EQUIVALENT ITERATIVE METHODS FOR p-CYCLIC MATRICES

OLIVER G. ERNST

Dedicated to Prof. Dr. Dr. h. c. mult. Richard S. Varga on the occasion of his 70th birthday.

ABSTRACT. An overview is given of the simplifications which arise when p-cyclic systems are solved by iterative methods. Besides the classic iterative methods, we treat the Chebyshev semi-iterative method and the OR and MR variants of the class of Krylov subspace methods. Particular emphasis is given to equivalent iterations applied to the cyclically reduced system.

1. INTRODUCTION

The systematic investigation of iterative methods for the solution of linear systems of equations with a p-cyclic coefficient matrix (p-cyclic systems, for short) seems to have begun with the thesis of Young [36], who, in the context of solving the linear systems which result from finite difference approximations of elliptic boundary value problems, discovered that systems which possess "Property A" allow one to determine the optimal SOR iteration parameter as a function of the spectral radius of the associated Jacobi matrix. A key achievement in the early work of Richard Varga is the realization that the theory of cyclic matrices of Frobenius [13] and Romanovsky [25] offers the key to the analysis of many iterative methods for the solution of such linear systems of equations. These results, which can be found in Varga's definitive 1962 monograph [35], generalize the results of Young, which apply in the case $p = 2$, to general p-cyclic problems. Some direct results of this generalization are the p-cyclic SOR theory, the cyclic Chebyshev method and the basis for the method of cyclic reduction. Many of these fundamental ideas were later developed further by other researchers.

Discretizations of partial differential equations by finite differences usually result in 2-cyclic systems, but exceptions can be found in Varga [33], where a nonstandard discretization of Laplace's equation on a triangular mesh leads to a 3-cyclic problem, and Tee [31], where the discretization of periodic parabolic initial-boundary value problems is considered, which leads to p-cyclic problems with p depending on the fineness of the discretization. Another well-known source of p-cyclic problems is the theory of Markov chain modeling (see Stewart [27]). We also mention linear least-squares problems, which can be formulated as 3-cyclic systems (see [11] and the references given there).

Date: August 8, 2000.

1991 *Mathematics Subject Classification.* 65N15.

Key words and phrases. SOR, cyclic reduction, Krylov subspace methods, GMRES, FOM, p-cyclic matrices, equivalent iterative methods.

One of the characteristics of p-cyclic systems is the existence of an equivalent system consisting of p uncoupled subsystems, the (cyclically) reduced systems. Many iterative solution methods can be reformulated to operate on one of these—often substantially smaller—reduced systems instead of the original system, often resulting in an iteration which is p times as fast while at the same time working with shorter vectors.

The purpose of this paper is to provide a systematic overview of which methods allow one to take advantage of the p-cyclic structure of a linear system of equations by solving reduced systems, and how this can be done. Most of the results in this paper are not new, but have, to the author's knowledge, not been presented under this perspective. Exceptions are the equivalence result for orthogonal residual (OR) methods for general p in Section 6.3 and some of the connections we point out.

The paper is organized as follows. Section 2 reviews the formal definitions of p-cyclic matrices and reduced systems and how these arise in classic stationary iterative methods. Before turning to iterative methods, however, Section 3 shows how the 2-cyclic property of certain discrete boundary value problems is exploited in the direct solution method known as (block) cyclic reduction. The classic iterative methods of Jacobi and Gauß-Seidel as well as SOR are treated in Section 4, followed by the Chebyshev Semi-Iterative Method for the Hermitian 2-cyclic case in Section 5. We conclude with the treatment of Krylov subspace methods for general p-cyclic problems in Section 6.

2. Iterative Methods and p-cyclic Matrices

The fundamental concept in this paper is that of a matrix which is *weakly cyclic of index* p. The following definition is taken from Varga [35], where it is attributed to Romanovsky [25].

Definition 1. A matrix $B \in \mathbb{C}^{N\times N}$ is *weakly cyclic of index* p if there exists a permutation matrix P such that

$$PBP^T = \begin{bmatrix} O & & \cdots & O & B_1 \\ B_2 & O & & \cdots & O \\ O & B_3 & O & \cdots & O \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ O & \cdots & O & B_p & O \end{bmatrix}, \tag{1}$$

where the null diagonal submatrices are square.

We shall henceforth assume that the matrix B has already been permuted to have the form (1), which is sometimes called the normal form for a matrix weakly cyclic of index p.

The most important results in the analysis of stationary iterative methods such as Jacobi, Gauß-Seidel and SOR iteration are formulated for linear systems of equations of the form

$$A\boldsymbol{x} = (I - B)\boldsymbol{x} = \boldsymbol{k}, \tag{2}$$

in which B is weakly cyclic of index p. For this reason, the following definition is also useful.

Definition 2. A matrix $A \in \mathbb{C}^{N \times N}$ is *p-cyclic* relative to a given partitioning

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,p} \\ \vdots & & \vdots \\ A_{p,1} & \cdots & A_{p,p} \end{bmatrix}$$

if the block diagonal matrix $D = \operatorname{blockdiag}(A_{1,1}, \ldots, A_{p,p})$ is nonsingular and $B := I - D^{-1}A$ is weakly cyclic of index p.

In other words, a matrix is p-cyclic relative to a partitioning if the iteration matrix of the associated (block) Jacobi method is weakly cyclic of index p. Since we assume B to be in the normal form (1), the matrix A is also *consistently ordered* in the terminology of Varga [35] and (for the case $p = 2$) Young [36].

We shall require two properties of p-cyclic matrices in the sequel, the first of which, again cited from [35, Theorem 2.4], states that the spectrum of these matrices is highly structured.

Theorem 2.1. *If $B \in \mathbb{C}^{N \times N}$ is weakly cyclic of index p, then its characteristic polynomial χ is given by*

$$\chi(\lambda) = \lambda^m \prod_{j=1}^{r} (\lambda^p - \beta_j^p)$$

in which each β_j is a nonzero eigenvalue of B.

As a consequence, the spectrum $\Lambda(B)$ of a matrix which is weakly cyclic of index p remains invariant under rotations of the complex plane about the origin through angles of $2\pi/p$.

The proof of Theorem 2.1 is based on the following observation for which we shall also find frequent use: The pth power of a matrix B of the form (1) is block diagonal and given by

$$B^p = \operatorname{blockdiag}(B_1^{(p)}, \ldots, B_p^{(p)}), \tag{3}$$

where we have introduced the notation $B_k^{(j)}$ to denote the cyclically permuted product

$$B_k^{(j)} = B_k \, B_{\pi(k)} \, B_{\pi^2(k)} \cdots B_{\pi^j(k)}, \qquad [\pi(1), \ldots, \pi(p)] = [p, 1, \ldots, p-1],$$

of j factors of the matrices $B_\ell, \ell = 1, \ldots, p$ in descending order starting with B_k. For the diagonal blocks of B^p this gives

$$\begin{aligned} B_1^{(p)} &= B_1 B_p B_{p-1} \cdots B_2, \\ B_2^{(p)} &= B_2 B_1 B_p \cdots B_3, \\ &\;\;\vdots \\ B_p^{(p)} &= B_p B_{p-1} \cdots B_1, \end{aligned}$$

all of which, since each is a product of the same matrix factors, share the same set of nonzero eigenvalues. Relation (3) combined with the identity

$$(I + B + B^2 + \cdots + B^{p-1})(I - B) = I - B^p$$

leads to a much simpler system by multiplying (2) from the left by $I+B+\cdots+B^{p-1}$, yielding

$$(I-B^p)\boldsymbol{x} = \widetilde{\boldsymbol{k}}, \qquad \widetilde{\boldsymbol{k}} := \boldsymbol{k} + B\boldsymbol{k} + \cdots + B^{p-1}\boldsymbol{k}. \tag{4}$$

The new system (4) is equivalent with (2), as follows from the rotational symmetry of $\Lambda(B)$, by which $1 \notin \Lambda(B^p)$ if $1 \notin \Lambda(B)$.

System (4) consists of the p uncoupled systems

$$(I - B_j^{(p)})\boldsymbol{x}_j = \widetilde{\boldsymbol{k}}_j, \qquad j = 1, \dots, p, \tag{5}$$

where $\widetilde{\boldsymbol{k}}_j$, the jth block of $\widetilde{\boldsymbol{k}}$, is given by

$$\widetilde{\boldsymbol{k}}_j = \sum_{i=0}^{p-1} B_j^{(\ell)} \boldsymbol{k}_{\pi^i(j)} = \boldsymbol{k}_j + B_j \boldsymbol{k}_{\pi(j)} + B_j^{(2)} \boldsymbol{k}_{\pi^2(j)} + \cdots + B_j^{(p-1)} \boldsymbol{k}_{\pi^{p-1}(j)}.$$

and $\boldsymbol{k}_j$ is the jth block of $\boldsymbol{k}$. We shall refer to each of the systems (5) as a *(cyclically) reduced system.* By the structure of B, solving any one of the reduced systems allows the remaining blocks of the solution vector $\boldsymbol{x}$ of (2) to be obtained through back substitution, which requires only multiplication by one of the matrices B_j per block. In this sense, solving any of the reduced systems—e.g. the smallest—is equivalent to solving the original system (2).

As an example, we consider the simplest case $p = 2$, which is familiar from discretizations of elliptic boundary value problems. Here the system (2) has the form

$$(I - B)\boldsymbol{x} = \boldsymbol{k}, \qquad B = \begin{bmatrix} O & B_1 \\ B_2 & O \end{bmatrix} \tag{6}$$

and its uncoupled form $(I+B)(I-B)\boldsymbol{x} = (I+B)\boldsymbol{k}$ is given by

$$\begin{bmatrix} I - B_1B_2 & O \\ O & I - B_2B_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{k}_1 + B_1\boldsymbol{k}_2 \\ \boldsymbol{k}_2 + B_2\boldsymbol{k}_1 \end{bmatrix}. \tag{7}$$

After solving either of the block systems in (7), say, the first for $\boldsymbol{x}_1$, the remaining block of unknowns $\boldsymbol{x}_2$ is obtained by substitution into the second block of equations in (6), yielding $\boldsymbol{x}_2 = B_2\boldsymbol{x}_1 + \boldsymbol{k}_2$.

3. Cyclic Reduction

Before turning to iterative methods for p-cyclic systems, we would first like to establish the connection to a popular direct solution algorithm known as *(block) cyclic reduction (CR)*, which also makes use of the equivalence of the original and reduced systems, in this case for $p = 2$. Although our focus is on iterative methods, we include this excursion to a direct method to emphasize how cyclic reduction, which is usually presented as an odd-even elimination scheme, really just consists of forming the reduced system. This technique, is designed for the block-tridiagonal systems of equations arising from the discretization of second-order elliptic boundary value problems to which the technique of separation of variables is applicable. Our intention is to show that one step of CR, which consists of combining adjacent block equations in order to eliminate the even-numbered blocks, is equivalent to forming one of the reduced systems (5). The CR method is most effective when the reduced system is itself 2-cyclic, so that the reduction process can be applied recursively.

For simplicity, we consider as an example the discrete five-point Laplacian on the unit square, whose (suitably scaled) coefficient matrix for six blocks of unknowns is given by

$$A = \begin{bmatrix} T & -I & & & & \\ -I & T & -I & & & \\ & -I & T & -I & & \\ & & -I & T & -I & \\ & & & -I & T & -I \\ & & & & -I & T \end{bmatrix}, \qquad T = \mathrm{tridiag}(-1, 4, -1).$$

One step of CR consists of eliminating every other block row—say, the odd ones—by multiplying the even ones by T and adding the adjacent odd rows. In our example, this leads to the new system with coefficient matrix

$$A_{\mathrm{e}} = \begin{bmatrix} T^2 - 2I & -I & \\ -I & T^2 - 2I & -I \\ & -I & T^2 - I \end{bmatrix}$$

of half the block size for the even unknowns. The CR algorithm proceeds recursively in this manner until the block system is small enough to be solved by other means, usually when this consists of only a single block. The solution of this small system is facilitated by the fact that the diagonal blocks of the successively smaller systems can be expressed as Chebyshev polynomials in the matrix T, and hence can be factored explicitly into a product of tridiagonal matrices. The solution of this small system is followed by a back substitution phase to compute the previously eliminated blocks of unknowns.

One elimination step can be expressed by multiplying A from the left by the appropriate elimination matrix

$$\begin{bmatrix} I & & & & & \\ I & T & I & & & \\ & & I & & & \\ & & I & T & I & \\ & & & & I & \\ & & & & I & T \end{bmatrix} A = \begin{bmatrix} T & -I & & & & \\ & T^2 - 2I & & -I & & \\ & -I & T & -I & & \\ & -I & & T^2 - 2I & & -I \\ & & & -I & T & -I \\ & & & -I & & T^2 - I \end{bmatrix}. \tag{8}$$

Reordering so that all odd and even blocks are contiguous yields

$$\widetilde{A} = \begin{bmatrix} T & & & -I & & \\ & T & & -I & -I & \\ & & T & & -I & -I \\ -I & -I & & T & & \\ & -I & -I & & T & \\ & & -I & & & T \end{bmatrix}$$

and, in this reordering, equation (8) becomes

$$\begin{bmatrix} I & & & & & \\ & I & & & & \\ & & I & & & \\ I & I & & T & & \\ & I & I & & T & \\ & & I & & & T \end{bmatrix} \widetilde{A} = \begin{bmatrix} T & & & -I & & \\ & T & & -I & -I & \\ & & T & & -I & -I \\ & & & T^2 - 2I & -I & \\ & & & -I & T^2 - 2I & -I \\ & & & & -I & T^2 - I \end{bmatrix}.$$

If we treat the odd unknowns the same way as the even ones and eliminate their coupling to the even unknowns in the upper right block, this yields

$$
\begin{bmatrix}
T & & & I & & \\
& T & & I & I & \\
& & T & & I & I \\
I & I & & T & & \\
& I & I & & T & \\
& & I & & & T
\end{bmatrix}
\widetilde{A} = \begin{bmatrix} A_{\mathrm{o}} & \\ & A_{\mathrm{e}} \end{bmatrix}
\tag{9}
$$

with

$$
A_{\mathrm{o}} = \begin{bmatrix} T^2 - I & -I & \\ -I & T^2 - 2I & -I \\ & -I & T^2 - 2I \end{bmatrix}, \quad A_{\mathrm{e}} = \begin{bmatrix} T^2 - 2I & -I & \\ -I & T^2 - 2I & -I \\ & -I & T^2 - I \end{bmatrix}.
$$

Finally, after factoring out the block diagonal matrix $D = \operatorname{diag}(T, \dots, T)$ on the left and right, equation (9) takes on the form

$$
\begin{bmatrix} A_{\mathrm{o}} & \\ & A_{\mathrm{e}} \end{bmatrix} = D(I + B)(I - B)D = D(I - B^2)D
$$

where

$$
B = \begin{bmatrix} O & B_1 \\ B_1^T & O \end{bmatrix}, \quad \text{and} \quad B_1 = \begin{bmatrix} T^{-1} & & \\ T^{-1} & T^{-1} & \\ & T^{-1} & T^{-1} \end{bmatrix}.
$$

Thus, except for the block diagonal scaling matrices needed because A is not yet in the normal form $I - B$, one step of CR is seen to be equivalent with forming the uncoupled system (4) and solving one of the reduced systems.

While the idea of solving the reduced system seems to have been proposed as early as 1954 by Schröder [26], the CR method was systematically developed as a recursive direct solution technique by Buzbee et al. [2], Buneman [1] and Hockney [20]. Further refinements of the method, including the implementation in the successful software suite FISHPAK is due to Swarztrauber and Sweet [28, 30, 29]. Recent applications of cyclic reduction include performing only one or a few steps and then applying an iterative method to the resulting reduced systems. This has been shown to be advantageous for discretized convection-diffusion problems in that convergence of the iteration is often faster for the reduced problem. Such an approach is taken e.g. in Thompson et al. [32] and Elman and Golub [8].

4. Classic Iterative Methods

4.1. **Jacobi Iteration.** The coefficient matrix $A = I - B$ of the system (2) is already in the form of the Jacobi splitting, hence the Jacobi iteration is given by

$$
\boldsymbol{x}^{(m+1)} = \boldsymbol{x}^{(m)} + [\boldsymbol{k} - (I - B)\boldsymbol{x}^{(m)}] = B\boldsymbol{x}^{(m)} + \boldsymbol{k},
\tag{10}
$$

which, since the diagonal of A is the identity, coincides with Richardson's method. After j steps, we obtain

$$
\boldsymbol{x}^{(m+j)} = B^j \boldsymbol{x}^{(m)} + (I + B + \cdots + B^{j-1})\boldsymbol{k},
$$

and, comparing with (4), observe that p steps of Jacobi iteration applied to the original system (2) are equivalent with one step of Richardson iteration applied to the uncoupled system (4). (The diagonal of $I - B^p$ is generally different from the identity, hence Jacobi iteration does not coincide with Richardson iteration for the uncoupled system.) As $I - B^p$ is block diagonal, Richardson's method applied to (4) amounts to simultaneously applying Richardson's method to each of the reduced equations (5) and, in view of the remarks in Section 2, we conclude that p steps of Jacobi iteration applied to the original system is equivalent with one step of Richardson iteration applied to a reduced system.

4.2. **Gauß-Seidel Iteration.** The Gauß-Seidel splitting $A = I - L - U$ with respect to the partitioning (1) involves the block matrices

$$L = \begin{bmatrix} O & & \cdots & & O \\ B_2 & O & & & \\ O & B_3 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \\ O & \cdots & O & B_p & O \end{bmatrix}, \qquad U = \begin{bmatrix} O & \cdots & O & B_1 \\ & & & O \\ \vdots & \ddots & & \vdots \\ & & & \\ O & \cdots & & O \end{bmatrix},$$

in terms of which the Gauß-Seidel iteration for the system (2) is

$$\boldsymbol{x}^{(m+1)} = (I - L)^{-1}(U\boldsymbol{x}^{(m)} + \boldsymbol{k}). \tag{11}$$

Since L is a nilpotent matrix ($L^p = O$), we have

$$(I - L)^{-1} = I + L + \cdots + L^{p-1},$$

which yields

$$(I - L)^{-1}U = \begin{bmatrix} O & \dots & O & B_1 \\ O & \dots & O & B_2^{(2)} \\ \vdots & & \vdots & \vdots \\ O & \dots & O & B_p^{(p)} \end{bmatrix}. \tag{12}$$

The powers L^j, $j = 1, \dots, p-1$, of L are nonzero only on the jth block subdiagonal, the entries of which are $B_{j+1}^{(j)}, B_{j+2}^{(j)}, \dots, B_p^{(j)}$. As a result, the last block of $L^j\boldsymbol{k}$ is given by $B_p^{(j)}\boldsymbol{k}^{(p-j)}$, and therefore the last block of $(I - L)^{-1}\boldsymbol{k}$ is

$$\boldsymbol{k}_p + B_p\boldsymbol{k}_{p-1} + B_pB_{p-1}\boldsymbol{k}_{p-2} + \cdots + B_pB_{p-1}\cdots B_2\boldsymbol{k}_1 = \widetilde{\boldsymbol{k}}_p,$$

which coincides with the right hand side of the last reduced system (5). Combined with (12), this reveals that, after *one* step of Gauß-Seidel iteration (11), the last block of the iterate coincides with that of Richardson's method applied to the last reduced system (5). In contrast to Jacobi iteration, only the last block of the Gauß-Seidel iterates is equivalent to an iteration on a reduced system. Which block is equivalent, of course, depends on the ordering in which the blocks are updated in the Gauß-Seidel iteration.

These observations regarding Jacobi and Gauß-Seidel iteration are well-known. The earliest reference known to the author is Schröder [26], who treated the case $p = 2$ in the context of solving Laplace's equation by finite difference methods. Schröder also conjectured that applying Jacobi or Gauß-Seidel iteration—i.e., not Richardson iteration—to the reduced system would result in faster convergence. That this is the case if B is in addition Hermitian, convergent, *nonnegative* and *irreducible* is shown in Varga [35, p. 157] using the theory of regular splittings. This result is extended to block iterative methods applied to the original system versus the induced block method applied to the reduced system by Hageman and Varga [19].

4.3. **Successive Overrelaxation.** Given the Gauß-Seidel splitting $A = I - L - U$, the successive overrelaxation (SOR) scheme with parameter ω for solving (2) is

$$(I - \omega L)\boldsymbol{x}^{(m+1)} = [(1-\omega)I + \omega U]\,\boldsymbol{x}^{(m)} + \omega \boldsymbol{k}, \qquad m = 0, 1, \ldots \tag{13}$$

with associated iteration matrix $\mathcal{L}_\omega = (I - \omega L)^{-1}[(1-\omega)I + \omega U]$. The analysis of SOR iteration in the p-cyclic case where $p \geq 2$ is given in Varga's celebrated 1959 paper [34], in which he determines the optimum relaxation parameter in the case $\Lambda(B^p) \subset [0, \rho(B^p)]$, $\rho(B) < 1$, thus generalizing Young's results for $p = 2$. For a current overview of SOR convergence theory the reader is referred to the contribution to this volume of Eiermann [4] and the references contained therein.

The "stepping stone" for analysis of p-cyclic SOR is Varga's generalization to $p > 2$ of the *fundamental relation*

$$(\lambda + \omega - 1)^p = \omega^p \beta^p \lambda^{p-1} \tag{14}$$

between the nonzero eigenvalues λ of $\mathcal{L}_\omega$ and those β of B, which holds when A is a consistently ordered p-cyclic matrix. (For a discussion of subsequent generalizations of the term 'consistently ordered', see [35, Section 4.6].)

Galanis et al. [14] show that, under the same assumptions, (14) holds also for the associated matrices themselves, i.e.,

$$\big(\mathcal{L}_\omega + (\omega - 1)I\big)^p = \omega^p B^p \mathcal{L}_\omega^{p-1}.$$

This matrix relation is used in [14] to show that one step of SOR iteration applied to (2) can be expressed as

$$\boldsymbol{x}^{(m+1)} = \omega^p \left(B^p \boldsymbol{x}^{(m)} + \widetilde{\boldsymbol{k}}\right) - \sum_{j=1}^{p} \binom{p}{j} (\omega - 1)^j \boldsymbol{x}^{(m+1-j)}. \tag{15}$$

The scheme in (15) is seen to be an iteration with an update involving the past p iterates applied to the uncoupled system (4). Such methods are known as *p-step polynomial acceleration methods* or *semi-iterative methods* and are thoroughly developed and analyzed by Niethammer and Varga in [22], which also contains a characterization of the asymptotic rate of convergence of such a method in terms of $\Lambda(B)$. This link of p-cyclic SOR with the class of p-step methods makes available the well-developed convergence theory of the latter for analyzing p-cyclic SOR, as carried out by Eiermann et al. [5, 7, 4].

This sequence of papers actually uses another p-step semi-iterative method, namely

$$\boldsymbol{y}^{(m+1)} = \omega\left(B\boldsymbol{y}^{(m)} + \boldsymbol{k}\right) + (1-\omega)\boldsymbol{y}^{(m+1-p)}, \quad m \geq 0, \tag{16}$$

which is equivalent to p-cyclic SOR iteration in the following sense: If the auxiliary vector $\widetilde{\boldsymbol{y}}^{(m)}$ is defined as consisting of blocks of successively more recent iterates generated by (16), i.e.,

$$\widetilde{\boldsymbol{y}}^{(m)} = \begin{bmatrix} \boldsymbol{y}_1^{(m-p+1)} \\ \boldsymbol{y}_2^{(m-p+2)} \\ \vdots \\ \boldsymbol{y}_p^{(m)} \end{bmatrix}$$

then, as is easily verified,

$$\widetilde{\boldsymbol{y}}^{(m+p)} = \mathcal{L}_\omega \widetilde{\boldsymbol{y}}^{(m)} + \omega(I - \omega L)^{-1}\boldsymbol{k}.$$

In other words, one SOR step applied to $\widetilde{\boldsymbol{y}}^{(m)}$ advances each of its block components by p steps of the semi-iterative method defined by (16). This equivalence was first noted in Gutknecht, Niethammer and Varga [17].

5. The Chebyshev Semi-Iterative Method

5.1. **Background.** The term semi-iterative method was introduced by Varga [33] (although this approach had been around some time before, cf. [10]) and refers to iteration schemes which accelerate the convergence of a given iteration by forming linear combinations of previous iterates. If the iteration to be accelerated for solving (2) is the Jacobi iteration (10), then the iterates $\boldsymbol{y}^{(m)}$ at step m of the accelerated sequence as well as the associated errors $\boldsymbol{e}^{(m)} = A^{-1}\boldsymbol{k} - \boldsymbol{y}^{(m)}$ have the form

$$\boldsymbol{y}^{(m)} = p_m(B)\boldsymbol{y}^{(0)}, \quad \boldsymbol{e}^{(m)} = p_m(B)\boldsymbol{e}^{(0)}, \qquad m \geq 0,$$

with a polynomial p_m of degree m normalized by $p_m(1) = 1$. For this reason, these methods are also called *polynomial acceleration methods.*

If B is Hermitian and satisfies $\Lambda(B) \subset [\beta_{\min}, \beta_{\max}]$ with $-1 < \beta_{\min} \leq \beta_{\max} < 1$, then the standard error analysis (see [35, Chapter 5]) reveals that the optimal choice of polynomials—in the sense of minimizing the spectral norm $\|p_m(B)\|_2$ over all polynomials based on the bounds on the spectrum—is the sequence of scaled and shifted Chebyshev polynomials

$$p_m(\zeta) = \frac{T_m(\eta(\zeta))}{T_m(\eta(1))}, \quad \eta(\zeta) = \frac{2\zeta - (\beta_{\max} + \beta_{\min})}{\beta_{\max} - \beta_{\min}}, \qquad m \geq 0. \tag{17}$$

The well-known three-term recurrence for the Chebyshev polynomials carries over to the iterates $\boldsymbol{y}^{(m)}$. If we simplify the spectral bounds to $\beta_{\max} = -\beta_{\min} = \rho(B)$, then $\eta(\zeta) = \zeta/\rho(B)$ and

$$\boldsymbol{y}^{(m+1)} = \omega_{m+1}\big(B\boldsymbol{y}^{(m)} + \boldsymbol{k} - \boldsymbol{y}^{(m-1)}\big) + \boldsymbol{y}^{(m-1)}, \qquad m \geq 0, \tag{18}$$

with $\boldsymbol{y}^{(0)} = \boldsymbol{x}^{(0)}$, $\boldsymbol{y}^{(1)} = B\boldsymbol{y}^{(0)} + \boldsymbol{k}$ and

$$\omega_{m+1} = 1 + \frac{T_{m-1}(1/\rho(B))}{T_{m+1}(1/\rho(B))}, \quad m \geq 1, \qquad \omega_1 = 1.$$

5.2. **The Cyclic Chebyshev Semi-Iterative Method.** Golub and Varga [16] introduced the *cyclic Chebyshev semi-iterative method* on applying the Chebyshev semi-iterative method to a Hermitian, positive definite, 2-cyclic matrix, in which case the system has the form (6) with $B_2 = B_1^H$, which, substituted in (18), yields

$$\begin{bmatrix} \boldsymbol{y}_1^{(m+1)} \\ \boldsymbol{y}_2^{(m+1)} \end{bmatrix} = \omega_{m+1} \begin{bmatrix} B_1\boldsymbol{y}_2^{(m)} + \boldsymbol{k}_1 \\ B_1^H\boldsymbol{y}_1^{(m)} + \boldsymbol{k}_2 \end{bmatrix} + (1 - \omega_{m+1}) \begin{bmatrix} \boldsymbol{y}_1^{(m-1)} \\ \boldsymbol{y}_2^{(m-1)} \end{bmatrix}. \tag{19}$$

Equation (19) suggests one way of exploiting the 2-cyclic structure: each of the two blocks of $\boldsymbol{y}^{(m+1)}$ depends only on the same block of $\boldsymbol{y}^{(m-1)}$ and on the opposite block of $\boldsymbol{y}^{(m)}$, i.e., the sequence $\{\boldsymbol{y}^{(m)}\}_{m\geq 0}$ of Chebyshev iterates consists of two completely uncoupled sequences of alternating blocks. The cyclic Chebyshev semi-iterative method computes only one of these sequences and, after a sufficiently accurate approximation of one block has been obtained, computes the remaining block, e.g. by back-substitution. The 2-cyclic structure thus allows one to compute only this staggered subsequence of the Chebyshev iterates, resulting in an iteration which is half as expensive in terms of work and storage than the straightforward Chebyshev semi-iterative method applied to (2).

It is further shown in [16] that the iteration operator of the cyclic Chebyshev semi-iterative method has a smaller spectral norm than the corresponding SOR operator, although for the problem under consideration both methods have the same asymptotic convergence rate.

We point out a further connection between the cyclic Chebyshev semi-iterative method and SOR: as shown by Golub [15], the sequence of parameters $\{\omega_m\}_{m\geq 2}$ in (19) is strictly decreasing with limit ω_b, which denotes the optimal SOR relaxation parameter for (2). Moreover, the equivalent 2-step semi-iterative method (16) in the Hermitian 2-cyclic case reduces to

$$\widetilde{\boldsymbol{y}}^{(m+2)} = \begin{bmatrix} \boldsymbol{y}_1^{(m+1)} \\ \boldsymbol{y}_2^{(m+2)} \end{bmatrix} = \omega \begin{bmatrix} B_1 \boldsymbol{y}_2^{(m)} + \boldsymbol{k}_1 \\ B_1^H \boldsymbol{y}_1^{(m+1)} + \boldsymbol{k}_2 \end{bmatrix} + (1-\omega) \begin{bmatrix} \boldsymbol{y}_1^{(m-1)} \\ \boldsymbol{y}_2^{(m)} \end{bmatrix},$$

which is precisely one (double-) step of the cyclic Chebyshev semi-iterative method with ω_m replaced by its limiting value.

5.3. **Accelerating the Uncoupled System.** An alternative to the staggering of the unknowns as in the cyclic Chebyshev semi-iterative method is to apply Chebyshev acceleration to the uncoupled system (4). With the same a priori knowledge $\Lambda(B) \subset [-\rho(B), \rho(B)]$, we conclude that the two diagonal blocks $B_1 B_1^H$ and $B_1^H B_1$ of B^2 in (4) have their eigenvalues in the interval $[0, \rho^2(B)]$, and therefore, in analogy with (17), we obtain

$$\widetilde{p}_m(\zeta) = \frac{T_m(\widetilde{\eta}(\zeta))}{T_m(\widetilde{\eta}(1))}, \quad \eta(\zeta) = 2\zeta/\rho^2(B) - 1, \qquad m \geq 0$$

for the iteration polynomials $\widetilde{p}_m$ of the Chebyshev semi-iterative method applied to (4). For $\xi \in [-1, 1]$, the Chebyshev polynomial is defined as

$$T_m(\xi) = \cos mx, \quad x = \arccos \xi, \qquad m \geq 0.$$

In view of

$$2 \arccos \xi = 2x = \arccos \cos 2x = \arccos(2\cos^2 x - 1) = \arccos(2\xi^2 - 1),$$

we see that $T_{2m}(\xi) = T_m(2\xi^2 - 1)$ and therefore

$$\widetilde{p}_m(B^2) = \frac{T_m(2\frac{B^2}{\rho^2(B)} - 1)}{T_m(2\frac{1}{\rho^2(B)} - 1)} = \frac{T_{2m}(\frac{B}{\rho(B)})}{T_{2m}(\frac{1}{\rho(B)})} = p_m(B)$$

from which we infer that the iterate at step m of the Chebyshev semi-iterative method applied to the uncoupled system (4) coincides with that of step $2m$ of the method applied to the original system (2). Of course it is only necessary to solve one of the two reduced systems in the uncoupled system by iteration and, since one iteration step for a reduced system requires two matrix-vector multiplications, the resulting approach is comparable to the cyclic Chebyshev semi-iterative method.

The equivalence of one step of Chebyshev iteration for the uncoupled/reduced system with two steps of Chebyshev applied to the original system is shown by Hageman et al. [18]. That Chebyshev iteration applied to a reduced system has the same asymptotic convergence rate as the cyclic Chebyshev method is shown much earlier in Varga [35].

6. Krylov Subspace Methods

In this section orthogonality can be interpreted with respect to any inner product on $\mathbb{C}^{N\times N}$, and $\|\cdot\|$ denotes the associated norm.

6.1. **Background.** Krylov subspace methods select approximations to the solution of (2) by determining corrections $\boldsymbol{c}^{(m)}$ to an initial guess $\boldsymbol{x}^{(0)}$ from successively generated subspaces

$$\mathcal{K}_m(A, \boldsymbol{r}^{(0)}) = \operatorname{span}\{\boldsymbol{r}^{(0)}, A\boldsymbol{r}^{(0)}, \dots, A^{m-1}\boldsymbol{r}^{(0)}\} \tag{20}$$

known as *Krylov spaces*, where $\boldsymbol{r}^{(0)} := \boldsymbol{k} - A\boldsymbol{x}^{(0)}$ is the residual associated with the initial guess. The two most common approaches can be characterized in terms of the residual $\boldsymbol{r}^{(m)}$ of the mth approximation $\boldsymbol{x}^{(0)} + \boldsymbol{c}^{(m)}$, i.e., $\boldsymbol{r}^{(m)} = \boldsymbol{r}^{(0)} - A\boldsymbol{c}^{(m)}$. Denoting $\mathcal{V}_m := \mathcal{K}_m(A, \boldsymbol{r}^{(0)})$ and $\mathcal{W}_m := A\mathcal{V}_m$, the *orthogonal residual (OR) method* is characterized by the condition $\boldsymbol{r}^{(m)} \perp \mathcal{V}_m$, which defines a unique approximation provided $\mathbb{C}^N = \mathcal{V}_m \oplus \mathcal{W}_m^{\perp}$, which for simplicity we assume to hold at each step in the sequel. The *minimal residual (MR) method*, on the other hand, selects the correction to minimize the norm of the residual $\boldsymbol{r}^{(m)}$ over all corrections $\boldsymbol{c}^{(m)} \in \mathcal{V}_m$, which is equivalent with $\boldsymbol{r}^{(m)} \perp \mathcal{W}_m$. Examples of implementations of OR methods are the conjugate gradient method, the full orthogonalization method and the biconjugate gradient method. Examples of MR algorithms are the conjugate residual method, the MINRES, GMRES and QMR.

The calculations are usually performed with regard to a convenient basis of the Krylov space $\mathcal{V}_m$ generated by some variant of the *Arnoldi process*[1], which, beginning with $\boldsymbol{v}^{(1)} = \boldsymbol{r}^{(0)}/\|\boldsymbol{r}^{(0)}\|$, generates an ascending sequence of orthonormal basis vectors $\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \dots\}$ of the sequence of spaces $\mathcal{V}_m$ by multiplying the most recently generated basis vector by A and orthogonalizing against the current basis. Writing $V^{(m)} := [\boldsymbol{v}^{(1)}\ \boldsymbol{v}^{(2)}\ \cdots\ \boldsymbol{v}^{(m)}]$, the action of A on this basis is described by the *Arnoldi decomposition*

$$AV^{(m)} = V^{(m+1)}\widetilde{H}_m \tag{21}$$

in terms of an unreduced upper Hessenberg matrix $\widetilde{H}_m \in \mathbb{C}^{(m+1)\times m}$. Denoting the correction vector at step m by $V^{(m)}\boldsymbol{z}$ with a coordinate vector $\boldsymbol{z} \in \mathbb{C}^m$, the residual may be written

$$\boldsymbol{r}^{(m)} = \boldsymbol{r}^{(0)} - AV^{(m)}\boldsymbol{z} = \boldsymbol{r}^{(0)} - V^{(m+1)}\widetilde{H}_m\boldsymbol{z} = V^{(m+1)}\left(\|\boldsymbol{r}^{(0)}\|\boldsymbol{u}_1 - \widetilde{H}_m\boldsymbol{z}\right) \tag{22}$$

where $\boldsymbol{u}_1 \in \mathbb{C}^{m+1}$ denotes the first unit coordinate vector. From this representation it is easily seen that the coordinate vector for the OR correction is obtained as the solution of the equation

$$H_m\boldsymbol{z} = \|\boldsymbol{r}^{(0)}\|\boldsymbol{u}_1$$

with H_m defined as the square Hessenberg matrix obtained by striking the last row of $\widetilde{H}_m$ and $\boldsymbol{u}_1$ the first unit coordinate vector in $\mathbb{C}^m$. Similarly, the coordinate vector of the MR correction is the solution to the full-rank least-squares problem

$$\widetilde{H}_m\boldsymbol{z} \approx \|\boldsymbol{r}^{(0)}\|\boldsymbol{u}_1$$

[1] There are also methods such as BCG and QMR based on the Lanczos process, but, as shown e.g. in [6], at least in the more difficult non-Hermitian case, the coordinate calculations are the same.

with $\boldsymbol{u}_1$ the first unit coordinate vector in $\mathbb{C}^{m+1}$. For a comprehensive overview of Krylov subspace methods we refer to Eiermann and Ernst [6] and the references given therein.

6.2. **Simplification of the Arnoldi Decomposition.** As is easily verified, replacing A by $A + \alpha I$ in (20) leaves the Krylov space unchanged. In particular, in the case $A = I - B$ under consideration, we have $\mathcal{K}_m(A, \boldsymbol{r}^{(0)}) = \mathcal{K}_m(B, \boldsymbol{r}^{(0)})$. Moreover, when A is p-cyclic with B in the normal form (1), we may assume the initial residual has the form

$$\boldsymbol{r}^{(0)} = \begin{bmatrix} \boldsymbol{r}_1^{(0)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \tag{23}$$

which results from fixing the first block $\boldsymbol{x}_1^{(0)}$ of the initial approximation and obtaining the remaining blocks $\boldsymbol{x}_2^{(0)}, \dots, \boldsymbol{x}_p^{(0)}$ by (forward) substitution. The special structure (23) of $\boldsymbol{r}^{(0)}$ combined with the weakly p-cyclic structure (1) of B now carries over to $\mathcal{K}_m(A, \boldsymbol{r}^{(0)}) = \mathcal{K}_m(B, \boldsymbol{r}^{(0)})$ in that

$$\mathcal{K}_m(B, \boldsymbol{r}^{(0)}) = \operatorname{span}\left\{ \begin{bmatrix} \boldsymbol{v}_1^{(1)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{v}_2^{(2)} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \dots, \begin{bmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{v}_p^{(p)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{v}_1^{(p+1)} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \dots \right\}$$

with suitable vectors $\boldsymbol{v}_j^{(k)}$, $j = 1, \dots, p$, $k \geq 1$. Consequently, we may write the Krylov space with index $m = kp + j$ as the orthogonal and direct sum

$$\mathcal{K}_m(B, \boldsymbol{r}^{(0)}) = \mathcal{V}_1^{(k+1)} \dot{\oplus} \dots \dot{\oplus} \mathcal{V}_j^{(k+1)} \dot{\oplus} \mathcal{V}_{j+1}^{(k)} \dot{\oplus} \dots \dot{\oplus} \mathcal{V}_p^{(k)} \tag{24}$$

in which $\mathcal{V}_q^{(r)} := \mathcal{K}_r(B^p, B^{q-1}\boldsymbol{r}^{(0)})$.

The decomposition (24) of $\mathcal{V}_m$ implies that the Hessenberg matrix $\widetilde{H}_m$ has a special structure: Its only nonzero entries lie on the first subdiagonal and, beginning with the main diagonal, on every p-th upper diagonal. It is convenient to reorder the basis vectors $V^{(m)}$ in such a way that those belonging to each of the spaces $\mathcal{V}_j^{(k)}, j = 1, \dots, p$, are adjacent. If this reordering is effected by the permutation matrix P, then we define $\hat{V}^{(m)}$ for $m = kp + j$ by

$$\hat{V}^{(kp+j)} := V^{(kp+j)} P = [V_1^{(k+1)} \dots V_j^{(k+1)} V_{j+1}^{(k)} \dots V_p^{(k)}].$$

in terms of which the Arnoldi decomposition (21) takes on the form

$$(I - B)\hat{V}^{(kp+j)} = \hat{V}^{(kp+j+1)} \hat{\widetilde{H}}_{kp+j}$$

with an accordingly reordered Hessenberg matrix $\hat{\tilde{H}}_m$ given by

$$\hat{\tilde{H}}_m = \begin{bmatrix} I_{k+1} & & & & & & -\widetilde{T}_1^{(k)} \\ -T_2^{(k+1)} & I_{k+1} & & & & & \\ & \ddots & \ddots & & & & \\ & & -T_j^{(k+1)} & \widetilde{I}_k & & & \\ & & & -T_{j+1}^{(k)} & I_k & & \\ & & & & \ddots & \ddots & \\ & & & & & -T_p^{(k)} & I_k \end{bmatrix}. \tag{25}$$

The matrix $\widetilde{T}_1^{(k)} \in \mathbb{C}^{(k+1)\times k}$ in the $(1,p)$ position is upper Hessenberg, the subdiagonal blocks $T_j^{(\ell)} \in \mathbb{C}^{\ell\times\ell}, j = 2,\ldots,p$, are upper triangular, I_ℓ is the $\ell\times\ell$ identity and $\widetilde{I}_\ell = \left[\begin{smallmatrix} I_\ell \\ \boldsymbol{0} \end{smallmatrix}\right]$.

Note that the Hessenberg matrix of the Arnoldi process applied to B in place of $A = I - B$, which we denote by $\hat{\tilde{T}}_m$, is obtained by deleting the identity blocks as well as $\widetilde{I}_k$ from $\hat{\tilde{H}}_m$. Furthermore, it is easily seen that the same set of basis vectors is generated. Thus, at step $m = kp$ there holds

$$B\,[V_1^{(k)} V_2^{(k)} \cdots V_p^{(k)}] = [V_1^{(k+1)} V_2^{(k)} \cdots V_p^{(k)}]\hat{\tilde{T}}_{kp},$$

from which we successively derive the block equations

$$\begin{aligned} BV_1^{(k)} &= V_2^{(k)} T_2^{(k)} \\ B^2 V_1^{(k)} &= BV_2^{(k)} T_2^{(k)} = V_3^{(k)} T_3^{(k)} T_2^{(k)} \\ &\;\;\vdots \\ B^p V_1^{(k)} &= V_1^{(k+1)} \widetilde{T}_1^{(k)} \underbrace{T_p^{(k)} \cdots T_2^{(k)}}_{=:T^{(k)}} =: V_1^{(k+1)} \widetilde{T}_1^{(k)} T^{(k)}, \end{aligned}$$

the last of which yields the relation

$$(I - B^p)V_1^{(k)} = V_1^{(k+1)} \left(\begin{bmatrix} I_k \\ \boldsymbol{0} \end{bmatrix} - \widetilde{T}_1^{(k)} T^{(k)} \right). \tag{26}$$

The identity (26) is the Arnoldi decomposition of the Krylov space $\mathcal{K}_k(I - B^p, \boldsymbol{r}^{(0)})$ associated with the uncoupled system (4) and initial residual as in (23). In particular, since $\boldsymbol{r}^{(0)}$ is only nonzero in the first block, the same is true for all vectors in the Krylov space $\mathcal{K}_k(I - B^p, \boldsymbol{r}^{(0)})$. As a result, all blocks in (26) but the first are zero, and, except for the $p-1$ zero blocks, (26) coincides with the Arnoldi decomposition after k steps of $\mathcal{K}_k(I - B_1^{(p)}, \boldsymbol{r}_1^{(0)})$ associated with the first reduced system.

As a result, we see that the Arnoldi decomposition of both $I - B$ and $I - B^p$ with B in the normal form (1) of a weakly cyclic matrix of index p and an initial vector of the form (23) can be computed by multiplying only small matrices with subvectors. Moreover, the orthogonality constraints across block boundaries are enforced automatically, requiring only orthogonalization within each block in the construction of $V^{(m)}$.

6.3. **The OR Method for p-cyclic Systems.** The expression (22) for the residual of a Krylov subspace method shows that $\boldsymbol{r}^{(m)} \in \mathcal{V}_{m+1}$. The OR residual additionally satisfies $\boldsymbol{r}^{(m)} \perp \mathcal{V}_m$, so that, if $\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \dots\}$ denotes an ascending orthonormal basis of the sequence of spaces $\{\mathcal{V}_m\}_{m\geq 1}$, then the OR residual is collinear with the basis vector $\boldsymbol{v}^{(m+1)}$.

In view of the decomposition (24), this implies that the OR residual in steps $m = kp$, $k \geq 0$ is therefore a multiple of $\boldsymbol{v}^{(kp+1)}$, which, if (23) holds, is zero in all blocks but the first. Consequently, the characterization

$$\boldsymbol{r}^{(kp)} \perp \mathcal{K}_{kp}(A, \boldsymbol{r}^{(0)}) = \mathcal{K}_{kp}(B, \boldsymbol{r}^{(0)})$$

of the OR residual at step $m = kp$ is equivalent with $\boldsymbol{r}^{(kp)} \perp \mathcal{K}_k(B^p, \boldsymbol{r}^{(0)})$, which in turn is equivalent to $\boldsymbol{r}_1^{(kp)} \perp \mathcal{K}_k(B_1^{(p)}, \boldsymbol{r}_1^{(0)})$. The last relation is recognized as the characterization of the OR residual at step $m = k$ of the OR method applied to the first reduced system (5). We summarize this result in the following theorem.

Theorem 6.1. *If the OR method is applied to the linear system* (2) *with a nonsingular p-cyclic matrix A and an initial residual of the form* (23)*, then the residual at step $m = pk$ is zero in all blocks but the first and the first block coincides with the residual at step $m = k$ of the OR method applied to the first block of the first reduced system* (5)*.*

The first result of this type is due to Reid [24], who considered the conjugate gradient method for Hermitian 2-cyclic matrices and showed that it is possible to compute only the first block of the iterates with even index

$$\boldsymbol{x}^{(2m)} \in \boldsymbol{x}^{(0)} + \mathcal{K}_{2m}(B, \boldsymbol{r}^{(0)}) = \boldsymbol{x}^{(0)} + \left(\mathcal{K}_m(B^2, \boldsymbol{r}^{(0)}) \dot{\oplus} \mathcal{K}_m(B^2, B\boldsymbol{r}^{(0)})\right)$$

in addition to all residuals, which possess a zero block which alternates in position. Reid's variant of CG is sometimes called the *2-cyclic conjugate gradient method*. Except for the residuals with odd index, these are the same quantities computed when solving the reduced system by conjugate gradients. Later, Hageman, Luk and Young [18] show that the iterate at step $m = 2k$ of the conjugate gradient method to a Hermitian 2-cyclic system coincides with the iterate at step k of the 2-cyclic conjugate gradient method as well as with the iterate at step k of these methods applied to the reduced system. Connections between the conjugate gradient method applied to the reduced problems and the original problem are also treated in Evans and Li [9], see also [3].

Another interesting equivalence result is due to Freund [11], where the iterative solution of least-squares problems is investigated, which can be reformulated as respectively 2-cyclic or 3-cyclic linear systems of equations. Freund shows that the conjugate gradient method applied to the suitably preconditioned normal equations produces iterates which lie in the same Krylov space as the iterates of two cyclic SOR schemes which had been proposed earlier by Plemmons [23] and Markham et al.[21]. By virtue of the minimization property of the conjugate gradient method, this equivalence result implies that the resulting algorithm is superior to the two SOR schemes.

6.4. **The MR Method for p-cyclic Systems.** That an analogous equivalence result to Theorem 6.1 fails to hold for the MR method was first observed by Freund, Golub and Hochbruck [12] and we formulate this result in Theorem 6.2 below. Although solving the reduced system involves working with short vectors in $\mathcal{K}_k(B_1^{(p)}, \boldsymbol{r}_1^{(0)})$, we identify these in the following with their counterparts in $\mathcal{K}_k(B^p, \boldsymbol{r}^{(0)})$.

Theorem 6.2. *If the MR method is applied to the linear system* (2) *with a nonsingular p-cyclic matrix A and an initial residual of the form* (23), *then the residual at step $m = pk$ only coincides with the residual at step $m = k$ of the MR method applied to the first reduced system* (5) *when the residual is zero.*

Proof. Selecting the MR correction $\boldsymbol{c}^{(m)} \in \mathcal{V}_m$ to minimize the norm of the residual $\boldsymbol{r}^{(m)} = \boldsymbol{r}^{(0)} - (I - B)\boldsymbol{c}^{(m)}$ is equivalent with determining $\boldsymbol{c}^{(m)}$ such that its image under $I - B$ is the best approximation to $\boldsymbol{r}^{(0)}$ from $(I - B)\mathcal{V}_m$. By (24) for $m = kp$, we have

$$\mathcal{V}_{kp} = \mathcal{V}_1^{(k)} \dot{\oplus} \dots \dot{\oplus} \mathcal{V}_p^{(k)}.$$

In addition, it is easily verified that

$$B\mathcal{V}_j^{(k)} = \mathcal{V}_{j+1}^{(k)}, \quad j = 1, \dots, p-1, \qquad B\mathcal{V}_p^{(k)} = B^p \mathcal{V}_1^{(k)}. \tag{27}$$

Therefore, since $\boldsymbol{r}^{(0)} \perp \mathcal{V}_j^{(k)}$ for $j = 2, \dots, p$, we have

$$(I - B)\mathcal{V}_j^{(k)} \perp \boldsymbol{r}^{(0)} \quad \text{for} \quad j = 2, \dots, p-1,$$

so that only the spaces $(I - B)\mathcal{V}_1^{(k)}$ and $(I - B)\mathcal{V}_p^{(k)}$ contribute to the approximation of $\boldsymbol{r}^{(0)}$ at step kp.

Denote by $\boldsymbol{r}_{\text{red}}^{(k)} \in \mathcal{V}_1^{(k+1)}$ the residual at step $m = k$ of the MR method applied to the first reduced system. In order for $\boldsymbol{r}_{\text{red}}^{(k)} = \boldsymbol{r}^{(kp)}$ to hold, it is necessary that $\boldsymbol{r}_{\text{red}}^{(k)} \perp (I - B)\mathcal{V}_{kp}$. By virtue of $\boldsymbol{r}_{\text{red}}^{(k)} \in \mathcal{V}_1^{(k+1)}$, its orthogonality against $(I - B)\mathcal{V}_j^{(k)}$ for $j = 2, \dots, p-1$ is evident by the same reasoning as for $\boldsymbol{r}^{(kp)}$. In view of (27), orthogonality against $(I - B)\mathcal{V}_1^{(k)}$ means

$$0 = (\boldsymbol{r}_{\text{red}}^{(k)}, (I - B)\boldsymbol{v}_1) = (\boldsymbol{r}_{\text{red}}^{(k)}, \boldsymbol{v}_1) \quad \text{for all} \quad \boldsymbol{v}_1 \in \mathcal{V}_1^{(k)},$$

and similarly, orthogonality against $(I - B)\mathcal{V}_p^{(k)}$ implies

$$0 = (\boldsymbol{r}_{\text{red}}^{(k)}, (I - B)\boldsymbol{v}_p) = (\boldsymbol{r}_{\text{red}}^{(k)}, B\boldsymbol{v}_p) \quad \text{for all} \quad \boldsymbol{v}_p \in \mathcal{V}_p^{(k)}.$$

Again by (27), the last two relations taken together imply $\boldsymbol{r}_{\text{red}}^{(k)} \perp (\mathcal{V}_1^{(k)} + B^p \mathcal{V}_1^{(k)})$. But the space on the right is equal to $\mathcal{V}_1^{(k+1)}$ and thus $\boldsymbol{r}_{\text{red}} = \boldsymbol{0}$.

The equivalence of the OR method applied to the reduced system implies that it finds the exact solution at step $m = k$ if and only if OR applied to the full system does so in step $m = kp$. The same holds for the termination with the exact solution of the MR method since MR terminates with the exact solution in the same step as OR.

In other words, until the MR method applied to either system has terminated with the exact solution, the approach of solving the reduced system is *not* equivalent with solving the full system. □

A natural question now is, of course, whether it is better to solve the reduced system when using the MR method. This question is still open to the author's knowledge. One difficulty is that both methods approximate $\boldsymbol{r}^{(0)}$ from the distinct spaces $(I - B^p)\mathcal{V}_1^{(k)}$ and $(I - B)(\mathcal{V}_1^{(k)} + \mathcal{V}_p^{(k)})$, neither of which is a subspace of the other. Of course, we have

$$(I - B)(\mathcal{V}_1^{(k)} + \mathcal{V}_p^{(k)}) \subset \mathcal{V}_1^{(k)} + B\mathcal{V}_1^{(k)} + \mathcal{V}_p^{(k)} + B\mathcal{V}_p^{(k)} = \mathcal{V}_1^{(k)} + \mathcal{V}_2^{(k)} + \mathcal{V}_p^{(k)} + B^p\mathcal{V}_1^{(k)}$$

and the slighly enlarged space on the right contains the approximation space used by MR applied to the reduced system, which shows that the two approximations cannot be too far apart. There is, however, nothing to prevent the MR approximation applied to the reduced system from yielding a smaller residual at a given step than when applied to the full system.

References

[1] Oscar Buneman. A compact non-iterative Poisson solver. Technical Report 294, Stanford University Institute for Plasma Research, Stanford, Ca., 1969.

[2] B. L. Buzbee, F. W. Dorr, J. A. George, and G.H. Golub. The direct solution of the discrete Poisson equation on irregular regions. *SIAM J. Numer. Anal.*, 8:722–736, 1974.

[3] Paul Concus, Gene H. Golub, and Dianne P. O'Leary. A generalized conjugate gradient method for solution of elliptic partial differential equations. In J. R. Bunch and D. J. Rose, editors, *Proc. Symp. on Sparse Matrix Computations.* Argonne National Lab.(Sept. 1975), Academic Press, New York, 1976.

[4] Michael Eiermann. Accelerating p-cyclic SOR. to appear.

[5] Michael Eiermann. Semiiterative Verfahren für nichtsymmetrische lineare Gleichungssysteme. Habilitationsschrift, Universität Karlsruhe, 1989.

[6] Michael Eiermann and Oliver G. Ernst. Geometric aspects in the theory of Krylov subspace methods. 1998. to appear in Acta Numerica 2001.

[7] Michael Eiermann, Wilhelm Niethammer, and Arden Ruttan. Optimal successive overrelaxation iterative methods for p-cyclic matrices. *Numer. Math.*, 57(6/7):593–606, 1990.

[8] Howard C. Elman and Gene H. Golub. Iterative methods for cyclically reduced non-selfadjoint linear systems. *Math. Comp.*, 54:167–700, 1990.

[9] D. J. Evans and C. Li. Connections between the PCG method applied to systems with red/black form and its reduced system. *BIT*, 29:741–747, 1989.

[10] George E. Forsythe. Solving linear equations can be interesting. *Bull. AMS*, 59:299–329, 1953.

[11] Roland W. Freund. A note on two block-SOR methods for sparse least squares problems. *Lin. Alg. Appl.*, 88/89:211–221, 1987.

[12] Roland W. Freund, Gene H. Golub, and Marlis Hochbruck. Krylov subspace methods for non-Hermitian p-cyclic matrices. unpublished manuscript.

[13] G. Frobenius. Über Matrizen aus nicht-negativen Elementen. *Sitzungsber. preuss. Akad. Wiss.*,, Math. Nat. Kl.:456–477, 1912.

[14] Sofoklis Galanis, Apostolos Hadjidimos, and Dimitrios Noutsos. On the equivalence of the k-step iterative Euler methods and successive overrelaxation (SOR) mathods for k-cyclic matrices. *Math. Comput. Simulation*, 30:213–230, 1988.

[15] Gene H. Golub. *The Use of Chebyshev Matrix Polynomials in the Iterative Solution of Linear Equations Compared with the Method of Successive Overrelaxation.* PhD thesis, University of Illinois, 1959.

[16] Gene H. Golub and Richard S. Varga. Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order Richardson iterative methods (parts I and II). *Numer. Math.*, 3:147–168, 1961.

[17] Martin H. Gutknecht, Wilhelm Niethammer, and Richard S. Varga. k-step iterative methods for solving nonlinear systems of equations. *Numer. Math.*, 48:699–712, 1986.

[18] Louis A. Hageman, Franklin T. Luk, and David M. Young. On the equivalence of certain iterative acceleration methods. *SIAM J. Numer. Anal.*, 17:852–873, 1980.

[19] Louis A. Hageman and Richard S. Varga. Block iterative methods for cyclically reduced matrix equations. *Numer. Math.*, 6:106–119, 1964.

[20] Roger W. Hockney. The potential calculation and some applications. *Meth. Comp. Physics*, 9:136–212, 1969.

[21] T. L. Markham, M. Neumann, and R. J. Plemmons. Convergence of a direct-iterative method for large-scale least-squares problems. *Lin. Alg. Appl.*, 69:155–167, 1985.

[22] Wilhelm Niethammer and Richard S. Varga. The analysis of k-step iterative methods for linear systems from summability theory. *Numer. Math.*, 41:177–206, 1983.

[23] Robert J. Plemmons. Adjustment by least squares in geodesy using block iterative methods for sparse matrices. In *Proceedings of the 1979 Army Numerical Analysis and Computers Conference, El Paso, Tex.*, pages 151–186. ARO Report 79-3, 1979.

[24] John K. Reid. The use of conjugate gradients for systems of linear equations possessing "Property A". *SIAM J. Numer. Anal.*, 9:325–332, 1972.

[25] V. Romanovsky. Recherches sur les chaînes de Markoff. *Acta Math.*, 66:147–251, 1936.

[26] Johannes Schröder. Zur Lösung von Potentialaufgaben mit Hilfe des Differenzenverfahrens. *Z. Angew. Math. Mech.*, 34:241–253, 1954.

[27] W. J. Stewart. *Introduction to the Numerical Solution of Markov Chains.* Princeton University Press, Princeton, NJ, 1994.

[28] Paul N. Swarztrauber. A direct method for the discrete solution of separable elliptic equations. *SIAM J. Numer. Analysis*, 11:1136–1150, 1974.

[29] Paul N. Swarztrauber and Roland A. Sweet. Efficient fortran subprograms for the solution of separable elliptic partial differential equations. *ACM Trans. Math. Software*, 5:351–371, 1979.

[30] Roland A. Sweet. A cyclic reduction algorithm for block tridiagonal systems of arbitrary dimension. *SIAM J. Numer. Analysis*, 14:706–720, 1977.

[31] G. J. Tee. An application of p-cyclic matrices for solving periodic problems. *Numer. Math.*, 6:142–159, 1964.

[32] M. C. Thompson, Joel H. Ferziger, and Gene H. Golub. Block SOR applied to cyclically-reduced equations as an efficient solution technique for convection-diffusion equations. In J. Noyce and C. Fletcher, editors, *Computational Techniques and Applications: CTAC-87.* Elsevier Science Publishers B. V. (North-Holland), 1988.

[33] Richard S. Varga. A comparison of the succesive overrelaxation method and the semi-iterative methods using chebyshev polynomials. *J. Soc. Indust. Appl. Math.*, 5(2):39–46, 1957.

[34] Richard S. Varga. p-cyclic matrices: a generalization of the Young-Frankel successive overrelaxation scheme. *Pacific J. Math.*, 9:617–628, 1959.

[35] Richard S. Varga. *Matrix Iterative Analysis.* Springer-Verlag, Berlin/Heidelberg, second edition, 1999.

[36] David M. Young. *Iterative Methods for Solving Partial Difference Equations of Elliptic Type.* PhD thesis, Harvard Univ., 1950.

Institut für Angewandte Mathematik II, TU Bergakademie Freiberg, 09596 Freiberg, Germany

E-mail address: `ernst@math.tu-freiberg.de`