MODEL REDUCTION USING THE VOROBYEV MOMENT PROBLEM

ZDENĚK STRAKOŠ*

Abstract. Given a nonsingular complex matrix $A \in \mathbb{C}^{N \times N}$ and complex vectors v and w of length N, one may wish to estimate the quadratic form $w^* A^{-1} v$, where w^* denotes the conjugate transpose of w. This problem appears in many applications, and Gene Golub was the key figure in its investigations for decades. He focused mainly on the case A Hermitian positive definite (HPD) and emphasized the relationship of the algebraically formulated problems with classical topics in analysis - moments, orthogonal polynomials and quadrature. The essence of his view can be found in his contribution *Matrix Computations and the Theory of Moments*, given at the International Congress of Mathematicians in Zürich in 1994. As in many other areas, Gene Golub has inspired a long list of coauthors for work on the problem, and our contribution can also be seen as a consequence of his lasting inspiration.

In this paper we will consider a general mathematical concept of *matching moments model reduction*, which as well as its use in many other applications, is the basis for the development of various approaches for estimation of the quadratic form above. The idea of model reduction via matching moments is well known and widely used in approximation of dynamical systems, but it goes back to Stieltjes, with some preceding work done by Chebyshev and Heine. The *algebraic* moment matching problem can for A HPD be formulated as a variant of the Stieltjes moment problem, and can be solved using Gauss-Christoffel quadrature. Using the operator moment problem suggested by Vorobyev, we will generalize model reduction based on matching moments to the non-Hermitian case in a straightforward way. Unlike in the model reduction literature, the presented proofs follow directly from the construction of the Vorobyev moment problem.

Key words. Matching moments, model reduction, Krylov subspace methods, conjugate gradient method, Lanczos method, Arnoldi method, Gauss-Christoffel quadrature, scattering amplitude.

AMS subject classifications. 15A06, 65F10, 65F25, 65G50.

1. Introduction. Given a nonsingular complex matrix $A \in \mathbb{C}^{N \times N}$ and complex vectors v and w of length N, estimates of the quadratic form $w^* A^{-1} v$, where w^* denotes the conjugate transpose of w, appear in many applications. If A is large, it seems natural to construct the estimates by projections of the original problem onto subspaces of small dimensions, with interpretation of the solution of the projected problem as the estimate (or bound, where appropriate) for the desired unknown quantity. Here the Krylov subspaces

$$\mathcal{K}_j(A, v) \equiv span\{v, Av, \dots, A^{j-1}v\}$$

come very naturally into play, since they tend to accumulate dominant information of A with respect to v. For description of the Krylov subspace methods as projection methods, see, e.g., [7, 11], [36, Chapter 5].

Asuming A is Hermitian positive definite (HPD) and using its spectral decomposition, different approaches were developed by relating the problem to the Gauss-Christoffel quadrature. For descriptions of the remarkable and rich work and achievements in that direction, we refer to [22, 24, 27, 12, 25], to the survey paper [34] and parts II. and IV. of the book [23], with Commentaries given by Anne Greenbaum

[1]Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 18207 Prague, and Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, emails: strakos@cs.cas.cz. The work was supported by the GAAS grant IAA100300802 and by the Institutional Research Plan AV0Z10300504.

and Walter Gautschi. A central role in the development was from the late Sixties played by Gene Golub, who has deeply influenced and inspired many others for work on the applications of the moment problem, and who was always looking for new connections [19].

In the paper [44] it was shown that the Gauss-Christoffel quadrature estimates are closely related to results which trace back to the original Hestenes and Stiefel paper on conjugate gradients (CG) [28]. The desired value $v^*A^{-1}v$ can be very efficiently estimated using the scalar quantities available in the CG algorithm. Moreover, the paper [44] shows that this estimate *is numerically stable* (a partial proof of numerical stability of the quadrature-based estimates has been given already in [27]). This has further been used in construction of stopping criteria for iterative solvers, see [5, 3, 4], the book [33], the survey paper [34] and the forthcoming paper [30]. The fundamental mathematical idea behind such estimates can be formulated as *matching moments model reduction*.

In this paper we present a description of the matching moments model reduction based on the *Vorobyev moment problem*, see [47], which allows us to exploit the direct and very natural relationship with the projection processes interpretation of Krylov subspace methods. The Vorobyev moment problem has been used, slightly generalized and popularized by Claude Brezinski [7]. Unfortunately, it still does not seem to be well-known.

Projection of the original problem to the subspace of small dimension can be viewed as a *reduction of the original model* represented by A, v and w. Reduced order modeling plays a central role in approximation of large-scale dynamical systems. Techniques based on Krylov subspaces have been used in that area for decades, see, e.g., the recent very thorough monograph [2] and the very nice survey paper [6], where one can find references to substantial work of many other authors. In [2, Chapter 10, p. 314], approximation of linear dynamical systems by moment matching is described as one of the three uses of Krylov subspace methods, in addition to iterative solution of $Ax = b$ and approximation of the eigenvalues of A, and it is treated in detail in Chapter 11 of that book. It should be emphasized that Krylov subspace methods can be viewed as much more than just *tools* for model reduction. Many Krylov subspace methods by their nature are actually model reductions based on matching moments. Such a view naturally complements the description using the projection processes framework, see [7].

Section 2 will very briefly outline the way from the Stieltjes moment problem to the Gauss-Christoffel quadrature. Section 3 will describe the matching moment property of the Lanczos and CG methods, and show how it can be formulated using the Vorobyev moment problem. Section 4 will prove the matching moment property of the non-Hermitian Lanczos and Arnoldi methods. The paper ends with concluding remarks. Appendix contains comments on the history which might be of interest independently of the rest of the paper.

2. Stieltjes moment problem and Gauss-Christoffel quadrature. In the Stieltjes moment problem [40, 42, 41, 43], a sequence of numbers ξ_k, $k = 0, 1, \dots$, is given and a non-decreasing distribution function $\omega(\lambda)$, $\lambda \geq 0$, is sought such that the Riemann-Stieltjes integral satisfies

$$\int_0^\infty \lambda^k d\omega(\lambda) = \xi_k, \quad k = 0, 1, \dots . \tag{2.1}$$

Here $\int_0^\infty \lambda^k d\omega(\lambda)$ represents the k-th moment of the distribution function $\omega(\lambda)$. For the definition and basic properties of the Riemann-Stieltjes integral see, e.g., [10, Section 1.6.5], and for a general orthogonal polynomial context see, e.g., [20]. This problem has a simple mechanical interpretation – it aims to find the distribution of positive mass on the half line $\lambda \geq 0$ given the (generalized) mechanical moments of the mass distribution with respect to 0 (here for $k = 1$ the moment divided by the total mass is the center of mass, $k = 2$ gives the center of inertia). A closely related problem was posed and studied by Chebyshev, Heine and Markov. Nice descriptions of the origins of the moment problem, mechanical motivations of Stieltjes and statistical motivations of Chebyshev, as well as the subsequent developments can be found, e.g., in [39, 1, 18]. A summary of some related later achievements motivated by oscillations of mechanical systems is given in [16], in particular in Chapter II, § 1, and in Appendix 2, which is devoted to the relationship with the work of Stieltjes on continued fractions.

In this and in the next section we briefly recall some main ideas which link the Stieltjes moment problem with the Gauss-Christoffel quadrature, and the Lanczos and CG methods. Technical details and proofs can be found in [44, 34].

Let the distribution function $\omega(\lambda)$ have N points of increase $0 < \lambda_1 < \cdots < \lambda_N$ with the corresponding positive weights $\omega_1, \ldots, \omega_N$, $\sum_{\ell=1}^N \omega_\ell \equiv 1$. Then the first $2N$ moments are given by

$$\int_0^\infty \lambda^k \, d\omega(\lambda) = \sum_{\ell=1}^{N} \omega_\ell \, \{\lambda_\ell\}^k \equiv \xi_k \, , \quad k = 0, 1, \ldots, 2N-1 \, . \tag{2.2}$$

For a given n between 1 and $N-1$ (the cases $n = 0$ and $n = N$ are trivial), one can look for a nondecreasing distribution function $\omega^{(n)}$ with n points of increase $0 < \theta_1^{(n)} < \cdots < \theta_n^{(n)}$ and positive weights $\omega_1^{(n)}, \ldots, \omega_n^{(n)}$, $\sum_{\ell=1}^n \omega_\ell^{(n)} \equiv 1$, such that its moments match the maximal number $2n$ of moments (2.2) given by $\omega(\lambda)$,

$$\int_0^\infty \lambda^k \, d\omega(\lambda) = \sum_{\ell=1}^{n} \omega_\ell^{(n)} \{\theta_\ell^{(n)}\}^k \, , \quad k = 0, 1, \ldots, 2n-1 \, . \tag{2.3}$$

This means that the Rieman-Stieltjes integral, determined by the distribution function $\omega(\lambda)$, is for *any* polynomial up to degree $2n-1$ given by the weighted sum of the polynomial values at the n points $\theta_\ell^{(n)}$ with the corresponding weights $\omega_\ell^{(n)}$. Equivalently, (2.3) is nothing but the n-point Gauss-Christoffel quadrature, see [10, sec. 2.7] and [18] for a basic description and a comprehensive survey of related topics. Let

$$p_1(\lambda) \equiv 1, \, p_2(\lambda), \, \ldots \, , p_{n+1}(\lambda)$$

be the first $n+1$ orthonormal polynomials corresponding to the inner product

$$(\phi, \psi) \equiv \int_0^\infty \phi(\lambda) \, \psi(\lambda) \, d\omega(\lambda) \tag{2.4}$$

determined by the distribution function $\omega(\lambda)$. Then, denoting

$$P_n(\lambda) = (p_1(\lambda), \ldots, p_n(\lambda))^T,$$

we get

$$\lambda P_n(\lambda) = T_n P_n(\lambda) + \delta_{n+1} p_{n+1}(\lambda) e_n \, , \tag{2.5}$$

which represents the matrix formulation of the Stieltjes recurrence for the orthogonal polynomials determined by (2.4). The recurrence coefficients form the Jacobi matrix

$$T_n \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_n \\ & & \delta_n & \gamma_n \end{pmatrix}, \quad \delta_\ell > 0\,, \ell = 2, \ldots, n\,. \tag{2.6}$$

The basic result about the Gauss-Christoffel quadrature states, see also [35, Section 4] and the references given there, that the nodes of the n-point quadrature are equal to the roots of $p_{n+1}(\lambda)$, i.e. the eigenvalues of T_n. The corresponding weights are given by

$$\omega_\ell^{(n)} = \frac{1}{p'_{n+1}(\theta_\ell^{(n)})} \int_0^\infty \frac{p_{n+1}(\lambda)}{\lambda - \theta_\ell^{(n)}}\, d\omega(\lambda)\,, \tag{2.7}$$

or, equivalently, by the sizes of the squared first entries of the corresponding normalized eigenvectors of T_n.

The n-point Gauss-Christoffel quadrature can be viewed as a matching moments model reduction, where the original model is represented by the distribution function $\omega(\lambda)$ with the N points of increase $\lambda_1, \ldots, \lambda_N$, and the reduced model by the distribution function $\omega^{(n)}(\lambda)$ with the n points of increase $\theta_1^{(n)}, \ldots, \theta_n^{(n)}$. The reduced model matches the first $2n$ moments given by the original model, see (2.3).

3. Matrix formulation of the moment problem for A Hermitian positive definite. In the following we will use some basic facts about the Lanczos and CG methods, and about the Arnoldi method, which can be found, e.g., in [44, 34, 36]. A well presented summary in [2, Chapter 10] might be preferred by dynamical systems oriented readers. While the Gauss-Christoffel quadrature describes the matching moments model reduction in *polynomial language*, the following formulation translates the description to matrix algebra.

Consider a linear algebraic system $Ax = b$ with a HPD matrix $A \in \mathbb{C}^{N\times N}$ and an initial vector x_0, giving the initial residual $r_0 = b - Ax_0$ and the Lanczos initial vector $v \equiv v_1 = r_0/\|r_0\|$. Consider the nondecreasing distribution function $\omega(\lambda)$ with the points of increase λ_ℓ equal to the eigenvalues of A and the weights ω_ℓ equal to sizes of the squared components of v_1 in the corresponding invariant subspaces. For simplicity of exposition we assume, following the notation in Section 2, that the eigenvalues of A are distinct and all weights are nonzero. With this setting, the moments (2.2) of the distribution function $\omega(\lambda)$ can be expressed in matrix language as

$$\int_0^\infty \lambda^k\, d\omega(\lambda) = \sum_{\ell=1}^{N} \omega_\ell \,\{\lambda_\ell\}^k = v_1^* A^k v_1\,, \tag{3.1}$$

and, analogously, using the spectral decomposition of the Jacobi matrix T_n and the well-known fact that all its eigenvectors have nonzero first components,

$$\sum_{\ell=1}^{n} \omega_\ell^{(n)} \{\theta_\ell^{(n)}\}^k = e_1^T T_n^k e_1\,. \tag{3.2}$$

It is easy to see that the Lanczos *algorithm* [31, 32] applied to A with v_1 gives in the nth step the Jacobi matrix T_n in (2.6), i.e.

$$AV_n = V_n T_n + \delta_{n+1} v_{n+1} e_n^T \,, \tag{3.3}$$

where V_n represents the matrix with orthonormal columns $v_1, \ldots, v_n$,

$$V_n = (v_1, \ldots, v_n) \,, \quad V_n^* V_n = I \,, \quad V_n^* v_{n+1} = 0. \tag{3.4}$$

The Lanczos *method* computes the eigenvalues of T_n and takes them as approximations to the eigenvalues of A.

Summarizing, let the original model be represented by the matrix A and the initial vector v_1. Then the Lanczos algorithm computes in steps 1 to n the model reduction of A with v_1 to T_n with e_1 such that the reduced model matches the first $2n$ moments of the original model, i.e.,

$$v_1^* A^k v_1 = e_1^T T_n^k e_1 \,, \quad k = 0, 1, \ldots, 2n-1 \,. \tag{3.5}$$

With A HPD, we can extend the considerations presented above to the CG method. Using the orthonormal basis V_n of the Krylov subspace $\mathcal{K}_n(A, v_1)$ determined by the Lanczos algorithm, one can write for the approximation $x_n = x_0 + V_n y_n$ generated by the CG method

$$0 = V_n^* r_n = V_n^*(b - Ax_n) = V_n^*(b - Ax_0 - AV_n y_n) \,,$$

which gives, using $V_n^* A V_n = T_n$,

$$x_n = x_0 + V_n y_n \,, \quad T_n y_n = \|r_0\| e_1 \,. \tag{3.6}$$

The nth CG approximation x_n can be considered as a result of the model reduction from $Ax = b$ to $T_n y_n = \|r_0\| e_1$ such that the first $2n$ moments (3.5) are matched. The eigenvalues and the squared first components of the corresponding normalized eigenvectors of T_n represent the nodes and weights of the related Gauss-Christoffel quadrature. Vice versa, T_n is determined by its eigenvalues and the first components of the corresponding normalized eigenvectors (for the rich history of the investigation of the last problem see [35, Section 3.1]), i.e. by the nodes and weights of the given Gauss-Christoffel quadrature. The Lanczos and CG methods can therefore be considered as *matrix formulations of the Gauss-Christoffel quadrature*. Moreover, considering the Gauss-Christoffel quadrature for $f(\lambda) = 1/\lambda$, we get

$$\int_0^\infty \lambda^{-1} \, d\omega(\lambda) \;=\; \frac{\|x - x_0\|_A^2}{\|r_0\|^2} \,,$$

and

$$\frac{\|x - x_0\|_A^2}{\|r_0\|^2} \;=\; \sum_{\ell=1}^{n} \omega_\ell^{(n)} \, \{\theta_\ell^{(n)}\}^{-1} \;+\; \frac{\|x - x_n\|_A^2}{\|r_0\|^2} \,. \tag{3.7}$$

That means that the error of the n-point Gauss-Christoffel quadrature is for $f(\lambda) = 1/\lambda$ given by the squared energy norm of the error at the nth step of the CG method scaled by $1/\|r_0\|^2$. For proof and review of some related results we refer to [24, 44, 34].

The moment matching model reduction described above can be easily extended to non-Hermitian Krylov subspace methods. For that purpose it will be useful to formulate the moment problem for A HPD in an operator (vector) form.

Consider the HPD matrix A as an operator on $\mathbb{C}^N$. Using the orthogonal projector $Q_n \equiv V_n V_n^*$ onto $\mathcal{K}_n(A, v_1)$, we can orthogonally restrict the operator A to the subspace $\mathcal{K}_n(A, v_1)$. Then the resulting orthogonally projected restriction A_n is given by

$$A_n = V_n V_n^* A V_n V_n^* = V_n T_n V_n^* \tag{3.8}$$

with its matrix in the orthonormal basis of $\mathcal{K}_n(A, v_1)$ represented by V_n

$$V_n^* A_n V_n = T_n \, . \tag{3.9}$$

The restricted operator A_n is determined by its action on the vectors generating $\mathcal{K}_n(A, v_1)$,

$$\begin{aligned} A_n v_1 &= A v_1, \\ A_n (A v_1) &= A^2 v_1, \\ &\vdots \\ A_n (A^{n-2} v_1) &= A^{n-1} v_1, \\ A_n (A^{n-1} v_1) &= Q_n (A^n v_1) \equiv V_n V_n^* A^n v_1 \, . \end{aligned} \tag{3.10}$$

Equivalently,

$$\begin{aligned} A_n v_1 &= A v_1, \\ A_n^2 v_1 &= A^2 v_1, \\ &\vdots \\ A_n^{n-1} v_1 &= A^{n-1} v_1, \\ A_n^n v_1 &= V_n V_n^* A^n v_1 \, . \end{aligned} \tag{3.11}$$

These relationships represent the operator (or vector) moment problem given by Vorobyev, see [47, Chapter III, Sections 2-4, in particular equation (11) on p. 53].

Using the Vorobyev formulation, proof of the matching moment property is easy and elegant. By construction, see (3.11),

$$v_1^* A^k v_1 = v_1^* A_n^k v_1 \, , \quad k = 1, \dots, n \, . \tag{3.12}$$

Since $\mathcal{K}_n(A, v_1) = \mathit{span}\ \{v_1, \dots, A^{n-1} v_1\}$ and $A_n^n v_1 \in \mathcal{K}_n(A, v_1)$, the orthogonal projection

$$0 = Q_n (A^n v_1) - A_n^n v_1 = Q_n (A^n v_1 - A_n^n v_1)$$

implies that the difference $A^n v_1 - A_n^n v_1$ must be orthogonal to all basis vectors v_1, $A v_1 = A_n v_1, \dots, A^{n-1} v_1 = A_n^{n-1} v_1$, which gives (for $j = 0$ trivially), using the properties $A^* = A$, $A_n^* = A_n$,

$$0 = (A^j v_1)^* (A^n v_1 - A_n^n v_1) = v_1^* A^{n+j} v_1 - v_1^* A_n^{n+j} v_1 \, , \quad j = 0, 1, \dots, n-1 \, .$$

Combining with (3.12) and (3.8) this gives

$$v_1^* A^k v_1 = v_1^* A_n^k v_1 = e_1^T T_n^k e_1 \, , \quad k = 0, 1, \dots, 2n-1 \, , \tag{3.13}$$

which proves (3.5).

Summarizing, the moment problems (2.3) and (3.5) can equivalently be represented by construction of the operator A_n on the n-dimensional subspace $\mathcal{K}_n(A, v_1)$ given by (3.10) or (3.11). Please note that apart from the relationship to the CG method, all statements of this subsection remain valid for A Hermitian.

4. Matching moments and non-Hermitian Krylov subspace methods. The nice paper [13] which focused on the HPD case motivated the Gauss-Christoffel quadrature interpretation related to the non-Hermitian Lanczos algorithm and to the Arnoldi algorithm presented in [15]. Further interpretations of moment matching as extensions of the Gauss-Christoffel quadrature to the complex plane were described in [38, 37], with references to computation of the scattering amplitude [48], see also [8, 26]. Developments of general polynomial-based extensions to the complex plane can lead to various interesting formulas. On the other hand, such general constructions typically include some nontrivial assumptions and formal polynomial relationships which do not have the quantitative impacts and interpretations of the same depth as the standard Gauss-Christoffel quadrature corresponding to the HPD case. In our view, the mathematical essence of the complex Gauss-Christoffel quadrature generalizations is given by the matching moment property, which can be formulated without further assumptions in a matrix form.

In this section we describe the extension of the matching moment model reduction to the non-Hermitian case without using the matrix of moments, cf. [2, Chapter 11], and without using polynomial-based generalizations of the Gauss-Christoffel quadrature formulas to the complex plane.

4.1. Non-Hermitian Lanczos process. Given a nonsingular N by N matrix A and two starting vectors $v \equiv v_1, w \equiv w_1$ of length N, $\|v_1\| = 1, w_1^* v_1 = 1$, the non-Hermitian Lanczos algorithm can be written in the form

$$\begin{aligned} AV_n &= V_n T_n + \delta_{n+1} v_{n+1} e_n^T \,, \\ A^* W_n &= W_n T_n^* + \beta_{n+1}^* w_{n+1} e_n^T \,, \end{aligned} \tag{4.1}$$

where $W_n^* V_n = I$, $T_n = W_n^* A V_n$, $\|v_{n+1}\| = 1, w_{n+1}^* v_{n+1} = 1$,

$$T_n = \begin{pmatrix} \gamma_1 & \beta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ & & \delta_n & \gamma_n \end{pmatrix} , \quad \delta_\ell > 0,\ \beta_\ell \neq 0,\ \ell = 2, \dots, n \,, \tag{4.2}$$

see, e.g. [36, Section 7.1], [2, Section 11.2.1], [6, Section 2.4]. Here *it is assumed* that the algorithm does not break down in steps 1 through n. The columns of V_n form a basis of $\mathcal{K}_n(A, v_1)$ while the columns of W_n a basis of $\mathcal{K}_n(A^*, w_1)$. Because of the biorthogonality $W_n^* V_n = I$, the oblique projector onto $\mathcal{K}_n(A, v_1)$ orthogonal to $\mathcal{K}_n(A^*, w_1)$ can be written as

$$Q_n = V_n W_n^* \,. \tag{4.3}$$

The discussion of breakdown in the non-Hermitian Lanczos algorithm is out of the scope of this paper. The assumption here means that the results on matching moments model reduction are valid for the step n providing that the algorithm does not breakdown before or on that step. It is not assumed, cf. [6, p. 16, relation (15)], that the Lanczos algorithm can be carried out to step N. Such an assumption means a loss of generality, since there are incurable breakdowns in the non-Hermitian Lanczos method, and mentioning a look-ahead scheme proposed in [14] as a possible general cure, see [6, p. 16] is not mathematically correct. Look-ahead techniques are important in practical computations, but the context of their use is different. They can not cure incurable breakdowns.

We will prove that under the given assumption on existence of the recurrence steps 1 through n, the non-Hermitian Lanczos algorithm represents the model reduction which matches the first $2n$ moments

$$w_1^* A^k v_1 = e_1^T T_n^k e_1 \,, \quad k = 0, 1, \ldots, 2n-1 \,, \tag{4.4}$$

with T_n given by (4.2). Our proof is based on the relationship with the corresponding Vorobyev moment problem, while the proof in [2] uses factorization of the matrix of moments, and the proof in [6] the additional assumption discussed above.

The restriction A_n of A to $\mathcal{K}_n(A, v_1)$ projected orthogonally to $\mathcal{K}_n(A^*, w_1)$ is given by

$$A_n = V_n W_n^* A V_n W_n^* = V_n T_n W_n^* \,. \tag{4.5}$$

It can be determined by its action on the generating vectors of $\mathcal{K}_n(A, v_1)$,

$$\begin{aligned} A_n v_1 &= A v_1, \\ A_n(A v_1) &= A^2 v_1, \\ &\vdots \\ A_n(A^{n-2} v_1) &= A^{n-1} v_1, \\ A_n(A^{n-1} v_1) &= V_n W_n^* A^n v_1 \,, \end{aligned} \tag{4.6}$$

or, equivalently,

$$\begin{aligned} A_n v_1 &= A v_1, \\ A_n^2 v_1 &= A^2 v_1, \\ &\vdots \\ A_n^{n-1} v_1 &= A^{n-1} v_1, \\ A_n^n v_1 &= V_n W_n^* A^n v_1 \,. \end{aligned} \tag{4.7}$$

Trivially, see (4.7),

$$w_1^* A^k v_1 = w_1^* A_n^k v_1 \,, \quad k = 0, 1, \ldots, n \,.$$

For the powers $n+1$ to $2n-1$ the situation is more subtle. Using

$$0 = V_n W_n^* (A^n v_1 - A_n^n v_1) \,,$$

$A^n v_1 - A_n^n v_1$ is orthogonal to the generating vectors $w_1, A^* w_1, \ldots, (A^*)^{n-1} w_1$ of the Krylov subspace $\mathcal{K}_n(A^*, w_1)$. We therefore get

$$\left((A^*)^\ell w_1\right)^* (A^n v_1 - A_n^n v_1) = 0 \,, \quad \ell = 0, \ldots, n-1 \,.$$

A simple rearrangement leads with (4.5) to

$$w_1^* A^{\ell+n} v_1 = w_1^* A^\ell A_n^n v_1 = w_1^* A^\ell V_n T_n^n e_1 \,, \quad \ell = 0, \ldots, n-1 \,. \tag{4.8}$$

Now consider the term $w_1^* A^\ell$ in (4.8), which we will view as $\big((A^*)^\ell w_1\big)^*$. Using the fact that the matrix T_n given by (4.2) is tridiagonal,

$$\begin{aligned}
(A^*)^{n-1} w_1 &= (A^*)^{n-2}(A^* w_1) = (A^*)^{n-2}(A^* W_n e_1)\\
&= (A^*)^{n-2}(W_n T_n^* + \beta_{n+1}^* w_{n+1} e_n^T)\, e_1 = (A^*)^{n-2} W_n T_n^* e_1\\
&= (A^*)^{n-3}(A W_n) T_n^* e_1 = (A^*)^{n-3}(W_n (T_n^*)^2 e_1 + \beta_{n+1}^* w_{n+1} e_n^T T_n^* e_1)\\
&= (A^*)^{n-3} W_n (T_n^*)^2 e_1\\
&= \quad \cdots\\
&= W_n (T_n^*)^{n-1} e_1 + \beta_{n+1}^* w_{n+1} e_n^T (T_n^*)^{n-2} e_1\,,
\end{aligned}$$

where we have repeatedly used

$$e_n^T (T_n^*)^\ell e_1 = 0, \quad \ell = 0, 1, \, \ldots \, , n-2.$$

Since $e_n^T (T_n^*)^{n-2} e_1 = 0$, we have

$$(A^*)^{n-1} w_1 = W_n (T_n^*)^{n-1} e_1\,,$$

and the analogous identity clearly holds for the powers less than $n-1$, i.e.

$$(A^*)^\ell w_1 = W_n (T_n^*)^\ell e_1\,, \quad \ell = 0, 1, \ldots, n-1\,.$$

Putting all pieces together and using the biorthogonality $W_n^* V_n = I$, (4.8) gives

$$w_1^* A^{\ell+n} v_1 = e_1^T T_n^{\ell+n} e_1\,, \quad \ell = 0, 1, \, \ldots \, , n-1\,,$$

which proves (4.4).

As suggested by Petr Tichý [46], the term $w_1^* A^\ell V_n$ in (4.8) can alternatively be expressed by considering the dual Vorobyev moment problem for $A_n^* = W_n T_n^* V_n^*$ which represent the restriction of A^* onto $\mathcal{K}_n(A^*, w_1)$ projected orthogonally to $\mathcal{K}_n(A, v_1)$,

$$\begin{aligned}
A_n^* w_1 &= A^* w_1,\\
(A_n^*)^2 w_1 &= (A^*)^2 w_1,\\
&\ \ \vdots\\
(A_n^*)^{n-1} w_1 &= (A^*)^{n-1} w_1,\\
(A_n^*)^n w_1 &= W_n V_n^* (A^*)^n w_1\,.
\end{aligned} \tag{4.9}$$

Clearly

$$w_1^* A^\ell V_n = w_1^* A^\ell V_n W_n^* V_n = w_1^* A_n^\ell V_n = e_1^* W_n^* A_n^\ell V_n = e_1^* T_n^\ell\,, \quad \ell = 0, 1, \ldots, n\,,$$

which again finishes the proof (the equality for the nth power given above is not needed).

Analogously to the Hermitian case, the non-Hermitian Lanczos algorithm and the non-Hermitian Lanczos method for approximation of the eigenvalues of A matches at the nth step the $2n$ moments (4.4). It is worth noticing that here we have *two* vectors v_1 and w_1, which represent the starting vectors for the two coupled recurrences with A and A^* respectively. In general v_1 is different from w_1.

4.2. Vorobyev moment problem and the Arnoldi method. Given a nonsingular N by N matrix A and an initial vector $v \equiv v_1$ of length N, $\|v_1\| = 1$, the Arnoldi algorithm can be seen as

$$AV_n = V_n H_n + h_{n+1,n} v_{n+1} e_n^T \,, \tag{4.10}$$

where

$$V_n^* V_n = I_n \,, \quad V_n^* v_{n+1} = 0 \,, \quad H_n = V_n^* A V_n \,,$$

and H_n is an upper Hessenberg matrix with positive entries on the first subdiagonal, see, e.g., [36, Section 6.3], [2, Chapter 10]. The n steps of the Arnoldi algorithm applied to the matrix A with the starting vector v_1 give the restriction A_n of A to $\mathcal{K}_n(A, v_1)$ projected orthogonally to $\mathcal{K}_n(A, v_1)$,

$$A_n = V_n V_n^* A V_n V_n^* = V_n H_n V_n^* \,. \tag{4.11}$$

Expressing this through the Vorobyev moment problem gives

$$\begin{aligned} A_n v_1 &= A v_1 \,, \\ A_n (A v_1) &= A^2 v_1 \,, \\ &\vdots \\ A_n (A^{n-2} v_1) &= A^{n-1} v_1 \,, \\ A_n (A^{n-1} v_1) &= V_n V_n^* (A^n v_1) \,, \end{aligned} \tag{4.12}$$

or, equivalently,

$$\begin{aligned} A_n v_1 &= A v_1 \,, \\ A_n^2 v_1 &= A^2 v_1 \,, \\ &\vdots \\ A_n^{n-1} v_1 &= A^{n-1} v_1 \,, \\ A_n^n v_1 &= V_n V_n^* A^n v_1 \,. \end{aligned} \tag{4.13}$$

Given *an additional* vector $u \equiv u_1$ of length N, multiplication of the first $n-1$ rows in (4.13) from the left by u_1^* results in

$$u_1^* A^k v_1 = u_1^* A_n^k v_1 \,, \quad k = 0, \dots, n-1 \,. \tag{4.14}$$

Unlike the non-Hermitian Lanczos case, the additional vector u_1 in the Arnoldi algorithm is not a part of the recurrence. Since u_1 is generally unrelated to v_1, and since A is non-Hermitian, the last row in (4.13) is, in general, of no help in extending the matching moments property beyond the first n moments. With (4.11)

$$A_n^k = V_n H_n^k V_n^* \,, \quad k = 1, \, \dots \, , n-1 \,.$$

Finally, using (4.14),

$$u_1^* A^k v_1 = u_1^* V_n H_n^k e_1 = t_n^* H_n^k e_1 \,, \quad k = 0, \dots, n-1 \,, \tag{4.15}$$

where t_n is defined by the orthogonal decomposition of u_1,

$$u_1 \equiv V_n t_n + u_1^{\perp} = V_n (V_n^* u_1) + u_1^{\perp} \,, \tag{4.16}$$

here $u_1^\perp$ is the component of u_1 orthogonal to $\mathcal{K}_n(A, v_1)$. Multiplication of the last row in (4.13) from the left by u_1^* gives

$$u_1^* V_n V_n^* A^n v_1 = (V_n^* u_1)(V_n^* A^n V_n) e_1 = u_1^* A_n^n v_1 \,. \tag{4.17}$$

With $u_1 = v_1$ we can therefore add one more moment and conclude,

$$v_1^* A^n v_1 = e_1 H_n^n e_1 \,. \tag{4.18}$$

Summarizing, the n steps of the Arnoldi algorithm and of the Arnoldi method for approximation of eigenvalues of A starting with v_1 can be viewed with a given additional vector u_1 as the reduction of the original model represented by A, v_1 and u_1 to the reduced model represented by H_n, e_1 and t_n which matches the n moments

$$u_1^* A^k v_1 = t_n^* H_n^k e_1 \,, \quad k = 0, 1, \ldots, n-1 \,. \tag{4.19}$$

The vector t_n is given by the orthogonal decomposition (4.16). If $u_1 = v_1$, then this model reduction matches $n+1$ moments (4.19) and (4.18).

5. Concluding remarks. Moment matching model reduction represents a fundamental mathematical concept used in many applications. The Vorobyev formulation of the moment problem makes in an elegant way a link between the matching moments model reduction and the approximations by projections onto Krylov subspaces. In our further work we will investigate how this approach leads to estimates for the form $w^* A^{-1} v$ which can be efficiently computed using various Krylov subspace methods for solving non-Hermitian linear algebraic systems. Results will be presented in the forthcoming paper [45].

Acknowledgement. I wish to express my deepest gratitude to Gene Golub, who introduced me into the beautiful world of moments, and who supported my interest in related topics by his knowledge, advice and friendly guidance for many years. My thanks belong also to Gerard Meurant for his help and lasting friendship. I am also greatly indebted to Claude Brezinski; without reading his book I would not realize the beauty of the Vorobyev moment problem and the work of Vorobyev would for me probably remain unknown. I thank Petr Tichý for careful reading the manuscript and for his valuable comments which have improved the presentation, Chris Paige for suggesting several changes of the text, and Jörg Liesen for useful discussions.

Appendix: Historical remarks. Different meanings of the term *moments* recalled in our short contribution reflect only a small part of the large variety of contexts in which moments are used throughout mathematics and in many areas of applications. As mentioned before, the motivation of Sieltjes [40, 42, 41, 43] was taken from mechanics; the Riemann-Stieltjes integral

$$\int_a^b d\omega(\lambda) \tag{5.1}$$

represented the total mass distributed over the interval $[a, b] \in [0, +\infty)$. As pointed out by Shohat and Tamarkin [39], this gave the name to $\omega(\lambda)$ as the *distribution function*, used nowadays. The first and the second moments also have the interpretation of the mechanical moments with respect to 0 of the total mass $\int_0^\infty d\omega(\lambda)$ distributed over the semi-axis $[0, +\infty)$. Stieltjes solved his formulation of the problem of moments by developing the theory of continued fractions. Prior to Stieltjes, analogous ideas can be found in the work of Chebyshev and Markov, see the Introduction in [39] and the Foreword in [1]. Though their motivation was not solving the problem of moments, their work led to the development of the closely related general theory of orthogonal polynomials.

As pointed out in our text, the idea of finding a distribution function with n points of increase which matches the first $2n$ moments of the *given* distribution function (in the sense of the Riemann-Stieltjes integral) can be viewed as a special form of the Stieltjes moment problem. In this sense this goes back to the idea of Gauss quadrature [17], and in particular to its formulation given by Jacobi [29], see the insightful commentary given by Goldstine in [21, Sections 4.11 and 5.2]. Gauss quadrature was further extended by Christoffel [9]. Therefore we refer to it as Gauss-Christoffel quadrature, while using the shorter term 'Gauss quadrature' in some cases where we refer to the work of other authors who use that term exclusively. A thorough historical description of related work can be found in the remarkable review paper by Gautschi [18, Introduction and Section 1].

In their seminal paper on the conjugate gradient method, Hestenes and Stiefel described the fundamental relationship of the algebraic formulation of the method, intended for solving systems of linear algebraic equations, to the Riemann-Stieltjes integral, orthogonal polynomials, Gauss(-Christoffel) quadrature and continued fractions [28, Sections 14–16]. As we see above, Vorobyev presented in his book [47] the algebraic formulation of the problem of moments which had allowed him to develop the so-called unified moment method for solving linear systems of equations and computing eigenvalues of linear operators. He used the equivalence of the scalar moment problem formulation recalling the work of Stieltjes, Chebyshev and Markov with his operator moment problem formulation for the Hermitian positive definite case. He was well aware of the relationship with the work of Lanczos, Hestenes and Stiefel (in addition to that he refers to the work of Ljusternik from 1956 which is not available to the author of this paper). Vorobyev did not formulate his moment problem as a model reduction in the sense presented in this paper. Our contribution can be considered an extension of his ideas in that direction, as well as an extension to the non-Hermitian case. It is worthwhile to point out that, apart from the book [47], published originally in Russian in 1958, we have found only five papers published by Vorobyev (all prior to 1965). We were unable to locate any bibliographical data about the author.

In [15] the moment matching in the non-Hermitian Lanczos process, described in Section 4.1, is interpreted as a Gauss quadrature in the complex plane. On the other hand, the Gauss quadrature associated with the Arnoldi algorithm given in [15] is not related in an analogous way to the moment matching described here. The Gauss quadrature interpretation of the non-Hermitian Lanczos process in [15] gives the essence of the Gauss quadrature extension, and it leaves out possible formal polynomial expression of the Gauss quadrature in the complex plane, which was given later under various assumptions (including existence of the spectral decomposition i.e. diagonalizability of the corresponding matrices) by several authors, see, e.g., [38, 37].

REFERENCES

[1] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, Translated by N. Kemmer, Hafner Publishing Co., New York, 1965.

[2] A. C. Antoulas, *Approximation of large-scale dynamical systems*, vol. 6 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA,

2005. With a foreword by Jan C. Willems.

[3] M. Arioli, *A stopping criterion for the conjugate gradient algorithms in a finite element method framework*, Numer. Math., 97 (2004), pp. 1–24.

[4] M. Arioli, D. Loghin, and A. J. Wathen, *Stopping criteria for iterations in finite element methods*, Numer. Math., 99 (2005), pp. 381–410.

[5] M. Arioli, E. Noulard, and A. Russo, *Stopping criteria for iterative methods: applications to PDE's*, Calcolo, 38 (2001), pp. 97–112.

[6] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*, Appl. Numer. Math., 43 (2002), pp. 9–44. 19th Dundee Biennial Conference on Numerical Analysis (2001).

[7] C. Brezinski, *Projection Methods for Systems of Equations*, vol. 7 of Studies in Computational Mathematics, North-Holland Publishing Co., Amsterdam, 1997.

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

[9] E. B. Christoffel, *Ueber die gaussische quadratur und eine verallgemeinerung derselben*, J. Reine Angew. Math., 55 (1858), pp. 61–82, Ges. Math. Abhandlungen I, 65–87.

[10] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Computer Science and Applied Mathematics, Academic Press Inc., Orlando, FL, second ed., 1984.

[11] M. Eiermann and O. G. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.

[12] B. Fischer, *Polynomial based iteration methods for symmetric linear systems*, Wiley-Teubner Series Advances in Numerical Mathematics, John Wiley & Sons Ltd., Chichester, 1996.

[13] B. Fischer and R. W. Freund, *On adaptive weighted polynomial preconditioning for Hermitian positive definite matrices*, SIAM J. Sci. Comput., 15 (1994), pp. 408–426. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).

[14] R. W. Freund, M. H. Gutknecht, and N. M. Nachtigal, *An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices*, SIAM J. Sci. Comput., 14 (1993), pp. 137–158.

[15] R. W. Freund and M. Hochbruck, *Gauss quadratures associated with the Arnoldi process and the Lanczos algorithm*, in Linear algebra for large scale and real-time applications, M. S. Moonen, G. H. Golub, and B. L. R. De Moor, eds., vol. 232 of NATO Advanced Science Institutes Series E: Applied Sciences, Dordrecht, 1993, Kluwer Academic Publishers Group, pp. 377–380.

[16] F. R. Gantmacher and M. G. Krein, *Ostsilljatsionnye matritsy i jadra i malye kolebania mekhanicheskikh sistem*, Gosudarstvjennoe Izdatjelstvo Techniko – Teoretitcheskoj Literatury, Moscow, 1950. (English translation based on the 1941 Russian original "Oscillation matrices and kernels and small vibrations of mechanical systems" edited and with a preface by Alex Eremenko, published by AMS Chelsea Publishing, Providence, RI, 2002).

[17] C. F. Gauss, *Methodus nova integralium valores per approximationem inveniendi*, Commentationes Societatis Regiae Scientarium Gottingensis Recentiores, 3 (1814), pp. Werke III, 163–196.

[18] W. Gautschi, *A survey of Gauss-Christoffel quadrature formulae*, in E. B. Christoffel (Aachen/Monschau, 1979), Birkhäuser, Basel, 1981, pp. 72–147.

[19] ———, *The interplay between classical analysis and (numerical) linear algebra—a tribute to Gene H. Golub*, Electron. Trans. Numer. Anal., 13 (2002), pp. 119–147.

[20] ———, *Orthogonal Polynomials, Computation and Approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2004.

[21] H. H. Goldstine, *A History of Numerical Analysis from the 16th through the 19th Century*, Springer-Verlag, New York, 1977. Studies in the History of Mathematics and Physical Sciences, Vol. 2.

[22] G. H. Golub, *Matrix computation and the theory of moments*, in Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 1440–1448.

[23] ———, *Milestones in Matrix Computation: Selected Works of Gene H. Golub, with Commentaries*, Oxford Science Publications, Oxford University Press, Oxford, 2007. Edited by Raymond H. Chan, Chen Greif and Dianne P. O'Leary.

[24] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical analysis 1993 (Dundee, 1993), vol. 303 of Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow, 1994, pp. 105–156.

[25] G. H. Golub and G. Meurant, *Matrices, moments and quadrature. II. How to compute the norm of the error in iterative methods*, BIT, 37 (1997), pp. 687–705. Direct methods, linear algebra in optimization, iterative methods (Toulouse, 1995/1996).

[26] G. H. Golub, M. Stoll, and A. Wathen, *Approximation of the scattering amplitude and linear systems*, ETNA, (2008).
[27] G. H. Golub and Z. Strakoš, *Estimates in quadratic formulas*, Numer. Algorithms, 8 (1994), pp. 241–268.
[28] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).
[29] C. G. J. Jacobi, *Ueber gauss neue methode, die werthe der integrale näherungsweise zu finden*, J. Reine Angew. Math., 1 (1826), pp. 301–308, Math. Werke III, 97–113.
[30] P. Jiránek, Z. Strakoš, and M. Vohralík, *A posteriori error estimates including algebraic error: Computable upper bounds and stopping criteria for iterative solvers*, (in preparation).
[31] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.
[32] ———, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.
[33] G. Meurant, *The Lanczos and Conjugate Gradient Algorithms - From Theory to Finite Precision Computations*, SIAM, Philadelphia, 2006.
[34] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numer., 15 (2006), pp. 471–542.
[35] D. P. O'Leary, Z. Strakoš, and P. Tichý, *On sensitivity of Gauss-Christoffel quadrature*, Numer. Math., 107 (2007), pp. 147–174.
[36] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, second ed., 2003.
[37] P. E. Saylor and D. C. Smolarski, *Addendum to: "Why Gaussian quadrature in the complex plane?" [Numer. Algorithms 26 (2001), no. 3, 251–280]*, Numer. Algorithms, 27 (2001), pp. 215–217.
[38] ———, *Why Gaussian quadrature in the complex plane?*, Numer. Algorithms, 26 (2001), pp. 251–280.
[39] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society Mathematical surveys, vol. II, American Mathematical Society, New York, 1943.
[40] T. J. Stieltjes, *Sur l'évaluation approchée des intégrales*, C. R. Acad. Sci Paris, 97 (1883), pp. 740–742, 798–799 Oeuvres I, 314–316, 317–318.
[41] ———, *Note sur quelques formules pour l'évaluation de certaines intégrales*, Bul. Astr. Paris, 1 (1884), pp. 568, Oeuvres I, 426–427.
[42] ———, *Quelques recherches sur la théorie des quadratures dites mécaniques*, Ann. Sci. École Norm. Sup. (3), 1 (1884), pp. 409–426, Oeuvres I, 377–396.
[43] ———, *Sur une généralisation de la théorie des quadratures mécaniques*, C. R. Acad. Sci Paris, 99 (1884), pp. 850–851, Oeuvres I, 428–429.
[44] Z. Strakoš and P. Tichý, *On error estimation in the conjugate gradient method and why it works in finite precision computations*, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80 (electronic).
[45] ———, *Estimation of $c^*A^{-1}b$ via matching moments*, (in preparation).
[46] P. Tichý, *Personal communication*, (2008).
[47] Y. V. Vorobyev, *Methods of moments in applied mathematics*, Translated from the Russian by Bernard Seckler, Gordon and Breach Science Publishers, New York, 1965.
[48] K. F. Warnick and W. C. Chew, *Numerical simulation methods for rough surface scattering*, Waves Random Media, 11 (2001), pp. R1–R30.