AMO - Advanced Modeling and Optimization, Volume 2, Number 2, 2000

Kaczmarz's Algorithm - Extensions and Preconditioning

CONSTANTIN POPA, "Ovidius" University, Faculty of Mathematics and Computer Science, Blvd. Mamaia 124, 8700 Constanta, Romania;
email: cpopa@univ-ovidius.ro

Keywords: Kaczmarz's algorithm, inconsistent linear least-squares problems, approximate orthogonalization, preconditioning.

Abstract. In this paper we present some recent developments concerning the Kaczmarz's projections algorithm. After the description of some extensions, we present some possibilities to improve its convergence properties. Numerical experiments are also described at the end of the paper.

1 Kaczmarz's projections method

Let A be an $m \times n$ real matrix (e.g., $n \leq m$ for simplifying the presentation) and $b \in I\!R^m$ a given vector. Let $A^t, R(A), N(A), r(A)$ be the transpose, range, null-space and rank of A respectively. If $S \subset I\!R^q$ is a closed convex subset, by P_S we shall denote the orthogonal projection on S (all these are considered with respect to the Euclidean scalar product and norm denoted by $<\cdot,\cdot>$ and $\|\cdot\|$, respectively).

Now, let us firstly suppose that $n = m$, the matrix A is invertible and consider the system

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

with $x^* \in I\!R^n$ its unique solution. By $H_i = \{x \in I\!R^n, < x, a_i >= b_i\}$ we shall denote the hyperplane determined by the i-th equation of (1). In [3], S. Kaczmarz proposed his well-known projections method for approximating x^*. This algorithm starts with an $x^0 \in I\!R^n$ and generates the sequence of approximations $(x^k)_{k\geq 0}$, by succesive orthogonal projections on $H_i, i = 1,\ldots,m$. These can be easily written as follows : let $x^0 \in I\!R^n$ and for $k = 0, 1, \ldots$ do

$$x^{k+1} = F(A; b; x^k), k \geq 0, \tag{2}$$

$$F(A; b; x) = (f_1 \circ \ldots \circ f_m)(A; b; x), \tag{3}$$

$$f_i(A; b; x) = x - \frac{< x, a_i > - b_i}{\| a_i \|^2} a_i, x \in I\!R^n \tag{4}$$

where a_i is the i-th row of A and b_i the i-th component of b. In [3] Kaczmarz proved the following result.

Theorem 1 *For any square nonsingular matrix A, any $b \in I\!R^m$ and $x^0 \in I\!R^n$, the sequence $(x^k)_{k\geq 0}$ generated by* $(2)-(4)$ *converges to x^*.*

The first important extension of Kaczmarz's result was presented by K. Tanabe in [10]. He considered a rectangular matrix A and, instead of (1) the least-squares problem : find $x^* \in I\!R^n$ such that

$$\| Ax^* - b \| = \min! \tag{5}$$

Let $LSS(A;b)$ be the set of all (least-squares) solutions of (5), x_{LS} the (unique) minimal norm solution and suppose that A has nonzero rows, i.e.

$$a_i \neq 0, \ i = 1, \ldots, m. \tag{6}$$

In [10] Tanabe proved the following result (partial statement).

Theorem 2 *Let us suppose that the matrix A satisfies* (6). *If the problem* (5) *is consistent then, for any $x^0 \in I\!R^n$ the sequence $(x^k)_{k\geq 0}$ generated by* $(2)-(4)$ *is convergent and*

$$\lim_{k \longrightarrow \infty} x^k = P_{N(A)}(x^0) + Gb, \tag{7}$$

where G is a generalized inverse of A (see e.g. [1]). Moreover, the minimal norm solution x_{LS} is obtained as a limit point in (7) *if and only if $x^0 \in I\!R(A^t)$. In this case $x_{LS} = Gb$.*

In the general inconsistent case for (5), the limit in (7) still exists, but its value does not represent a solution any more. Thus, in order to extend Tanabe's results to the inconsistent case, we had to add some steps to the original Kaczmarz algorithm (2)-(3). In this sense, we considered the orthogonal projections φ_j, Φ, similar with those from (3)-(4), but with respect to the columns $\alpha_j, j = 1, \ldots, n$ of A, defined by

$$\begin{aligned} \varphi_j(A;y) &= y - \frac{< y, \alpha_j >}{\| \alpha_j \|^2} \alpha_j, j = 1, \ldots, n, & (8) \\ \Phi(A;y) &= (\varphi_1 \circ \ldots \circ \varphi_n)(A;y), y \in I\!R^m. & (9) \end{aligned}$$

If we introduce an assumption for α_j similar to (6), i.e.

$$\alpha_j \neq 0, \ j = 1, \ldots, n, \tag{10}$$

we can define the following algorithm (denoted by **(KE)** : set
$y^0 = b, x^0 \in I\!R^n$ and for $k = 0, 1, 2, \ldots$ succesively compute

$$y^{k+1} = \Phi(A; y^k),\ b^{k+1} = b - y^{k+1},\ x^{k+1} = F(A; b^{k+1}; x^k). \tag{11}$$

In [5] we proved the following result.

Theorem 3 *Let us suppose that the matrix A satisfies* (6) *and* (10)*. Then, for any $b \in I\!R^m$ and $x^0 \in I\!R^n$ the sequence $(x^k)_{k\geq 0}$ generated by* (11) *is convergent and*

$$\lim_{k\longrightarrow\infty} x^k = P_{N(A)}(x^0) + Gb_A, \tag{12}$$

where $b_A = P_{R(A)}(b)$. Moreover, the minimal norm solution x_{LS} is obtained as a limit point in (12) *if and only if $x^0 \in I\!R(A^t)$. In this case, $x_{LS} = Gb_A$.*

Remark 1 *Other extensions of the form* (11) *(e.g. with relaxation parameters, block-row and block-column versions, etc.) have been obtained and described in* [*6*].

Remark 2 *Unfortunately, the above described Kaczmarz-like algorithms usually converge (very) slow. Moreover, it was observed in numerical experiments (and partially proved) that the convergence properties are strongly related to the values of the angles between the rows (and columns) of A, i.e. "closer" to 90^0 are these values, better will be the convergence properties.*

2 Kovarik's orhtogonalization algorithm

We shall begin this section of the paper with the presentation of an "approximate orthogonalization" method proposed by Z. Kovarik in [4]. Let $(a_k)_{k\geq 0}$ be the sequence of real numbers defined by the Taylor's expansion

$$(1-t)^{-\frac{1}{2}} = a_0 + a_1 t + a_2 t^2 + \ldots,\quad t \in (-1, 1), \tag{13}$$

i.e.

$$a_j = \frac{1}{2^{2j}} \frac{(2j)!}{(j!)^2},\quad j \geq 0 \tag{14}$$

and $(q_k)_{k\geq 0}$ a given sequence of positive integers. Kovarik's method ([4], Algorithm A, page 386) is the following : let $A_0 = A$ and recursively define the sequences of matrices $(H_k)_{k\geq 0}$, $(A_k)_{k\geq 0}$ by

$$H_k = I - A_k A_k^t,\ A_{k+1} = (I + a_1 H_k + a_2 H_k^2 + \ldots + a_{q_k} H_k^{q_k}) A_k \tag{15}$$

(where by I we denoted the corresponding unit matrix). In the above mentioned paper, Kovarik proved the following results.

Theorem 4 *If $m \leq n$ and the rows of A are linearly independent, the sequence $(A_k)_{k\geq 0}$ generated in* (15) *converges to A_{infty}, given by*

$$A_{infty} = ((AA^t)^{\frac{1}{2}})^{-1} A. \tag{16}$$

Moreover $\parallel K_0 \parallel < 1$ and

$$\parallel A_{infty} - A_k \parallel \ \ \leq \ \ \parallel K_0 \parallel^{2^k}, \forall\ k \geq 1, \tag{17}$$

where by $\parallel B \parallel$ we denoted the spectral norm of the matrix B.

Remark 3 *In the particular case for A from Theorem 4, it can be proved that the matrix A_{infty} has mutually orthogonal rows.*

Unfortunately, the result from the above Remark 3 can not be true for a general matrix A, thus a generalization of Kovarik's algorithm (15) had to be obtained. For this, we had to replace A_{infty} from (16) with its ("natural") generalization A_∞, defined by

$$A_\infty = [(AA^t)^{\frac{1}{2}}]^+ A, \tag{18}$$

(where B^+ is the Moore-Penrose pseudoinverse of B; see e.g. [1]) and to add the supplementary assumption

$$\parallel AA^t \parallel \ < \ 1. \tag{19}$$

In [9] we proved the following result.

Theorem 5 *Let A be an arbitrary $m \times n$ (real) matrix which satisfies* (19)*. Then, the sequence $(A_k)_{k\geq 0}$ generated in* (15) *converges to A_∞ from* (18)*. Moreover, it exists $\delta \in (0, 1)$ such that*

$$\| A_\infty - A_k \| \leq \delta^{2^k}, \; \forall \; k \geq 1. \tag{20}$$

Remark 4 *The assumption* (19) *is not at all restrictive. It can be fulfilled by an appropriate scaling of the elements of* A*, e.g.*

$$A = \frac{1}{\sqrt{\| A^t A \|}} \cdot A, \tag{21}$$

where $\| \cdot \|$ *is some matrix norm (see e.g. [1]). Moreover it can be easily proved that transformations of the form* (21) *do not change the set* $LSS(A; b)$*.*

Remark 5 *As we have already observed, the matrix* A_∞ *does not (generally) have mutually orthogonal rows. But, an improvement of the angles between the rows of* A *has been observed in numerical experiments (see e.g. [8], [9]).*

We are now in the position to present two preconditioned techniques for improving the convergence properties of Kaczmarz-like algorithms described in the first section of the paper of the paper.

Algorithm 1. Preconditioned Kaczmarz

Part I. Set $A_0 = A, b^0 = b$ and let $N(A) \geq 1$ be a fixed integer. For $k = 0, 1, \ldots, N(A) - 1$ apply the transformations from **(KO)** algorithm (15) to A and b and obtain $A_{N(A)}$ and $b^{N(A)}$.

Part II. Apply the Kaczmarz-extended algorithm **(KE)** from (11), with respect to $A_{N(A)}$ and $b^{N(A)}$.

Algorithm 2. Kaczmarz-Kovarik

Set $x^0 = 0, b^0 = b, A_0 = A$ and for $k \geq 0$ do

$$A_{k+1} = (\mathbf{KO})(A_k), \quad b^{k+1} = (\mathbf{KO})(b^k); \tag{22}$$

$$\beta^{k+1} = b^{k+1} - \Phi^{k+1}(A_{k+1}, b^{k+1}); \tag{23}$$

$$x^{k+1} = F(A_{k+1}; \beta^{k+1}; x^k). \tag{24}$$

Remark 6 *Both the above described methods are "well-defined" because of the comments in the second part of Remark* 4*. Moreover, convergence results can be easily obtained or have already been proved for both the above algorithms (see e.g. [7], [8]).*

3 Numerical experiments

We considered in our numerical experiments, as model problem, the following first kind integral equation (see e.g. [2]): find
$x^*(t) \in L^2([0,1])$ such that

$$(Tx^*)(s) = \int_0^1 k(s,t)\ x^*(t)dt = y(s),\ \ s \in [0,1], \tag{25}$$

where

$$k(s,t) = \frac{1}{\sqrt{1+(2s-1)^2(2t-1)^2}},\ \ y(s) = 2s-1+w(s), \tag{26}$$

$$w(s) = \begin{cases} \frac{ln((2s-1)+\sqrt{1+(2s-1)^2})}{2s-1}, & s \neq 0.5 \\ 1, & s = 0.5. \end{cases} \tag{27}$$

$L^2([0,1])$ is the vector space of Lebesgue square integrable functions $f : [0,1] \longrightarrow I\!R$ equiped with the norm

$$\| f \|_{L^2([0,1])} = \sqrt{\int_0^1 (f(t))^2 dt}. \tag{28}$$

If by $R(T), R(T)^\perp$ we denote the range of T and its orthogonal complement in $L^2([0,1])$, respectively, then the function y from (26)-(27) (which is differentiable on $(0,1)$) satisfies

$$y \in R(T)^\perp \oplus R(T) = D(T^+), \tag{29}$$

(because $w(s) = (Tx_0)(s)$, with $x_0 : [0,1] \longrightarrow [0,1]$ defined by $x_0(t) = 1$, $\forall t \in [0,1]$), where $D(T^+)$ is the domain of the Moore-Penrose pseudoinverse of T. Then, we reformulated the problem (25) in the least-squares sense (see [2]) as follows: find $x^* \in L^2([0,1])$ the minimal $L^2([0,1])$ - norm solution of

$$\| Tx^* - y \|_{L^2([0,1])} = \min! \tag{30}$$

We discretized the problem (30) using the least-squares collocation algorithm described in [2], i.e. for two given integers $m, n \geq 2$ we defined the $m \times n$ real matrix $\hat{A}$ and the vector $b \in I\!R^m$ by

$$(\hat{A})_{ij} = \int_0^1 k(s_i,t)\ k(\sigma_j,t)\ dt,\ \ (b)_i = y(s_i). \tag{31}$$

where $s_i = \frac{i-1}{m-1}$, $\sigma_j = \frac{j-1}{n-1}$ are the collocation points in $[0,1]$. But, because the integrals in (31) can not be exactly computed, we approximated them using the rectangles quadrature formula ("midle point") with 16 equally spaced points in $[0,1]$ obtaining an $m \times n$ matrix A. With this matrix and the vector b from (31) we considered a problem of the form (5). We then used both the preconditioned algorithms described before for approximating its minimal norm solution x_{LS}. We started with $x^0 = 0$ and used in the **(KO)** part of them (see (15)) the values

$$q_k = 3, \ \forall k \geq 0.$$

Because the components of x_{LS} were "big numbers" (of order $10^9 - 10^{11}$) we used the following "relative-error" stopping test:

$$\frac{\| x^{k+1} - x^k \|}{\| x^k \|} \leq 10^{-6}. \tag{32}$$

The results, for $n = 9, m = 17$ and different values for $N(A)$ with **Algorithm 1** are described in Table 1 and, for different values of n and m with **Algorithm 2** in Table 2.

Final remarks.

1. The results from the Tables 1 and 2 show that both preconditioned algorithms have very good convergence properties which are independent on the dimensions m and n of A.

2. The "bad thing" in **Algorithm 1** is the choice of $N(A)$. Unfortunately, we don't have yet a "search method" and the results in Table 1 seem to tell us that the convergence properties are very "sensitive" with respect to $N(A)$ (we used in our tests a stopping rule of the form $\| A_{k+1} - A_k \| \leq 10^{-6}$ in **Part I** of **Algorithm 1**, which gave us the "good" values $N(A) \in \{35, \ldots, 50\}$).

3. The "bad thing" in **Algorithm 2** is the big number of operations in step (22) (i.e. the application of Φ, $(k+1)$ - times).

4. In both algorithms, the "parts" corresponding to Kovarik's transformations (15) are "very costly" from a computational point of view. But, it seems that this difficulty can be diminished by parallelization. Work on this subject is in progress.

Table 1 Numerical experiments with Preconditioned Kaczmarz

$N(A)$	*Number of iterations for* (32)
50	10
40	10
35	11
30	137
25	≥ 5000
20	≥ 10.000

Table 2 Numerical experiments with Kaczmarz - Kovarik

n	m	*Number of iterations for* (32)
33	33	35
65	65	37
129	129	39
257	257	40
9	33	43
9	65	44
9	129	45
9	257	47

References

[1] Å. Björck, *Numerical methods for least squares problems*, SIAM Philadelphia, 1996.

[2] H. W. Engl, *Regularization methods for the stable solution of inverse problems*, Surv. Math. Ind. **3**(1993), 71 - 143.

[3] S. Kaczmarz, *Angenaherte Auflosung von Systemen linearer Ghleichungen*, Bull. Acad. Polon. Sci. ct Lettres A, 1937, 355 - 357.

[4] Z. Kovarik, *Some iterative methods for improving orthogonality*, SIAM J. Numer. Anal., **7(3)**(1970), 386 - 389.

[5] C. Popa, *Characterization of the solutions set of inconsistent least-squares problems by an extended Kaczmarz algorithm*, Korean Journal on Computational and Applied Mathematics, **6(3)**(1999), 523 - 535.

[6] C. Popa, *Extensions of block-projections methods with relaxation parameters to inconsistent and rank-defficient least-squares problems*; *B I T*, **38(1)**(1998), pp. 151-176.

[7] C. Popa, *Projections and approximate orthogonalization for numerical solution of inconsistent least - squares problems*, Preprint Series of the Institute of Math. of the Romanian Academy of Sciences, Bucharest, no. **6**, 1999.

[8] C. Popa, *A fast Kaczmarz-Kovarik algorithm for consistent least squares problems*, to appear in Korean J. Comp. Appl. Math., 2000.

[9] C. Popa, *Extensions of an approximate orthogonalization algorithm to arbitrary rectangular matrices*, Preprint Series of the Institute of Math. of the Romanian Academy of Sciences, Bucharest, no. **18**, 1998.

[10] K. Tanabe, *Projection method for solving a singular system of linear equations and its applications*, Numer. Math., **17** (1971), pp. 203-214.

Constantin Popa graduated the Faculty of Mathematics and obtained the Ph.D at the University of Bucharest, Romania. Since 1990 he is teaching at the "Ovidius" University of Constanta, having now a full professor position. He had two DAAD research grants in Germany (Kiel and Augsburg) and a one year postdoc grant (1996-1997) at Department of Mathematics, Weizmann Institute, Israel. His research interests are centered on preconditioning techniques, iterative solutions of least - squares formulations and regularization techniques for inverse problems.