The shadow vector in the Lanczos Method

Petr Tichý, Praha

1. Introduction. Consider a system of linear algebraic equations

$$\mathbf{A}x = b \tag{1}$$

with a nonsingular matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Let x_0 be a given initial approximation and r_0 the corresponding residual $r_0 = b - \mathbf{A}x_0$ and let $\mathcal{K}_k(\mathbf{A}, r_0) \stackrel{\text{def}}{=} \text{span}\{r_0, \mathbf{A}r_0, \ldots, \mathbf{A}^{k-1}r_0\}$ denote the k-th Krylov subspace spanned by matrix $\mathbf{A}$ and vector r_0.

The *Lanczos method* (LM) constructs approximate solutions x_k^L and residuals $r_k^L \stackrel{\text{def}}{=} b - \mathbf{A}x_k^L$ such that

$$x_k^L \in x_0 + \mathcal{K}_k(\mathbf{A}, r_0), \quad r_k^L \perp \mathcal{K}_k(\mathbf{A}^T, \widetilde{r}_0), \tag{2}$$

where $\widetilde{r}_0$ is an auxiliary vector often called the *shadow vector*, using two three-term recurrences

$$r_k^L = \gamma_k(\mathbf{A}r_{k-1}^L + \alpha_k r_{k-1}^L + \beta_k r_{k-2}^L), \tag{3}$$

$$x_k^L = \gamma_k(-r_{k-1}^L + \alpha_k x_{k-1}^L + \beta_k x_{k-2}^L). \tag{4}$$

To satisfy the condition $r_k^L \in r_0 + \mathbf{A}\mathcal{K}_k(\mathbf{A}, r_0)$, which is equivalent to the condition $x_k^L \in x_0 + \mathcal{K}_k(\mathbf{A}, r_0)$, we choose $\gamma_k = (\alpha_k + \beta_k)^{-1}$. Coefficients α_k and β_k determine the direction of the computed vector. If also vectors r_{k-1}^L and r_{k-2}^L satisfy the relevant orthogonality conditions then it is possible to choose α_k and β_k such that the orthogonal condition in (2) is satisfied. For effective computing these coefficients, we generate an auxiliary sequence of vectors

$$\widetilde{r}_k = \gamma_k(\mathbf{A}^T\widetilde{r}_{k-1} + \alpha_k\widetilde{r}_{k-1} + \beta_k\widetilde{r}_{k-2}),$$

which represents the base of the space $\mathcal{K}_k(\mathbf{A}^T, \widetilde{r}_0)$ and vectors $\widetilde{r}_k$ satisfy $\widetilde{r}_k \perp \mathcal{K}_k(\mathbf{A}, r_0)$. Using these vectors, we can compute numbers α_k and β_k as

$$\alpha_k = -\frac{(\widetilde{r}_{k-1}, \mathbf{A}r_{k-1}^L)}{(\widetilde{r}_{k-1}, r_{k-1}^L)}, \quad \beta_k = -\frac{(\widetilde{r}_{k-2}, \mathbf{A}r_{k-1}^L)}{(\widetilde{r}_{k-2}, r_{k-2}^L)}. \tag{5}$$

This paper was supported under the Grant No. 201/19/0267 of the Grant Agency of the Czech Republic and under the Grant No CEZ J13/98: 113200007.

As we can see, the *breakdown* occurs if the scalar product $(\widetilde{r}_i, r_i^L)$ that appears in the denominators of α's and β's is equal to zero for some i or if $\alpha_i + \beta_i = 0$. It is possible to show that if no breakdown occurs during the run of the LM then residuals r_k^L and approximations x_k^L are determined uniquely by the conditions (2) ([Gutkn–97]).

Let us focus on one parameter of the LM named the shadow vector. There are two choices of this vector which are commonly used by practical computing. The first is to choose $\widetilde{r}_0$ to be equal to the initial residual vector. This choice is inspired by the symmetric case, because if $\mathbf{A}$ is a symmetric matrix then the LM computes the same residuals and approximations as CG algorithm. The second choice, nowadays the most supported one, is to choose $\widetilde{r}_0$ randomly. In 1997, at the Czech-US Workshop in Milovy, Mrs. Greenbaum asked the question *"What is the role of the shadow vector $\widetilde{r}_0$?"* and she found a new theorem which she presented as a disturbing result:

Known result (Greenbaum). *If 3-term recurrence (α's and β's can be almost anything) is run for no more than $n/2$ steps, there is a vector $\widetilde{r}_0$ such that recurrence came from the Lanczos method.*

This result shows that the LM need not converge in the first $n/2$ steps and can behave almost arbitrarily.

We wanted to use this freedom of the LM in the first $n/2$ steps to make this method closer to another Krylov subspace method like GMRES. We will try to answer the question *"Is there a vector $\widetilde{r}_0$ such that the Lanczos method computes residuals of another Krylov subspace method?"*

2. Definition and properties of space $\mathcal{W}_k$. We used the idea of Greenbaum for determining the shadow vector. We exploit the fact that

$$0 = (r_k^L, (\mathbf{A}^T)^i \widetilde{r}_0) = (\mathbf{A}^i r_k^L, \widetilde{r}_0)$$

and hence

$$r_k^L \perp \mathcal{K}_k(\mathbf{A}^T, \widetilde{r}_0) \Leftrightarrow \widetilde{r}_0 \perp \mathcal{K}_k(\mathbf{A}, r_k^L), \quad k = 1, 2, \ldots.$$

From the orthogonality condition above it follows that the shadow vector $\widetilde{r}_0$

is orthogonal to space

$$\mathcal{W}_k(\mathbf{A}, r_1^L, \ldots, r_k^L) \stackrel{\text{def}}{=} \text{span} \begin{pmatrix} r_1^L, & r_2^L, & r_3^L, & r_4^L & \ldots & r_k^L, \\ & \mathbf{A}r_2^L, & \mathbf{A}r_3^L, & \mathbf{A}r_4, & \ldots & \mathbf{A}r_k^L, \\ & & \mathbf{A}^2 r_3^L, & \mathbf{A}^2 r_4^L, & \ldots & \mathbf{A}^2 r_k^L, \\ & & & \ddots & & \vdots \\ & & & & \ddots & \vdots \\ & & & & & \mathbf{A}^{k-1} r_k^L \end{pmatrix}$$

generated by matrix $\mathbf{A}$ and residuals $r_1^L, \ldots, r_k^L$. In the following paragraphs, we will explore the properties of space $\mathcal{W}_k(\mathbf{A}, r_1, \ldots, r_k)$ in dependence on parameters $r_1, \ldots, r_k$.

General assumptions and denomination. We will consider k to be some integer, $1 \le k \le [n/2]$, $\dim(\mathcal{K}_{2k}(\mathbf{A}, r_0)) = 2k$, $r_i = \mathcal{P}_i(\mathbf{A}) r_0$ $(i = 1, \ldots, k)$ where $\mathcal{P}_i(\xi)$ is a polynomial satisfying $\mathcal{P}_i(0) = 1$ (this is consistent with $r_i \in r_0 + \mathbf{A}\mathcal{K}_i(\mathbf{A}, r_0)$) and $\deg(\mathcal{P}_i(\xi)) = i$ (nontrivial leading coefficient provides linear independence of vectors $r_0, r_1, \ldots, r_k$) and we will write only $\mathcal{W}_k$ instead of $\mathcal{W}_k(\mathbf{A}, r_1, \ldots, r_k)$.

Lemma 1. *The dimension of space $\mathcal{W}_k$ is equal to $2k-1$ or $2k$. The initial residual lies in $\mathcal{W}_k$ if and only if* $\dim(\mathcal{W}_k) = 2k$.

Proof. Any vector from $\mathcal{W}_k$ lies in $\mathcal{K}_{2k}(\mathbf{A}, r_0)$ and, therefore, $\dim(\mathcal{W}_k) \le 2k$. Since we can find $2k-1$ linearly independent vectors, for example $r_1, \ldots, r_k$, $\mathbf{A}r_k, \ldots, \mathbf{A}^{k-1} r_k$, it is $\dim(\mathcal{W}_k) \ge 2k-1$.

If $r_0 \in \mathcal{W}_k$ we can find in this space $2k$ linearly independent vectors: $r_0, r_1, \ldots, r_k$, $\mathbf{A}r_k, \ldots, \mathbf{A}^{k-1} r_k$, and, for that reason, $\dim(\mathcal{W}_k) = 2k$.

On the other hand, we know that $\mathcal{W}_k \subseteq \mathcal{K}_{2k}(\mathbf{A}, r_0)$ and if $\dim(\mathcal{W}_k) = 2k$ then $\mathcal{W}_k = \mathcal{K}_{2k}(\mathbf{A}, r_0)$ and $r_0 \in \mathcal{W}_k$. □

In the following lemma, we will say more about the dimension of $\mathcal{W}_k$ in dependence on recurrence which generates residuals $r_1, \ldots, r_k$.

Lemma 2. *Let us suppose a Krylov subspace method whose residuals are computed by recurrence*

$$(6) \qquad r_i = \alpha_i^{(i)} \mathbf{A} r_{i-1} + \sum_{j=0}^{i-1} \alpha_j^{(i)} r_j, \quad \sum_{j=0}^{i-1} \alpha_j^{(i)} = 1, \quad \alpha_i^{(i)} \ne 0, \quad i = 1, \ldots, k.$$

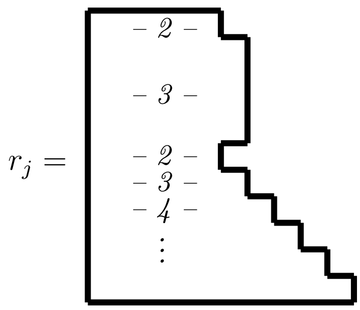
Then vector r_0 does not lie in the space $\mathcal{W}_k$ if and only if the recurrences of the Krylov subspace method are

(a) *three-term or*

(b) *of the form*

$$
\begin{array}{llcl}
(7) & r_1 & = & \alpha_1^{(1)}\mathbf{A}r_0 + r_0 \\
(8) & r_j & = & \alpha_j^{(j)}\mathbf{A}r_{j-1} + \alpha_{j-1}^{(j)}r_{j-1} + \alpha_{j-2}^{(j)}r_{j-2},\ j = 2, \dots, i-1, \\
(9) & r_i & = & \alpha_i^{(i)}\mathbf{A}r_{i-1} + r_{i-1} \\
(10) & r_j & = & \alpha_j^{(j)}\mathbf{A}r_{j-1} + \alpha_{j-1}^{(j)}r_{j-1} + \dots + \alpha_{i-1}^{(j)}r_{i-1},\ j = i+1, \dots, k.
\end{array}
$$

where $2 \le i \le k$ *and* $\alpha_{j-2}^{(j)} \ne 0$ *for* $j = 2, \dots, i-1$*. The shape of the recurrences is depicted in the following scheme.*



Proof. $\boxed{\Rightarrow}$ Consider the case $r_0 \notin \mathcal{W}_k$. We must show that some coefficients $\alpha_i^{(j)}$ of the Krylov subspace method (6) are zeros.

• *Coefficients* $\alpha_0^{(i)}$*,* $i = 3, \dots, k$ *are zeros.* If we suppose that some $\alpha_0^{(i)} \ne 0$, $3 \le i \le k$, we can express vector r_0 as a linear combination of vectors from space $\mathcal{W}_k$ and it implies that $r_0 \in \mathcal{W}_k$, which is a contradiction.

• *If* $\alpha_{j-2}^{(j)} \ne 0$ *for every* $j = 2, \dots, i-1$ *then it is* $\alpha_{j-1}^{(s)} = 0$*,* $s = j+2, \dots, k$. For better understanding the proof of this proposition we write down the recurrences:

$$
\begin{array}{lrcl}
(11) & r_2 & = & \alpha_2^{(2)}\mathbf{A}r_1 + \alpha_1^{(2)}r_1 + \alpha_0^{(2)}r_0 \\
(12) & r_3 & = & \alpha_3^{(3)}\mathbf{A}r_2 + \alpha_2^{(3)}r_2 + \alpha_1^{(3)}r_1 \\
(13) & r_4 & = & \alpha_4^{(4)}\mathbf{A}r_3 + \alpha_3^{(4)}r_3 + \alpha_2^{(4)}r_2 + \alpha_1^{(4)}r_1 \\
(14) & r_5 & = & \alpha_5^{(5)}\mathbf{A}r_4 + \alpha_4^{(5)}r_4 + \alpha_3^{(5)}r_3 + \alpha_2^{(5)}r_2 + \alpha_1^{(5)}r_1 \\
 & & \vdots & \\
(15) & r_i & = & \alpha_i^{(i)}\mathbf{A}r_{i-1} + \alpha_{i-1}^{(i)}r_{i-1} + \dots + \alpha_2^{(i)}r_2 + \alpha_1^{(i)}r_1 \\
 & & \vdots & \\
(16) & r_k & = & \alpha_k^{(k)}\mathbf{A}r_{k-1} + \alpha_{k-1}^{(k)}r_{k-1} + \dots + \alpha_2^{(k)}r_2 + \alpha_1^{(k)}r_1.
\end{array}
$$

Proof by induction on i.

Initial step: ($i = 3$) If $\alpha_0^{(2)} \neq 0$ we obtain from (11) that $\mathbf{A}r_1 \notin \mathcal{W}_k$ (else $r_0 \in \mathcal{W}_k$). If we multiply equations (13)-(16) by matrix $\mathbf{A}$ we obtain from the facts $\mathbf{A}r_j \in \mathcal{W}_k$, $j = 2, \ldots, k$, $\mathbf{A}^2 r_j \in \mathcal{W}_k$, $j = 3, \ldots, k$ and $\mathbf{A}r_1 \notin \mathcal{W}_k$ that $\alpha_1^{(j)} = 0$, $j = 4, \ldots, k$.

Induction step: ($i > 3$) Let us suppose that $\alpha_{j-2}^{(j)} \neq 0$, $j = 2, \ldots, i-1$, $\mathbf{A}^j r_j \notin \mathcal{W}_k$, $j = 1, \ldots, i-3$ and $\alpha_{j-1}^{(s)} = 0$, $s = j+2, \ldots, k$, $j = 2, \ldots, i-2$. From $\alpha_{i-3}^{(i-1)} \neq 0$ we have $\mathbf{A}^{i-2} r_{i-2} \notin \mathcal{W}_k$ (multiply recurrence for vector r_{i-1} by $\mathbf{A}^{i-3}$ and use the fact $\mathbf{A}^{i-3} r_{i-3} \notin \mathcal{W}_k$). If we multiply equations for vectors r_j, $j = i+1, \ldots, k$ by matrix $\mathbf{A}^{i-2}$ we obtain from the fact $\mathbf{A}^{i-2} r_{i-2} \notin \widetilde{\mathcal{W}}_k$ that $\alpha_{i-2}^{(j)} = 0$, $j = i+1, \ldots, k$, i.e., $\alpha_{j-1}^{(s)} = 0$, $s = j+2, \ldots, k$, $j = 2, \ldots, i-1$ and $A^j r_j \notin \mathcal{W}_k$, $j = 1, \ldots, i-2$.

• From the previous two points we see that under assumption $r_0 \notin \mathcal{W}_k$ the recurrences of the Krylov subspace method can be written in the form (b) or that they are three-term.

$\boxed{\Leftarrow}$ On the other hand, let us suppose that the recurrences of some Krylov subspace method are three-term or of form (b). We shall show that in this case $r_0 \notin \mathcal{W}_k$ or, equivalently, if we denote the space generated by vectors $r_1, \ldots, r_k$, $\mathbf{A}r_k, \ldots, \mathbf{A}^{k-1} r_k$ by $\widetilde{\mathcal{W}}_k$ we have to prove the equality $\mathcal{W}_k = \widetilde{\mathcal{W}}_k$ (here it is sufficient to show that $\mathbf{A}^s r_j \in \widetilde{\mathcal{W}}_k$, $j = 2, \ldots, k-1$, $s = 1, \ldots, j-1$).

• If the recurrences are three-term ones we have

$$\mathbf{A}^s r_{j-1} \in \operatorname{span}\{\mathbf{A}^{s-1} r_j, \mathbf{A}^{s-1} r_{j-1}, \mathbf{A}^{s-1} r_{j-2}\}, \tag{17}$$

where $j = 3, \ldots, k$, $s = 1, \ldots, k-2$. By induction on s we can see that $\mathbf{A}^s r_{j-1} \in \widetilde{\mathcal{W}}_k$, $j = 3, \ldots, k$, $s = 1, \ldots, j-2$.

• If they are of form (b) then for every integer $s > 0$ it holds

$$\begin{aligned}
(18)\quad \mathbf{A}^s r_{j-1} &\in \operatorname{span}\{\mathbf{A}^{s-1} r_j, \mathbf{A}^{s-1} r_{j-1}, \mathbf{A}^{s-1} r_{j-2}\}, \quad j = 3, \ldots, i-1,\\
(19)\quad \mathbf{A}^s r_{i-1} &\in \operatorname{span}\{\mathbf{A}^{s-1} r_i, \mathbf{A}^{s-1} r_{i-1}\},\\
(20)\quad \mathbf{A}^s r_{j-1} &\in \operatorname{span}\{\mathbf{A}^{s-1} r_j, \mathbf{A}^{s-1} r_{j-1}, \ldots, \mathbf{A}^{s-1} r_{i-1}\}, \quad j = i+1, \ldots, k.
\end{aligned}$$

We will show the fact $\mathbf{A}^s r_{j-1} \in \widetilde{\mathcal{W}}_k$, $j = 3, \ldots, k$, $s = 1, \ldots, j-2$ in two steps. At first we prove that $\mathbf{A}^s r_{i-1} \in \widetilde{\mathcal{W}}_k, \ldots, \mathbf{A}^s r_{k-1} \in \widetilde{\mathcal{W}}_k$ for $s = 1, \ldots, k$.

It follows from (19) that $\mathbf{A}r_{i-1} \in \widetilde{\mathcal{W}}_k$ and from (20) that $\mathbf{A}r_j \in \widetilde{\mathcal{W}}_k$, $j = i, \ldots, k-1$. By induction, let us suppose that $\mathbf{A}^{s-1} r_j \in \widetilde{\mathcal{W}}_k$, $j = i-1, \ldots, k-1$. According to (19) it holds that $\mathbf{A}^s r_{i-1} \in \widetilde{\mathcal{W}}_k$ and from (20) that $\mathbf{A}^s r_j \in \widetilde{\mathcal{W}}_k$, $j = i, \ldots, k-1$, $s \leq k$.

In the second step we shall prove that $\mathbf{A}^s r_j \in \widetilde{\mathcal{W}}_k,\ j = 2,\ldots,i-2,$ $s = 1,\ldots,j-1$. It can be easily shown by induction, using (18) and the result of the first step ($A^s r_{i-1} \in \widetilde{\mathcal{W}}_k$). □

Lemma 3. *Consider a Krylov subspace method whose residuals are linearly independent and generated by three-term recurrence as in* (3). *Then*

$$\mathcal{W}_k(\mathbf{A}, r_1, \ldots, r_k) = \operatorname{span}\{r_1, \ldots, r_{2k-1}\}. \tag{21}$$

Proof. The dimension of both spaces is equal to $2k-1$. It is sufficient to show that residuals $r_{k+1}, \ldots, r_{2k-1}$ lie in $\mathcal{W}_k$.

All vectors on the right hand side of the recurrence

$$r_{k+1} = \gamma_k(\mathbf{A} r_k + \alpha_k r_k + \beta_k r_{k-1})$$

lie in $\mathcal{W}_k$ and so $r_{k+1} \in \mathcal{W}_k$. If we multiply this recurrence by matrices $\mathbf{A}, \ldots, \mathbf{A}^{k-2}$ we obtain $\mathbf{A}^j r_{k+1} \in \mathcal{W}_k,\ j = 0, \ldots, k-2$ and, by induction, $\mathbf{A}^j r_i \in \mathcal{W}_k$, $i = k+1, \ldots, 2k-1$, $j = 0, \ldots, 2k-1-i$. □

3. Possible choices of the shadow vector. Now we can use space $\mathcal{W}_k$ or its subspace to make the Lanczos method compute the given residual vectors.

Theorem 1. *Choose vector $\widetilde{r}_0$ to satisfy the condition $\widetilde{r}_0 \perp \mathcal{W}_k$. If no breakdown occurs during the run of the Lanczos method then the residuals computed by the Lanczos method are the same as the given residuals $r_1, \ldots, r_k$.*

Proof. The residual of the Lanczos method r_i^L $(i \leq k)$ and the given residual r_i lie both in variety $r_0 + \mathbf{A}\mathcal{K}_i(\mathbf{A}, r_0)$, are orthogonal to the space $\mathcal{K}_i(\mathbf{A}^T, \widetilde{r}_0)$ and, because of uniqueness, have to be equal. □

Lemma 4. *Consider a method generated by recurrences* (7)−(10) *and choose a shadow vector such that $\widetilde{r}_0 \perp \mathcal{W}_k$. Then the Lanczos method will break down at the latest at the i-th step by computing the residual r_i.*

Proof. Multiplying (9) by matrix $\mathbf{A}^{i-2}$ we obtain $\mathbf{A}^{i-1} r_{i-1} \in \mathcal{W}_k$. It means that

$$(\mathcal{P}_{i-1}(\mathbf{A}^T)\widetilde{r}_0, r_{i-1}) = (\widetilde{r}_0, \mathcal{P}_{i-1}(\mathbf{A}) r_{i-1}) = 0$$

for any matrix polynomial $\mathcal{P}_{i-1}(\mathbf{A})$ of degree $i-1$ at most. Since we can write the auxiliary vector $\widetilde{r}_{i-1}$ in the form $\widetilde{r}_{i-1} = \mathcal{P}_{i-1}(\mathbf{A}^T)\widetilde{r}_0$ we have $(\widetilde{r}_{i-1}, r_{i-1}) = 0$ and the breakdown occurs by computing the residual r_i. □

Theorem 2. *Consider any Krylov subspace method whose residual vectors* $r_1, \ldots, r_k$ *are computed by the three-term recurrence* (3) *where* $\gamma_i = \frac{1}{\alpha_i+\beta_i}$*,* $\alpha_i + \beta_i \neq 0$*,* $\beta_i \neq 0$*,* $i \leq k$*. Choose the shadow vector such that* $(\widetilde{r}_0, r_0) \neq 0$ *and* $\widetilde{r}_0 \perp \mathcal{W}_k$*. Then the Lanczos method computes residuals* r_i*,* $i \leq k$*, without breakdown.*

Proof. From the previous theorem we know that if no breakdown occurs then the Lanczos method computes residuals $r_1, \ldots, r_k$. We have to show that no breakdown occurs.

• First, we will show by induction that $\mathbf{A}^i r_i \notin \mathcal{W}_k$ for $i = 0, \ldots, k-1$.
Initial step: We know that $\mathbf{A}^0 r_0 = r_0$ does not lie in $\mathcal{W}_k$.
Induction step: If we multiply the three-term recurrence

$$r_i = \alpha_i^{(i)} \mathbf{A} r_{i-1} + \alpha_{i-1}^{(i)} r_{i-1} + \alpha_{i-2}^{(i)} r_{i-2}$$

by $\mathbf{A}^{i-2}$ we obtain from the facts $\mathbf{A}^{i-2} r_i \in \mathcal{W}_k$, $\mathbf{A}^{i-2} r_{i-1} \in \mathcal{W}_k$, $\mathbf{A}^{i-2} r_{i-2} \notin \mathcal{W}_k$ and $\alpha_i^{(i)} \neq 0$, $\alpha_{i-2}^{(i)} \neq 0$ that $\mathbf{A}^{i-1} r_{i-1} \notin \mathcal{W}_k$.

• Second, we will show that $(r_i, \mathcal{P}_i(\mathbf{A}^T)\widetilde{r}_0) \neq 0$ for any matrix polynomial of degree i. It holds that

$$(r_i, \mathcal{P}_i(\mathbf{A}^T)\widetilde{r}_0) = (r_i, \sum_{j=0}^{i} \xi_j (\mathbf{A}^T)^j \widetilde{r}_0) = \sum_{j=0}^{i} \xi_j (\mathbf{A}^j r_i, \widetilde{r}_0) = \xi_i (\mathbf{A}^i r_i, \widetilde{r}_0)$$

and $\xi_i \neq 0$ because $\mathcal{P}_i(\xi)$ is a polynomial of degree i. Let us suppose for a moment that $(\mathbf{A}^i r_i, \widetilde{r}_0) = 0$. Then from the facts $\mathbf{A}^i r_i \notin \mathcal{W}_k$, $\mathbf{A}^i r_i \in \mathcal{K}_{2k}(\mathbf{A}, r_0)$, it follows that $\mathcal{K}_{2k}(\mathbf{A}, r_0) = \mathcal{W}_k \cup \mathbf{A}^i r_i$ and then $\widetilde{r}_0$ is orthogonal to space $\mathcal{K}_{2k}(\mathbf{A}, r_0)$, it means $(\widetilde{r}_0, r_0) = 0$, which is a contradiction.

Therefore, if we run the Lanczos method with this shadow vector we can compute coefficients α_i and β_i without breakdown $(r_i^L, \widetilde{r}_i) \neq 0$. By the choice of α_i and β_i, the direction of the computed vector is determined uniquely and, hence, $\mathbf{A} r_{i-1}^L + \alpha_i r_{i-1}^L + \beta_i r_{i-2}^L$ is a multiple of r_i. It means that there exists an intersection of this direction with the variety $r_0 + \mathbf{A}\mathcal{K}_i(\mathbf{A}, r_0)$ and it is possible to compute γ_i. □

Notice. This theorem, in a little different form, was presented by Greenbaum. We have only specified its assumptions and have said when the LM computes residuals of another method without breakdown.

Now we will choose the shadow vector to be orthogonal to the particular subspace of $\mathcal{W}_k$ in order to reach equality of some Lanczos' residuals and residuals of another Krylov subspace method.

Theorem 3. *Let* $r_k = \mathcal{P}_k(\mathbf{A})r_0$, $\mathcal{P}_k(0) = 1$, $\deg(\mathcal{P}_k(\xi)) = k$ *where* $k < n$ *and let* $\dim(\mathcal{K}_k(\mathbf{A}, r_0)) = k$. *Choose a shadow vector such that* $(\widetilde{r}_0, r_0) \neq 0$ *and* $\widetilde{r}_0 \perp \mathcal{K}_k(\mathbf{A}, r_k)$. *If no breakdown occurs during the run of the Lanczos Method then* $r_k^L = r_k$.

Proof. If no breakdown occurs during the run of the Lanczos Method and if $\dim(\mathcal{K}_k(\mathbf{A}, r_0)) = k$ then also $\dim(\mathcal{K}_k(\mathbf{A}^T, \widetilde{r}_0)) = k$, residuals r_k^L and r_k satisfy the same determining conditions, they lie in the same variety and are orthogonal to $\mathcal{K}_k(\mathbf{A}^T, \widetilde{r}_0)$ of dimension k, and, because of uniqueness, $r_k = r_k^L$. □

Theorem 4. *Choose* $k \leq [n/2]$ *and let* $r_i = \mathcal{P}_i(\mathbf{A})r_0$, $\mathcal{P}_i(0) = 1$, $\deg(\mathcal{P}_i(\xi)) = i$, $i = 1, 2, 4, 8, \ldots, 2^l$, $2^l \leq k$. *Let* $\dim(\mathcal{K}_{2k}(\mathbf{A}, r_0)) = 2k$. *Define the space*

$$\mathcal{Z}_{2^l} \stackrel{def}{=} \bigcup_{i=0}^{l} \mathcal{K}_{2^i}(\mathbf{A}, r_{2^i}),$$

and choose a shadow vector orthogonal to this space, $\widetilde{r}_0 \perp \mathcal{Z}_{2^l}$, $(\widetilde{r}_0, r_0) \neq 0$. *If no breakdown occurs during the run of the Lanczos method then* $r_1^L = r_1$, $r_2^L = r_2$, $r_4^L = r_4$, $\ldots$, $r_{2^l}^L = r_{2^l}$.

Proof. Residual r_i^L and r_i $(i = 2^j)$ satisfy the same determining conditions. If no breakdown occurs in Lanczos method, residuals r_i^L are determined uniquely and, hence $r_i^L = r_i$. □

Notice. Let us suppose that $k = 2^l$ for some integer l. Space $\mathcal{Z}_k$ is formed by $2k - 1$ linearly independent vectors and $r_0 \notin \mathcal{Z}_k \subseteq \mathcal{W}_k$. It is not always possible to find the shadow vector such that the Lanczos method computes residuals r_{2^i}, $i = 0, \ldots, l$ without breakdown. It can be readily seen from Lemma 2. If the recurrences of a method are of form (b) then $\mathcal{Z}_k = \mathcal{W}_k$ and according to Lemma 4, the Lanczos method will break down at the latest at the step i, $i < k$.

Conclusion. For every set of residuals $r_1, \ldots, r_k$ it is always possible to find $\widetilde{r}_0$ such that the orthogonal condition in (2) is satisfied. The Lanczos method can compute all these residuals only if they have come from 3-term recurrences and satisfy the assumptions of Theorem 2. It means that by the LM we can map all three-term Krylov subspace methods (3), such that $\beta_i \neq 0$, $\gamma_i \neq 0$, $\alpha_i + \beta_i \neq 0$, in the first $n/2$ steps.

It is possible to choose a shadow vector such that if the LM does not break down then some Lanczos' residuals are the same as those of another Krylov subspace method. We have not shown yet whether there is $\widetilde{r}_0$ such that the LM computes particular residuals of another method without breakdown.

4. Numerical experiments. For testing the methods we used the convection-diffusion equation

$$-10^{-3} \triangle u(x,y) + u_x(x,y) = 0$$

on the unit square $\Omega = (0,1) \times (0,1)$ with the homogeneous Dirichlet boundary condition on $\partial\Omega$. For discretization we considered a uniform mesh, our matrix is of the rank 10000 and the right-hand side vector is a zero vector. We took the vector $x_0 = (1,\ldots,1)^T$ as the initial approximation.

In our numerical experiments, we used various choices of the shadow vector:

(1) $\widetilde{r}_0 = r_0$,
(2) $\widetilde{r}_0$ is chosen randomly,
(3) $\widetilde{r}_0$ is chosen as the orthogonal projection of a randomly chosen vector to space $\mathcal{K}_k(\mathbf{A}, r_k^{GMRES})$,
(4) $\widetilde{r}_0$ is chosen as the orthogonal projection of vector r_0 to space $\mathcal{Z}_{2^l}$ generated by vector $r_{2^i}^{GMRES}$ and by matrix $\mathbf{A}$, $i = 0,\ldots,l$, $k = 2^l \leq [n/2]$.

The x-axis shows the number of iterations and the y-axis the size of quantity $\|r_k\|/\|r_0\|$. For computing Lanczos' residuals and approximations we used BiCG method; see for instance [Tichý–98].

In the first figure we can see in detail the behaviour of the BiCG method with choices (3) and (4) in the first 64 iterations ($k = 64$). Choice (3) caused equality of the 64-th residuals. If we choose the shadow vector according to (4) we can observe that the BiCG behaves very sensibly, in $2^i th$-iterations are residuals $r_{2^i}^L$ equal to those of the GMRES and that, among these iterations, the curve of the BiCG is quite close to the curve of the GMRES.

The second figure shows the whole convergence curves of the GMRES method and the BiCG with various choices of the shadow vector. The first choice leads to the creation of a big peak and to a slow convergence. The second choice, a randomly chosen vector, creates quite a big peak but the convergence is then fast. By the third choice we decreased the height of the peak by conserving the speed of convergence and in the last case the big peak almost vanished. We know that the height of the peak is connected with the reduction of ultimate accuracy of the approximate solution [Green–97b] and

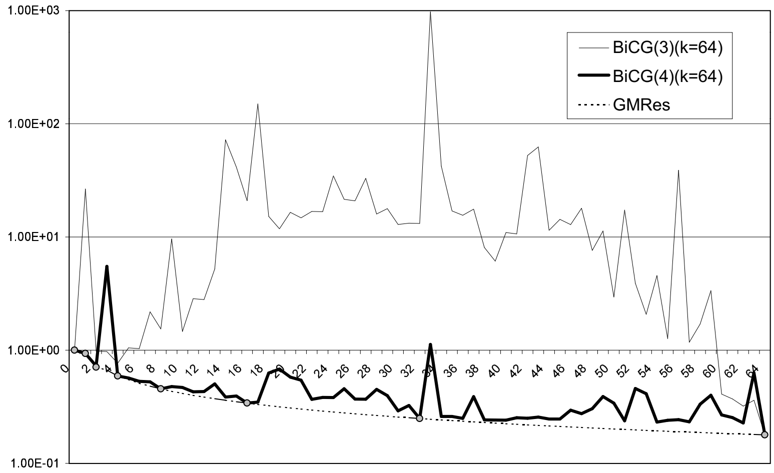


Figure 1: The BiCG and the GMRES in detail

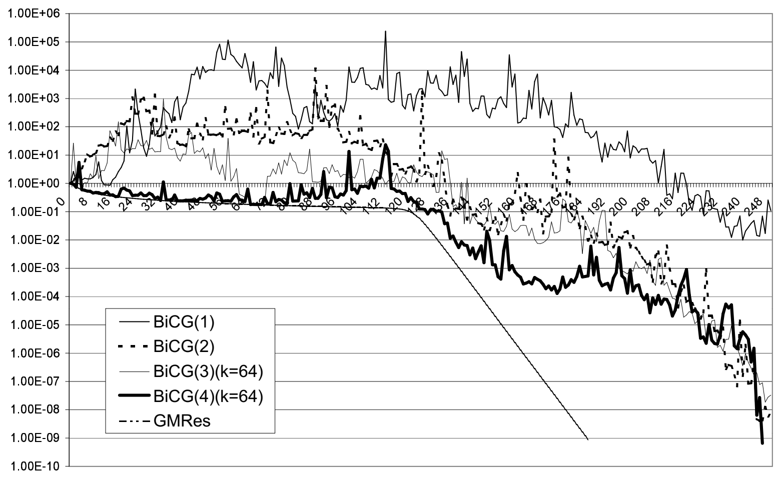


Figure 2: The whole convergence curves

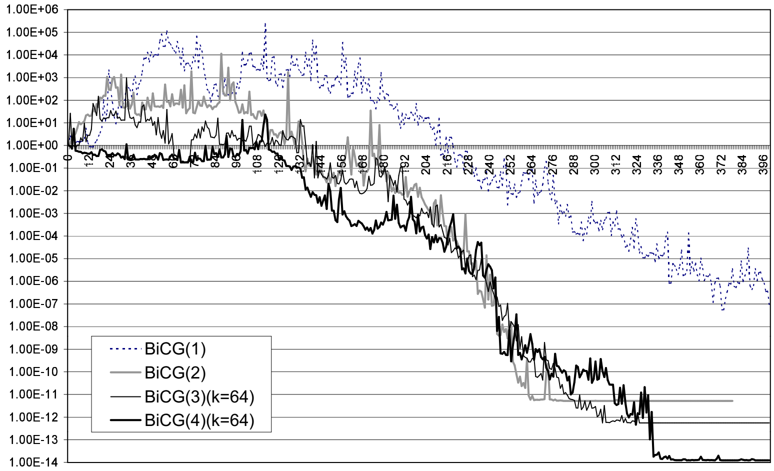


Figure 3: The level of accuracy

it means that the BiCG with the last choice computes most accurately of all choices as we can see in Figure 3 which shows the norm of the true residual $b - \mathbf{A}x_k$ divided by the norm of the initial residual r_0.

A similar behaviour, as that of the BiCG, we can observe in the last figure by the CGS method; see e.g. [Gutkn–97]. The CGS with choices (3) and (4) works very well and the number of steps is smaller than when the GMRES is used.

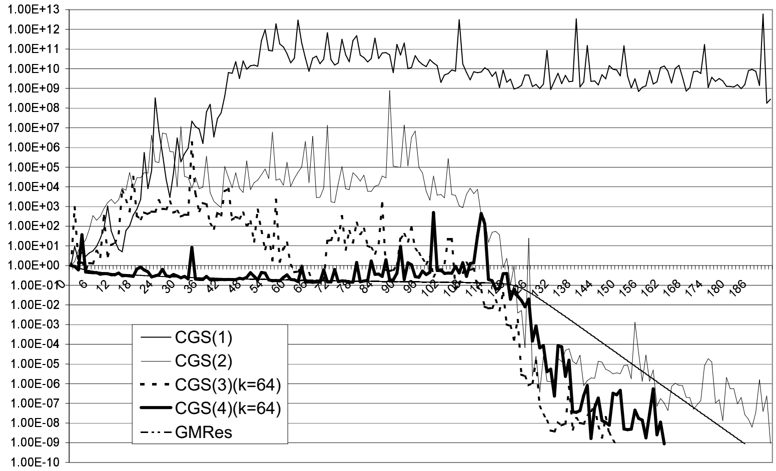


Figure 4: The CGS method and various shadow vectors

References

[Green–97a] *A. Greenbaum*: On the role of the left starting vector in the two-sided Lanczos algorithm, Proceedings of the Dundee meeting, in Numerical Analysis 1997, D.F. Griffiths, D.J. Higham, G.A. Watson, eds., Pitman Research Notes in Mathematics Series 380, Longman, 1997.

[Green–97b] *A. Greenbaum*: Estimating the attainable accuracy of recursively computed residual methods. SIAM Journal on Matrix Analysis and Applications, 18, 535–551, 1997.

[Gutkn–97] *M. H. Gutknecht*: Lanczos-type Solvers for Nonsymmetric Linear System of Equations. Technical Report TR–97–04, Swiss Center for Scientific Computing ETH-Zentrum, Switzerland, 1997.

[Tichý–98] *P. Tichý, J. Zítko*: Derivation of BiCG from the conditions defining Lanczos' method for solving a system of linear equations, Application of Mathematics 5, vol. 43 (1998), 381-388.