} - \langle a^{r(i)}, x^k \rangle}{\|a^{r(i)}\|^2} a^{r(i)}, \quad (2)$$

where  $r(i)$  is chosen from the set  $\{1, 2, \dots, m\}$  at random with probability proportional to  $\|a^{r(i)}\|^2$ .

Obviously, the equations can be scaled, i.e., both sides of each equation  $\langle a^i, x \rangle = b_i$  can be divided through by an arbitrary nonzero real number, say,  $c_i$ , without changing the hyperplane  $H_i$  but changing only its algebraic presentation. The question is how do such permissible changes affect the behavior of Algorithm 1, if at all, and what does that mean for the superiority claim of this algorithm in [24].

### 3 The results of Strohmer and Vershynin

Strohmer and Vershynin use the *scaled condition number* of a matrix  $A$ , introduced by Demmel [10],

$$\kappa(A) := \|A\|_F \|A^{-1}\|_2, \quad (3)$$

where  $\|A\|_2$  and  $\|A\|_F$  are the spectral norm and the Frobenius norm of  $A$ , respectively. Their main result [24, Theorem 2] states that any iterative sequence  $\{x^k\}_{k=0}^\infty$ , generated by Algorithm 1, converges to the solution of the system  $Ax = b$  with an expected exponential rate, and that the rate of convergence depends on the scaled condition number  $\kappa(A)$ .

In their numerical experiments they compare what they call the standard Kaczmarz method (in which the rows of  $A$  are selected in their “natural” order 1, 2, ...) with a simple randomized Kaczmarz method (in which the rows of  $A$  are selected at random with equal probability and with the *randomized Kaczmarz* method of Algorithm 1 (where the rows of  $A$  are selected at random with probability proportional to the 2-norm of the rows)). They plot the least squares error versus the number of projections for each algorithm and claim that “Algorithm 1 significantly outperforms the other Kaczmarz methods, demonstrating not only the power of choosing the projections at random, but also the importance of choosing the projections according to their relevance.”

### 4 Theory and practice

Clearly, Kaczmarz’s method is a *geometric algorithm* (this term was used by Gordon and Mansour [14]) in the sense that the sequence of iterates generated by it depends only on the hyperplanes defined by the equations and not on any particular algebraic representation of the hyperplanes. One

can always use real nonzero numbers, say  $c_i$ ,  $i = 1, 2, \dots, m$ , and divide through each equation  $\langle a^i, x \rangle = b_i$ , without affecting the hyperplanes defined by the equations and the solution set of the system. Therefore, the rate of convergence of a geometric algorithm must depend on properties of the underlying geometric objects (the hyperplanes) and not on their algebraic representation. This geometric approach has been mathematically studied in several works, see, e.g., Hamaker and Solmon [16], Kayalar and Weinert [22], Bauschke, Borwein and Lewis [2] and Deutsch and Hundal [11] which consider, in various settings, the connection between the angles among the underlying sets and the rate of convergence of the alternating projections method. When specialized to hyperplanes, that algorithm coincides with Kaczmarz's method and the results apply to the linear system of equations  $Ax = b$ .

A scaling of the equations will change the system matrix  $A$  and its scaled condition number  $\kappa(A)$  and, in the light of [24], it might be tempting to think that it is possible to control in such a way the convergence rate of Kaczmarz's method. However, the geometric nature of Kaczmarz's method precludes such a possibility. Given any probability distribution whatsoever to be used for selecting the next hyperplane in Algorithm 1, we can always use real nonzero numbers to divide through the equations so that the rule for selecting the rows in Algorithm 1 will be exactly according to that probability distribution.<sup>1</sup> Thus, the rule proposed in Algorithm 1 cannot be in general optimal (or even in any sense superior).

Thus, the results shown in Figure 1 of [24], cannot be fully explained by the presented theory there. It is true that the convergence behavior depends on the order-selection of hyperplanes (as is well-known), but this dependence is determined by the geometry of the hyperplanes (the angles among them) and not by their algebraic representation, see, e.g., [2, 11, 16, 18, 19, 22].

Stated differently, if, for any reason whatsoever, we believe that at iteration  $k$  hyperplane  $H_i$  should be picked with higher probability than hyperplane  $H_j$ , then this is the case however the equations representing  $H_i$  and  $H_j$  are scaled (since the step of projecting onto the selected hyperplane is independent of the scaling). Hence, a claim that states that an order-selection methodology that depends on the norms of the  $a^i$ 's in the representation of

---

<sup>1</sup>For any distribution  $\lambda_i > 0$ ,  $i = 1, 2, \dots, m$ , the equations may be scaled as follows. Let  $\tilde{a}^i = \lambda_i \frac{a^i}{\|a^i\|_2}$ . Then  $\|\tilde{a}^i\|_2 = \lambda_i$  for the scaled equation. This scaling method also works for any other norm in the Euclidean space.

the hyperplanes is generally superior to another order-selection methodology must be inherently false.

To avoid inferior behavior by the randomized Kaczmarz algorithm, it is essential that the system of algebraic equations that represent the set of hyperplanes be carefully chosen. Indeed, this was done by Strohmer and Vershynin [24] in their numerical simulation in Section 4.1, see their equation (18). Had they selected a different algebraic representation, they would have obtained different convergence behavior. In particular, if they scaled their equations so that  $\|a^i\| = 1$  after scaling, then the reported difference between the behaviors of simple randomized Kaczmarz and of Algorithm 1 would have disappeared. And, if by the luck of the draw, the algebraic representation happened to be such that the norm associated with one equation is very much larger than the norms associated with the other equations, then the progress made by Algorithm 1 towards a solution would be poor due to the fact that the random selection would keep selecting the same equation most of the time.

**Acknowledgments.** This work was supported by grant No. 2003275 of the United States-Israel Binational Science Foundation (BSF), by a Chinese NKBRSF grant No. 2003CB716101 and by National Science Foundation of China (NSFC) grants No. 60325101, 60532080, and 60628102 and by Award Number R01HL070472 from the National Heart, Lung, And Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

## References

- [1] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review*, **38** (1996), 367–426.
- [2] H. H. Bauschke, J. M. Borwein and A. S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, *Contemporary Mathematics*, **204** (1997), 1–38.
- [3] L. M. Bregman, The method of successive projections for finding a common point of convex sets, *Soviet Mathematics Doklady*, **6** (1965), 688–692.

- [4] D. Butnariu, R. Davidi, G. T. Herman and I. G. Kazantsev, Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, *IEEE Journal of Selected Topics in Signal Processing*, **1** (2007), 540–547.
- [5] C. L. Byrne, *Applied Iterative Methods*, A. K. Peters Ltd., Wellesley, MA, USA, 2008.
- [6] Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Review*, **23** (1981), 444–466.
- [7] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.
- [8] Y. Censor, T. Elfving, G. T. Herman and T. Nikazad, On diagonally-relaxed orthogonal projection methods, *SIAM Journal on Scientific Computing*, **30** (2008), 473–504.
- [9] P. L. Combettes, The convex feasibility problem in image recovery, *Advances in Imaging and Electron Physics*, **95** (1996), 155–270.
- [10] J. W. Demmel, The probability that a numerical analysis problem is difficult, *Mathematics of Computation*, **50** (1988), 449–480.
- [11] F. Deutsch and H. Hundal, The rate of convergence for the cyclic projections algorithm I: Angles between convex sets, *Journal of Approximation Theory*, **142** (2006), 36–55.
- [12] P. P. B. Eggermont, G. T. Herman and A. Lent, Iterative algorithms for large partitioned linear systems, with applications to image reconstruction, *Linear Algebra and Its Applications*, **40** (1981), 37–67.
- [13] R. Gordon, R. Bender, and G. T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography, *Journal of Theoretical Biology*, **29** (1970), 471–481.
- [14] D. Gordon and R. Mansour, A geometric approach to quadratic optimization: an improved method for solving strongly underdetermined systems in CT, *Inverse Problems in Science & Engineering*, **15** (2007), 811–826.

- [15] L. Gubin, B. Polyak and E. Raik, The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, **7** (1967), 1–24.
- [16] C. Hamaker and D. Solmon, The angles between the null space of X-rays, *Journal of Mathematical Analysis and Applications*, **62** (1978), 1–23.
- [17] G. T. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.
- [18] G. T. Herman, Algebraic reconstruction techniques in medical imaging, in: C.T. Leondes (Editor), *Medical Imaging Systems Techniques and Applications: Computational Techniques*, Gordon and Breach, Overseas Publishers Association (OPA), Amsterdam, The Netherlands, 1998, pp. 1–42.
- [19] G. T. Herman and L. B. Meyer, Algebraic reconstruction techniques can be made computationally efficient, *IEEE Transactions on Medical Imaging*, **12** (1993), 600–609.
- [20] M. Jiang and G. Wang, Convergence studies on iterative algorithms for image reconstruction, *IEEE Transactions on Medical Imaging*, **22** (2003), 569–579.
- [21] S. Kaczmarz, Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin de l'Académie Polonaise des Sciences et Lettres*, **A35** (1937), 355–357.
- [22] S. Kayalar and H. L. Weinert, Error bounds for the method of alternating projections, *Mathematics of Control, Signals, and Systems*, **1** (1988), 43–59.
- [23] T. Strohmer and R. Vershynin, A randomized solver for linear systems with exponential convergence, in: J. Diaz, K. Jansen, J. D. P. Rolim and U. Zwick (Editors), *Approximation, Randomization and Combinatorial Optimization: Algorithms and Techniques*, Lecture Notes in Computer Science (LNCS), Vol. **4110**, Springer-Verlag, Berlin, Heidelberg, Germany, 2006, pp. 499–507.



- [24] T. Strohmer and R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, *Journal of Fourier Analysis and Applications*, to appear. The OnlineFirst version is available at:  
<http://www.springerlink.com/content/8028365666841332/>
- [25] D. C. Youla, Mathematical theory of image restoration by the method of convex projections, in: H. Stark (Editor), *Image Recovery: Theory and Applications*, Academic Press, Orlando, Florida, USA, 1987, pp. 29–77.