Electronic Transactions on Numerical Analysis.
Volume 29, pp. 150-162, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ALGEBRAIC MULTIGRID SMOOTHING PROPERTY OF KACZMARZ'S RELAXATION FOR GENERAL RECTANGULAR LINEAR SYSTEMS*

CONSTANTIN POPA†

Abstract. In this paper we analyze the smoothing property from classical Algebraic Multigrid theory, for general rectangular systems of linear equations. We prove it for Kaczmarz's projection algorithm in the consistent case and obtain in this way a generalization of the classical well-known result by A. Brandt. We then extend this result for the Kaczmarz Extended algorithm in the inconsistent case.

Key words. algebraic multigrid, smoothing property, Kaczmarz relaxation, inconsistent least squares problems

AMS subject classifications. 65F10, 65F20, 65N55

1. Algebraic Multigrid (AMG) - some historical comments. In the early 80's the AMG methods have been designed for the solution of (sparse) linear systems of equations using classical (geometric) multigrid ideas. The starting point and initial paper on the subject seems to be the 1982 Report [4] by Brandt, McCormick, and Ruge; see the references from [6]. In the same period of time, at the occasion of the 1983 International MG Conference at Copper Mountain, three other basic papers were presented; see [5, 16, 19]. Not far from this moment, in [17] Stüben and Ruge developed both theoretical aspects together with implementation issues of the classical AMG, as we think of it today. Moreover, although the initial theoretical results and efficient implementations were concerned with the class of symmetric M-matrices, recent developments, for which a rigorous convergence and optimality theory (level independence, dimension independence, etc.) have been obtained, are related to structured matrices (Toeplitz, circulants, sine/cosine, transform matrices, Hartley matrices, etc.). These matrices are characterized by the shift invariance property joint with proper boundary conditions; see [1, 2, 18] and references therein.

2. The smoothing property - definition and classical results. According to the basic principles of AMG, as described in the papers mentioned before, the *smoother* (AMG relaxation) is usually fixed among the classical iterative methods (Gauss - Seidel, Jacobi, Kaczmarz etc.). But, in order to design an efficient AMG code the *smoothing property* of this smoother has to be properly formulated and proved. In this respect, we shall briefly replay in what follows the basic ideas and results from the classical AMG theory. Let B be an $n \times n$ symmetric and positive definite matrix (SPD, for short), and $c \in I\!R^n$ a given vector. By B_i, B_{ij} we shall denote the i-th row and (i, j)-th element of B. All the vectors that will appear will be column vectors and the superscript T will indicate the transpose. For the purposes of this section we consider the system of linear equations

$$Bx^* = c, \tag{2.1}$$

where $x^* = B^{-1}c$ is its unique solution. Let

$$x^0 \in I\!R^n, \; x^{k+1} = Gx^k + g, \;\; k \geq 0, \tag{2.2}$$

*Received October 12, 2006. Accepted for publication July 3, 2007. Published online on July 22, 2008. Recommended by A. Frommer.

†Faculty of Mathematics and Computer Science, Ovidius University of Constanta, Blvd. Mamaia 124, 900527 Constanta, Romania (cpopa@univ-ovidius.ro).

ETNA
Kent State University
etna@mcs.kent.edu

or componentwise

$$x_i^{k+1} = g_i + \sum_{j=1}^{n} G_{ij} x_j^k,$$

be a (convergent) relaxation scheme for (2.1). Denoting by $\langle \cdot, \cdot \rangle, \| \cdot \|$ the Euclidean scalar product and norm, respectively, we define the energy norms $\| \cdot \|_B$ and $\| \cdot \|_{D^{-1}}$

$$\| z \|_B = \sqrt{\langle Bz, z \rangle}, \quad \| z \|_{D^{-1}} = \sqrt{\langle D^{-1}z, z \rangle}, \; z \in I\!R^n, \tag{2.3}$$

where

$$D = \mathrm{diag}(B) = \mathrm{diag}(B_{11}, \ldots, B_{nn}).$$

REMARK 2.1. This special (diagonal) choice of D is related to the classical approach considered in [17]. But, it can be replaced by any positive definite matrix D (see Remark 2 in [1] and also [2]). This further degree of flexibility could be useful for refining the convergence results.

Let x be a given approximation of x^*, $\bar{x}$ the one obtained after one step of the relaxation scheme (2.2) applied to x and $e, \bar{e}, r$ the corresponding errors and residual defined by

$$e = x - x^*, \quad r = Be = Bx - c, \quad \bar{e} = \bar{x} - x^*. \tag{2.4}$$

DEFINITION 2.2. *We say that the relaxation scheme* (2.2) *satisfies the* **smoothing property** *(SP, for short) for the system* (2.1) *if there is a constant* $\alpha > 0$ *(independent of the dimension* n *of* B*) such that*

$$\| \bar{e} \|_B^2 \;\leq\; \| e \|_B^2 - \alpha \| r \|_{D^{-1}}^2 . \tag{2.5}$$

In what follows we shall present classical results about the SP property for some well-known relaxation schemes of the type (2.2).

Gauss-Seidel. Let $x^0 \in I\!R^n$; for $k = 0, 1, \ldots$ do

$$x_i^{k+1} = \frac{1}{B_{ii}} \left[c_i - \sum_{j<i} B_{ij} x_j^{k+1} - \sum_{j>i} B_{ij} x_j^k \right], \forall i = 1, \ldots, n. \tag{2.6}$$

SOR. Let $x^0 \in I\!R^n$; for $k = 0, 1, \ldots$ do

$$x_i^{k+1} = (1-\omega) x_i^k + \frac{\omega}{B_{ii}} \left[c_i - \sum_{j<i} B_{ij} x_j^{k+1} - \sum_{j>i} B_{ij} x_j^k \right], \forall i = 1, \ldots, n. \tag{2.7}$$

Kaczmarz. Let $x^0 \in I\!R^n$; for $k = 0, 1, \ldots$ do

$$\left\{ \begin{array}{lcl} x^{k,0} & = & x^k \\ x^{k,i} & = & x^{k,i-1} - \frac{\langle x^{k,i-1}, B_i \rangle - c_i}{\langle B_i, B_i \rangle} B_i, i = 1, \ldots, n. \\ x^{k+1} & = & x^{k,n} \end{array} \right. \tag{2.8}$$

The following result analyzes property (2.5) for the Gauss-Seidel relaxation (for the proof see [7] and [21, Theorem A.3.1, p. 436]).

ETNA
Kent State University
etna@mcs.kent.edu

THEOREM 2.3. *Gauss-Seidel relaxation* (2.6) *for the system* (2.1) *satisfies* (2.5) *with* $\alpha = \gamma_0$ *given by*

$$\gamma_0 = \frac{1}{(1+\gamma_-(B))(1+\gamma_+(B))} \tag{2.9}$$

and

$$\gamma_-(B) = \max_{1\le i\le n} \sum_{j<i} \frac{|B_{ij}|}{B_{ii}};\ \ \gamma_+(B) = \max_{1\le i\le n} \sum_{j>i} \frac{|B_{ij}|}{B_{ii}}. \tag{2.10}$$

In [11] we extended the above theorem for SOR relaxation (2.7) in the following way:

THEOREM 2.4. *The SOR relaxation* (2.7) *for the system* (2.1) *satisfies* (2.5) *with* $\alpha = \delta_0$ *given by*

$$\delta_0 = \frac{\omega(2-\omega)}{(1+\delta_-(B))(1+\delta_+(B))}$$

and

$$\delta_-(B) = \max_{1\le i\le n} \sum_{j<i} \frac{|B_{ij}|}{\sqrt{B_{ii}B_{jj}}};\ \ \delta_+(B) = \max_{1\le i\le n} \sum_{j>i} \frac{|B_{ij}|}{\sqrt{B_{ii}B_{jj}}}. \tag{2.11}$$

REMARK 2.5. If the matrix B satisfies $B_{ii} = B_{jj}, \forall i,j = 1,\ldots,n$ (as is the case in some finite differences approximations of boundary value problems or Toeplitz circulant problems), then the constants $\delta_-(B), \delta_+(B)$ from (2.11) are equal with $\gamma_-(B), \gamma_+(B)$ from (2.10), respectively. Thus, in this case Theorem 2.4 is an extension of Theorem 2.3 for SOR relaxation.

In order to derive an SP for Kaczmarz relaxation (2.8), we shall consider a general invertible matrix A (not necessarily SPD), $b \in \mathbb{R}^n$ a given vector and the linear system

$$Ax^* = b, \tag{2.12}$$

with x^* its unique solution. For a given approximation x of x^* from (2.12), let $\bar{x}$ be the approximation of x^* obtained after a Kaczmarz step (2.8) applied to x and $e = x - x^*$, $\bar{e} = \bar{x} - x^*$ and $r = Ae$ the corresponding errors and residual (see (2.4)). The following theorem was first proved in [7]. But, we shall present in what follows another, more detailed proof required for the results described in section 2 of the paper; see also Remark 2.9 below.

THEOREM 2.6. *Using the above definitions and notation, Kaczmarz relaxation* (2.8) *for the system* (2.12) *satisfies the following smoothing property (of the type* (2.5)*)*

$$\parallel \bar{e} \parallel^2 \ \ \le \ \ \parallel e \parallel^2 - \tilde{\gamma}_0 \parallel \tilde{D}^{\frac{1}{2}} r \parallel^2, \tag{2.13}$$

where

$$\tilde{D} = \mathrm{diag}\left(\frac{1}{\parallel A_1 \parallel^2}, \ldots, \frac{1}{\parallel A_n \parallel^2}\right), \tag{2.14}$$

$$\tilde{\gamma}_0 = \frac{1}{(1+\tilde{\gamma}_-(A))(1+\tilde{\gamma}_+(A))} \tag{2.15}$$

and

$$\tilde{\gamma}_-(A) = \max_{1\le i\le n} \sum_{j<i} \frac{|\langle A_i, A_j\rangle|}{\parallel A_i \parallel^2},\ \ \tilde{\gamma}_+(A) = \max_{1\le i\le n} \sum_{j>i} \frac{|\langle A_i, A_j\rangle|}{\parallel A_i \parallel^2}. \tag{2.16}$$

ETNA
Kent State University
etna@mcs.kent.edu

Proof. Let $g_i = (0, \dots, 1, \dots, 0)^T \in I\!R^n, i = 1, \dots, n$ be the canonical basis. We first observe that, because of the symmetry of B, $Bg_i = B_i$ the expression (2.6) can be written as follows: $x^0 \in I\!R^n$ given; for $k = 0, 1, \dots$ do

$$\left\{ \begin{array}{lcl} x^{k,0} & = & x^k \\ x^{k,i} & = & x^{k,i-1} - \frac{\langle Bx^{k,i-1} - c, g_i \rangle}{\langle Bg_i, g_i \rangle} g_i, \; i = 1, \dots, n. \\ x^{k+1} & = & x^{k,n} \end{array} \right. \tag{2.17}$$

Now, if we write the system (2.12) in the form

$$(AA^T) y^* = b, \;\; x^* = A^T y^*, \tag{2.18}$$

and we apply to it Gauss-Seidel relaxation (2.17) ($B = AA^T, c = b$) with the initial approximation y^0, we get

$$\left\{ \begin{array}{lcl} y^{k,0} & = & y^k \\ y^{k,i} & = & y^{k,i-1} - \frac{\langle A^T y^{k,i-1}, A_i \rangle - b_i}{\| A_i \|^2} g_i, \; i = 1, \dots, n. \\ y^{k+1} & = & y^{k,n} \end{array} \right. \tag{2.19}$$

If we multiply from the left in (2.19) by A^T and replace the terms of the form $A^T y^{k,i}$ by $x^{k,i}$, we obtain exactly the Kaczmarz step (2.8) applied to the system (2.12). Thus, the Kaczmarz step (2.8) applied to the system (2.12) with an initial approximation of the form $x^0 = A^T y^0$, for some $y^0 \in I\!R^n$ is equivalent to the Gauss-Seidel step (2.17) applied to the system (2.18), with the initial approximation y^0 and setting $x^{k+1} = A^T y^{k+1}$. But, because the matrix AA^T is SPD, we can apply Theorem 2.3 for the Gauss-Seidel iteration and get (see (2.5) and (2.3))

$$\langle (AA^T) \bar{f}, \bar{f} \rangle \leq \langle (AA^T) f, f \rangle - \; \tilde{\gamma}_0 \langle \tilde{D} AA^T f, AA^T f \rangle \tag{2.20}$$

with $\tilde{D}$ from (2.14), $\tilde{\gamma}_0$ computed as in (2.9)–(2.10) with AA^T instead of B, and with $f, \bar{f}$ the corresponding errors with respect to the system (2.18). But, if $e, \bar{e}$ and r are the corresponding errors and residual, respectively for the Kaczmarz relaxation, we have the following relations

$$e = A^T f, \;\; \bar{e} = A^T \bar{f},$$

which when substituted into (2.20) give us (2.13) with the elements from (2.14)–(2.16) and the proof is complete. ☐

A similar result as in Theorem 2.6 can be proved for the Kaczmarz iteration with relaxation parameter (for short, ω-Kaczmarz relaxation), as described below.
ω-Kaczmarz relaxation. Let $x^0 \in I\!R^n$; for $k = 0, 1, \dots$ do

$$\left\{ \begin{array}{lcl} x^{k,0} & = & x^k \\ x^{k,i} & = & x^{k,i-1} - \omega \frac{\langle x^{k,i-1}, B_i \rangle - c_i}{\langle B_i, B_i \rangle} B_i, i = 1, \dots, n. \\ x^{k+1} & = & x^{k,n} \end{array} \right. \tag{2.21}$$

THEOREM 2.7. *Using the above definitions and notation, the ω-Kaczmarz relaxation (2.21) for the system (2.12) satisfies the following smoothing property (of the type (2.5))*

$$\| \bar{e} \|^2 \;\; \leq \;\; \| e \|^2 - \tilde{\delta}_0 \| \tilde{D}^{\frac{1}{2}} r \|^2, \tag{2.22}$$

with $\tilde{D}$ from (2.14) and

$$\tilde{\delta}_0 = \frac{\omega(2-\omega)}{(1+\tilde{\delta}_-(A))(1+\tilde{\delta}_+(A))}, \tag{2.23}$$

$$\tilde{\delta}_-(A) = \max_{1\le i\le n} \sum_{j<i} \frac{|\langle A_i, A_j\rangle|}{\parallel A_i \parallel\parallel A_j \parallel}, \quad \tilde{\delta}_+(A) = \max_{1\le i\le n} \sum_{j>i} \frac{|\langle A_i, A_j\rangle|}{\parallel A_i \parallel\parallel A_j \parallel}. \tag{2.24}$$

Proof. As in the proof of Theorem 2.6 we first show the equivalence between the ω-Kaczmarz and the SOR relaxation (2.7) written in the form (2.17); see also [10]. □

REMARK 2.8. The Kaczmarz iteration (2.8) or (2.21) is less efficient as smoother than the Gauss-Seidel one (2.6) (see for a detailed discussion Remark 4.7.2, page 128 in [21]). But, inspite of this, it has the advantage of being applicable to a much larger class of systems (than the classical square invertible ones).

REMARK 2.9. We have to observe that in (2.13) and (2.22), for the errors $\bar{e}$ and e we have the Euclidean norm, instead of the energy one from (2.5). Moreover, the proofs of Theorems 2.6 and 2.7 are based on Theorems 2.3 and 2.4, i.e., the matrix $B = AA^T$ must be SPD; thus, we cannot expect such a simple and direct extension of Theorems 2.6 and 2.7 to arbitrary noninvertible systems like (2.12), because in such a case, the matrix AA^T is no longer SPD. This extension will be proved in the next section of the paper, using a special technique.

3. Smoothing property of Kaczmarz relaxation for arbitrary consistent systems. Let A be an $m \times n$ matrix with $A_i \neq 0, \forall i = 1,\dots m$ and $b \in I\!R^m$ such that the system

$$Ax = b \tag{3.1}$$

is consistent. We shall denote by $S(A;b), N(A), R(A)$ the solutions set for (3.1), null space and range of A, respectively. For a given vector subspace $E \subset I\!R^q$, $P_E(x)$ will be the orthogonal projection onto E of an element $x \in I\!R^q$ and $E^\perp$ will denote its orthogonal complement. If x_{LS} is the (unique) minimal norm solution of (3.1) it is well known that (see, e.g., [3], [8])

$$x_{LS} \in N(A)^\perp = R(A^T), \;\; S(A;b) = x_{LS} + N(A). \tag{3.2}$$

Thus, for a vector $z \in I\!R^n$ we shall denote by $s(z)$ the solution vector (see (3.2))

$$s(z) = P_{N(A)}(z) + x_{LS} \in S(A;b). \tag{3.3}$$

The Kaczmarz relaxation for (3.1) can be written as (see (2.8)): let $x^0 \in I\!R^n$ be given; for $k = 0, 1, 2\dots$ do

$$\left\{ \begin{array}{lcl} x^{k,0} & = & x^k \\ x^{k,i} & = & x^{k,i-1} - \frac{\langle x^{k,i-1}, A_i\rangle - b_i}{\|A_i\|^2} A_i, i = 1,\dots,m. \\ x^{k+1} & = & x^{k,m} \end{array} \right. \tag{3.4}$$

The following results are known (see, e.g., [20]).

THEOREM 3.1. *For any $x^0 \in I\!R^n$, the sequence $(x^k)_{k\ge 0}$ generated by the algorithm (3.4) has the properties*

$$P_{N(A)}(x^k) = P_{N(A)}(x^0), \;\; \forall k \ge 0 \tag{3.5}$$

and

$$\lim_{k\to\infty} x^k = P_{N(A)}(x^0) + x_{LS} = s(x^0).$$

Let now $x \in I\!R^n$ be a current approximation of $s(x^0)$ (generated by (3.4), for $k = 0, x = x^0$) and $\bar{x}$ the approximation after one step of (3.4) applied to x. Let $e, \bar{e}, r$ be the corresponding errors and residual, defined by (according to (3.3) and (3.5))

$$e = x - s(x^0), \;\; \bar{e} = \bar{x} - s(x^0), \;\; r = Ae = Ax - Ax_{LS} = Ax - b.$$

Then, the generalization of Theorem 2.6 is the following.

THEOREM 3.2. *Using the above definitions and notation, Kaczmarz relaxation (3.4) for the system (3.1) satisfies the smoothing property (2.13)–(2.16).*

Proof. **Step 1.** According to (3.4), the computation of $\bar{x}$ from x can be written as

$$\left\{ \begin{array}{lcl} x^0 & = & x \\ x^i & = & x^{i-1} - \frac{\langle x^{i-1}, A_i\rangle - b_i}{\|A_i\|^2} A_i, \; i = 1, \dots, m. \\ \bar{x} & = & x^m \end{array} \right. \tag{3.6}$$

Let $\{f_1, f_2, \dots, f_m\}$ be the canonical basis in $I\!R^m$ and e^i, r^i the errors and residuals defined by

$$e^i = x^i - s(x^0), \;\; r^i = Ae^i, \; i = 1, \dots, m. \tag{3.7}$$

Then, for $i = 1, \dots, m$, we obtain (by also using the equality $A_i = A^T f_i$)

$$\begin{array}{lcl} e^i & = & x^i - s(x^0) = x^{i-1} - \frac{\langle x^{i-1}, A_i\rangle - b_i}{\|A_i\|^2} A_i - s(x^0) \\ & = & e^{i-1} - \frac{\langle x^{i-1}, A^T f_i\rangle - \langle b, f_i\rangle}{\|A_i\|^2} A_i = e^{i-1} - \frac{\langle r^{i-1}, f_i\rangle}{\|A_i\|^2} A_i \end{array} \tag{3.8}$$

and

$$r^i = Ae^i = r^{i-1} - \frac{\langle r^{i-1}, f_i\rangle}{\| A_i \|^2} AA_i. \tag{3.9}$$

From (3.8), taking the Euclidean norm and using again the equality $A_i = A^T f_i$ and (3.7) we obtain

$$\begin{array}{lcl} \| e^i \|^2 & = & \| e^{i-1} \|^2 - 2\frac{\langle r^{i-1}, f_i\rangle \langle e^{i-1}, A_i\rangle}{\|A_i\|^2} + \frac{\langle r^{i-1}, f_i\rangle^2}{\|A_i\|^2} \\ & = & \| e^{i-1} \|^2 - \frac{\langle r^{i-1}, f_i\rangle^2}{\|A_i\|^2} = \| e^{i-1} \|^2 - \frac{(r_i^{i-1})^2}{\|A_i\|^2}, \; \forall i = 1, \dots, m, \end{array} \tag{3.10}$$

where r_i^{i-1} is the i-th component of the residual vector r^{i-1}. But, because of (3.6) we have $x^0 = x, \bar{x} = x^m$, then $e^0 = e, \bar{e} = e^m$. Thus, summing up in (3.10), we get

$$\| \bar{e} \|^2 \;\; = \;\; \| e \|^2 - \sum_{i=1}^{m} \frac{(r_i^{i-1})^2}{\| A_i \|^2}. \tag{3.11}$$

Step 2. Let now $r^* = (r_1^*, \dots, r_m^*)^T \in I\!R^m$ be the *dynamic residual* (as called in [7]), defined by

$$r_i^* = r_i^{i-1}, i = 1, \dots, m, \tag{3.12}$$

i.e., r_i^* is the i-th component of the residual r^{i-1} (that is, the residual before the projection on the i-th equation of (3.6)). From (3.11) and (3.12), we then get

$$\| \bar{e} \|^2 = \| e \|^2 - \sum_{i=1}^{m} \frac{(r_i^*)^2}{\| A_i \|^2} = \| e \|^2 - \| \tilde{D}^{\frac{1}{2}} r^* \|^2, \tag{3.13}$$

with $\tilde{D}$ from (2.14). On the other hand, from the equation (3.9), we successively obtain

$$\begin{aligned} r^i &= r^{i-1} - \tfrac{\langle r^{i-1}, f_i \rangle}{\| A_i \|^2} A A^T f_i = r^{i-1} - (\hat{A}\hat{A}^T) f_i \langle r^{i-1}, f_i \rangle \\ &= r^{i-1} - (\hat{A}\hat{A}^T) f_i f_i^T r^{i-1} = r^{i-1} - B E_i r^{i-1}, \end{aligned} \tag{3.14}$$

where

$$\hat{A} = \tilde{D}^{\frac{1}{2}} A, \quad B = \hat{A}\hat{A}^T, \quad E_i = f_i f_i^T. \tag{3.15}$$

Thus,

$$\begin{cases} r^1 & = r^0 - BE_1 r^0 & = r - BE_1 r, \\ r^2 & = r^1 - BE_2 r^1 & = r - (BE_1 r + BE_2 r^1), \\ \ldots & \ldots & \ldots \\ r^{m-1} & = r^{m-2} - BE_{m-1} r^{m-2} & = r - (BE_1 r + BE_2 r^1 + \ldots + BE_{m-1} r^{m-2}). \end{cases}$$

Then, from the definition of the matrices E_i in (3.15) and (3.12), we get $r_1^* = r_1^0 = r_1$ and for $i = 2, \ldots, m$,

$$r_i^* = r_i - (B_{i1} r_1^0 + B_{i2} r_2^1 + \ldots + B_{i,i-1} r_i^{i-1})$$

or in matrix form

$$r^* = \begin{bmatrix} r_1^0 \\ r_2^1 \\ \cdot \\ \cdot \\ \cdot \\ r_m^{m-1} \end{bmatrix} = r - \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ -B_{21} & 0 & 0 & \ldots & 0 & 0 \\ -B_{31} & -B_{32} & 0 & \ldots & 0 & 0 \\ .. & .. & .. & .. & .. & .. \\ -B_{m,1} & -B_{m,2} & -B_{m,3} & \ldots & -B_{m,m-1} & 0 \end{bmatrix} \begin{bmatrix} r_1^0 \\ r_2^1 \\ \cdot \\ \cdot \\ \cdot \\ r_m^{m-1} \end{bmatrix} \tag{3.16}$$

$$= r - L r^*,$$

or (see also (3.13))

$$\tilde{D}^{\frac{1}{2}} r = \left(I + \tilde{D}^{\frac{1}{2}} L \tilde{D}^{-\frac{1}{2}} \right) \left(\tilde{D}^{\frac{1}{2}} r^* \right), \tag{3.17}$$

where L is the corresponding strictly lower triangular matrix from (3.16).

Step 3. If C is a square matrix and $\| C \|_2, \| C \|_\infty$ are, respectively its spectral and infinity norm (see, e.g., [3]), we have

$$\| C \|_2^2 \ = \ \rho(C^T C) \ \leq \ \| C^T C \|_\infty \ \leq \ \| C^T \|_\infty \| C \|_\infty \,. \tag{3.18}$$

Using (3.18) and taking the Euclidean norm in (3.17), we obtain

$$\begin{aligned} \| \tilde{D}^{\frac{1}{2}} r \|^2 \ &\leq \ \| I + \tilde{D}^{\frac{1}{2}} L \tilde{D}^{-\frac{1}{2}} \|_\infty \| I + \tilde{D}^{-\frac{1}{2}} L^T \tilde{D}^{\frac{1}{2}} \|_\infty \| \tilde{D}^{\frac{1}{2}} r^* \|^2 \\ &= \left(1 + \| \tilde{D}^{\frac{1}{2}} L \tilde{D}^{-\frac{1}{2}} \|_\infty \right) \left(1 + \| \tilde{D}^{-\frac{1}{2}} L^T \tilde{D}^{\frac{1}{2}} \|_\infty \right) \| \tilde{D}^{\frac{1}{2}} r^* \|^2 \,. \end{aligned} \tag{3.19}$$

ETNA
Kent State University
etna@mcs.kent.edu

But, a simple computation gives us

$$1+ \| \tilde{D}^{\frac{1}{2}} L \tilde{D}^{-\frac{1}{2}} \|_\infty \; = \; \tilde{\gamma}_-(A), \;\; 1+ \| \tilde{D}^{-\frac{1}{2}} L^T \tilde{D}^{\frac{1}{2}} \|_\infty \; = \; \tilde{\gamma}_+(A), \tag{3.20}$$

with $\tilde{\gamma}_-(A), \tilde{\gamma}_+(A)$ from (2.16). Then, using (3.13), (3.19) and (3.20), we get (2.13) and the proof is complete. ☐

A similar result can be proved for ω-Kaczmarz relaxation.

THEOREM 3.3. *Using the above definitions and notation, the ω-Kaczmarz relaxation for the system (3.1) satisfies the smoothing property (2.22)–(2.24).*

Proof. The ω-Kaczmarz relaxation (2.21) for the system (3.1) can be written as

$$\begin{cases} x^0 & = & x \\ x^i & = & x^{i-1} - \omega \frac{\langle x^{i-1}, A_i \rangle - b_i}{\|A_i\|^2} A_i, \; i = 1, \ldots, m. \\ \bar{x} & = & x^m \end{cases}$$

As in the proof of Theorem 3.2, we get

$$e^i = e^{i-1} - \omega(2-\omega) \frac{\langle r^{i-1}, f_i \rangle}{\| A_i \|^2} A_i$$

and

$$r^i = A e^i = r^{i-1} - \omega(2-\omega) \frac{\langle r^{i-1}, f_i \rangle}{\| A_i \|^2} A A_i.$$

Then we proceed in exactly the same way as in the above mentioned proof. ☐

4. Smoothing property of Kaczmarz Extended relaxation for arbitrary inconsistent systems. Let A and b be as in section 2. Instead of the consistent system (3.1), we shall consider in this section the linear least squares formulation (inconsistent system)

$$\| Ax - b \| = \min!, \tag{4.1}$$

for which we shall denote by $LSS(A; b)$ and x_{LS} the set of its solutions and the minimal norm one, respectively. Let b_A, b_A^* be defined by

$$b_A = P_{R(A)}(b), \;\; b_A^* = P_{N(A^T)}(b).$$

Then (see, e.g., [3])

$$b = b_A + b_A^*, \; LSS(A; b) = S(A; b_A) = x_{LS} + N(A). \tag{4.2}$$

Moreover, (see (3.3)) for any $z \in I\!R^n$, we have

$$s(z) = P_{N(A)}(z) + x_{LS} \in LSS(A; b) = S(A; b_A).$$

Thus we define (as in section 2) the error and residual by

$$e = e(z) = z - s(z), \;\; r = Ae = Az - b_A. \tag{4.3}$$

The Kaczmarz Extended algorithm for (4.1) (KE, for short), introduced in [12] (see also [13]) is the following.

Kaczmarz Extended. Let $x^0 \in R^n, y^0 = b$; for $k = 0, 1, \ldots$ do

$$\begin{cases} y^{k+1} & = & \Phi(y^k) = (\phi_1 \cdot \ldots \cdot \phi_n)(y^k), \\ b^{k+1} & = & b - y^{k+1}, \\ x^{k+1} & = & \text{Kaczmarz}(b^{k+1}; x^k), \end{cases} \tag{4.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

where

$$\phi_j(y) = y - \frac{\langle y, A^j \rangle}{\| A^j \|^2} A^j$$

and $A^j \neq 0, j = 1, \ldots, n$, are the columns of A. In [14] we proved that the sequence $(x^k)_{k \geq 0}$ generated with the above KE algorithm satisfies the relation (3.5). Then, if $x = x^k$ (for some $k \geq 0$) is a current approximation, we shall define the error (see (4.3)) by

$$e = x - s(x^0).$$

Let $y = y^k$ be the corresponding element from the first step of (4.4) and $\bar{y} = y^{k+1}$, i.e.,

$$\bar{y} = \Phi(y) = P_{N(A^T)}(y) + \tilde{\Phi}(y), \tag{4.5}$$

where (see for details [13])

$$\tilde{\Phi}(y) = \Phi P_{R(A)}(y) \in R(A). \tag{4.6}$$

From (4.5) and (4.6), we get that

$$P_{N(A^T)}(\bar{y}) = P_{N(A^T)}(y).$$

Thus (see also (4.2))

$$P_{N(A^T)}(\bar{y}) = P_{N(A^T)}(b) = b_A^*. \tag{4.7}$$

Then, if $\bar{b} = b^{k+1}$ (see (4.4)), from (4.5) and (4.7), we obtain

$$\bar{b} = b - \bar{y} = b_A - \tilde{y}, \tag{4.8}$$

where

$$\tilde{y} = \tilde{\Phi}(y). \tag{4.9}$$

REMARK 4.1. Returning to the original notation, from the above equalities (4.7)–(4.9), we obtain

$$\tilde{y} = \tilde{\Phi}(y) = \tilde{\Phi}^k(b).$$

Moreover, in [20] it is proved that $\| \tilde{\Phi} \|_2 < 1$. Thus

$$\lim_{k \to \infty} \tilde{\Phi}^k(b) = 0.$$

The extension of Theorem 3.2 to the problem (4.1), for the KE algorithm (4.4) is the following.

THEOREM 4.2. *Using the above definitions and notation, the KE relaxation (4.4) for the (inconsistent) problem (4.1) satisfies the following smoothing property*

$$\| \bar{e} \|^2 \leq \| e \|^2 - \frac{\tilde{\gamma}_0}{2} \| \tilde{D}^{\frac{1}{2}} r \|^2 + 2 \| \tilde{D}^{\frac{1}{2}} \tilde{y} \|^2, \tag{4.10}$$

with $\tilde{D}$ and $\tilde{\gamma}_0$ from (2.14)–(2.16).

ETNA
Kent State University
etna@mcs.kent.edu

Proof. **Step 1.** The third step in (4.4) can be written as (see also (2.8))

$$\begin{cases} x^0 &= x \\ x^i &= x^{i-1} - \frac{\langle x^{i-1}, A_i\rangle - \bar{b}_i}{\|A_i\|^2} A_i,\ i = 1, \dots, m. \\ \bar{x} &= x^m \end{cases}$$

Then, by defining the errors (see also (3.7))

$$e^i = x^i - s(x^0),\ i = 1, \dots, m,\ \ \bar{e} = e^m, \tag{4.11}$$

we get

$$\begin{aligned} e^i &= x^i - s(x^0) = x^{i-1} - \frac{\langle x^{i-1}, A_i\rangle - (b_A)_i + \tilde{y}_i}{\|A_i\|^2} A_i - s(x^0) \\ &= e^{i-1} - \frac{\langle Ax^{i-1} - b_A, f_i\rangle + \tilde{y}_i}{\|A_i\|^2} A_i = e^{i-1} - \frac{r_i^{i-1}}{\|A_i\|^2} A_i - \frac{\tilde{y}_i}{\|A_i\|^2} A_i. \end{aligned} \tag{4.12}$$

From (4.12) we obtain (by also using the relation $A_i = A^T f_i$)

$$\begin{aligned} \| e^i \|^2 &= \| e^{i-1} - \frac{r_i^{i-1}}{\|A_i\|^2} A_i \|^2 + \frac{|\tilde{y}_i|^2}{\|A_i\|^2} A_i - 2\langle e^{i-1} - \frac{r_i^{i-1}}{\|A_i\|^2} A_i, \frac{\tilde{y}_i}{\|A_i\|^2} A_i\rangle \\ &= \| e^{i-1} \|^2 - 2\langle e^{i-1}, \frac{r_i^{i-1}}{\|A_i\|^2} A_i\rangle + \frac{(r_i^{i-1})^2}{\|A_i\|^2} + \frac{(\tilde{y}_i)^2}{\|A_i\|^2} - 2\langle e^{i-1}, \frac{\tilde{y}_i}{\|A_i\|^2} A_i\rangle \\ &\quad + 2\langle \frac{r_i^{i-1}}{\|A_i\|^2} A_i, \frac{\tilde{y}_i}{\|A_i\|^2} A_i\rangle = \| e^{i-1} \|^2 - \frac{(r_i^{i-1})^2}{\|A_i\|^2} + \frac{(\tilde{y}_i)^2}{\|A_i\|^2},\ \ \forall i = 1, \dots, m. \end{aligned} \tag{4.13}$$

Then by summing in (4.13), using (4.11) and the notation from section 2, we get

$$\| \bar{e} \|^2 \ = \ \| e \|^2 - \| \tilde{D}^{\frac{1}{2}} r^* \|^2 + \| \tilde{D}^{\frac{1}{2}} \tilde{y} \|^2 . \tag{4.14}$$

Step 2. From (4.12) we obtain (see again the notation in section 2)

$$\begin{aligned} r^i &= Ae^i = Ae^{i-1} - \frac{r_i^{i-1}}{\|A_i\|^2} AA_i - \frac{\tilde{y}_i}{\|A_i\|^2} AA_i \\ &= r^{i-1} - BE_i r^{i-1} - BE_i \tilde{y},\ \ i = 1, \dots, m, \end{aligned}$$

thus

$$\begin{cases} r^1 &= r - BE_1 r - BE_1 \tilde{y}, \\ r^2 &= r - (BE_1 r + BE_2 r^1) - (BE_1 + BE_2)\tilde{y}, \\ \dots & \dots \\ r^{m-1} &= r - (BE_1 r + BE_2 r^1 + \dots + BE_{m-1} r^{m-2}) \\ & \quad -(BE_1 + BE_2 + \dots + BE_{m-1})\tilde{y}. \end{cases}$$

Using the equality (see section 2)

$$r^* = (r_1, r_2^1, \dots, r_m^{m-1})^T,$$

we obtain

$$\begin{cases} r_1^* &= r_1, \\ r_2^* &= r_2 - B_{21} r_1^* - B_{21}\tilde{y}_1, \\ r_3^* &= r_3 - (B_{31} r_1^* + B_{32} r_2^*) - (B_{31}\tilde{y}_1 + B_{32}\tilde{y}_2), \\ \dots & \dots \\ r_m^* &= r_m - (B_{m1} r_1^* + B_{m2} r_2^* + \dots + B_{m,m-1} r_{m-1}^*) \\ & \quad -(B_{m1}\tilde{y}_1 + B_{m2}\tilde{y}_2 + \dots + B_{m,m-1}\tilde{y}_{m-1}). \end{cases} \tag{4.15}$$

Writing (4.15) in matrix form

$$\begin{bmatrix} r_1^* \\ r_2^* \\ \cdot \\ \cdot \\ \cdot \\ r_m^* \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_m \end{bmatrix} - L \begin{bmatrix} r_1^* \\ r_2^* \\ \cdot \\ \cdot \\ \cdot \\ r_m^* \end{bmatrix} - L \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}_m \end{bmatrix}$$

with L from (3.16), we get

$$r^* = r - Lr^* - L\tilde{y}. \tag{4.16}$$

From (4.16) we obtain as in section 2

$$(I+L)r^* + L\tilde{y} = r \Leftrightarrow (I+\tilde{L})(\tilde{D}^{\frac{1}{2}} r^*) + \tilde{L}(\tilde{D}^{\frac{1}{2}}\tilde{y}) = \tilde{D}^{\frac{1}{2}} r, \tag{4.17}$$

where

$$\tilde{L} = \tilde{D}^{\frac{1}{2}} L \tilde{D}^{-\frac{1}{2}}. \tag{4.18}$$

Using (4.17) and (4.18), it follows that

$$\begin{array}{rcl} & \parallel \tilde{D}^{\frac{1}{2}} r \parallel^2 \;\leq & \parallel (I+\tilde{L})(\tilde{D}^{\frac{1}{2}} r^*) + \tilde{L}(\tilde{D}^{\frac{1}{2}}\tilde{y}) \parallel^2 \\ \leq & 2 & \left[\parallel I+\tilde{L} \parallel_2^2 \ \parallel \tilde{D}^{\frac{1}{2}} r^* \parallel^2 + \parallel \tilde{L} \parallel_2^2 \ \parallel \tilde{D}^{\frac{1}{2}} \tilde{y} \parallel^2 \right] \\ \leq & 2 & \left[(1+\parallel \tilde{L} \parallel_\infty)(1+\parallel \tilde{L}^T \parallel_\infty) \parallel \tilde{D}^{\frac{1}{2}} r^* \parallel + \parallel \tilde{L} \parallel_\infty \parallel \tilde{L}^T \parallel_\infty \parallel \tilde{D}^{\frac{1}{2}} \tilde{y} \parallel^2 \right] \\ \leq & 2 & \left[(1+\parallel \tilde{L} \parallel_\infty)(1+\parallel \tilde{L}^T \parallel_\infty) \right] \left(\parallel \tilde{D}^{\frac{1}{2}} r^* \parallel^2 + \parallel \tilde{D}^{\frac{1}{2}} \tilde{y} \parallel^2 \right). \end{array} \tag{4.19}$$

From (4.19) and (2.15), we then obtain

$$\parallel \tilde{D}^{\frac{1}{2}} r^* \parallel^2 \;\geq\; \frac{\tilde{\gamma}_0}{2} \parallel \tilde{D}^{\frac{1}{2}} r \parallel^2 - \parallel \tilde{D}^{\frac{1}{2}} \tilde{y} \parallel^2,$$

which together with (4.14) gives us (4.10) and the proof is complete. □

A similar result can be derived for the Kaczmarz Extended algorithm with Relaxation Parameters (KERP, for short), introduced in [13].

KERP. Let $x^0 \in R^n, y^0 = b$; for $k = 0, 1, \ldots$ do

$$\left\{ \begin{array}{rcl} y^{k+1} & = & \Phi(\alpha; y^k) = (\phi_1 \cdot \ldots \cdot \phi_n)(\alpha; y^k), \\ b^{k+1} & = & b - y^{k+1}, \\ x^{k+1} & = & \omega - \text{Kaczmarz}(b^{k+1}; x^k), \end{array} \right. \tag{4.20}$$

where

$$\phi_j(\alpha; y) = y - \alpha \frac{\langle y, A^j \rangle}{\parallel A^j \parallel^2} A^j.$$

THEOREM 4.3. *Using the above definitions and notation, KERP relaxation (4.20) for the (inconsistent) problem (4.1) satisfies the following smoothing property*

$$\parallel \bar{e} \parallel^2 \;\leq\; \parallel e \parallel^2 - \frac{\tilde{\delta}_0}{2} \parallel \tilde{D}^{\frac{1}{2}} r \parallel^2 + 2 \parallel \tilde{D}^{\frac{1}{2}} \tilde{y}(\alpha) \parallel^2,$$

with $\tilde{D}$ *and* $\tilde{\delta}_0$ *from* (2.14), (2.23), *and* $\tilde{y}(\alpha) = \Phi(\alpha; y)$.

Proof. Apply Theorems 4.3 and 2.7 to ω-Kaczmarz relaxation. ∎

REMARK 4.4. The result from Theorem 3.2 is not a particular case of Theorem 4.2. Indeed, if $b \in R(A)$ for the problem (4.1), we have $b_A^* = 0$. Thus, $\bar{y} = \tilde{y} \in R(A)$, but we cannot drop the term $2 \parallel \tilde{D}^{\frac{1}{2}} \tilde{y} \parallel^2$ in (4.10). Thus, in the consistent case Theorems 3.2 and 4.2 provide two smoothing properties for two different algorithms: Kaczmarz and Kaczmarz Extended.

5. Final comments and further developments. In this paper we formulated and proved AMG smoothing properties for the classical Kaczmarz and Kaczmarz Extended algorithms in the general case of arbitrary rectangular systems, either consistent or in least squares formulation. In the consistent case, our result for classical Kaczmarz relaxation (Theorem 3.2) generalizes the well-known result for square invertible matrices (Theorem 2.6). The general character of the matrices involved in the above theory makes extension to AMG to general least squares problems possible. Some steps in this direction have been taken in [15] and [9], but both the theoretical development and experiments are in initial phases.

Acknowledgement. The author thanks the two referees for their comments and suggestions which improved some important parts of the original version of the paper.

REFERENCES

[1] A. ARICÓ, M. DONATELLI, AND S. SERRA-CAPIZZANO, *V-cycle optimal convergence for certain (multilevel) structured linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 186-214.
[2] A. ARICÓ AND M. DONATELLI, *V-cycle multigrid for multilevel algebras: proof of optimality*, Numer. Math., 105 (2007), pp. 511-547.
[3] A. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.
[4] A. BRANDT, S. MCCORMICK, AND J. RUGE, *Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations*, Technical Report, Institute for Computational Studies, Colorado State University, 1982.
[5] A. BRANDT, *Algebraic multigrid theory: the symmetric case (a preliminary version)*, in Preliminary Proceedings for the International Multigrid Conference (Copper Mountain, Colorado), April, 1983.
[6] A. BRANDT, S. MCCORMICK, AND J. RUGE, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and Its Applications, D.J. Evans, ed., Cambridge University Press, Cambridge, 1984.
[7] A. BRANDT, *Algebraic multigrid theory: the symmetric case*, Appl. Math. Comp., 19 (1986), pp. 23–56.
[8] Y. CENSOR AND A. Z. STAVROS, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford Univ. Press, New York, 1997.
[9] H. KÖSTLER, C. POPA, M. PRÜMMER, AND U. RÜDE, *Towards an algebraic multigrid method for tomographic image reconstruction - improving convergence of ART*, Paper 476 in Electronic Proceedings of the European Conference on Computational Fluid Dynamics ECCOMAS CFD 2006, The Netherlands, 2006.
[10] F. NATTERER, *The Mathematics of Computerized Tomography*, John Wiley and Sons, New York, 1986.
[11] C. POPA, *On smoothing property of the SOR relaxation*, Studii si Cercetari Matematice, 41 (1989), pp. 399–406.
[12] C. POPA, *Least-squares solution of overdetermined inconsistent linear systems using Kaczmarz's relaxation*, Internat. J. Comput. Math., 55 (1995), pp. 79–89.
[13] C. POPA, *Extensions of block-projections methods with relaxation parameters to inconsistent and rank-defficient least-squares problems*, BIT, 38 (1998), pp. 151–176.
[14] C. POPA, *Characterization of the solutions set of inconsistent least-squares problems by an extended Kaczmarz algorithm*, Korean J. Comp. Appl. Math., 6 (1999), pp. 51–64.
[15] M. PRÜMMER, H. KOSTLER, J. HORNEGGER, AND U. RÜDE, *A full multigrid technique to accelerate an ART scheme for tomographic image reconstruction*, in Proceedings of ASIM Conference, September 12-15 2005, SCS Publishing House, Erlangen, Germany, pp. 632–637.
[16] J. RUGE, *Algebraic multigrid (AMG) for geodetic survey problems*, in Preliminary Proceedings for the International Multigrid Conference (Copper Mountain, Colorado), April, 1983, 24 pages.
[17] J. RUGE AND K. STÜBEN, *Algebraic Multigrid (AMG)*, Arbeitspapiere der GMD Bonn 210, GMD, Bonn, Germany, 1986.

[18] S. SERRA-CAPIZZANO, *Convergence analysis of two-grid methods for elliptic Toeplitz and PDEs matrix-sequences*, Numer. Math., 92 (2002), pp. 433–465.
[19] K. STÜBEN, *Algebraic multigrid (AMG): experiences and comparisons*, Appl. Math. Comput. 13 (1983), pp. 419–451.
[20] K. TANABE, *Projection method for solving a singular system of linear equations and its applications*, Numer. Math., 17 (1971), pp. 203–214.
[21] U. TROTTENBERG, C. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, London, 2001.
[22] D. M. YOUNG, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.