A randomized solver for linear systems with exponential convergence

Thomas Strohmer and Roman Vershynin*
Department of Mathematics, University of California
Davis, CA 95616-8633, USA.
strohmer@math.ucdavis.edu, vershynin@math.ucdavis.edu

Abstract

The Kaczmarz method for solving linear systems of equations $Ax = b$ is an iterative algorithm that has found many applications ranging from computer tomography to digital signal processing. Despite the popularity of this method, useful theoretical estimates for its rate of convergence are still scarce. We introduce a randomized version of the Kaczmarz method for overdetermined linear systems and we prove that it converges with expected exponential rate. Furthermore, this is the first solver whose *rate does not depend on the number of equations* in the system. The solver does not even need to know the whole system, but only its small random part. It thus outperforms all previously known methods on extremely overdetermined systems. Even for moderately overdetermined systems, numerical simulations reveal that our algorithm can converge faster than the celebrated conjugate gradient algorithm.

1 Introduction and state of the art

We study a linear system of equations

$$Ax = b, \tag{1}$$

where A is a full rank $m \times n$ matrix with $m \geq n$, and $b \in \mathbb{C}^m$. One of the most popular solvers for such overdetermined systems is *Kaczmarz's method* [12], which is a form of alternating projection method. This method is also known under the name *Algebraic Reconstruction Technique* (ART) in computer tomography [9, 13], and in fact, it was implemented in the very first medical scanner [11]. It can also

*T.S. was supported by the NSF DMS grant 0511461. R.V. was supported by Alfred P. Sloan Foundation and by the NSF DMS grant 0401032.

be considered as a special case of the POCS (Projection onto Convex Sets) method, which is a prominent tool in signal and image processing [15, 1].

We denote the rows of A by $a_1^*, \ldots, a_m^*$, where $a_1, \ldots, a_m \in \mathbb{C}^n$, and let $b = (b_1, \ldots, b_m)^T$. The classical scheme of Kaczmarz's method sweeps through the rows of A in a cyclic manner, projecting in each substep the last iterate orthogonally onto the solution hyperplane of $\langle a_i, x\rangle = b_i$ and taking this as the next iterate. Given some initial approximation x_0, the algorithm takes the form

$$x_{k+1} = x_k + \frac{b_i - \langle a_i, x_k\rangle}{\|a_i\|^2} a_i, \tag{2}$$

where $i = k \mod m+1$. Note that we refer to one projection as one iteration, thus one sweep in (2) through all m rows of A consists of m iterations. We will refer to this as *one cycle*.

While conditions for convergence of this method are readily established, useful theoretical estimates of the *rate of convergence* of the Kaczmarz method (or more generally of the alternating projection method for linear subspaces) are difficult to obtain, at least for $m > 2$. Known estimates for the rate of convergence are based on quantities of the matrix A that are hard to compute and difficult to compare with convergence estimates of other iterative methods (see e.g. [2, 3, 6] and the references therein). What numerical analysts would like to have is estimates of the convergence rate with respect to standard quantities such as $\|A\|$ and $\|A^{-1}\|$. The difficulty that no such estimates are known so far stems from the fact that the rate of convergence of (2) depends strongly on the *ordering* of the equations in (1), while quantities such as $\|A\|, \|A^{-1}\|$ are independent of the ordering of the rows of A.

It has been observed several times in the literature that using the rows of A in Kaczmarz's method in random order, rather than in their given order, can greatly improve the rate of convergence, see e.g. [13, 1, 10]. While this randomized Kaczmarz method is thus quite appealing for applications, no guarantees of its rate of convergence have been known.

In this paper, we propose the first randomized Kaczmarz method with *exponential expected rate of convergence*, cf. Section 2. Furthermore, this *rate does not depend on the number of equations* in the system. The solver does not even need to know the whole system, but only its small random part. Thus our solver outperforms all previously known methods on extremely overdetermined systems. We analyze the optimality of the proposed algorithm as well as of the derived estimate, cf. Section 3. Our numerical simulations demonstrate that even for moderately overdetermined systems, this random Kaczmarz method can outperform the celebrated conjugate gradient algorithm, see Section 4.

Notation: For a matrix A, $\|A\| := \|A\|_2$ denotes the spectral norm of A, $\|A\|_F$ is the Frobenious norm, i.e. the square root of the trace of A^*A, where the superscript *

stands for the conjugate transpose of a vector or matrix. The left inverse of A (which we always assume to exist) is written as A^{-1}. Thus $\|A^{-1}\|$ is the smallest constant M such that the inequality $\|Ax\| \geq \frac{1}{M}\|x\|$ holds for all vectors x. As usual, $\kappa(A) := \|A\|\|A^{-1}\|$ is the condition number of A. The linear subspace spanned by a vector x is written as $\mathrm{lin}(x)$. Finally, $\mathbb{E}$ denotes expectation.

2 Randomized Kaczmarz algorithm and its rate of convergence

It has been observed [13, 1, 10] that the convergence rate of the Kaczmarz method can be significantly improved when the algorithm (2) sweeps through the rows of A in a random manner, rather than sequentially in the given order. Here we propose a specific version of this randomized Kaczmarz method, which chooses each row of A with probability proportional to its relevance – more precisely, proportional to the square of its Euclidean norm. This method of sampling from a matrix was proposed in [5] in the context of computing a low-rank approximation of A, see also [14] for subsequent work and references. Our algorithm thus takes the following form:

Algorithm 1 (Random Kaczmarz algorithm)**.** *Let $Ax = b$ be a linear system of equations as in* (1) *and let x_0 be arbitrary initial approximation to the solution of* (1)*. For $k = 0, 1, \ldots$ compute*

$$x_{k+1} = x_k + \frac{b_{r(i)} - \langle a_{r(i)}, x_k \rangle}{\|a_{r(i)}\|^2} a_{r(i)}, \tag{3}$$

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

Our main result states that x_k converges exponentially fast to the solution of (1), and the rate of convergence depends *only* on the norms of the matrix and its inverse.

Theorem 2. *Let x be the solution of* (1)*. Then Algorithm 1 converges to x in expectation, with the average error*

$$\mathbb{E}\|x_k - x\|^2 \leq \Big(1 - \frac{1}{R}\Big)^k \cdot \|x_0 - x\|^2, \tag{4}$$

where $R = \|A^{-1}\|^2 \, \|A\|_F^2$.

Proof. There holds

$$\sum_{j=1}^{m} |\langle z, a_j \rangle|^2 \geq \frac{\|z\|^2}{\|A^{-1}\|^2} \qquad \text{for all } z \in \mathbb{C}^n. \tag{5}$$

Using the fact that $\|A\|_F^2 = \sum_{j=1}^m \|a_j\|^2$ we can write (5) as

$$\sum_{j=1}^m \frac{\|a_j\|^2}{\|A\|_F^2} \left|\left\langle z, \frac{a_j}{\|a_j\|}\right\rangle\right|^2 \geq \frac{1}{R}\|z\|^2 \quad \text{for all } z \in \mathbb{C}^n. \tag{6}$$

The main point in the proof is to view the left hand side in (6) as an expectation of some random variable. Namely, recall that the solution space of the j-th equation of (1) is the hyperplane $\{y : \langle y, a_j\rangle = b\}$, whose normal is $\frac{a_j}{\|a_j\|}$. Define a random vector Z whose values are the normals to all the equations of (1), with probabilities as in our algorithm:

$$Z = \frac{a_j}{\|a_j\|} \quad \text{with probability} \quad \frac{\|a_j\|^2}{\|A\|_F^2}, \quad j = 1, \dots, m. \tag{7}$$

Then (6) says that

$$\mathbb{E}|\langle z, Z\rangle|^2 \geq \frac{1}{R}\|z\|^2 \quad \text{for all } z \in \mathbb{C}^n. \tag{8}$$

The orthogonal projection P onto the solution space of a random equation of (1) is given by $Pz = z - \langle z - x, Z\rangle\, Z$.

Now we are ready to analyze our algorithm. We want to show that the error $\|x_k - x\|^2$ reduces at each step in average (conditioned on the previous steps) by at least the factor of $(1 - \frac{1}{R})$. The next approximation x_k is computed from x_{k-1} as $x_k = P_k x_{k-1}$, where $P_1, P_2, \dots$ are independent realizations of the random projection P. The vector $x_{k-1} - x_k$ is in the kernel of P_k. It is orthogonal to the solution space of the equation onto which P_k projects, which contains the vector $x_k - x$ (recall that x is the solution to all equations). The orthogonality of these two vectors then yields

$$\|x_k - x\|^2 = \|x_{k-1} - x\|^2 - \|x_{k-1} - x_k\|^2.$$

To complete the proof, we have to bound $\|x_{k-1} - x_k\|^2$ from below. By the definition of x_k, we have

$$\|x_{k-1} - x_k\| = \langle x_{k-1} - x, Z_k\rangle$$

where $Z_1, Z_2, \dots$ are independent realizations of the random vector Z. Thus

$$\|x_k - x\|^2 \leq \left(1 - \left|\left\langle \frac{x_{k-1} - x}{\|x_{k-1} - x\|}, Z_k\right\rangle\right|^2\right) \|x_{k-1} - x\|^2.$$

Now we take the expectation of both sides conditional upon the choice of the random vectors $Z_1, \dots, Z_{k-1}$ (hence we fix the choice of the random projections $P_1, \dots, P_{k-1}$ and thus the random vectors $x_1, \dots, x_{k-1}$). Then

$$\mathbb{E}|_{\{Z_1,\dots,Z_{k-1}\}}\|x_k - x\|^2 \leq \left(1 - \mathbb{E}_{\{Z_1,\dots,Z_{k-1}\}}\left|\left\langle \frac{x_{k-1} - x}{\|x_{k-1} - x\|}, Z_k\right\rangle\right|^2\right) \|x_{k-1} - x\|^2.$$

By (8) and the independence,

$$\mathbb{E}|_{\{Z_1,\dots,Z_{k-1}\}}\|x_k - x\|^2 \le \Big(1 - \frac{1}{R}\Big)\,\|x_{k-1} - x\|^2.$$

Taking the full expectation of both sides, by induction we complete the proof. ∎

Remark (Dimension-free perspective, robustness): The rate of convergence in Theorem 2 does not depend on the number of equations nor the number of variables, and obviously also not on the order of the projections. It is only controlled by the intrinsic and stable quantity R of the matrix A. This continues the dimension free approach to operators on finite dimensional normed spaces, see [14].

2.1 Quadratic time

Let n denote the number of variables in (1). Clearly, $n \le R \le \kappa(A)^2 n$, where $\kappa(A)$ is the condition number of A. Then as $k \to \infty$,

$$\mathbb{E}\|x_k - x\|^2 \le \exp\Big([1 - o(1)]\frac{k}{\kappa(A)^2 n}\Big) \cdot \|x_0 - x\|^2. \tag{9}$$

Thus the algorithm converges exponentially fast to the solution in $O(n)$ iterations (projections). Each projection can be computed in $O(n)$ time; thus the algorithm takes $O(n^2)$ operations to converge to the solution. This should be compared to the Gaussian elimination, which takes $O(mn^2)$ time. (Strassen's algorithm and its improvements reduce the exponent in Gaussian elimination, but these algorithms are, as of now, of no practical use). Of course, we have to know the (approximate) Euclidean lengths of the rows of A before we start iterating; computing them takes $O(nm)$ time. But the lengths of the rows may in many cases be known a priori. For example, all of them may be equal to one (as is the case for Vandermonde matrices arising in trigonometric approximation) or be tightly concentrated around a constant value (as is the case for Gaussian random matrices).

The number m of equations is essentially irrelevant for our algorithm, as seen from (9). The algorithm does not even need to know the whole matrix, but only $O(n)$ random rows. Such Monte-Carlo methods have been successfully developed for many problems, even with precisely the same model of selecting a random submatrix of A (proportional to the squares of the lengths of the rows), see [5] for the original discovery and [14] for subsequent work and references.

3 Optimality

We discuss conditions under which our algorithm is optimal in a certain sense, as well as the optimality of the estimate on the expected rate of convergence.

3.1 General lower estimate

For any system of linear equations, our estimate can not be improved beyond a constant factor of R, as shown by the following theorem.

Theorem 3. *Consider the linear system of equations* (1) *and let x be its solution. Then there exists an initial approximation x_0 such that*

$$\mathbb{E}\|x_k - x\|^2 \geq \left(1 - \frac{2k}{R}\right) \cdot \|x_0 - x\|^2 \tag{10}$$

for all k, where $R = R(A) = \|A^{-1}\|^2 \, \|A\|_F^2$.

Proof. For this proof we can assume without loss of generality that the system (1) is homogeneous: $Ax = 0$. Let x_0 be a vector which realizes R, that is $R = \|A^{-1}x_0\|^2 \, \|A\|_F^2$ and $\|x_0\| = 1$. As in the proof of Theorem 2, we define the random normal Z associated with the rows of A by (7). Similar to (8), we have $\mathbb{E}|\langle x_0, Z\rangle|^2 = 1/R$. We thus see $\mathrm{lin}(x_0)$ as an "exceptional" direction, so we shall decompose $\mathbb{R}^n = \mathrm{lin}(x_0) \oplus (x_0)^\perp$, writing every vector $x \in \mathbb{R}^n$ as

$$x = x' \cdot x_0 + x'', \quad \text{where} \quad x' \in \mathbb{R}, \quad x'' \in (x_0)^\perp.$$

In particular,

$$\mathbb{E}|Z'|^2 = 1/R. \tag{11}$$

We shall first analyze the effect of one random projection in our algorithm. To this end, let $x \in \mathbb{R}^n$, $\|x\| \leq 1$, and let $z \in \mathbb{R}^n$, $\|z\| = 1$. (Later, x will be the running approximation x_{k-1}, and z will be the random normal Z). The projection of x onto the hyperplane whose normal is z equals

$$x_1 = x - \langle x, z\rangle z.$$

Since

$$\langle x, z\rangle = x'z' + \langle x'', z''\rangle, \tag{12}$$

we have

$$|x_1' - x'| = |\langle x, z\rangle z'| \leq |x'||z'|^2 + |\langle x'', z''\rangle z'| \leq |z'|^2 + |\langle x'', z''\rangle z'| \tag{13}$$

because $|x'| \leq \|x\| \leq 1$. Next,

$$\begin{aligned} \|x_1''\|^2 - \|x''\|^2 &= \|x'' - \langle x, z\rangle z''\|^2 - \|x''\|^2 \\ &= -2\langle x, z\rangle\langle x'', z''\rangle + \langle x, z\rangle^2\|z''\|^2 \leq -2\langle x, z\rangle\langle x'', z''\rangle + \langle x, z\rangle^2 \end{aligned}$$

because $\|z''\| \le \|z\| = 1$. Using (12), we decompose $\langle x, z\rangle$ as $a + b$, where $a = x'z'$ and $b = \langle x'', z''\rangle$ and use the identity $-2(a+b)b + (a+b)^2 = a^2 - b^2$ to conclude that

$$\|x_1''\|^2 - \|x''\|^2 \le |x'|^2|z'|^2 - \langle x'', z''\rangle^2 \le |z'|^2 - \langle x'', z''\rangle^2 \tag{14}$$

because $|x'| \le \|x\| \le 1$.

Now we apply (13) and (14) to the running approximation $x = x_{k-1}$ and the next approximation $x_1 = x_k$ obtained with a random $z = Z_k$. Denoting $p_k = \langle x_k'', Z_k''\rangle$, we have by (13) that $|x_k' - x_{k-1}'| \le |Z_k'|^2 + |p_k Z_k'|$ and by (14) that $\|x_k''\|^2 - \|x_{k-1}''\|^2 \le |Z_k'|^2 - |p_k|^2$. Since $x_0' = 1$ and $x_0'' = 0$, we have

$$|x_k' - 1| \le \sum_{j=1}^k |x_j' - x_{j-1}'| \le \sum_{j=1}^k |Z_j'|^2 + \sum_{j=1}^k |p_j Z_j'| \tag{15}$$

and

$$\|x_k''\|^2 = \sum_{j=1}^k \left(\|x_j''\|^2 - \|x_{j-1}''\|^2\right) \le \sum_{j=1}^k |Z_j'|^2 - \sum_{j=1}^k |p_j|^2.$$

Since $\|x_k''\|^2 \ge 0$, we conclude that $\sum_{j=1}^k |p_j|^2 \le \sum_{j=1}^k |Z_j'|^2$. Using this, we apply Cauchy-Schwartz inequality in (15) to obtain

$$|x_k' - 1| \le \sum_{j=1}^k |Z_j'|^2 + \Big(\sum_{j=1}^k |Z_j'|^2\Big)^{1/2}\Big(\sum_{j=1}^k |Z_j'|^2\Big)^{1/2} = 2\sum_{j=1}^k |Z_j'|^2.$$

Since all Z_j are copies of the random vector Z, we conclude by (11) that $\mathbb{E}|x_k' - 1| \le 2k\mathbb{E}|Z'|^2 \le \frac{2k}{R}$. Thus $\mathbb{E}\|x_k\| \ge \mathbb{E}|x_k'| \ge 1 - \frac{2k}{R}$. This proves the theorem, actually with the stronger conclusion

$$\mathbb{E}\|x_k - x\| \ge \Big(1 - \frac{2k}{R}\Big) \cdot \|x_0 - x\|.$$

(the actual conclusion follows by Jensen's inequality). ∎

3.2 The upper estimate is attained

If $\kappa(A) = 1$ then the estimate in Theorem 2 becomes an equality. This follows directly from the proof of Theorem 2.

Furthermore, there exist arbitrarily large systems and with arbitrarily large $\kappa(A)$ for which the estimate in Theorem 2 becomes an equality. More precisely, let n and $m \ge n$, $R \ge n$ be arbitrary positive numbers such that $\frac{1}{R}m$ is an integer. Then

there exists a system (1) of m equations in n variables and with $R(A) = R$ for which the estimate in Theorem 2 becomes an equality.

To see this, we define the matrix A with the help of any orthogonal set $e_1, \dots, e_n$ in $\mathbb{R}^n$. Let the first $\frac{1}{R}m$ rows of A be equal to e_1, the other rows of A be equal to one of the vectors e_j, $j > 1$, so that every vector from this set repeats at least $\frac{1}{R}m$ times as a row (this is possible because $R \geq n$). Then $R(A) = R$ (note that (5) is attained for $z = e_1$).

Let us test our algorithm on the system $Ax = 0$ with the initial approximation $x_0 = e_1$ to the solution $x = 0$. Every step of the algorithm brings the running approximation to 0 with probability $\frac{1}{R}$ (the probability of picking the row of A equal to e_1 in uniform sampling), and leaves the running approximation unchanged with probability $1 - \frac{1}{R}$. By the independence, for all k

$$\mathbb{E}\|x_k - x_0\|^2 = \left(1 - \frac{1}{R}\right)^k \cdot \|x_0 - x\|^2.$$

4 Numerical experiments and comparisons

In recent years conjugate gradient (CG) type methods have emerged as the leading iterative algorithms for solving large linear systems of equations, since they often exhibit remarkably fast convergence. How does the proposed random Kaczmarz method compare to CG algorithms?

It is not surprising, that one can easily construct examples for which CG (or its variations, such as CGLS or LSQR [8]) will clearly outperform the proposed method. For instance, take a matrix whose singular values, all but one, are equal to one, while the remaining singular value is ε, a number close to zero, say 10^{-8}. It follows from well known properties of the CG method (cf. [16]) that CGLS will converge in two steps, while the proposed Kaczmarz method will converge extremely slow, since $R \approx \varepsilon^{-2}$ and thus $1 - \frac{1}{R} \approx 1$ in this example.

On the other hand, the proposed algorithm outperforms CGLS in cases for which CGLS is actually quite well suited. We consider a Gaussian random matrix with $m \geq n$. While one iteration of CG requires $\mathcal{O}(mn)$ operations, one iteration (i.e., one projection) of Kaczmarz takes $\mathcal{O}(n)$ operations. Thus a cycle of m Kaczmarz iterations corresponds to one iteration of CG. Therefore, for a fair comparison, in the following we will compare the number of iteration cycles (1 iteration cycle for CGLS equals one standard CGLS iteration, and 1 iteration cycle for Kaczmarz equals m random projections). We let $m = 400, n = 100$ and construct 1000 random matrices. For each of them we run CGLS and the random Kaczmarz method described in Algorithm 1 (which does not require any preprocessing in this case since all rows of A have approximately the same norm). The resulting average rate of convergence

for both methods is displayed in Figure 1.

Somewhat surprisingly, Algorithm 1 gives faster convergence than CGLS. Classical results about Gaussian random matrices [7, 4], combined with convergence estimates for the CG algorithm [8] and a little algebra yield that the (expected) convergence rate of CG for Gaussian $m \times n$ matrices is governed by $(\sqrt{\frac{n}{m}})^k$. Whereas for Algorithm 1 the expected convergence rate is bounded by $(1 - \frac{(\sqrt{m}-\sqrt{n})^2}{mn})^{\frac{k}{2}}$ which is inferior to the value computed for CG. Yet, numerical experiments clearly demonstrate the better performance of Algorithm 1. We will give a more thorough discussion of this performance gain compared to its theoretical prediction elsewhere.

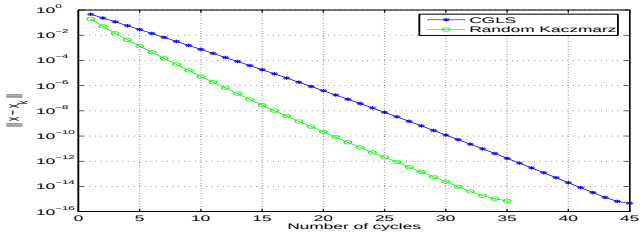

Figure 1: Comparison of rate of convergence for the random Kaczmarz method described in Algorithm 1 and the conjugate gradient least squares algorithm.

References

[1] C. Cenker, H. G. Feichtinger, M. Mayer, H. Steier, and T. Strohmer. New variants of the POCS method using affine subspaces of finite codimension, with applications to irregular sampling. In *Proc. SPIE: Visual Communications and Image Processing*, pages 299–310, 1992.

[2] F. Deutsch. Rate of convergence of the method of alternating projections. In *Parametric optimization and approximation (Oberwolfach, 1983)*, volume 72 of *Internat. Schriftenreihe Numer. Math.*, pages 96–107. Birkhäuser, Basel, 1985.

[3] F. Deutsch and H. Hundal. The rate of convergence for the method of alternating projections. II. *J. Math. Anal. Appl.*, 205(2):381–405, 1997.

[4] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[5] A. Frieze, R. Kannan and S. Vempala, *Fast Monte-Carlo Algorithms for finding low-rank approximations*, Proceedings of the Foundations of Computer Science, 1998, pp. 378–390, journal version in Journal of the ACM 51 (2004), 1025-1041

[6] A. Galántai. On the rate of convergence of the alternating projection method in finite dimensional spaces. *J. Math. Anal. Appl.*, 310(1):30–44, 2005.

[7] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[8] G.H. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins, Baltimore, third edition, 1996.

[9] G.T. Herman. *Image reconstruction from projections*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. The fundamentals of computerized tomography, Computer Science and Applied Mathematics.

[10] G.T. Herman and L.B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Transactions on Medical Imaging*, 12(3):600–609, 1993.

[11] G.N. Hounsfield. Computerized transverse axial scanning (tomography): Part I. description of the system. *British J. Radiol.*, 46:1016–1022, 1973.

[12] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Internat. Acad. Polon.Sci. Lettres A*, pages 335–357, 1937.

[13] F. Natterer. *The Mathematics of Computerized Tomography*. Wiley, New York, 1986.

[14] M. Rudelson and R. Vershynin. Sampling from large matrices: an approach through geometric functional analysis, 2006. preprint.

[15] 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*, pages 415–462. Acad. Press, 1987.

[16] A. van der Sluis and H.A. van der Vorst. The rate of convergence of conjugate gradients. *Numer. Math.*, 48:543–560, 1986.