On prescribing residual norms in restarted GMRES

Jurjen Duintjer Tebbens

Institute of Computer Science

Academy of Sciences of the Czech Republic

joint work with

Gérard Meurant

SNA 2012, Liberec, January 23, 2012.

Outline

1. Prescribed convergence behavior for full GMRES.

2. Prescribed convergence behavior for restarted GMRES.

1. Prescribed convergence for full GMRES

Throughout we will consider a system of linear algebraic equations

$$Ax = b, \qquad A \in \mathbb{C}^{n\times n}, \quad b \in \mathbb{C}^n,$$

with a non-hermitian nonsingular matrix to be solved by the popular GMRES method. Starting with initial guess $x_0 = 0$, GMRES iterates x_k minimize the residual $r_k = b - Ax_k$,

$$\|r_k\| = \|b - Ax_k\| = \min \|b - As\| \quad \text{over all} \quad s \in \mathcal{K}_k(A, b),$$

where the Krylov subspace $\mathcal{K}_k(A, b)$ is defined as

$$\mathcal{K}_k(A, b) \equiv \text{span}\{b, Ab, \ldots, A^{k-1}b\}, \quad b \in \mathbb{C}^n, \qquad k = 1, 2, \ldots.$$

Because of the residual minimizing property, GMRES convergence curves do not increase.

In the standard implementation, an orthogonal basis for the Krylov subspace $\mathcal{K}_k(A, b)$ is computed by the Arnoldi process, yielding the Arnoldi decomposition

$$AV_k = V_{k+1}\tilde{H}_k,$$

where the columns of V_k contain the basis and $\tilde{H}_k \in \mathbb{C}^{(k+1)\times k}$ is upper Hessenberg.

1. Prescribed convergence for full GMRES

If we write the kth iterate as

$$x_k = V_k y_k \in \mathcal{K}_k(A, r_0),$$

it follows the residual is minimized through

$$\begin{aligned} \|b - Ax_k\| &= \|b - AV_k y_k\| = \|V_{k+1}\|b\|e_1 - AV_k y_k\| \\ &= \|V_{k+1}(\|b\|e_1 - \tilde{H}_k y_k)\| = \min_{y \in \mathbb{C}^k} \|\|b\|e_1 - \tilde{H}_k y\|. \end{aligned}$$

We see that the residual norm is fully determined by the Hessenberg matrix $\tilde{H}_k$ (and by $\|b\|$).

Convergence behavior of the methods is not fully understood, analysis is particularly challenging with highly non-normal input matrices. In particular, eigenvalues do not dominate convergence behavior as is the case with Krylov subspace methods for hermitian problems.

1. Prescribed convergence for full GMRES

It is known since 1994 [Greenbaum, Strakoš], that if a GMRES convergence curve is generated by some matrix and right hand side $\{A, b\}$, the same curve can be generated by a pair $\{C, d\}$ where the matrix C has arbitrary nonzero spectrum.

In 1996, Greenbaum, Pták and Strakoš complemented this result by showing that any non-increasing convergence curve is possible with any nonzero spectrum.

Finally, in 1998, Arioli, Pták and Strakoš closed this series of papers with a parametrization of the pairs $\{A, b\}$ generating arbitrary Arnoldi behavior. Here is this parametrization:

Theorem 1 [Arioli, Pták and Strakoš - 1998]. Let n complex nonzero numbers $(\lambda_1, \dots, \lambda_n)$ and n positive numbers

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

be given. Let A be a square matrix of size n and let b be a nonzero n-dimensional vector. The following assertions are equivalent:

1. Prescribed convergence for full GMRES

1. The matrix A has the eigenvalues $\lambda_1, \ldots, \lambda_n$, and the GMRES method applied to A and right-hand side b with zero initial guess yields residuals $r^{(k)},\ k = 0, \ldots, n-1$ such that

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1.$$

2. The pair $\{A, b\}$ is of the form

$$A = W \begin{bmatrix} h & \begin{matrix} R \\ 0 \end{matrix} \end{bmatrix} \begin{bmatrix} 0 & -\alpha_0 \\ I_{n-1} & \vdots \\ & -\alpha_{n-1} \end{bmatrix} \begin{bmatrix} h & \begin{matrix} R \\ 0 \end{matrix} \end{bmatrix}^{-1} W^*, \qquad b = Wh,$$

where W is a unitary matrix, R is a nonsingular upper triangular matrix of order $n-1$,

$$h = [\eta_1, \ldots, \eta_n]^T, \qquad \eta_k = \sqrt{f(k-1)^2 - f(k)^2}, \quad k < n, \quad \eta_n = f(n-1)$$

and $\alpha_0, \ldots, \alpha_{n-1}$ are the coefficients of the polynomial $q(\lambda)$ with roots $\lambda_1, \ldots, \lambda_n$,

$$q(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = \lambda^n + \sum_{j=0}^{n-1} \alpha_j \lambda^j.$$

1. Prescribed convergence for full GMRES

The previous parametrization contains the unitary matrix $W = [w_1, w_2, \ldots, w_n]$ representing the orthogonal basis of the Krylov *residual* space $A\mathcal{K}_k(A, r_0)$,

$$\text{span}\{w_1, w_2, \ldots, w_n\} = \text{span}\{Ab, A^2b, \ldots, A^n b\}.$$

From [DT, Meurant - 2012?] we easily obtain an analogue parametrization working with a unitary matrix $V = [v_1, v_2, \ldots, v_n]$ representing the orthogonal basis of the Krylov subspace $\mathcal{K}_k(A, r_0)$,

$$\text{span}\{v_1, v_2, \ldots, v_n\} = \text{span}\{b, Ab, \ldots, A^{n-1}b\} :$$

Theorem 2 [DT, Meurant - 2012?]. Let n complex nonzero numbers $(\lambda_1, \ldots, \lambda_n)$ and n positive numbers

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

be given. Let A be a square matrix of size n and let b be a nonzero n-dimensional vector. The following assertions are equivalent:

1. Prescribed convergence for full GMRES

1. The matrix A has the eigenvalues $\lambda_1, \ldots, \lambda_n,$ and the GMRES method applied to A and right-hand side b with zero initial guess yields residuals $r^{(k)},\ k = 0, \ldots, n-1$ such that

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1, .$$

2. The pair $\{A, b\}$ is of the form

$$A = V \begin{bmatrix} g^T & \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} 0 & -\alpha_0 \\ I_{n-1} & \vdots \\ & -\alpha_{n-1} \end{bmatrix} \begin{bmatrix} g^T & \\ 0 & T \end{bmatrix} V^*, \quad b = f(0) V e_1,$$

where V is a unitary matrix, T is nonsingular upper triangular of size $n-1$,

$$g_1 = \frac{1}{f(0)}, \qquad g_k = \frac{\sqrt{f(k-2)^2 - f(k-1)^2}}{f(k-2)f(k-1)}, \qquad k = 2, \ldots, n$$

and $\alpha_0, \ldots, \alpha_{n-1}$ are the coefficients of the polynomial $q(\lambda)$ with roots $\lambda_1, \ldots, \lambda_n$,

$$q(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = \lambda^n + \sum_{j=0}^{n-1} \alpha_j \lambda^j.$$

1. Prescribed convergence for full GMRES

This parametrization shows what the Hessenberg matrix

$$H_n = \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}^{-1} \begin{bmatrix} 0 & -\alpha_0 \\ I_{n-1} & \vdots \\ & -\alpha_{n-1} \end{bmatrix} \begin{bmatrix} & g^T & \\ 0 & & T \end{bmatrix}$$

generated in the standard implementation with an orthogonal basis for $\mathcal{K}_n(A,b)$ looks like.

We will use this parametrization to study the Hessenberg matrices generated during subsequent cycles of restarted GMRES.

Prescribing residual norms in restarted GMRES was already considered in the paper [Vecharinsky, Langou - 2011].

2. Prescribed convergence for restarted GMRES

The paper [Vecharinsky, Langou - 2011] assumes a rather special situation in GMRES(m) (GMRES restarted after every mth iteration):

1. During every restart cycle, all residual norms stagnate except for the very last iteration inside the cycle.
2. In this very last iteration it is assumed that the residual norm is strictly decreasing.

Theorem 4 [Vecharinsky, Langou - 2011]. With the assumptions 1. and 2. above, let the very last residual norm at the end of the kth cycle be denoted by $\|\bar{r}_k\|$. If $km < n$, there exists a matrix of order n with a right hand side such that the residual norms after $m-1$ stagnating iterations in every cycle,

$$\|\bar{r}_0\|, \|\bar{r}_1\|, \ldots, \|\bar{r}_k\|$$

generated by GMRES(m) applied to the corresponding linear system, can assume any strictly decreasing nonnegative values. Moreover, it can have arbitrary nonzero eigenvalues.

2. Prescribed convergence for restarted GMRES

In this talk we try to generalize the result of Vecharinsky and Langou. First we wish to

1. eliminate the condition that during every restart cycle, all residual norms stagnate except for the very last iteration inside the cycle, i.e. we wish to prescribe residual norms inside restart cycles,
2. but we keep the condition that the residual norm in the last iteration of every cycle is strictly decreasing.

For the moment we focus on prescribing residual norms in the initial and in the second restart cycle. Let their residuals be denoted as

$$\begin{aligned} r_0^{(1)} &= b, r_1^{(1)}, \quad \dots \quad, r_m^{(1)}, \\ r_0^{(2)} &= r_m^{(1)}, r_1^{(2)}, \quad \dots \quad, r_m^{(2)}. \end{aligned}$$

2. Prescribed convergence for restarted GMRES

The system matrix A and right hand side b will be written as

$$A = VHV^*, \qquad b = \|b\| V e_1,$$

where H is unreduced upper Hessenberg and V is unitary. We will investigate what entries H must have in order to prescribe the behavior of restarted GMRES.

The m iterations of the initial cycle will give the Arnoldi decomposition

$$AV_m^{(1)} \quad = \quad V_{m+1}^{(1)} \hat{H}_m^{(1)}, \quad \text{where} \quad V_{m+1}^{(1)} = V \begin{bmatrix} I_{m+1} \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} \hat{H}_m^{(1)} \\ 0 \end{bmatrix} = H \begin{bmatrix} I_m \\ 0 \end{bmatrix}.$$

The m iterations of the second cycle will give the Arnoldi decomposition

$$AV_m^{(2)} \quad = \quad V_{m+1}^{(2)} \hat{H}_m^{(2)}, \quad \text{where} \quad V_{m+1}^{(2)^*} V_{m+1}^{(2)} = I_{m+1}, \qquad V_{m+1}^{(2)} e_1 = \frac{r_m^{(1)}}{\|r_m^{(1)}\|} \equiv V_{m+1}^{(1)} z^{(1)}.$$

It can be proved easily that

$$z^{(1)} = \left(I_{m+1} - \hat{H}_m^{(1)} (\hat{H}_m^{(1)})^\dagger \right) e_1 / \left\| \left(I_{m+1} - \hat{H}_m^{(1)} (\hat{H}_m^{(1)})^\dagger \right) e_1 \right\|.$$

2. Prescribed convergence for restarted GMRES

Lemma 1 The Hessenberg matrix $\hat{H}_m^{(2)}$ is the Hessenberg matrix generated by m iterations of the Arnoldi process with input matrix H and initial vector $\begin{bmatrix} z^{(1)T} & 0 \end{bmatrix}^T$, i.e.

$$HZ_m = Z_{m+1}\hat{H}_m^{(2)}, \qquad \text{where} \quad Z_{m+1}e_1 = \begin{bmatrix} z^{(1)} \\ 0 \end{bmatrix}, \qquad Z_{m+1}^* Z_{m+1} = I_{m+1}. \tag{1}$$

Knowing that the columns $1, \ldots, m$ of H are

$$H \begin{bmatrix} I_m \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{H}_m^{(1)} \\ 0 \end{bmatrix},$$

can we construct the columns $m+1, m+2, \ldots$ of H such that (1) is satisfied with a prescribed Hessenberg matrix $\hat{H}_m^{(2)}$?

This will depend on the number of non-zeroes in $\begin{bmatrix} z^{(1)} \\ 0 \end{bmatrix}$!

2. Prescribed convergence for restarted GMRES

Lemma 2 Let $r_m^{(1)} = V_{m+1}^{(1)} z^{(1)}$. Then we have

$$e_i^T z^{(1)} = 0, \qquad i = m+3-j, \ldots, m+1$$

for an integer j if and only if

$$\|r_0^{(1)}\| \geq \|r_1^{(1)}\| \geq \cdots \geq \|r_{m-j}^{(1)}\| > \|r_{m-j+1}^{(1)}\| = \cdots = \|r_m^{(1)}\|.$$

Hence with our assumption that the residual norm in the last iteration of every cycle is strictly decreasing, we always have $e_{m+1}^T z^{(1)} \neq 0$. By choosing appropriately the columns $m+1, m+2, \ldots$ of H such that

$$HZ_m = Z_{m+1} \hat{H}_m^{(2)}, \qquad \text{where} \quad Z_{m+1} e_1 = \begin{bmatrix} z^{(1)} \\ 0 \end{bmatrix}, \qquad Z_{m+1}^* Z_{m+1} = I_{m+1},$$

we can force the Hessenberg matrix of the second cycle $\hat{H}_m^{(2)}$ to have prescribed entries.

2. Prescribed convergence for restarted GMRES

Theorem 5 [DT, Meurant - 2012?]. Let $A \in \mathbb{C}^{n \times n}$ be a matrix, $b \in \mathbb{C}^n$ be a nonzero vector and $\hat{H}_m^{(1)}, \hat{H}_m^{(2)} \in \mathbb{C}^{(m+1) \times m}$ be unreduced upper Hessenberg matrices with positive subdiagonal. The following assertions are equivalent.

1. The initial cycle of GMRES(m) applied to A and b does not stagnate in its last iteration and generates the Hessenberg matrix $\hat{H}_m^{(1)}$ and the second cycle generates the Hessenberg matrix $\hat{H}_m^{(2)}$.

2. The matrix A and the vector b have the form

$$A = VHV^*, \qquad b = \|b\|Ve_1,$$

where V is unitary and H is upper Hessenberg and its first $2m$ columns are:

2. Prescribed convergence for restarted GMRES

$$H \begin{bmatrix} I_{2m} \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{H}_m^{(1)} & \vdots & z^{(1)} e_1^T \hat{H}_m^{(2)} \begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \\ 0 & \hat{h}^{(2)} & \\ 0 & 0 & \begin{bmatrix} 0 & I_m \end{bmatrix} \hat{H}_m^{(2)} \begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \\ 0 & 0 & 0 \end{bmatrix},$$

with the vector $z^{(1)}$ being

$$z^{(1)} = \left(I_{m+1} - \hat{H}_m^{(1)} (\hat{H}_m^{(1)})^\dagger \right) e_1 / \left\| \left(I_{m+1} - \hat{H}_m^{(1)} (\hat{H}_m^{(1)})^\dagger \right) e_1 \right\|$$

and the vector $\hat{h}^{(2)} = [\hat{h}_1^{(2)}, \ldots, \hat{h}_{m+2}^{(2)}]^T$ having the entries

$$\begin{bmatrix} \hat{h}_1^{(2)} \\ \vdots \\ \hat{h}_{m+1}^{(2)} \end{bmatrix} = \frac{1}{z_{m+1}^{(1)}} \left(h_{1,1}^{(2)} z^{(1)} - \hat{H}_m^{(1)} \begin{bmatrix} z_1^{(1)} \\ \vdots \\ z_m^{(1)} \end{bmatrix} \right), \quad \hat{h}_{m+2}^{(2)} = \frac{h_{2,1}^{(2)}}{z_{m+1}^{(1)}}.$$

2. Prescribed convergence for restarted GMRES

This result can be easily generalized for k restart cycles, as long as $k \cdot m < n$.

Theorem 6 [DT, Meurant - 2012?]. Let $A \in \mathbb{C}^{n \times n}$ be a matrix, $b \in \mathbb{C}^n$ be a nonzero vector and let for $k \cdot m < n$,

$$\hat{H}_m^{(1)}, \ldots, \hat{H}_m^{(k)} \in \mathbb{C}^{(m+1)\times m}$$

be k unreduced upper Hessenberg matrices with positive subdiagonal. The following assertions are equivalent.

1. The kth cycle of GMRES(m) applied to A and b does not stagnate in its last iteration and generates the Hessenberg matrix $\hat{H}_m^{(k)}$.
2. The matrix A and the vector b have the form

$$A = VHV^*, \qquad b = \|b\| V e_1,$$

where V is unitary, H is upper Hessenberg and with the vectors $z^{(i)}$ defined through

$$z^{(i)} = \left(I_{m+1} - \hat{H}_m^{(i)} (\hat{H}_m^{(i)})^\dagger \right) e_1 / \left\| \left(I_{m+1} - \hat{H}_m^{(i)} (\hat{H}_m^{(i)})^\dagger \right) e_1 \right\|, \quad i = 1, \ldots, k-1,$$

its columns $(k-1)m + 1$ till km are:

2. Prescribed convergence for restarted GMRES

$$H\left[e_{(k-1)m+1},\ldots,e_{km}\right] = \begin{bmatrix} & (\prod_{i=2}^{k-1} z_1^{(i)}) z^{(1)} e_1^T \hat{H}_m^{(k)} \\ & \vdots \\ & z_1^{(k-1)} z^{(k-2)} e_1^T \hat{H}_m^{(k)} \\ \hat{h}^{(k)} & z^{(k-1)} e_1^T \hat{H}_m^{(k)} \begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \\ 0 & \begin{bmatrix} 0 & I_m \end{bmatrix} \hat{H}_m^{(k)} \begin{bmatrix} 0 \\ I_{m-1} \end{bmatrix} \\ 0 & 0 \end{bmatrix}, \qquad \text{where}$$

$$\hat{h}^{(k)} = [\hat{h}_1^{(k)},\ldots,\hat{h}_{m+1}^{(k)}]^T = \frac{1}{z_{m+1}^{(k-1)}}\left(h_{1,1}^{(k)} z^{(k-1)} - \hat{H}_m^{(k-1)}[z_1^{(k-1)},\ldots,z_m^{(k-1)}]^T\right), \quad \hat{h}_{m+2}^{(k)} = \frac{h_{2,1}^{(k)}}{z_{m+1}^{(k-1)}}.$$

2. Prescribed convergence for restarted GMRES

Thus we know how to prescribe *all* the entries of the Hessenberg matrices generated at *all* restarts (as long as $km < n$ and no cycle stagnates in its last iteration). In particular, we can prescribe GMRES residual norms inside each cycle.

Corollary The residual vectors for the kth restart cycle $r_0^{(k)}, \ldots, r_m^{(k)}$ satisfy

$$\|r_i^{(k)}\| = f(i),\ i = 0, \ldots, m$$

if and only if the initial residual of the cycle has norm $f(0)$ and the generated Hessenberg matrix is of the form

$$\hat{H}_m^{(k)} = \begin{bmatrix} g_1^{(k)} & \cdots & g_{m+1}^{(k)} \\ 0 & T_m^{(k)} & \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I_m \end{bmatrix} \begin{bmatrix} g_1^{(k)} & \cdots & g_m^{(k)} \\ 0 & T_{m-1}^{(k)} & \end{bmatrix} \in \mathbb{C}^{(m+1)\times m}$$

for a nonsingular upper triangular matrix $T_m^{(k)}$ of size m with leading principal submatrix $T_{m-1}^{(k)}$ of size $m-1$ and

$$g_1^{(k)} = \frac{1}{f(0)}, \qquad g_i^{(k)} = \frac{\sqrt{f(i-2)^2 - f(i-1)^2}}{f(i-2)f(i-1)}, \qquad i = 2, \ldots, m.$$

Remark: Note that prescribing k restarts under the condition $km < n$ means that in the parametrization of the matrix A and the vector b,

$$A = VHV^*, \qquad b = \|b\|Ve_1,$$

we prescribe km residual norms and put conditions on the first km columns of H only. In particular, the last column can be chosen arbitrarily. It can be easily checked, that any nonzero spectrum of A is possible with an appropriate choice of the last column.

Now we come to the case of stagnation at the end of the cycles. Recall that with

$$\|r_0^{(1)}\| \geq \|r_1^{(1)}\| \geq \cdots \geq \|r_{m-j}^{(1)}\| > \|r_{m-j+1}^{(1)}\| = \cdots = \|r_m^{(1)}\|,$$

we have

$$e_i^T z^{(1)} = 0, \qquad i = m+3-j, \ldots, m+1$$

and that the Hessenberg matrix $\hat{H}_m^{(2)}$ of the second cycle satisfies

$$HZ_m = Z_{m+1}\hat{H}_m^{(2)}, \qquad \text{where} \quad Z_{m+1}e_1 = \begin{bmatrix} z^{(1)} \\ 0 \end{bmatrix}, \qquad Z_{m+1}^* Z_{m+1} = I_{m+1}.$$

2. Prescribed convergence for restarted GMRES

Therefore, with $j-1$ stagnation steps at the end of the first restart cycle,

- the first $j-1$ columns of the Hessenberg matrix of the second cycle $\hat{H}_m^{(2)}$ are fully determined by $\hat{H}_m^{(1)}$ and $z^{(1)}$ - they cannot be prescribed.
- We can also prove that the first $j-1$ columns of the Hessenberg matrix of the second cycle have the first row zero, i.e. they correspond to iterations with stagnation!

Corollary If the residual norms in the initial cycle satisfy

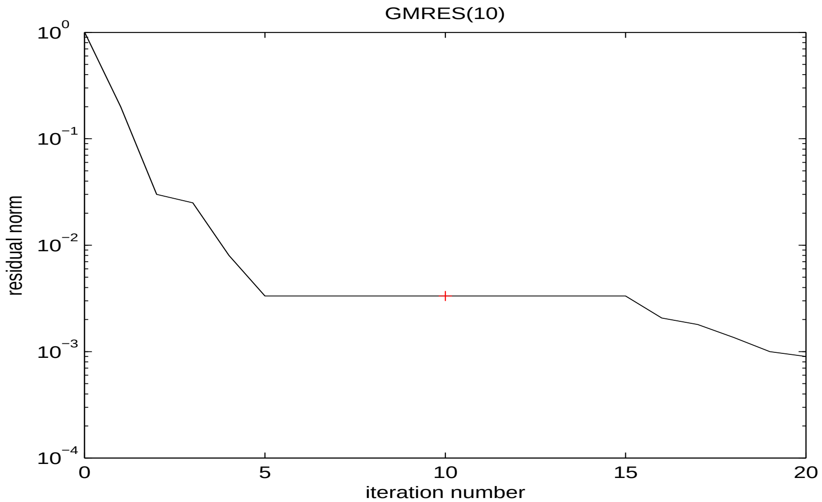
$$\|r_0^{(1)}\| \geq \|r_1^{(1)}\| \geq \cdots \geq \|r_{m-j}^{(1)}\| > \|r_{m-j+1}^{(1)}\| = \cdots = \|r_m^{(1)}\|$$

for an integer j, then the residual norms in the next cycle satisfy

$$\|r_0^{(2)}\| = \|r_1^{(2)}\| = \cdots = \|r_{j-1}^{(2)}\|.$$

2. Prescribed convergence for restarted GMRES

Hence stagnation in one cycle is literally mirrored in the next cycle, for example:



$\hat{H}_m^{(1)} \quad | \quad \hat{H}_{j-1}^{(2)} \quad | \quad \hat{H}_{j,m}^{(2)}$

Summarizing our results, under the assumption $km < n$, we showed that:

- Any non-increasing convergence curve is possible for restarted GMRES with any nonzero spectrum if there is no stagnation at the end of each cycle.
- With prescribed stagnation at the end of one cycle, we must prescribe stagnation at the beginning of the next cycle.

Thank you for your attention!

Supported by projects numbers M100300901 and IAA100300802 of the grant agency of the Academy of Sciences of the Czech Republic.