Electronic Transactions on Numerical Analysis.
Volume 33, pp. 17-30, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ANALYSIS OF SOME KRYLOV SUBSPACE METHODS FOR NORMAL MATRICES VIA APPROXIMATION THEORY AND CONVEX OPTIMIZATION*

M. BELLALIJ†, Y. SAAD‡, AND H. SADOK§

Dedicated to Gérard Meurant on the occasion of his 60th birthday

Abstract. Krylov subspace methods are strongly related to polynomial spaces and their convergence analysis can often be naturally derived from approximation theory. Analyses of this type lead to discrete min-max approximation problems over the spectrum of the matrix, from which upper bounds on the relative Euclidean residual norm are derived. A second approach to analyzing the convergence rate of the GMRES method or the Arnoldi iteration, uses as a primary indicator the (1,1) entry of the inverse of $K_m^H K_m$ where K_m is the Krylov matrix, i.e., the matrix whose column vectors are the first m vectors of the Krylov sequence. This viewpoint allows us to provide, among other things, a convergence analysis for normal matrices using constrained convex optimization. The goal of this paper is to explore the relationships between these two approaches. Specifically, we show that for normal matrices, the Karush-Kuhn-Tucker (KKT) optimality conditions derived from the convex maximization problem are identical to the properties that characterize the polynomial of best approximation on a finite set of points. Therefore, these two approaches are mathematically equivalent. In developing tools to prove our main result, we will give an improved upper bound on the distances of a given eigenvector from Krylov spaces.

Key words. Krylov subspaces, polynomials of best approximation, min-max problem, interpolation, convex optimization, KKT optimality conditions

AMS subject classifications. 65F10, 65F15

1. Introduction. This paper is concerned with the study of convergence of Krylov subspace methods for solving linear systems of equations,

$$Ax = b, \tag{1.1}$$

or eigenvalue problems

$$Au = \lambda u. \tag{1.2}$$

Here, A is a given matrix of size $N \times N$, possibly complex. These are projection methods onto Krylov subspaces

$$\mathcal{K}_m(A, v) = \operatorname{span}\left\{v, Av, \cdots, A^{m-1}v\right\},$$

generated by v and A, where $v \in \mathbb{C}^N$ is a given initial vector.

A wide variety of iterative methods fall within the Krylov subspace framework. This paper focuses on two methods for non-Hermitian matrix problems: Arnoldi's method [15], which computes eigenvalues and eigenvectors of A; and the generalized minimal residual method (GMRES) [14], which solves linear systems of equations. GMRES extracts an approximate solution $x^{(m)}$ from the affine subspace $x^{(0)} + \mathcal{K}_m(A, r^{(0)})$, where $r^{(0)} = b - Ax^{(0)}$

* Received December 13, 2007. Accepted August 14, 2008. Published online on October 17, 2008. Recommended by M. Sadkane.

†Université de Valenciennes, Le Mont Houy, 59313 Valenciennes Cedex, France. E-mail: Mohammed.Bellalij@univ-valenciennes.fr.

‡University of Minnesota, Dept. of Computer Science and Engineering. E-mail: saad@cs.umn.edu.
Work supported by NSF grant ACI-0305120, DOE grant DE-FG-08ER25841, and by the Minnesota Supercomputing Institute.

§L.M.P.A, Université du Littoral, 50 rue F. Buisson BP699, F-62228 Calais Cedex, France. E-mail: sadok@lmpa.univ-littoral.fr.

ETNA
Kent State University
etna@mcs.kent.edu

is the initial residual and $x^{(0)} \in \mathbb{C}^N$ is a given initial approximate solution of (1.1), by requiring that this approximation yields a minimum residual norm. The question of estimating the convergence rate of iterative methods of this type has received much attention in the past and is still an active area of research. Researchers have taken different paths to provide answers to the question, which is a very hard one in the non-normal case.

Numerous papers deal with this issue by deriving upper bounds for the residual norm, which reveal some intrinsic links between the convergence properties and the spectral information available for A. The standard technique in most of these works [2, 3, 5, 6, 8, 15, 16] is based on a polynomial approach. More precisely, the link between residual vectors and polynomials inspired a search for bounds on the residual norm that are derived from analytic properties of some normalized associated polynomials as functions defined on the complex plane. In recent years, a different approach taking a purely algebraic point of view was advocated for studying the convergence of the GMRES method. This approach, discussed initially by Sadok [17] (see also [18]) and Ipsen [7], and followed by Zavorin, O'Leary and Elman [20], and Liesen and Tichy [10], is distinct from that based on approximation theory. Related theoretical residual bounds have been established, by exploring certain classes of matrices, in trying to explain the obscure behavior of this method, in particular the stagnation phenomenon. Nevertheless, a great number of open questions remain.

Exploiting results shown in [18], we have recently presented in [1] an alternative way to analyze the convergence of the GMRES and Arnoldi methods, based on an expression for the residual norm in terms of determinants of Krylov matrices. An appealing feature of this viewpoint is that it allows us to thoroughly analyze convergence for normal matrices using results from constrained convex optimization. Our approach provides an upper bound for the residual norm, at any step, which can be expressed as a product of relative eigenvalue differences.

The purpose of the present work is to show the connection between these two approaches: the min-max polynomial approximation viewpoint and the constrained convex optimization viewpoint. Specifically, we establish that the Karush-Kuhn-Tucker (KKT) optimality conditions derived from the convex maximization problem are mathematically equivalent to the properties that characterize the best polynomial approximation on a finite set of points.

The paper is organized as follows. Section 2 sets the notation and states the main result which will be key to showing the connection between the approximation theory viewpoint and convex optimization. In the same section is a useful lemma whose application to the GMRES and Arnoldi cases leads to the introduction of the optimization viewpoint. Sections 3 and 4 begin with brief reviews of the two Krylov methods under consideration and then derive upper bounds for GMRES and for Arnoldi algorithms respectively. Results concerning the characterization of the polynomials of best approximation on finite point sets are discussed in Section 5 and they are then applied to our situation. Section 6 outlines the proof of the main result by examining the KKT optimality conditions derived from the convex maximization problem and establishes the equivalence of the two formulations. Finally, a few concluding remarks are made in Section 7.

2. Preliminaries and statement of the main result. Throughout the paper it is assumed that the matrix under consideration, namely A, is normal. In addition, all results are under the assumption that exact arithmetic is used. The Euclidean two-norm on $\mathbb{C}^N$ and the matrix norm induced by it will be denoted by $\|\cdot\|$. The identity matrix of order m (respectively N) is denoted by I_m (respectively I). We use e_i to denote the ith column of the identity of appropriate order.

Let $A \in \mathbb{C}^{N\times N}$ be a complex matrix with spectrum $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$. Since A

ETNA
Kent State University
etna@mcs.kent.edu

is normal, there exists a unitary matrix U such that

$$A = U\Lambda U^H,$$

where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$. The superscripts "$T$" and "$H$" indicate the transpose and the conjugate transpose respectively. The diagonal matrix with diagonal entries β_i, $i = 1, \ldots, m$ will be denoted by

$$D_\beta = \operatorname{diag}(\beta_1, \beta_2, \ldots, \beta_m). \tag{2.1}$$

The Krylov matrix whose columns are $r^{(0)}, A\, r^{(0)}, \ldots, A^{m-1}\, r^{(0)}$ is denoted by K_m. Polynomials which are closely related to Krylov subspaces will play an important role in our analysis. We denote the set of all polynomials of degree not exceeding m by $\mathbb{P}_m$ and the set of polynomials of $\mathbb{P}_m$ with value one at ω by $\mathbb{P}_m^{(\omega)}$. Recall that for any $p \in \mathbb{P}_m$ we have $p(A) = p(U\Lambda U^H) = Up(\Lambda)U^H$.

For any vector $\mu = [\mu_1, \mu_2, \ldots, \mu_M]^T$, we denote by $V_m(\mu)$ the rectangular Vandermonde matrix:

$$V_m(\mu) = \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_1^{m-1} \\ 1 & \mu_2 & \cdots & \mu_2^{m-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \mu_M & \cdots & \mu_M^{m-1} \end{bmatrix}. \tag{2.2}$$

For example we will denote by $V_m(\lambda)$ the matrix of size $N \times m$ whose entry (i,j) is λ_i^{j-1}, where $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvalues of A.

We will also need a notation for a specific row of a matrix of the form (2.2). We define the vector function

$$s_m(\omega) = \begin{bmatrix} 1 & \omega & \ldots & \omega^{m-1} \end{bmatrix}^H. \tag{2.3}$$

Note that the ith row of the Vandermonde matrix (2.2) is $s_m(\mu_i)^H$.

Finally, for a complex number $z = \rho e^{\imath\theta}$, $\overline{z} = \rho e^{-\imath\theta}$ denotes the complex conjugate of z, $|z| = \rho$ its modulus, and $\operatorname{sgn}(z) = z/|z|$ its sign.

2.1. Main result. The result to be presented next will be used in the convergence analysis of both GMRES and the Arnoldi method. For this reason it is stated in general terms without reference to a specific algorithm. We denote by Δ_M the standard simplex of $\mathbb{R}^M$:

$$\Delta_M = \left\{ \gamma \in \mathbb{R}^M : \gamma \geq 0 \text{ and } e^T\gamma = 1 \right\},$$

where $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^M$. The common notation $\gamma \geq 0$ means that $\gamma_i \geq 0$ for $i = 1, \ldots, M$. Let $\omega, \mu_1, \ldots, \mu_M$ be $(M+1)$ distinct complex numbers, and m an integer such that $m + 1 < M$. Define the following function of γ:

$$F_{m,\omega}(\gamma) = \frac{1}{s_{m+1}^H(\omega)\left(V_{m+1}(\mu)^H D_\gamma V_{m+1}(\mu)\right)^{-1} s_{m+1}(\omega)}. \tag{2.4}$$

Then, we can state the following.

THEOREM 2.1. *Let $\widetilde{\Delta}_M \subset \Delta_M$ be the domain of definition of $F_{m,\omega}$. Then the supremum of $F_{m,\omega}$ over $\widetilde{\Delta}_m$ is reached and we have*

$$\left(\min_{p \in \mathbb{P}_m^{(\omega)}} \max_{j=1,\ldots,M} |p(\mu_j)| \right)^2 = \max_{\gamma \in \widetilde{\Delta}_M} F_{m,\omega}(\gamma). \tag{2.5}$$

The above theorem will be helpful in linking two approaches, one based on approximation theory and the other on optimization. It shows in effect that the min-max problem is equivalent to a convex optimization problem.

The proof of this theorem is based on the Karush-Kuhn-Tucker (KKT) conditions applied to the maximization problem stated above. From these KKT conditions, we will derive two linear systems which appear in the characterization of the polynomial of best approximation.

2.2. A lemma on the projection error. Next, we state a known lemma which will be key in establishing some relations between various results. This lemma gives a simple expression for the projection error, i.e., the difference between a given vector and its orthogonal projection onto a subspace.

LEMMA 2.2. *Let X be an arbitrary subspace with a basis represented by the matrix B and let $c \notin X$. Let $\mathcal{P}$ be the orthogonal projector onto X. Then*

$$\|(I-\mathcal{P})c\|^2 = \frac{1}{e_1^T C^{-1} e_1}, \tag{2.6}$$

where

$$C = \begin{bmatrix} c^H c & c^H B \\ B^H c & B^H B \end{bmatrix}.$$

Proof. The proof, given in [1], is reproduced here for completeness. Given an arbitrary vector $c \in \mathbb{C}^N$, observe that

$$\|(I-\mathcal{P})c\|^2 = c^H(I-\mathcal{P})(I-\mathcal{P})c = c^H(I-\mathcal{P})c = c^H c - c^H \mathcal{P} c$$

with $\mathcal{P} = B(B^H B)^{-1} B^H$. From this it follows that

$$\|(I-\mathcal{P})c\|^2 = c^H c - c^H B(B^H B)^{-1} B^H c. \tag{2.7}$$

The right-hand side of (2.7) is simply the Schur complement of the (1,1) entry of C, which is the inverse of the (1,1) entry of C^{-1}. □

The expression (2.6) can also be derived from basic identities satisfied by least squares residuals, as was first shown in [19]. This was later formulated explicitly and proved in [9] by exploiting properties of the Moore-Penrose generalized inverse.

3. Convergence analysis for GMRES for normal matrices. The idea of the GMRES algorithm for solving linear systems is to project the problem onto the Krylov subspace of dimension $m \le N$. The GMRES algorithm starts with an initial guess $x^{(0)}$ for the solution of (1.1) and seeks the mth approximate solution $x^{(m)}$ in the affine subspace $x^{(0)} + \mathcal{K}_m(A, r^{(0)})$ satisfying the residual norm minimization property

$$\|b - Ax^{(m)}\| = \min_{u \in x^{(0)} + \mathcal{K}_m(A, r^{(0)})} \|b - Au\| = \min_{z \in \mathcal{K}_m(A, r^{(0)})} \|r^{(0)} - Az\|. \tag{3.1}$$

As can be readily seen, this approximation is of the form $x^{(m)} = x^{(0)} + p^*_{m-1}(A) r^{(0)}$, where $p^*_{m-1} \in \mathbb{P}_{m-1}$. Therefore, the residual $r^{(m)}$ has the polynomial representation $r^{(m)} = (I - Ap^*_{m-1}(A)) r^{(0)}$, and the problem (3.1) translates to

$$\|r^{(m)}\| = \min_{p \in \mathbb{P}_m^{(0)}} \|p(A) r^{(0)}\|. \tag{3.2}$$

ETNA
Kent State University
etna@mcs.kent.edu

Characteristic properties of GMRES are that the norm of $r^{(m)}$ is a non-increasing function of m and that it terminates in m steps if $r^{(m)} = 0$ and $r^{(m-1)} \neq 0$. Moreover, we have $r^{(m)} \neq 0$ if and only if $\dim(\mathcal{K}_{m+1}(A, r^{(0)})) = m+1$. Therefore, while analyzing the convergence of GMRES, we will assume that the Krylov matrix K_{m+1} is of rank $m+1$. Before we turn to bounds for analyzing convergence, we will explore two ways to study the convergence rate: using so-called optimal polynomials and using the spectral decomposition of Krylov matrices.

3.1. Analysis based on the best uniform approximation. The contribution of the initial residual in (3.2) is usually simplified using the inequality $\|p(A)r^{(0)}\| \leq \|p(A)\|\|r^{(0)}\|$. Then the issue becomes one of finding an upper bound for $\|p(A)\|$ for all $p \in \mathbb{P}_m^{(0)}$. It follows that an estimate of the relative Euclidean residual norm $\|r^{(m)}\|/\|r^{(0)}\|$ is associated with the so-called ideal GMRES polynomial of a matrix problem: $\min_{p\in\mathbb{P}_m^{(0)}} \|p(A)\|$. If we expand $r^{(0)}$ in the eigenbasis $r^{(0)} = U\alpha$ then

$$\|p(A)r^{(0)}\| = \|Up(\Lambda)U^H r^{(0)}\| = \|Up(\Lambda)\alpha\| = \|p(\Lambda)\alpha\| \leq \|\alpha\| \max_{i=1,\ldots,N} |p(\lambda_i)|,$$

which then shows that

$$\|r^{(m)}\| \leq \|r^{(0)}\| \min_{p\in\mathbb{P}_m^{(0)}} \max_{\lambda\in\sigma(A)} |p(\lambda)|\,.$$

3.2. Analysis based on convex optimization. Next, an alternative viewpoint for analyzing residual norms will be formulated. This alternative, developed in [1, 18], uses as a primary indicator the $(1,1)$ entry of the inverse of $K_l^H K_l$, where K_l is the Krylov matrix with l columns associated with the GMRES method.

It is assumed that $\mathrm{rank}(K_{m+1}) = m+1$. Setting $c = r^{(0)}$ and $B = AK_m$ in Lemma 2.2 yields the following expression for the residual norm $r^{(m)}$:

$$\|r^{(m)}\|^2 = \frac{1}{e_1^T\,(K_{m+1}^H K_{m+1})^{-1}\,e_1}. \tag{3.3}$$

Assume that $r^{(0)}$ has the eigen-expansion $r^{(0)} = U\alpha$. A little calculation shows (see [20]) that we can write K_{m+1} as $K_{m+1} = U\,D_\alpha\,V_{m+1}(\lambda)$ (spectral factorization of K_{m+1}). We refer to (2.1) and (2.2) for the definitions of D_α and $V_{m+1}(\lambda)$. Thus, since A is normal ($U^H U = I$), we obtain $K_{m+1}^H K_{m+1} = \|\alpha\|^2 V_{m+1}^H(\lambda) D_\beta V_{m+1}(\lambda)$ where $\beta_i = |\alpha_i|^2/\|\alpha\|^2 = |\alpha_i|^2/\|r^{(0)}\|^2$. We therefore can rewrite (3.3) as follows:

$$\frac{\|r^{(m)}\|^2}{\|r^{(0)}\|^2} = \frac{1}{e_1^T\,\left(V_{m+1}^H(\lambda) D_\beta V_{m+1}(\lambda)\right)^{-1}\,e_1}. \tag{3.4}$$

Note that $\beta \in \widetilde{\Delta}_N$. Thus, an optimal bound for $\|r^{(m)}\|^2/\|r^{(0)}\|^2$ can be obtained by choosing $\beta \in \widetilde{\Delta}_N$ to maximize the right-hand side of (3.4). In fact, thanks to Theorem 2.1, the two bounds given in this section coincide. Indeed, substituting $M = N$, $\mu_j = \lambda_j$ and $\omega = 0$ in (2.5) yields

$$\max_{\beta\in\Delta_N} \frac{1}{\left(e_1^T\,\left(V_{m+1}^H(\lambda) D_\beta V_{m+1}(\lambda)\right)^{-1} e_1\right)} = \max_{\beta\in\widetilde{\Delta}_N} F_{m,0}(\beta) = \left(\min_{p\in\mathbb{P}_m^{(0)}} \max_{j=1,\ldots,N} |p(\lambda_j)|\right)^2.$$

We end this section by stating the following assertion. Let $\Im(\beta)$ be the set of indices j such that $\beta_j \neq 0$. If $\mathrm{rank}(K_{m+1}) = m+1$ ($\|r^{(m)}\| \neq 0$) then clearly the cardinality of $\Im(\beta)$ is at least $m+1$.

ETNA
Kent State University
etna@mcs.kent.edu

4. Outline of the convergence analysis for Arnoldi's method. Arnoldi's method approximates solutions of the eigenvalue problem (1.2) by computing an approximate eigenpair $(\widetilde{\lambda}^{(m)}, \widetilde{u}^{(m)})$ obtained from the Galerkin condition

$$\widetilde{u}^{(m)} \in \mathcal{K}_m(A, v_1),$$
$$\text{and} \quad (A\widetilde{u}^{(m)} - \widetilde{\lambda}^{(m)}\widetilde{u}^{(m)}, A^i v_1) = 0 \text{ for } i = 0, \ldots, m-1.$$

Let $\mathcal{P}_m$ be the orthogonal projector onto the Krylov subspace $\mathcal{K}_m(A, v_1)$, and let (λ, u) be an exact eigenpair of A. In [15], the following result was shown to analyze the convergence of the process in terms of the projection error $\|u - \mathcal{P}_m u\|$ of a given eigenvector u from the subspace $\mathcal{K}_m(A, v_1)$.

THEOREM 4.1. *Let $A_m = \mathcal{P}_m A \mathcal{P}_m$ and let $\gamma_m = \|\mathcal{P}_m(A - \lambda I)(I - \mathcal{P}_m)\|$. Then the residual norms of the pairs $\lambda, \mathcal{P}_m u$ and λ, u for the linear operator A_m satisfy, respectively*

$$\|(A_m - \lambda I)\mathcal{P}_m u)\| \leq \gamma_m \|(I - \mathcal{P}_m)u\|,$$
$$\|(A_m - \lambda I)u\| \leq \sqrt{|\lambda|^2 + \gamma_m^2}\,\|(I - \mathcal{P}_m)u\|.$$

This theorem establishes that the convergence of the Arnoldi method can be analyzed by estimating $\|(I - \mathcal{P}_m)u\|$. Note that $\gamma_m \leq \|A\|$. The result shows that under the condition that the projected problem is not too ill-conditioned, there will be an approximate eigenpair close to the exact one when $\|(I - \mathcal{P}_m)u\|$ is small.

4.1. Analysis based on (uniform) approximation theory. The result just discussed above shows how the convergence analysis of the Arnoldi method can be stated in terms of the projection error $\|(I - \mathcal{P}_m)u\|$ between the exact eigenvector u and the Krylov subspace. The usual technique to estimate this projection error assumes that A is diagonalizable and expands the initial vector v_1 in the eigenbasis as $v_1 = \sum_{j=1}^{N} \alpha_j u_j$. We examine the convergence of a given eigenvalue which is indexed by 1; i.e., we consider u_1, the first column of U. Adapting Lemma 6.2 from [15] stated for diagonalizable matrices to the special situation of normal matrices gives the following theorem.

THEOREM 4.2. *Let A be normal ($A = U\Lambda U^H$, $U^H U = I$) and let $v_1 = \sum_{j=1}^{N} \alpha_j u_j = U\alpha$; then*

$$\|(I - \mathcal{P}_m)u_1\| \leq \frac{\sqrt{\sum_{j=2}^{N} |\alpha_j|^2}}{|\alpha_1|} \epsilon_1^{(m)}, \tag{4.1}$$

where $\epsilon_1^{(m)} = \min_{p \in \mathbb{P}_{m-1}^{(\lambda_1)}} \max_{j=2,\ldots,N} |p(\lambda_j)|$.

The right-hand side of (4.1) may exceed one, but we know that $\|(I - \mathcal{P}_m)u_1\| \leq 1$ since $\mathcal{P}_m$ is an orthogonal projector and $\|u_1\| = 1$. The new bound provided next for the left part of (4.1) does not exceed one. This result is based on optimization theory.

4.2. Analysis based on convex optimization. Let $L_{m+1} \in \mathbb{C}^{N \times (m+1)}$ be the matrix with columns $\alpha_1 u_1, v, A\,v, \ldots, A^{m-1}\,v$. Lemma 2.2 with $B = [v_1, A\,v_1, \ldots, A^{m-1}\,v_1]$ and $c = \alpha_1 u_1$ yields

$$\|(I - \mathcal{P}_m)\alpha_1 u_1\|^2 = \frac{1}{e_1^T (L_{m+1}^H L_{m+1})^{-1} e_1},$$

where it is assumed that L_{m+1} is of full rank so that $L_{m+1}^H L_{m+1}$ is nonsingular. As with the Krylov matrix, we can write L_{m+1} as $L_{m+1} = U\,D_\alpha\,W_{m+1}$ with $W_{m+1} \equiv [e_1, V_m(\lambda)]$.

Thus, in the normal case ($U^H U = I$), we have

$$\frac{\|(I-\mathcal{P}_m)u_1\|^2\,|\alpha_1|^2}{\|\alpha\|^2} = \frac{1}{e_1^T\left(Z_{m+1}^{(W)}(\beta)\right)^{-1} e_1}, \tag{4.2}$$

where $Z_{m+1}^{(W)}(\beta) \equiv W_{m+1}^H D_\beta W_{m+1}$ is in $\mathbb{C}^{(m+1)\times(m+1)}$ and $\beta_i = (|\alpha_i|/\|\alpha\|)^2$. Another formulation of $\|(I-\mathcal{P}_m)u_1\|$ is given next. Definitions of $V_m(\lambda)$, s_m, and other quantities may be found in Section 2.

COROLLARY 4.3. *If the matrix function $Z_{m+1}^{(W)}(\beta)$ is nonsingular with $\beta_1 > 0$, then*

$$e_1^T\left(Z_{m+1}^{(W)}(\beta)\right)^{-1} e_1 = \frac{1}{\beta_1} + \varphi_m(\widetilde{\beta}),$$

where $\widetilde{\beta} = (\beta_2, \ldots, \beta_N)^T$ and the function φ_m is independent of β_1. More precisely, we have

$$\|(I-\mathcal{P}_m)\,u_1\|^2 = \frac{1}{1+\beta_1\varphi_m(\widetilde{\beta})},$$

where

$$\varphi_m(\widetilde{\beta}) = s_m^H(\lambda_1)\left(\widetilde{V}_m^H\, D_{\widetilde{\beta}}\widetilde{V}_m\right)^{-1} s_m(\lambda_1),$$

in which $\widetilde{V}_m = V_m(\tilde{\lambda})$, where $\tilde{\lambda} = [\lambda_2, \lambda_3, \cdots, \lambda_N]^T$.

Proof. We write the matrix $Z_{m+1}^{(W)}(\beta)$ in the partitioned form

$$Z_{m+1}^{(W)}(\beta) = \begin{bmatrix} \beta_1 & \beta_1 s_m^H(\lambda_1) \\ \beta_1 s_m(\lambda_1) & V_m^H(\lambda)\,D_\beta V_m(\lambda)\end{bmatrix}.$$

First, using the matrix inversion in block form and then applying the Sherman-Morrison-Woodbury formula to the $(1,1)$ block, see, e.g., [21], leads to:

$$e_1^T\left(Z_{m+1}^{(W)}(\beta)\right)^{-1} e_1 = \frac{1}{\beta_1} + s_m^H(\lambda_1)\,\left(V_m^H(\lambda)\,D_\beta V_m(\lambda) - \beta_1 s_m(\lambda_1)\,s_m^H(\lambda_1)\right)^{-1} s_m(\lambda_1).$$

Note that $V_m(\lambda) = \begin{bmatrix} s_m^H(\lambda_1) \\ \widetilde{V}_m \end{bmatrix}$ and $V_m^H(\lambda)\,D_\beta V_m(\lambda) = \sum_{i=1}^N \beta_i\, s_m(\lambda_i)\, s_m(\lambda_i)^H$, hence

$$V_m^H(\lambda)\,D_\beta V_m(\lambda) - \beta_1 s_m(\lambda_1)\,s_m^H(\lambda_1) = \sum_{i=2}^N \beta_i\, s_m(\lambda_i)\, s_m(\lambda_i)^H = \widetilde{V}_m^H\, D_{\widetilde{\beta}}\widetilde{V}_m.$$

Therefore,

$$e_1^T\left(Z_{m+1}^{(W)}(\beta)\right)^{-1} e_1 = \frac{1}{\beta_1} + s_m^H(\lambda_1)(\widetilde{V}_m^H\, D_{\widetilde{\beta}}\widetilde{V}_m)^{-1} s_m(\lambda_1).$$

Applying the relation (4.2) results in

$$\|(I-\mathcal{P}_m)u_1\|^2 = \frac{1}{1+\beta_1\varphi_m(\widetilde{\beta})}. \qquad \square$$

Next, we state a bound for $\|(I-\mathcal{P}_m)\,u_1\|$ which slightly improves the one given in Theorem 4.2.

THEOREM 4.4. *If* $m < N$, $\|(I-\mathcal{P}_j)\ u_1\| \neq 0$ *for* $j \in \{1,\ldots,m\}$ *and the matrix* A *is normal then*

$$\|(I-\mathcal{P}_m)\ u_1\| \leq \frac{\|\widetilde{\alpha}\|\,\epsilon_1^{(m)}}{\sqrt{\|\widetilde{\alpha}\|^2\,(\epsilon_1^{(m)})^2+|\alpha_1|^2}} \leq 1,$$

where $\widetilde{\alpha} = (\alpha_2,\ldots,\alpha_N)^T$.

Proof. First observe that

$$\begin{aligned}\beta_1\varphi_m(\widetilde{\beta}) &= |\alpha_1|^2\,\varphi_m\left(|\alpha_2|^2,\ldots,|\alpha_N|^2\right)\\ &= \frac{|\alpha_1|^2}{\|\widetilde{\alpha}\|^2}\varphi_m\left(\frac{|\alpha_2|^2}{\|\widetilde{\alpha}\|^2},\ldots,\frac{|\alpha_N|^2}{\|\widetilde{\alpha}\|^2}\right) = \frac{|\alpha_1|^2}{\|\widetilde{\alpha}\|^2}\varphi_m(\gamma),\end{aligned}$$

where $\gamma = (\gamma_1,\ldots,\gamma_{N-1})$ with $\gamma_i = |\alpha_{i+1}|^2/\|\widetilde{\alpha}\|^2$. It is easy to see that $\gamma \in \widetilde{\Delta}_{N-1}$. Invoking Corollary 4.3, we obtain

$$\|(I-\mathcal{P}_m)\ u_1\|^2 \leq \frac{1}{1+\frac{|\alpha_1|^2}{\|\widetilde{\alpha}\|^2}\min_{\gamma\in\widetilde{\Delta}_{N-1}}\varphi_m(\gamma)}.$$

Using Theorem 2.1 with $M = N-1$, $\mu_j = \lambda_{j+1}$ and $\omega = \lambda_1$ leads to

$$\|(I-\mathcal{P}_m)\ u_1\|^2 \leq \frac{1}{1+\frac{|\alpha_1|^2}{\|\widetilde{\alpha}\|^2}\left(\epsilon_1^{(m)}\right)^{-2}} \leq 1,$$

and this completes the proof. □

5. Characterization of the polynomial of best approximation.

To characterize polynomials of best uniform approximation, we will follow the treatment of Lorentz [11, Chap. 2]. Our main goal is to derive two linear systems which characterize the optimal polynomial. These systems are fundamental in establishing the link with the optimization approach to be covered in the next section.

5.1. General context.

We begin this section with some additional notation. Let $\mathcal{C}(S)$ denote the space of complex or real continuous functions on a compact subset S of $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$). If $g \in \mathcal{C}(S)$, then the uniform norm of g is $\|g\|_\infty = \max_{z\in S}|g(z)|$. We set

$$\mathcal{E}(g,S) \equiv \{z : |g(z)| = \|g\|_\infty\,, z\in S\}\,.$$

A set $\Psi = \{\psi_1,\psi_2,\ldots,\psi_m\}$ from $\mathcal{C}(S)$ is a Chebyshev system if it satisfies the Haar condition, i.e., if each polynomial

$$p = a_1\psi_1 + a_2\psi_2 + \ldots + a_m\psi_m,$$

with the coefficients a_i not all equal to zero, has at most $(m-1)$ distinct zeros on S. The m-dimensional space E spanned by such a Ψ is called a Chebyshev space. We can verify that Ψ is a Chebyshev system if and only if for any m distinct points $z_i \in S$ the following determinant is nonzero:

$$\det(\psi_j(z_i)) := \begin{vmatrix}\psi_1(z_1) & \cdots & \psi_1(z_m)\\ \vdots & \ldots & \vdots\\ \psi_m(z_1) & \cdots & \psi_m(z_m)\end{vmatrix}.$$

ETNA
Kent State University
etna@mcs.kent.edu

Let $f \in \mathcal{C}(S)$, $f \notin \mathrm{E}$. We say that $q^* \in \mathrm{E}$ is a best approximation to f from E if $\|f - q^*\|_\infty \leq \|f - p\|_\infty, \forall p \in \mathrm{E}$. In other words,

$$\|f - q^*\|_\infty = \min_{p \in \mathrm{E}} \max_{z \in S} |f(z) - p(z)| \,.$$

Our first result exploits the following well-known characterization of the best uniform approximation [11]. An elegant characterization of best approximations is also given in [13].

THEOREM 5.1. *A polynomial q^* is a polynomial of best approximation to $f \in \mathcal{C}(S)$ from E if and only if there exist r extremal points, i.e., r points $z_1, z_2, \ldots, z_r \in \mathcal{E}(f - q^*, S)$, and r positive scalars β_i, $i = 1, \cdots, r$, such that $\sum_{l=1}^{r} \beta_l = 1$, with $r \leq 2m + 1$ in the complex case and $r \leq m + 1$ in the real case, which satisfy the equations:*

$$\sum_{l=1}^{r} \beta_l \left[f(z_l) - q^*(z_l) \right] \overline{\psi_j(z_l)} = 0, j = 1, \ldots, m. \tag{5.1}$$

It is important to make two remarks about this result. First, it can be applied to characterize the best uniform approximation over any finite subset σ of S, with at least $m + 1$ points. Second, the uniqueness of the polynomial of best approximation is guaranteed if E is a Chebyshev space [11, Chap. 2, p. 26]. Moreover, we have $r = m + 1$ if $S \subset \mathbb{R}$ and $m + 1 \leq r \leq 2m + 1$ if $S \subset \mathbb{C}$. This will be the case because we will deal with polynomials in $\mathbb{P}_m$.

The above result can be applied to our finite min-max approximation problem: $\min_{p \in \mathbb{P}_m^{(\omega)}} \max_{j=1,\ldots,M} |p(\mu_j)|$. This is the goal of the next section.

5.2. Application to the min-max problem. Let σ denote the set of the complex numbers $\mu_1, \ldots, \mu_M$ and let $\mathbb{Q}_m^{(\omega)} (m < M)$ denote the set of all polynomials of degree not exceeding m with value zero at ω. Let $(\psi_j)_{j=1}^m$ be the set of polynomials of $\mathbb{Q}_m^{(\omega)}$ defined by $\psi_j(z) = z^j - \omega^j$. Our problem corresponds to taking $f \equiv \mathbf{1}$ and $\mathrm{E} = \mathbb{Q}_m^{(\omega)}$. This yields:

$$\min_{p \in \mathbb{P}_m^{(\omega)}} \max_{\mu \in \sigma} |p(\mu)| = \min_{q \in \mathbb{Q}_m^{(\omega)}} \max_{\mu \in \sigma} |1 - q(\mu)| \equiv \|f - q^*\|_\infty \,.$$

According to the remarks made following Theorem 5.1, the polynomial of best approximation $q^* \in \mathbb{Q}_m^{(\omega)}$ for $f \equiv \mathbf{1}$ (with respect to σ) exists and is unique.

The following corollary, which gives an equivalent formulation of the relation (5.1), can now be stated. This corollary will lead to auxiliary results from which the maximum and the polynomial of best approximation can be characterized.

COROLLARY 5.2. *The polynomial $q^*(z) = a_1^* \psi_1(z) + a_2^* \psi_2(z) + \ldots + a_m^* \psi_m(z)$ is the best approximation polynomial for the function $f(z) = 1$ on σ from $\mathbb{Q}_m^{(\omega)}$ if and only if there exist r extremal points, i.e., r points $\mu_1, \mu_2, \ldots, \mu_r \in \mathcal{E}(f - q^*, S)$, and r positive scalars β_i, $i = 1, \cdots, r$, such that $\sum_{l=1}^{r} \beta_l = 1$, with $r \leq 2m + 1$ in the complex case and $r \leq m + 1$ in the real case, satisfying*

$$t_1^* + \sum_{j=2}^{m+1} t_j^* \mu_l^{j-1} = \varepsilon_l \frac{1}{\sqrt{\delta^*}}, l = 1, \ldots, r; \tag{5.2}$$

with $\delta^* = \|f - q^*\|_\infty^2$, $t_{j+1}^* = -\dfrac{a_j^*}{\delta^*}$, $j = 1, \ldots, m$, *and* $t_1^* = \dfrac{1}{\delta^*} - \sum_{j=2}^{m+1} t_j^* \omega^{j-1}$ *and*

$$\sum_{l=1}^{r} \beta_l \varepsilon_l = \sqrt{\delta^*}, \tag{5.3}$$

$$\sum_{l=1}^{r} \beta_l \varepsilon_l (\overline{\mu}_l^j - \overline{\omega}^j) = 0, \quad j = 1, \ldots, m; \tag{5.4}$$

where $\varepsilon_l = \operatorname{sgn}(1 - q^*(\mu_l))$.

Proof. Let $\delta^* = \|f - q^*\|_\infty^2$. Then $\mu_l \in \mathcal{E}(f - q^*, S)$ is equivalent to $|1 - q^*(\mu_l)| = \|f - q^*\|_\infty = \sqrt{\delta^*}$. Without loss of generality, we can assume the r extremal points are the r first items of σ. According to the above definition of ε_l, the polynomial q^* satisfies the following interpolation conditions:

$$1 - q^*(\mu_l) = \sqrt{\delta^*} \varepsilon_l \text{ for } l = 1, \ldots, r. \tag{5.5}$$

Setting $t_1^* = 1/\delta^* - \sum_{j=2}^{m+1} t_j^* \omega^{j-1}$ and $t_{j+1}^* = -a_j^*/\delta^*$, $j = 1, \ldots, m$, we obtain (5.2).

Equation (5.5) shows that $\sum_{l=1}^{r} \beta_l \varepsilon_l \overline{\psi_j(\mu_l)} = 0$, $j = 1, \ldots, m$ is a restatement of (5.1). We then have

$$\sum_{l=1}^{r} \beta_l \varepsilon_l \overline{q(\mu_l)} = 0 \quad \text{for all polynomials } q \in \mathbb{Q}_m^{(\omega)}. \tag{5.6}$$

Furthermore, from (5.5) we have the relation $\sqrt{\delta^*}\beta_l = \beta_l \overline{\varepsilon_l} f(\mu_l) - \beta_l \overline{\varepsilon_l} q^*(\mu_l)$, $l = 1, \ldots, r$. To see that $\sum_{l=1}^{r} \beta_l = 1$ is equivalent to (5.3), it suffices to sum the terms in this relation and apply the conjugate of (5.6). □

As a remark, we may transform the interpolation conditions (5.5) to $\pi_m(\mu_j) = \varepsilon_j$ for all $j = 1, \ldots, r$, with $\pi_m(z) = (1 - q^*(z))/\sqrt{\delta^*}$. Then π_m can be written in the form of the Lagrange interpolation formula

$$\pi_m(z) = \sum_{j=1}^{m+1} \varepsilon_j l_j^{(m+1)}(z), \quad \text{with} \quad l_j^{(m+1)}(z) = \prod_{k=1, k \neq j}^{m+1} \frac{z - \mu_k}{\mu_j - \mu_k}.$$

Note that $l_j^{(m+1)}(z)$ is the jth Lagrange interpolating polynomial of degree m associated with $\{\mu_1, \ldots, \mu_{m+1}\}$. Finally, recalling that $\pi_m(\omega) = 1/\sqrt{\delta^*}$, we obtain

$$\sum_{j=1}^{m+1} \varepsilon_j l_j^{(m+1)}(\omega) = \frac{1}{\sqrt{\delta^*}}. \tag{5.7}$$

A consequence of this is that:

$$\min_{p \in \mathbb{P}_m^{(\omega)}} \max_{\mu \in \sigma} |p(\mu)| = \|f - q^*\|_\infty = \sqrt{\delta^*} = \frac{1}{\sum_{j=1}^{m+1} \varepsilon_j l_j^{(m+1)}(\omega)}.$$

6. Proof of the main result. In this section, we will show that

$$\max_{\beta \in \widetilde{\Delta}_M} F_{m,\omega}(\beta) = \left(\epsilon_1^{(m)}(\omega)\right)^2,$$

where $F_{m,\omega}(\beta)$ is defined in (2.4) and

$$\epsilon_1^{(m)}(\omega) = \min_{p\in\mathbb{P}_m^{(\omega)}} \max_{j=1,\dots,M} |p(\mu_j)|\,. \tag{6.1}$$

The proof will be based on applying the Karush-Kuhn-Tucker (KKT) optimality conditions to our convex maximization problem. We begin by establishing the following lemma which shows a few important properties of the function $F_{m,\omega}$. Since there is no ambiguity, we will simplify notation by writing V_{m+1} for $V_{m+1}(\mu)$ in the remainder of this section.

LEMMA 6.1. *Let $\widetilde{\Delta}_M \subset \Delta_M$ be the domain of definition of the function $F_{m,\omega}$ in (2.4). Then the following properties hold:*

1. *$F_{m,\omega}$ is a concave function defined on the convex set $\widetilde{\Delta}_M$.*
2. *$F_{m,\omega}$ is differentiable at $\beta \in \widetilde{\Delta}_M$ and we have*

$$\frac{\partial F_{m,\omega}}{\partial \beta_j}(\beta) = -\,(F_{m,\omega}(\beta))^2\left|e_j^T V_{m+1} t\right|^2,$$

where $t = [t_1, t_2, \dots, t_{m+1}]^T$ is such that $V_{m+1}^H D_\beta V_{m+1} t = s_{m+1}(\omega)$. Moreover, we have

$$\sum_{i=1}^M \beta_i \frac{\partial F_{m,\omega}(\beta)}{\partial \beta_i} = F_{m,\omega}(\beta). \tag{6.2}$$

Proof. We will begin by proving the first property. Let r be a real positive scalar. Since $D_{r\beta} = r\, D_\beta$, then

$$F_{m,\omega}(r\beta) = \frac{1}{s_{m+1}^H(\omega)(r(V_{m+1}^H D_\beta V_{m+1}))^{-1} s_{m+1}(\omega)} = r\, F_{m,\omega}(\beta).$$

Thus $F_{m,\omega}$ is homogeneous of degree 1. Let $\beta, \beta' \in \widetilde{\Delta}_M$ and $0 \le r \le 1$. It is easy to see that $r\beta + (1-r)\beta' \in \widetilde{\Delta}_M$. We now take $x = s_{m+1}(\omega)$, $G_1 = V_{m+1}^H D_{r\beta} V_{m+1}$ and $G_2 = V_{m+1}^H D_{(1-r)\beta'} V_{m+1}$, in the following version of Bergstrom's inequality given in [12, p. 227]:

$$\left(x^H\,(G_1+G_2)^{-1}\,x\right)^{-1} \ge \left(x^H G_1^{-1} x\right)^{-1} + \left(x^H G_2^{-1} x\right)^{-1},$$

where G_1 and G_2 are positive definite Hermitian matrices. Then from the homogeneity of $F_{m,\omega}$ it follows that

$$F_{m,\omega}(r\beta + (1-r)\beta') \ge r\, F_{m,\omega}(\beta) + (1-r)\, F_{m,\omega}(\beta').$$

Hence, $F_{m,\omega}$ is concave.

Next we establish the second part. Let us define $Z_{m+1}(\beta) = V_{m+1}^H D_\beta V_{m+1} \in \mathbb{C}^{m+1,m+1}$. We have $Z_{m+1}^H(\beta) = Z_{m+1}(\beta)$ and $F_{m,\omega}(\beta) = \left[s_{m+1}^H(\omega) Z_{m+1}^{-1}(\beta) s_{m+1}(\omega)\right]^{-1}$ for $\beta \in \widetilde{\Delta}_M$. Clearly, $F_{m,\omega}$ is differentiable at $\beta \in \widetilde{\Delta}_M$.

By using the derivative of the inverse of the matrix function, we have

$$\frac{\partial Z_{m+1}^{-1}(\beta)}{\partial \beta_i} = -Z_{m+1}^{-1}(\beta)\frac{\partial Z_{m+1}(\beta)}{\partial \beta_i}\, Z_{m+1}^{-1}(\beta).$$

It follows that

$$s_{m+1}^H(\omega)\frac{\partial Z_{m+1}^{-1}(\beta)}{\partial \beta_i} s_{m+1}(\omega) = -t^H V_{m+1}^H e_i e_i^T V_{m+1} t,$$

where t is such that $Z_{m+1}(\beta)t = s_{m+1}(\omega)$. As a consequence,

$$\frac{\partial F_{m,\omega}}{\partial \beta_i}(\beta) = -\frac{s_{m+1}^H(\omega)\dfrac{\partial Z_{m+1}^{-1}(\beta)}{\partial \beta_i}s_{m+1}(\omega)}{\left(s_{m+1}^H(\omega)Z_{m+1}^{-1}(\beta)s_{m+1}(\omega)\right)^2} = \frac{t^H V_{m+1}^H e_i e_i^T V_{m+1} t}{\left(s_{m+1}^H(\omega)Z_{m+1}^{-1}(\beta)s_{m+1}(\omega)\right)^2}.$$

We can write more succinctly

$$\frac{\partial F_{m,\omega}}{\partial \beta_i}(\beta) = (F_{m,\omega}(\beta))^2\, |e_i^T V_{m+1} t|^2.$$

The equality (6.2) follows from a simple algebraic manipulation. Indeed, we have

$$\begin{aligned}\sum_{i=1}^{M} \beta_i \frac{\partial F_{m,\omega}(\beta)}{\partial \beta_i} &= (F_{m,\omega}(\beta))^2 \sum_{i=1}^{M} t^H V_{m+1}^H (\beta_i e_i e_i^T) V_{m+1}\, t \\ &= (F_{m,\omega}(\beta))^2 \left(t^H V_{m+1}^H D_\beta V_{m+1}\, t\right) = F_{m,\omega}(\beta). \qquad \square\end{aligned}$$

With the lemma proved, we now proceed with the proof of the theorem. Let us consider the characterization of the solution of the convex maximization problem $\max_{\beta\in\widetilde{\Delta}_M} F_{m,\omega}(\beta)$ by using the standard KKT conditions. We begin by defining the functions $g_i(\beta) = -\beta_i$ and $g(\beta) = e^T\beta - 1$. Thus $\beta \in \Delta_M$ means that $g(\beta) = 0$ and $g_i(\beta) \le 0$. So the Lagrangian function is formulated as

$$\mathcal{L}_m(\beta, \delta, \eta) = F_{m,\omega}(\beta) - \delta g(\beta) - \sum_{i=1}^{M} \eta_i g_i(\beta),$$

where $\delta \in \mathbb{R}$ and $\eta = (\eta_1, \ldots, \eta_M) \in \mathbb{R}^M$. Note that the functions g, g_i are convex (affine functions) and by Lemma 6.1, the objective function $-F_m$ is also convex. It follows that this maximization problem can be viewed as a constrained convex optimization problem. Thus according to the KKT conditions [4], if $F_{m,\omega}(\beta)$ has a local maximizer β^* in $\widetilde{\Delta}_M$ then there exist Lagrange multipliers $\delta^*, \eta^* = (\eta_1^*, \ldots, \eta_M^*)$ such that $(\beta^*, \delta^*, \eta^*)$, satisfy the following conditions:

(i) $\dfrac{\partial \mathcal{L}_m}{\partial \beta_j}(\beta^*, \delta^*, \eta^*) = 0$, for $j = 1, \ldots, M$;

(ii) $g(\beta^*) = 0$ and $(g_j(\beta^*) \le 0$ and $\eta_j^* \ge 0$ for all $j = 1, \ldots, M)$;

(iii) $\eta_j^* g_j(\beta^*) = 0$ for all $j = 1, \ldots, M$.

As β^* is in $\widetilde{\Delta}_M$, we have at least $m+1$ components of β^* which are nonzero. Thus, there exist $m+\kappa$, $(\kappa \ge 1)$ components of β^* which we label $\beta_1^*, \beta_2^*, \ldots, \beta_{m+\kappa}^*$, for simplicity, such that $\beta_j^* \ne 0$ for all $j = 1, \ldots, m+\kappa$. The complementarity condition (iii), yields $\eta_j^* = 0$ for all $j = 1, \ldots, m+\kappa$ and $\eta_j^* > 0$ for all $j = m+\kappa+1, \ldots, M$. Hence the condition (i) can be re-expressed as

$$\begin{cases} \dfrac{\partial F_{m,\omega}}{\partial \beta_j}(\beta^*) = \delta^* & \text{for } j = 1, \ldots, m+\kappa, \text{ and} \\ \dfrac{\partial F_{m,\omega}}{\partial \beta_j}(\beta^*) = \delta^* - \eta_j^* & \text{for } j = m+\kappa+1, \ldots, M. \end{cases} \tag{6.3}$$

Again by using the formula of $\dfrac{\partial F_{m,\omega}}{\partial \beta_j}(\beta^*)$ given in the lemma, the relations in (6.3) become

$$F_{m,\omega}(\beta^*)^2 \left|e_j^T V_{m+1} t^*\right|^2 = \delta^* \quad \text{for} \quad j = 1, \ldots, m+\kappa, \tag{6.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

and

$$F_{m,\omega}(\beta^*)^2 \left|e_j^T V_{m+1} t^*\right|^2 = \delta^* - \eta_j^* \quad \text{for} \quad j = m+\kappa+1, \ldots, M. \tag{6.5}$$

Note that t^* is such that

$$V_{m+1}^H D_{\beta^*} V_{m+1} t^* = s_{m+1}(\omega). \tag{6.6}$$

Now by observing that $\beta_j^* = 0$ for all $j = m+\kappa+1, \ldots, M$, $\sum_{j=1}^{m+\kappa} \beta_j^* = 1$, and by using the first part of (6.3), equation (6.2) of the lemma shows that $F_{m,\omega}(\beta^*) = \delta^*$. The remaining part of the proof is devoted to establishing the same formulas given in the theorem which characterize the polynomial of best approximation. In view of (6.4), we then have

$$e_j^T V_{m+1} t^* = \varepsilon_j \frac{1}{\sqrt{\delta^*}} \text{ for all } j = 1, \ldots, m+\kappa. \tag{6.7}$$

Here we have written $\varepsilon_j = e^{\iota\theta_j}$ in the complex case and $\varepsilon_j = \pm 1$ in the real case.

Combining $\left[s_{m+1}^H(\omega)\left(V_{m+1}^H D_{\beta^*} V_{m+1}\right)^{-1} s_{m+1}(\omega)\right]^{-1} = F_{m,\omega}(\beta^*)$ with (6.6), we obtain $s_{m+1}^H(\omega)\, t^* = 1/\delta^*$. On the other hand, the optimal solutions β_j^* and the numbers ε_j can be derived from (6.6). Indeed, using (6.6) and (6.7), we have

$$e_j^T D_{\beta^*} V_{m+1} t^* = \beta_j^* \varepsilon_j \frac{1}{\sqrt{\delta^*}} \text{ for all } j = 1, \ldots, m+\kappa.$$

Therefore, by applying V_{m+1}^H we have $\sum_{j=1}^{m+\kappa} \beta_j^* \varepsilon_j = \sqrt{\delta^*}$ and $\sum_{j=1}^{m+\kappa} \beta_j^* \varepsilon_j \overline{\mu}_j^l = \sqrt{\delta^*}\overline{\omega}^l = \sum_{j=1}^{m+\kappa} \beta_j^* \varepsilon_j \overline{\omega}^l$ for $l = 1, \ldots, m$. Hence, we find that $\sum_{j=1}^{m+\kappa} \beta_j^* \varepsilon_j (\overline{\mu_j^l - \omega^l}) = 0$ for $l = 0, \ldots, m$. The relations (5.2), (5.4) and (5.3) in Corollary 5.2 are all satisfied, with $f \equiv \mathbf{1}$, $z_j = \mu_j$, $\psi_j(z) = z^j - \omega^j$ and $r = m+\kappa$. It follows that the Lagrange multiplier δ^* is unique and is the same as in the previous section. As a consequence we have $\delta^* = F_{m,\omega}(\beta^*) = \|\mathbf{1} - q^*\|_\infty^2$ and (2.5) is established.

7. Concluding remarks. For normal matrices, we have established the equivalence of the approximation theory approach and the optimization approach to solving a min-max problem that arises in convergence studies of the GMRES and Arnoldi methods. Our work reveals a strong connection between the two points of view. Because of their convenient properties, the KKT conditions allow us to completely characterize the residual bounds at each step for both methods; and the KKT conditions give more precise information about the extremal points than approximation theory. We also point out the importance of the form of the KKT conditions. Only the non-active part (6.4) is needed to prove equivalence of the approximation theory and optimization approaches; while for the GMRES method, for example, we infer from the active part (6.5) that

$$\eta_j^* = \delta^* - F_{m,\omega}(\beta^*)^2 \left|e_j^T V_{m+1}(\lambda) t^*\right|^2 > 0 \text{ for } j = m+\kappa+1, \ldots, M,$$

so that

$$\left|e_j^T V_{m+1}(\lambda) t^*\right| < \frac{1}{\sqrt{\delta^*}} \text{ for } j = m+\kappa+1, \ldots, M.$$

This shows that the extremal points can be characterized by

$$\frac{1}{\sqrt{\delta^*}} = \left|e_i^T V_{m+1}(\lambda) t^*\right| = \max_{1 \le j \le M} \left|e_j^T V_{m+1}(\lambda) t^*\right| \text{ for } i = 1, \ldots, m+\kappa.$$

The connections established in this paper provide new insights into the different ways in which residual bounds can be derived. We hope that the development of the optimization approach will pave the way for extensions beyond the case of normal matrices.

Acknowledgment. We would like to thank the anonymous referees and the editor for their valuable comments and suggestions.

REFERENCES

[1] M. BELLALIJ, Y. SAAD, AND H. SADOK, *On the convergence of the Arnoldi process for eigenvalue problems*, Technical Report UMSI-2007-12, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN, 2007.

[2] M. EIERMANN AND O. G. ERNST, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.

[3] M. EIERMANN, O. G. ERNST, AND O. SCHNEIDER, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 345–357.

[4] R. FLETCHER, *Practical Methods of Optimization*, 2nd Edition, Wiley, Chichester, 1987.

[5] A. GREENBAUM AND L. GURVITS, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.

[6] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.

[7] I. C. F. IPSEN, *Expressions and bounds for the GMRES residuals*, BIT, 40 (2000), pp. 524–535.

[8] W. JOUBERT, *A robust GMRES-based adaptive polynomial preconditioning algorithm for nonsymmetric linear systems*, SIAM J. Sci. Comput., 15 (1994), pp. 427–439.

[9] J. LIESEN, M. ROŽLOZNÍK AND Z. STRAKOŠ, *Least squares residuals and minimal residual methods*, SIAM J. Sci. Comput., 23 (2002), pp. 1503–1525.

[10] J. LIESEN AND P. TICHY, *The worst-case GMRES for normal matrices*, BIT, 44 (2004), pp. 79–98.

[11] G. G. LORENTZ, *Approximation of Functions*, 2nd Edition, Chelsea Publishing Company, New York, 1986.

[12] J. R. MAGNUS AND H. NEUDECKER, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd Edition, Wiley, Chichester, 1999.

[13] T. J. RIVLIN AND S. SHAPIRO, *A unified approach to certain problems of approximation and minimization*, J. Soc. Indust. Appl. Math., 9 (1961), pp. 670–699.

[14] Y. SAAD AND M. H. SCHULTZ, *GMRES: A Generalized Minimal Residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.

[15] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, Halstead Press, New York, 1992.

[16] Y. SAAD, *Iterative Methods for Solving Sparse Linear Systems*, SIAM publications, Philadelphia, 2003.

[17] H. SADOK, *Méthodes de projections pour les systèmes linéaires et non linéaires*, Habilitation thesis, University of Lille1, Lille, France, 1994.

[18] H. SADOK, *Analysis of the convergence of the minimal and the orthogonal residual methods*, Numer. Algorithms, 40 (2005), pp. 101–115.

[19] G. W. STEWART, *Collinearity and least squares regression*, Statist. Sci., 2 (1987), pp. 68–100.

[20] I. ZAVORIN, D. O'LEARY, AND H. ELMAN. *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.

[21] F. ZHANG, *Matrix Theory*, Springer Verlag, New York, 1999.