Chapter 6

# Is There a Short Recurrence for a Near-Optimal Approximation?

Of the many non-Hermitian iterative methods described in the previous chapter, none can be shown to generate a near-optimal approximate solution for every initial guess. It sometimes happens that the QMR approximation at step $k$ is almost as good as the (optimal) GMRES approximation, but sometimes this is not the case. It was shown by Faber and Manteuffel [45] that if "optimal" is taken to mean having the smallest possible error in some inner product norm that is *independent* of the initial vector, then the optimal approximation cannot be generated with a short recurrence. The details of this result are provided in section 6.1. The result should not necessarily be construed as ruling out the possibility of a clear "method of choice" for non-Hermitian problems. Instead, it may suggest directions in the search for such a method. Possibilities are discussed in section 6.2.

## 6.1. The Faber and Manteuffel Result.

Consider a recurrence of the following form. Given $x_0$, compute $p_0 = b - Ax_0$, and for $k = 1, 2, \ldots$, set

$$x_k = x_{k-1} + a_{k-1}p_{k-1}, \tag{6.1}$$

$$p_k = Ap_{k-1} - \sum_{j=k-s+1}^{k-1} b_{k-1,j}p_j \tag{6.2}$$

for some coefficients $a_{k-1}$ and $b_{k-1,j}$, $j = k-s+1, \ldots, k-1$, where $s$ is some integer less than $n$. It is easy to show by induction that the approximate solution $x_k$ generated by this recurrence is of the form

$$x_k \in x_0 + \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\} \tag{6.3}$$

and that the direction vectors $p_0, \ldots, p_{k-1}$ form a basis for the Krylov space

$$\text{span}\{p_0, p_1, \ldots p_{k-1}\} = \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}.$$

The recurrence Orthodir(3) is of the form (6.1–6.2), with $s = 3$, as is Orthomin(2). To see that Orthomin(2) is of this form, note that in that

algorithm we have

$$p_k = r_k - b_{k-1}p_{k-1}, \quad \text{where } r_k = r_{k-1} - a_{k-1}Ap_{k-1}.$$

Substituting for $r_k$ in the recurrence for $p_k$ gives

$$p_k = r_{k-1} - a_{k-1}Ap_{k-1} - b_{k-1}p_{k-1},$$

and using the fact that $r_{k-1} = p_{k-1} + b_{k-2}p_{k-2}$ gives

$$p_k = -a_{k-1}Ap_{k-1} + (1 - b_{k-1})p_{k-1} + b_{k-2}p_{k-2}.$$

The normalization of $p_k$ is of no concern, since that can be accounted for by choosing the coefficient $a_{k-1}$ appropriately, so if $p_k$ is replaced by $[(-1)^k \prod_{j=1}^{k-1} a_j^{-1}]p_k$, then Orthomin(2) fits the pattern (6.1–6.2). The MINRES algorithm for Hermitian problems is also of this form, and it has the desirable property of generating, at each step, the approximation of the form (6.3) for which the 2-norm of the residual is minimal. The CG algorithm for Hermitian positive definite problems is also of the form (6.1–6.2), and at each step it generates the approximation of the form (6.3) for which the $A$-norm of the error is minimal.

For what matrices $A$ can one construct a recurrence of the form (6.1–6.2) with the property that for any initial vector $x_0$, the approximation $x_k$ at step $k$ is the "optimal" approximation from the space (6.3), where "optimal" means that the error $e_k \equiv A^{-1}b - x_k$ is minimal in some *inner product* norm, the inner product being *independent* of the initial vector? This is essentially the question answered by Faber and Manteuffel [45]. See also [46, 5, 86, 138]. We will not include the entire proof, but the answer is that for $s < \sqrt{n}$, except for a few anomalies, the matrices for which such a recurrence exists are those of the form $B^{-1/2}CB^{1/2}$, where $C$ is either Hermitian or of the form $C = e^{i\theta}(dI + F)$, with $d$ real and $F^H = -F$, and $B$ is a Hermitian positive definite matrix. Equivalently (Exercise 6.1), such a recurrence exists for matrices $A$ of the form

$$A = e^{i\phi}(icI + G), \quad c \geq 0, \quad 0 \leq \phi \leq 2\pi, \quad B^{-1}G^HB = G. \tag{6.4}$$

If $A$ is of the form (6.4), then $B^{1/2}AB^{-1/2}$ is just a shifted and rotated Hermitian matrix.

To see why this class of matrices is special, note that the error $e_k$ in a recurrence of the form (6.1–6.2) satisfies

$$e_k \in e_0 + \text{span}\{p_0, p_1, \ldots, p_{k-1}\}. \tag{6.5}$$

If $\langle\langle \cdot, \cdot \rangle\rangle$ denotes the inner product in which the norm of $e_k$ is minimized, then $e_k$ must be the unique vector of the form (6.5) satisfying

$$\langle\langle e_k, p_j \rangle\rangle = 0, \quad j = 0, 1, \ldots, k-1.$$

It follows that since $e_k = e_{k-1} - a_{k-1}p_{k-1}$, the coefficient $a_{k-1}$ must be

$$a_{k-1} = \frac{\langle\langle e_{k-1}, p_{k-1}\rangle\rangle}{\langle\langle p_{k-1}, p_{k-1}\rangle\rangle}.$$

For $j < k-1$, we have

$$\langle\langle e_k, p_j\rangle\rangle = \langle\langle e_{k-1}, p_j\rangle\rangle - a_{k-1}\langle\langle p_{k-1}, p_j\rangle\rangle,$$

so if $\langle\langle e_{k-1}, p_j\rangle\rangle = 0$, then in order to have $\langle\langle e_k, p_j\rangle\rangle = 0$, it is necessary that either $a_{k-1} = 0$ or $\langle\langle p_{k-1}, p_j\rangle\rangle = 0$. If it is required that $\langle\langle p_k, p_j\rangle\rangle = 0$ for all $k$ and all $j < k$, then the coefficients $b_{k-1,j}$ must be given by

$$b_{k-1,j} = \frac{\langle\langle Ap_{k-1}, p_j\rangle\rangle}{\langle\langle p_j, p_j\rangle\rangle}.$$

A precise statement of the Faber and Manteuffel result is given in the following definition and theorem.

DEFINITION 6.1.1. *An algorithm of the form* (6.1–6.2) *is an* $s$-term CG method *for* $A$ *if, for every* $p_0$, *the vectors* $p_k$, $k = 1, 2, \ldots, m-1$, *satisfy* $\langle\langle p_k, p_j\rangle\rangle = 0$ *for all* $j < k$, *where* $m$ *is the number of steps required to obtain the exact solution* $x_m = A^{-1}b$.

THEOREM 6.1.1 (Faber and Manteuffel [45]). *An* $s$*-term CG method exists for the matrix* $A$ *if and only if either*

(i) *the minimal polynomial of* $A$ *has degree less than or equal to* $s$, *or*

(ii) $A^\star$ *is a polynomial of degree less than or equal to* $s-2$ *in* $A$, *where* $A^\star$ *is the adjoint of* $A$ *with respect to some inner product, that is,* $\langle\langle Av, w\rangle\rangle = \langle\langle v, A^\star w\rangle\rangle$ *for all vectors* $v$ *and* $w$.

*Proof (of sufficiency only).* The choice of coefficients $a_i$ and $b_{i,j}$, $i = 0, \ldots, s-1$, $j \le i$ not only forces $\langle\langle p_k, p_j\rangle\rangle = 0$, $k = 1, \ldots, s-1$, $j < k$, but also ensures that the error at steps 1 through $s$ is minimized in the norm corresponding to the given inner product. Since the error at step $k$ is equal to a certain $k$th-degree polynomial in $A$ times the initial error, if the minimal polynomial of $A$ has degree $k \le s$, then the algorithm will discover this minimal polynomial (or another one for which $e_k = p_k(A)e_0 = 0$), and the exact solution will be obtained after $k \le s$ steps. In this case, then, iteration (6.1–6.2) is an $s$-term CG method.

For $k \ge s$ and $i < k-s+1$, it follows from (6.2) that

$$\langle\langle p_k, p_i\rangle\rangle = \langle\langle Ap_{k-1}, p_i\rangle\rangle - \sum_{j=k-s+1}^{k-1} b_{k-1,j}\langle\langle p_j, p_i\rangle\rangle.$$

If $\langle\langle p_j, p_i\rangle\rangle = 0$ for $j = k-s+1, \ldots, k-1$, then we will have $\langle\langle p_k, p_i\rangle\rangle = 0$ if and only if

$$\langle\langle Ap_{k-1}, p_i\rangle\rangle \equiv \langle\langle p_{k-1}, A^\star p_i\rangle\rangle = 0. \tag{6.6}$$

If $A^\star = q_{s-2}(A)$ for some polynomial $q_{s-2}$ of degree $s-2$ or less, then (6.6) will hold, since $p_{k-1}$ is orthogonal to the space

$$\text{span}\{p_0, \ldots, p_{k-2}\} = \text{span}\{p_0, Ap_0, \ldots, A^{k-2}p_0\},$$

which contains $q_{s-2}(A)p_i$ since $i + s - 2 \leq k - 2$. □

To clarify condition (ii) in Theorem 6.1.1, first recall (section 1.3.1) that for any inner product $\langle\langle\cdot,\cdot\rangle\rangle$ there is a Hermitian positive definite matrix $B$ such that

$$\langle\langle v, w\rangle\rangle = \langle v, Bw\rangle$$

for all vectors $v$ and $w$, where $\langle\cdot,\cdot\rangle$ denotes the standard Euclidean inner product. The $B$-adjoint of $A$, denoted $A^\star$ in the theorem, is the unique matrix satisfying

$$\langle Av, Bw\rangle = \langle v, BA^\star w\rangle$$

for all $v$ and $w$. From this definition it follows that

$$A^\star = B^{-1}A^H B,$$

where the superscript $^H$ denotes the adjoint in the Euclidean norm $A^H = \bar{A}^T$. The matrix $A$ is said to be *$B$-normal* if and only if $A^\star A = AA^\star$. If $B^{1/2}$ denotes the Hermitian positive definite square root of $B$, then this is equivalent to the condition that

$$(B^{-1/2}A^H B^{1/2})(B^{1/2}AB^{-1/2}) = (B^{1/2}AB^{-1/2})(B^{-1/2}A^H B^{1/2}),$$

which is the condition that $B^{1/2}AB^{-1/2}$ be *normal*.

Let $B$ be fixed and let $\tilde{A}$ denote the matrix $B^{1/2}AB^{-1/2}$. It can be shown (Exercise 6.2) that $\tilde{A}$ is normal ($A$ is $B$-normal) if and only if $\tilde{A}^H$ can be written as a polynomial (of some degree) in $\tilde{A}$. If $\eta$ is the smallest degree for which this is true, then $\eta$ is called the *$B$-normal degree* of $A$. For any integer $t \geq \eta$, $A$ is said to be *$B$-normal($t$)*. With this notation, condition (ii) of Theorem 6.1.1 can be stated as follows:

(ii′) $A$ is $B$-normal($s-2$).

Condition (ii′) still may seem obscure, but the following theorem, also from [45], shows that matrices $A$ with $B$-normal degree $\eta$ greater than 1 but less than $\sqrt{n}$ also have minimal polynomials of degree less than $n$. These matrices belong to a subspace of $\mathbf{C}^{n\times n}$ of dimension less than $n^2$, so they might just be considered anomalies. The more interesting case is $\eta = 1$ or the $B$-normal(1) matrices in (ii′).

THEOREM 6.1.2 (Faber and Manteuffel [45]). *If $A$ has $B$-normal degree $\eta > 1$, then the minimal polynomial of $A$ has degree less than or equal to $\eta^2$.*

*Proof.* The degree $d(A)$ of the minimal polynomial of $A$ is the same as that of $\tilde{A} \equiv B^{1/2}AB^{-1/2}$. Since $\tilde{A}$ is normal, it has exactly $d(A)$ distinct eigenvalues, and we will have $\tilde{A}^H = q(A)$ if and only if

$$q(\lambda_i) = \bar{\lambda}_i, \quad i = 1, \ldots, d(A).$$

How many distinct complex numbers $z$ can satisfy $q(z) = \bar{z}$? Note that $\bar{q}(\bar{z}) = z$ or $\bar{q}(q(z)) = z$. The expression $\bar{q}(q(z)) - z$ is a polynomial of degree exactly $\eta^2$ if $q$ has degree $\eta > 1$. (If the degree of $q$ were 1, this expression could be identically zero.) It follows that there are at most $\eta^2$ distinct roots, so $d(A) \leq \eta^2$. □

The $B$-normal(1) matrices, for which a 3-term CG method exists, are characterized in the following theorem.

THEOREM 6.1.3 (Faber and Manteuffel [45]). *If $A$ is $B$-normal(1) then $d(A) = 1$, $A^\star = A$, or*

$$\tilde{A} \equiv B^{1/2}AB^{-1/2} = e^{\imath\theta}\left(\frac{r}{2}I + F\right),$$

*where $r$ is real and $F = -F^H$.*

*Proof.* Since $\tilde{A}$ is normal, if $\tilde{A}$ has all real eigenvalues, then $\tilde{A}^H = \tilde{A}$ or $A^\star = A$.

Suppose $\tilde{A}$ has at least one complex eigenvalue. There is a linear polynomial $q$ such that each of the eigenvalues $\lambda_i$ of $\tilde{A}$ satisfies $q(\lambda_i) = \bar{\lambda}_i$. This implies that $\bar{q}(\bar{\lambda}_i) = \lambda_i$ or $\bar{q}(q(\lambda_i)) - \lambda_i = 0$. In general, this equation has just one root $\lambda_i$, and if this is the case then $d(A) = 1$.

Let $q(z) = az - b$. The expression $\bar{q}(q(\lambda_i)) - \lambda_i = 0$ can be written as

$$(\bar{a}a - 1)\lambda_i - (\bar{a}b + \bar{b}) = 0.$$

There is more than one root $\lambda_i$ only if the expression on the left is identically zero, which means that $a = -b/\bar{b}$. Let $b = re^{\imath\theta}$, $\imath \equiv \sqrt{-1}$. Then

$$q(z) = -e^{\imath\theta}(ze^{-\imath\theta} - r).$$

If $q(z) = \bar{z}$, then

$$-(ze^{-\imath\theta} - r) = \bar{z}e^{\imath\theta} = \overline{ze^{-\imath\theta}},$$

which yields

$$r = ze^{-\imath\theta} + \overline{ze^{-\imath\theta}}.$$

Thus, if $\lambda$ is an eigenvalue of $\tilde{A}$, the real part of $\lambda e^{-\imath\theta}$ is $r/2$. This implies that

$$F = e^{-\imath\theta}\tilde{A} - \frac{r}{2}I$$

has only pure imaginary eigenvalues; hence, since $\tilde{A}$ is normal, $F = -F^H$. □

### 6.2. Implications.

The class of $B$-normal(1) matrices of the previous section are matrices for which CG methods are already known. They are diagonalizable matrices whose spectrum is contained in a line segment in the complex plane. See [26, 142].

Theorems 6.1.1–6.1.3 imply that for most non-Hermitian problems, one cannot expect to find a short recurrence that generates the optimal approximation from successive Krylov spaces, if "optimality" is defined in terms of an inner product norm that is independent of the initial vector. It turns out that most non-Hermitian iterative methods actually do find the optimal approximation in some norm [11] (see Exercise 6.3). Unfortunately, however, it is a norm that cannot be related easily to the 2-norm or the $\infty$-norm or any other norm that is likely to be of interest. For example, the BiCG approximation is optimal in the $P_n^{-H}P_n^{-1}$-norm, where the columns of $P_n$ are the biconjugate direction vectors. The QMR approximation is optimal in the $A^H V_n^{-H} V_n^{-1} A$-norm, where the columns of $V_n$ are the biorthogonal basis vectors.

The possibility of a short recurrence that would generate optimal approximations in some norm that depends on the initial vector but that can be shown to differ from, say, the 2-norm by no more than some moderate size factor remains. This might be the best hope for developing a clear "method of choice" for non-Hermitian linear systems.

It should also be noted that the Faber and Manteuffel result deals only with a *single* recurrence. It is still an open question whether coupled short recurrences can generate optimal approximations. For some preliminary results, see [12].

It remains a major open problem to find a method that generates provably "near-optimal" approximations in some standard norm while still requiring only $O(n)$ work and storage (in addition to the matrix–vector multiplication) at each iteration—or to prove that such a method does not exist.

### Exercises.

6.1. Show that a matrix $A$ is of the form (6.4) if and only if it is of the form $B^{-1/2}CB^{1/2}$, where $C$ is either Hermitian or of the form $e^{i\theta}(dI + F)$, with $d$ real and $F^H = -F$.

6.2. Show that a matrix $A$ is normal if and only if $A^H = q(A)$ for some polynomial $q$. (Hint: If $A$ is normal, write $A$ in the from $A = U\Lambda U^H$, where $\Lambda$ is diagonal and $U$ is unitary, and determine a polynomial $q$ for which $q(\Lambda) = \bar{\Lambda}$.)

6.3. The following are special instances of results due to Barth and Manteuffel [11]:

(a) Assume that the BiCG iteration does not break down or find the exact solution before step $n$. Use the fact that the BiCG error at

step $k$ is of the form

$$e_k \in e_0 + \text{span}\{p_0, p_1, \ldots, p_{k-1}\}$$

and the residual satisfies

$$r_k \perp \text{span}\{\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_{k-1}\}$$

to show that the BiCG approximation at each step is optimal in the $P_n^{-H}P_n^{-1}$-norm, where the columns of $P_n$ are the biconjugate direction vectors.

(b) Assume that the two-sided Lanczos recurrence does not break down or terminate before step $n$. Use the fact that the QMR error at step $k$ is of the form $e_k = e_0 - V_k y_k$ and that the QMR residual satisfies $r_k^H V_{k+1} V_{k+1}^H A V_k = 0$ to show that the QMR approximation at each step is optimal in the $A^H V_n^{-H} V_n^{-1} A$-norm, where the columns of $V_n$ are the biorthogonal basis vectors.

6.4. Write down a CG method for matrices of the form $I-F$, where $F = -F^H$, which minimizes the 2-norm of the residual at each step. (Hint: Note that one can use a 3-term recurrence to construct an orthonormal basis for the Krylov space $\text{span}\{q_1, (I-F)q_1, \ldots, (I-F)^{k-1}q_1\}$, when $F$ is skew-Hermitian.)