DOI 10.1007/s10958-017-3459-4
Journal of Mathematical Sciences, Vol. 224, No. 6, August, 2017

ITERATIVE PROCESSES IN THE KRYLOV–SONNEVELD SUBSPACES

V. P. Il'in* UDC 519.6

The paper presents a generalized block version of the Induced Dimension Reduction (IDR) methods in comparison with the Multi–Preconditioned Semi-Conjugate Direction (MPSCD) algorithms in Krylov subspaces with deflation and low-rank matrix approximation. General and individual orthogonality and variational properties of these two methodologies are analyzed. It is demonstrated, in particular, that for any sequence of Krylov subspaces with increasing dimensions there exists a sequence of the corresponding shrinking subspaces with decreasing dimensions. The main conclusion is that the IDR procedures, proposed by P. Sonneveld and other authors, are not an alternative to but a further development of the general principles of iterative processes in Krylov subspaces. Bibliography: 29 *titles.*

1. Introduction

In 1980, a new iterative method [1] for solving nonsymmetric systems of linear algebraic equations (SLAEs) was published in the proceedings of a symposium on numerical solution of Navier–Stokes problems. This Induced Dimension Reduction (IDR) method hardly received any attention. It was based on looking for approximate solutions in embedded subspaces of decreasing dimensions. Almost 30 years later, the author, P. Sonneveld, published (jointly with van Gijzen) the paper [2] devoted to a family of algorithms, which follow from his old idea. Later, a number of papers on this topic were published, see, for example, [3–9]. Some researchers proposed to call them "methods in Sonneveld subspaces," and these algorithms were considered an alternative to methods in Krylov subspaces.

It should be indicated that in 1989 P. Sonneveld proposed the conjugate gradient squared method, CGS, for solving nonsymmetric SLAEs, which was a transpose-free version of the bi-conjugate gradient algorithm BiCG, see [10, 11] and the references therein. This approach has provided for the appearance of a similar stabilized bi-conjugate gradient method BiCGStab [10], which was followed by new generalizations, in particular, BiCGStab(l) [12]. In contrast to the widely used generalized minimal residual method GMRES [10], the algorithms proposed (with different types of preconditioning) are based on using short recursions for computing residuals and other vectors, which is possible owing to biorthogonalization. Note also that in [13] a similar family of algorithms with various preconditioners is constructed. This family is obtained by changing biorthogonalization of vectors for their A-biorthogonalization. The corresponding methods, referred to as the A-BiCG (or BiCR, from Biconjugate Residual), A-CGS and A-BiCGStab, or CGR and BiCRStab, respectively, demonstrate a good stability on model problems, which is even superior to that of their prototypes with the classical biorthogonalization.

In the subsequent papers (see, e.g., [14–17] and many others), different authors have investigated various versions of the IDR(s) methods. Their connections with BiCGStab(l) have been established, and their generalization IDRStab(s,l), which reduces to known algorithms in some particular cases, has been proposed. These algorithms have been used not only for solving SLAEs, but also for computing eigenvalues and the corresponding eigenvectors. Formally, the parameter s in the latter algorithm means the dimension of the space S of "shadow residuals,"

*Institute of Computational Mathematics and Mathematical Geophysics, SB RAS and Novosibirsk State University, Novosibirsk, Russia, e-mail: ilin@sscc.ru.

Translated from *Zapiski Nauchnykh Seminarov POMI*, Vol. 453, 2016, pp. 114–130. Original article submitted November 21, 2016.

to which the Sonneveld subspaces are orthogonal. The second parameter l defines the order of auxiliary matrix polynomials, or the corresponding subspaces, which are used in minimizing the residual norm at every iteration.

The goal of this paper is, first, to provide a block representation of the IDRStab(s,l) methods and, based on it, give some known variants and compare them with the modern modifications of the classical iterative processes in Krylov subspaces. Second, we want to eliminate a misunderstanding related to contraposition of algorithms in Krylov and Sonneveld subspaces. This is done on the example of the proposed family of Multi-Preconditioned Semi-Conjugate Direction (MPSCD, see [18, 19]) methods, which use the following three approaches. The first one consists in the possibility of applying the A-biorthogonalization of vectors along with the classical biorthogonalization. The second one consists in using the multipreconditioned semi-conjugate direction algorithms (in particular, the Semi-Conjugate Residual, SCR), which extend the methods for realization of matrix polynomials minimizing the norm of the residual considered in [20, 21].

Another principal aspect of the considered MPSCD methods consists in orthogonalizing, at each iteration, the residual vectors to an a priori fixed "deflation" subspace, which actually is an analog of the "shadow residual" subspace S in the IDR methods. The idea of deflation in Krylov subspaces was first proposed in 1987 by R. Nicolaides [22]; later, it has been developed in different directions (quite independently of the induced dimension reduction methods) under the names of aggregation, augmentation, coarse grid correction, and low-rank matrix approximation, see [23–27].

The third aspect of the algorithms considered consists in applying, at every iteration, several preconditioning matrices. In application to the conjugate gradient method, this was originally suggested by R. Bridson and C. Greif in [28]. Note that in our treatment, deflation (or aggregation) is interpreted as an additional preconditioning of a special type.

The paper is organized as follows. In Sec. 2, we present a generalized block representation and a short review of the IDR methods, with an emphasis on their natural interpretation as specific algorithms in Krylov subspaces. Section 3 describes the MPSCD methods, which possess some properties in the shrink Sonneveld subspaces. They are considered in the context of classical modified deflated iterations in Krylov subspaces. In conclusion, we discuss the results obtained.

2. A block representation of the IDRStab(s,l) methods

Let a real nonsingular system of linear algebraic equations

$$Au = f, \;\; A \in \mathcal{R}^{N,N}, \;\; u, f \in \mathcal{R}^N, \tag{1}$$

must be solved by an iterative method, given an initial guess u^0 and the corresponding residual vector $r^0 = f - Au^0$. In this section, we restrict ourselves to properties of the induced dimension reduction methods only, not dwelling on the possibility of applying various preconditionings. In particular, the coefficient matrix A in (1) may be viewed as an already preconditioned one.

Assume that two subspaces $\mathcal{G}_0$ and $\mathcal{S}$, which contain no nontrivial common invariant subspaces of the matrix A, are defined by the relations

$$\begin{aligned} &\mathcal{G}_0 = \mathcal{K}_N(A, g_1) = \text{Span}\,(g_1, Ag_1, \ldots, A^{N-1}g_1), \;\; g_1 \in \mathcal{R}^N, \;\; g_1 \neq 0, \\ &\mathcal{S} = \mathcal{N}(\widetilde{R}^T), \;\; \widetilde{R} = (\widetilde{r}_1\, \widetilde{r}_2 \ldots \widetilde{r}_s) \in \mathcal{R}^{N,s}, \;\; \widetilde{r}_k \in \mathcal{R}^N, \;\; k = 1, \ldots, s. \end{aligned} \tag{2}$$

Here, g_1 is a nonzero vector, $\mathcal{K}_N$ is the full Krylov subspace, $\mathcal{S}$ is the left null space of a given rectangular matrix $\widetilde{R}$ of full rank, $s \ll N$, whose columns $\widetilde{r}_k$ form a basis of $\mathcal{S}$ and are referred to as the initial shadow residual vectors; the superscript T means transposition. In other

words, if a certain vector g belongs to $\mathcal{S}$, then it is orthogonal to all $\widetilde{r}_k$ ($g \perp \widetilde{r}_k$, or $(g, \widetilde{r}_k) = 0$, $k = 1, \ldots, s$), which can also be written as $\mathcal{S} = \widetilde{R}^{\perp}$.

The assumption made on the subspace in (2) means, in particular, that no eigenvector of the matrix A belongs to $\mathcal{G}_0$ and $\mathcal{S}$ simultaneously, i.e., any egeinvector from $\mathcal{G}_0$ is orthogonal to all the columns of the matrix $\widetilde{R}$.

Now consider matrix polynomials of the form

$$P_{l_n}(A) = \prod_{k=1}^{l_n} (\mu_k I - A), \;\; l_n \ll N, \tag{3}$$

where I is the identity matrix, l_n are the degrees of the corresponding polynomials, and μ_k are some real values, or shift parameters of the matrix A, whose choice will be discussed later. We construct a sequence of the following subspaces, called the Sonneveld subspaces:

$$\mathcal{G}_n = P_{l_n}(A)(\mathcal{G}_{n-1} \cap \mathcal{S}); \tag{4}$$

their dimensions will be denoted by d_n. Properties of these embedded subspaces provide a foundation for the iterative methods considered below, and they are based on the so-called IDR theorem, which is proved in [2] for $l_n = 1$ and in [6] for a fixed $\ell_n = \ell > 1$.

Theorem 1. *Let $A \in \mathcal{R}^{N,N}$ and $\widetilde{R} \in \mathcal{R}^{N,s}$ be a nonsingular matrix and a matrix of full rank, respectively, and let $\mathcal{G}_0$ and $\mathcal{S}$ be the subspaces defined in* (2) *that contain no eigenvalues of the matrix A; let $P_{l_n}(A)$ be a polynomial of order $l_n < s$ of the form* (3)*. Then the Sonneveld subspaces* (4) *satisfy the condition of embedding $\mathcal{G}_n \subseteq \mathcal{G}_{n-1}$ if $\mathcal{G}_{n-1} \neq \{0\}$, and the sequence of differences $d_n - d_{n-1}$ of their dimensions is monotone nonicreasing, i.e.,*

$$0 \leq d_n - d_{n+1} \leq d_{n+1} - d_{n+2} \leq s. \tag{5}$$

Remark 1. The Sonneveld subspaces are shrinking, and their dimensions decrease, i.e., $\mathcal{G}_n \subset \mathcal{G}_{n-1}$ and $d_n < d_{n-1}$ unless $\mathcal{G}_{n-1} = \{0\}$. If for $0 < n < N$ we have $\mathcal{G}_n = \mathcal{G}_{n-1} \neq \{0\}$, then $\mathcal{G}_{n-1} \cap \mathcal{S}$ contains an eigenvector of the matrix A. Actually, instead of $\mathcal{G}_0$ of the form (2) one can take an arbitrary linear subspace invariant with respect to multiplication by the matrix A ($A\mathcal{G}_0 \subset \mathcal{G}_0$).

Remark 2. The Sonneveld subspaces $\mathcal{G}_n$ (as is shown in [4]) can be represented using the following special block Krylov subspaces, expanding as n grows:

$$\mathcal{K}_n(A^T, \widetilde{R}) = \left\{ \sum_{k=0}^{n-1} (A^T)^k \widetilde{R} \bar{c}_k |\; \bar{c}_k \in \mathcal{R}^s \right\}; \tag{6}$$

here, c_k are the coefficient vectors, which form linear combinations of the columns of the matrix $\widetilde{R}$ and are determined below from the orthogonality conditions. In accordance with (2)–(4), for $n = 1$ we have

$$\mathcal{G}_n = \{P_{l_n}(A)g |\; P_{l_n}(A)g \perp \widetilde{R} \in \mathcal{K}_n(A^T, \widetilde{R})\}. \tag{7}$$

However, the orthogonality condition $P_{l_n}(A)g \perp \widetilde{R}$ is satisfied if and only if $A^k g \perp \widetilde{R}$ for all $k \leq l_n$, which amounts to the relation $g \perp (A^T)^k \widetilde{R}$, from which we derive, by induction,

$$\mathcal{G}_n = \{P_{l_n}(A)g |\; g \perp \mathcal{K}_n(A^T, \widetilde{R})\}. \tag{8}$$

The latter relation means that with every sequence of expanding block Krylov subspaces one can associate a sequence of shrinking Sonneveld subspaces.

The above preliminary geometric considerations about the Sonneveld subspaces allow one to construct a general scheme of an iterative process, in which the residual vectors $r^n = f - Au^n$

belong to $\mathcal{G}_n$. If the vectors u^{n-1} and $r^{n-1} \in \mathcal{G}_{n-1}$ are known, then an iteration of the IDRStab(s,l) method can be split into two stages, at which the vectors

$$r^{n-1/2} \in \mathcal{G}_{n-1} \cap \mathcal{S}, \quad r^n = P_{l_n}(A) r^{n-1/2} \tag{9}$$

are defined. Here, the first stage includes orthogonalization of the residual r^{n-1} to the shadow residual vectors $\widetilde{r}_k$, whereas the second one is responsible for the minimization of r^n. Observe that the conventional BiCGStab algorithm corresponds to the case $s = \ell = 1$. The orthogonalization stage can be performed by applying the numerically stable modified Gram–Schmidt method. To this end, in accordance with (9), in order to compute the vector $r^{n-1/2} \in \mathcal{G}_{n-1} \cap \mathcal{S}$, we first define $s+1$ vectors $g_{-s}, \ldots, g_0 \in \mathcal{G}_0 = \mathcal{K}_N(A, g_1)$ for $n = 0$. Assuming that this is already done, we use the representation

$$r^{-1/2} = g_0 - \sum_{k=-s}^{-1} g_k c_k,$$

where the unknown coefficients c_k are determined from the orthogonality condition

$$\widetilde{R}^T r^{-1/2} = 0, \quad \widetilde{R} \in \mathcal{R}^{N,s}. \tag{10}$$

This yields a system of ℓ linear equations

$$C c = R^T G\, c = R^T g_0, \quad G = (g_{-1} \cdots g_{-s}) \in \mathcal{R}^{N,s}, \quad c = \{c_k\} \in \mathcal{R}^s, \tag{11}$$

which has a unique solution, provided that the matrix $C = R^T G \in \mathcal{R}^{s,s}$ is nonsingular.

Remark 3. The computation of the vector c from the orthogonality condition (10) reduces to solving a least squares problem [29]. Its normal solution (with the minimal norm) can be computed from the overdetermined system $G\, c = g_0$ with the rectangular matrix G. If one of the matrices G and C is singular, then it is reasonable to use the generalized inverse matrices or, for example, the singular value decomposition (SVD).

In order to solve system (11), one must preliminarily find s vectors $g_k \in \mathcal{K}_N(A, g_1)$. To this end, any conventional algorithm in Krylov subspaces can be used. One of the common ways consists in choosing $g_0 = r^0$ and computing the subsequent g_k by applying the Arnoldi orthogonalization to the vectors $A^k g_0$.

The next stage of the algorithm is the computation of the vector $r^1 = P_{l_1}(A) r^{-1/2} \in \mathcal{G}_1$. Multiplication by a matrix polynomial in A is carried out after determining the shift parameters μ_k in (3), which will be discussed later. When the first step is completed, the second stage begins with ensuring the fulfillment of the orthogonalization condition (11). Then the computational process proceeds in a similar way. In fact, this is a two-level iterative process. The upper level corresponds to the sequence of Sonneveld subspaces, indexed by n, and the lower level corresponds to inner iterations in Krylov subspaces.

Thus, for $n = 1, 2, \ldots$, the following two kinds of operations, occurring in (9), are performed:

(a) Multiplication of the residual vector r^{n-1} from the previous Sonneveld subspace by the polynomial $P_{l_n}(A)$,

$$\widehat{r}^n = P_{l_n}(A) r^{n-1}, \quad r^{n-1} \in \mathcal{G}_{n-1} \cap \mathcal{S}.$$

(b) Solution of the SLAE for the coefficients c_k, for which the following orthogonality condition is fulfilled:

$$R^T r^n = 0, \quad r^n = \widehat{r}^n - \sum_{k=-s}^{-1} g_k c_k = \widehat{r} - G\, c.$$

The number of the basis vectors $g_k \in \mathcal{G}_n$ is not necessarily equal to s. In such a case, one should consider the generalized (least squares) solution of the equations obtained, but we do not dwell on this.

Note that here and below we use a block representation of the iterative process, different from the conventional one, in which the entire process of orthogonalizing the residual vector to all the columns of the matrix $\widetilde{R}$ is performed in one iteration.

Other interesting issues related to the IDR methods concern the choice of the matrix $\widetilde{R}$, initialization of the iterative process (i.e., the choice of the subspace $\mathcal{G}_0$ and computation of the vector $r^0 \in \mathcal{G}_0$), and also the choice of an algorithm for computing the subsequent residual vectors $r^n \in \mathcal{G}_n$. In all the algorithms considered below, we use the following natural stopping criterion:

$$||r^{n(\varepsilon)}|| \le \varepsilon ||f||, \quad \varepsilon \ll 1, \quad ||f||^2 = (f, f). \tag{12}$$

Remark 4. As was mentioned in [6], the IDR(s, l) methods are related to the Petrov–Galerkin algorithms in the following sense. Let the vectors $u^n - u^0$ and r^n belong to the Krylov subspaces $\mathcal{K}_n(A, r^0)$ and $\mathcal{K}_{n+1}(A, r^0)$, respectively (if the degree of the minimal polynomial $P_n(A)$ equals $n = N$, then $\mathcal{K}_N = \mathcal{R}^N$). Under these assumptions, the Petrov–Galerkin method with respect to the set of embedded subspaces $\{\mathcal{L}_m\}$ is defined by the condition of orthogonality of the residual vector r^n to the subspace $\mathcal{L}_m$ of dimension m, i.e., $r^n \perp \mathcal{L}_m$, or $r^n \in \mathcal{L}_m^{\perp}$. In particular, if $\mathcal{L}_m = A\mathcal{K}_n$, then this scheme includes the GMRES algorithm. Since $\mathcal{R}^N = \mathcal{K}_{n+1} \oplus \mathcal{K}_{n+1}^{\perp}$ and since the dimensions of $\mathcal{K}_{n+1}$ and $\mathcal{K}_{n+1}^{\perp}$ increase and decrease, respectively, as n grows, the orthogonal complement $\mathcal{K}_{n+1}^{\perp}$ to the Krylov subspace plays, in a sense, the role of the Sonneveld subspace.

In the papers presented in the reference list and the references therein, one can find extensive surveys and known interpretations of various IDR methods, which show that the methodology of these iterative processes is not so far completely developed.

3. Multi-preconditioned semi-conjugate direction methods

In this section, we consider application of a block semi-conjugate direction method [18, 19], which is an extension of the generalized conjugate residual (GCR) algorithm [20].

If in (2) we set $g_1 = r^0$, then the suggested family of methods MPSCD in the Krylov subspaces $\mathcal{K}_m(A, r^0)$ can be written as follows:

$$u^{m+1} = u^m + P_m \bar{\alpha}_m, \quad r^{m+1} = r^m - AP_m \bar{\alpha}_m, \quad m = 0, 1, \ldots; \tag{13}$$

here, $P_m = (p_1^m \ldots p_{M_m}^m) \in \mathcal{R}^{N, M_m}$ is the matrix composed of the direction vectors p_k^m, and $\bar{\alpha}_m = (\alpha_{m,1} \ldots \alpha_{m,M_m})^T$ is the vector of parameters, which are determined from the orthogonality conditions

$$\begin{aligned} &(Ap_k^m, A^\gamma p_{k'}^{m'}) = \rho_{m,k}^{(\gamma)} \delta_{m,m'}^{k,k'}, \quad \rho_{m,k}^{(\gamma)} = (Ap_k^m, A^\gamma p_k^m), \quad \gamma = 0, 1, ; \\ &m' = 0, 1, \ldots, m-1; \quad k, k' = 1, 2, \ldots, M_m; \end{aligned} \tag{14}$$

here, $\delta_{m,m'}^{k,k'}$ is the Kronecker symbol, which is equal to unity if $m = m'$ and $k = k'$, and to zero otherwise; the values $\gamma = 0, 1$ specify the semi-conjugate gradient and semi-conjugate residual methods, respectively. Note that in contrast to the conventional semi-conjugate direction algorithms, formulas (13) involve, at every mth iteration, M_n direction vectors rather than one, and, in general, their number may vary from iteration to iteration.

Under assumptions (14), the coefficient vectors $\bar{\alpha}_m = \{\alpha_{m,l}^{(\gamma)}\}$ in (13) for $\gamma = 0, 1$ are determined from the extremum condition

$$\partial \Phi_m^{(\gamma)} / \partial \alpha_{m,l} = 0, \quad \Phi_m^{(\gamma)}(r^{m+1}) \equiv (r^{m+1}, A^{\gamma-1} r^{m+1}), \tag{15}$$

and are given by

$$\alpha_{m,l}^{(\gamma)} = (A^\gamma B_{m,l}^{-1} r^m, r^m)/\rho_{m,l}^{(\gamma)}. \tag{16}$$

For $\gamma = 0, 1$, the direction vectors p_ℓ^m are determined from (14) in the following block form:

$$\begin{aligned}
&P_0 = \{p_l^0 = B_{0,l}^{-1} r^0\},\\
&P_{m+1} = P_{m+1,0} - \sum_{k=0}^{m} P_k \bar{\beta}_{m,k}^{(\gamma)} = \left\{ p_l^{m+1} = B_{m+1,l}^{-1} r^{m+1} - \sum_{k=o}^{m} \sum_{l=1}^{M_k} \beta_{m,k,l}^{(\gamma)} p_l^k \right\}, \quad m = 0, 1, \ldots;\\
&B_{m,l} \in \mathcal{R}^{N,N}, \quad l = 1, \ldots, M_m, \ \ \gamma = 0, 1;
\end{aligned} \tag{17}$$

here, $\bar{\beta}_{m,k}^{(\gamma)} = \{\beta_{m,k,l}^{(\gamma)}\} = (\beta_{m,k,1}^{(\gamma)} \ldots \beta_{m,k,M_m}^{(\gamma)})^T \in \mathcal{R}^{M_m}$ are the coefficient vectors, and $B_{m,l} \in \mathcal{R}^{N,N}$ are the preconditioning matrices, which should be nonsingular, easily invertible, and provide for an efficient acceleration of the resulting iterative process. Note that the preconditioners $B_{m,\ell}$ are dynamic, or flexible (as in FGMRES [10]) because, for any ℓ, they depend on the iteration number m.

On substituting (17) into the orthogonality conditions (14), we obtain the following formulas for the coefficients:

$$\begin{gathered}
\beta_{m,k,l}^{(\gamma)} = (A^\gamma p_l^k, A B_{m+1,l}^{-1} r^{m+1})/\rho_{m,l}^{(\gamma)} = (A^\gamma p_l^k, A p_l^{m,k})/\rho_{m,l}^{(\gamma)},\\
m = 0, 1, \ldots; \quad k = 0, \ldots, m; \quad l = 1, \ldots, M_m.
\end{gathered} \tag{18}$$

In (18), the vectors $p_\ell^{m,k}$ are determined from the relations

$$\begin{aligned}
&p_l^{m,k} = p_l^{m,k-1} - \beta_{m,k-1,l}^{(\gamma)} p_l^{k-1} = B_{m+1,l}^{-1} r^{m+1} - \sum_{i=0}^{k-1} \beta_{m,i,l}^{(\gamma)} p_l^m,\\
&l = 1, \ldots, M_m, \quad k = 0, 1, \ldots, m+1;\\
&p_l^{m,0} = B_{m+1,l}^{-1} r^{m+1}, \quad p_l^{m,m+1} = p_l^{m+1},
\end{aligned} \tag{19}$$

by the modified Gram–Schmidt method.

The following assertion is valid (for the proof, see [19]).

Theorem 2. *For the iterative MPSCD process defined by relations* (13) *and* (16)–(19)*, where* $\gamma = 0, 1$ *and the matrices* A *and* $B_{m,\ell}$ *are nonsingular, the following assertions hold*:

- *the direction vectors* p_k^m *satisfy the orthogonality conditions* (14);
- *the residual vectors* r^m *are generalized semi-conjugate, i.e.,*

$$(A^\gamma B_{k,l}^{-1} r^m, r^k) = \begin{cases} 0, & k < m,\\ \sigma_m^{(\gamma)} = (A^\gamma B_{m,l}^{-1} r^m, r^m), & k = m; \end{cases} \tag{20}$$

- *the functionals* $\Phi_m^{(\gamma)}(r^{m+1})$ *satisfy the extremum conditions* (15)*, and for* $q = 0, 1, \ldots, m$ *the following relations are valid*:

$$\Phi_m^{(\gamma)}(r^{m+1}) = \Phi_q^{(\gamma)}(r^q) - \sum_{k=q}^{m} \sum_{l=1}^{Mm} (A^\gamma B_{q,l}^{-1} r^q, r^q)^2/\rho_{k,l}^{(\gamma)}. \tag{21}$$

Remark 5. If $\gamma = 1$ or the matrix A is symmetric, then the method MPSCD ensures minimization of the norm $||r^{m+1}|| = (r^{m+1}, r^{m+1})^{1/2}$ in the "multi-preconditioned" Krylov subspace

$$\begin{aligned}
\mathcal{K}_{\sum_{m+1}}(r^0, A) = \mathrm{Span}\Big\{ &B_{0,1}^{-1} r^0, \ldots, B_{0,M_0}^{-1} r^0,\\
&A B_{1,1}^{-1} r^1, \ldots, A B_{1,M_1}^{-1} r^1, \ldots, A^m B_{m,1}^{-1} r^m, \ldots, A^m B_{m,M_m}^{-1} r^m \Big\},
\end{aligned} \tag{22}$$

whose dimension equals $\sum_{m+1} = M_0 + \cdots + M_m$. In semi-conjugate gradient methods, i.e., for $\gamma = 0$, the functional $\Phi_m^{(0)}(r^{m+1}) = (A^{-1}r^{m+1}, r^{m+1})$ does not, in general, attains its minimum unless A is a symmetric matrix.

The block semi-conjugate direction methods considered above use the so-called "long" recursions, i.e., all the direction vectors computed at the previous iterations, or the direction matrices $P_1, \ldots, P_m$ must be stored. In order to decrease the memory requirements, we will use the known restart procedure: upon performing a certain number of iterations, the residual vector is computed from the original equation rather than from (13), and the Krylov process is started anew with the current approximation as the initial guess.

In order to accelerate the resulting iterative process, at each restart with number n we orthogonalize the residual vector to the subspace similar to $\mathcal{S}$ in (2). However, this procedure is carried out not by the way conventional for the IDR approach but by applying deflation or aggregation, see [24] and the references therein. In this case, given a rectangular matrix $\widetilde{R}_n = (\widetilde{r}_1 \ldots \widetilde{r}_{s_n})^T \in \mathcal{R}^{N,s_n}$, $s_n \ll N$, an additional preconditioning matrix is computed,

$$B_{n,0}^{-1} = \widetilde{R}_n(\widehat{A}_n)^{-1}\widetilde{R}_n^T, \quad \widehat{A}_n = \widetilde{R}_n^T A \widetilde{R}_n \in \mathcal{R}^{s_n,s_n}. \tag{23}$$

Here, $B_{n,0}^{-1}$ is a low-rank approximation of the matrix A^{-1}, inverse to the original one. It is nonsingular if A is nonsingular, and the deflation matrix $\widetilde{R}_n$ has full rank s_n. Here, for the sake of generality, we allow the subspaces $\mathcal{S}_n$ and their dimensions s_n to depend on the restart number n. If the iteration number of the MPSCD method (13) is denoted by m_n ($m_0 = 0$), where each iteration includes a restart and orthogonalization, then the resulting algorithm for $n = 0$ can be written as follows:

$$\begin{aligned} &u^0 = u^{-1} + B_{0,0}^{-1}r^{-1}, \quad r^{-1} = f - Au^{-1}, \\ &r^0 = f - Au^0, \quad p_l^0 = (I - B_{0,0}^{-1}A)B_{0,l}^{-1}r^0, \quad l = 1, \ldots, M_0. \end{aligned} \tag{24}$$

In this case, u^{-1} denotes an arbitrary initial guess, and u^0 is the "corrected" approximation of the solution to be sought. As is readily verified, from (24) it follows that

$$\widetilde{R}_0^T r^0 = 0, \quad \widetilde{R}_0^T A P_0 = 0, \tag{25}$$

where the matrix P_0 consists of the M_0 columns p_ℓ^0, defined in (24).

Realization of formulas (24) is sufficiently simple because multiplication by $B_{0,0}^{-1} \in \mathcal{R}^{N,N}$ requires solution of an auxiliary SLAE with the matrix $\widehat{A}_0 = \widetilde{R}_0^T A \widetilde{R}_0 \in \mathcal{R}^{s_0,s_0}$ of small dimension. Note that in order to emphasize the similarity with the methods in Sonneveld subspaces considered in Sec. 2, in (23) we preserve the notation for $\widetilde{R}_n$.

In order to find the linearly independent columns $\widetilde{r}_k$, $k = 1, \ldots, s_n$, of the matrix $\widetilde{R}_n$, the simplest piecewise-constant basis vectors can be used. To this end, the set of indices $\Omega = \{i = 1, \ldots, N\}$ is partitioned into approximately equal disjoint subsets $\Omega_k^{(n)}$ (these partitions can be different at different iterations), and we set $\widetilde{r}_k(i) = 1$ for $i \in \Omega_k^{(n)}$ and $\widetilde{r}_k(i) = 0$ for $i \notin \Omega_k^{(n)}$, $k = i, \ldots, s_n$. Observe that the basis vectors $\widetilde{r}_k$ can be chosen smoother, as is done in [26]. In some papers, the columns of the matrix $\widetilde{R}_n$ are chosen randomly or by solving an auxiliary eigenproblem. The problem of selecting the most informative basis is the subject of a separate research.

If at each restart the approximate solution u^{m_n-1} is known, then formulas (24)–(25) obviously take the following form:

$$\begin{aligned} &u^{m_n} = u^{m_n-1} + B_{m_n,0}^{-1}r^{m_n-1}, \quad r^{m_n-1} = f - Au^{m_n-1}, \\ &r^{m_n} = f - Au^{m_n}, \quad p_l^{m_n} = (I - B_{m_n,0}^{-1}A)B_{m_n,l}^{-1}r^{m_n}. \end{aligned} \tag{26}$$

In this case, the preconditioners $B_{m_n,l}$ are defined by (23), where n is changed for m_n, and the orthogonality conditions are replaced by

$$\widetilde{R}_{m_n}^T r^{m_n} = 0, \quad \widetilde{R}_{m_n}^T A P_{m_n} = 0. \tag{27}$$

Remark 6. It can be straightforwardly verified that

$$(C_1^{(n)})^2 \equiv (B_{m_n,0}^{-1}A)^2 = C_1^{(n)}, \quad (C_2^{(n)})^2 \equiv (I - B_{m_n,0}^{-1}A)^2 = C_2^{(n)},$$

which means that the matrices $C_1^{(n)}$ and $C_2^{(n)}$ are projectors.

For the orthogonality conditions (27) to be fulfilled not only at points of restarts but at every iteration, formulas (17) for the direction vectors in MPSCD for $m \neq m_n$ must be changed for the following ones:

$$\begin{aligned} &p_l^{m+1} = \widetilde{B}_{m+1,l}^{-1} r^{m+1} - \sum_{k=0}^{m} \sum_{l=1}^{M_k} \beta_{m,k,l}^{(\gamma)} p_l^k, \quad \widetilde{B}_{m+1,l}^{-1} = (I - B_{m+1,0}^{-1}A) B_{m+1,l}^{-1}, \\ &l = 1, \ldots, M_m, \;\; m = m_{n-1}, \;\; m_{n-1}+1, \ldots, m_n - 1. \end{aligned} \tag{28}$$

In this case, as is readily verified by induction, the following relations are valid:

$$\widetilde{R}_m^T r^m = 0, \quad \widetilde{R}_m^T A P_m = 0, \quad m = 0, 1, \ldots. \tag{29}$$

However, here one should bear in mind that relations (28), which formally differ from (17) in the preconditioning matrices only ($\widetilde{B}_{m+1,l}^{-1}$ instead of $B_{m+1,l}^{-1}$), require that in order to preserve the orthogonality of the direction vectors (14), the formulas for the coefficients $\beta_{m,k.l}^{(\gamma)}$ in relations (18)–(19) be modified. Namely, the matrices $B_{m+1,l}^{-1}$ must be changed for $\widetilde{B}_{m+1,l}^{-1}$.

As is known from theoretical estimation and numerical results, restarts lead to reducing the dimensions of the Krylov subspaces and, as a consequence, to slowing down the convergence of an iterative process. This is an inevitable price for saving memory in all algorithms with short recursions.

In the above-considered MPSCD methods with deflational restarts, it is natural to choose equidistant iteration numbers m_n, where $m_n = m_{n-1} + \varkappa$, the integer $\varkappa$ being fixed a priori. On the other hand, it is possible to consider the problem of optimally choosing the values m_n based on some a posteriori data.

Remark 7. As has been mentioned in Remark 4, to every sequence $\mathcal{K}_{n+1}$ of Krylov subspaces, whose dimensions increase as n grows, a sequence of shrinking subspaces correspond, and the latter are the orthogonal complements $\mathcal{K}_{n+1}^{\perp}$ in the space $\mathcal{R}^N$. It is readily seen that modifications of the Krylov subspaces, which consist in additional orthogonalization of the computed vectors to the "trial" subspace $\mathcal{S}$, do not influence the duality relations. The orthogonality conditions (25), (27), and (29) naturally allow one to define, for MPSCD, subspaces with decreasing dimensions similar to the subspaces $\mathcal{G}_n$ in the IDR methods.

An alternative way to save memory and to shorten recursions is the approach with limited orthogonality in the sense that in constructing iterations for solving nonsymmetric SLAEs, only a few last direction vectors in Krylov subspaces are stored and used in recursions. If the number of such vectors is constant and equals $q + 1$, then formulas (28) take the following form:

$$p_l^{m+1} = \widetilde{B}_{m+1,l}^{-1} r^{m+1} - \sum_{k=m-q}^{m} \sum_{l=1}^{M_k} \beta_{m,k,l}^{(\gamma)} p_l^k, \quad m = 0, 1, \ldots. \tag{30}$$

If, in this case, the vectors p_l^0 are computed by formulas (24), then the orthogonality properties (29) remain valid. In this approach, further orthogonalization can be carried out using deflation or aggregation.

4. Conclusion

The aim of this paper is to dethrone, in a sense, the myth about the exclusivity of induced dimension reduction methods, which are opposed, by misunderstanding, to the classical iterative processes in Krylov subspaces. On the example of multi-preconditioned semi-conjugate residual methods proposed, it is shown that modified Krylov type algorithms using additional orthogonalization (deflation, aggregation, etc.) can naturally be associated with shrinking subspaces of decreasing dimensions, which are referred to by some authors as the Sonneveld subspaces. In addition, between the subspaces of "shadow residuals" in the IDR methods and the deflation subspaces in the "classical" Krylov methods there is a close analogy, which is not mentioned by the authors of various publications. Moreover, publications on these two approaches surprisingly include no cross references, although these areas of computational algebra are intimately related.

In general, active discussions on various approaches to optimization of iterative processes confirm good prospects of developing iterative methods in Krylov subspaces.

This work was supported by the Russian Science Foundation (grant No. 14-11-00485) and the RFBR (grant No. 16-29-15122).

Translated by V. P. Il'in.

REFERENCES

1. P. Wesseling and P. Sonneveld, "Numerical experiments with a multiple grid and a preconditioned Lanczos type method," *Lect. Notes Math.*, **771**, 543–562 (1980).
2. P. Sonneveld and M. B. Van Gijzen, "IDR(s): a family of simple and fast algorithms for solving large nonsymmetric systems of linear equations," *SIAM J. Sci. Comput.*, **31**, No. 2, 1035–1062 (2008/09).
3. M. H. Gutknecht, "IDR explained," *Electron. Trans. Numer. Anal.*, **36**, 126–148 (2009/10).
4. Y. Onoue, S. Fujino, and N. Nakashima, "An overview of a family of new iterative methods based on IDR theorem and its estimation," in: *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2009, Vol. 11, IMECS, Hong Kong (2009), pp. 2129–2134.
5. V. Simoncini and D. B. Szyld, "Interpreting IDR as a Petrov–Galerkin method," *SIAM J. Sci. Comput.*, **32**, No. 4, 1898–1912 (2010).
6. M. Tanio and M. Sugihara, "GBi-CGSTAB(s,l): IDR(s) with higher-order stabilization polynomials," *J. Comput. Appl. Math.*, **235**, No. 3, 765–784 (2010).
7. G. L. Sleijpen, P. Sonneveld, and M. B. Van Gijzen, "Bi-CGStab as an induced dimension reduction method," *Appl. Numer. Math.*, **60**, No. 11, 1100–1114 (2010).
8. G. L. Sleijpen and M. B. Van Gijzen, "Exploiting BiCGStab(l) strategies to induce dimension reduction," *SIAM J. Sci. Comput.*, **32** (5), 2687–2709 (2010).
9. P. Sonneveld, "On the convergence behavior of IDR(s) and related methods," *SIAM J. Sci. Comput.*, **34** (5), 2576–2598 (2012).
10. Y. Saad, *Iterative Methods for Sparse Linear Systems* (Second edition), Society for Industrial and Applied Mathematics (2003).
11. V. P. Il'in, *Methods and Technologies of Finite Elements* [in Russian], Izd. IVMiMG SO RAN, Novosibirsk (2007).
12. G. L. Sleijpen and D. R. Fokkema, "BiCGStab(l) for linear equations involving unsymmetric matrices with complex spectrum," *Electron. Trans. Numer. Anal.*, **1**, 11–32 (1993).

13. V. P. Il'in, "Biconjugate direction methods in Krylov subspaces," *Sib. Zh. Ind. Mat.*, **11**, No. 4, 47–60 (2008).
14. M. B. Van Gijzen, G. I. Sleijpen, and J. -P. Zemke, "Flexible and multi-shift induced dimension reduction algorithms for solving large sparse linear systems," *Numer. Linear Algebra Appl.*, **22**, 1–25 (2014).
15. M. B. Van Gijzen and P. Sonneveld, "Algorithm 913: an elegant IDR(s) variant that efficiently exploits biorthogonality properties," *ACM Trans. Math. Software*, **38**, No. 1, 5.1–5.19 (2011).
16. O. Rendel, A. Rizvanolli, and J.-P. Zemke, "IDR: A new generation of Krylov subspace methods," *Linear Algebra Appl.*, **439**, No. 4, 1040–1061 (2013).
17. R. Astudillo and M. B. Van Gijzen, "A restarted induced dimension reduction method to approximate eigenpairs of large unsymmetric matrices," *J. Comput. Appl. Math.*, **296**, 24–35 (2016).
18. V. P. Il'in, "Methods of semiconjugate directions," *Russ. J. Numer. Anal. Math. Model.*, **23**, No. 4, 369–387 (2008).
19. V. P. Il'in and E. A. Itskovich, "On the semi-conjugate direction methods with dynamic preconditioning," *J. Appl. Ind. Math.*, **3**, No. 2, 222–233 (2009).
20. S. C. Eisenstat, H. C. Elman, and M. H. Schultz, "Variational iterative methods for nonsymmetric systems of linear equations," *SIAM J. Numer. Anal.*, **20**, No. 3, 345–357 (1983).
21. J. Y. Yuan, G. H. Golub, R. J. Plemmons, and W. A. Cecilio, "Semi-conjugate direction methods for real positive definite systems," *BIT*, **44**, No. 1, 189–207 (2004).
22. R. Nicolaides, "Deflation of conjugate gradients with applications to boundary value problems," *SIAM J. Numer. Anal.*, **24**, 355–365 (1987).
23. L. Yu. Kolotilina, "Twofold deflation preconditioning of linear algebraic systems. I. Theory," *Zap. Nauchn. Semin. POMI*, **229**, 95–152 (1995).
24. O. Coulaud, L. Giraud, P. Ramet, and X. Vasseur, *Deflation and augmentation techniques in Krylov subspace methods for the solution of linear systems*, INRIA, Bordeaux-Sud Ouest, France, Res. Rep. No. 8265 (2013).
25. M. H. Gutknecht, "Deflated and augmented Krylov subspace methods: A framework for deflated BICG and related solvers," *SIAM J. Matrix Anal. Appl.*, **35**, No. 4, 1444–1466 (2014).
26. Y. L. Gurieva and V. P. Il'in, "Parallel approaches and technologies of domain decomposition methods," *Zap. Nauchn. Semin. POMI*, **428**, 89–106 (2014).
27. V. P. Il'in, "Problems of parallel solution of large systems of linear algebraic equations," *Zap. Nauchn. Semin. POMI*, **439**, 112–127 (2015).
28. R. Bridson and C. Greif, "A multipreconditioned conjugate gradient algorithm," *SIAM J. Matrix Anal. Appl.*, **27**, No. 4, 1056–1068 (2006).
29. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* [Russian translation], Nauka, Moscow (1986).