# LAPACK working note 66
# A Characterization of Polynomial Iterative Methods*

**Victor Eijkhout**
**University of Tennessee**
**Department of Computer Science**
**107 Ayres Hall, Knoxville, TN 37996-1301**
`eijkhout@cs.utk.edu`

**February 10, 1994**

**Abstract**

Polynomial iterative methods, such as methods of the conjugate gradient type, involve a starting vector, a right hand side vector, a coefficient matrix, possibly a preconditioning matrix, and for methods based on conjugacy, an inner product. In this paper, we give a rigorous definition of vector sequences that are generated by polynomial methods, and we characterize those methods in terms of the above-mentioned elements.

## 1 Introduction

We start by defining polynomial iterative methods.

**Definition 1** *A polynomial iterative method is a sequence of vectors* $\{x_i\}_{i\geq 1}$*, denoted by a 4-tuple* $\langle\{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$ *where* $x_1$*,* $f$ *are vectors in* $R^n$*,* $A$ *is an* $n\times n$ *matrix, and the* $\pi_i$ *are polynomials with* $\deg(\pi_i) = i \Leftrightarrow 1$*; the sequence is defined by*

$$x_{i+1} \Leftrightarrow x_1 = \pi_i(A)\{Ax_1 \Leftrightarrow f\}. \tag{1}$$

Next we will define polynomial sequences independent of the particular choices for $A$, $\{\pi_i\}$, and $f$, but only dependent on a solution vector.

*. This work was supported by DARPA under contract number DAAL03-91-C-0047

**Definition 2** *A sequence $\{x_i\}_{i\geq 1}$ is called a polynomial sequence for the vector $\bar{x}$ if it is a polynomial method $\langle\{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$, and $A$ and $f$ are such that $A\bar{x} = f$. The vector $\bar{x}$ is called the solution vector of the sequence.*

Such methods can be motivated informally from the following observations. First of all

$$r_1 = Ax_1 \Leftrightarrow f \Rightarrow \bar{x} = A^{-1}f = x_1 \Leftrightarrow A^{-1}r_1.$$

Then, there is a polynomial $\phi$ such that $\phi(A) = 0$, and without loss of generality we can write $\phi(x) = 1 + x\pi(x)$ with $\pi$ an inhomogenous polynomial. Then

$$A^{-1} = \Leftrightarrow \pi(A) \qquad \text{so} \qquad \bar{x} \Leftrightarrow x_1 = \pi(A)r_1.$$

Polynomial iterative methods then construct subsequent polynomials than in some sense approximate this polynomial $\pi$.

**Lemma 1** *If $X$ is a polynomial sequence for $\bar{x}$ and $B$ is an invertible matrix, then $BX$ is a polynomial sequence for $B\bar{x}$, specifically, if $X$ is $\langle\{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$ then $BX$ is $\langle\{\pi_i\}_{i\geq 1}, BAB^{-1}, Bx_1, Bf\rangle$.*

Proof. This follows from

$$Bx_{i+1} \Leftrightarrow Bx_1 = \pi_i(BAB^{-1})\{(BAB^{-1})Bx_1 \Leftrightarrow Bf\}.$$

•

## 2 Tools

In this section we will develop some tools that will facilitate further presentation and analysis.

First of all, we will often abbreviate vector sequences as a matrix:

$$X = (x_1, x_2, \ldots).$$

Next we introduce the 'left-shift' operator $J$ for sequences:

$$J = (\delta_{i,j+1}) = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & 1 & 0 & \\ & & \ddots & \ddots \end{pmatrix}$$

so that for sequences $X$ and $Y$ the statement $Y = XJ$ implies $y_i = x_{i+1}$. Also, Krylov sequences $y_{i+1} = Ay_i$ can conveniently be denoted as $AY = YJ$. Furthermore, we introduce the matrix

$$E_1 = \begin{pmatrix} 1 & \dots \\ 0 & \dots \\ \vdots & \end{pmatrix}$$

which picks the first element of a sequence: if $Y = XE_1$ then $y_i = x_1$ for all $i$.

The matrices $J$ and $E_1$ are convenient in talking about updating a sequence:

$$Y = X(J \Leftrightarrow I) \qquad \Leftrightarrow \qquad y_i = x_{i+1} \Leftrightarrow x_i$$

and

$$Y = X(J \Leftrightarrow E_1) \qquad \Leftrightarrow \qquad y_i = x_{i+1} \Leftrightarrow x_1.$$

The relation between $J \Leftrightarrow I$ and $J \Leftrightarrow E_1$ is as follows:

$$\begin{aligned} J \Leftrightarrow E_1 &= (J \Leftrightarrow I)(I \Leftrightarrow J^t)^{-1} \\ J \Leftrightarrow I &= (J \Leftrightarrow E_1)(I \Leftrightarrow J^t) \end{aligned}$$

The following auxiliary lemma shows that constructing a sequence by

$$x_{i+1} \Leftrightarrow x_1 = \sum_{j \leq i} k_j c_{ji}$$

is equivalent to updating it as

$$x_{i+1} \Leftrightarrow x_i = \sum_{j \leq i} k_j \tilde{c}_{ji}.$$

**Lemma 2** *If $X$ and $K$ are sequences, $U$ is upper triangular, then*

$$X(J \Leftrightarrow I) = KU \qquad \textit{iff} \qquad X(J \Leftrightarrow E_1) = KV$$

*for some upper triangular matrix $V$.*

Proof: Choose $V = U(I \Leftrightarrow J^t)$. •

The right hand side in (1) can be described differently in terms of a Krylov sequence.

**Lemma 3** *A sequence $Y$ is generated by applying successive polynomials to an initial vector $k_1$ as*

$$y_i = \pi_i(A)k_1; \qquad \text{degree}(\pi_i) = i \Leftrightarrow 1,$$

*iff there is an upper triangular matrix $U$ such that*

$$Y = KU$$

*where $K$ is the Krylov sequence $k_{i+1} = Ak_i$. The polynomials $\pi_i$ have coefficients in the $i$-th column of $U$; specifically,*

$$\pi_i(x) = u_{ii}x^{i-1} + \cdots + u_{2i}x + u_{1i}.$$

Proof. See [1]. •

Occasionally we will use the vector $e = (1, \ldots)^t$; for instance, we can denote residuals $r_i = Ax_i \Leftrightarrow f$ as a sequence by $R = AX \Leftrightarrow fe^t$.

The subject of Hessenberg matrices also comes up in the discussion of polynomial iterative methods. The following auxiliary lemma states the connection between Hessenberg matrices and Krylov sequences.

**Lemma 4** *If $AR = RH$ and $r_1 \parallel k_1$, then $H$ is an irreducible upper Hessenberg matrix iff there is a nonsingular upper triangular matrix $U$ such that $R = KU$, with $K$ the Krylov sequence satisfying $AK = KJ$; $U$ and $H$ are related by $H = U^{-1}JU$.*

Proof. See [1]. •

We will have occasion to use the following lemmas characterizing Hessenberg matrices.

**Lemma 5** *Let $U$ be a non-singular upper triangular matrix and $H = U^{-1}JU$. Then the first row of $U$ is constant iff $H$ has zero column sums.*

Proof. With the zero vector and the all-ones vector $e$ we can formulate the zero column sums as $e^tH = 0^t$. Then

$$\begin{aligned} e^tH = e^tU^{-1}JU = 0^t &\Leftrightarrow e^tU^{-1}J = 0^t \\ &\Leftrightarrow e^tU^{-1} = (\alpha, 0, 0, \ldots) \qquad \text{some nonzero } \alpha \\ &\Leftrightarrow \alpha^{-1}e^t = (1, 0, 0, \ldots)U \end{aligned}$$

which proves the statement. •

**Lemma 6** *Let $H$ be a Hessenberg matrix that allows factorization without pivoting to $H = (I \Leftrightarrow L)U$ form where $L$ contains a single nonzero lower subdiagonal. Then the column sums of $H$ are zero iff $L = J$.*

Proof. Since the diagonal elements of $U$ are nonzero, we have

$$x^tU = 0^t \Leftrightarrow x^t = 0^t.$$

Expressing the zero column sums of $H$ as $e^tH = 0^t$, we then find

$$e^t H = 0^t \Leftrightarrow e^t(I \Leftrightarrow L) = 0^t \Leftrightarrow L = J$$

which concludes the proof. •

## 3 Characterization

Using the results of the previous section, we can now give some equivalent definitions of polynomial iterative methods.

**Lemma 7** *Let $X$ be a sequence, and let the matrix $A$ and the vector $f$ be given. Define $k_1 = Ax_1 \Leftrightarrow f$ and let $K$ be the Krylov sequence $AK = KJ$. Then the following statements are equivalent.*

- *There are polynomials $\{\pi_i\}_{i\geq 1}$ such that*
  $X(J \Leftrightarrow I) = (\pi_1(A)k_1, \pi_2(A)k_1, \ldots)$.
- *There are polynomials $\{\pi_i\}_{i\geq 1}$ such that*
  $X(J \Leftrightarrow E_1) = (\pi_1(A)k_1, \pi_2(A)k_1, \ldots)$.
- *There is an upper triangular matrix $U$ such that*
  $X(J \Leftrightarrow I) = KU$.
- *There is an upper triangular matrix $U$ such that*
  $X(J \Leftrightarrow E_1) = KU$.

The above lemma states that polynomial iterative methods use combinations of a Krylov sequence for updating. The following lemma shows that the residuals of the iterative method are then themselves combinations of this Krylov sequence; there is a normalization condition on these combinations.

**Lemma 8** *Let a matrix $A$ a vector $f$ and a sequence $X$ be given. Let $R$ be the sequence of residuals $R = AX \Leftrightarrow fe^t$, and let $K$ be the Krylov sequence satisfying $AK = KJ$, $k_1 = r_1$[1]. Then there are polynomials $\{\pi_i\}_{i\geq 1}$ such that $X$ is generated by a polynomial iterative method $\langle\{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$, iff there is an upper triangular matrix $\tilde{U}$ such that $R = K\tilde{U}$, with $\tilde{u}_{1j} \equiv 1$.*

Proof. Suppose $X$ is generated by a polynomial iterative method $\langle\{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$, and by lemma 7 write the generating equation as $X(J \Leftrightarrow I) = KU$ with $U$ an upper triangular matrix. From $R(J \Leftrightarrow I) = AKU = KJU$ it follows that $R(J \Leftrightarrow I)J^t = K\bar{U}$ where $\bar{u}_{i+1j+1} = u_{ij}$ and $\bar{u}_{1j} = \bar{u}_{i1} = 0$. Since $r_1 = k_1$, we can extend $\bar{U}$ to $\hat{U}$ and $(J \Leftrightarrow I)J^t$ to $(I \Leftrightarrow J^t)$ by putting a 1 in the $(1,1)$ position. This gives $R(I \Leftrightarrow J^t) = K\hat{U}$, or $R = K\tilde{U}$ with $\tilde{U} = \hat{U}(I \Leftrightarrow J^t)^{-1}$. It is easy to see that $\tilde{U}$ satisfies $\tilde{u}_{1j} \equiv 1$.

1. There is no loss of generality over assuming $k_1 \| r_1$.

Inspection of this proof shows that all implications can be reversed. $\bullet$

Combining this lemma and lemma 3 we find that there are polynomials associated with the sequence $R$ of residuals, and the polynomials are normalized at zero.

**Corollary 9** *Let $A$, $f$, $X$, $R$ be given as in the previous lemma. Then $X$ is generated by a polynomial iterative method $\langle \{\pi_i\}_{i\geq 1}, A, x_1, f\rangle$ iff there are polynomials $\{\tilde{\pi}_i\}_{i\geq 1}$ satisfying*

$$r_i = \tilde{\pi}_i(A)r_1, \qquad \deg(\tilde{\pi}_i) = i \Leftrightarrow 1, \qquad \tilde{\pi}_i(0) = 1.$$

*These polynomials are called the* residual polynomials*.*

Proof. Lemma 8 states that the residuals are combinations of the Krylov sequence, and it states the condition on the first row of the triangular matrix describing the combinations. Use lemma 3 to translate this to polynomial terms. $\bullet$

We shall now characterize all polynomial iterative methods for a given solution vector, by relating them to the residuals with respect to a given system.

**Theorem 1** *Let a vector $\bar{x}$ and a sequence $X$ be given, and let $A$ and $f$ be such that $A\bar{x} = f$. Define residuals by $R = AX \Leftrightarrow fe^t$, then the following statements are equivalent:*

1. *The sequence $X$ is a polynomial method for $\bar{x}$.*
2. *There are a nonsingular matrix $M$ and Hessenberg matrix $H$ with zero column sums such that*
   $$AMR = RH \quad \text{that is,} \quad AMr_i = h_{i+1i}r_{i+1} + \cdots + h_{1i}r_1.$$
3. *There are a nonsingular matrix $M$ and upper triangular matrix $U$ such that*
   $$X(J \Leftrightarrow I) = MRU, \quad \text{that is,} \quad x_{i+1} \Leftrightarrow x_i = \sum_{j\leq i} Mr_j u_{ji}.$$
4. *There are a sequence $P$ (the 'search directions'), a nonsingular matrix $M$, a diagonal matrix $D$, and a normalized upper triangular matrix $U$ such that*
   $$APD = R(I \Leftrightarrow J) \quad \text{that is,} \quad r_{i+1} = r_i \Leftrightarrow Ap_i d_{ii},$$
   and
   $$PU = MR \quad \text{that is,} \quad p_i = Mr_i \Leftrightarrow \sum_{j<i} p_j u_{ji}.$$
5. *There are a nonsingular matrix $M$ and polynomials $\{\pi_i\}_{i\geq 1}$ such that*
   $$r_i = \pi_i(AM)r_1, \qquad \deg(\pi_i) = i \Leftrightarrow 1, \qquad \pi_i(0) = 1.$$

Proof. Let $X$ be the specific polynomial method $\langle \{\pi_i\}_{i\geq 1}, B, x_1, g\rangle$ with $B\bar{x} = g$. Since $A\bar{x} = f$, there is a matrix $M$ such that $B = MA$ and $g = Mf$. If $\{\pi_i\}_{i\geq 1}$ is the sequence of polynomials of the method, then

$$\begin{aligned} x_{i+1} \Leftrightarrow x_i &= \pi_i(B)(Bx_1 \Leftrightarrow g) = \pi_i(MA)(MAx_1 \Leftrightarrow Mf) \\ &= M\pi_i(AM)(Ax_1 \Leftrightarrow f) \\ \Leftrightarrow M^{-1}(x_{i+1} \Leftrightarrow x_i) &= \pi_i(AM)(Ax_1 \Leftrightarrow f) \\ &= \pi_i(AM)\big((AM)(M^{-1}x_1) \Leftrightarrow f\big) \end{aligned}$$

Equivalently, we find from lemma 3

$$M^{-1}X(J \Leftrightarrow I) = K\bar{U} \tag{2}$$

with $\bar{U}$ upper triangular, and $K$ the Krylov sequence satisfying

$$AMK = KJ, \qquad k_1 = Ax_1 \Leftrightarrow f.$$

From $R = AX \Leftrightarrow fe^t = (AM)(M^{-1}X) \Leftrightarrow fe^t$ this is equivalent (see lemma 8) to the fact that $R = K\tilde{U}$ for some upper triangular matrix $\tilde{U}$ with first row identical 1. By lemma 5 this is equivalent to $AMR = RH$ with $H$ a Hessenberg matrix with zero column sums.[2] It also follows that $X(J \Leftrightarrow I) = MRU$ with $U = \tilde{U}^{-1}\bar{U}$ upper triangular.

From lemma 6 we know that $H$ can be factored as $(I \Leftrightarrow J)DU$ with $D$ diagonal and $U$ normalized. Introducing $P = MRU^{-1}$ gives the equivalence of 3 and 4.

For the proof $3 \Rightarrow 1$, note that

$$X(J \Leftrightarrow I) = MRU \Rightarrow RH = AMR$$

with $H = (J \Leftrightarrow I)U^{-1}$ an upper Hessenberg matrix. It then follows from lemma 4 that $R = K\tilde{U}$ with the Krylov sequence $K$ as above. Hence

$$M^{-1}X(J \Leftrightarrow I) = K\bar{U}$$

with $\bar{U} = \tilde{U}U$. We can finish now the proof by following the equivalences starting with equation (2) in reverse.

In order to show the equivalence of 1 and 5, note that by corollary 9, 1 is equivalent to the existence of polynomials $\{\pi_i\}_{i\geq 1}$ (with degree and normalization as indicated) such that the residuals $MR = BX \Leftrightarrow ge^t$ satisfy

$$\begin{aligned} Mr_i &= \pi_i(B)(Bx_1 \Leftrightarrow g) \\ &= \pi_i(MA)M(Ax_1 \Leftrightarrow f) \\ \Leftrightarrow r_i &= \pi_i(AM)(Ax_1 \Leftrightarrow f). \end{aligned}$$

•

2. Actually, the lemma only implies that the first row of $\tilde{U}$ is constant; it is identical 1 since $r_1 = k_1$.

## 4 Left and right preconditioning

The proof of theorem 1 noted that all systems with $A^{-1}f$ as solution can be written as $MAx = Mf$. The matrix $M$ is commonly called the 'left-preconditioner'.

A right preconditioner can be employed as follows. If $\bar{x} = A^{-1}f$, then a polynomial method $X = \langle\{\pi_i\}_{i\geq 1}, AN, x_1, f\rangle$ is a method for $N^{-1}x$, and we need to transform the iterates – which was not necessary in the case of a left preconditioner – to obtain a method for the original system.

Specifically, we are interested in the sequence $NX = \{Nx_i\}_{i\geq 1}$. From lemma 1 we already know that this is again a polynomial method, so by theorem 1 above it can be characterized by a single *left* preconditioner, but we will derive this fact in a second way.

Consider any polynomial method for $N^{-1}A^{-1}f$, and let residuals be defined by $R = ANX \Leftrightarrow fe^t$. By theorem 1 above, we can compute iterates, residuals, and search directions by

$$X(I \Leftrightarrow J) = PD, \qquad ANPD = R(I \Leftrightarrow J), \qquad MR = P(I \Leftrightarrow U).$$

Since $X$ is a method for $N^{-1}\bar{x}$, we introduce the sequences

$$\tilde{X} = NX, \qquad \tilde{P} = NP,$$

with which we get the method

$$\tilde{X}(I \Leftrightarrow J) = PD, \qquad A\tilde{P}D = R(I \Leftrightarrow J), \qquad NMR = \tilde{P}(I \Leftrightarrow U).$$

We see that the right preconditioner is simply aborbed as part of the total preconditioner $NM$. Note also that $R = ANX \Leftrightarrow fe^t = A\tilde{X} \Leftrightarrow fe^t$, that is, the residuals of the right preconditioned method are also the residuals of the sequence for $\bar{x}$.

## 5 Inner product

It remains to describe the role of the polynomials. From lemma 8 it is clear that we can equivalently talk about the upper triangular matrix $U$ for which $R = KU$, where $K$ is the Krylov sequence generated. There are no a priori restrictions on the matrix $U$, but for methods based on conjugacy it is equivalent to choosing an inner product for orthogonalizing the residuals.

**Lemma 10** *If $N$ is a symmetric nonsingular matrix, it is possible to construct $U$ such that $R^tNR$ is diagonal, where the sequence $R$ is constructed from $R = KU$.*

Proof. Let $R_n$, $U_n$ be the initial $n$ columns of $R$, $U$. Suppose inductively that $R_n^tNR_n$ is diagonal. In order to let $r_{n+1}$ be $N$-orthogonal to $R_n$, we need to solve the $n+1$-st column, $u_{n+1}$, of $U$ from the overdetermined system

$$U_n^t \bar{N} u_{n+1} = \left.\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}\right\} n$$

where $\bar{N}$ is the $n \times n+1$ primary subblock of $N$. This determines $u_{n+1}$ up to scaling. We scale it so that $u_{1,n+1} = 1$, so the system to be solved now becomes

$$\bar{N} \begin{pmatrix} 0 \\ u_{2n+1} \\ \vdots \\ u_{n+1n+1} \end{pmatrix} = \bar{N} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Now it follows that $R_n^t N r_{n+1} = 0$, and by symmetry $R_{n+1}^t N R_{n+1}$ is diagonal.
•

## References

[1] Victor Eijkhout. Lapack working note 51: Qualitative properties of the conjugate gradient and lanczos methods in a matrix framework. Technical Report CS 92-170, Computer Science Department, University of Tennessee, 1992.