RANDOMIZED KACZMARZ SOLVER FOR NOISY LINEAR SYSTEMS

DEANNA NEEDELL

Abstract. The Kaczmarz method is an iterative algorithm for solving systems of linear equations $Ax = b$. Theoretical convergence rates for this algorithm were largely unknown until recently when work was done on a randomized version of the algorithm. It was proved that for overdetermined systems, the randomized Kaczmarz method converges with expected exponential rate, independent of the number of equations in the system. Here we analyze the case where the system $Ax = b$ is corrupted by noise, so we consider the system $Ax \approx b + r$ where r is an arbitrary error vector. We prove that in this noisy version, the randomized method reaches an error threshold dependent on the matrix A with the same rate as in the error-free case. We provide examples showing our results are sharp in the general context.

1. Introduction

The Kaczmarz method [8] is one of the most popular solvers of overdetermined linear systems and has numerous applications from computer tomography to image processing. It is an iterative method, and so therefore is practical in the realm of very large systems of equations. The algorithm consists of a series of alternating projections, and is often considered a type of *Projection on Convex Sets* (POCS) method. Given a consistent system of linear equations of the form

$$Ax = b,$$

the Kaczmarz method iteratively projects onto the solution spaces of each equation in the system. That is, if $a_1, \dots, a_m \in \mathbb{R}^n$ denote the rows of A, the method cyclically projects the current estimate orthogonally onto the hyperplanes consisting of solutions to $\langle a_i, x\rangle = b_i$. Each iteration consists of a single orthogonal projection. The algorithm can thus be described using the recursive relation,

$$x_{k+1} = x_k + \frac{b_i - \langle a_i, x_k\rangle}{\|a_i\|_2^2} a_i,$$

where x_k is the k^{th} iterate and $i = (k \bmod m) + 1$.

Although the Kaczmarz method is popular in practice, theoretical results on the convergence rate of the method have been difficult to obtain. Most known estimates depend on properties of the matrix A which may be time consuming to compute, and are not easily comparable to those of other iterative methods (see e.g. [3], [4], [5]).

Date: Jan. 18, 2009.

D.N. is with the Dept. of Statistics, Stanford University, 390 Serra Mall, Stanford CA 94305, USA. e-mail: dneedell@stanford.edu.

Since the Kaczmarz method cycles through the rows of A sequentially, its convergence rate depends on the order of the rows. Intuition tells us that the order of the rows of A does not change the difficulty level of the system as a whole, so one would hope for results independent of the ordering. One natural way to overcome this is to use the rows of A in a random order, rather than sequentially. Several observations were made on the improvements of this randomized version [9, 6], but only recently have theoretical results been obtained [11, 13].

1.1. **Randomized Kaczmarz.** In designing a random version of the Kaczmarz method, it is necessary to set the probability of each row being selected. Strohmer and Vershynin propose in [11, 13] to set the probability proportional to the Euclidean norm of the row. Their revised algorithm can then be described by the following:

$$x_{k+1} = x_k + \frac{b_{p(i)} - \langle a_{p(i)}, x_k \rangle}{\|a_{p(i)}\|_2^2} a_{p(i)},$$

where $p(i)$ takes values in $\{1, \ldots, m\}$ with probabilities $\frac{\|a_{p(i)}\|_2^2}{\|A\|_F^2}$. Here and throughout, $\|A\|_F$ denotes the Frobenius norm of A and $\|\cdot\|_2$ denotes the usual Euclidean norm or spectral norm for vectors or matrices, respectively. We note here that of course, one needs some knowledge of the norm of the rows of A in this version of the algorithm. In general, this computation takes $\mathrm{O}(mn)$ time. However, in many cases such as the case in which A contains Gaussian entries, this may be approximately or exactly known.

In [11, 13], Strohmer and Vershynin prove the following exponential bound on the expected rate of convergence for the randomized Kaczmarz method,

$$\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2, \tag{1.1}$$

where $R = \|A^{-1}\|^2 \|A\|_F^2$, x_0 is an arbitrary initial estimate, and $\mathbb{E}$ denotes the expectation (over the choice of the rows). Here and throughout, we will assume that A has full column rank so that $\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\}$ is well defined. We comment here that this particular mixed condition number comes as an immediate consequence of the simple probabilities used within the randomized algorithm.

The first remarkable note about this result is that it is essentially independent of the number m of equations in the system. Indeed, by the definition of R, R is proportional to n within a square factor of $\kappa(A)$, the condition number of A ($\kappa(A)$ is defined as the ratio of the largest to smallest singular values of A). This bound also demonstrates, however, that the Kaczmarz method is an efficient alternative to other methods only when the condition number is very small. If this is not the case, then other alternative methods may offer improvements in practice.

The bound (1.1) and the relationship of R to n shows that the estimate x_k converges exponentially fast to the solution in just $\mathrm{O}(n)$ iterations. Since each iteration requires $\mathrm{O}(n)$ time, the method overall has a $\mathrm{O}(n^2)$ runtime. Being an iterative algorithm, it is clear that the randomized Kaczmarz method is competitive only for very large systems. For such large systems, the runtime of $\mathrm{O}(n^2)$ is clearly superior to,

for example, Gaussian elimination which has a runtime of $O(mn^2)$. Also, since the algorithm needs only access to the randomly chosen rows of A, the method need not know the entire matrix A, which for very large systems is a clear advantage. Thus the interesting cases for the randomized method are those in which n and m are large, and especially those in which m is extremely large. Strohmer and Vershynin discuss in detail in Section 4.2 of [13] cases where the randomized Kaczmarz method even outperforms the conjugate gradient method (CGLS). They show that for example, randomized Kaczmarz computationally outperforms CGLS for Gaussian matrices when $m > 3n$. Numerical experiments in [13] also demonstrate advantages of the randomized Kaczmarz method in many cases.

Since the results of [11, 13], there has been some further discussion about the benefits of this randomized version of the Kaczmarz method (see [2, 12]). The Kaczmarz method has been studied for over seventy years, and is useful in many applications. The notion of selecting the rows randomly in the method has been proposed before (see [9, 1, 6]), and improvements over the standard method were observed. However, the work by Strohmer and Vershynin in [11, 13] provides the first proof on the rate of convergence. The rate is exponential in expectation and is in terms of standard matrix properties. We are not aware of any other Kaczmarz method that provably achieves exponential convergence.

It is important to note that the method of row selection proposed in this version of the randomized Kaczmarz method is *not* optimal, and an example that demonstrates this is given in [13]. However, under this selection strategy, the convergence rates proven in [11, 13] are optimal, and there are matrices that satisfy the proven bounds exactly. The selection strategy in this method was chosen because it often yields very good results, allows a provable guarantee of exponential convergence, and is computationally efficient.

Since the algorithm selects rows based on their row norms, it is natural to ask whether one can simply scale the rows any way one wishes. Indeed, choosing the rows based on their norms is related to the notion of applying a diagonal preconditioner. However, since finding the optimal diagonal preconditioner for a system $Ax = b$ is itself a task that is often more costly than inverting the entire matrix, we select an easier, although not optimal, preconditioner that simply scales by the (square of the) row norms. This type of preconditioner yields a balance of computational cost and optimality (see [14, 10]). The distinction between the effect of an alternative diagonal preconditioner on the Kaczmarz method versus the randomized method discussed here is important. If the system is multiplied by a diagonal matrix, the standard Kaczmarz method will not change, since the angles between all rows do not change. However, such a multiplication to the system in our randomized setting changes the probabilities of selecting the rows (by definition). It is then not a surprise that this will also affect the convergence rate proved for this method (since multiplication will affect the value of R in (1.1)).

This randomized version of the Kaczmarz method provides clear advantages over the standard method in many cases. Using the selection strategy above, Strohmer and Vershynin were able to provide a proof for the expected rate of convergence

that shows exponential convergence. No such convergence rate for any Kaczmarz method has been proven before. These benefits lead one to question whether the method works in the more realistic case where the system is corrupted by noise. In this paper we provide theoretical and empirical results to suggest that in this noisy case the method converges exponentially to the solution within a specified error bound. The error bound is proportional to $\sqrt{R}$, and we also provide a simple example showing this bound is sharp in the general setting.

2. Main Results

Theoretical and empirical studies have shown the randomized Kaczmarz algorithm to provide very promising results. Here we show that it also performs well in the case where the system is corrupted with noise. In this section we consider the consistent system $Ax = b$ after an error vector r is added to the right side:

$$Ax \approx b + r.$$

Note that we do not require the perturbed system to be consistent. First we present a simple example to gain intuition about how drastically the noise can affect the system. To that end, let A be the $n \times n$ identity matrix, $b = 0$, and suppose the error is the vector whose entries are all one, $r = (1, 1, \ldots, 1)$. Then the solution to the noisy system is clearly $x = r = (1, 1, \ldots, 1)$, and the solution to the unperturbed problem is $x = 0$. By Jensen's inequality, we have

$$\Big(\mathbb{E}\|x_k - r\|_2\Big)^2 \leq \mathbb{E}\Big(\|x_k - r\|_2^2\Big).$$

Now considering the noisy problem, we may substitute r for x in (1.1). Combining this with Jensen's inequality above, we obtain

$$\mathbb{E}\|x_k - r\|_2 \leq \Big(1 - \frac{1}{R}\Big)^{k/2} \|x_0 - r\|_2. \tag{2.1}$$

Then by the triangle inequality, we have

$$\|r - x\|_2 \leq \|r - x_k\|_2 + \|x_k - x\|_2.$$

Next, by taking expectation and using (2.1) above, we have

$$\mathbb{E}\|x_k - x\|_2 \geq \|r - x\|_2 - \Big(1 - \frac{1}{R}\Big)^{k/2} \|x_0 - r\|_2.$$

Finally by the definition of r and R, this implies

$$\mathbb{E}\|x_k - x\|_2 \geq \sqrt{R} - \Big(1 - \frac{1}{R}\Big)^{k/2} \|x_0 - r\|_2.$$

This means that the limiting error between the iterates x_k and the original solution x is $\sqrt{R}$. In [11, 13] it is shown that the bound provided in (1.1) is optimal, so even this trivial example demonstrates that if we wish to maintain a general setting, the best error bound for the noisy case we can hope for is proportional to $\sqrt{R}$. Our main result proves this exact theoretical bound.

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. *Proc. SPIE: Visual Communications and Image Processing*, pages 299–310, 1992.

[2] Y. Censor, G.T. Herman, and M. Jiang. A note on the behavior of the randomized Kaczmarz algorithm of Strohmer and Vershynin. *J. Fourier Anal. Appl.*, 15:431–436, 2009.

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

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

[5] M. Hanke and W. Niethammer. On the acceleration of Kaczmarz's method for inconsistent linear systems. *Linear Alg. Appl.*, 130:83–98, 1990.

[6] 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.

[7] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1985.

[8] S. Kaczmarz. Angen¨aherte aufl¨osung von systemen linearer gleichungen. *Bull. Internat. Acad. Polon.Sci. Lettres A*, pages 335–357, 1937.

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

[10] A. Shapiro. Upper bounds for nearly optimal diagonal scaling of matrices. *Linear and Multilinear Algebra*, 29:145–147, 1991.

[11] T. Strohmer and R. Vershynin. A randomized solver for linear systems with exponential convergence. In *RANDOM 2006 (10th International Workshop on Randomization and Computation)*, number 4110 in Lecture Notes in Computer Science, pages 499–507. Springer, 2006.

[12] T. Strohmer and R. Vershynin. Comments on the randomized Kaczmarz method. *J. Fourier Anal. Appl.*, 15:437–440, 2009.

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

[14] A. van der Sluis. Condition numbers and equilibration of matrices. *Numer. Math.*, 14:14–23, 1969.