Functions of a Matrix and Krylov Matrices

Hongguo Xu*

Abstract

For a given nonderogatory matrix A, formulas are given for functions of A in terms of Krylov matrices of A. Relations between the coefficients of a polynomial of A and the generating vector of a Krylov matrix of A are provided. With the formulas, linear transformations between Krylov matrices and functions of A are introduced, and associated algebraic properties are derived. Hessenberg reduction forms are revisited equipped with appropriate inner products and related properties and matrix factorizations are given.

Keywords Keywords Krylov matrix, Krylov subspace, function of a matrix, polynomial of a matrix, Hessenberg matrix, companion matrix
AMS subject classification. 15A21, 65F15.

1   Introduction

Let $A \in \mathbb{C}^{n,n}$. The Krylov matrix of A generated by a vector $b \in \mathbb{C}^n$ is given by

$$\begin{bmatrix} b & Ab & \dots & A^{n-1}b \end{bmatrix} \in \mathbb{C}^{n,n}.$$

Given a scalar function $f(t)$ that is well defined on the spectrum of A, one defines a matrix $f(A) \in \mathbb{C}^{n,n}$, which is usually called a *function of A*, e.g., [10, 11].

Both functions of a matrix and Krylov matrices play a fundamental role in matrix computations. They are key tools in understanding and developing numerical methods for solving eigenvalue problems and systems of linear equations, including the QR algorithm and Krylov subspace methods, e.g., [8, 9, 22]. Functions of a matrix arise from a variety of applications. The development of numerical algorithms is still a challenging topic. See the recently published book [10] for details.

In this paper we investigate the Krylov matrices and functions of a matrix. We focus on the situation where the associated matrix A is nonderogatory, i.e., the geometric multiplicity of every eigenvalue of A is one. We provide formulas to express a function of A in terms of Krylov matrices and vise versa, based on a simple observation. We use the formulas to study the relations and properties of these two objects.

Krylov matrices and functions of a matrix have been studied extensively in the past several decades. Still, it seems that their behaviors have not been fully understood. The goal of this study is to use a new angle to interpret the existing properties and provide new insight that may be potentially useful for the development of numerical methods.

The paper is organized as follows. In Section 2 we give definitions of functions of a matrix and Krylov matrices, and some basic properties that are necessary for deriving main results. In Section 3 we show relations between functions of a matrix and Krylov matrices by providing explicit formulas. In Section 4 we interpret the relations in terms of linear transformations and subspaces. In Section 5 we study the Hessenberg reduction forms, and derive some related properties and matrix factorizations. In Section 6 we give conclusions.

The spectrum of A is denoted by $\lambda(A)$. $\|\cdot\|$ stands for both the Euclidian norm of a vector and the spectral norm of a matrix. I_n is the $n \times n$ identity matrix, and e_j is the jth column of I_n. N_r is the $r \times r$ nilpotent matrix with 1 on the super diagonal and 0 elsewhere, and $N_r(\lambda) = \lambda I_r + N_r$. A square matrix is called *unreduced upper Hessenberg* if it is upper Hessenberg with nonzero subdiagonal elements. $\mathbb{P}_m$ denotes the space of the polynomials with degree no greater than m.

2 Functions of a matrix and Krylov matrices

In this paper we only consider the functions defined as follows. Let A be a square matrix and have the Jordan canonical form

$$Z^{-1}AZ = \mathrm{diag}(N_{r_{1,1}}(\lambda_1), \ldots, N_{r_{1,s_1}}(\lambda_1), \ldots, N_{r_{\eta,1}}(\lambda_\eta), \ldots, N_{r_{\eta,s_\eta}}(\lambda_\eta))$$

where $\lambda_1, \ldots, \lambda_\eta \in \lambda(A)$ are distinct. Let $f(t)$ be a scalar function. If for each λ_i, $f(\lambda_i)$ and the derivatives $f^{(k)}(\lambda_i)$ $(k = 1, \ldots, \max_{1 \leq j \leq s_i} r_{i,j} - 1)$ are defined, we define

$$f(A) := Z \, \mathrm{diag}(f(N_{r_{1,1}}(\lambda_1)), \ldots, f(N_{r_{1,s_1}}(\lambda_1)), \ldots, f(N_{r_{\eta,1}}(\lambda_\eta)), \ldots, f(N_{r_{\eta,s_\eta}}(\lambda_\eta)))Z^{-1},$$

where

$$f(N_{r_{i,j}}(\lambda_i)) = \begin{bmatrix} f(\lambda_i) & \frac{f'(\lambda_i)}{1!} & \cdots & \frac{f^{(r_{i,j}-1)}(\lambda_i)}{(r_{i,j}-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{bmatrix}.$$

For a scalar polynomial $p(t) = \sum_{j=0}^m \alpha_j t^j \in \mathbb{P}_m$ we simply have $p(A) = \sum_{j=0}^m \alpha_j A^j$.

We provide below some basic properties of functions of a matrix.

Proposition 2.1 ([10, 11]) *Suppose that μ is the degree of the minimal polynomial of A. For any function $f(t)$ such that $f(A)$ is defined, there exists a unique polynomial $p(t) \in \mathbb{P}_{\mu-1}$ such that*

$$f(A) = p(A).$$

The unique polynomial $p(t)$ can be constructed by the Lagrange-Hermite interpolation with $p^{(k)}(\lambda_i) = f^{(k)}(\lambda_i)$ for $k = 0, 1, \ldots, \max_{1 \leq j \leq s_i} r_{i,j} - 1$, and $i = 1, 2, \ldots, \eta$.

Proposition 2.2 ([10, 11])

(i) $Af(A) = f(A)A.$

(ii) $f(X^{-1}AX) = X^{-1}f(A)X.$

The Schur-Parlett algorithm for computing $f(A)$ is based on these two properties. See [16, 17, 12, 3] and [10, Ch. 4, 9]. The properties will be used frequently in the rest of the paper.

Suppose that $A \in \mathbb{C}^{n,n}$ and $b \in \mathbb{C}^n$. We define the Krylov matrix

$$K_{n,m}(A, b) = \begin{bmatrix} b & Ab & \dots & A^{m-1}b \end{bmatrix} \in \mathbb{C}^{n,m}.$$

When $m = n$, we will simply use the notation $K_n(A, b)$ or $K(A, b)$.

A polynomial with degree no greater than $m - 1$ is characterized uniquely by its m coefficients. We use the following polynomial notation to emphasize the coefficients.

Definition 2.3 *For $x = [x_1, \dots, x_m]^T \in \mathbb{C}^m$,*

$$p(t; x) := x_1 + x_2 t + \dots + x_m t^{m-1} \in \mathbb{P}_{m-1}.$$

It is obvious that

$$p(A; x)b = K_{n,m}(A, b)x, \quad x \in \mathbb{C}^m. \tag{1}$$

So $K_{n,m}(A, b)x = 0$ if and only if $p(A; x)b = 0$. The *minimal polynomial of b with respect to A* is a nonzero polynomial $p(t)$ of the lowest degree such that $p(A)b = 0$, [22, pp. 36 - 37]. Let ν be the degree of this minimal polynomial $p(t)$. Then based on (1),

$$\operatorname{rank} K_{n,m}(A, b) = \min\{m, \nu\}.$$

More precisely, $b, Ab, \dots, A^{\nu-1}b$ are linearly independent, and for any $k \geq \nu$, $A^k b$ can be expressed as a linear combination of $b, Ab, \dots, A^{\nu-1}b$, [19, Ch. VI].

Proposition 2.4 *Suppose $\operatorname{rank} K_{n,m}(A, b) = r$. Then there exists a nonsingular matrix $X = [X_1, X_2] \in \mathbb{C}^{n,n}$ with $X_1 \in \mathbb{C}^{n,r}$ and $\operatorname{range} X_1 = \operatorname{range} K_{n,r}(A, b)$ such that*

$$K_{n,m}(A, b) = X_1 \begin{bmatrix} R_{11} & R_{12} \end{bmatrix}, \tag{2}$$

where $R_{11} \in \mathbb{C}^{r,r}$ is nonsingular, and

$$X^{-1}AX = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad X^{-1}b = \begin{bmatrix} b_1 \\ 0 \end{bmatrix} \tag{3}$$

with $A_{11} \in \mathbb{C}^{r,r}$ and $b_1 \in \mathbb{C}^r$.

Moreover, R_{11} is upper triangular if and only if A_{11} is unreduced upper Hessenberg and $X^{-1}b = \gamma e_1$.

Proof. It is trivial when $b = 0$. So we only consider the case when $b \neq 0$.

Since $\operatorname{rank} K_{n,m}(A, b) = r$, based on the above arguments,

$$K_{n,m}(A, b) = K_{n,r}(A, b) \begin{bmatrix} I_r & T \end{bmatrix},$$

for some $T \in \mathbb{C}^{r,m-r}$. If $X_1 \in \mathbb{C}^{n,r}$ satisfies $\operatorname{range} X_1 = \operatorname{range} K_{n,r}(A, b)$, then $X_1 = K_{n,r}(A, b)Z$ for some nonsingular matrix $Z \in \mathbb{C}^{r,r}$. So we have (2) with

$$\begin{bmatrix} R_{11} & R_{12} \end{bmatrix} = Z^{-1} \begin{bmatrix} I_r & T \end{bmatrix}.$$

3

Clearly, $b = X_1 b_1$ with $b_1 = Z^{-1} e_1$. Because $A^r b$ is a linear combination of $b, Ab, \ldots, A^{r-1} b$, we have

$$AK_{n,r}(A,b) = K_{n,r}(A,b)C_r, \quad C_r = \begin{bmatrix} 0 & & & c_{1,r} \\ 1 & \ddots & & c_{2,r} \\ & \ddots & \ddots & \vdots \\ & & 1 & c_{r,r} \end{bmatrix}.$$

Then

$$AX_1 = X_1 A_{11}, \qquad A_{11} = Z^{-1} C_r Z.$$

So for a nonsingular matrix $X = [X_1, X_2]$ we have (3).

Suppose that R_{11} in (2) is upper triangular. Then $Z = R_{11}^{-1}$ is also upper triangular. So A_{11} is unreduced upper Hessenberg. Because $b_1 = Z^{-1} e_1 = R_{11} e_1 = r_{11} e_1$, we have $X^{-1} b = r_{11} e_1 =: \gamma e_1$.

Conversely, if A_{11} is unreduced upper Hessenberg and $X^{-1} b = \gamma e_1$, using (3),

$$K_{n,r}(A,b) = X \left(\gamma \begin{bmatrix} e_1 & A_{11} e_1 & \ldots & A_{11}^{r-1} e_1 \\ 0 & 0 & \ldots & 0 \end{bmatrix} \right) =: \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} R_{11} \\ 0 \end{bmatrix} = X_1 R_{11},$$

and it is straightforward to show that R_{11} is nonsingular and upper triangular. $\square$

We now turn to a square Krylov matrix $K(A,b)$ $(m = n)$. Suppose that the characteristic polynomial of A is

$$\det(\lambda I - A) =: \lambda^n - c_n \lambda^{n-1} - \ldots - c_2 \lambda - c_1.$$

We define the companion matrix[1] of A as

$$C = \begin{bmatrix} 0 & & & c_1 \\ 1 & \ddots & & c_2 \\ & \ddots & \ddots & \vdots \\ & & 1 & c_n \end{bmatrix}. \tag{4}$$

Proposition 2.5 $AX = XC$ *if and only if* $X = K(A,b)$ *for some* $b \in \mathbb{C}^n$.

Proof. For sufficiency, it is straightforward to show $AK(A,b) = K(A,b)C$ for any $b \in \mathbb{C}^n$, by using the Cayley-Hamilton Theorem.

For necessity, $X = K(A,b)$ follows simply by comparing the columns of the matrices AX and XC with $b = Xe_1$. $\square$

The rank of $K(A,b)$ is ν, the degree of the minimal polynomial of b with respect to A, which is no greater than the degree of the minimal polynomial of A. In order for $K(A,b)$ to be nonsingular, it is necessary for the minimal polynomial of A to be the same as its characteristic polynomial, or equivalently, A has to be nonderogatory, i.e., the geometric multiplicity for every eigenvalue is one [6, 7]. Still, the nonsingularity of $K(A,b)$ depends on the vector b. There are numerous equivalence conditions based on canonical forms [1] and the controllability from linear system theory [13, 18, 4, 5]. We list a few of them in the following proposition.

Proposition 2.6 *Suppose* $A \in \mathbb{C}^{n,n}$ *and* $b \in \mathbb{C}^n$. *The following statements are equivalent.*

---

[1]Usually the transpose of C is also called a companion matrix of A. In this paper we always refer the companion matrix of A to C.

(a) $K(A, b)$ *is nonsingular.*

(b) $y^*b \neq 0$ *for any vector* $y \in \mathbb{C}^n$ *satisfying* $y^*A = \lambda y^*$ *with some* $\lambda \in \mathbb{C}$.

(c) $\operatorname{rank} \begin{bmatrix} A - \lambda I & b \end{bmatrix} = n$ *for all* $\lambda \in \mathbb{C}$.

(d) $b \neq 0$ *and there exists a nonsingular (or unitary)* X *such that* $X^{-1}b = \gamma e_1$ *and* $X^{-1}AX$ *is unreduced upper Hessenberg.*

(e) *The only polynomial* $p(t; x) \in \mathbb{P}_{n-1}$ *that satisfies*

$$p(A; x)b = 0$$

is $p(t; 0) \equiv 0$.

Proof. (a), (b), (c) are just three equivalence conditions for (A, b) to be controllable [13, 18, 4, 5]. The equivalence between (a) and (d) is from Proposition 2.4. The equivalence between (a) and (e) can be shown by using (1). □

Generically, a square matrix A is nonderogatory. When A is nonderogatory, the left eigenvector space of each eigenvalue is one-dimensional. So the set of b that satisfies Proposition 2.6 (b) is dense in $\mathbb{C}^n$, and the Krylov matrix $K(A, b)$ is generically nonsingular.

When A is derogatory, it is impossible for A to be similar to its companion matrix. Instead, A has a Frobenius form ([22, pp. 15 - 16]),

$$X^{-1}AX = \operatorname{diag}(C_1, \ldots, C_q),$$

where $C_1, \ldots, C_q$ are in a companion matrix form, and the characteristic polynomial of each C_j divides the characteristic polynomials of $C_1, \ldots, C_{j-1}$. Suppose the size of C_j is $n_j \times n_j$, for $j = 1, \ldots, q$. Then the similarity matrix X can be expressed as

$$X = [K_{n,n_1}(A, b_1); \ldots, K_{n,n_q}(A, b_q)],$$

for some $b_1, \ldots, b_q \in \mathbb{C}^n$, which generalizes the result in Proposition 2.5. In this paper, however, we focus on the nonderogatory case only, although some results can be generalized to the derogatory case by using the above observation.

3 Relations between Krylov matrices and functions of a matrix

The formulation of a function of A in terms of Krylov matrices of A is based on the following simple observation. For any $A \in \mathbb{C}^{n,n}$ and and $b \in \mathbb{C}^n$, using the fact $f(A)A = Af(A)$ (Proposition 2.2) we have

$$f(A)K(A, b) = K(A, d), \quad d = f(A)b. \tag{5}$$

We first use this fact to show relations between polynomials $p(A; x)$ and Krylov matrices.

Theorem 3.1 *For any* $x \in \mathbb{C}^n$,

$$K(A, d) = p(A; x)K(A, b), \qquad d = K(A, b)x, \tag{6}$$

If $K(A, b)$ is nonsingular (so A is necessarily nonderogatory), then for any $d \in \mathbb{C}^n$,

$$p(A; x) = K(A, d)K(A, b)^{-1}, \qquad x = K(A, b)^{-1}d, \tag{7}$$

and in this case (6) can be also expressed as

$$K(A, d) = K(A, b)p(C; x), \qquad d = K(A, b)x, \tag{8}$$

where C is the companion matrix of A.

Proof. The formula (6) follows simply from (5) with $f(t) = p(t; x)$, and (1).
The formula (7) is simply from (6) and the nonsingularity of $K(A, b)$. (8) follows from

$$p(A; x)K(A, b) = p(K(A, b)CK(A, b)^{-1}; x)K(A, b) = K(A, b)p(C; x),$$

based on Proposition 2.2 (ii) and Proposition 2.5. □
We now consider a general function $f(t)$ and we have to following results.

Theorem 3.2 *Suppose that $K(A, b)$ is nonsingular and C is the companion matrix of A. Let $f(t)$ be a scalar function and $\tau \in \mathbb{C}$ such that $f(\tau A)$ is defined. Then*

$$f(\tau A) = K(A, d(\tau))K(A, b)^{-1}, \qquad d(\tau) = f(\tau A)b, \tag{9}$$

and

$$f(\tau A) = p(A; x(\tau)), \qquad x(\tau) = K(A, b)^{-1}f(\tau A)b = f(\tau C)e_1. \tag{10}$$

Also, when $\tau \neq 0$,

$$f(\tau A) = p(\tau A; y(\tau)), \quad y(\tau) = K(\tau A, b)^{-1}f(\tau A)b = f(C(\tau))e_1, \tag{11}$$

where

$$C(\tau) = \begin{bmatrix} 0 & & & \tau^n c_1 \\ 1 & \ddots & & \tau^{n-1}c_2 \\ & \ddots & \ddots & \vdots \\ & & 1 & \tau c_n \end{bmatrix} \tag{12}$$

is the companion matrix of τA.

Proof. Because $f(\tau A)$ and A^i $(i = 0, \ldots, n-1)$ commute,

$$f(\tau A) = f(\tau A)K(A, b)K(A, b)^{-1} = K(A, f(\tau A)b)K(A, b)^{-1} = K(A, d(\tau))K(A, b)^{-1}.$$

By (9) and (7),

$$f(\tau A) = K(A, d(\tau))K(A, b)^{-1} = p(A; x(\tau)),$$

where

$$x(\tau) = K(A, b)^{-1}d(\tau) = K(A, b)^{-1}f(\tau A)b = f(\tau C)e_1.$$

For (11), define $D = \mathrm{diag}(1, \tau, \ldots, \tau^{n-1})$, which is nonsingular. From the simple relation

$$K(\tau A, b) = K(A, b)D, \tag{13}$$

6

$K(\tau A, b)$ is also nonsingular. By applying (5) with A replaced by τA, we get

$$f(\tau A) = K(\tau A, d(\tau))K(\tau A, b)^{-1}, \quad d(\tau) = f(\tau A)b.$$

Again, using (7) we have

$$f(\tau A) = p(\tau A; y(\tau)),$$

where

$$y(\tau) = K(\tau A, b)^{-1}d(\tau) = K(\tau A, b)^{-1}f(\tau A)b = f(C(\tau))e_1,$$

and $C(\tau)$ is the companion matrix of τA.

It remains to derive the formula for $C(\tau)$. By using (13) we have

$$C(\tau) = K(\tau A, b)^{-1}(\tau A)K(\tau A, b) = D^{-1}K(A, b)^{-1}(\tau A)K(A, b)D = \tau D^{-1}CD,$$

which has the form (12). □

Note that when $\tau = 0$, (11) may not hold, since on the left-hand side it only requires $f(0)$ to be defined while on the right-hand side $f(C(0))$ has to be defined. Even if (11) holds, it usually doesn't give a polynomial corresponding to $f(0)$ with minimal degree. This is because $y(0) = f(C(0))e_1$ may not be a scalar multiple of e_1, resulting a polynomial $p(t; y(0))$ with degree greater than 0. Note also that when $\tau = 1$, (10) and (11) are identical.

The following results are directly from Theorem 3.2.

Theorem 3.3 *Suppose that $K(A, b)$ is nonsingular and C is the companion matrix of A. Let $f(t)$ be a scalar function such that $f(A)$ is defined. Then*

$$f(A) = K(A, d)K(A, b)^{-1}, \quad d = f(A)b, \tag{14}$$

and

$$f(A) = p(A; x), \quad x = K(A, b)^{-1}f(A)b = f(C)e_1. \tag{15}$$

If in addition $f(A)$ is nonsingular, then

$$[f(A)]^{-1} = p(A; y), \quad y = K(A, d)^{-1}b. \tag{16}$$

Proof. The first part is from Theorem 3.2 with $\tau = 1$. So we only need to prove (16).

If $f(A)$ is nonsingular, then from (14), $K(A, d)$ is also nonsingular. So $[f(A)]^{-1} = K(A, b)K(A, d)^{-1}$, and (16) is from (15). □

Formula (15) not only restates the result given in Proposition 2.1 in the nonderogatory case, i.e., $f(A) \in \mathbb{P}_{n-1}(A)$, but also provides an explicit formula for the polynomial $p(t; x)$. Formula (16) shows that the same properties hold for the inverse of $f(A)$.

Formula (14) holds true for all $f(A)$. When $f(t)$ is a rational function, we have the following additional formula.

Theorem 3.4 *Suppose that $K(A, b)$ is nonsingular and $r(t) = p(t)/q(t)$ is a rational function with $q(A)$ nonsingular. Then*

$$r(A) = K(A, d_1)K(A, d_2)^{-1}, \quad d_1 = p(A)b, \quad d_2 = q(A)b, \tag{17}$$

and

$$r(A) = p(A; x), \quad x = K(A, d_2)^{-1}p(A)b. \tag{18}$$

Proof. Since $K(A,b)$ is nonsingular,

$$r(A) = p(A)q(A)^{-1} = p(A)K(A,b)K(A,b)^{-1}q(A)^{-1} = K(A,p(A)b)K(A,q(A)b)^{-1},$$

which gives (17). The relations in (18) are from (17) and (7). $\square$

Remark 3.5 Computing a Krylov matrix $K(A,b)$ has the same cost of a matrix-matrix multiplication. So if $f(A)b$ is available, computing $f(A)$ with (14) or (17) requires two matrix-matrix multiplications and one matrix equation solving.

In general, computing the vector $f(A)b$ is far from trivial, but it is straightforward when $f(t)$ is a polynomial or a rational function. So this approach may have advantages in symbolic or exact arithmetic computations. For numerical computations, however, it is well-known that a Krylov matrix is usually ill-conditioned. A method that uses (14) or (17) directly may be numerically unstable.

The above formulations may be used to derive some interesting results. For instance, let $f(t) = e^t$. Then from Theorem 3.2, we have

$$e^{\tau A} = K(A, d(\tau))K(A,b)^{-1}, \quad d(\tau) = e^{\tau A}b.$$

This shows that the fundamental matrix of the linear system $dx/d\tau = Ax$ is completely determined by the solution to the initial value problem $dx/d\tau = Ax$, $x(0) = b$.

In the end of this section, we consider the case where Krylov matrices are slightly generalized.

Let

$$g_j(t) = p(t; \gamma_j) \in \mathbb{P}_{n-1}, \quad \gamma_j \in \mathbb{C}^n,$$

for $j = 1, \ldots, n$. Define

$$G(A,b) = \begin{bmatrix} g_1(A)b & g_2(A)b & \ldots & g_n(A)b \end{bmatrix}. \tag{19}$$

By using (1),

$$g_j(A)b = K(A,b)\gamma_j,$$

for $j = 1, \ldots, n$. Define

$$\Gamma = \begin{bmatrix} \gamma_1 & \ldots & \gamma_n \end{bmatrix}.$$

Then

$$G(A,b) = K(A,b)\Gamma, \tag{20}$$

Clearly for any $\Gamma \in \mathbb{C}^{n,n}$ a matrix $G(A,b)$ can be generated by using (20). When $K(A,b)$ is nonsingular, it defines an isomorphism from Γ to $G(A,b)$. Note also that $G(A,b)$ is nonsingular if and only if both $K(A,b)$ and Γ are nonsingular.

Corollary 3.6 *Suppose that $G(A,b)$ defined in (19) with $g_1(t), \ldots, g_n(t) \in \mathbb{P}_{n-1}$ is nonsingular. Let $f(t)$ be a scalar function and $\tau \in \mathbb{C}$ be a scalar such that $f(\tau A)$ is defined. Then*

$$f(\tau A) = G(A, d(\tau))G(A,b)^{-1}, \quad d(\tau) = f(\tau A)b. \tag{21}$$

Proof. The proof is trivial. $\square$

Remark 3.7 All the results established in this section apply to the matrices and vectors defined over any field as long as $f(\tau A)$ is defined and satisfies $f(\tau A)A = Af(\tau A)$.

4 Connections to subspaces and linear transformations

In this section, we interpret Krylov matrices and polynomials of a matrices in terms of linear transformations.

For any vectors $b_1, b_2 \in \mathbb{C}^n$ and scalars α, β, we have

$$K(A, \alpha b_1 + \beta b_2) = \alpha K(A, b_1) + \beta K(A, b_2).$$

So the matrix A introduces a linear transformation: $\mathbb{C}^n \to \mathbb{C}^{n,n}$ defined by $b \to K(A, b)$. The range of the transformation is the set of the Krylov matrices of A:

$$\mathbb{K}(A) = \{K(A, b) \,|\, b \in \mathbb{C}^n\},$$

which is a subspace of $\mathbb{C}^{n,n}$. Clearly, $\dim \mathbb{K}(A) = n$.

Let $L(\mathbb{K}(A))$ be the space of the linear operators on $\mathbb{K}(A)$. It has the dimension n^2. Suppose $\mathbf{T} \in L(\mathbb{K}(A))$ and $T \in \mathbb{C}^{n,n}$ is its matrix with the basis $\{K(A, e_j)\}_{j=1}^n$. Then

$$\mathbf{T}K(A, b) = K(A, Tb), \quad \forall b \in \mathbb{C}^n.$$

So we may identify $L(\mathbb{K}(A))$ with $\mathbb{C}^{n,n}$ based on the above relation.

Now consider a subspace of $L(\mathbb{K}(A))$ defined by

$$L_c(\mathbb{K}(A)) = \{\mathbf{T} \,|\, \mathbf{T}K(A, b) = K(A, Tb) = TK(A, b), \quad \forall b \in \mathbb{C}^n\}.$$

Define

$$\mathbb{P}_{n-1}(A) = \{p(A; x) \,|\, x \in \mathbb{C}^n\}.$$

From (6), $p(A; x) \in L_c(\mathbb{K}(A))$ for any $x \in \mathbb{C}^n$. Hence $\mathbb{P}_{n-1}(A) \subseteq L_c(\mathbb{K}(A))$.

Theorem 4.1 *Suppose* $A \in \mathbb{C}^{n,n}$. *Then* $\mathbb{P}_{n-1}(A) = L_c(\mathbb{K}(A))$ *if and only if* $A \in \mathbb{C}^{n,n}$ *is nonderogatory.*

Proof. For any $\mathbf{T} \in L_c(\mathbb{K}(A))$, the corresponding matrix T satisfies $K(A, Tb) = TK(A, b)$ for all $b \in \mathbb{C}^n$ if and only if $TA = AT$. Without distinguishing $\mathbf{T}$ and its matrix T we have

$$L_c(\mathbb{K}(A)) = \{T \,|\, TA = AT, \ T \in \mathbb{C}^{n,n}\},$$

which is called the centralizer of A [11, pp. 275]. With this connection, the equivalence relations follow from [11, Corollary 4.4.18]. $\square$

Because $\mathbb{K}(A)$ and $\mathbb{C}^n$ are isomorphic, when A is nonderogatory, $\mathbb{P}_{n-1}(A)$ and $\mathbb{C}^n$ are also isomorphic. So $\mathbb{P}_{n-1}(A)$ and $\mathbb{K}(A)$ are isomorphic. Then $L(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$, the space of linear transformations from $\mathbb{P}_{n-1}(A)$ to $\mathbb{K}(A)$, has the dimension n^2, and it is isomorphic to $\mathbb{C}^{n,n}$. For any $S \in \mathbb{C}^{n,n}$ we may introduce $\mathbf{S} \in L(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$ defined by

$$\mathbf{S}p(A; x) = K(A, Sx), \quad \forall x \in \mathbb{C}^n.$$

(Again, S is considered as the matrix of $\mathbf{S}$ with the bases $\{A^j\}_{j=0}^{n-1}$ and $\{K(A, e_j)\}_{j=1}^n$.)

Define

$$L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A)) = \{\mathbf{S} \,|\, \mathbf{S}p(A; x) = K(A, Sx) = p(A; x)S, \quad x \in \mathbb{C}^n\} \subseteq L(\mathbb{P}_{n-1}(A), \mathbb{K}(A)).$$

Theorem 4.2 *Suppose A is nonderogatory. Then*

$$\mathbb{K}(A) = L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A)).$$

Proof. Formula (6) shows that for any $b \in \mathbb{C}^n$, the linear transformation **S** corresponding to $S := K(A, b)$ is in $L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$. So if we don't distinguish **S** with S, we have

$$\mathbb{K}(A) \subseteq L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A)).$$

On the other hand, for each $\mathbf{S} \in L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$, the corresponding matrix S satisfies $K(A, Sx)e_j = p(A; x)Se_j$ for $j = 1, \ldots, n$. Using these relations and (1), we have

$$A^{j-1}Sx = K(A, Sx)e_j = p(A; x)Se_j = K(A, Se_j)x, \quad \forall x \in \mathbb{C}^n,$$

which implies

$$A^{j-1}S = K(A, Se_j), \tag{22}$$

for $j = 1, \ldots, n$. Let $b = Se_1$. Then setting $j = 1$ in (22) we have $S = K(A, b)$, and with which (22) holds for $j = 2, \ldots, n$. We then have shown

$$L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A)) \subseteq \mathbb{K}(A).$$

Therefore, the two spaces are the same. $\square$

Theorems 4.1 and 4.2 show that when A is nonderogatory, a linear operator on $\mathbb{K}(A)$ is just a polynomial $p(A; x)$ and a linear transformation from $\mathbb{P}_{n-1}$ to $\mathbb{K}(A)$ is just a Krylov matrix $K(A, b)$, both described by (6). A common technique to generate a new Krylov matrix from $K(A, b)$ is to choose a new initial vector $d = p(A; x)b$ for an appropriate polynomial $p(t; x)$. Such a technique is widely used in the QR algorithm and Krylov subspace methods [19, 8, 9, 20]. Theorem 4.1 shows that in order to obtain $K(A, d)$ expressed as $TK(A, b)$, this is the only way when A is nonderogatory.

For any b such that $K(A, b)$ is nonsingular, the corresponding $\mathbf{S}_b \in L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$ is an isomorphism of $\mathbb{P}_{n-1}(A)$ onto $\mathbb{K}(A)$ defined by

$$\mathbf{S}_b p(A; x) = K(A, d) = p(A; x)K(A, b), \quad d = K(A, b)x, \quad \forall x \in \mathbb{C}^n.$$

Its inverse is

$$\mathbf{S}_b^{-1}K(A, d) = p(A; x) = K(A, d)K(A, b)^{-1}, \quad x = K(A, b)^{-1}d, \quad \forall d \in \mathbb{C}^n, \tag{23}$$

which is just (7).

Using $L_c(\mathbb{K}(A))$ and the isomorphisms of $\mathbb{K}(A)$ and $L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$, we are also able to define a subspace of linear operators on $\mathbb{P}_{n-1}(A)$. Let $\mathbf{S}_{b_1}, \mathbf{S}_{b_2} \in L_c(\mathbb{P}_{n-1}(A), \mathbb{K}(A))$ be invertible. Define

$$L_c(\mathbb{P}_{n-1}(A)) = \left\{ \mathbf{W} \;\middle|\; \begin{array}{l} \mathbf{W}p(A; x) = \mathbf{S}_{b_2}^{-1}\mathbf{T}\mathbf{S}_{b_1}p(A; x) = p(A; y), \quad \mathbf{T} = p(A; z) \in L_c(\mathbb{K}(A)), \\ \quad\quad y = K(A, b_2)^{-1}p(A; z)K(A, b_1)x, \quad \forall x \in \mathbb{C}^n \end{array} \right\}$$

Clearly, $L_c(\mathbb{P}_{n-1}(A))$ is isomorphic to $L_c(\mathbb{K}(A)) = \mathbb{P}_{n-1}(A)$. So its dimension is n.

When $K(A, b)$ is nonsingular, by Proposition 2.4 we have a Hessenberg reduction form

$$Q^*AQ = H, \quad Q^*b = \gamma e_1, \quad |\gamma| = \|b\|, \tag{24}$$

where Q is unitary and H is unreduced upper Hessenberg.

The matrices Q and H can be computed by using the Arnoldi process.

Arnoldi process.

Input: Matrix A and vector b

Output: Unitary matrix $Q = [q_1, \ldots, q_n]$ and upper Hessenberg matrix $H = [h_{ij}]$

Choose $q_1 = b/\|b\|$.

For $k = 1, \ldots, n$

$$
\begin{aligned}
h_{jk} &= q_j^* A q_k, \quad j = 1, \ldots, k \\
h_{k+1,k} &= \|A q_k - h_{1k} q_1 - \ldots - h_{kk} q_k\| \\
q_{k+1} &= (A q_k - h_{1k} q_1 - \ldots - h_{kk} q_k)/h_{k+1,k}
\end{aligned}
$$

End

In practice, one uses the modified Arnoldi process [8, Sec. 9.4], or the numerically stable Hessenberg reduction method with Householder transformations [8, Sec. 7.4]. Note that with the above Arnoldi process $\gamma = \|b\|$ and all the subdiagonal elements of H are positive. With the Hessenberg reduction form (24), by Proposition 2.4, one has the QR factorization

$$
K(A, b) = QR, \tag{25}
$$

where $R = K(H, \gamma e_1)$ is nonsingular and upper triangular with $r_{kk} = \gamma \prod_{j=1}^{k-1} h_{j+1,j}$ for $k = 1, \ldots, n$. From (25),

$$
Q = K(A, b)R^{-1} =: \begin{bmatrix} g_1(A)b & g_2(A)b & \ldots & g_n(A)b \end{bmatrix}, \tag{26}
$$

and it is easily verified that $g_j(t) \in \mathbb{P}_{n-1}$ and $\deg g_j(t) = j - 1$ for $j = 1, \ldots, n$. So Q is a generalized Krylov matrix of the form (19). In fact the polynomials $g_i(t)$ has the following properties.

Theorem 4.3 *The polynomials $g_1(t), \ldots, g_n(t)$ satisfy*

$$
g_1(t) = \frac{1}{\gamma}, \quad g_j(t) = \frac{1}{\gamma \prod_{k=2}^{j} h_{k,k-1}} \det(tI - H_{j-1}), \quad j = 2, \ldots, n, \tag{27}
$$

where H_k is the leading principal $k \times k$ matrix of H given in (24).

Also,
$$
[g_1(t), g_2(t), \ldots, g_n(t)] = [1, t, \ldots, t^{n-1}] R^{-1}. \tag{28}
$$

Proof. From the Arnoldi process it is not difficult to get the recurrence

$$
g_1(t) = \frac{1}{\gamma}, \quad g_{j+1}(t) = \frac{1}{h_{j+1,j}}(t g_j(t) - h_{1j} g_1(t) - \ldots - h_{jj} g_j(t)), \quad j = 1, \ldots, n-1.
$$

We now prove (27) by induction. When $j = 2$,

$$
g_2(t) = \frac{1}{h_{21}}(t g_1(t) - h_{11} g_1(t)) = \frac{1}{\gamma h_{21}}(t - h_{11}) = \frac{1}{\gamma h_{21}} \det(tI_1 - H_1).
$$

So (27) holds for $j = 2$.

Assume (27) is true for $1, \ldots, j$. Expanding $\det(tI - H_j)$ based on the last column we get

$$
\begin{aligned}
\det(tI - H_j) \;=\;& (t - h_{jj}) \det(tI - H_{j-1}) - h_{j-1,j} h_{j,j-1} \det(tI - H_{j-2}) \\
& - h_{j-2,j} \left(\prod_{k=j-1}^{j} h_{k,k-1} \right) \det(tI - H_{j-3}) - \ldots - h_{1j} \left(\prod_{k=2}^{j} h_{k,k-1} \right).
\end{aligned}
$$

By dividing $\gamma \prod_{k=2}^{j+1} h_{k,k-1}$ on both side, and using the assumption we have

$$
\frac{1}{\gamma \prod_{k=2}^{j+1} h_{k,k-1}} \det(tI - H_j) = \frac{1}{h_{j+1,j}} ((t - h_{jj}) g_j(t) - h_{j-1,j} g_{j-1}(t) - \ldots - h_{1j} g_1(t)) = g_{j+1}(t).
$$

So (27) hold also for $j + 1$.

The relation (28) is simply form (26). □

Since $K(A, b)$ is nonsingular, one may introduce the following inner product in $\mathbb{P}_{n-1}(A)$.

$$
\langle p(A; x), p(A; y) \rangle_b = b^* p(A; x)^* p(A; y) b = x^* K(A, b)^* K(A, b) y, \quad \forall p(A; x), p(A; y) \in \mathbb{P}_{n-1}(A).
$$

The last relation is due to (1). With this inner product we define the norm

$$
\| p(A; x) \|_b = \langle p(A; x), p(A; x) \rangle_b^{1/2} = \| p(A; x) b \| = \| K(A, b) x \|.
$$

Then the matrices $g_1(A), \ldots, g_n(A)$ that determine Q in (26) are orthonormal, which can be viewed as being generated from $I, A, \ldots, A^{n-1} \in \mathbb{P}_{n-1}(A)$ by applying the Gram-Schmidt process with respect to the above defined inner product [2]. So $g_1(A), \ldots, g_n(A)$ form an orthonormal basis for $\mathbb{P}_{n-1}(A)$. Also, the polynomials $g_1(t), \ldots, g_n(t)$ form an orthonormal basis for $\mathbb{P}_{n-1}$ with respect to the inner product

$$
\langle p(t; x), p(t; y) \rangle_{A,b} := \langle p(A; x), p(A; y) \rangle_b, \quad \forall p(t; x), p(t; y) \in \mathbb{P}_{n-1},
$$

which can be interpreted as being generated by applying the Gram-Schmidt process to $1, t, \ldots, t^{n-1}$.

Because $\mathbb{K}(A)$ and $\mathbb{P}_{n-1}(A)$ are isomorphic, if $K(A, b)$ is nonsingular, using the isomorphism $\mathbf{S}_b^{-1}$ defined in (23), an inner product in $\mathbb{K}(A)$ can be induced from the inner product $\langle \cdot, \cdot \rangle_b$ with

$$
\begin{aligned}
\langle K(A, u), K(A, v) \rangle \;=\;& \left\langle \mathbf{S}_b^{-1} K(A, u), \mathbf{S}_b^{-1} K(A, v) \right\rangle_b = \left\langle p(A, K(A, b)^{-1} u), p(A, K(A, b)^{-1} v) \right\rangle_b \\
=\;& (K(A, b)^{-1} u)^* (K(A, b)^* K(A, b)) (K(A, b)^{-1} v) = u^* v,
\end{aligned}
$$

which is just the standard inner product in $\mathbb{C}^n$.

5   More properties related to Hessenberg reductions

Given a nonsingular matrix W, in a similar way one can determine a W-unitary matrix X, i.e., $X^* W^* W X = I$, such that

$$
X^{-1} A X = \widehat{H}, \quad X^{-1} b = \hat{\gamma} e_1, \tag{29}
$$

where $\widehat{H}$ is unreduced upper Hessenberg (by Proposition 2.4). The matrix X can be obtained by applying the Arnoldi process. The only difference is to make the columns of X to be W-orthonormal. By Proposition 2.4,

$$K(A,b) = X\widehat{R}, \qquad \widehat{R} = K(\widehat{H}, \hat{\gamma}e_1).$$

So we also have

$$X = K(A,b)\widehat{R}^{-1} = \begin{bmatrix} \hat{g}_1(A)b & \hat{g}_2(A)b & \dots & \hat{g}_n(A)b \end{bmatrix},$$

where $\hat{g}_j(t) \in \mathbb{P}_{n-1}$ with $\deg \hat{g}_j(t) = j - 1$, and $\hat{g}_1(A), \dots, \hat{g}_n(A)$ form an orthonormal basis for $\mathbb{P}_{n-1}(A)$ with the generalized inner product

$$\langle p(A;x), p(A;y)\rangle_{W,b} = b^* p(A;x)^* W^* W p(A;y)b = x^* K(A,b)^* W^* W K(A,b)y. \qquad (30)$$

Theorem 5.1 *Suppose $K(A,b)$ is nonsingular. Let W be nonsingular, X and $\widehat{H}$ satisfy (29), and Q, H satisfy (24). Define $\widehat{Q} = WX$ and $T = \widehat{R}R^{-1} = K(\widehat{H}, \hat{\gamma}e_1)K(H, \gamma e_1)^{-1}$. Then $\widehat{Q}$ is unitary, T is upper triangular, and*

$$
\begin{aligned}
T &= \hat{\gamma}[g_1(\widehat{H})e_1, \dots, g_n(\widehat{H})e_1] = \gamma^{-1}[\hat{g}_1(H)e_1, \dots, \hat{g}_n(H)e_1]^{-1}, \\
Q &= XT, \\
W &= \widehat{Q}TQ^*, \\
H &= T^{-1}\widehat{H}T.
\end{aligned}
$$

Proof. It is obvious that $\widehat{Q} = WX$ is unitary and T is upper triangular.

As (28), we have
$$[\hat{g}_1(t), \dots, \hat{g}_n(t)] = [1, t, \dots, t^{n-1}]\widehat{R}^{-1}.$$

Then

$$T = \widehat{R}R^{-1} = \hat{\gamma}[g_1(\widehat{H})e_1, \dots, g_n(\widehat{H})e_1] = \gamma^{-1}[\hat{g}_1(H)e_1, \dots, \hat{g}_n(H)e_1]^{-1} = (R\widehat{R}^{-1})^{-1},$$

which is upper triangular, and from

$$[g_1(t), \dots, g_n(t)] = [1, t, \dots, t^{n-1}]R^{-1} = [1, t, \dots, t^{n-1}]\widehat{R}^{-1}(\widehat{R}R^{-1}) = [\hat{g}_1(t), \dots, \hat{g}_n(t)]T,$$

we have

$$Q = XT.$$

Then

$$W = \widehat{Q}X^{-1} = \widehat{Q}TQ^*,$$

and from (24) and (29) we have $H = T^{-1}\widehat{H}T$. $\square$

This theorem shows the relations between the Hessenberg reduction forms (24) and (29). In fact, with $\widehat{A} = WAW^{-1}$ and $\hat{b} = Wb$, (29) can be rewritten as

$$\widehat{Q}^*\widehat{A}\widehat{Q} = \widehat{H}, \quad \widehat{Q}^*\hat{b} = \hat{\gamma}e_1, \quad |\hat{\gamma}| = \|\hat{b}\|.$$

So (29) is the same as (24) but with Q, A, b replaced by $\widehat{Q}, \widehat{A}, \hat{b}$.

The next result shows that for any sequence of n polynomials in $\mathbb{P}_{n-1}$ with degrees in increasing order, a unitary matrix can be constructed to reduces WAW^{-1} and Wb to a Hessenberg reduction form for an appropriate W.

Theorem 5.2 *Suppose that $K(A, b)$ is nonsingular and $p(t; r_1), \ldots, p(t; r_n) \in \mathbb{P}_{n-1}$ with $\deg p(t; r_j) = j - 1$ for $j = 1, \ldots, n$. There exists a nonsingular matrix W such that for*

$$\widehat{A} = WAW^{-1}, \quad \hat{b} = Wb,$$

the matrix

$$\widehat{Q} = [p(\widehat{A}; r_1)\hat{b}, \ldots, p(\widehat{A}; r_n)\hat{b}]$$

is unitary and satisfies

$$\widehat{Q}^* \widehat{A} \widehat{Q} = \widehat{H}, \quad \widehat{Q}^* \hat{b} = \hat{\gamma} e_1, \quad |\hat{\gamma}| = \|\hat{b}\|,$$

where $\widehat{H}$ is unreduced upper Hessenberg.

Proof. Let

$$R = [r_1, \ldots, r_n], \quad X = [p(A; r_1)b, \ldots, p(A; r_n)b].$$

By the assumptions R is upper triangular and nonsingular, and $X = K(A, b)R$ is nonsingular. From $AK(A, b) = K(A, b)C$, where C is the companion matrix of A, we have

$$X^{-1}AX = R^{-1}CR =: \widehat{H}, \qquad X^{-1}b = \hat{\gamma} e_1,$$

where $\widehat{H}$ is unreduced upper Hessenberg and $\hat{\gamma}$ is a scalar.

Let

$$X = W^{-1}\widehat{Q},$$

where W is a nonsingular matrix and $\widehat{Q}$ is unitary. Such a factorization always exists, for instance, an RQ factorization. Then, with this W and the corresponding $\widehat{A}$, $\hat{b}$, we have

$$\widehat{Q} = WX = W[p(A; r_1)b, \ldots, p(A; r_n)b] = [p(\widehat{A}; r_1)\hat{b}, \ldots, p(\widehat{A}; r_n)\hat{b}],$$

$$\widehat{Q}^* \widehat{A} \widehat{Q} = X^{-1}W^{-1}(WAW^{-1})WX = X^{-1}AX = \widehat{H},$$

and

$$\widehat{Q}^* \hat{b} = X^{-1}W^{-1}Wb = X^{-1}b = \hat{\gamma} e_1.$$

Obviously, $|\hat{\gamma}| = \|\widehat{Q}^* \hat{b}\| = \|\hat{b}\|$. □

We now use the Hessenberg reduction form (24) to give a factorization for $f(A)$.

Theorem 5.3 *Suppose that $K(A, b)$ is nonsingular and A, b have the forms in (24). Then for any $f(t)$ such that $f(A)$ is defined,*

$$f(A) = QK(H, \tilde{d})K(H, e_1)^{-1}Q^*, \quad \tilde{d} = f(H)e_1. \tag{31}$$

Moreover, let $\widetilde{Q} = [\widetilde{Q}_1, \widetilde{Q}_2]$ be unitary with $\widetilde{Q}_1 \in \mathbb{C}^{n,r}$ such that

$$H\widetilde{Q} = \widetilde{Q}\widetilde{H}, \quad \widetilde{H} = \begin{bmatrix} \widetilde{H}_{11} & \widetilde{H}_{12} \\ 0 & \widetilde{H}_{22} \end{bmatrix}, \qquad \widetilde{Q}^* \tilde{d} = \tilde{\gamma} e_1, \quad |\tilde{\gamma}| = \|\tilde{d}\|,$$

where $\widetilde{H}_{11} \in \mathbb{C}^{r,r}$ is an $r \times r$ unreduced upper Hessenberg matrix. Then $\operatorname{rank} f(A) = r$ and

$$f(A) = (Q\widetilde{Q}_1)\widetilde{R}Q^*, \tag{32}$$

where

$$\widetilde{R} = \tilde{\gamma} K_{r,n}(\widetilde{H}_{11}, e_1)K(H, e_1)^{-1},$$

is upper triangular.

Proof. Since $A = QHQ^*$, we have $f(A) = Qf(H)Q^*$. Applying (14) to $f(H)$ with $A = H$ and $b = \gamma e_1$,

$$f(H) = K(H, \tilde{d})K(H, e_1)^{-1}, \quad \tilde{d} = f(H)e_1.$$

So we have (31).

The factorization (32) follows by applying Proposition 2.4 to $K(H, \tilde{d})$ in (31). $\square$

Corollary 5.4 *Suppose $K(A, b)$ is nonsingular. If* rank $f(A) = r < n$, *then all the eigenvalues of $\widetilde{H}_{22}$ are the roots of $f(t) = 0$ (counting multiplicity), and* range $Q\widetilde{Q}_2 = \mathrm{null}([f(A)]^*)$.

Proof. From (32),

$$(Q\widetilde{Q})^* f(A)(Q\widetilde{Q}) = \left[\begin{array}{c} \widetilde{R} \\ 0 \end{array}\right] \widetilde{Q}.$$

On the other hand, by using the Hessenberg reduction forms we have

$$(Q\widetilde{Q})^* f(A)(Q\widetilde{Q}) = f(\widetilde{H}) = \left[\begin{array}{cc} f(\widetilde{H}_{11}) & * \\ 0 & f(\widetilde{H}_{22}) \end{array}\right].$$

So we have $f(\widetilde{H}_{22}) = 0$, and $(Q\widetilde{Q}_2)^* f(A) = 0$. Clearly, all the eigenvalues of $\widetilde{H}_{22}$ are the roots of $f(t)$. Since rank $f(A) = r$, we have range $Q\widetilde{Q}_2 = \mathrm{null}([f(A)]^*)$. $\square$

When $f(A)$ is nonsingular, (32) can be also derived by the orthogonalization argument using the weighted inner product (30) with $W = f(A)$. In this case, (WAW^{-1}, Wb) becomes (A, d). From Theorem 5.3, the matrices and scalar in (29) are

$$\widehat{Q} = Q\widetilde{Q}, \quad \widehat{H} = \widetilde{H}, \quad \hat{\gamma} = \gamma\tilde{\gamma}.$$

By Theorem 5.1, $f(A)$ has the URV decomposition ([21]),

$$f(A) = Q\widetilde{Q}TQ^*,$$

where

$$T = K(\widetilde{H}, \gamma\tilde{\gamma}e_1)K(H, \gamma e_1)^{-1} = \tilde{\gamma}K(\widetilde{H}, e_1)K(H, e_1)^{-1} = \widetilde{R}.$$

The formula (32) is more generalized, since it holds when $f(A)$ is singular as well.

Remark 5.5 Using $Af(A) = f(A)A$, the matrix $\widetilde{R}$ in (32) satisfies $\widetilde{H}_{11}\widetilde{R} = \widetilde{R}H$. So $\widetilde{R}$ can be computed column by column with the recurrence

$$\tilde{r}_1 = \tilde{\gamma}e_1, \quad \tilde{r}_{k+1} = (\widetilde{H}_{11}\tilde{r}_k - h_{1k}\tilde{r}_1 - \ldots - h_{kk}\tilde{r}_k)/h_{k+1,k}, \quad k = 1, \ldots, n-1,$$

which is the Arnoldi process for q_k, but with A replaced by $\widetilde{H}_{11}$.

More generally, using

$$Hf(H) = f(H)H,$$

one may use the same recurrence to compute $f(H)$, provided $f(H)e_1$ is given. This approach was mentioned in [14, 15] for $f(t) = e^t$.

The next result shows how the unitary matrix is related to Q if it is generated by another vector.

Theorem 5.6 *Suppose that $K(A,b)$ is nonsingular and A, b have the forms in (24). Let $d \in \mathbb{C}^n$ and $\widehat{Q} = [\widehat{Q}_1, \widehat{Q}_2]$ be unitary with $\widehat{Q}_1 \in \mathbb{C}^{n,r}$ and satisfy*

$$A\widehat{Q} = \widehat{Q}\widehat{H}, \quad \widehat{H} = \begin{bmatrix} \widehat{H}_{11} & \widehat{H}_{12} \\ 0 & \widehat{H}_{22} \end{bmatrix}, \qquad \widehat{Q}^* d = \hat{\gamma} e_1, \quad |\hat{\gamma}| = \|d\|,$$

where $\widehat{H}_{11} \in \mathbb{C}^{r,r}$ is unreduced upper Hessenberg. Let

$$Q = [Q_1, Q_2], \quad H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad Q_1 \in \mathbb{C}^{n,r}, \quad H_{11} \in \mathbb{C}^{r,r}.$$

Then

$$\widehat{Q}_1 K_{r,n}(\widehat{H}_{11}, \hat{\gamma} e_1) = p(A; x) Q K(H, \gamma e_1), \tag{33}$$

and

$$\widehat{Q}_1 T_r = p(A; x) Q_1,$$

where

$$x = K(A, b)^{-1} d, \qquad T_r = K_r(\widehat{H}_{11}, \hat{\gamma} e_1) K_r(H_{11}, \gamma e_1)^{-1}.$$

Proof. Because $x = K(A, b)^{-1} d$, by (6),

$$K(A, d) = p(A; x) K(A, b).$$

By the Hessenberg reduction forms we have

$$K(A, d) = \widehat{Q}_1 K_{r,n}(\widehat{H}_{11}, \hat{\gamma} e_1), \quad K(A, b) = Q K(H, \gamma e_1).$$

So we have (33). The second equation follows by equating the first r columns in (33) to get

$$\widehat{Q}_1 K_r(\widehat{H}_{11}, \hat{\gamma} e_1) = p(A; x) Q_1 K_r(H_{11}, \gamma e_1),$$

and by using the fact that $K_r(H_{11}, \gamma e_1)$ is nonsingular. $\square$

The relation (33) shows that the unitary matrix corresponding to d in the Hessenberg reduction is just the unitary factor of the QR factorization of $p(A; x)Q$ for an appropriate polynomial $p(t; x)$.

Any nonsingular Krylov matrix $K(A, b)$ has a QR factorization (25) with R nonsingular. If $Q = I$, from (24) A has to be unreduced upper Hessenberg and $b = \gamma e_1$. In this case (32) becomes a QR factorization of $f(A)$. If further $K(A, b) = I$, then $b = e_1$ and A is the companion matrix C defined in (4). In this case, $f(A)$ and $K(A, b)$ have simpler relations.

Corollary 5.7 *Suppose $C \in \mathbb{C}^{n,n}$ is a companion matrix defined in (4). Then*

$$p(C; d) = K(C, d), \quad \forall d \in \mathbb{C}^n.$$

Proof. It is simply from Theorem 3.1 with $A = C$ and $b = e_1$, and $K(C, e_1) = I_n$. $\square$

If we choose d as $c = [c_1, \ldots, c_n]^T$, the last column of C, then by the Cayley-Hamilton Theorem,

$$C^n = p(C; c) = K(C, c).$$

Corollary 5.8 *Suppose C is the companion matrix (4) and $f(t)$ is a scalar function such that $f(C)$ is defined. Then*

$$f(C) = K(C, d) = p(C; d), \quad d = f(C)e_1.$$

Proof. It follows from Theorem 3.3 and Corollary 5.7. □

If in (25), $R = I$ or equivalently $K(A, b) = Q$, then $Q^* K(A, b) = K(Q^* A Q, Q^* b) = I$. This implies $b = Q e_1$ and $A = Q C Q^*$. In this case, we have a simple formula $f(A) = K(A, d)Q^*$. Note that this is one of the situations where the GMRES stagnates when it is used to solve associated linear systems or eigenvalue problems, [9, Sec. 3.2]. Note also that when $W = K(A, b)^{-1}$ in the inner product (30), (WAW^{-1}, Wb) becomes (C, e_1). So for any A, b such that $K(A, b)$ is nonsingular, the above situation occurs (with $Q = I$) if we consider the orthonormal polynomials with respect to $\langle \cdot, \cdot \rangle_{K(A,b)^{-1}, b}$ in $\mathbb{P}_{n-1}$.

6 Conclusions

Starting from a simple observation, we derived formulas to show relations between functions of a matrix and Krylov matrices. By introducing subspaces and linear transformations, we interpreted the relations at an abstract level. We provided several properties of Hessenberg reductions that can be used to understand some common techniques used in Krylov subspace methods and eigenvalue algorithms. How to use the results to improve existing methods and develop new methods? That needs more work.

Acknowledgment

References

[1] M. Arioli, V. Pták, and Z. Strakoš. Krylov sequences of maximal length and convergence of GMRES. *BIT*, 38(4):636–643, 1998.

[2] D. Calvetti, S.-M. Kim, and L. Reichel. Quadrature rules based on the arnoldi process. *SIAM J. Matrix Anal. Appl.*, 26(3):765–781, 2005.

[3] P. I. Davies and N. J. Higham. A Schur-Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[4] R. Eising. The distance between a systems and the set of uncontrollable systems. *Proc. MTNS,* 303–314, Beer-sheva, June 1983.

[5] R. Eising. Between controllable and uncontrollable. *Systems Control Lett.*, 4(5):263–264, 1984.

[6] F. R. Gantmacher. *Theory of Matrices. Vol. 1.* Chelsea, New York, 1959.

[7] F. R. Gantmacher. *Theory of Matrices. Vol. 2.* Chelsea, New York, 1959.

[8] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[9] A. Greenbaum. *Iterative Methods for Solving Linear Systems.* SIAM Publisher, Philadelphia, PA, 1997.

[10] N. J. Higham. *Functions of Matrices, Theory and Computation.* SIAM Publications, Philadelphia, PA, 2008.

[11] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1991.

[12] B. Kågström. Numerical computation of matrix functions. Report UMIN-58.77, Department of Information processing, University of Umeå, Sweden, 1977.

[13] R. E. Kalman. On the general theory of control systems. In *Proc. 1st IFAC Congr.*, 1:481–491, London, Butterworth, 1960.

[14] C. B. Moler and C. F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.*, 29:801–837, 1978.

[15] C. B. Moler and C. F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.*, 45:3–49, 2003.

[16] B. N. Parlett. Computation of functions of triangular matrices. Memorandum ERL-M481, Electronics Research Laboratory, College of Engineering, UC Berkeley, 1974.

[17] B. N. Parlett. A recurrence among the elements of functions of triangular matrices. *Linear Algebra Appl.*, 14:117–121, 1976.

[18] H. H. Rosenbrock. *State-Space and Multivariable Theory.* London, Nelson, 1970.

[19] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* Manchester University Press, Manchester, UK, 1992.

[20] Y. Saad. *Iterative Methods for Sparse Linear Systems (2nd edition).* SIAM Publications, Philadelphia, PA, 2004.

[21] G. W. Stewart. Updating a rank-revealing URV decompositions in parallel. *Parallel Computing*, 20:151–172, 1994.

[22] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford University Press, Oxford, 1965.