to be submitted. 

# ABSTRACT PERTURBED KRYLOV METHODS*

JENS-PETER M. ZEMKE†

**Abstract.** We introduce the framework of "abstract perturbed Krylov methods". This is a new and unifying point of view on Krylov subspace methods based solely on the matrix equation $AQ_k + F_k = Q_{k+1}\underline{C}_k = Q_kC_k + q_{k+1}c_{k+1,k}e_k^T$ and the assumption that the matrix $C_k$ is unreduced Hessenberg. We give polynomial expressions relating the Ritz vectors, (Q)OR iterates and (Q)MR iterates to the starting vector $q_1$ and the perturbation terms $\{f_l\}_{l=1}^k$. The properties of these polynomials and similarities between them are analyzed in some detail. The results suggest the interpretation of abstract *perturbed* Krylov methods as additive overlay of several abstract *exact* Krylov methods.

**Key words.** Abstract perturbed Krylov method, inexact Krylov method, finite precision, Hessenberg matrix, basis polynomial, adjugate polynomial, residual polynomial, quasi-kernel polynomial, Ritz vectors, (Q)OR, (Q)MR.

**AMS subject classifications.** 15A24 (primary), 65Y20, 15A15, 65F10, 65F15, 65F20

## 1. Introduction.

We consider the matrix equation

$$AQ_k + F_k = Q_{k+1}\underline{C}_k = Q_kC_k + M_k = Q_kC_k + q_{k+1}c_{k+1,k}e_k^T, \tag{1.1}$$

and thus implicitly every perturbed Krylov subspace method that can be written in this form. We refer to equation (1.1) as a *perturbed* Krylov *decomposition* and to any instance of such an equation as an *abstract perturbed* Krylov *method*. In the remainder of the introduction we clarify the rôle the particular ingredients take in this perturbed Krylov decomposition, motivate the present note and introduce notations.

The matrix $A \in \mathbb{C}^{n\times n}$ is either the system matrix of a linear system of equations, or we are interested in some of its eigenvalues and maybe the corresponding eigenvectors,

$$Ax = b - Ax_0 = r_0 \qquad \text{or} \qquad Av = v\lambda. \tag{1.2}$$

The matrix $Q_k \in \mathbb{C}^{n\times k}$ and its expanded counterpart $Q_{k+1} \in \mathbb{C}^{n\times k+1}$ collect as column vectors the vectors $q_1, q_2, \ldots, q_k \in \mathbb{C}^n$ (and $q_{k+1} \in \mathbb{C}^n$). In some of the methods under consideration, these vectors form at least in some circumstances a basis of the underlying unperturbed Krylov subspace. To smoothen the understanding process we will laxly speak in *all* cases of them as the basis vectors of the possibly perturbed Krylov subspace method. The gain lies in the simplicity of this notion; the justification is that we are seldom interested in the true basis of the unperturbed Krylov subspace and that in most cases there does not exist a nearby Krylov subspace at all, the reasons becoming obvious in Theorem 2.3 and Theorem 2.4.

The matrix $F_k \in \mathbb{C}^{n\times k}$ is to be considered as a perturbation term. This perturbation term may be zero; in this case all results we derive in this note remain valid and make statements about the unperturbed Krylov subspace methods. The perturbation term may be due to a balancing of the equation necessary because of execution in finite precision; in this case, the term will frequently in some sense be small and we usually have bounds or estimates on the norms of the column vectors.

*Report 89 des Arbeitsbereiches Mathematik 4-13, © Juli 2005 Jens-Peter M. Zemke

†Technische Universität Hamburg-Harburg, Schwarzenbergstraße 95, Arbeitsbereich Mathematik 4-13, 21073 Hamburg, Germany (zemke@tu-harburg.de)

The term may arise from a so-called *inexact* KRYLOV subspace method [1, 7, 10]; in this case the columns of $F_k$ are due to inexact matrix-vector multiplies, we have control on the norms of the perturbation terms and are interested in the upper bounds on the magnitudes that do not spoil the convergence of the method. Of course the perturbation terms in the latter two settings have to be combined when one wishes to understand the properties of inexact KRYLOV subspace methods executed in finite precision.

The matrix $C_k \in \mathbb{C}^{k\times k}$ is an unreduced upper HESSENBERG matrix, frequently an approximation to a projection of $A$ onto the space spanned by $Q_k$. The capital letter $C$ should remind of *condensed* and, especially in the perturbed variants, *computed*. The square matrix $C_k$ is used to construct approximations to eigenvalues and eigenvectors and, in context of (Q)OR KRYLOV subspace methods, to construct approximations to the solution of a linear system. It is essential for our investigations that $C_k$ is an unreduced upper HESSENBERG matrix. The properties of unreduced upper HESSENBERG matrices have recently been investigated by ZEMKE [12] and allow the subsequent results for KRYLOV subspace methods, which are a refinement of the results proven in ZEMKE's Dissertation [11].

The matrix $\underline{C}_k \in \mathbb{C}^{k+1\times k}$ is an *extended* unreduced upper HESSENBERG matrix. The rectangular matrix $\underline{C}_k$ is used in context of (Q)MR KRYLOV subspace methods to construct approximations to the solution of a linear system. The notation $\underline{C}_k$ should remind of an additional *row* which is appended to the *bottom* of the HESSENBERG matrix $C_k$ and seems to have been coined independently by SLEIJPEN [8] and GUTKNECHT [4]. We feel that this notation should be preferred against other attempts of notation like $\overline{C}_k$, $\tilde{C}_k$ or even $C_k^e$.

**1.1. Motivation.** In this note we consider some of the interesting properties of quantities related to equation (1.1). The only and crucial assumption is that the matrix $C_k$ is unreduced HESSENBERG. The good news is that most simple KRYLOV subspace methods are captured by equation (1.1). The startling news is that additionally some methods with a rather strange behavior are also covered. For a brief account of some of the methods covered we refer to [11].

Most numerical analysts will agree that there is a interest in the proper understanding of KRYLOV subspace methods, especially of the finite precision and inexact variants. The "usual" branch of investigation picks *one* variant of *one* method for *one* task in *one* "flavor", say the *finite precision* variant of the *symmetric* method of LANCZOS for the solution of the partial *eigenvalue* problem, implemented in the "stablest" *variant (A1)* as categorized by PAIGE. The beautiful analysis relies heavily on the properties of this particular method, in the case mentioned the so-called *local orthogonality* of the computed basis vectors, the *symmetry* of the computed unreduced *tridiagonal matrices* $T_k \equiv C_k \in \mathbb{R}^{k\times k}$ and the underlying *short-term recurrence.*

The subsumption of *several* methods that are quite distinct in nature under *one* common *abstract* framework undertaken in this paper will most probably be considered to be rather strange, if not useless, or even harmful. Quoting the Merriam-Webster Online Dictionary, the verb "abstract" means "to consider apart from application to or association with a particular instance" and the adjective "abstract" means "disassociated from any specific instance". The framework developed in this paper tries to strike a balance between the benefits of such an abstraction, e.g. unification and derivation of *qualitative* results, and the loss of knowledge necessary to give any *quantitative* results, e.g. the convergence of a method in finite precision.

Below we prove that the quantities associated to KRYLOV subspace methods, i.e., the Ritz vectors, the (Q)OR and (Q)MR iterates, and their corresponding residuals and errors, of *any* Krylov subspace method covered by equation (1.1) can be described in terms of polynomials related *solely* to the computed $C_k$ or $\underline{C}_k$. These results could have been achieved without the setting of abstract perturbed Krylov methods, but focusing from the *beginning* on a particular instance, e.g. the inexact variant of the method of Arnoldi or unperturbed BiCGSTAB clouds the view for such intrinsic properties and would presumably result in yet another large amount of articles proving essentially the same result for every particular method.

The qualitative results achieved in this paper might be considered as companion to the "classical" results; focusing on a particular instance, e.g. the aforementioned variant (A1) of the method of Lanczos executed in finite precision, we can utilize results about the convergence of Ritz values to make predictions on the expected behavior of the other quantities, e.g. the Ritz vectors and their relation to eigenvectors of $A$.

**1.2. Notation.** As usual, $I = I_k$ denotes the identity of appropriate dimensions. The columns are the standard unit vectors $e_j$, the elements are Kronecker delta $\delta_{ij}$. The letter $O = O_k$ denotes a zero matrix of appropriate dimensions. Zero vectors of length $j$ are denoted by $o_j$. Standard unit vectors padded with one additional zero element at the bottom are denoted by

$$\underline{e}_j \equiv \begin{pmatrix} e_j \\ 0 \end{pmatrix} \in \mathbb{C}^{k+1}, \quad e_j \in \mathbb{C}^k. \tag{1.3}$$

The identity matrix of size $k \times k$ padded with an additional zero row at the bottom is denoted by $\underline{I}_k$.

The characteristic matrices $zI - A$ and $zI - C_k$ are denoted by ${}^{z}\!A$ and ${}^{z}C_k$. The characteristic polynomials $\chi(z)$ of $A$ and $\chi_k(z)$ of $C_k$ are defined by $\chi(z) \equiv \det({}^{z}\!A)$ and $\chi_k(z) \equiv \det({}^{z}C_k)$, respectively. Let $\theta$ be an eigenvalue of $C_k$ of algebraic multiplicity $\alpha \equiv \alpha(\theta)$. The reduced characteristic polynomial $\omega_k(z)$ of $C_k$ corresponding to $\theta$ is defined by

$$\chi_k(z) = (z - \theta)^\alpha \omega_k(z). \tag{1.4}$$

We remark that $\omega_k(\theta) \neq 0$. We make extensive use of other characteristic polynomials denoted by $\chi_{i:j}$ and defined by

$$\chi_{i:j}(z) \equiv \det({}^{z}C_{i:j}) \equiv \det(zI - C_{i:j}), \tag{1.5}$$

where $C_{i:j}$ denotes the principal submatrix of $C_k$ consisting of the columns and rows $i$ to $j$, for all admissible indices $1 \leqslant i \leqslant j \leqslant k$. We extend the notation to the case $i = j + 1$ by setting $\chi_{j+1:j}(z) \equiv 1$ for all $j$. Additionally we define the shorthand notation $\chi_j = \chi_{1:j}$.

We denote products of subdiagonal elements of the unreduced Hessenberg matrices $C_k$ by $c_{i:j} \equiv \prod_{l=i}^{j} c_{l+1,l}$. Polynomial vectors $\nu$ and $\check{\nu}$ are defined by

$$\nu(z) \equiv \left( \frac{\chi_{j+1:k}(z)}{c_{j:k-1}} \right)_{j=1}^{k} \quad \text{and} \quad \check{\nu}(z) \equiv \left( \frac{\chi_{1:j-1}(z)}{c_{1:j-1}} \right)_{j=1}^{k}. \tag{1.6}$$

The elements are denoted by $\nu_j(z)$ and $\check{\nu}_j(z)$, where $j$ runs from 1 to $k$. The notation is extended to include the polynomial $\check{\nu}_{k+1}(z)$, defined by

$$\check{\nu}_{k+1}(z) \equiv \frac{\chi_{1:k}(z)}{c_{1:k}}. \tag{1.7}$$

We extend the polynomial vector $\check{\nu}$ by padding it in the last position with the polynomial $\check{\nu}_{k+1}$, denoted by $\underline{\check{\nu}}$,

$$\underline{\check{\nu}}(z) \equiv \begin{pmatrix} \check{\nu}(z)^T & \check{\nu}_{k+1}(z) \end{pmatrix}^T = \begin{pmatrix} \check{\nu}_1(z) & \cdots & \check{\nu}_k(z) & \check{\nu}_{k+1}(z) \end{pmatrix}^T. \tag{1.8}$$

We denote the complex conjugates of $\check{\nu}$, $\check{\nu}_j$, $\underline{\check{\nu}}$ by $\hat{\nu}$, $\hat{\nu}_j$, $\underline{\hat{\nu}}$, respectively. The operation "complex conjugate" can be memorized as a reflection on the real axis turning hat to vee and vice versa.

The Jordan normal forms of $A$ and $C_k$ are denoted by $J$ and $J_k$, respectively. Similarity transformations $V$ and $S_k$ are chosen to satisfy

$$V^{-1}AV = J, \qquad S_k^{-1}C_kS_k = J_k. \tag{1.9}$$

We call *any* matrices $V$ and $S_k$ that satisfy equation (1.9) *(right) eigenmatrices* and define corresponding (special) *left eigenmatrices* by

$$\hat{V}^H \equiv \check{V}^T \equiv V^{-1}, \qquad \hat{S}_k^H \equiv \check{S}_k^T \equiv S_k^{-1}. \tag{1.10}$$

This definition ensures the biorthogonality of left and right eigenmatrices, which is a partial normalization of the eigen- and principal vectors. When $A$ is normal, we set $\hat{V} \equiv V$, when $C_k$ is normal, we set $\hat{S}_k \equiv S_k$. In any of these cases the eigenvectors are normalized to have unit length, since the *bi*orthogonality simplifies to orthogonality.

The eigenvalues of $A$ and $C_k$ are distinguished by the Greek letters $\lambda$ and $\theta$, respectively. For reasons of simplicity, we refer to the eigenvalues $\theta$ of $C_k$ as Ritz values. These values are *only* Ritz values of $A$ when an underlying projection exists, but they are *always* Ritz values of $C_{k+\ell}$, for any $\ell \in \mathbb{N}$.

The Jordan matrix $J$ of $A$ is the direct sum of Jordan blocks. The direct sum of the $\gamma$ Jordan blocks to an eigenvalue $\lambda$, where $\gamma = \gamma(\lambda)$ denotes the geometric multiplicity of $\lambda$, is called a Jordan box and denoted by $J_\lambda$. The Jordan blocks are denoted by $J_{\lambda\iota}$, where $\iota = 1, 2, \ldots, \gamma$. A single Jordan block $J_{\lambda\iota}$ has dimension $\sigma = \sigma(\lambda, \iota)$ and is *upper* triangular,

$$J_{\lambda\iota} = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} = \lambda I_\sigma + N_\sigma \in \mathbb{C}^{\sigma\times\sigma}. \tag{1.11}$$

Since $C_k$ is unreduced Hessenberg, $J_k$ is the direct sum of Jordan boxes $J_\theta$ that collapse to Jordan blocks. These notations are summarized by

$$J = \bigoplus_\lambda J_\lambda, \quad J_\lambda = \bigoplus_{\iota=1}^{\gamma} J_{\lambda\iota}, \qquad J_k = \bigoplus_\theta J_\theta. \tag{1.12}$$

We split the eigenmatrices according to the splitting of the Jordan matrices into rectangular matrices,

$$V = \bigoplus_\lambda V_\lambda, \quad V_\lambda = \bigoplus_{\iota=1}^{\gamma} V_{\lambda\iota}, \qquad S_k = \bigoplus_\theta S_\theta. \tag{1.13}$$

These matrices are named *partial* eigenmatrices. Similarly, left partial eigenmatrices are defined.

The adjugate of ${}^zC_k$, i.e., the transposed matrix of cofactors of ${}^zC_k$, is sometimes denoted by $P_k(z) \equiv \operatorname{adj}({}^zC_k)$ to emphasize that this matrix is polynomial in $z$. The

MOORE-PENROSE inverse of a (possibly rectangular) matrix $A$ is denoted by $A^\dagger$, the DRAZIN inverse of a square matrix $A$ is denoted by $A^D$. The matrix $P^D \equiv AA^D$ satisfies

$$P^D = \sum_{\lambda \neq 0} V_\lambda \hat{V}_\lambda^H \tag{1.14}$$

and is known as the *eigenprojection*. When $A$ is singular, $I - P^D = V_0 \hat{V}_0^H$. Some of the results are based on representing linear equations using the KRONECKER product of matrices, denoted by $A \otimes B$, and the vec-operator, denoted by $\mathrm{vec}(A)$, cf. [5, Chapter 4, Definition 4.2.1 and Definition 4.2.9]. The EUCLIDian norm is denoted by $\|\cdot\|$.

**2. Basis transformations.** In the fifties KRYLOV subspace methods like the methods of ARNOLDI and LANCZOS were considered as a means to compute the leading part $Q_k$ of a basis transformation $Q$ that brings $A$ to HESSENBERG and tridiagonal form $C = Q^{-1}AQ$, respectively, with $C_k$ as its leading part. Even though this property is frequently lost in finite precision and is not present in the general case, consider the power method applied to an eigenvector, this point of view is helpful in the construction of more elaborate KRYLOV subspace methods, e.g. (Q)OR and (Q)MR methods for the solution of linear systems.

**2.1. Basis vectors.** In this section we give an expression of the "basis" vectors of abstract perturbed KRYLOV methods that reveals the dependency on the starting vector and the perturbation terms. This result is used in succeeding sections to give similar expressions for the other quantities of interest.

For consistency with later sections we define the *basis polynomials* of an abstract perturbed KRYLOV method by

$$\mathcal{B}_k(z) \equiv \frac{\chi_{1:k}(z)}{c_{1:k}} = \check{\nu}_{k+1}(z), \qquad \mathcal{B}_{l+1:k}(z) \equiv \frac{\chi_{l+1:k}(z)}{c_{l+1:k}} = \frac{c_{l+1,l}}{c_{k+1,k}}\, \nu_l(z). \tag{2.1}$$

THEOREM 2.1 (the basis vectors). *The basis vectors of an abstract perturbed* KRYLOV *method* (1.1) *can be expressed in terms of the starting vector $q_1$ and all perturbation terms $\{f_l\}_{l=1}^k$ as follows:*

$$q_{k+1} = \mathcal{B}_k(A) q_1 + \sum_{l=1}^{k} \mathcal{B}_{l+1:k}(A) \frac{f_l}{c_{l+1,l}}. \tag{2.2}$$

*Proof.* We start with the abstract perturbed KRYLOV method (1.1). First we insert ${}^zA \equiv zI_n - A$ and ${}^zC_k \equiv zI_k - C_k$, since $Q_k z I_k = z I_n Q_k$ and thus scalar multiples of $Q_k$ are in the null space, to introduce a dependency on variable $z$,

$$M_k = Q_k(zI - C_k) - (zI - A)Q_k + F_k. \tag{2.3}$$

We use the definition of the adjugate and the LAPLACE expansion of the determinant to obtain

$$M_k \mathrm{adj}({}^zC_k) = Q_k \chi_k(z) - (zI - A) Q_k \mathrm{adj}({}^zC_k) + F_k \mathrm{adj}({}^zC_k). \tag{2.4}$$

We know from [12, Lemma 5.1, equation (5.5)] that the adjugate of unreduced HESSENBERG matrices satisfies

$$\mathrm{adj}({}^zC_k) e_1 = c_{1:k-1} \cdot \nu(z). \tag{2.5}$$

Together with equation (2.4) this gives the simplified equation

$$c_{k+1,k}q_{k+1} = M_k\nu(z) = \frac{q_1\chi_k(z)}{c_{1:k-1}} - (zI - A)Q_k\nu(z) + F_k\nu(z). \tag{2.6}$$

We reorder equation (2.6) slightly to obtain an equation where only scalar polynomials are involved:

$$c_{k+1,k}q_{k+1} = \frac{\chi_k(z)q_1}{c_{1:k-1}} + \sum_{l=1}^{k}\nu_l(z)Aq_l - \sum_{l=1}^{k}\nu_l(z)zq_l + \sum_{l=1}^{k}\nu_l(z)f_l. \tag{2.7}$$

Substituting $A$ for $z$ gives

$$\frac{\chi_k(A)q_1}{c_{1:k-1}} + \sum_{l=1}^{k}\nu_l(A)f_l = c_{k+1,k}q_{k+1}, \tag{2.8}$$

which is upon division by nonzero $c_{k+1,k}$, a corresponding cosmetic division by nonzero $c_{l+1,l}$, $l = 1, \ldots, k$ and by definition of $\nu(z)$ the result to be proved. ☐

**2.2. A closer look.** We need some additional knowledge from the report [12] on Hessenberg eigenvalue-eigenmatrix relations (HEER) in the sequel. We state a relation in the form of the next lemma that is of imminent importance in the proof of the following theorems. We remark that this lemma can not be found 'as is' in the aforementioned report.

Lemma 2.2 (HEER). *Let $C_k$ be unreduced upper* Hessenberg. *Then we can choose the eigenmatrices $S_k$ and $\check{S}_k$ such that the partial eigenmatrices satisfy*

$$e_1^T\check{S}_\theta = e_\alpha^T(\omega_k(J_\theta))^{-T} \quad \text{and} \quad S_\theta^T e_l = c_{1:l-1}\chi_{l+1:k}(J_\theta)^T e_1. \tag{2.9}$$

*The* Hessenberg *eigenvalue-eigenmatrix relations tailored to diagonalizable $C_k$ state that*

$$\check{s}_{lj}s_{\ell j} = \frac{\chi_{1:l-1}(\theta_j)c_{l:\ell-1}\chi_{\ell+1:k}(\theta_j)}{\chi'_{1:k}(\theta_j)} \quad \forall\ l \leqslant \ell. \tag{2.10}$$

*Proof.* The choice mentioned above corresponds to [12, eqn. (5.35)] with $c_{1:k-1}$ brought from the right to the left and is given by

$$S_\theta \equiv c_{1:k-1}\mathcal{V}_{\alpha-1}(\theta), \qquad \check{S}_\theta \equiv \check{\mathcal{V}}_{\alpha-1}(\theta)\omega_k(J_\theta)^{-T}, \tag{2.11}$$

where the unknown quantities are defined in [12, eqns. (5.28), (5.3)] and are given by

$$\mathcal{V}_{\alpha-1}(\theta) \equiv \left(\nu(\theta), \nu'(\theta), \frac{\nu''(\theta)}{2}, \ldots, \frac{\nu^{(\alpha-1)}(\theta)}{(\alpha-1)!}\right), \tag{2.12}$$

$$\check{\mathcal{V}}_{\alpha-1}(\theta) \equiv \left(\frac{\check{\nu}^{(\alpha-1)}(\theta)}{(\alpha-1)!}, \ldots, \frac{\check{\nu}''(\theta)}{2}, \check{\nu}'(\theta), \check{\nu}(\theta)\right). \tag{2.13}$$

By definition of $\nu$ and $\check{\nu}$ it is easy to see that

$$e_1^T\check{S}_\theta = e_1^T\check{\mathcal{V}}_{\alpha-1}(\theta)(\omega_k(J_\theta))^{-T} = e_\alpha^T(\omega_k(J_\theta))^{-T}, \tag{2.14}$$

$$\begin{aligned} e_l^T S_\theta &= c_{1:k-1}e_l^T\left(\nu(\theta), \nu'(\theta), \frac{\nu''(\theta)}{2} \ldots, \frac{\nu^{(\alpha-1)}(\theta)}{(\alpha-1)!}\right) \\ &= \frac{c_{1:k-1}}{c_{l:k-1}}\left(\chi_{l+1:k}(\theta), \chi'_{l+1:k}(\theta), \frac{\chi''_{l+1:k}(\theta)}{2}, \ldots, \frac{\chi^{(\alpha-1)}_{l+1:k}(\theta)}{(\alpha-1)!}\right). \end{aligned} \tag{2.15}$$

The row vector $e_l^T S_\theta / c_{1:l-1}$ consists of TAYLOR expansion terms that can also be written as

$$\left(\chi_{l+1:k}(\theta), \chi'_{l+1:k}(\theta), \frac{\chi''_{l+1:k}(\theta)}{2}, \dots, \frac{\chi^{(\alpha-1)}_{l+1:k}(\theta)}{(\alpha-1)!}\right) = e_1^T \chi_{l+1:k}(J_\theta),$$

i.e., interpreted as the first row of the polynomial $\chi_{l+1:k}$ evaluated at the JORDAN block $J_\theta$. The statement in equation (2.10) is just a rewritten version of [12, Theorem 5.6, eqn. (5.29)]. □

We remark that most of the results proven in this note do *not* depend on the actual choice of the corresponding right and left partial eigenmatrices. This is due to the fact that upper triangular TOEPLITZ matrices commute and that we mostly derive expressions involving inner products of both right and left partial eigenmatrices in the appropriate order.

In most cases we will compute diagonalizable matrices and are interested in eigenvectors solely. In the following, we refer to the coefficients of a vector in the eigenbasis, i.e., in the basis spanned by the columns of $V$, as its eigenparts.

THEOREM 2.3 (the eigenparts of the basis vectors). *Let $\hat{v}^H$ be a left eigenvector of $A$ to eigenvalue $\lambda$ and let $s$ be a right eigenvector of $C_k$ to eigenvalue $\theta$.*

*Then the eigenpart $\hat{v}^H q_{k+1}$ of the basis vector $q_{k+1}$ of an abstract perturbed* KRYLOV *subspace method* (1.1) *can be expressed in terms of the* RITZ *value $\theta$ and the* RITZ *vector $y \equiv Q_k s$ as follows:*

$$\hat{v}^H q_{k+1} = \frac{(\lambda-\theta)\hat{v}^H y}{c_{k+1,k} e_k^T s} + \frac{\hat{v}^H F_k s}{c_{k+1,k} e_k^T s}. \tag{2.16}$$

*Let furthermore $C_k$ be diagonalizable and suppose that $\lambda \neq \theta_j$ for all $j = 1, \dots, k$.*

*Then we can express the dependency of the eigenpart $\hat{v}^H q_{k+1}$ of $q_{k+1}$ on the starting vector $q_1$ and the perturbation terms $\{f_l\}_{l=1}^k$ in three equivalent forms. In terms of the distances of the* RITZ *values $\theta_j$ to $\lambda$ and the left and right eigenvectors of the matrix $C_k$:*

$$\left(\sum_{j=1}^k \frac{c_{k+1,k}\check{s}_{1j} s_{kj}}{\lambda-\theta_j}\right)\hat{v}^H q_{k+1} = \hat{v}^H q_1 + \sum_{l=1}^k \left(\sum_{j=1}^k \frac{c_{l+1,l}\check{s}_{1j} s_{lj}}{\lambda-\theta_j}\right)\frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.17}$$

*In terms of the distances of the* RITZ *values to $\lambda$, the trailing characteristic polynomials of $C_k$ and the derivative of the characteristic polynomial of $C_k$, all evaluated at the* RITZ *values:*

$$\left(\sum_{j=1}^k \frac{c_{1:k}}{\chi'_{1:k}(\theta_j)(\lambda-\theta_j)}\right)\hat{v}^H q_{k+1} = \hat{v}^H q_1 + \sum_{l=1}^k \left(\sum_{j=1}^k \frac{c_{1:l}\chi_{l+1:k}(\theta_j)}{\chi'_{1:k}(\theta_j)(\lambda-\theta_j)}\right)\frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.18}$$

*Without the restriction on $\lambda$, in terms of trailing characteristic polynomials of $C_k$ evaluated at the eigenvalue $\lambda$ of $A$:*

$$\hat{v}^H q_{k+1} = \left(\frac{\chi_{1:k}(\lambda)}{c_{1:k}}\right)\hat{v}^H q_1 + \sum_{l=1}^k \left(\frac{\chi_{l+1:k}(\lambda)}{c_{l+1:k}}\right)\frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.19}$$

*Proof.* We multiply equation (1.1) from the left by $\hat{v}^H$ and from the right by $s$ and obtain

$$\hat{v}^H q_{k+1} c_{k+1,k} e_k^T s = (\lambda - \theta)\hat{v}^H y + \hat{v}^H F_k s. \tag{2.20}$$

The constant $c_{k+1,k}$ is non-zero because $C_k$ is unreduced HESSENBERG. The last component of $s$ is non-zero because $s$ is a right eigenvector of an unreduced HESSENBERG matrix [12, Corollary 5.2]. Equation (2.16) follows upon division by $c_{k+1,k} e_k^T s$.

When $C_k$ is diagonalizable, the columns of the eigenmatrix $S_k$ form a complete set of eigenvectors $s_j$, $j = 1, \ldots, k$. We use this basis to express the standard unit vectors $e_\ell$ by use of the following (trivial) identity:

$$e_\ell = I e_\ell = S S^{-1} e_\ell = S \check{S}^T e_\ell = \sum_{j=1}^k \check{s}_{\ell j} s_j \tag{2.21}$$

Equation (2.20) holds true for all pairs $(\theta_j, s_j)$. We assume that $\lambda \neq \theta_j$ for all RITZ values $\theta_j$ and divide by $\lambda - \theta_j$ to obtain the following set of equations:

$$\left(\frac{c_{k+1,k} s_{kj}}{\lambda - \theta_j}\right) \hat{v}^H q_{k+1} = \hat{v}^H Q_k s_j + \sum_{l=1}^k \left(\frac{c_{l+1,l} s_{lj}}{\lambda - \theta_j}\right) \frac{\hat{v}^H f_l}{c_{l+1,l}}, \quad j = 1, \ldots, k. \tag{2.22}$$

Here we have chosen for cosmetic reasons to divide by $c_{l+1,l}$ in the perturbation terms. All $c_{l+1,l}$ are nonzero since the HESSENBERG matrix $C_k$ is unreduced. We sum up the equations (2.22) using the identity (2.21) for the case $\ell = 1$ to obtain (2.17). We insert the HESSENBERG eigenvalue-eigenmatrix relations tailored to diagonalizable $C_k$ given by Lemma 2.2, equation (2.10) into the first term of (2.17) and obtain

$$\left(\sum_{j=1}^k \frac{c_{1:k}}{\chi'_{1:k}(\theta_j)(\lambda - \theta_j)}\right) \hat{v}^H q_{k+1} = \hat{v}^H q_1 + \sum_{l=1}^k \left(\sum_{j=1}^k \frac{c_{l+1,l} \check{s}_{1j} s_{lj}}{\lambda - \theta_j}\right) \frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.23}$$

When we insert equation (2.10) in repeated manner into the second term of equation (2.17) we obtain (2.18). Now it is time to recognize that an expression like

$$\sum_{j=1}^k \frac{f(\theta_j)}{\chi'_{1:k}(\theta_j)(\lambda - \theta_j)} = \frac{1}{\chi_{1:k}(\lambda)} \sum_{j=1}^k \frac{\prod_{s \neq j}(\lambda - \theta_s)}{\prod_{s \neq j}(\theta_j - \theta_s)} f(\theta_j) \tag{2.24}$$

is just the LAGRANGE form of the interpolation of a function $f$ at nodes $\{\theta_j\}_{j=1}^k$ divided by constant $\chi_{1:k}(\lambda)$. Recognizing that the first term is the interpolation of the constant function $f \equiv 1$ at the RITZ values, we obtain

$$\left(\frac{c_{1:k}}{\chi_{1:k}(\lambda)}\right) \hat{v}^H q_{k+1} = \hat{v}^H q_1 + \sum_{l=1}^k \left(\sum_{j=1}^k \frac{c_{1:l} \chi_{l+1:k}(\theta_j)}{\chi'_{1:k}(\theta_j)(\lambda - \theta_j)}\right) \frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.25}$$

The repeated use of this argument for the second term shows that this is the LAGRANGE interpolation of the polynomials $\chi_{l+1:k}$ of degrees less than $k$,

$$\left(\frac{c_{1:k}}{\chi_{1:k}(\lambda)}\right) \hat{v}^H q_{k+1} = \hat{v}^H q_1 + \sum_{l=1}^k \left(\frac{c_{1:l} \chi_{l+1:k}(\lambda)}{\chi_{1:k}(\lambda)}\right) \frac{\hat{v}^H f_l}{c_{l+1,l}}. \tag{2.26}$$

Division by the first factor results in (2.19). It follows from Theorem 2.1 by multiplication of equation (2.2) from the left by $\hat{v}^H$ that equation (2.19) remains valid without any artificial restrictions on the eigenvalue $\lambda$ and without the need for a diagonalizable HESSENBERG matrix $C_k$. ∎

The validity of Theorem 2.1 for general matrices $A$ and *general* unreduced HESSENBERG matrices $C_k$ suggests that we can also derive expressions for the eigenparts in the general (i.e., not necessarily diagonalizable) case. To proceed we introduce some additional abbreviating notations. We define

$$\operatorname{dist}(J_{\lambda\iota}, J_\theta) \equiv \left(I \otimes J_{\lambda\iota} - J_\theta^T \otimes I\right) \tag{2.27}$$

as shorthand notation for the KRONECKER sum of $J_{\lambda\iota}$ and $-J_\theta$. This may be interpreted as a measure of the "distance" between the two JORDAN blocks. Here and in the following, $J_{\lambda\iota}$ is an arbitrary JORDAN block of $A$ and $J_\theta$ is what we will call a RITZ JORDAN block of $C_k$.

The eigenvector components $\hat{s}_{1j}$ and $s_{lj}$, $l = 1, \ldots, k$, have to be replaced in the general case by block column vectors. The column vectors arise naturally when $C_k$ has principal vectors, the blocks arise naturally when $A$ has principal vectors. Let $\sigma = \sigma(\lambda, \iota)$ be the length of the $\iota$th JORDAN chain to eigenvalue $\lambda$ of $A$. To stress the similarity we will denote the block column vectors in the sequel by

$$\check{S}_{1,\theta}^T \equiv e_1^T \check{S}_\theta \otimes I_\sigma, \qquad S_{l,\theta} \equiv S_\theta^T e_l \otimes I_\sigma, \quad l = 1, \ldots, k. \tag{2.28}$$

The next theorem is the generalization of Theorem 2.3 to the case of not necessarily diagonalizable $C_k$.

THEOREM 2.4. *Let $\hat{V}_{\lambda\iota}^H$ be a left partial eigenmatrix of $A$ to eigenvalue $\lambda$ and let $S_\theta$ be a right partial eigenmatrix of $C_k$ to eigenvalue $\theta$.*

*Then the eigenpart $\hat{V}_\lambda^H q_{k+1}$ of the basis vector $q_{k+1}$ of an abstract perturbed* KRYLOV *subspace method* (1.1) *can be expressed in terms of the* RITZ JORDAN *block $J_\theta$ and the partial* RITZ *matrix $Y_\theta \equiv Q_k S_\theta$ as follows:*

$$\hat{V}_\lambda^H q_{k+1} c_{k+1,k} e_k^T S_\theta = (J_\lambda \hat{V}_\lambda^H Y_\theta - \hat{V}_\lambda^H Y_\theta J_\theta) + \hat{V}_\lambda^H F_k S_\theta. \tag{2.29}$$

*Suppose furthermore that $\lambda \neq \theta_j$ for all $j = 1, \ldots, k$.*

*Then we can express the dependency of the eigenpart $\hat{V}_\lambda^H q_{k+1}$ of $q_{k+1}$ on the starting vector $q_1$ and the perturbation terms $\{f_l\}_{l=1}^k$ in three equivalent forms. In terms of the distances of the* RITZ JORDAN *blocks $J_\theta$ to the* JORDAN *block $J_{\lambda\iota}$ and the left and right partial eigenmatrices of the matrix $C_k$:*

$$\sum_\theta c_{k+1,k} \left(\check{S}_{1,\theta}^T \operatorname{dist}(J_{\lambda\iota}, J_\theta)^{-1} S_{k,\theta}\right) \hat{V}_{\lambda\iota}^H q_{k+1} =$$
$$\hat{V}_{\lambda\iota}^H q_1 + \sum_{l=1}^k \sum_\theta c_{l+1,l} \left(\check{S}_{1,\theta}^T \operatorname{dist}(J_{\lambda\iota}, J_\theta)^{-1} S_{l,\theta}\right) \frac{\hat{V}_{\lambda\iota}^H f_l}{c_{l+1,l}} \tag{2.30}$$

*In terms of the distances of the* RITZ JORDAN *blocks to the* JORDAN *block $J_{\lambda\iota}$ and the trailing characteristic polynomials of $C_k$ and the reduced polynomials $\omega_k$ of $C_k$, all evaluated at the* RITZ JORDAN *blocks:*

$$\sum_\theta c_{1:k} (e_\alpha^T \omega_k(J_\theta)^{-T} \otimes I) \operatorname{dist}(J_{\lambda\iota}, J_\theta)^{-1} (e_1 \otimes I) \hat{V}_{\lambda\iota}^H q_{k+1} = \hat{V}_{\lambda\iota}^H q_1 +$$
$$\sum_{l=1}^k \sum_\theta c_{1:l} (e_\alpha^T \omega_k(J_\theta)^{-T} \otimes I) \operatorname{dist}(J_{\lambda\iota}, J_\theta)^{-1} (\chi_{l+1:k}(J_\theta)^T e_1 \otimes I) \frac{\hat{V}_{\lambda\iota}^H f_l}{c_{l+1,l}} \tag{2.31}$$

*Without the restriction on $\lambda$ in terms of the trailing characteristic polynomials of $C_k$ evaluated at the* JORDAN *block $J_{\lambda_\iota}$ of $A$:*

$$\hat{V}_{\lambda_\iota}^H q_{k+1} = \left(\frac{\chi_{1:k}(J_{\lambda_\iota})}{c_{1:k}}\right)\hat{V}_{\lambda_\iota}^H q_1 + \sum_{l=1}^{k}\left(\frac{\chi_{l+1:k}(J_{\lambda_\iota})}{c_{l+1:k}}\right)\frac{\hat{V}_{\lambda_\iota}^H f_l}{c_{l+1,l}}. \tag{2.32}$$

*Proof.* We multiply equation (1.1) from the left by $\hat{V}_{\lambda_\iota}^H$ and from the right by $S_\theta$ to obtain

$$\hat{V}_{\lambda_\iota}^H q_{k+1} c_{k+1,k} e_k^T S_\theta = J_{\lambda_\iota}\hat{V}_{\lambda_\iota}^H Q_k S_\theta - \hat{V}_{\lambda_\iota}^H Q_k S_\theta J_\theta + \hat{V}_{\lambda_\iota}^H F_k S_\theta. \tag{2.33}$$

This is rewritten to exhibit its linear form utilizing the KRONECKER product and the vec operator. This results in

$$\operatorname{vec}(\hat{V}_{\lambda_\iota}^H q_{k+1} c_{k+1,k} e_k^T S_\theta) = \operatorname{dist}(J_{\lambda_\iota}, J_\theta)\operatorname{vec}(\hat{V}_{\lambda_\iota}^H Q_k S_\theta) + \operatorname{vec}(\hat{V}_{\lambda_\iota}^H F_k S_\theta). \tag{2.34}$$

Suppose that $\lambda$ is not in the spectrum of $C_k$. Inversion of the KRONECKER sum establishes

$$\begin{aligned}&\operatorname{dist}(J_{\lambda_\iota}, J_\theta)^{-1}(c_{k+1,k} S_{k,\theta})\hat{V}_{\lambda_\iota}^H q_{k+1} = \\ &\qquad (S_\theta^T \otimes I_\sigma)\operatorname{vec}(\hat{V}_{\lambda_\iota}^H Q_k) + \operatorname{dist}(J_{\lambda_\iota}, J_\theta)^{-1}(S_\theta^T \otimes I_\sigma)\operatorname{vec}(\hat{V}_{\lambda_\iota}^H F_k).\end{aligned} \tag{2.35}$$

Splitting the summation due to the matrix multiplication in the last term gives

$$\begin{aligned}&\operatorname{dist}(J_{\lambda_\iota}, J_\theta)^{-1}(c_{k+1,k} S_{k,\theta})\hat{V}_{\lambda_\iota}^H q_{k+1} = \\ &\qquad (S_\theta^T \otimes I_\sigma)\operatorname{vec}(\hat{V}_{\lambda_\iota}^H Q_k) + \operatorname{dist}(J_{\lambda_\iota}, J_\theta)^{-1}\left(\sum_{l=1}^{k} S_{l,\theta}\,\hat{V}_{\lambda_\iota}^H f_l\right)\end{aligned} \tag{2.36}$$

At this point we need again the representation of the first standard unit vector $e_1$, the representation of the standard unit vectors $e_\ell$ this time given by

$$e_\ell = I e_\ell = S\check{S}^T e_\ell = \sum_\theta S_\theta \check{S}_\theta^T e_\ell. \tag{2.37}$$

Thus, in KRONECKER product form:

$$\sum_\theta \hat{V}_{\lambda_\iota}^H Q_k S_\theta \check{S}_\theta^T e_1 = \sum_\theta \check{S}_{1,\theta}^T (S_\theta^T \otimes I_\sigma)\operatorname{vec}(\hat{V}_{\lambda_\iota}^H Q_k) = V_{\lambda_\iota}^H q_1. \tag{2.38}$$

A summation over all distinct RITZ values proves (2.30). To proceed, we consider the terms

$$c_{l+1,l}\sum_\theta \left(\check{S}_{1,\theta}^T \operatorname{dist}(J_{\lambda_\iota}, J_\theta)^{-1} S_{l,\theta}\right), \qquad l \in \{1, \dots, k\} \tag{2.39}$$

separately. These correspond to the terms

$$\sum_{j=1}^{k}\frac{c_{l+1,l}\check{s}_{1j}s_{lj}}{\lambda - \theta_j} = \sum_{j=1}^{k}\frac{c_{1:l}\chi_{l+1:k}(\theta_j)}{\chi_{1:k}'(\theta_j)(\lambda - \theta_j)}$$

in the diagonalizable case. We transform the terms by insertion of the by virtue of Lemma 2.2 chosen partial eigenmatrices. Inserting the relations (2.9) results in equation (2.31). We could rewrite equation (2.31) using the generalization of LAGRANGE interpolation to multiple nodes, namely HERMITE interpolation, more precisely SCHWERDTFEGER's formula, to obtain equation (2.32). But (2.32) follows more easily and also for the general case that we allow $\lambda = \theta$ from Theorem 2.1 by multiplication of equation (2.2) from the left by $\hat{V}_{\lambda_\iota}^H$. □

**3. Eigenvalue problems.** In the last section we have shown how the convergence of the Ritz values to an eigenvalue results in the amplification of the error terms. The resulting nonlinearity of the convergence behavior of Ritz values to eigenvalues of $A$ reveals that it is hopeless to ask for results on the convergence of the Ritz values, at least in this abstract setting. One branch of investigation uses the properties of the constructed basis vectors, e.g. (local) orthogonality, to make statements about the convergence of Ritz values. We simply drop the convergence analysis for the Ritz *values* and ask for expressions revealing conditions for a convergence of the Ritz *vectors* and Ritz *residuals*.

Residuals directly give information on the backward errors of the corresponding quantities. The removal of the dependency on the condition that is related to the forward error makes the Ritz residuals slightly more appealing as a point to start the examination.

**3.1. Ritz residuals.** The analysis of the Ritz residuals is simplified since we can easily compute an a posteriori bound involving the next basis vector $q_{k+1}$. Adding the expression for the basis vectors of the last section we can prove the following theorem.

THEOREM 3.1 (the Ritz residuals). *Let $\theta$ be an eigenvalue of $C_k$ with* Jordan *block $J_\theta$ and $S_\theta$ any corresponding partial eigenmatrix. Define the partial* Ritz *matrix by $Y_\theta \equiv Q_k S_\theta$.*

*Then*

$$AY_\theta - Y_\theta J_\theta = \left(\frac{\chi_{1:k}(A)}{c_{1:k-1}}\right) q_1 e_k^T S_\theta + \sum_{l=1}^{k} \left(\frac{\chi_{l+1:k}(A)}{c_{l:k-1}}\right) f_l e_k^T S_\theta - f_l e_l^T S_\theta. \tag{3.1}$$

*Let $S_\theta$ be the (unique) partial eigenmatrix given in Lemma 2.2.*

*Then*

$$AY_\theta - Y_\theta J_\theta = \chi_{1:k}(A) q_1 e_1^T + \sum_{l=1}^{k} c_{1:l-1} \left(\chi_{l+1:k}(A) f_l e_1^T - f_l e_1^T \chi_{l+1:k}(J_\theta)\right). \tag{3.2}$$

*Proof.* We start with the backward expression for the Ritz residual

$$AY_\theta - Y_\theta J_\theta = q_{k+1} c_{k+1,k} e_k^T S_\theta - \sum_{l=1}^{k} f_l e_l^T S_\theta. \tag{3.3}$$

obtained from (1.1) by multiplication by $S_\theta$. A scaled variant of the result in equation (2.2) of Theorem 2.1 is used to replace the next basis vector, namely,

$$q_{k+1} c_{k+1,k} = \left(\frac{\chi_{1:k}(A)}{c_{1:k-1}}\right) q_1 + \sum_{l=1}^{k} \left(\frac{\chi_{l+1:k}(A)}{c_{l:k-1}}\right) f_l. \tag{3.4}$$

Inserting equation (3.4) into equation (3.3) gives equation (3.1). Lemma 2.2 states that with our (special) choice of the biorthogonal partial eigenmatrices

$$e_k^T S_\theta = (c_{1:k-1})\, e_1^T, \qquad e_l^T S_\theta = (c_{1:l-1})\, e_1^T \chi_{l+1:k}(J_\theta). \tag{3.5}$$

Inserting the expressions stated in equation (3.5) into equation (3.1) we obtain equation (3.2). ∎

**3.2. Ritz vectors.** We observe that the adjugate of the family ${}^zC_k \equiv zI - C_k$ is given by the matrix $P_k(z)$ of cofactors, which is polynomial in $z$ and simultaneously a polynomial in $C_k$ for every fixed $z$. This is used in the following lemma to define a bivariate polynomial denoted by $\mathcal{A}_k(z, C_k)$, such that

$$P_k(z) = \mathcal{A}_k(z, C_k) = \operatorname{adj}({}^zC_k). \tag{3.6}$$

It is well-known that whenever we insert a simple eigenvalue $\theta$ of $C_k$ into $P_k(z)$ we obtain a multiple of the spectral projector. More generally, when the eigenvalue has multiplicity $\alpha$ we might consider the evaluation of the matrix $P_k(z)$ along with the derivatives up to order $\alpha - 1$ at $\theta$ to gain information about the eigen- *and* principal vectors of $C_k$, compare with [12, Corollary 3.1].

LEMMA 3.2 (the adjugate polynomial). *Let the bivariate polynomial $\mathcal{A}_k(\theta, z)$ be given by*

$$\mathcal{A}_k(\theta, z) \equiv \begin{cases} (\chi_k(\theta) - \chi_k(z))\,(\theta - z)^{-1}, & z \neq \theta, \\ \chi_k'(z), & z = \theta. \end{cases} \tag{3.7}$$

*Then $\mathcal{A}_k(\theta, C_k)$ is the adjugate of the matrix $\theta I_k - C_k$.*

*Proof.* Inserting the TAYLOR expansion of the polynomial $\chi_k$ at $\theta$ (at $z$) shows that the function given by the right-hand side is a polynomial of degree $k-1$ in $z$ (in $\theta$). For $\theta$ not in the spectrum of $C_k$ we have

$$\begin{aligned} \mathcal{A}_k(\theta, C_k) &= (\chi_k(\theta) I_k - \chi_k(C_k))\,(\theta I_k - C_k)^{-1} && (3.8)\\ &= \det(\theta I_k - C_k)\,(\theta I_k - C_k)^{-1} = \operatorname{adj}(\theta I_k - C_k). && (3.9) \end{aligned}$$

The result for $\theta$ in the spectrum follows by continuity. ◻

REMARK 3.1. *What we call adjugate polynomial appears (without a name) in* GANTMACHER*'s treatise [3, Kapitel 4.3, Seite 111, Formel (28)] published 1954 (in Russian).* TAUSSKY *[9] published a similar result 14 years later.*

We extend the notation to all trailing characteristic polynomials.

DEFINITION 3.3 (trailing adjugate polynomials). *We define the bivariate polynomials $\mathcal{A}_{l+1:k}(\theta, z)$, $l = 1, \ldots, k$, that give the adjugate of a shifted matrix at the* RITZ *values of the trailing submatrices $C_{l+1:k}$ by*

$$\mathcal{A}_{l+1:k}(\theta, z) \equiv \begin{cases} (\chi_{l+1:k}(\theta) - \chi_{l+1:k}(z))\,(\theta - z)^{-1}, & z \neq \theta, \\ \chi_{l+1:k}'(z), & z = \theta. \end{cases} \tag{3.10}$$

In the sequel we need an alternative expression for the adjugate polynomials which clearly reveals their polynomial structure. To proceed we first prove what we call the first adjugate identity (because of its close relation to the first resolvent identity) and specialize it to HESSENBERG structure.

PROPOSITION 3.4. *The adjugates of any matrix family ${}^z\!A$ satisfy the* first adjugate identity *given by*

$$(z - \theta)\operatorname{adj}({}^z\!A)\operatorname{adj}({}^\theta\!A) = \det({}^z\!A)\operatorname{adj}({}^\theta\!A) - \det({}^\theta\!A)\operatorname{adj}({}^z\!A). \tag{3.11}$$

*For unreduced* HESSENBERG *matrices $C_k$ this implies the following important relation*

$$(z - \theta)\sum_{j=1}^{k} \chi_{1:j-1}(z)\chi_{j+1:k}(\theta) = \chi_k(z) - \chi_k(\theta). \tag{3.12}$$

*Proof.* We start with the obvious relation

$$(z-\theta)I = (zI - A) - (\theta I - A) = {}^{z}\!A - {}^{\theta}\!A. \tag{3.13}$$

The multiplication by the adjugates of ${}^{z}\!A$ and ${}^{\theta}\!A$ results in equation (3.11). Now consider the case of an unreduced HESSENBERG matrix $C_k$. By [12, Lemma 5.1, equation (5.4)] we can rewrite the component $(k,1)$ of the equation (3.11) in the lower left corner to obtain

$$(z-\theta)c_{1:k-1}\check{\nu}(z)^T\nu(\theta) = \det({}^{z}C_k)\check{\nu}(\theta)^T e_1 - \det({}^{\theta}C_k)e_k^T\nu(z). \tag{3.14}$$

By definition of $\chi_k$, $\check{\nu}$ and $\nu$ we have proven equation (3.12). ∎

Dividing equation (3.12) by the scalar factor $(z-\theta)$ (and taking limits) proves the following lemma for the adjugate polynomials $\mathcal{A}_k(\theta,z)$. Since trailing submatrices $C_{l+1:k}$ of unreduced HESSENBERG matrices $C_k$ are also unreduced HESSENBERG matrices, the previous arguments also apply to the trailing adjugate polynomials.

LEMMA 3.5. *The adjugate polynomial $\mathcal{A}_k(\theta,z)$ and the trailing adjugate polynomials $\mathcal{A}_{l+1:k}(\theta,z)$ can be expressed in polynomial terms as follows:*

$$\mathcal{A}_{l+1:k}(\theta,z) = \sum_{j=l+1}^{k} \chi_{l+1:j-1}(z)\chi_{j+1:k}(\theta), \quad l = 0,\dots,k. \tag{3.15}$$

*Their $\ell$th derivatives for all $\ell \geqslant 0$ with respect to $\theta$ are given by*

$$\mathcal{A}^{(\ell)}_{l+1:k}(\theta,z) = \sum_{j=l+1}^{k} \chi_{l+1:j-1}(z)\chi^{(\ell)}_{j+1:k}(\theta), \quad l = 0,\dots,k. \tag{3.16}$$

*The relations (3.15) and (3.16) hold also true when $z$ is replaced by a square matrix $A$, in which case we obtain a parameter dependent family of matrices along with their derivatives with respect to the parameter $\theta$.*

We remark that the relations stated in Lemma 3.5 rely strongly on the unreduced HESSENBERG structure of the matrix $C_k$.

THEOREM 3.6 (the RITZ vectors). *Let $\theta$ be an eigenvalue of $C_k$ with JORDAN block $J_\theta$ and let $S_\theta$ be the corresponding unique right eigenmatrix from Lemma 2.2. Let the corresponding partial RITZ matrix be given by $Y_\theta \equiv Q_k S_\theta$. Let $\mathcal{A}_k(\theta,z)$ and $\mathcal{A}_{l+1:k}(\theta,z)$ denote the bivariate adjugate polynomials defined above.*

*Then*

$$\operatorname{vec}(Y_\theta) = \begin{pmatrix} \mathcal{A}_k(\theta,A) \\ \mathcal{A}'_k(\theta,A) \\ \vdots \\ \dfrac{\mathcal{A}^{(\alpha-1)}_k(\theta,A)}{(\alpha-1)!} \end{pmatrix} q_1 + \sum_{l=1}^{k} c_{1:l-1} \begin{pmatrix} \mathcal{A}_{l+1:k}(\theta,A) \\ \mathcal{A}'_{l+1:k}(\theta,A) \\ \vdots \\ \dfrac{\mathcal{A}^{(\alpha-1)}_{l+1:k}(\theta,A)}{(\alpha-1)!} \end{pmatrix} f_l, \tag{3.17}$$

*where the derivation of the bivariate adjugate polynomials is to be understood with respect to the shift $\theta$.*

*Proof.* We know by Theorem 2.1 that the basis vectors $\{q_j\}_{j=1}^k$ are given due to equation (2.2) by

$$q_j = \left(\frac{\chi_{1:j-1}(A)}{c_{1:j-1}}\right) q_1 + \sum_{l=1}^{j-1} \left(\frac{\chi_{l+1:j-1}(A)}{c_{l:j-1}}\right) f_l. \tag{3.18}$$

Using the representation

$$Y_\theta = Q_k S_\theta = \sum_{j=1}^{k} q_j e_j^T S_\theta \tag{3.19}$$

of the partial Ritz matrix and the representation

$$e_j^T S_\theta = c_{1:j-1} e_1^T \chi_{j+1:k}(J_\theta) \tag{3.20}$$

of the by virtue of Lemma 2.2 chosen partial eigenmatrix $S_\theta$ we obtain

$$Y_\theta = \sum_{j=1}^{k} q_j c_{1:j-1} e_1^T \chi_{j+1:k}(J_\theta). \tag{3.21}$$

We insert the expression (3.18) for the basis vectors into equation (3.21) to obtain

$$\begin{aligned} Y_\theta &= \sum_{j=1}^{k} \chi_{1:j-1}(A) q_1 e_1^T \chi_{j+1:k}(J_\theta) \\ &+ \sum_{j=1}^{k} \sum_{l=1}^{j-1} c_{1:l-1} \chi_{l+1:j-1}(A) f_l e_1^T \chi_{j+1:k}(J_\theta) \end{aligned} \tag{3.22}$$

and make use of the alternative expression of the (trailing) shifted adjugate polynomials $\{\mathcal{A}_{l+1:k}\}_{l=0}^{k-1}$ and their derivatives with respect to $\theta$ stated in Lemma 3.5 to obtain equation (3.17). ∎

**3.3. Angles.** In this section we use the last result to express the matrix of angles $\hat{V}_\lambda^H Y_\theta$ between a right partial Ritz matrix $Y_\theta$ and a left partial eigenmatrix $\hat{V}_\lambda$ of $A$.

Theorem 3.7. *Let all notations be given as in Theorem 3.6 and let $Y_\theta \equiv Q_k S_\theta$, where $S_\theta$ is the unique right eigenmatrix from Lemma 2.2.*

*Then the angles between this right partial* Ritz *matrix $Y_\theta$ and any left partial eigenmatrix $\hat{V}_\lambda$ of $A$ are given by*

$$\operatorname{vec}(\hat{V}_\lambda^H Y_\theta) = \begin{pmatrix} \mathcal{A}_k(\theta, J_\lambda) \\ \mathcal{A}'_k(\theta, J_\lambda) \\ \vdots \\ \dfrac{\mathcal{A}_k^{(\alpha-1)}(\theta, J_\lambda)}{(\alpha-1)!} \end{pmatrix} \hat{V}_\lambda^H q_1 + \sum_{l=1}^{k} c_{1:l-1} \begin{pmatrix} \mathcal{A}_{l+1:k}(\theta, J_\lambda) \\ \mathcal{A}'_{l+1:k}(\theta, J_\lambda) \\ \vdots \\ \dfrac{\mathcal{A}_{l+1:k}^{(\alpha-1)}(\theta, J_\lambda)}{(\alpha-1)!} \end{pmatrix} \hat{V}_\lambda^H f_l. \tag{3.23}$$

*Proof.* The result follows by multiplication of the result (3.17) of Theorem 3.6 with any left partial eigenmatrix $\hat{V}_\lambda^H$. ∎

**4. Linear systems: (Q)OR.** The (Q)OR approach is used to approximately solve a linear system $Ax = r_0$ when a square matrix $C_k$ approximating $A$ in some sense is at hand. The (Q)OR approach in context of Krylov subspace methods is based on the choice $q_1 = r_0/\|r_0\|$ and the prolongation $x_k \equiv Q_k z_k$ of the solution $z_k$ of the linear system of equations

$$C_k z_k = \|r_0\| e_1. \tag{4.1}$$

A solution does only exist when $C_k$ is regular, the solution in this case given by $z_k = C_k^{-1}\|r_0\|e_1$. We call $z_k$ the (Q)OR *solution* and $x_k$ the (Q)OR *iterate*. We need another formulation for $z_k$ based on polynomials. We denote the LAGRANGE (the HERMITE) interpolation polynomial that interpolates $z^{-1}$ (and its derivatives) at the RITZ values by $\mathcal{L}_k[z^{-1}](z)$.

LEMMA 4.1 (the LAGRANGE interpolation of the inverse). *The interpolation polynomial $\mathcal{L}_k[z^{-1}](z)$ of the function $z^{-1}$ at the* RITZ *values is defined for nonsingular unreduced* HESSENBERG *$C_k$, can be expressed in terms of the characteristic polynomial $\chi_k(z)$, and is given explicitely by*

$$\begin{aligned}\mathcal{L}_k[z^{-1}](z) &= \begin{cases} \dfrac{\chi_k(0)-\chi_k(z)}{\chi_k(0)} z^{-1}, & z \neq 0, \\ -\dfrac{\chi_k'(0)}{\chi_k(0)}, & z = 0, \end{cases} \\ &= -\frac{\mathcal{A}_k(0,z)}{\chi_k(0)}.\end{aligned} \tag{4.2}$$

*Proof.* It is easy to see that the right-hand side is a polynomial of degree $k-1$, since we explicitely remove the constant term and divide by $z$. Let us denote the right-hand side for the moment by $p_{k-1}$. By CAYLEY-HAMILTON the polynomial evaluated at the nonsingular matrix $C_k$ gives

$$p_{k-1}(C_k) = \frac{\chi_k(0)I - \chi_k(C_k)}{\chi_k(0)} C_k^{-1} = \frac{\chi_k(0)}{\chi_k(0)} C_k^{-1} = C_k^{-1}. \tag{4.3}$$

Thus we have found a polynomial of degree $k-1$ taking the right values at $k$ points (counting multiplicity). The result is proven since the interpolation polynomial is unique. ∎

We extend the definition and notation to trailing submatrices.

DEFINITION 4.2 (trailing interpolations of the inverse). *The* trailing interpolations *of the function $z^{-1}$ are defined for nonsingular $C_{l+1:k}$ to be the interpolation polynomials $\mathcal{L}_{l+1:k}[z^{-1}](z)$ of the function $z^{-1}$ at the* RITZ *values of the trailing submatrices $C_{l+1:k}$ and are given explicitely due to the preceeding Lemma 4.1 by*

$$\begin{aligned}\mathcal{L}_{l+1:k}[z^{-1}](z) &= \begin{cases} \dfrac{\chi_{l+1:k}(0)-\chi_{l+1:k}(z)}{\chi_{l+1:k}(0)} z^{-1}, & z \neq 0, \\ -\dfrac{\chi_{l+1:k}'(0)}{\chi_{l+1:k}(0)}, & z = 0, \end{cases} \\ &= -\frac{\mathcal{A}_{l+1:k}(0,z)}{\chi_{l+1:k}(0)}.\end{aligned} \tag{4.4}$$

We are also confronted with interpolations of the singularly perturbed identity function $1-\delta_{z0}$, where $\delta_{z0}(z)$ is defined by

$$\delta_{z0}(z) \equiv \begin{cases} 1, & z = 0, \\ 0, & z \neq 0. \end{cases} \tag{4.5}$$

DEFINITION 4.3 (interpolations of a perturbed identity). *For nonsingular $C_{l+1:k}$ we define the interpolation polynomials $\mathcal{L}^0_{l+1:k}[1-\delta_{z0}](z)$ that interpolate the identity* *at the* RITZ *values of the trailing submatrices* $C_{l+1:k}$ *and have an additional zero at the node* 0 *by*

$$\mathcal{L}^0_{l+1:k}[1-\delta_{z0}] \equiv \frac{\chi_{l+1:k}(0)-\chi_{l+1:k}(z)}{\chi_{l+1:k}(0)} = \mathcal{L}_{l+1:k}[z^{-1}]z. \tag{4.6}$$

The last equality in equation (4.6) better reveals the characteristics to be expected from such a singular interpolation. We observe that the resulting polynomials are of degree $k-l$ and behave like $z^{k-l}/\det(C_{l+1:k})$ for $z$ outside the field of values of $C_{l+1:k}$ and like $\chi'_{l+1:k}(0)z/\det(C_{l+1:k})$ for $z$ close to zero. These observations help to understand how (Q)OR KRYLOV subspace methods choose RITZ values.

**4.1. Residuals.** It is well known that in unperturbed KRYLOV subspace methods the (Q)OR residual vector $r_k$ is related to the starting residual vector by the so-called *residual polynomial* $\mathcal{R}_k(z)$, $r_k=\mathcal{R}_k(A)r_0$, which is given by

$$\begin{aligned}\mathcal{R}_k(z) &\equiv \det(I_k - zC_k^{-1}) = \frac{\chi_k(z)}{\chi_k(0)}\\ &= 1 - z\mathcal{L}_k[z^{-1}](z) = 1-\mathcal{L}^0_k[1-\delta_{z0}](z)\\ &= \prod_{j=1}^k\left(1-\frac{z}{\theta_j}\right) = \prod_\theta\left(1-\frac{z}{\theta}\right)^{\alpha(\theta)}.\end{aligned} \tag{4.7}$$

This result is a byproduct of the following result that applies to *all* abstract perturbed KRYLOV subspace methods (1.1).

THEOREM 4.4 (the (Q)OR residual vectors). *Suppose an abstract perturbed* KRYLOV *method* (1.1) *is given with* $q_1 = r_0/\|r_0\|$. *Suppose that* $C_k$ *is invertible such that the (Q)OR approach can be applied. Let* $x_k$ *denote the (Q)OR iterate and* $r_k = r_0 - Ax_k$ *the corresponding residual.*

*Then*

$$r_k = \mathcal{R}_k(A)r_0 + \|r_0\|\sum_{l=1}^k c_{1:l-1}\frac{\chi_{l+1:k}(A)-\chi_{l+1:k}(0)}{\chi_{1:k}(0)}f_l. \tag{4.8}$$

*Suppose further that all* $C_{l+1:k}$ *are regular. Define* $\mathcal{R}_{l+1:k}(z)\equiv\chi_{l+1:k}(z)/\chi_{l+1:k}(0)$.

*Then*

$$\begin{aligned} r_k &= \mathcal{R}_k(A)r_0 + \sum_{l=1}^k z_{lk}\,\mathcal{L}^0_{l+1:k}[1-\delta_{z0}](A)\,f_l\\ &= \mathcal{R}_k(A)r_0 - \sum_{l=1}^k z_{lk}\,\mathcal{R}_{l+1:k}(A)\,f_l + F_kz_k.\end{aligned} \tag{4.9}$$

REMARK 4.1. *The last line of equation* (4.9) *is a key result to understand inexact* KRYLOV *subspace methods using the (Q)OR approach. As long as the residual polynomials* $\mathcal{R}_k$ *and* $\mathcal{R}_{l+1:k}$ *are such that the corresponding terms decay to zero, or at least until reaching some threshold below the desired accuracy, the term* $F_kz_k$ *dominates the size of the reachable exact residual.*

*Proof.* Again, our starting point is the KRYLOV decomposition

$$Q_kC_k - AQ_k = -q_{k+1}c_{k+1,k}e_k^T + F_k. \tag{4.10}$$

We compute the residual by applying $z_k/\|r_0\| = C_k^{-1}e_1$ from the right,

$$\frac{r_k}{\|r_0\|} = \frac{-c_{k+1,k}z_{kk}}{\|r_0\|}q_{k+1} + \sum_{l=1}^{k} \frac{z_{lk}}{\|r_0\|} f_l. \tag{4.11}$$

We represent the inverse of $-C_k$ in terms of the adjugate and the determinant,

$$\frac{z_{lk}}{\|r_0\|} = -e_l^T(-C_k)^{-1}e_1 = -\frac{e_l^T \operatorname{adj}(-C_k)e_1}{\det(-C_k)} = -\frac{c_{1:l-1}\chi_{l+1:k}(0)}{\chi_k(0)}. \tag{4.12}$$

Thus,

$$\frac{r_k}{\|r_0\|} = \frac{c_{1:k}}{\chi_{1:k}(0)}q_{k+1} - \sum_{l=1}^{k} \frac{c_{1:l-1}\chi_{l+1:k}(0)}{\chi_{1:k}(0)} f_l. \tag{4.13}$$

When we insert the representation of the next basis vector from equation (2.2) we obtain equation (4.8). When $C_{l+1:k}$ is regular and thus $\chi_{l+1:k}(0) \neq 0$, the first line of equation (4.9) follows with equation (4.12),

$$\begin{aligned} c_{1:l-1}\frac{\chi_{l+1:k}(A) - \chi_{l+1:k}(0)}{\chi_{1:k}(0)} &= \left(-\frac{c_{1:l-1}\chi_{l+1:k}(0)}{\chi_{1:k}(0)}\right) \cdot \left(\frac{\chi_{l+1:k}(0) - \chi_{l+1:k}(A)}{\chi_{l+1:k}(0)}\right) \\ &= \frac{z_{lk}}{\|r_0\|} \cdot \mathcal{L}^0_{l+1:k}[1-\delta_{z0}](A), \end{aligned} \tag{4.14}$$

where we have used the definition of the interpolation of the perturbed identity from equation (4.6). The last line follows, since

$$\mathcal{R}_{l+1:k}(z) = 1 - z\mathcal{L}_{l+1:k}[z^{-1}](z) = 1 - \mathcal{L}^0_{l+1:k}[1-\delta_{z0}](z) \tag{4.15}$$

and thus $\mathcal{L}^0_{l+1:k} = 1 - \mathcal{R}_{l+1:k}$. □

**4.2. Iterates.** In this section we shift our focus to the iterates $x_k = Q_k z_k$. We prove that the iterates are connected to a simple polynomial interpolation problem.

THEOREM 4.5. *Suppose that $C_k$ is regular. Define $z_k = C_k^{-1}e_1\|r_0\|$ and denote the kth (Q)OR iterate by $x_k \equiv Q_k z_k$.*

*Then*

$$x_k = \mathcal{L}_k[z^{-1}](A)r_0 - \|r_0\| \sum_{l=1}^{k} c_{1:l-1}\frac{\mathcal{A}_{l+1:k}(0,A)}{\chi_{1:k}(0)} f_l. \tag{4.16}$$

*Suppose further that all $C_{l+1:k}$ are regular.*

*Then*

$$x_k = \mathcal{L}_k[z^{-1}](A)r_0 - \sum_{l=1}^{k} z_{lk}\, \mathcal{L}_{l+1:k}[z^{-1}](A)\, f_l. \tag{4.17}$$

REMARK 4.2. *This proves that the kth iterate is a linear combination of $k+1$ approximations to the inverse of $A$ obtained from distinct* KRYLOV *subspaces spanned by the same matrix $A$ and different starting vectors, namely $r_0$ and $\{-z_{lk}f_l\}_{l=1}^k$, the latter changing in every step.*

*Proof.* We know that $x_k = Q_k z_k = \sum_{j=1}^k q_j z_{jk}$. Inserting the expression for the basis vectors given in equation (3.18) and the expression for the elements $z_{jk}$ of $z_k$ given in equation (4.12),

$$\begin{aligned}\frac{x_k}{\|r_0\|} &= -\sum_{j=1}^k \frac{\chi_{j-1}(A)\chi_{j+1:k}(0)}{\chi_k(0)}\, q_1 \\ &\quad - \sum_{j=1}^k \sum_{l=1}^{j-1} \frac{c_{1:l-1}\chi_{l+1:j-1}(A)\chi_{j+1:k}(0)}{\chi_k(0)}\, f_l.\end{aligned} \tag{4.18}$$

We switch the order of summation according to $\sum_{j=1}^k \sum_{l=1}^{j-1} = \sum_{l=1}^k \sum_{j=l+1}^k$. By Lemma 3.5,

$$\sum_{j=l+1}^k \chi_{l+1:j-1}(A)\chi_{j+1:k}(0) = \mathcal{A}_{l+1:k}(0, A) \qquad \forall\ l = 0, 1, \ldots, k. \tag{4.19}$$

Thus, equation (4.18) simplifies further to

$$\frac{x_k}{\|r_0\|} = -\frac{\mathcal{A}_k(0, A)}{\chi_k(0)}\, q_1 - \sum_{l=1}^k c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_k(0)}\, f_l. \tag{4.20}$$

Now, equation (4.16) follows from equation (4.2). Similarly to the transformation used in the case of the residuals, when $C_{l+1:k}$ is regular and thus $\chi_{l+1:k}(0) \neq 0$, by equations (4.12) and (4.4),

$$\begin{aligned}c_{1:l-1}\frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_k(0)} &= \left(-\frac{c_{1:l-1}\chi_{l+1:k}(0)}{\chi_k(0)}\right) \cdot \left(-\frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{l+1:k}(0)}\right) \\ &= \frac{z_{lk}}{\|r_0\|} \cdot \mathcal{L}_{l+1:k}[z^{-1}](A),\end{aligned} \tag{4.21}$$

we obtain equation (4.17). ∎

**4.3. Errors.** We derive two theorems that give an explicit expression for the error vectors $x - x_k$. In Theorem 4.6, $A$ is supposed to be regular and we set $x \equiv A^{-1}r_0$. In Theorem 4.7, $A$ is allowed to be singular and the inverse is replaced by a generalized inverse. Due to the polynomial structure we use the Drazin inverse and we set $x \equiv A^D r_0$. For the use of the Drazin inverse in the unperturbed case we refer the reader to [6].

Theorem 4.6 (the (Q)OR error vectors in case of regular $A$). *Suppose an abstract perturbed* Krylov *method* (1.1) *is given with $q_1 = r_0/\|r_0\|$. Suppose that $C_k$ is invertible such that the (Q)OR approach can be applied. Let $x_k$ denote the (Q)OR solution and $r_k = r_0 - Ax_k$ the corresponding residual. Suppose further that $A$ is invertible and let $x = A^{-1}r_0$ denote the unique solution of the linear system $Ax = r_0$. Then*

$$(x - x_k) = \mathcal{R}_k(A)(x - 0) + \|r_0\| \sum_{l=1}^k c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{1:k}(0)} f_l. \tag{4.22}$$

*Suppose further that all trailing submatrices $C_{l+1:k}$, $l = 1, \ldots, k-1$ are nonsingular.*

*Then*

$$
\begin{aligned}
(x - x_k) &= \mathcal{R}_k(A)(x - 0) + \sum_{l=1}^{k} z_{lk}\, \mathcal{L}_{l+1:k}[z^{-1}](A)\, f_l \\
&= \mathcal{R}_k(A)(x - 0) - \sum_{l=1}^{k} z_{lk}\, \mathcal{R}_{l+1:k}[z^{-1}](A)\, A^{-1} f_l + A^{-1} F_k z_k.
\end{aligned}
\tag{4.23}
$$

*Proof.* The results (4.22) and (4.23) follow by subtracting both sides of equation (4.16) and the first line of equation (4.17) from the trivial equation $x = x$ and the observation that $\mathcal{L}_k(z) = (1 - \mathcal{R}_k[z^{-1}](z))z^{-1}$. The last line uses the similar transformations $\mathcal{L}_{l+1:k}(z) = (1 - \mathcal{R}_{l+1:k}[z^{-1}](z))z^{-1}$ to express the error terms in residual form, which results in the additional term $+A^{-1}F_k z_k$. ∎

When $A$ is singular, we can not use the equation $\mathcal{L}_k(z) = (1 - \mathcal{R}_k[z^{-1}](z))z^{-1}$ to simplify the expression for the iterates. Instead, we split the relations into a part where $A$ is regular and another where $A$ is nilpotent. Afterwards, both parts are added to obtain an overall expression.

Theorem 4.7 (the (Q)OR error vectors in case of singular $A$). *Suppose an abstract perturbed* Krylov *method* (1.1) *is given with $q_1 = r_0/\|r_0\|$. Suppose that $C_k$ is invertible such that the (Q)OR approach can be applied. Let $x_k$ denote the (Q)OR solution and $r_k = r_0 - Ax_k$ the corresponding residual. Set $x = A^D r_0$ where $A^D$ denotes the* Drazin *inverse.*

*Then*

$$
\begin{aligned}
(x - x_k) &= \mathcal{R}_k(A)(x - 0) - \mathcal{L}_k[z^{-1}](A) V_0 \hat{V}_0^H r_0 \\
&\qquad + \|r_0\| \sum_{l=1}^{k} c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{1:k}(0)} f_l.
\end{aligned}
\tag{4.24}
$$

*Suppose that all $C_{l+1:k}$ are regular.*

*Then*

$$
\begin{aligned}
(x - x_k) &= \mathcal{R}_k(A)(x - 0) - \mathcal{L}_k[z^{-1}](A) V_0 \hat{V}_0^H r_0 + \sum_{l=1}^{k} z_{lk}\, \mathcal{L}_{l+1:k}[z^{-1}](A)\, f_l \\
&= \mathcal{R}_k(A)(x - 0) - \mathcal{L}_k[z^{-1}](A) V_0 \hat{V}_0^H r_0 \\
&\quad - \sum_{l=1}^{k} z_{lk}\, \mathcal{R}_{l+1:k}[z^{-1}](A)\, A^D f_l \\
&\qquad + \sum_{l=1}^{k} z_{lk}\, \mathcal{L}_{l+1:k}[z^{-1}](A)\, V_0 \hat{V}_0^H f_l + A^D F_k z_k.
\end{aligned}
\tag{4.25}
$$

Remark 4.3. *The term $\mathcal{L}_k[z^{-1}](A) V_0 \hat{V}_0^H r_0$ in equation* (4.24) *is small as long as the starting residual has small components in direction of the partial eigenmatrix to the eigenvalue zero of $A$ and as long as no* Ritz *value comes close to zero. This is the reason why, at least under favorable circumstances,* Krylov *subspace methods can be interpreted as* regularization methods.

*Proof.* We multiply the expression (4.8) for the residuals by the Drazin inverse to obtain

$$A^D r_k = \mathcal{R}_k(A) A^D r_0 + \|r_0\| \sum_{l=1}^{k} c_{1:l-1} \frac{\chi_{l+1:k}(A) - \chi_{l+1:k}(0)}{\chi_{1:k}(0)} A^D f_l \tag{4.26a}$$

$$= \mathcal{R}_k(A)(x - 0) + \|r_0\| \sum_{l=1}^{k} c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{1:k}(0)} A A^D f_l. \tag{4.26b}$$

Now, $A^D r_k = A^D r_0 - A A^D x_k = x - P^D x_k$, $AA^D = P^D$ and thus we have proven

$$x - P^D x_k = \mathcal{R}_k(A)(x - 0) + \|r_0\| \sum_{l=1}^{k} c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{1:k}(0)} P^D f_l. \tag{4.27}$$

We use the expression (4.16) for the iterates times the projection $I - P^D$,

$$(I - P^D) x_k = \mathcal{L}_k[z^{-1}](A)(I - P^D) r_0 - \|r_0\| \sum_{l=1}^{k} c_{1:l-1} \frac{\mathcal{A}_{l+1:k}(0, A)}{\chi_{1:k}(0)} (I - P^D) f_l \tag{4.28}$$

Equation (4.24) follows by subtracting equation (4.28) from equation (4.27) and rewriting $(I - P^D) r_0 = V_0 \hat{V}_0^H r_0$. The first line of equation (4.25) follows from equation (4.21). The remaining part of equation (4.25) follows from a similar splitting, we reformulate equation (4.26a) using the identity

$$\begin{aligned} \|r_0\| c_{1:l-1} \frac{\chi_{l+1:k}(A) - \chi_{l+1:k}(0)}{\chi_{1:k}(0)} &= -z_{lk} \frac{\chi_{l+1:k}(A) - \chi_{l+1:k}(0)}{\chi_{l+1:k}(0)} \\ &= -z_{lk} \left( \mathcal{R}_{l+1:k}(A) - I \right) \end{aligned} \tag{4.29}$$

and equation (4.28) using equation (4.21). Subtracting the reformulated equations and rewriting the terms $(I - P^D) f_l = V_0 \hat{V}_0^H f_l$ gives the remaining part of equation (4.25). □

The terms preventing a convergence of the (Q)OR iterates to the vector $x = A^D r_0$ can only become large when some Ritz values come close to zero. In that case the vector $z_k$ should be modified to have no components in the directions to very small Ritz values to regularize the iterates. For further investigation we remark that only the *smallest* $\sigma_{\max} \equiv \max_\iota \{\sigma(0, \iota)\}$ monomials of the interpolation polynomials of the inverse are active, since the Jordan box $J_0$ of $A$ to the eigenvalue zero is nilpotent with ascent $\sigma_{\max}$ and for all $\ell \geqslant \alpha_{\max}$, $J_0^\ell = O$.

**5. Linear systems: (Q)MR.** The (Q)MR approach is used to approximately solve a linear system $Ax = \underline{r}_0$ when a *rectangular* approximation to $A$ is at hand. To better distinguish the (Q)MR approach from the (Q)OR approach we denote the starting residual by $\underline{r}_0$ instead of $r_0$. Mostly, $\underline{r}_0 \equiv r_0$. The (Q)MR approach in context of Krylov subspace methods is based on the choice $q_1 = \underline{r}_0 / \|\underline{r}_0\|$ and the prolongation $\underline{x}_k \equiv Q_k \underline{z}_k$ of the solution $\underline{z}_k$ of the least-squares problem

$$\|\underline{C}_k \underline{z}_k - \|\underline{r}_0\| \underline{e}_1\| = \min. \tag{5.1}$$

We call $\underline{z}_k$ the (Q)MR *solution* and $\underline{x}_k$ the (Q)MR *iterate*. Since $\underline{C}_k$ is extended unreduced Hessenberg, obviously $\underline{C}_k$ has full rank $k$. By definition of $\check{\underline{\nu}}$ and [12, Lemma 5.1, equation (5.5)], compare with [2, equations (3.4) and (3.8)], the extended nonzero vector $\underline{\check{\nu}}(0)^T$ spans the left null space of $\underline{C}_k$,

$$\underline{\check{\nu}}(0)^T \underline{C}_k = \underline{\hat{\nu}}(0)^H \underline{C}_k = o_k^T. \tag{5.2}$$

The unique solution $\underline{z}_k$ of the least-squares problem (5.1) and its residual $\mathfrak{r}_k$ are given by

$$\underline{z}_k \equiv \|\underline{r}_0\| \underline{C}_k^\dagger \underline{e}_1, \qquad \mathfrak{r}_k \equiv \underline{e}_1 \|\underline{r}_0\| - \underline{C}_k \underline{z}_k = \|\underline{r}_0\| (I_{k+1} - \underline{C}_k \underline{C}_k^\dagger) \underline{e}_1. \tag{5.3}$$

We call the residual $\mathfrak{r}_k$ of the HESSENBERG least-squares problem (5.1) the *quasi-residual* of the linear system $Ax = \underline{r}_0$. The residual $\underline{r}_k$ of the (Q)MR approximation $\underline{x}_k \equiv Q_k \underline{z}_k$ is related to the quasi-residual as follows,

$$\begin{aligned} \underline{r}_k &= \underline{r}_0 - A\underline{x}_k = Q_{k+1} \underline{e}_1 \|\underline{r}_0\| - AQ_k \underline{z}_k \\ &= Q_{k+1}(\underline{e}_1 \|\underline{r}_0\| - \underline{C}_k \underline{z}_k) + F_k \underline{z}_k = Q_{k+1} \mathfrak{r}_k + \sum_{l=1}^{k} f_l \underline{z}_{lk}. \end{aligned} \tag{5.4}$$

To proceed, we construct another expression for the least-squares solution $\underline{z}_k$ and the quasi-residual $\mathfrak{r}_k$.

LEMMA 5.1 (the vectors $\underline{z}_k$ and $\mathfrak{r}_k$). *Let $\underline{C}_k \in \mathbb{C}^{(k+1)\times k}$ be an unreduced extended upper* HESSENBERG *matrix. Let $C_{k+1}^{\triangle} \in \mathbb{C}^{k\times k}$ denote the regular upper triangular matrix $\underline{C}_k$ without its first row, and let $M_{k+1}^{\triangle}(0)$ denote the inverse of $-C_{k+1}^{\triangle}$, compare with [12, Lemma 5.4 and Theorem 5.5].*

*Then*

$$(M_{k+1}^{\triangle}(0))_{lj} = \begin{cases} -\dfrac{\chi_{l+1:j}(0)}{c_{l:j}}, & l \leqslant j, \\ \quad 0, & l > j, \end{cases} \tag{5.5}$$

$$\frac{\underline{z}_k}{\|\underline{r}_0\|} = \begin{pmatrix} o_k & M_{k+1}^{\triangle}(0) \end{pmatrix} \frac{\underline{\hat{\nu}}(0)}{\|\underline{\check{\nu}}(0)\|^2} \qquad \textit{and} \qquad \frac{\mathfrak{r}_k}{\|\underline{r}_0\|} = \frac{\underline{\hat{\nu}}(0)}{\|\underline{\check{\nu}}(0)\|^2}. \tag{5.6}$$

*Proof.* The expression for the matrix $M_{k+1}^{\triangle}$ is given in [12, Lemma 5.4] and is included here merely for the sake of completeness. We first prove the expression for the quasi-residual. The nonzero vector $\underline{\hat{\nu}}(0)^H$ spans the left null space and $\underline{C}_k$ has full rank $k$. Thus, the matrix $\underline{C}_k \underline{C}_k^\dagger$ is given by

$$\underline{C}_k \underline{C}_k^\dagger = I_{k+1} - \frac{\underline{\hat{\nu}}(0)\underline{\hat{\nu}}(0)^H}{\|\underline{\check{\nu}}(0)\|^2}. \tag{5.7}$$

By definition of $\mathfrak{r}_k$, equation (5.3), and since $\hat{\nu}_1(0) = 1$, the quasi-residual is given by the expression in equation (5.6). The relation $\mathfrak{r}_k = \underline{e}_1 \|\underline{r}_0\| - \underline{C}_k \underline{z}_k$ can be embedded into

$$\begin{aligned} \begin{pmatrix} \underline{e}_1 & -\underline{C}_k \end{pmatrix} \begin{pmatrix} 0 \\ \underline{z}_k \end{pmatrix} &= \mathfrak{r}_k - \underline{e}_1 \|\underline{r}_0\| \\ \Leftrightarrow \quad \begin{pmatrix} 0 \\ \underline{z}_k \end{pmatrix} &= \begin{pmatrix} 1 & -c_{1,1:k} M_{k+1}^{\triangle}(0) \\ o_k & M_{k+1}^{\triangle}(0) \end{pmatrix} (\mathfrak{r}_k - \underline{e}_1 \|\underline{r}_0\|). \end{aligned} \tag{5.8}$$

We remove the first column and use the fact that $\underline{e}_1$ is in the null space of the matrix with $M_{k+1}^{\triangle}(0)$ in its lower block to obtain the expression for $\underline{z}_k$ in equation (5.6). □

When all leading submatrices $C_j$ are regular, which is the case, e.g. in the unperturbed CG method applied to a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$, we can rewrite part of the results of Lemma 5.1 in terms of the (Q)OR quantities as follows.

LEMMA 5.2 (the relation between $\underline{z}_k$ and $z_j$, $j \leqslant k$). *Suppose that all leading $C_j$, $j = 1, \dots, k$ are regular and that $r_0 \equiv \underline{r}_0$.*

*Then*

$$\underline{z}_k = \frac{\displaystyle\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2 \begin{pmatrix} z_j \\ o_{k-j} \end{pmatrix}}{\|\underline{\check{\nu}}(0)\|^2}, \qquad \underline{x}_k = \frac{\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2\, x_j}{\|\underline{\check{\nu}}(0)\|^2}, \tag{5.9}$$

*where for convenience we interpret $z_0$ as the empty matrix with dimensions $0 \times 1$.*

REMARK 5.1. *The preceeding lemma states that the (Q)MR solution (iterate) is a convex combination of all prior (Q)OR solutions (iterates). The representation of (Q)OR solutions and (Q)OR iterates with interpolation polynomials suggests a representation of the polynomials associated with the (Q)MR approach as convex combinations of all corresponding prior (Q)OR polynomials. Because of the close relations, the same holds true for the associated residual and perturbed identity polynomials.*

*Proof.* The $j$th (Q)OR solution $z_j$ can be embedded into

$$\begin{aligned} \begin{pmatrix} \underline{e}_1 & -\underline{C}_k \end{pmatrix} \begin{pmatrix} 0 \\ z_j \\ o_{k-j} \end{pmatrix} &= -c_{j+1,j} z_{jj} e_{j+1} - \underline{e}_1 \|\underline{r}_0\| \\ &= \frac{c_{1:j}}{\chi_{1:j}(0)} e_{j+1} - \underline{e}_1 \|\underline{r}_0\| \\ &= \frac{1}{\check{\nu}_{j+1}(0)} e_{j+1} - \underline{e}_1 \|\underline{r}_0\|. \end{aligned} \tag{5.10}$$

Multiplication by $|\check{\nu}_{j+1}(0)|^2$, summation over $j$ and division by $\|\underline{\check{\nu}}(0)\|^2$ results in

$$\begin{pmatrix} \underline{e}_1 & -\underline{C}_k \end{pmatrix} \frac{\displaystyle\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2 \begin{pmatrix} 0 \\ z_j \\ o_{k-j} \end{pmatrix}}{\|\underline{\check{\nu}}(0)\|^2} = \frac{\underline{\hat{\nu}}(0)}{\|\underline{\check{\nu}}(0)\|^2} - \underline{e}_1 \|\underline{r}_0\|. \tag{5.11}$$

The matrix $\begin{pmatrix} \underline{e}_1 & -\underline{C}_k \end{pmatrix}$ is regular, which proves the first part of equation (5.9) by comparison of equation (5.11) with equation (5.8). The second part of equation (5.9) follows upon multiplication by $Q_k$ from the left. □

This indicates how the residual and other polynomials associated with the (Q)MR approach might be constructed. The expressions given in the following apply also to cases where the submatrices $C_j$ are not necessarily regular.

DEFINITION 5.3 (the polynomials $\underline{\mathcal{R}}_k$, $\underline{\mathcal{L}}_k$ and $\underline{\mathcal{L}}_k^0$). *We define the polynomials*

$\underline{\mathcal{R}}_k(z)$, $\underline{\mathcal{L}}_k[z^{-1}](z)$ *and* $\underline{\mathcal{L}}_k^0[1-\delta_{z0}](z)$ *by*

$$\underline{\mathcal{R}}_k(z) \equiv \frac{\sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(z)}{\|\underline{\check{\nu}}(0)\|^2}, \tag{5.12}$$

$$\underline{\mathcal{L}}_k[z^{-1}](z) \equiv \frac{\sum_{j=1}^{k+1} \hat{\nu}_j(0)(c_{1:j-1})^{-1}(-\mathcal{A}_{j-1}(0,z))}{\|\underline{\check{\nu}}(0)\|^2}, \tag{5.13}$$

$$\underline{\mathcal{L}}_k^0[1-\delta_{z0}](z) \equiv \frac{\sum_{j=1}^{k+1} \hat{\nu}_j(0)(\check{\nu}_j(0)-\check{\nu}_j(z))}{\|\underline{\check{\nu}}(0)\|^2}. \tag{5.14}$$

With these definitions,

$$\begin{aligned} \underline{\mathcal{R}}_k(z) &= 1-\underline{\mathcal{L}}_k^0[1-\delta_{z0}](z) = 1-z\underline{\mathcal{L}}_k[z^{-1}](z), \\ \deg(\underline{\mathcal{R}}_k(z)) &= \deg(\underline{\mathcal{L}}_k^0[1-\delta_{z0}](z)) = \deg(\underline{\mathcal{L}}_k[z^{-1}](z))+1. \end{aligned} \tag{5.15}$$

When all $C_j$ are regular, by inspection

$$\underline{\mathcal{R}}_k(z) = \frac{\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2\,\mathcal{R}_j(z)}{\|\underline{\check{\nu}}(0)\|^2}, \tag{5.16}$$

$$\underline{\mathcal{L}}_k[z^{-1}](z) = \frac{\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2\,\mathcal{L}_j[z^{-1}](z)}{\|\underline{\check{\nu}}(0)\|^2}, \tag{5.17}$$

$$\underline{\mathcal{L}}_k^0[1-\delta_{z0}](z) = \frac{\sum_{j=0}^{k} |\check{\nu}_{j+1}(0)|^2\,\mathcal{L}_j^0[1-\delta_{z0}](z)}{\|\underline{\check{\nu}}(0)\|^2}. \tag{5.18}$$

To better understand the *interpolation* properties of the polynomials defined in Definition 5.3 we need another expression for the residual polynomial that gives more insight. To obtain this expression, we need the following auxiliary simple lemma.

LEMMA 5.4. *Suppose that $A \in \mathbb{C}^{n\times n}$ is given. Let $A_{n-1}$ denote its leading principal submatrix consisting of the columns and rows indexed from 1 to $n-1$.*

*Then*

$$\det(A + ze_ne_n^T) = \det(A) + z\det(A_{n-1}). \tag{5.19}$$

*Proof.* Equation (5.19) follows from the multilinearity of the determinant. □

We define the characteristic matrix of $\underline{C}_k$ by ${}^z\underline{C}_k \equiv z\underline{I}_k - \underline{C}_k$. In the sequel we need some knowledge about matrices of the form $-\underline{C}_k^H\,{}^z\underline{C}_k$ expressed in terms of rank-one modified square matrices. It is easy to see that

$$-\underline{C}_k^H\,{}^z\underline{C}_k = \begin{pmatrix} -C_k \\ -c_{k+1,k}e_k^T \end{pmatrix}^H \begin{pmatrix} {}^zC_k \\ -c_{k+1,k}e_k^T \end{pmatrix} = -C_k^H\,{}^zC_k + |c_{k+1,k}|^2 e_ke_k^T. \tag{5.20}$$

We need the leading principal submatrix of size $k-1\times k-1$ of $C_k^H\,{}^zC_k$, denoted by $(C_k^H\,{}^zC_k)_{k-1}$. Obviously,

$$(C_k^H\,{}^zC_k)_{k-1} = \underline{C}_{k-1}^H\,{}^z\underline{C}_{k-1}. \tag{5.21}$$

Now we can characterize the so-called *quasi-kernel polynomials* $\underline{\mathcal{R}}_k(z)$ further in the following lemma and corollary, compare with [2, Theorem 3.2, Corollary 5.3].

LEMMA 5.5 (the quasi-kernel polynomials). *Let ${}^z\underline{C}_k \equiv z\underline{I}_k - \underline{C}_k$ denote the characteristic matrix of $\underline{C}_k$ and define ${}^0\underline{C}_k \equiv -\underline{C}_k$. Let letter $\mu$ denote the eigenvalues* *of the generalized eigenvalue problem*

$$C_k^H u = \mu \underline{C}_k^H \underline{C}_k u \tag{5.22}$$

*and let letter* $\vartheta = 1/\mu$ *denote the eigenvalues of the generalized eigenvalue problem*

$$\underline{C}_k^H \underline{C}_k u = \vartheta C_k^H u, \tag{5.23}$$

*where the algebraic multiplicity of* $\mu$ *and* $\vartheta$ *is denoted by* $\alpha(\mu)$ *and* $\alpha(\vartheta)$*, respectively.*

*Then*

$$\begin{aligned}\underline{\mathcal{R}}_k(z) &\equiv \frac{\sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(z)}{\sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(0)} = \frac{\underline{\check{\nu}}(0)^H \underline{\check{\nu}}(z)}{\underline{\check{\nu}}(0)^H \underline{\check{\nu}}(0)} = \frac{\underline{\check{\nu}}(0)^H \underline{\check{\nu}}(z)}{\|\underline{\check{\nu}}(0)\|^2} = \frac{\det(\underline{C}_k^H {}^z\underline{C}_k)}{\det(\underline{C}_k^H {}^0\underline{C}_k)} \\ &= \frac{\det(\underline{C}_k^H \underline{C}_k - z C_k^H)}{\det(\underline{C}_k^H \underline{C}_k)} = \prod_\mu (1 - z\mu)^{\alpha(\mu)} = \prod_\vartheta \left(1 - \frac{z}{\vartheta}\right)^{\alpha(\vartheta)}.\end{aligned} \tag{5.24}$$

REMARK 5.2. *It is an interesting observation that the matrix* $(\underline{C}_k^H \underline{C}_k)^{-1} C_k^H$ *occurs naturally as the "best" lower-dimensional approximation to the pseudoinverse in the sense that*

$$\underline{C}_k^\dagger \underline{I}_k = (\underline{C}_k^H \underline{C}_k)^{-1} \underline{C}_k^H \underline{I}_k = (\underline{C}_k^H \underline{C}_k)^{-1} C_k^H. \tag{5.25}$$

*Especially,*

$$\underline{\mathcal{R}}_k(z) = \frac{\det(\underline{C}_k^H \underline{C}_k - z C_k^H)}{\det(\underline{C}_k^H \underline{C}_k)} = \det(I_k - z\underline{C}_k^\dagger \underline{I}_k) \tag{5.26}$$

*and*

$$\underline{z}_k = \underline{C}_k^\dagger \underline{e}_1 = (\underline{C}_k^H \underline{C}_k)^{-1} C_k^H e_1. \tag{5.27}$$

*Thus, the eigenvalues* $\mu$ *of* $(\underline{C}_k^H \underline{C}_k)^{-1} C_k^H$ *are related in this sense to the behavior of the matrix* $\underline{C}_k^\dagger$*. The singular values of* $(\underline{C}_k^H \underline{C}_k)^{-1} C_k^H$ *interlace those of* $\underline{C}_k^\dagger$*, the closeness of the two sets of singular values is related to the magnitude of* $c_{k+1,k}$*.*

*Proof.* We only need to prove the equation

$$\frac{\sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(z)}{\sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(0)} = \frac{\det(\underline{C}_k^H {}^z\underline{C}_k)}{\det(\underline{C}_k^H {}^0\underline{C}_k)} = \frac{\det(-\underline{C}_k^H {}^z\underline{C}_k)}{\det(-\underline{C}_k^H {}^0\underline{C}_k)}, \tag{5.28}$$

since the remaining parts of equation (5.24) consist of trivial rewritings. By iterated application of Lemma 5.4 and equations (5.20) and (5.21),

$$\begin{aligned}\det(-\underline{C}_k^H {}^z\underline{C}_k) &= \det(-C_k^H {}^z C_k) + |c_{k+1,k}|^2 \det(-\underline{C}_{k-1}^H {}^z\underline{C}_{k-1}) \\ &= \sum_{j=0}^{k} |c_{j+1:k}|^2 \overline{\chi_j(0)} \chi_j(z) = |c_{1:k}|^2 \sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(z).\end{aligned} \tag{5.29}$$

Similarly,

$$\det(-\underline{C}_k^H {}^0\underline{C}_k) = |c_{1:k}|^2 \sum_{j=1}^{k+1} \hat{\nu}_j(0)\check{\nu}_j(0). \tag{5.30}$$

Thus we have proven equation (5.28). □

In the following, we focus on the similarities and differences of the residual polynomials $\mathcal{R}_k$ and $\underline{\mathcal{R}}_k$. To reveal the similarity, we state once again the following two akin expressions for the residual polynomials,

$$\underline{\mathcal{R}}_k(z) = \det(I_k - z\underline{C}_k^{\dagger}\underline{I}_k) \quad \text{and} \quad \mathcal{R}_k(z) = \det(I_k - zC_k^{-1}). \tag{5.31}$$

The residual polynomials $\mathcal{R}_k$ cease to exist when $C_k$ is singular, the residual polynomials $\underline{\mathcal{R}}_k$ do always exist since by assumption $\underline{C}_k$ always has full rank $k$. When the polynomials $\mathcal{R}_k$ exist they are of exact degree $k$. The degree of $\underline{\mathcal{R}}_k$ might be $k-r$ for any $r$ in $\{0,\dots,k\}$. We show that the non-existence of certain residual polynomials $\mathcal{R}_j$ is related to a drop in the degree of $\underline{\mathcal{R}}_k$. The precise conditions for a drop of the degree by $r$ are summarized in the next corollary.

COROLLARY 5.6 (the degree of $\underline{\mathcal{R}}_k(z)$). *The polynomial $\underline{\mathcal{R}}_k(z)$ has degree $k-r$ precisely when the matrices $C_k,\dots,C_{k-r+1}$ are singular and $C_{k-r}$ is regular, the generalized eigenvalue problem (5.22) has zero as eigenvalue with multiplicity $r = \alpha(0)$ and the generalized eigenvalue problem (5.23) has infinity as eigenvalue with multiplicity $r = \alpha(\infty)$.*

*Proof.* By definition the polynomials $\check{\nu}_{j+1}(z)$ have exact degree $j$. A drop in the degree of the polynomial $\underline{\mathcal{R}}_k(z)$ by $r$ can occur only when the $r$ trailing factors $\hat{\nu}_{k+2-r}(0),\dots,\hat{\nu}_{k+1}(0)$ are zero and the factor $\hat{\nu}_{k+1-r}(0)$ is nonzero. Thus, by definition of $\hat{\nu}_j(0)$, the matrices $C_k,\dots,C_{k-r+1}$ are singular and the matrix $C_{k-r}$ is regular. Since $\underline{C}_k^H\underline{C}_k$ is regular, exactly $k$ finite eigenvalues $\mu$ counting multiplicity of the generalized eigenvalue problem (5.22) exist. The zero eigenvalues obviously cause the drop in the degree, thus $r = \alpha(0)$. The infinite eigenvalues of the generalized eigenvalue problem (5.23) are the inverses of the zero eigenvalues of the generalized eigenvalue problem (5.22). □

The inverses $\vartheta \equiv 1/\mu$ of the eigenvalues $\mu$ (counting multiplicity) are the so-called *harmonic* RITZ *values*. The harmonic RITZ values are the eigenvalues of the generalized eigenvalue problem $\vartheta C_k^H u = \underline{C}_k^H\underline{C}_k u$. There are precisely $k-\alpha(\infty)$ finite eigenvalues, the infinite eigenvalues case the rank-drop. The name follows from their interpretation as harmonic mean to eigenvalues of $A$. In the general case these values are *not* harmonic RITZ values of $A$, but of all $C_{k+\ell}$, $\ell \in \mathbb{N}$.

**5.1. Residuals.** It is known that in unperturbed (Q)MR KRYLOV subspace methods the residual vector $\underline{r}_k$ is related to the starting residual vector by $\underline{r}_k = \underline{\mathcal{R}}_k(A)\underline{r}_0$. This result for the unperturbed methods is a byproduct of the following result that applies to *all* abstract perturbed KRYLOV subspace methods (1.1). We use the expression for the residuals $\underline{r}_k$, the representation of the basis vectors $\{q_j\}_{j=1}^{k+1}$ and the representation of the quasi-residual $\mathfrak{r}_k$ and the vector $\underline{z}_k$ to prove the following theorem.

THEOREM 5.7 (the (Q)MR residual vectors). *Suppose an abstract* KRYLOV *method (1.1) is given with $q_1 = \underline{r}_0/\|\underline{r}_0\|$. Let $\underline{x}_k = Q_k\underline{z}_k$ denote the (Q)MR iterate and $\underline{r}_k = \underline{r}_0 - A\underline{x}_k$ the corresponding residual.*

*Then*

$$\underline{r}_k = \underline{\mathcal{R}}_k(A)\underline{r}_0 + \frac{\|\underline{r}_0\|}{\|\check{\underline{\nu}}(0)\|^2}\sum_{l=1}^{k}\left(\sum_{j=l}^{k}\hat{\nu}_{j+1}(0)\frac{\chi_{l+1:j}(A)-\chi_{l+1:j}(0)}{c_{l:j}}\right)f_l. \tag{5.32}$$

*Proof.* We use the expression for the residual given in equation (5.4),

$$\underline{r}_k = Q_{k+1}\mathfrak{r}_k + \sum_{l=1}^{k} f_l \underline{z}_{lk} = \sum_{j=1}^{k+1} q_j \mathfrak{r}_{jk} + \sum_{l=1}^{k} f_l \underline{z}_{lk}. \tag{5.33}$$

and insert the expression (5.6) for the quasi-residual $\mathfrak{r}_k$ and the solution $z_k$ of the least-squares problem obtained in Lemma 5.1. We already have noted that by Theorem 2.1 the basis vectors $\{q_j\}_{j=1}^k$ are given by equation (3.18),

$$q_j = \left(\frac{\chi_{1:j-1}(A)}{c_{1:j-1}}\right) q_1 + \sum_{l=1}^{j-1} \left(\frac{\chi_{l+1:j-1}(A)}{c_{l:j-1}}\right) f_l. \tag{5.34}$$

Putting pieces together results in

$$\begin{aligned}
\underline{r}_k &= \underline{\mathcal{R}}_k(A)\underline{r}_0 + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2} \sum_{j=1}^{k+1}\sum_{l=1}^{j-1} \hat{\nu}_j(0) \frac{\chi_{l+1:j-1}(A)}{c_{l:j-1}} f_l \\
&\quad + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2} F_k \begin{pmatrix} 0 & M_{k+1}^{\triangle}(0) \end{pmatrix} \underline{\hat{\nu}}(0) \\
&= \underline{\mathcal{R}}_k(A)\underline{r}_0 + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2} \sum_{j=0}^{k}\sum_{l=1}^{j} \hat{\nu}_{j+1}(0) \frac{\chi_{l+1:j}(A) - \chi_{l+1:j}(0)}{c_{l:j}} f_l.
\end{aligned} \tag{5.35}$$

Switching the order of summation according to $\sum_{j=0}^k \sum_{l=1}^j = \sum_{l=1}^k \sum_{j=l}^k$ results in equation (5.32). ∎

**5.2. Iterates.** By a direct multiplication of the by virtue of Theorem 2.1 known basis vectors $\{q_j\}_{j=1}^k$ and the explicit expression for the (Q)MR solutions $\underline{z}_k$ we can prove the following theorem.

THEOREM 5.8 (the (Q)MR iterates). *Suppose an abstract* KRYLOV *method* (1.1) *is given with $q_1 = \underline{r}_0/\|\underline{r}_0\|$. Let $\underline{x}_k = Q_k\underline{z}_k$ denote the kth (Q)MR iterate.*

*Then*

$$\underline{x}_k = \underline{\mathcal{L}}_k(A)\underline{r}_0 + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2} \sum_{l=1}^{k} \left( \sum_{j=l}^{k} \hat{\nu}_{j+1}(0) \frac{-\mathcal{A}_{l+1:j}(0,A)}{c_{l:j}} \right) f_l. \tag{5.36}$$

*Proof.* We use equation (5.34) to obtain

$$\begin{aligned}
\frac{\underline{x}_k}{\|r_0\|} &= Q_k \frac{\underline{z}_k}{\|r_0\|} = \sum_{\ell=1}^{k} q_\ell \frac{\underline{z}_{\ell k}}{\|r_0\|} \\
&= \sum_{\ell=1}^{k} \left( \left(\frac{\chi_{1:\ell-1}(A)}{c_{1:\ell-1}}\right) q_1 + \sum_{l=1}^{\ell-1} \left(\frac{\chi_{l+1:\ell-1}(A)}{c_{l:\ell-1}}\right) f_l \right) \frac{\underline{z}_{\ell k}}{\|r_0\|}
\end{aligned} \tag{5.37}$$

By application of Lemma 5.1, equations (5.5) and (5.6), we can write the elements $\underline{z}_{\ell k}$ of the vector $\underline{z}_k$ in the form

$$\frac{\underline{z}_{\ell k}}{\|\underline{r}_0\|} = \sum_{j=\ell}^{k} \frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}} \frac{\hat{\nu}_{j+1}(0)}{\|\underline{\check{\nu}}(0)\|^2}. \tag{5.38}$$

Thus, by switching the order of summation and the alternate description of the adjugate polynomials given in Lemma 3.5, equation (3.15),

$$\begin{aligned}
\frac{\|\underline{\check{\nu}}(0)\|^2}{\|\underline{r}_0\|}\underline{x}_k &= \sum_{\ell=1}^{k}\sum_{j=\ell}^{k}\left(\frac{\chi_{1:\ell-1}(A)}{c_{1:\ell-1}}\,q_1 + \sum_{l=1}^{\ell-1}\frac{\chi_{l+1:\ell-1}(A)}{c_{l:\ell-1}}f_l\right)\frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}}\hat{\nu}_{j+1}(0)\\
&= \sum_{\ell=1}^{k}\sum_{j=\ell}^{k}\frac{\chi_{1:\ell-1}(A)}{c_{1:\ell-1}}\cdot\frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}}\hat{\nu}_{j+1}(0)\,q_1\\
&\qquad + \sum_{\ell=1}^{k}\sum_{j=\ell}^{k}\sum_{l=1}^{\ell-1}\frac{\chi_{l+1:\ell-1}(A)}{c_{l:\ell-1}}\cdot\frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}}\hat{\nu}_{j+1}(0)f_l\\
&= \sum_{j=1}^{k}\left(\sum_{\ell=1}^{j}\frac{\chi_{1:\ell-1}(A)}{c_{1:\ell-1}}\cdot\frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}}\right)\hat{\nu}_{j+1}(0)\,q_1\\
&\qquad + \sum_{l=1}^{k-1}\sum_{j=l}^{k}\left(\sum_{\ell=l+1}^{j}\frac{\chi_{l+1:\ell-1}(A)}{c_{l:\ell-1}}\cdot\frac{-\chi_{\ell+1:j}(0)}{c_{\ell:j}}\right)\hat{\nu}_{j+1}(0)f_l\\
&= \sum_{j=1}^{k}\hat{\nu}_{j+1}(0)\frac{-\mathcal{A}_j(0,A)}{c_{1:j}}\,q_1 + \sum_{l=1}^{k-1}\left(\sum_{j=l}^{k}\hat{\nu}_{j+1}(0)\frac{-\mathcal{A}_{l+1:j}(0,A)}{c_{l:j}}\right)f_l.
\end{aligned} \tag{5.39}$$

We multiply by $\|\underline{r}_0\|/\|\underline{\check{\nu}}(0)\|^2$ and insert $\underline{\mathcal{L}}_k[z^{-1}](z)$ from its definition in equation (5.13). Equation (5.36) follows since by definition $\mathcal{A}_{k+1:k}\equiv 0$. ◻

**5.3. Errors.** When the system matrix $A\in\mathbb{C}^{n\times n}$ underlying an abstract perturbed KRYLOV method is regular, we define $\underline{x}\equiv A^{-1}\underline{r}_0$. We can use both the theorem on the residuals and the theorem on the iterates to obtain the following expression for the errors $\underline{x}-\underline{x}_k$.

THEOREM 5.9 (the (Q)MR error vectors in case of regular $A$). *Suppose an abstract* KRYLOV *method* (1.1) *is given with* $q_1=\underline{r}_0/\|\underline{r}_0\|$. *Let* $\underline{x}_k=Q_k\underline{z}_k$ *denote the kth (Q)MR iterate and define* $\underline{x}\equiv A^{-1}\underline{r}_0$.

*Then*

$$\underline{x}-\underline{x}_k = \underline{\mathcal{R}}_k(A)(\underline{x}-0) + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2}\sum_{l=1}^{k}\left(\sum_{j=l}^{k}\hat{\nu}_{j+1}(0)\frac{\mathcal{A}_{l+1:j}(0,A)}{c_{l:j}}\right)f_l. \tag{5.40}$$

*Proof.* We start with $\underline{x}=\underline{x}$ and subtract equation (5.36). We group the leading terms and use the identity

$$\underline{\mathcal{R}}_k(A)(\underline{x}-0) = (I-\underline{\mathcal{L}}_k[z^{-1}](A)A)\underline{x} = \underline{x}-\underline{\mathcal{L}}_k[z^{-1}](A)\underline{r}_0. \tag{5.41}$$

This results in equation (5.40). ◻

When $A$ is singular we use again the DRAZIN inverse. A splitting into the regular and the nilpotent part like in the (Q)OR case proves the following theorem.

THEOREM 5.10 (the (Q)MR error vectors in case of singular $A$). *Suppose an abstract* KRYLOV *method* (1.1) *is given with* $q_1=\underline{r}_0/\|\underline{r}_0\|$. *Let* $\underline{x}_k=Q_k\underline{z}_k$ *denote the kth (Q)MR iterate and define* $\underline{x}\equiv A^{D}\underline{r}_0$.

*Then*

$$\begin{aligned}\underline{x} - \underline{x}_k &= \underline{\mathcal{R}}_k(A)(\underline{x} - 0) - \underline{\mathcal{L}}_k(A)V_0\hat{V}_0^H\underline{r}_0 \\ &\quad + \frac{\|\underline{r}_0\|}{\|\underline{\check{\nu}}(0)\|^2}\sum_{l=1}^{k}\left(\sum_{j=l}^{k}\hat{\nu}_{j+1}(0)\frac{\mathcal{A}_{l+1:j}(0,A)}{c_{l:j}}\right)f_l.\end{aligned} \tag{5.42}$$

*Proof.* We start with $\underline{x} = \underline{x}$ and subtract equation (5.36). The multiplication of the Drazin "solution" $\underline{x} = A^D\underline{r}_0$ by $A$ results in

$$A\underline{x} = AA^D\underline{r}_0 = r_0 - V_0\hat{V}_0 r_0. \tag{5.43}$$

Thus a slightly modified variant of equation (5.41) holds for the Drazin "solution" $\underline{x}$, namely

$$\begin{aligned}\underline{\mathcal{R}}_k(A)(\underline{x} - 0) &= (I - \underline{\mathcal{L}}_k[z^{-1}](A)A)\underline{x} \\ &= \underline{x} - \underline{\mathcal{L}}_k[z^{-1}](A)\underline{r}_0 + \underline{\mathcal{L}}_k[z^{-1}](A)V_0\hat{V}_0 r_0.\end{aligned} \tag{5.44}$$

Replacing the occurrence of $\underline{x} - \underline{\mathcal{L}}_k[z^{-1}](A)\underline{r}_0$ in the equation resulting from subtracting equation (5.36) from $\underline{x} = \underline{x}$ proves equation (5.42). □

**6. Conclusion and Outlook.** We have successfully applied the Hessenberg eigenvalue-eigenmatrix relations derived in [12] to abstract perturbed Krylov subspace methods. The investigation carried out in this paper sheds some light on the methods and introduces a new point of view. This new abstract point of view on perturbed Krylov subspace methods enables no detailed convergence analysis but unifies important parts of the analysis considerably. In this abstract setting, without any additional knowledge on the properties of the computed matrices $C_k$ ($\underline{C}_k$), we cannot make any statements on the convergence of, say, the Ritz *values*, but the convergence of the Ritz *vectors* and the (Q)OR iterates can be described independently of the particular method under investigation in terms of the unknown Ritz values. Even though we cannot compute bounds on the distance of the eigenvalues and the computed Ritz values, Theorem 2.3, and its refinement Theorem 2.4, clearly reveal that in case of random errors the Ritz values can only accidentally come arbitrarily close to the eigenvalues of and that the occurrence of multiple Ritz values is extremely unlikely.

At least in the humble opinion of the author, the main achievement of the paper lies not in working out the polynomial structure presented in the results, but in the deeper understanding of most of the polynomials involved and their close connection to approximation problems. In this sense the author has failed with respect to the polynomials amplifying the perturbation terms in the theorems on the (Q)MR case. Here is room for improvements.

Missing, mainly for reasons of space is the application of the results to a single instance of a Krylov subspace method, which may be in the form of the derivation of bounds, backward errors, convergence theorems or the stabilization of existing algorithms or even the development of entirely new methods based on the abstract insights given here. A detailed application of the results to the symmetric finite precision algorithm of Lanczos will follow.

The generalization of the approach of abstraction to Krylov subspace methods *not* covered by equation (1.1) must be based on a corresponding generalization of the underlying results on HESSENBERG matrices. A typical candidate for such a generalization are block HESSENBERG matrices which would allow for a treatment of many block KRYLOV subspace methods.

## REFERENCES


[1] A. BOURAS AND V. FRAYSSÉ, *Inexact matrix-vector products in Krylov methods for solving linear systems: a relaxation strategy*, 26 (2005), pp. 660–678.

[2] R. W. FREUND, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, Journal of Computational and Applied Mathematics, 43 (1992), pp. 135–158. Orthogonal polynomials and numerical methods.

[3] F. R. GANTMACHER, *Matrizentheorie*, Springer-Verlag, Berlin, 1986.

[4] M. H. GUTKNECHT, *Lanczos-type solvers for nonsymmetric linear systems of equations*, Acta Numerica, 6 (1997), pp. 271–397.

[5] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1994.

[6] I. C. F. IPSEN AND C. D. MEYER, *The idea behind Krylov methods*, American Mathematical Monthly, 105 (1998), pp. 889–899.

[7] V. SIMONCINI AND D. B. SZYLD, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.

[8] G. L. SLEIJPEN AND H. A. VAN DER VORST, *Krylov subspace methods for large linear systems of equations*, Preprint 803, Department of Mathematics, University Utrecht, 1993.

[9] O. TAUSSKY, *The factorization of the adjugate of a finite matrix*, Linear Algebra and Its Applications, 1 (1968), pp. 39–41.

[10] J. VAN DEN ESHOF AND G. L. G. SLEIJPEN, *Inexact Krylov subspace methods for linear systems*, 26 (2004), pp. 125–153.

[11] J.-P. M. ZEMKE, *Krylov Subspace Methods in Finite Precision: A Unified Approach*, Dissertation, Technische Universität Hamburg-Harburg, Inst. f. Informatik III, Technische Universität Hamburg-Harburg, Schwarzenbergstraße 95, 21073 Hamburg, Germany, 2003.

[12] ———, *(Hessenberg) eigenvalue-eigenmatrix relations*, Bericht 78, Arbeitsbereich Mathematik 4-13, Technische Universität Hamburg-Harburg, Schwarzenbergstraße 95, 21073 Hamburg, September 2004. Submitted to Linear Algebra and Its Applications.