The Block Grade of a Block Krylov Space

Martin H. Gutknecht and Thomas Schmelzer[†]

[†]Winton Capital Management, Magdalen Centre, The Oxford Science Park, Oxford
OX4 4GA, UK; Email: thomas.schmelzer@gmail.com

The Block Grade of a Block Krylov Space

Martin H. Gutknecht and Thomas Schmelzer[†]

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Abstract

The aim of the paper is to compile and compare basic theoretical facts on Krylov subspaces and block Krylov subspaces. Many Krylov (sub)space methods for solving a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ have the property that in exact computer arithmetic the true solution is found after ν iterations, where ν is the dimension of the largest Krylov subspace generated by $\mathbf{A}$ from $\mathbf{r}_0$, the residual of the initial approximation $\mathbf{x}_0$. This dimension is called the grade of $\mathbf{r}_0$ with respect to $\mathbf{A}$. Though the structure of block Krylov subspaces is more complicated than that of ordinary Krylov subspaces, we introduce here a block grade for which an analogous statement holds when block Krylov space methods are applied to linear systems with multiple, say s, right-hand sides. In this case, the s initial residuals are bundled into a matrix $\mathbf{R}_0$ with s columns. The possibility of linear dependence among columns of the block Krylov matrix ($\mathbf{R}_0$ $\mathbf{A}\mathbf{R}_0$ $\ldots$ $\mathbf{A}^{\nu-1}\mathbf{R}_0$), which in practical algorithms calls for the deletion (or, deflation) of some columns, requires extra care. Relations between grade and block grade are also established, as well as relations to the corresponding notions of a minimal polynomial and its companion matrix.

Keywords: sparse linear systems, multiple right-hand sides, several right-hand sides, block Krylov space method, block Krylov space solver, block size reduction, deflation, grade, minimal polynomial

Subject Classification: 65F10; 15A03, 65F15.

[†]Winton Capital Management, Magdalen Centre, The Oxford Science Park, Oxford OX4 4GA, UK; Email: `thomas.schmelzer@gmail.com`

1 Krylov spaces and their grade

Most currently used iterative methods for solving nonsingular linear systems of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ are Krylov space solvers, which generate a sequence of *approximate solutions* $\mathbf{x}_n$, also called *iterates*, which are chosen from an affine space that grows with n:

$$\mathbf{x}_n - \mathbf{x}_0 \in \mathcal{K}_n := \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) := \mathsf{span}\,(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0)\,, \qquad (1)$$

where $\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$ is the *initial residual* and $\mathcal{K}_n$ is the *n*th *Krylov subspace* generated by $\mathbf{A}$ from $\mathbf{r}_0$. There are some Krylov space solvers, where $\mathbf{x}_n$ may not exist for some exceptional values of n: the iteration may break down temporarily, but there may be a chance to recover, like in the biconjugate gradient method with the option of applying look-ahead [10, 11], or when the conjugate gradient method is applied to symmetric indefinite matrices [6]. But even in this case many of the results we cite here remain true as they are results about subspaces, not special methods.

We assume $\mathbf{A} \in \mathbb{C}^{N \times N}$ nonsingular, $\mathbf{b} \in \mathbb{C}^N$, $\mathbf{x} \in \mathbb{C}^N$, so $\mathcal{K}_n \subseteq \mathbb{C}^N$. Although Krylov space solvers are iterative in spirit, many of them actually produce in exact arithmetic the exact solution in at most N steps, and, as we see in a moment, this bound can then be replaced by a well determined integer ν that depends on $\mathbf{A}$ and $\mathbf{b}$, but, under a weak assumption, not on the method.

In practice, this theoretical bound on the number of iterations, which is valid in exact arithmetic, may be rather irrelevant: on the one hand, iterative methods are often strongly contaminated by roundoff and therefore do not stop at the νth step, on the other hand, they typically produce sufficiently good approximate solutions in much fewer than ν iterations. Nevertheless the basic facts on Krylov spaces deserve to be better known among the many users of Krylov space methods, and so do the basic facts on block Krylov spaces. Many of the results we cite, in particular those that do not explicitly refer to $\mathbf{A}^{-1}$ or the solution of linear system, persist if $\mathbf{A}$ is singular. However, a full treatment of the singular case that would involve the distinction between consistent and inconsistent systems, is beyond the scope of this paper.

By definition, for given $\mathbf{A}$ and $\mathbf{r}_0$, the subspaces $\mathcal{K}_n$ are clearly nested, and $\mathcal{K}_n$ can have at most dimension $\min\{n, N\}$. However, one can say more. The following results are well known and easy to prove. An early classical text that addresses the topic is Section 8 of Chapter VII of Gantmacher's first volume [7], but there the aim is the description of Krylov's method [16] for computing the characteristic polynomial, from which one would compute its roots, the eigenvalues of the matrix. Our treatment is much shorter, and the target is different.

For generality and simplicity we replace $\mathbf{r}_0$ by $\mathbf{y}$ when appropriate.

The key observation [7] is that for any $\mathbf{y} \neq \mathbf{o}$ there is a smallest positive integer $\nu := \nu(\mathbf{A}, \mathbf{y})$ such that

$$\mathbf{A}^\nu \mathbf{y} = -\,\mathbf{y}\gamma_0 - \mathbf{A}\mathbf{y}\gamma_1 - \cdots - \mathbf{A}^{\nu-1}\mathbf{y}\gamma_{\nu-1}\,, \qquad (2)$$

and this can be reformulated as

$$\psi(\mathbf{A})\mathbf{y} = \mathbf{o}\,, \quad \text{where} \quad \psi(t) := \psi_{\mathbf{A},\mathbf{y}}(t) := t^\nu + \gamma_{\nu-1}\,t^{\nu-1} + \cdots + \gamma_1\,t + \gamma_0\,. \ (3)$$

Following Wilkinson [29] we call the positive integer $\nu := \nu(\mathbf{A}, \mathbf{y})$ the *grade of* $\mathbf{y}$ *with respect to* $\mathbf{A}$ (or, *grade of* $\mathbf{A}$ *with respect to* $\mathbf{y}$). The polynomial $\psi_{\mathbf{A},\mathbf{y}}$ of (3) is the *minimum polynomial of* $\mathbf{y}$ *with respect to* $\mathbf{A}$. Algebraist use a different terminology[1].

Clearly, $\nu \leq N$, and by choosing $\mathbf{y}$ as an eigenvector we see that, for every $\mathbf{A}$, ν can be as small as 1.

In (2), $\gamma_0 \neq 0$ because otherwise we could multiply by $\mathbf{A}^{-1}$ and replace ν by $\nu - 1$. For $n > \nu$ we can then clearly write all terms $\mathbf{A}^n \mathbf{y}$ as linear combinations of $\mathbf{y}, \mathbf{A}\mathbf{y}, \ldots, \mathbf{A}^{\nu-1}\mathbf{y}$. This establishes the first of a number of simple statements that we are summarizing in the following proposition.

PROPOSITION 1 *The grade* $\nu := \nu(\mathbf{A}, \mathbf{y})$ *and the polynomial* $\psi_{\mathbf{A},\mathbf{y}}$ *of exact degree* ν *are characterized by any of the following respective statements:*

(i) $\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = n$ *if* $n \leq \nu$, $\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \nu$ *if* $n \geq \nu$.

(ii) $\nu(\mathbf{A}, \mathbf{y}) = \min \left\{ n \mid \dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\}$.

(iii) $\nu(\mathbf{A}, \mathbf{y}) = \min \left\{ n \mid \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\}$.

(iv) $\nu(\mathbf{A}, \mathbf{y}) = \min \left\{ n \mid \mathbf{A}^{-1}\mathbf{y} \in \mathcal{K}_n(\mathbf{A}, \mathbf{y}) \right\}$.

(v) $\mathcal{K}_\nu(\mathbf{A}, \mathbf{y})$ *is the smallest* $\mathbf{A}$*–invariant subspace that contains* $\mathbf{y}$.

(vi) *Among the monic polynomials* ψ *satisfying* $\psi(\mathbf{A})\mathbf{y} = \mathbf{o}$, *the polynomial* $\psi = \psi_{\mathbf{A},\mathbf{y}}$ *has the smallest degree. (In particular,* ν *is bounded by the degree of the minimal polynomial* $\widehat{\chi}_{\mathbf{A}}$ *of* $\mathbf{A}$.)

There is no generally accepted preference for using the term "Krylov space" or the term "Krylov subspace", but one could argue that $\mathcal{K}_\nu(\mathbf{A}, \mathbf{y})$ is the Krylov space while $\mathcal{K}_n(\mathbf{A}, \mathbf{y})$ with $n \leq \nu$ are Krylov subspaces.

It is important to note that for Krylov spaces the minimal polynomial $\widehat{\chi}_{\mathbf{A}}$ and not the characteristic polynomial $\chi_{\mathbf{A}}$ matters. In particular, by Krylov space methods we cannot determine the geometric multiplicity of an eigenvalue. This follows from Statement (v) of Proposition 1 and was the main reason for introducing block Krylov space methods for solving eigenvalue problems with multiple eigenvalues; see, e.g., [3, 5]. Note in particular that if $\mathbf{A}$ is diagonalizable, a basis for the smallest $\mathbf{A}$–invariant subspace containing $\mathbf{y}$ is given by a minimal set of eigenvectors such that $\mathbf{y}$ is contained in their span. There can be no two linearly independent eigenvectors associated with the same eigenvalue in this set.

---

[1]Algebraists call $\mathcal{K}_\nu(\mathbf{A}, \mathbf{y})$ an $\mathbf{A}$–cyclic (sub)space or the cyclic $\mathbb{C}[x]$–submodule induced by $\mathbf{A}$ and generated by $\mathbf{y}$; see, e.g., page 356 of [13] or pages 146–147 of [20]. Based on that, Ilić and Turner [14] call ν the algebraic grade of $\mathbf{A}$ with respect to $\mathbf{y}$. Unlike numerical analysts, algebraists seem mainly interested in the cases where either $\mathcal{K}_\nu(\mathbf{A}, \mathbf{y})$ is an invariant subspace that belongs to a single Jordan block (i.e., to a single elementary divisor) or it is the whole space (i.e., here $\mathbb{C}^N$). In the latter case the starting vector $\mathbf{y}$ is called an $\mathbf{A}$–cyclic vector or a cyclic vector for $\mathbf{A}$. Algebraists call $\psi_{\mathbf{A},\mathbf{y}}$ the $\mathbf{A}$–annihilator of $\mathbf{y}$, and the same name is given to the ideal that is generated by $\psi_{\mathbf{A},\mathbf{y}}$ in $\mathbb{C}[x]$. In systems and control theory $\mathcal{K}_\nu(\mathbf{A}, \mathbf{y})$ is the reachable subspace of a single-input-single-output (SISO) system, and in case of a minimal realization, $\nu(\mathbf{A}, \mathbf{y})$ is equal to the McMillan degree of the system; see. e.g., [4, 27].

From Statement (iv) of Proposition 1 it follows in one line that

$$\mathbf{x}_\star - \mathbf{x}_0 = \mathbf{A}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_0) = -\mathbf{A}^{-1}\mathbf{r}_0 \in \mathcal{K}_\nu(\mathbf{A}, \mathbf{r}_0),$$

and that ν is the smallest integer with this property. This means that once we have constructed a basis of $\mathcal{K}_\nu(\mathbf{A}, \mathbf{r}_0)$ we can find the exact solution of the linear system there.

THEOREM 2 *Let $\mathbf{x}_\star$ be the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ and let $\mathbf{x}_0$ be any initial approximation of it and $\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$ the corresponding residual. Moreover, let $\nu := \nu(\mathbf{A}, \mathbf{r}_0)$. Then*

$$\mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{K}_\nu(\mathbf{A}, \mathbf{r}_0),$$

and ν is the smallest index for which this holds.

This well-known fact was proved, for example, in [12] and in the technical report associated with [30]. It has the following consequence.

COROLLARY 3 *Any Krylov space solver that does not break down prematurely and produces, for a particular starting vector, residuals that are linearly independent (unless zero) or in some norm minimal with respect to the full search space $\mathbf{x}_0 + \mathcal{K}_n$ for $\mathbf{x}_n$, and which applies only one matrix-vector product with $\mathbf{A}$ per iteration will find (in exact arithmetic) the exact solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ in exactly $\nu := \nu(\mathbf{A}, \mathbf{r}_0)$ iterations: $\mathbf{x}_\nu = \mathbf{x}_\star$.*

PROOF. We assume by definition that $\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n$, so $\mathbf{r}_n \in \mathbf{r}_0 + \mathbf{A}\mathcal{K}_n \in \mathcal{K}_{n+1}$. By the minimality of ν in Theorem 2 we need at least $n := \nu$ steps to be able to find $\mathbf{x}_\star$. But if the residuals are linearly independent, $\mathbf{r}_\nu \in \mathcal{K}_{\nu+1} = \mathcal{K}_\nu$ must be the zero vector since $\dim \mathcal{K}_\nu = \nu$ only. So ν iterations are enough. The same is clearly true if the residuals are minimal in some norm. $\qquad\square$

Examples for methods where Corollary 3 applies are the conjugate gradient method [12], the biconjugate gradient method [6] (if it does not break down), and GMRES [23]. These and many others have this so-called finite-termination property. In contrast, Chebyshev iteration [9] does not have this property: the first ν residuals it generates are linearly independent, but the $(\nu+1)$th is nonzero, except under very special circumstances. GMRES may stagnate, that is, two successive residuals may be identical (and therefore not linearly independent), but since $\|\mathbf{r}_n\|$ it will nevertheless terminate with $\mathbf{x}_\nu = \mathbf{x}_\star$, $\mathbf{r}_\nu = \mathbf{o}$

Admittedly, Theorem 2 and Corollary 3 are theoretical results that are of limited practical value. Normally ν is so large that we do not want to spend the ν matrix-vector products that are needed to construct a basis of $\mathcal{K}_\nu(\mathbf{A}, \mathbf{r}_0)$. We want to find very good approximate solutions with much fewer matrix-vector products. Typically, Krylov space solvers provide that; but there are always exceptions of particularly hard problems. On the other hand, there exist situations where ν is very small compared to N, and then Krylov space solvers do particularly well, because ν iterations are enough; see, e.g., [18]. (In contrast, if $\mathbf{A}$ is just one big Jordan block, so has just one eigenvalue of geometric multiplicity 1, we may still need up to N iterations.)

However, if we replaced the target of finding the exact solution $\mathbf{x}_\star$ by the one of finding a good approximate solution, we could deduce from a correspondingly adapted notion of grade results that are truly relevant in practice. An interesting approach in this direction is due to Ilić and Turner [14], but alternatives exist too.

2 Linear systems with multiple right-hand sides and block Krylov spaces

A nonsingular linear system with s right-hand sides (RHSs) can be written as

$$\mathbf{AX} = \mathbf{B} \qquad \text{with} \quad \mathbf{A} \in \mathbb{C}^{N \times N}, \quad \mathbf{B} \in \mathbb{C}^{N \times s}, \quad \mathbf{X} \in \mathbb{C}^{N \times s}. \qquad (4)$$

We will refer to the "tall and skinny" $N \times s$ matrices of the unknowns and the RHSs as *block vectors*. Their s columns will be distinguished by a superscript when they are referred to individually.

We gather the s initial approximations for the s systems in the block vector $\mathbf{X}_0 \in \mathbb{C}^{N \times s}$ and determine the *initial block residual*

$$\mathbf{R}_0 := \mathbf{B} - \mathbf{AX}_0 \ \in \mathbb{C}^{N \times s}. \qquad (5)$$

A *block Krylov space solver* has the property that each of the nth iterates gathered in $\mathbf{X}_n$ is up to the corresponding shift stored in $\mathbf{X}_0$ a linear combination of *all* the ns columns of $\mathbf{R}_0, \mathbf{AR}_0, \ldots, \mathbf{A}^{n-1}\mathbf{R}_0$:

$$\mathbf{X}_n - \mathbf{X}_0 = \sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{R}_0 \, \boldsymbol{\gamma}_k, \qquad \text{where} \quad \boldsymbol{\gamma}_k \in \mathbb{C}^{s \times s} \ (k = 0, \ldots, n-1). \qquad (6)$$

This leads to the following definition.[2]

DEFINITION. Given $\mathbf{A} \in \mathbb{C}^{N \times N}$ nonsingular and $\mathbf{Y} := \left(\begin{array}{ccc} \mathbf{y}^{(1)} & \ldots & \mathbf{y}^{(s)} \end{array} \right) \in \mathbb{C}^{N \times s}$ with $\mathbf{y}^{(i)} \neq \mathbf{o}$ for $i = 1, \ldots, s$, the *block Krylov (sub)spaces* $\mathcal{B}_n^\square$ $(n \in \mathbb{N}^+)$ generated by $\mathbf{A}$ from $\mathbf{Y}$ are

$$\mathcal{B}_n^\square := \mathcal{B}_n^\square(\mathbf{A}, \mathbf{Y}) := \textsf{block span} \, (\mathbf{Y}, \mathbf{AY}, \ldots, \mathbf{A}^{n-1}\mathbf{Y}) \ \subseteq \mathbb{C}^{N \times s}, \qquad (7)$$

where 'block span' is defined such that

$$\mathcal{B}_n^\square(\mathbf{A}, \mathbf{Y}) = \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{Y} \boldsymbol{\gamma}_k \, ; \ \boldsymbol{\gamma}_k \in \mathbb{C}^{s \times s} \ (k = 0, \ldots, n-1) \right\}. \qquad (8)$$

▲

In this notation, (6) can be written as

$$\mathbf{X}_n \in \mathbf{X}_0 + \mathcal{B}_n^\square(\mathbf{A}, \mathbf{R}_0). \qquad (9)$$

---

[2] In systems and control $\mathcal{B}_\nu^\square(\mathbf{A}, \mathbf{Y})$ is the reachable subspace of a multiple-input-multiple-output (MIMO) system [27].

Again, in practice there exist methods where $\mathbf{X}_n$ (or some columns of it) may not exist for some n, e.g., block BiCG with look-ahead [17]. On the other hand, (9) alone is too general a definition for block Krylov space solvers, as it does not reflect some of their essential properties. For example, typically $\mathbf{X}_n \notin \mathbf{X}_0 + \mathcal{B}_{n-1}^{\square}(\mathbf{A}, \mathbf{R}_0)$.

From now on we will for simplicity always suppose that the assumptions of the definition hold. So, in particular, all initial residuals are assumed to be nonzero.

Each column of an element of $\mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{Y})$ is itself an element of

$$\mathcal{B}_n := \mathcal{B}_n(\mathbf{A}, \mathbf{Y}) := \left\{ \sum_{i=1}^{s} \sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{y}^{(i)} \beta_{k,i} \, ; \; \beta_{k,i} \in \mathbb{C} \; (\forall k, i) \right\} \subseteq \mathbb{C}^N \, , \qquad (10)$$

and this subspace of $\mathbb{C}^N$ is just the sum of the s Krylov subspaces $\mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(i)})$:

$$\mathcal{B}_n(\mathbf{A}, \mathbf{Y}) = \mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(1)}) + \cdots + \mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(s)}) \, . \qquad (11)$$

$\mathcal{B}_n^{\square}$ is then the Cartesian product of s copies of $\mathcal{B}_n$:

$$\mathcal{B}_n^{\square} = \underbrace{\mathcal{B}_n \times \cdots \times \mathcal{B}_n}_{s \text{ times}} \, . \qquad (12)$$

Let us return to the solution of the s linear systems $\mathbf{A}\mathbf{X} = \mathbf{B}$. Now, $\mathbf{x}_0^{(i)} + \mathcal{B}_n(\mathbf{A}, \mathbf{R}_0)$ is the affine space where the nth approximation $\mathbf{x}_n^{(i)}$ of the solution of the ith system $\mathbf{A}\mathbf{x}^{(i)} = \mathbf{b}^{(i)}$ is constructed from:

$$\mathbf{x}_n^{(i)} \in \mathbf{x}_0^{(i)} + \mathcal{B}_n(\mathbf{A}, \mathbf{R}_0) \, . \qquad (13)$$

So, we should learn more about this space.

3 The block grade

Clearly, if the ns vectors $\mathbf{A}^k \mathbf{y}^{(i)} \in \mathbb{C}^N$ in (10) are linearly independent,

$$\dim \mathcal{B}_n = ns \, . \qquad (14)$$

But $\dim \mathcal{B}_n$ can be less than ns because the sum (11) need not be a direct sum and because $\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(i)}) < n$ may hold for some i. This is where the difficulties of block Krylov space solvers come from, but also some of their merits.

Like the Krylov subspaces, the subspaces $\mathcal{B}_n$ and $\mathcal{B}_n^{\square}$ are nested:

$$\mathcal{B}_n \subseteq \mathcal{B}_{n+1} \, , \qquad \mathcal{B}_n^{\square} \subseteq \mathcal{B}_{n+1}^{\square} \, .$$

We are going to show that, again, for sufficiently large n equality holds. Based on this fact, Schmelzer [24] introduced a generalization of the grade discussed in Section 1 to block Krylov spaces. It is defined by an adaptation of the Statements (ii) and (iii) of Proposition 1 and will allow us to establish a number of results that are analogous to those for the ordinary grade.

DEFINITION. The positive integer $\nu := \nu(\mathbf{A}, \mathbf{Y})$ defined by

$$\nu(\mathbf{A}, \mathbf{Y}) = \min \left\{ n \mid \dim \mathcal{B}_n(\mathbf{A}, \mathbf{Y}) = \dim \mathcal{B}_{n+1}(\mathbf{A}, \mathbf{Y}) \right\}$$
$$= \min \left\{ n \mid \mathcal{B}_n(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{n+1}(\mathbf{A}, \mathbf{Y}) \right\}$$

is called *block grade of* $\mathbf{Y}$ *with respect to* $\mathbf{A}$. ▲

Clearly, $n \leq \dim \mathcal{B}_n \leq ns$ if $n \leq \nu(\mathbf{A}, \mathbf{Y})$. Moreover, in analogy to the case $n \geq \nu$ of Statement (i) of Proposition 1 we have then:

LEMMA 4 *For* $n \geq \nu(\mathbf{A}, \mathbf{Y})$,

$$\mathcal{B}_n(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y}), \qquad \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}^{\square}(\mathbf{A}, \mathbf{Y}). \qquad (15)$$

PROOF. By definition of $\nu(\mathbf{A}, \mathbf{Y})$, (15) holds for $n = \nu = \nu(\mathbf{A}, \mathbf{Y})$. Since for any of the individual Krylov spaces $\mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(j)})$ in (11) we have clearly $\mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}^{(j)}) = \mathcal{K}_1(\mathbf{A}, \mathbf{y}^{(j)}) + \mathbf{A}\mathcal{K}_n(\mathbf{A}, \mathbf{y}^{(j)})$ it holds likewise that $\mathcal{B}_{n+1}(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_1(\mathbf{A}, \mathbf{Y}) + \mathbf{A}\mathcal{B}_n(\mathbf{A}, \mathbf{Y})$. So, in view of the nonsingularity of $\mathbf{A}$ and the dimensions of the subspaces involved, $\mathcal{B}_1(\mathbf{A}, \mathbf{Y}) \subseteq \mathbf{A}\mathcal{B}_\nu(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_\nu(\mathbf{A}, \mathbf{Y})$, that is, $\mathcal{B}_\nu(\mathbf{A}, \mathbf{Y})$ is an invariant subspace of $\mathbf{A}$. So, applying $\mathbf{A}$ to any element of $\mathcal{B}_\nu(\mathbf{A}, \mathbf{Y})$ does not lead beyond this space, and this does not change if we replace ν by $n > \nu$. Consequently, the equality on the left side of (15) holds for all $n \geq \nu$. The one on the right side follows then from (12). □

REMARK. Note that the inequality $\nu(\mathbf{A}, \mathbf{y}^{(i)}) \leq \nu(\mathbf{A}, \mathbf{Y})$ need not hold for all $i = 1, \ldots, s$. For example, let $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ be two eigenvectors of $\mathbf{A}$ that belong to different eigenvalues, and let $\mathbf{y}^{(3)} = \mathbf{y}^{(1)} + \mathbf{y}^{(2)}$ and $\mathbf{Y} := \left(\begin{array}{ccc} \mathbf{y}^{(1)} & \mathbf{y}^{(2)} & \mathbf{y}^{(3)} \end{array} \right)$. Then $\nu(\mathbf{A}, \mathbf{y}^{(1)}) = \nu(\mathbf{A}, \mathbf{y}^{(2)}) = 1$ and $\nu(\mathbf{A}, \mathbf{y}^{(3)}) = 2$, but $\nu(\mathbf{A}, \mathbf{Y}) = 1$. ▼

On the other hand, the following inequalities hold:

LEMMA 5 *The block grade of the block Krylov space and the grades of the corresponding individual Krylov spaces are related by*

$$\nu(\mathbf{A}, \mathbf{Y}) \leq \max_{i=1,\ldots,s} \nu(\mathbf{A}, \mathbf{y}^{(i)}). \qquad (16)$$

PROOF. The claim follows from (11) and (15). If we had $m := \max_{i=1,\ldots,s} \nu(\mathbf{A}, \mathbf{y}^{(i)}) < \nu(\mathbf{A}, \mathbf{Y})$, it would follow that $\mathcal{B}_m(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$ in contrast to the definition of $\nu(\mathbf{A}, \mathbf{Y})$. □

In light of the above remark the following result is not completely trivial:

LEMMA 6 *A block Krylov space and the corresponding individual Krylov spaces are related by*

$$\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y}) = \mathcal{K}_{\nu(\mathbf{A}, \mathbf{y}^{(1)})}(\mathbf{A}, \mathbf{y}^{(1)}) + \cdots + \mathcal{K}_{\nu(\mathbf{A}, \mathbf{y}^{(s)})}(\mathbf{A}, \mathbf{y}^{(s)}), \qquad (17)$$

and $\nu(\mathbf{A}, \mathbf{Y})$ *is the smallest index for which this holds.*

6

PROOF. We choose in (11) n larger than the indices ν of all the spaces that appear in there and apply Lemma 4 on the left-hand side and s times Statement (i) of Proposition 1 on the right-hand side. The minimality of $\nu(\mathbf{A}, \mathbf{Y})$ is a consequence of its definition: $\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})-1}(\mathbf{A}, \mathbf{Y}) \subsetneq \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$, so (17) cannot hold for $\nu(\mathbf{A}, \mathbf{Y}) - 1$ if it holds for $\nu(\mathbf{A}, \mathbf{Y})$. $\qquad\square$

Now we easily obtain the analog of the first part of Statement (v) of Proposition 1:

LEMMA 7 $\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$ *is the smallest* $\mathbf{A}$*–invariant subspace of* $\mathbb{C}^N$ *that contains* $\mathbf{y}^{(i)}$, $i = 1, \ldots, s$.

$\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}^{\square}(\mathbf{A}, \mathbf{Y})$ *is the smallest* $\mathbf{A}$*–invariant subspace of* $\mathbb{C}^{N \times s}$ *that contains* $\mathbf{Y}$.

PROOF. We have seen in the proof of Lemma 4 that $\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$ is an $\mathbf{A}$–invariant subspace. But any $\mathbf{A}$–invariant subspace that contains $\mathbf{y}^{(i)}$, $i = 1, \ldots, s$, must contain $\mathcal{K}_{\nu(\mathbf{A}, \mathbf{y}^{(i)})}(\mathbf{A}, \mathbf{y}^{(i)})$, $i = 1, \ldots, s$, so by (17) it must contain $\mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$. $\qquad\square$

Before we come to the adaptation of Statement (iv) of Proposition 1, we note the following generalization of (2). By definition of $\nu = \nu(\mathbf{A}, \mathbf{Y})$, the columns of $\mathbf{A}^\nu \mathbf{Y}$ are linear combinations of the columns of $\mathbf{Y}, \mathbf{A}\mathbf{Y}, \ldots, \mathbf{A}^{\nu-1}\mathbf{Y}$, and this does not hold for all columns of $\mathbf{A}^n \mathbf{Y}$ for any $n < \nu$. That means that there are matrices $\boldsymbol{\gamma}_0, \ldots, \boldsymbol{\gamma}_{\nu-1} \in \mathbb{C}^{s \times s}$, such that

$$\mathbf{A}^\nu \mathbf{Y} = -\mathbf{Y}\boldsymbol{\gamma}_0 - \mathbf{A}\mathbf{Y}\boldsymbol{\gamma}_1 - \cdots - \mathbf{A}^{\nu-1}\mathbf{Y}\boldsymbol{\gamma}_{\nu-1}. \tag{18}$$

Here, $\boldsymbol{\gamma}_0 \neq \boldsymbol{o}$, because of the minimality of ν, but unfortunately we cannot be sure that $\boldsymbol{\gamma}_0$ is nonsingular. (In fact, although we could have assumed that $\mathbf{Y}$ has linearly independent columns, we did not, because in practice, checking the rank of $\mathbf{Y} = \mathbf{R}_0$ is fully analogous to checking the rank of the block residuals $\mathbf{R}_n$ $(n > 0)$ and thus needs to be implemented within a good block Krylov solver anyway. So, here, $\boldsymbol{\gamma}_0$ need not even be uniquely determined and may have rows where all elements are zero.) So, we cannot solve (18) easily for $\mathbf{Y}$ and then apply $\mathbf{A}^{-1}$ to it. Nevertheless, by an alternative, more complicated argument we can still prove the following analog of Statement (iv) of Proposition 1.

LEMMA 8 *The block grade* $\nu(\mathbf{A}, \mathbf{Y})$ *is characterized by*

$$\nu(\mathbf{A}, \mathbf{Y}) = \min\left\{n \mid \mathbf{A}^{-1}\mathbf{Y} \in \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{Y})\right\}.$$

PROOF. Defining $m := \min\{n \mid \mathbf{A}^{-1}\mathbf{Y} \in \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{Y})\}$ we show first that $m \leq \nu(\mathbf{A}, \mathbf{Y})$ and then that $m \geq \nu(\mathbf{A}, \mathbf{Y})$. We know from Statement (iv) of Proposition 1 that $\mathbf{A}^{-1}\mathbf{y}^{(i)} \in \mathcal{K}_{\nu(\mathbf{A}, \mathbf{y}^{(i)})}(\mathbf{A}, \mathbf{y}^{(i)})$, and so by (17) we have also $\mathbf{A}^{-1}\mathbf{y}^{(i)} \in \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}(\mathbf{A}, \mathbf{Y})$ for all i, which is the same as $\mathbf{A}^{-1}\mathbf{Y} \in \mathcal{B}_{\nu(\mathbf{A}, \mathbf{Y})}^{\square}(\mathbf{A}, \mathbf{Y})$. Hence, $m \leq \nu(\mathbf{A}, \mathbf{Y})$.

If we knew that $\mathcal{B}_m^{\square}(\mathbf{A}, \mathbf{Y})$ is an $\mathbf{A}$–invariant subspace of $\mathbb{C}^{N \times s}$, we could conclude that $m \geq \nu(\mathbf{A}, \mathbf{Y})$. By an argument similar to one in the proof of Lemma 4 we show that the former is indeed true. As $\mathbf{A}^{-1}\mathbf{Y} \in \mathcal{B}_m^{\square}(\mathbf{A}, \mathbf{Y})$ we have $\mathbf{Y} \in \mathbf{A}\mathcal{B}_m^{\square}(\mathbf{A}, \mathbf{Y}) \subseteq \mathcal{B}_{m+1}^{\square}(\mathbf{A}, \mathbf{Y})$. But $\mathcal{B}_{m+1}^{\square}(\mathbf{A}, \mathbf{Y})$ is the sum of the subspaces $\mathbf{A}\mathcal{B}_m^{\square}(\mathbf{A}, \mathbf{Y})$

and $\mathcal{B}_1^\square(\mathbf{A}, \mathbf{Y}) = \mathsf{block\ span}\ \{\mathbf{Y}\}$. However, as $\mathbf{Y} \in \mathbf{A}\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y})$ we know that $\mathcal{B}_1^\square(\mathbf{A}, \mathbf{Y}) \subseteq \mathbf{A}\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y})$ and therefore

$$\mathbf{A}\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{m+1}^\square(\mathbf{A}, \mathbf{Y}) .$$

As $\mathbf{A}$ is an invertible linear map acting on $\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y})$ the dimension of $\mathcal{B}_{m+1}^\square(\mathbf{A}, \mathbf{Y})$ is the same as that of $\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y})$. Knowing that those subspaces form a nested sequence we end up with

$$\mathbf{A}\mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_{m+1}^\square(\mathbf{A}, \mathbf{Y}) = \mathcal{B}_m^\square(\mathbf{A}, \mathbf{Y}) .$$

So, by definition of $\nu(\mathbf{A}, \mathbf{Y})$ or by Lemma 7 we have $m \geq \nu(\mathbf{A}, \mathbf{Y})$. □

Next we are looking for an analog of Theorem 2. Amazingly, we do not need Lemma 8 for its proof.

THEOREM 9 *Let $\mathbf{X}_\star$ be the block solution of $\mathbf{A}\mathbf{X} = \mathbf{B}$ and let $\mathbf{X}_0$ be any initial block approximation of it and $\mathbf{R}_0 := \mathbf{B} - \mathbf{A}\mathbf{X}_0$ the corresponding block residual. Moreover, let $\nu := \nu(\mathbf{A}, \mathbf{R}_0)$. Then*

$$\mathbf{X}_\star \in \mathbf{X}_0 + \mathcal{B}_\nu^\square(\mathbf{A}, \mathbf{R}_0) , \tag{19}$$

and ν is the smallest index for which this holds.

PROOF. To prove (19) we just combine Theorem 2 and the relations (17) and (12). The minimality of ν follows from Theorem 2 and Lemma 6. □

Unfortunately, the generalization of the polynomial relation (3) and the related second part of Statement (v) of Proposition 1 is not straightforward since we cannot write $\mathbf{A}^\nu \mathbf{Y} + \mathbf{A}^{\nu-1} \mathbf{Y} \boldsymbol{\gamma}_{\nu-1} + \cdots + \mathbf{A}\mathbf{Y}\boldsymbol{\gamma}_1 + \mathbf{Y}\boldsymbol{\gamma}_0 = \mathbf{O}$ as $\psi_{\mathbf{A}, \mathbf{Y}}(\mathbf{A})\mathbf{Y} = \mathbf{O}$. We will return to this point in Section 5.

4 Block Krylov bases

In the case of a single right-hand side we know from Statement (i) of Proposition 1 that for $n \leq \nu$ the spanning set $\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0$ in (1) is a basis of $\mathcal{K}_n(\mathbf{A}, \mathbf{r}_0)$, the so-called *Krylov basis*. The corresponding $N \times n$ *Krylov matrix*

$$\mathbf{K}_n := \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 & \ldots & \mathbf{A}^{n-1}\mathbf{r}_0 \end{pmatrix} \tag{20}$$

is typically very ill-conditioned, but there are situations where the basis or the matrix are used for theoretical derivations.

In the multiple right-hand side case, where we want to switch again from $\mathbf{R}_0$ to an arbitrary $\mathbf{Y}$ with s nonzero columns, the columns of the $N \times ns$ *block Krylov matrix*

$$\mathbf{B}_n := \begin{pmatrix} \mathbf{Y} & \mathbf{A}\mathbf{Y} & \ldots & \mathbf{A}^{n-1}\mathbf{Y} \end{pmatrix} \tag{21}$$

are still a spanning set of $\mathcal{B}_n(\mathbf{A}, \mathbf{Y})$, but they are in general no longer linearly independent. Clearly, it may happen that $ns > N$. For example, assume $\nu(\mathbf{A}, \mathbf{y}^{(1)}) = N$ and let $\mathbf{y}^{(2)} = \mathbf{A}\mathbf{y}^{(1)}$; then, for $s = 2$ and $n = N$, $\mathcal{B}_n(\mathbf{A}, \mathbf{Y})$ has $ns = 2N$ columns, but, of course, only N of them are linearly independent.

Kent [15] assumed in his treatment of block Krylov matrices that N is a multiple of s and that $\mathbf{B}_{N/s}$ is nonsingular. In this case we have

$$\mathbb{C}^N = \mathcal{K}_{N/s}(\mathbf{A}, \mathbf{y}^{(1)}) \oplus \mathcal{K}_{N/s}(\mathbf{A}, \mathbf{y}^{(2)}) \oplus \cdots \oplus \mathcal{K}_{N/s}(\mathbf{A}, \mathbf{y}^{(s)}), \qquad (22)$$

which is a very special situation, far from the reality of typical applications. We would like to discuss the general case instead.

To get an analog of the Krylov basis in general, we may need to delete some of the columns of $\mathbf{B}_n$, and we want to do this in such a way that for $n = 1, \ldots, \nu$, the columns of the constructed *reduced block Krylov matrix* $\mathbf{B}_n^\circ$ are the basis vectors of a sequence of nested *block Krylov bases* for $\mathcal{B}_k(\mathbf{A}, \mathbf{Y})$, $k = 1, \ldots, n$. This reduction corresponds to the need for deflation in block Krylov space methods. Assume $\mathbf{B}_{n-1}^\circ$ has been constructed for some $n - 1 < \nu$. To get $\mathbf{B}_n^\circ$ we append to $\mathbf{B}_{n-1}^\circ$ the columns of $\mathbf{A}^{n-1}\mathbf{Y}$ and check for linear dependence of the new columns from those of $\mathbf{B}_{n-1}^\circ$ and from each other. For example this could be done by updating the QR decomposition of $\mathbf{B}_{n-1}^\circ$, possibly using column pivoting. If we detect linear dependence, we delete one or more of the columns of $\mathbf{A}^{n-1}\mathbf{Y}$, but we do not make any changes to the columns that are inherited from $\mathbf{B}_{n-1}^\circ$.

This procedure is in general not unique. If, for example, the first column of $\mathbf{A}^{n-1}\mathbf{Y}$ appears in a linear combination that expresses the second column in terms of this column and those of $\mathbf{B}_{n-1}^\circ$, we have the option of deleting either the first or the second column. Of course, we could opt for a specific rule, but we prefer to leave the choice open. This means that, in general, $\mathbf{B}_n^\circ$ is not uniquely determined. So, there exists not just one sequence of nested block Krylov bases for $\mathcal{B}_1, \ldots, \mathcal{B}_\nu$, but a whole tree of such bases. If we denote the dimension of $\mathcal{B}_n(\mathbf{A}, \mathbf{Y})$ by $d_n = d_n(\mathbf{A}, \mathbf{Y})$ then, clearly, any $\mathbf{B}_n^\circ$ has d_n columns and, for $k = 1, \ldots, n - 1$, its first d_k columns form a basis of $\mathcal{B}_k(\mathbf{A}, \mathbf{Y})$ and are thus a possible choice for $\mathbf{B}_k^\circ$.

If $d_\nu = N$, $\mathbf{B}_\nu^\circ$ is nonsingular and hence its columns form a basis of $\mathbb{C}^N$. In analogy to the terminology mentioned in the Footnote 1, we suggest to call $\mathbf{Y}$ *block* $\mathbf{A}$*–cyclic* in this case.

What we have described here is basically a recursive, exact-rank-revealing QR decomposition of the block Krylov matrix $\mathbf{B}_n$ of (21). In finite-precision arithmetic, this would be a very bad approach since the Krylov and block Krylov matrices are typically extremely ill-conditioned. Therefore, in practice, other algorithms are used to build up a block Krylov space basis. For example, there are block versions of the Arnoldi process (or the symmetric Lanczos process if $\mathbf{A}$ is Hermitian); they come in two flavors: the natural one, where a block of up to s orthonormal basis vectors is appended at once, and Ruhe's approach [21], where basis vectors are added one after the other; see, e.g., [19], [22], [28]. The nonsymmetric Lanczos process has also been implemented in these two flavors; see, e.g., [1], [2] or Sections 4.6, 7.9, and 7.10 of [3]. We cannot go into the details here. There are well over a hundred publications on block Krylov space solvers for linear systems of equations and eigenvalue problems. Only a minority among them discusses in detail the delicate problem of how to treat approximate linear dependence among the columns of block vectors when it occurs, and there is no generally accepted answer. To detect it, one should apply refined versions of the rank-revealing QR decompositions or even the more expensive singular

value decomposition.

5 Block minimal polynomials

The minimum polynomial $\psi_{\mathbf{A},\mathbf{y}}$ of $\mathbf{y}$ with respect to $\mathbf{A}$ defined in (3) had the property that $\psi_{\mathbf{A},\mathbf{y}}(\mathbf{A})\mathbf{y} = \mathbf{o}$. How can we extend this to the block case? Can we also write (18) in terms of a polynomial?

Using the block Krylov matrix $\mathbf{B}_\nu$ defined by (21) with $n = \nu$ we can write (18) as

$$\mathbf{A}^\nu\mathbf{Y} = -\mathbf{B}_\nu\mathbf{C}, \qquad \text{where} \quad \mathbf{C} := \begin{pmatrix} \boldsymbol{\gamma}_0 \\ \vdots \\ \boldsymbol{\gamma}_{\nu-1} \end{pmatrix} \in \mathbb{C}^{\nu s \times s}, \qquad (23)$$

or, if we use instead the reduced block Krylov matrix $\mathbf{B}_\nu^\circ$,

$$\mathbf{A}^\nu\mathbf{Y} = -\mathbf{B}_\nu^\circ\mathbf{C}^\circ, \qquad \text{where} \quad \mathbf{C}^\circ := \begin{pmatrix} \boldsymbol{\gamma}_0^\circ \\ \vdots \\ \boldsymbol{\gamma}_{\nu-1}^\circ \end{pmatrix} \in \mathbb{C}^{d_\nu \times s}. \qquad (24)$$

Here, $\gamma_i^\circ \in \mathbb{C}^{(d_{i+1}-d_i)\times s}$ (with $d_0 := 0$), that is, these blocks of coefficients are in general of different size. Whereas in (23) $\mathbf{B}_\nu$ is uniquely defined, but $\mathbf{C}$ may be non-unique, in (24) $\mathbf{B}_\nu^\circ$ is in general non-unique, but $\mathbf{C}^\circ$ is unique once $\mathbf{B}_\nu^\circ$ has been chosen.

Given a matrix polynomial φ of degree m and with coefficients in $\mathbb{C}^{s \times s}$,

$$\varphi(t) = \boldsymbol{\beta}_0 + t\boldsymbol{\beta}_1 + \cdots + t^m\boldsymbol{\beta}_m, \qquad (25)$$

it is unfortunately impossible to form $\varphi(\mathbf{A})\mathbf{Y}$, because the dimensions do not match. But let us recall from Kent [15] and Simoncini and Gallopoulos [25] the following notation (which, in [25], is attributed to William B. Gragg):

$$\varphi(\mathbf{A}) \circ \mathbf{Y} := \mathbf{Y}\boldsymbol{\beta}_0 + \mathbf{A}\mathbf{Y}\boldsymbol{\beta}_1 + \cdots + \mathbf{A}^m\mathbf{Y}\boldsymbol{\beta}_m. \qquad (26)$$

In this notation, and with $\boldsymbol{\iota}$ the $s \times s$ unit matrix, (18) and (23) can be written as

$$\psi(\mathbf{A}) \circ \mathbf{Y} = \mathbf{O}, \qquad (27)$$

where

$$\psi(t) := \psi_{\mathbf{A},\mathbf{Y}}(t) := t^\nu\boldsymbol{\iota} + t^{\nu-1}\boldsymbol{\gamma}_{\nu-1} + \cdots + t\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_0. \qquad (28)$$

This looks very similar to the non-block case in (3), but it does not reflect the possible need for deflation that led to (24). The difficulty is that the coefficients $\boldsymbol{\gamma}_k$ need to be of the same size to define a matrix polynomial. We can bail out by enforcing that those rows of $\mathbf{C}$ that do not appear in $\mathbf{C}^\circ$ are set to zero. Generalizing [15] we may call the corresponding (non-unique) polynomial $\psi_{\mathbf{A},\mathbf{Y}}$ a *matrix-valued minimal polynomial of* $\mathbf{Y}$ *with respect to* $\mathbf{A}$. It has a uniquely determined degree and a minimum number of nonzero rows in $\mathbf{C}$.

Matrix polynomials have been studied extensively by Gohberg, Lancaster, and Rodman [8], but the construction (26) seems to have received little attention. But they have been shown to be of practical value in the discussion of block Krylov spaces, see Simoncini and Gallopoulos [26].

Due to the special structure of $\mathbf{B}_\nu$ Eq. (23) implies that

$$\mathbf{AB}_\nu = \mathbf{B}_\nu \mathbf{F}\,, \tag{29}$$

where

$$\mathbf{F} := \begin{pmatrix} \boldsymbol{o} & \cdots & \boldsymbol{o} & -\boldsymbol{\gamma}_0 \\ \boldsymbol{\iota} & \ddots & \vdots & -\boldsymbol{\gamma}_1 \\ & \ddots & \boldsymbol{o} & \vdots \\ & & \boldsymbol{\iota} & -\boldsymbol{\gamma}_{\nu-1} \end{pmatrix} \tag{30}$$

is a $\nu s \times \nu s$ *block companion matrix*. Particularly beneficial is the case where $\nu s = N$ and $\mathbf{B}_\nu$ is nonsingular, like under the assumption (22). Then $\mathbf{A}$ is similar to $\mathbf{B}_\nu$. More general results are covered in [8].

6   Conclusions

The block grade that we introduced here is not the dimension of an exhausted block Krylov space and yet it is the natural generalization of the grade of a Krylov space. As we have shown, almost all properties of the grade can be reformulated for the block grade, although alternative approaches for most of the proofs have to be used. There are further related mathematical concepts, like the block Krylov basis, the corresponding block companion matrix, and the matrix polynomial that is in a certain sense a block minimal polynomial for the starting vector that generates the block Krylov space. For all these constructions the possibility of linear dependence of columns of the block Krylov matrix — which is related to the possible necessity for deflation in block Krylov methods — causes extra difficulties.

There are also close connections to system and control theory, to certain areas of pure linear algebra, and to the theory of matrix polynomials, but these have only been noted in the margin here.

Acknowledgment. The authors appreciate the referees' careful reading of the manuscript and a large number of specific suggestions for improvements.

References

[1] J. I. ALIAGA, D. L. BOLEY, R. W. FREUND, AND V. HERNÁNDEZ, *A Lanczos-type method for multiple starting vectors*, Math. Comp., 69 (2000), pp. 1577–1601.

[2] Z. BAI, D. DAY, AND Q. YE, *ABLE: an adaptive block Lanczos method for non-Hermitian eigenvalue problems*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 1060–1082 (electronic).

[3] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE, AND H. VAN DER VORST, *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[4] A. BULTHEEL AND M. VAN BAREL, *Linear Algebra, Rational Approximation and Orthogonal Polynomials*, Elsevier, Amsterdam, 1997.

[5] J. CULLUM AND W. E. DONATH, *A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace of large, sparse, symmetric matrices*, in Proc. 1974 IEEE Conference on Decision and Control, 1974, pp. 505–509.

[6] R. FLETCHER, *Conjugate gradient methods for indefinite systems*, in Numerical Analysis, Dundee, 1975, G. A. Watson, ed., vol. 506 of Lecture Notes in Mathematics, Springer, Berlin, 1976, pp. 73–89.

[7] F. GANTMACHER, *The Theory of Matrices*, vol. 1, Chelsea, New York, 1959.

[8] I. GOHBERG, P. LANCASTER, AND L. RODMAN, *Matrix Polynomials*, Academic Press, New York, 1982.

[9] G. H. GOLUB AND R. S. VARGA, *Chebyshev semiiterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods*, Numer. Math., 3 (1961), pp. 147–168.

[10] M. H. GUTKNECHT, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part I*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 594–639.

[11] ———, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part II*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 15–58.

[12] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 409–435.

[13] T. W. HUNGERFORD, *Algebra*, Springer-Verlag, New York, 1974.

[14] M. ILIĆ AND I. W. TURNER, *Krylov subspaces and the analytic grade*, Numer. Linear Algebra Appl., 12 (2005), pp. 55–76.

[15] M. D. KENT, *Chebyshev, Krylov, Lanczos: Matrix Relationships and Computations*, PhD thesis, Stanford University, 1989.

[16] A. N. KRYLOV, *On the numerical solution of the equation by which, in technical questions, frequencies of small oscillations of material systems are determined*, Izv. Akad. Nauk SSSR, ser. fis.-mat., VII (1931), pp. 491–539. (in Russian).

[17] D. LOHER, *New block Lanczos solvers for linear systems with multiple right-hand side*, PhD thesis, Diss. no. 16337, ETH Zurich, Zurich, Switzerland, 2006.

[18] M. F. MURPHY, G. H. GOLUB, AND A. J. WATHEN, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972 (electronic).

[19] M. ROBBÉ AND M. SADKANE, *Exact and inexact breakdowns in the block GMRES method*, Linear Algebra Appl., 419 (2006), pp. 265–285.

[20] S. ROMAN, *Advanced Linear Algebra, 2nd ed.*, Springer, New York, 2005.

[21] A. RUHE, *Implementation aspects of band Lanczos algorithms for computation of eigenvalues of large sparse symmetric matrices*, Math. Comp., 33 (1979), pp. 680–687.

[22] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd. ed., 2003.

[23] Y. SAAD AND M. H. SCHULTZ, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comp., 44 (1985), pp. 417–424.

[24] T. SCHMELZER, *Block Krylov methods for Hermitian linear systems.* Diploma thesis, Department of Mathematics, University of Kaiserslautern, Germany, 2004.

[25] V. SIMONCINI AND E. GALLOPOULOS, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.

[26] ——, *A hybrid block GMRES method for nonsymmetric systems with multiple right-hand sides*, in Proceedings of the Sixth International Congress on Computational and Applied Mathematics (Leuven, 1994), vol. 66, 1996, pp. 457–469.

[27] E. D. SONTAG, *Mathematical Control Theory*, Springer-Verlag, New York, 1990.

[28] B. VITAL, *Etude de quelques méthodes de résolution de problèmes linéaires de grande taille sur multiprocesseur*, PhD thesis, Université de Rennes, 1990.

[29] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[30] D. M. YOUNG AND K. C. JEA, *Generalized conjugate-gradient acceleration of nonsymmetrizable iterative methods*, Linear Algebra Appl., 34 (1980), pp. 159–194.