FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
INSTITUT FÜR INFORMATIK (MATHEMATISCHE MASCHINEN UND DATENVERARBEITUNG)

Lehrstuhl für Informatik 10 (Systemsimulation)

Algebraic multigrid for general inconsistent linear systems: The correction step

Harald Köstler, Constantin Popa, Silke Bergler, Ulrich Rüde

Lehrstuhlbericht 06-4

Algebraic multigrid for general inconsistent linear systems: The correction step

Harald Köstler*, Constantin Popa†, Silke Bergler ‡, Ulrich Rüde§

May 10, 2006

Abstract

This paper is a continuation of the previous Technical Report [2]. We are now focusing on the properties of the correction step of our AMG approach. We included also numerical experiments for the 2D case to support our theoretical results.

2000 MS Classification: 65F10, 65F20, 65N55

Key words and phrases: algebraic multigrid, inconsistent least squares problems, correction step

1 Problem formulation

Let A be an $m \times n$ matrix and $b \in I\!R^m$. We shall denote by $(A)_i, (A)^j, (A)_{ij}, A^T, N(A), R(A), A^+$ the i-th row, j-th column, (i,j)-th element, transpose, null space, range and Moore-Penrose pseudoinverse 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, with respect to the euclidean scalar product and norm, denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. The following properties are known (see e.g. [1])

$$AA^+A = A,\ A^+AA^+ = A^+,\ (AA^+)^T = AA^+,\ (A^+A)^T = A^+A,\ (A^T)^+ = (A^+)^T, \tag{1}$$

$$P_{R(A)} = AA^+,\ P_{R(A^T)} = A^+A,\ P_{N(A^T)} = I - AA^+,\ P_{N(A)} = I - A^+A, \tag{2}$$

where I are the corresponding unit matrices and

$$N(A^+) = N(A^T),\quad R(A^+) = R(A^T). \tag{3}$$

We consider the linear least squares problem: find $x \in I\!R^n$ such that

$$\| Ax - b \| = \min! \tag{4}$$

and denote by $LSS(A; b), x_{LS}$ its solutions set and the (unique) minimal norm one. It is also well known that (see e.g. [1])

$$x_{LS} = A^+b \in N(A)^\perp = R(A^T),\quad LSS(A; b) = x_{LS} + N(A). \tag{5}$$

Thus, for a vector $z \in I\!R^n$ we shall denote by $s(z)$ the solution vector (see (5))

$$s(z) = P_{N(A)}(z) + x_{LS} \in LSS(A; b). \tag{6}$$

*FAU Erlangen-Nürnberg, IMMD10, Germany; e-mail: harald.koestler@informatik.uni-erlangen.de

†"Ovidius" University of Constanta, Faculty of Mathematics and Computer Science, Romania; e-mail: cpopa@univ-ovidius.ro. For this author the paper was supported by the CNMP CEEX Grant 05-D11-25/2005 and by the DAAD via a grant as a visiting professor at the Friedrich-Alexander Universität Erlangen- Nürnberg, Germany, in the period December 2005 - January 2006.

‡FAU Erlangen-Nürnberg, Germany; e-mail: Silke.Bergler@gmx.de

§FAU Erlangen-Nürnberg, IMMD10, Germany; e-mail: ulrich.ruede@informatik.uni-erlangen.de

In the previous report [2] we proved the Algebraic Multigrid (AMG) smoothing property for Kaczmarz and Kaczmarz Extended algorithms applied to the problem (4). In the present one we shall concentrate on the properties of the AMG correction step. Thus, let $p < n$ be a fixed integer, A_p $(m \times p)$ - the coarse grid matrix and I_p^n $(n \times p)$ - the interpolation operator (which we suppose to be full column rank). In [2] we considered the following form of the correction process for the problem (4).

$$\begin{cases} d = b - Ax \\ \parallel A_p v_p - d \parallel = \min! \Rightarrow \quad v_p = A_p^+ d = A_p^+ \left(P_{R(A_p)}(d) \right) \\ \bar{x} = x + I_p^n v_p \end{cases} \tag{7}$$

With respect to the above mentioned scope our paper is organized as follows. In section 2 we analyze the properties of the correction step (7). In section 3 we consider a special construction for the elements of the correction step, according to the paper [3] and then present some numerical experiments for the 2D case.

2 General properties of the correction step

Let

$$b_A = P_{R(A)}(b), \quad b_A^* = P_{N(A^T)}(b). \tag{8}$$

We know that (see e.g. [1])

$$x \in LSS(A; b) \Leftrightarrow Ax = b_A. \tag{9}$$

From (8) we obtain that the correction vector v_p in (7) satisfies

$$A_p v_p = P_{R(A_p)}(d). \tag{10}$$

We shall now introduce our main assumption on the elements of the correction step (7).

Assumption 1. *The matrices A, A_p and I_p^n satisfy the equality*

$$A_p = A I_p^n. \tag{11}$$

Based on this assumption we shall derive in what follows some properties of the correction process (7). For this, let first denote by $x, \bar{x}$ be the approximation before and after the correction step, respectively and define the corresponding errors (see (6)) by

$$e = x - s(x), \quad \bar{e} = \bar{x} - s(\bar{x}). \tag{12}$$

Proposition 1 *If $x \in LSS(A; b)$, then $\bar{x} = x$.*

Proof. From our hypothesis about x and by using (3), (9) and (11) it results

$$d = b - Ax = b - b_A = b_A^* \in N(A^T) \tag{13}$$

and

$$N(A_p^+) = N(A_p^T) = N((I_p^n)^T A^T) \supset N(A^T). \tag{14}$$

From (13) and (14) we obtain that $d = b_A^* \in N(A_p^+)$, thus

$$v_p = A_p^+ d = 0,$$

i.e. $\bar{x} = x$ and the proof is complete.

Proposition 2 *The correction process (7) is idempotent.*

Proof. Let $\bar{x}'$ the vector obtained after we apply one more the correction step (7) to $\bar{x}$. We have (by also using (10) and (11))

$$\bar{d} = b - A\bar{x} = b - Ax - AI_p^n v_p =$$
$$d - A_p v_p = d - P_{R(A_p)}(d) = P_{N(A_p^T)}(d), \tag{15}$$

thus, following (7)

$$\bar{v}_p = A_p^+ \bar{d} = A_p^+ P_{N(A_p^T)}(d) = 0,$$

i.e. $\bar{x}' = \bar{x} + I_p^n \bar{v}_p = \bar{x}$ and the proof is complete.

Proposition 3 *Let $r, \bar{r}$ be the residuals before and after the correction step (7), i.e. (see (12))*

$$r = Ae = A(x - s(x)) = Ax - b_A, \quad \bar{r} = A\bar{e} = A(\bar{x} - s(\bar{x})) = A\bar{x} - b_A. \tag{16}$$

Then

$$A_p^T \bar{r} = 0, \tag{17}$$

$$\| \bar{r} \| \leq \| r \| . \tag{18}$$

Proof. By using (16), (7), (10) and (11) we successively obtain

$$\bar{r} = A\bar{x} - b_A = Ax + AI_p^n v_p - b_A = r + A_p v_p =$$

$$r + P_{R(A_p)}(d) = r + P_{R(A_p)}(-r + b_A^*) =$$

$$r - P_{R(A_p)}(r) + P_{R(A_p)}(b_A^*) = P_{N(A_p^T)}(r) + P_{R(A_p)}(b_A^*). \tag{19}$$

But, $b_A^* \in N(A_p^T) = R(A_p)^\perp$, thus $P_{R(A_p)}(b_A^*) = 0$, then from (19) it results

$$\bar{r} = P_{N(A_p^T)}(r) \Rightarrow A_p^T \bar{r} = 0.$$

For the inequality (18) we have

$$\| \bar{r} \|^2 = \| P_{N(A_p^T)}(r) \|^2 \leq \| P_{N(A_p^T)}(r) \|^2 + \| P_{R(A_p)}(r) \|^2 =$$

$$\| P_{N(A_p^T)}(r) + P_{R(A_p)}(r) \|^2 = \| r \|^2 .$$

A special case for the errors in (12) is when we refer to the minimal norm solution of (4), i.e.

$$e' = x - x_{LS}, \quad \bar{e}' = \bar{x} - x_{LS}. \tag{20}$$

Proposition 4 *The following equalities hold.*

$$e' = -A^+ d + P_{N(A)}(x) = A^+ r + P_{N(A)}(x),$$

$$\| e' \|^2 = \| A^+ d \|^2 + \| P_{N(A)}(x) \|^2 = \| A^+ r \|^2 + \| P_{N(A)}(x) \|^2, \tag{21}$$

$$\bar{e}' = -A^+ \bar{d} + P_{N(A)}(\bar{x}) = A^+ \bar{r} + P_{N(A)}(\bar{x}),$$

$$\| \bar{e}' \|^2 = \| A^+ \bar{d} \|^2 + \| P_{N(A)}(\bar{x}) \|^2 = \| A^+ \bar{r} \|^2 + \| P_{N(A)}(\bar{x}) \|^2, \tag{22}$$

where $d, \bar{d}, r, \bar{r}$ are the corresponding defects and residuals.

Proof. It is enough to prove the corresponding relations for e'. From (10) and (8) we have $d = b - Ax = -r + b_A^*$, thus (by also using (2))

$$A^+ d = -A^+ r = A^+ b - A^+ A x = x_{LS} - P_{R(A^T)}(x) = x_{LS} - x + P_{N(A)}(x) = -e + P_{N(A)}(x) \tag{23}$$

from which the first two equalities in (21) hold. For the last two, we just take norms in (23) and use the fact that the vectors $A^+ d$ and $P_{N(A)}(x)$ are orthogonal. This completes the proof.
Beside the above described properties of the coarse grid correction step (7) the following one is compulsory for the convergence analysis of a two grid AMG.

$$P_{N(A)}(\bar{x}) = P_{N(A)}(x). \tag{24}$$

It ensures the fact that, after the correction step, the new approximation $\bar{x}$, generates an error $\bar{e}$ with respect to the same LSS solution. Else, at each application of (7) the solution according to which the error is computed (see (12)) would be changed. In what follows we shall give three sufficient assumptions for that (24) holds.
Assumption 2.1 *The matrices A, A_p, I_p^n satisfy the following relation*

$$(A^+ A) I_p^n = I_p^n (A_p^+ A_p). \tag{25}$$

Assumption 2.2 *The matrices A, A_p, I_p^n satisfy the following relation*

$$A^+ A_p A_p^+ A = I_p^n A_p^+ A. \tag{26}$$

Assumption 2.1 *The interpolation I_p^n satisfies the following relation*

$$R(I_p^n) = R(A^T). \tag{27}$$

Proposition 5 *Each of the above assumptions ensures the property* (24).

Proof. For Assumption 2.2 From (7) we have $\bar{x} = x + w$, with $w = I_p^n v_p$, thus (24) is equivalent with $P_{N(A)}(w) = 0$. Then (by also using (2), (10), (7)) we have the following sequence of relations.

$$P_{N(A)}(w) = 0 \Leftrightarrow w \in R(A^T) \Leftrightarrow P_{R(A^T)}(w) = w \Leftrightarrow A^+ A w = w \Leftrightarrow$$

$$A^+ A I_p^n v_p = w \Leftrightarrow A^+ A_p v_p = w \Leftrightarrow A^+ P_{R(A_p)}(d) = I_p^n A_p^+ P_{R(A_p)}(d) \Leftrightarrow$$

$$A^+ P_{R(A_p)}(r) = I_p^n A_p^+ P_{R(A_p)}(r) \Leftrightarrow A^+ A_p A_p^+ A z = I_p^n A_p^+ A z, \tag{28}$$

for some $z \in I\!R^n$. But, (28) holds from the (stronger) assumption (26).
For Assumption 2.1 As before we obtain that (24) is equivalent with

$$A^+ P_{R(A_p)}(d) = I_p^n A_p^+ P_{R(A_p)}(d) \Leftrightarrow A^+ A_p u = I_p^n A_p^+ A_p u,$$

for some $u \in I\!R^p$, which is ensured by the stronger condition from (25).
For Assumption 2.3 It directly results from the relation $\bar{x} = x + I_p^n v_p$ and the fact that the subspaces $N(A)$ and $R(A^T)$ are orthogonal.

Proposition 6 *(i) If the matrix A has full column rank, then* (25) *and* (26) *are true.*
(ii) If the interpolation operator is of the form

$$I_p^n = A^T E, \tag{29}$$

for some $m \times p$ matrix E, then (27) *is true.*

Proof. From our hypothesis it results that (see e.g. [1])

$$A^+ = (A^T A)^{-1} A^T, \quad A^+ A = I_n, \tag{30}$$

where I_n is the $n \times n$ unit matrix. Thus,

$$A^+ A_p A_p^+ A = A^+ A I_p^n A_p^+ A = I_p^n A_p^+ A,$$

i.e. relation (26). Now, by using the fact that (26) is true we successively get

$$I_p^n A_p^+ A_p = (I_p^n A_p^+ A) I_p^n = A^+ A_p A_p^+ A_p = A^+ A_p = (A^+ A) I_p^n$$

and the proof of (i) is complete. For (ii) the proof is obvious.

Corollary 1 *If* (25) *holds, then the errors e and $\bar{e}$ from* (12) *satisfy*

$$\bar{e} = e + I_p^n v_p. \tag{31}$$

Proof. It results by directly combining (12), (6) and (24).

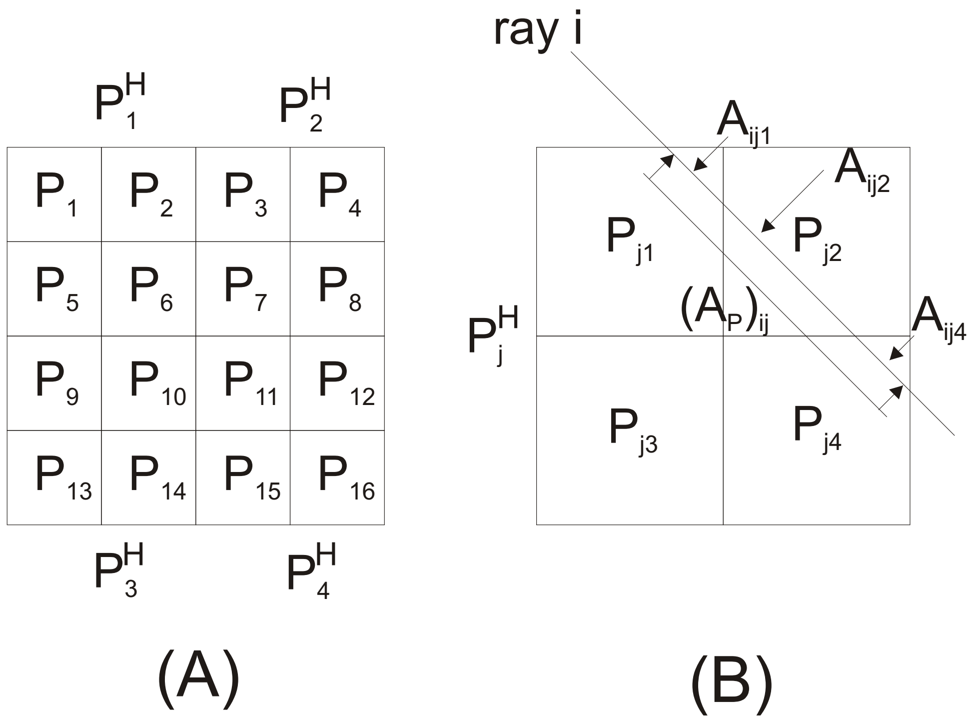

Figure 1: Relation between fine and coarse grid matrix entries.

3 An example from image reconstruction in computerized tomography

3.1 Construction of coarse grid matrix

We consider the approach in [3]. Let A be the $m \times n$ matrix obtained by scanning the given image (we shall consider only the 2D case here). We shall suppose that

$$n = 4p \tag{32}$$

and let $P_1, \dots, P_n$ be the pixels on the "fine grid". The "coarse grid" is obtained by considering the "bigger" pixels formed (each) by 4 adjacent ones from the fine grid, $P_1^H, \dots, P_p^H$ (see Figure 1 (A)).

For any $j \in \{1, \dots, p\}$ we define $S(j)$ as the set of indices of fine grid pixels that form the coarse grid one P_j^H, i.e.

$$S(j) = \{j_1, j_2, j_3, j_4\}, \quad \forall j = 1, \dots, p. \tag{33}$$

such that

$$P_j^H = P_{j_1} \cup P_{j_2} \cup P_{j_3} \cup P_{j_4}. \tag{34}$$

We may suppose that the pixels P_i and P_j^H are numbered such that (e.g. as in Figure 1 (A))

$$j_1 < j_2 < j_3 < j_4. \tag{35}$$

We construct the above coarse grid matrix A_p following the formulas (see Figure 1 (B))

$$(A_p)_{ij} = \sum_{k \in S(j)} A_{ik}, \ \forall i = 1, \dots, m, \ j = 1, \dots, p \tag{36}$$

and the $n \times p$ interpolation operator I_p^n by

$$(I_p^n)_{ij} = \begin{cases} 1, & \text{if } i \ \in S(j) \\ 0, & \text{if } i \ \notin S(j), \end{cases} \tag{37}$$

$i = 1, \ldots, n,\ j = 1, \ldots, p$.

Proposition 7 *The above matrices A, A_p and I_p^n satisfy* (11)*. Moreover, the interpolation operator I_p^n has full column rank.*

Proof. We shall prove that $\forall i = 1, \ldots, m,\ \forall j = 1, \ldots, p$

$$(AI_p^n)_{ij} = (A_p)_{ij}. \tag{38}$$

Let $\{g_j\} \subset I\!R^p$, $\{f_i\} \subset I\!R^m$ be the canonical basis. By the definition of I_p^n, for a fixed $j \in \{1, \ldots, p\}$ the j-th column of I_p^n has 1's only on the positions $j_1, j_2, j_3, j_4 \in S(j)$ and 0's in rest. Then we have (by also using (36))

$$(AI_p^n)_{ij} = \langle AI_p^n g_j, f_i \rangle = \langle I_p^n g_j, A^T f_i \rangle = \langle \text{column } j \text{ of } I_p^n, \text{ row } i \text{ of } A \rangle =$$

$$A_{i,j_1} + A_{i,j_2} + A_{i,j_3} + A_{i,j_4} = \sum_{k \in S(j)} A_{ik} = (A_p)_{ij}.$$

Concerning the second assertion, we first observe that from (33)-(34) we obtain

$$S(j) \cap S(j') = \emptyset,\ \forall j \neq j'. \tag{39}$$

From (39) it results that the j-th and j'-th columns of I_p^n are "structurally orthogonal", thus they are orthogonal and the proof is complete.

Proposition 8 *If A from before has full column rank and A_p and I_p^n are defined as in* (36) *and* (37)*, then A_p has also full column rank.*

Proof. From (36) and (37) it results that the column $(A_p)^j$ of the coarse grid matrix A_p are given by (see also (33))

$$(A_p)^j = A(I_p^n)^j = \sum_{k \in S(j)} (A)^k. \tag{40}$$

Let $\sum_{j=1}^p \alpha_j (A_p)^j = 0$ be a null linear combination of them. Then

$$0 = \sum_{j=1}^p \alpha_j A_p^j = \sum_{j=1}^p \alpha_j \left(\sum_{k \in S(j)} (A)^k \right) = \sum_{q=1}^n \beta_q (A)^q, \tag{41}$$

where the last equality holds because of the structural orthogonality of the column of I_p^n (see the proof of Proposition 7) and β_q are various α_j values (according to the numbering in the fine and coarse grids; see Figure 1 (A)). But, from our hypothesis and (41) it results that $\beta_q = 0,\ \forall q = 1, \ldots, n$, thus $\alpha_j = 0,\ \forall j = 1, \ldots, p$ and the proof is complete.

Remark 1 *From above it results that our approach* $(36) - (37)$ *satisfies all the properties from section* 2 *when A is full column rank. One disadvantage of this approach would be the fact that the number of rows (i.e. rays) in all the coarse grid matrices remains the same as for A. But, this is not a very unpleasant fact because, although the dimension m is the same, the number of columns in the coarse grid matrices is divided by* 4 *with each discretization level, which will give as on the coarsest one a very "high", but "thin" matrix for which a problem like* (4) *can be easily solved. A very good aspect, beside the above mentioned properties of the correction is that the coarse grid matrices keep the sparsity of the problem (on the corresponding discretization levels).*

Remark 2 *In the paper* [3] *the authors considered also the possibility to reduce the number of rows (rays) in A_p. This is mathematically equivalent with constructing the coarse grid matrix A_p by (see for comparison* (11)*)*

$$A_p = I_m^q A I_p^n, \tag{42}$$

where I_m^q is an $q \times m$ matrix. It can be a "pick - up" one (i.e. it contains only one 1 *in each row, in a prescribed position) or an "interpolation-like" matrix (as e.g. the linear restriction operator on an* 1D *multigrid for Poisson equation; see e.g.* [4]*). Unfortunately, such a construction doesn't always satisfy Proposition* 1*, which can destroy the efficiency of the correction process* (7) *(more* *clear, the correction step* (7) *is no more compatible with the solutions set* $LSS(A;b)$*). Indeed, in this case* (7) *becomes*

$$\begin{cases} d = b - Ax; \; d^q = I_m^q d \\ \parallel A_p v_p - d^q \parallel = \min! \Rightarrow \quad v_p = A_p^+ d^q = A_p^+ \left(P_{R(A_p)}(d^q) \right) \\ \bar{x} = x + I_p^n v_p \end{cases} \tag{43}$$

Then, if $x \in LSS(A;b)$*, by also using* (13) *we obtain* $d^q = I_m^q b_A^*$*, which doesn't always belong to the subspace* $N(A^T) \subset N(A_p^+)$*, i.e.* v_p *and* $I_p^n v_p$ *are not* 0*, thus* $\bar{x} \neq x$.

Remark 3 *The* $m \times p$ *matrix* E *from* (29) *can be a "pick-up" one, i.e. only with 1's and 0's as entries. defined in the following way (see* (33) − (34)*)*

$$(E)_{ij} = \begin{cases} 1, & \text{if the } i-th \text{ ray intersects at least} \\ & \text{one pixel } P_k \text{ with } k \in S(j) \\ 0, & \text{else.} \end{cases} \tag{44}$$

In this way we can get an (enough) sparse interpolation operator I_p^n*. But, in order to keep the Assumption 1, we have to define the coarse grid matrix* A_p *as in* (11)*, which gives us (see* (29)*)*

$$A_p = A I_p^n = A A^T E. \tag{45}$$

Then, beside the fact that its elements have to be computed as scalar products of the form $\langle A_i, A_j \rangle$*, the sparsity structure can be different that one for* A_p *from section* 5.1 (see (36)).

3.2 Numerical Results

We implemented the setup of the projection matrix A and the solution of the least squares problem in 2D in Matlab.

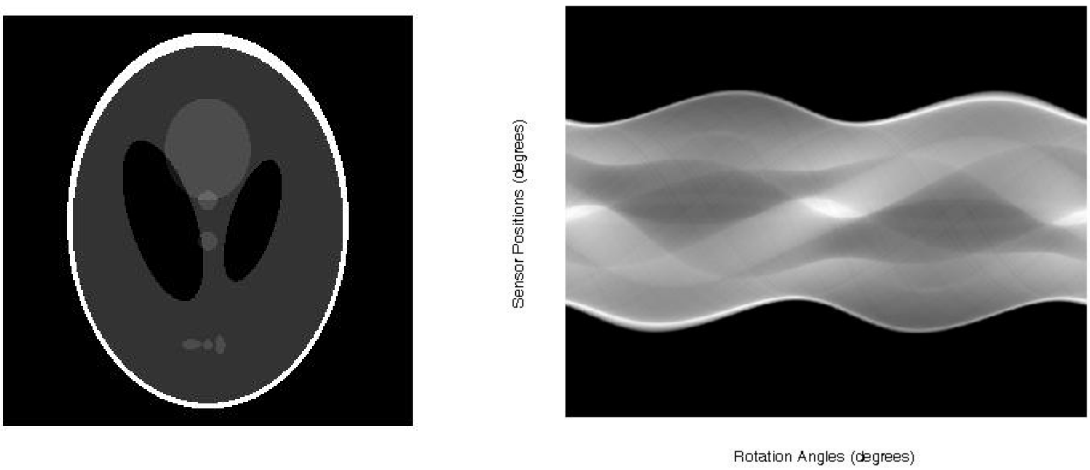

Figure 2: Modified Shepp-Logan phantom (left) and sinogram (right).

Figure 2 shows the original image (size 256^2), a Shepp-Logan phantom available in Matlab and the corresponding right hand side b, the sinogram, computed by the Matlab routine fanbeam. For our experiments we resized the image to $n = 24^2$ and used $m = RP = 2808$ rays in all projections with $R = 39$ and $P = 72$. The structure of the projection matrix $A \in R^{2808 \times 576}$ that has full column rank for this setup is shown in Figure 3.

The results in Figure 4 show that the multigrid method is able to reduce both the error (the L_2-norm of the difference between original image and reconstructed image) and the residual (the L_2-norm of $b - Ax$).

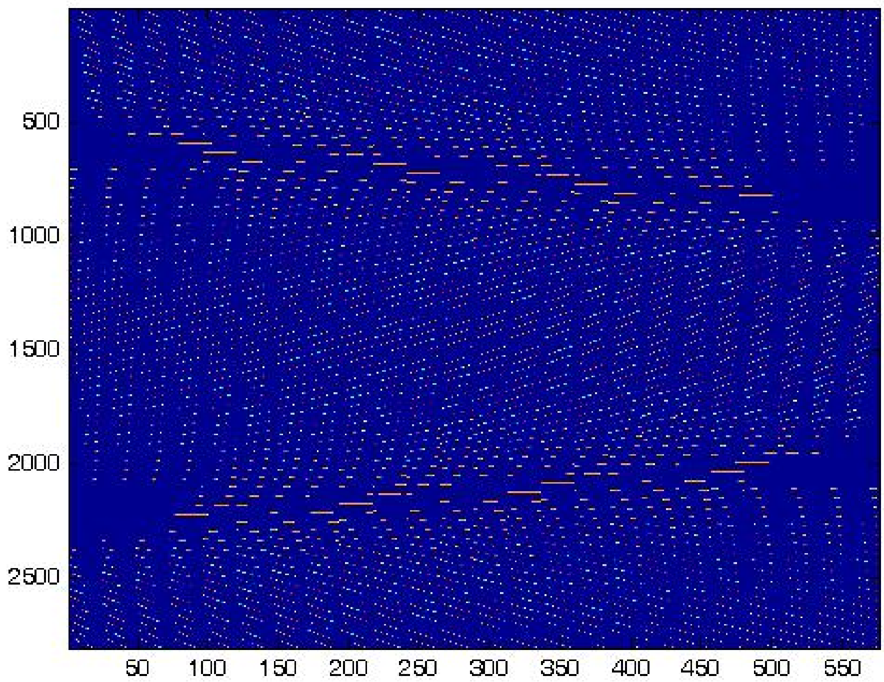

Figure 3: Structure of projection matrix

4 Conclusions and Future work

We have derived a AMG method for general least square systems. The next steps are now a more detailed experimental validation of our theoretical results. This includes a study of the multigrid convergence rates for different matrix sizes in 2D and 3D and an evaluation of the amount of the computational time that can be saved by our method compared to Kaczmarz or other standard algebraic image reconstruction algorithms.

References

[1] Björk A., *Numerical methods for least squares problems*, SIAM Philadelphia, 1996.

[2] Popa C., *Algebraic multigrid for general inconsistent linear systems: Preliminary results*, Technical Report **06-2**(2006), Lehrstuhl für Informatik 10 (Systemsimulation), FAU Erlangen-Nürnberg.

[3] Prümmer, M., Köstler, H. et al., *A full multigrid technique to accelerate an ART scheme for tomographic image reconstruction*, Proceedings of ASIM Conference, September 12-15 2005, Erlangen, Germany, SCS Publishing House e.V., 632-637.

[4] Ruge, J., Stüben, K., *Algebraic multigrid (AMG)*, Arbeitspapiere der GMD **210**(1986), Bonn.

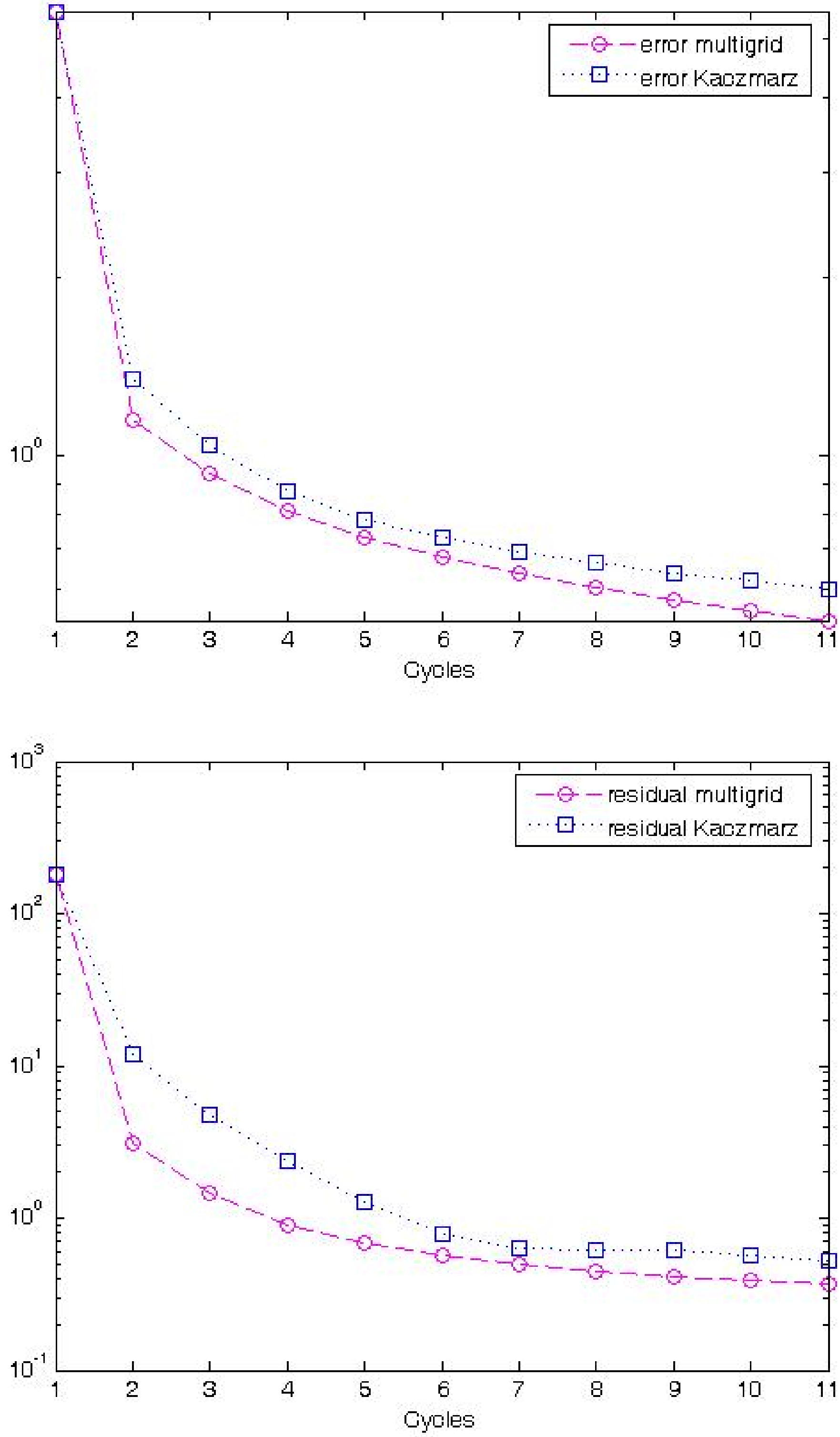

Figure 4: Comparison of errors and residuals for V(2,2)-cycles using only 1 level (Kaczmarz) and using 2 levels with a direct solver on the coarse grid.