Open access • Journal Article • DOI:10.1137/08073144X

The Lanczos Method for Parameterized Symmetric Linear Systems with Multiple Right-Hand Sides — Source link

Karl Meerbergen, Zhaojun Bai

Institutions: Katholieke Universiteit Leuven, University of California, Davis

Published on: 01 Jan 2010 - SIAM Journal on Matrix Analysis and Applications (Society for Industrial and Applied Mathematics)

Topics: Lanczos approximation, Lanczos resampling, Lanczos algorithm, Linear system and Linear algebra

The Lanczos method for parameterized symmetric linear systems with multiple right-hand sides

Zhaojun Bai *Karl Meerbergen*

Report TW 527, July 2008

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

The Lanczos method for parameterized symmetric linear systems with multiple right-hand sides

Zhaojun Bai *Karl Meerbergen*

Report TW 527, July 2008

Department of Computer Science, K.U.Leuven

Abstract

The solution of linear systems with a parameter is an important problem in engineering applications, especially in structural dynamics, acoustics, and electronic circuit simulations and related modelreduction methods such as Padé via Lanczos. In this paper, we present a method for solving parameterized symmetric linear systems with multiple right-hand sides, based on the Lanczos method. We show that for this class of applications, a simple deflation method can be used.

Keywords : Parameterized linear system, Lanczos method, recycling Ritz vectors.
MSC : Primary : 65F15,65F50

THE LANCZOS METHOD FOR PARAMETERIZED SYMMETRIC LINEAR SYSTEMS WITH MULTIPLE RIGHT-HAND SIDES

ZHAOJUN BAI* AND KARL MEERBERGEN†

Abstract. The solution of linear systems with a parameter is an important problem in engineering applications, especially in structural dynamics, acoustics, and electronic circuit simulations and related modelreduction methods such as Padé via Lanczos. In this paper, we present a method for solving parameterized symmetric linear systems with multiple right-hand sides, based on the Lanczos method. We show that for this class of applications, a simple deflation method can be used.

Key words. Parameterized linear system, Lanczos method, recycling Ritz vectors.

AMS subject classification. 65F15,65F50

1. Introduction. In engineering applications including structural dynamics and acoustics, the computation of the frequency response function of a vibrating system over a given frequency range $\Omega = [\omega_{\min}, \omega_{\max}]$ can be a time consuming operation. In applications on closed domains without damping, the frequency response function often is the solution of the parameterized linear system

$$Ax = f \quad \text{with} \quad A = K - \omega^2 M \tag{1.1}$$

with K real symmetric, and M symmetric positive definite. The matrices K and M are large and sparse. We have chosen ω to be the frequency, but it could also be the angular frequency, the wave number or the characteristic (dimensionless) wave number. The frequency range Ω is discretized into the set $\{\omega_1, \omega_2, \ldots, \omega_m\}$ where m can be of the order 100 or 1000. This solution process is called frequency sweeping. Since A is large, the solution of (1.1) is expensive when (1.1) is solved for the frequency points $\omega = \omega_1, \omega_2, \ldots, \omega_m$ independently. In this paper, we study frequency sweeping with multiple right-hand sides, i.e. f can take different values $f_1, f_2, \ldots, f_s$ in (1.1).

Efficient solution methods to this parameterized linear system have been proposed in the literature. The most famous method in engineering is undoubtfully modal superposition. References can be found in text books on engineering, e.g. [10]. The method projects the right-hand side and solution vectors on a basis of eigenvectors of the underlying eigenvalue problem

$$Ku = \lambda M u \tag{1.2}$$

in $\Lambda = [\lambda_{\min}, \lambda_{\max}]$, where $\lambda_{\min} \ll \min(\Omega^2)$ and $\lambda_{\max} \gg \max(\Omega^2)$. This method is usually experienced as efficient when the eigenvectors and eigenvalues are available, since (1.1) is transformed to a diagonal linear system. The practical problem is that it is not always clear how $\lambda_{\min}$ and $\lambda_{\max}$ need to be chosen. For example, when $\min(\Omega) = 0$, we could use $\lambda_{\min} = 0$ and $\lambda_{\max} = \eta \max(\Omega^2)$ with $\eta \in [2, 10]$. The eigenvalues and eigenvectors of (1.2) are in practice computed by the (block) Lanczos method, see e.g. [17] or the automated multi-level substructuring (AMLS) method [6, 5, 7] which is advocated for very large scale problems.

*Department of Computer Science, University of California, Davis, CA 95616, USA. bai@cs.ucdavis.edu

†Department of Computer Science, K.U.Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium. Karl.Meerbergen@cs.kuleuven.be

Another technique that received quite some attention is based on Krylov iterative methods on the preconditioned system

$$(K - \sigma M)^{-1} A x = (K - \sigma M)^{-1} f, \tag{1.3}$$

where σ is a properly chosen point within the interval Ω^2. We do not discuss how to choose σ It should be chosen somewhere inside or close to Ω^2. See the discussion for the associated eigenvalue problem [17, 21] and the rational Krylov method for modelreduction [15]. The solution vector x is approximated by a truncated vector Padé series, which is computed by the Lanczos method [13, 2, 3]. See also [31, 22] in the context of parameterized iterative linear system solvers and [23] for Rayleigh damping. This method is related to the Lanczos eigenvalue solver and the conjugate gradient method, which we will discuss further in the paper. The limitation of the method is that the right-hand side vector f should not have a spatial dependency on ω, i.e. $f = \widetilde{f}\phi(\omega)$ where $\widetilde{f}$ does not depend on ω and ϕ is a scalar function. In the engineering literature, this is called the Ritz vector technique [33]. The connection between the Lanczos method and a vector Padé series around σ is important in applications, since the frequency response function is a rational function with the eigenvalues as poles. The approximation preferably respects this rational nature. This is one reason why the Lanczos method is preferred to the MINRES method [22].

In contrast to the Lanczos method, modal superposition does not pose a condition on the right-hand side. The difficulty with the modal superposition method is that a relatively large number of eigenvectors may be required with a relatively high precision. In [5, 18], the AMLS frequency sweeping method is studied which makes a combination of modal superposition and a stationary iterative method. The method only uses the modes corresponding to the frequency range of interest, i.e. eigenvalues outside this interval are not used in modal superposition. Superposition on this reduced set of eigenvectors does produce the peaks in the frequency response function, but the zeroes are wrong. Therefore, the AMLS frequency sweeping method uses an iterative method to improve the solution obtained by superposition on a reduced modal basis. The iterative method uses the eigenbasis as a preconditioner. This is also known as deflation preconditioning or augmented Krylov subspace methods [29]. It can be viewed as an iterative method on the subspace orthogonal to the given eigenvectors. Since the eigenvalues play an important role in the convergence of iterative methods, eliminating the eigenvalues that hinder convergence can be very effective indeed.

The use of deflation preconditioning and recycling Ritz vectors is not new in the context of iterative linear system solvers [8, 29, 26, 32]. In [29], a sequence of linear systems $\{Ax_k = f_k \text{ for } k = 1, 2, \ldots\}$ is solved. The Ritz values of A, computed by the conjugate gradient method for solving the first system, are used in a deflation preconditioner for the second. Recycling subspaces for linear systems with a parameter was recently introduced in [12, 11]. In both papers, the Ritz vectors for the solution of the first right-hand side are recycled for preconditioning the systems with the remaining right-hand sides. In [12], the GCRO method was extended to linear systems with a parameter. The authors prefer the situation where eigenvectors are recycled (or Ritz vectors with small residual norm), since this simplifies the method. In [11], recycling was proposed for the GMRES method. It was argued that the fact the residual vectors for the additional right-hand sides are not parallel to the GMRES residual vector, creates difficulties in recycling. A remedy was proposed to this problem. The key point in the current paper is that we can show that the

recycled Ritz vectors have a small residual norm, which allows for efficient deflation and recycling. The difference with the cited work and frequency sweeping is that the number of iterations is typically low for frequency sweeping, i.e. typically less than a hundred.

Our contributions can be summarized as follows. First, we show that the Lanczos method converges very quickly when Ritz vectors are recycled. This allows us to propose a simple algorithm. Second, we show that the residual norms of the Ritz pairs associated with the interval $\Omega^2 = [\min(\Omega)^2, \max(\Omega)^2]$ computed by the Lanczos method are small. Third, we show a connection with the Padé via Lanczos method (PVL) [13, 2, 3] for the model order reduction problem. If the deflated Ritz values have zero residual norms, we have an exact vector Padé approximation.

The plan of the paper is as follows. In §2, we introduce the numerical method based on deflation and the Lanczos method. In §3, we perform a spectral analysis and we show spectral properties of the deflated linear system. In §4, we discuss how Ritz values can be computed using the Lanczos method and how accurate they are: we show that the residual norm is small for the eigenvalues corresponding to the interval Ω. Section 5 presents a practical procedure for solving (1.1) with multiple right-hand sides recycling Ritz vectors from the first right-hand side. Section 6 shows a numerical example for an application from structural engineering. We close the paper with concluding remarks.

We introduce the following notation. The interval of ω's for which x needs to be computed is denoted by $\Omega = [\omega_{\min}, \omega_{\max}]$. In our applications, $\omega_{\min} \geq 0$. We also define $\Omega^2 = [\omega_{\min}^2, \omega_{\max}^2]$. The transpose is denoted by x^T. The M norm $\|x\|_M$ is defined as the induced norm from the M inner product: $\sqrt{x^T M x}$.

2. Deflation in parameterized linear systems. In this section, we explain the ideas of deflation for solving (1.1). We start from the viewpoint of rational approximation, since x is a rational function in ω^2. Next, we deflate a part of the spectrum and obtain a linear system for the remainder which will be solved iteratively.

2.1. Rational function splitting. Let

$$KU = MU\Lambda \tag{2.1}$$

be an eigendecomposition of (1.2), where Λ is a diagonal matrix and $U^T M U = I$. By (2.1), we have that $M = U^{-T}U^{-1}$ and $K = U^{-T}\Lambda U^{-1}$. Therefore, the solution x of (1.1) can be written as

$$x = U(\Lambda - \omega^2 I)^{-1} U^T f = \sum_{j=1}^{n} u_j \frac{u_j^T f}{\lambda_j - \omega^2} \ . \tag{2.2}$$

This is a rational function with the eigenvalues of (1.2) as poles. The vector x, as a function of ω^2, has a vertical asymptote in $\lambda_j \in \Omega^2$.

The idea in [5, 18] is to first compute the eigenvalues in Ω^2 and then compute the solution vector x as the sum of

$$x = x^{(1)} + x^{(2)}$$

with

$$x^{(1)} = \sum_{j=1}^{p} u_j \frac{u_j^T f}{\lambda_j - \omega^2} \quad \text{and} \quad x^{(2)} = \sum_{j=p+1}^{n} u_j \frac{u_j^T f}{\lambda_j - \omega^2}$$

where the $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of (1.2) in Ω^2. The first term $x^{(1)}$ is then computed straightforwardly as a sum whereas the second term $x^{(2)}$ is computed by an iterative process.

2.2. Deflated linear system. Denote by $\mathbb{L}$ the set of eigenvalues of (1.2). Let the columns of $U_p \in \mathbf{R}^{n\times p}$ be a selection of eigenvectors of (1.2). The associated eigenvalues are the set $\mathbb{L}_p$. We define the projector $P_p = U_p U_p^T M$ that maps $\mathbf{R}^n$ onto $\mathcal{U}_p = \text{range}(U_p)$. Similarly, the projector $P_\perp = I - U_p U_p^T M$ maps $\mathbf{R}^n$ onto the M-orthogonal complement of $\mathcal{U}_p$.

LEMMA 2.1. *Consider the linear systems*

$$P_p^T A P_p x^{(1)} = P_p^T f \tag{2.3}$$

$$P_\perp^T A P_\perp x^{(2)} = P_\perp^T f \tag{2.4}$$

then $x = x^{(1)} + x^{(2)}$.

Proof. The proof follows from the spectral decomposition of A. □

By Lemma 2.1, we can solve the problem in two steps:

1. Compute $x^{(1)} = U_p z$ with $(\Lambda_p^2 - \omega^2 I)z = U_p^T f$; this is a $p \times p$ diagonal system.
2. Solve $x^{(2)}$ from (2.4) using an iterative method.

For the iterative solution of (2.4), we use the following preconditioner:

$$P_\perp K_\sigma^{-1} P_\perp^T \ , \tag{2.5}$$

with $K_\sigma = K - \sigma M$, where σ is a shift chosen in Ω^2.

Preconditioning (2.4) by (2.5) leads to

$$(P_\perp K_\sigma^{-1} P_\perp^T)(P_\perp^T A P_\perp) x^{(2)} = (P_\perp K_\sigma^{-1} P_\perp^T) P_\perp^T f \ . \tag{2.6}$$

LEMMA 2.2. *The solution of (2.6) corresponds to solving*

$$P_\perp K_\sigma^{-1} A P_\perp x^{(2)} = P_\perp K_\sigma^{-1} f \tag{2.7}$$

or

$$K_\sigma^{-1} A x^{(2)} = P_\perp K_\sigma^{-1} f \ . \tag{2.8}$$

Proof. The projectors $P_\perp^T$ can be dropped from (2.6) since

$$P_\perp K_\sigma^{-1} P_\perp^T = P_\perp K_\sigma^{-1} \ ,$$

This leads to (2.7). Since $P_\perp x^{(2)} = x^{(2)}$ and $P_\perp K_\sigma A x^{(2)} = K_\sigma A x^{(2)}$, (2.8) follows. □

We define

$$B = P_\perp K_\sigma^{-1} A P_\perp \ , \tag{2.9}$$

so that (2.7) becomes

$$B x^{(2)} = P_\perp K_\sigma^{-1} f \ .$$

In [18], a stationary iterative solver was used for each value of ω for which x is computed. Experiments showed that only a few iterations (less than ten) for each

ω_j are usually sufficient for convergence, where $x(\omega_{j-1})$ is used as starting vector for computing $x(\omega_j)$. The explanation relies on the fact that $x^{(2)}$ does not have vertical asymptotes in Ω^2 and that $x^{(2)}$ is a much smoother function than $x^{(1)}$. When the number of ω's is a few hundred, the total cost is still significant. Instead of a stationary solver, we can use the parameterized Lanczos method [22] when f is independent of ω. This reduces the cost even more since x is computed for all ω's at once with a marginal additional cost per ω. We will discuss this in §2.3 and §2.4.

2.3. Lanczos method. In this section, we present the parameterized Lanczos method for the solution of (1.1), first without deflation and then with deflation. The method starts with the following so-called Lanczos procedure [19, 20]:

ALGORITHM 2.1 (Lanczos procedure).
1. Let $v_0 \neq 0$ be an initial vector, and set $\beta_0 = 0$
2. Solve $K_\sigma b = f$ for b.
3. Normalize $v_1 = b/\|b\|_M$.
4. For $j = 1, 2, \ldots, k$ do:
   4.1. Solve $K_\sigma w_j = M v_j$ for w_j.
   4.2. Compute $\widehat{w}_j = w_j - v_{j-1}\beta_{j-1}$.
   4.3. Compute $\alpha_j = v_j^T M \widehat{w}_j$.
   4.4. Compute $\widetilde{w}_j = \widehat{w}_j - v_j \alpha_j$.
   4.5. Normalize $\beta_j = \|\widetilde{w}_j\|_M$ and $v_{j+1} = \widetilde{w}_j/\beta_j$.

This is the spectral transformation Lanczos procedure using M orthogonalization. The computation of w_j in Step 4.1 requires a linear system solve with K_σ. In frequency response function computations in structural dynamics usually a direct solver is used, since linear systems are often strongly ill-conditioned. Alternatively, the AMLS method leads to a diagonal K_σ.

Define the tridiagonal matrix

$$T_k = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{bmatrix}$$

and $V_k = [v_1, \ldots, v_k]$. The following equations readily follow from Algorithm 2.1:

$$\begin{aligned} K_\sigma^{-1} M V_k &= V_k T_k + v_{k+1}\beta_k e_k^T \\ V_{k+1}^T M V_{k+1} &= I \ . \end{aligned} \tag{2.10}$$

The parameterized Lanczos method for the solution of (1.1) is first proposed in papers on model order reduction by the Padé via Lanczos method [13, 2, 3] and later studied in the context of frequency response computation by the shifted Lanczos method [14, 30, 31, 22]. The solution of (1.1) is computed by first preconditioning into (1.3). Applying the Lanczos procedure with M–inner product to solving (1.3) produces

$$K_\sigma^{-1}(K - \omega^2 M)V_k = V_k(I - (\omega^2 - \sigma)T_k) - (\omega^2 - \sigma)v_{k+1}\beta_k e_k^T$$

with V_k and T_k satisfying (2.10). An approximate solution $\widetilde{x}$ of (1.1) is given by

$$\widetilde{x} = V_k z, \tag{2.11}$$

where z is the solution of the linear system

$$(I - (\omega^2 - \sigma)T_k)z = e_1\|b\|_M \ .$$

This requires the solution of a $k \times k$ tridiagonal linear system. If k is small, its cost is low. The residual is

$$r = K_\sigma^{-1}(f - (K - \omega^2 M)\widetilde{x}) = (\omega^2 - \sigma)v_{k+1}\beta_k e_k^T z \ .$$

We assume that the Lanczos method is able to compute $\widetilde{x}$ with a small enough residual norm for $\omega \in \Omega$. If not, a larger value of k could be used or a restarting strategy with a new σ as for the eigenvalue problem [17]. This is out of the scope of this paper. Therefore, we assume that $\|r(\omega)\|_M$ is 'small' for all $\omega \in \Omega$.

Note that the vectors V_k need not be stored when $\widetilde{x}$ is updated on each Lanczos iteration. Often, only a few components of x are wanted and so we only need to store the desired components of $\widetilde{x}$ for all wanted ω's. In finite precision arithmetic, the columns of V_k lose orthogonality. Reorthogonalization can be used to restore the orthogonality although this is strictly speaking not required for convergence. Reorthogonalization, however, can produce a smaller residual norm: see e.g. the numerical examples in [22].

2.4. Deflation. In exact arithmetic, the Lanczos method can be used for solving the deflated system (2.4) when applied to $K_\sigma^{-1}M$ with starting vector $P_\perp b$, since $\mathcal{U}_p^\perp$ is an invariant subspace of $K_\sigma^{-1}M$ and $P_\perp b \in \mathcal{U}_p^\perp$ all columns of V_k lie in $\mathcal{U}_p^\perp$. See also (2.8). The iteration vectors stay in the invariant subspace M-orthogonal to the columns of U_p. In practical computations, the eigenpairs are not available to full accuracy. Alternatively, if we do not need accurate eigenvalue estimates, we may save computation time in the eigenvalue computation.

Let $\widehat{U}_p$ and $\widehat{\Theta}_p$ be approximate eigenpairs of $K_\sigma^{-1}M$, e.g. computed with the Lanczos procedure. We introduce the projectors $\widehat{P}_p = \widehat{U}_p\widehat{U}_p^T M$ and $\widehat{P}_\perp = I - \widehat{U}_p\widehat{U}_p^T M$. The Lanczos method can be used to solve the deflation system (2.8) using the projectors $\widehat{P}_p$ and $\widehat{P}_\perp$. However, if the basis vectors $\widehat{U}_p$ are no eigenvectors, the spectral splitting from Lemma 2.1 is no longer valid.

Furthermore, the deflated matrix B from the deflated equation from (2.8) becomes

$$\widehat{B} = \widehat{P}_\perp K_\sigma^{-1} A \widehat{P}_\perp \ .$$

Note that the residual of the approximate eigenpairs $\widehat{U}_p$ and $\widehat{\Theta}_p$ is

$$R_p = K_\sigma^{-1}M\widehat{U}_p - \widehat{U}_p\widehat{\Theta}_p \tag{2.12}$$

where $\widehat{U}_p^T M R_p = 0$. Therefore, $\widehat{\Theta}_p = \widehat{U}_p^T M K_\sigma^{-1} M \widehat{U}_p$.

2.4.1. Deflated right-hand side Lanczos (DRHSL). The simplest method is to apply the parameterized Lanczos method to the preconditioned equation with deflated right-hand side:

$$K_\sigma^{-1}Ax^{(2)} = \widehat{P}_\perp b.$$

By the Lanczos procedure, we have the recurrence relation

$$K_\sigma^{-1}MV_k = V_k T_k + \beta_k v_{k+1} e_k^T$$

where V_k is not necessarily orthogonal to $\widehat{U}_p$. An approximate solution $\widetilde{x}$ to (1.1) takes the form

$$\widetilde{x} = \widehat{U}_p z^{(1)} + V_k z^{(2)}$$

with

$$z^{(1)} = (I + (\sigma - \omega^2)\widehat{\Theta}_p)^{-1}\widehat{U}_p^T M b$$
$$z^{(2)} = (I + (\sigma - \omega^2)T_k)^{-1} e_1 \|\widehat{P}_\perp b\|_M$$

The residual for this solution takes the form

$$r = K_\sigma^{-1}(A\widetilde{x} - f) = (\sigma - \omega^2)R_p z^{(1)} + (\sigma - \omega^2)v_{k+1}\beta_k e_k^T z^{(2)} \tag{2.13}$$

where the second term can be made arbitrarily small by increasing k. The first term is only small if R_p is small enough, i.e. if the Ritz values are chosen accurately enough. We shall see in §3.5 that this is usually the case, when the Ritz values arise from a recycling process in the Lanczos method.

2.4.2. Deflated Matrix Lanczos (DML). In finite precision arithmetic and large k, the latter method is not advised since due to rounding errors the components in the deflated eigenvectors can grow in the Lanczos method. It is therefore usually wise to explicitly orthogonalize, i.e. the Lanczos method is now applied to the deflated matrix

$$\widehat{C} = \widehat{P}_\perp K_\sigma^{-1} M \widehat{P}_\perp.$$

We then have the recurrence relation

$$(\widehat{P}_\perp K_\sigma^{-1} M \widehat{P}_\perp)V_k = V_k T_k + \beta_k v_{k+1} e_k^T$$

where $\widehat{P}_\perp V_k = V_k$. We thus obtain

$$(I - \widehat{U}_p \widehat{U}_p^T M)K_\sigma^{-1} M V_k = V_k T_k + \beta_k v_{k+1} e_k^T$$

and by moving the second term of the left-hand side to the right-hand side

$$K_\sigma^{-1} M V_k = V_k T_k + \widehat{U}_p C + \beta_k v_{k+1} e_k^T \tag{2.14}$$

Since $\widehat{U}_p^T M V_{k+1} = I$ is imposed, this leads us to $C = \widehat{U}_p^T M K_\sigma^{-1} M V_k$. From (2.12), we have that $C = R_p^T M V_k$.

We compute the solution by projecting (1.1) on the space spanned by the Krylov vectors and the Ritz vectors [12], i.e. compute z from

$$\left[\widehat{U}_p \; V_k\right]^T M K_\sigma^{-1} A \left[\widehat{U}_p \; V_k\right] z = \begin{bmatrix} \widehat{U}_p^T M b \\ V_k^T M b \end{bmatrix},$$

which is equivalent to the system

$$\begin{bmatrix} I + (\sigma - \omega^2)\widehat{\Theta}_p & (\sigma - \omega^2)R_p^T M V_k \\ (\sigma - \omega^2)V_k^T M R_p & I + (\sigma - \omega^2)T_k \end{bmatrix} \begin{bmatrix} z^{(1)} \\ z^{(2)} \end{bmatrix} = \begin{bmatrix} \widehat{U}_p^T M b \\ V_k^T M b \end{bmatrix}. \tag{2.15}$$

It is easy to see that the residual is

$$\begin{aligned} r &= K_\sigma^{-1}(f - A(\widehat{U}_p z^{(1)} + V_k z^{(2)})) \\ &= (\omega^2 - \sigma)\beta_k v_{k+1} e_k^T z^{(2)} + (\omega^2 - \sigma)(I - V_k V_k^T M)R_p z^{(1)} \; . \end{aligned} \tag{2.16}$$

There is no guarantee that it is small unless $\|R_p z^{(1)}\|$ is small. We will discuss this in §3.5.

3. Convergence analysis. We show spectral properties in order to understand the convergence behaviour of the iterative process. Since in practical computations, we do not have exact eigenvalues at hand, but approximations, often called Ritz values, we analyze the impact on the quality of deflation.

3.1. Backward error analysis on deflation. When we use $\widehat{U}_p$ for deflation, we introduce an error because of the residual term R_p in (2.12). In this section, we show that this introduces a backward error on the linear system. The theory for deflation of exact eigenvectors can be applied to this perturbed system.

THEOREM 3.1. *Let*

$$\widetilde{K}_\sigma = (K_\sigma^{-1} - \widehat{U}_p R_p^T - R_p \widehat{U}_p^T)^{-1}$$

and let the Lanczos method be applied to $P_\perp K_\sigma^{-1} M P_\perp$ so that (2.14) holds, then

$$\begin{aligned}\widetilde{K}_\sigma^{-1} M \widehat{U}_p &= \widehat{U}_p \widehat{\Theta}_p \\ \widetilde{K}_\sigma^{-1} M V_k &= V_k T_k + \beta_k v_{k+1} e_k^T \ .\end{aligned}$$

Proof. The proof follows from (2.12) and (2.14). ∎

3.2. Vector Padé connection. The Lanczos method makes an approximation to (2.2) of the form

$$\widetilde{x} = \sum_{j=1}^{k} \widehat{u}_j \frac{\widehat{u}_j^T f}{\widehat{\lambda}_j - \omega^2}$$

where $(\widehat{\lambda}_j, \widehat{u}_j)$ is a Ritz pair of (1.2), i.e. $\widetilde{x}$ is a rational function whose poles correspond to Ritz values. In the Padé–via–Lanczos (PVL) method, we have that the first k derivatives of $x(\alpha)$ and $\widetilde{x}(\alpha)$ with $\alpha = \omega^2 - \sigma$ evaluated at $\alpha = 0$ match, see e.g. [4, 22]. In [4], it was argued, that the error grows with $|\alpha|$, i.e. when ω^2 moves away from σ.

In the presence of deflation, the first k derivatives of $x^{(2)}$ match with those of the solution computed by the Lanczos method. As a result, the computed solution x matches the first k derivatives in σ with the exact solution of (1.1). In addition, the solution is an exact interpolation in the poles $\lambda_1, \ldots, \lambda_p$.

When quasi-deflation is used as in §2.4.2, we lose the connection with the vector Padé approximation. However, we build a Padé approximation for the perturbed matrix $\widetilde{K}_\sigma$ defined in Theorem 3.1.

3.3. Spectral convergence analysis. The spectrum of $K_\sigma^{-1} A$ is

$$\phi_j = \frac{\lambda_j - \omega^2}{\lambda_j - \sigma} \quad , \quad j = 1, \ldots, n \ . \tag{3.1}$$

The spectral condition number of $K_\sigma^{-1} A$ is

$$\frac{\max_j\{\phi_j\}}{\min_j\{\phi_j\}}$$

and grows when ω^2 moves away from σ. This can easily be seen as follows. If $\omega^2 = \sigma$, then all $\phi_j = 1$. If ω^2 approaches λ_j, then $|\phi_j|$ becomes large.

FIG. 3.1. *Situation where σ and ω^2 lie in $\mathbb{L}_p$*

LEMMA 3.2. *The matrix B defined in (2.9) is self–adjoint with respect to the M inner product, i.e. $x^T MBy = (By)^T Mx$. In addition, if $\mathbb{L}_p$ contains all eigenvalues of (1.2) between σ and ω^2, then the matrix B, restricted to $\mathcal{U}_p$, is positive definite.* The convergence rate of the Lanczos method for the positive definite matrix B (i.e. the conjugate gradients method) is then bounded from above by

$$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k , \tag{3.2}$$

where κ is the condition number of the matrix B, see e.g. [16, Theorem 3.1.1.]. This leads to the same conclusion as §3.2: the solution for large $|\omega^2 - \sigma|$ is harder to compute than for small $|\omega^2 - \sigma|$.

The condition number of B restricted to $\mathcal{U}_p^{\perp}$ is defined by

$$\begin{aligned}\kappa_M(B) &= \frac{\max_{\det(B-\phi I)=0} |\phi|}{\min_{\det(B-\phi I)=0, \phi\neq 0} |\phi|} \\ &= \max_{\lambda\in\mathbb{L}\backslash\mathbb{L}_p} \left|\frac{\lambda-\omega^2}{\lambda-\sigma}\right| \Big/ \min_{\lambda\in\mathbb{L}\backslash\mathbb{L}_p} \left|\frac{\lambda-\omega^2}{\lambda-\sigma}\right| .\end{aligned}$$

Figure 3.1 shows the situation where σ and ω^2 lie in $\mathcal{I}_p = [\min(\mathbb{L}_p), \max(\mathbb{L}_p)]$. We see that if ω^2 is somewhere in the middle of the interval $\mathcal{I}_p$, $|\lambda - \omega^2|$ and $|\lambda - \sigma|$ are both large so that their ratio is almost one, leading to a small κ. When ω^2 lies close to $\min(\mathbb{L}_p)$, then some $|\phi_j|$ may be small.

Consider as an example, the interval $\mathcal{I}_p = [0, 100]$ and $\sigma = 80$. Let the eigenvalues be $\lambda_j = 5 + 10j$ for $j = 1, \ldots, n = 1000$. Note that $\lambda_1, \ldots, \lambda_{10}$ lie in $\mathcal{I}_p$. Then for $\omega^2 = 10$,

$$\phi_j = \frac{-5+10j}{-75+10j} .$$

We find that $\max_j\{|\phi_j|\} = 25$ and $\min_j\{|\phi_j|\} = 1/13$, so $\kappa = 325$. When we restrict to $\mathbf{R}\backslash\mathcal{I}_p$, we have $\max_j\{|\phi_j|\} = 3$ and $\min_j\{|\phi_j|\} = 9995/9925 \simeq 1$, so that $\kappa \simeq 3$.

LEMMA 3.3. *Let $\mathbb{L}_p$ contain all eigenvalues in $\mathcal{I}_p = [\min(\mathbb{L}_p), \max(\mathbb{L}_p)]$. Let*

$$\begin{aligned}\lambda_m &= \max\{\lambda : \lambda \in \mathbb{L}\backslash\mathbb{L}_p, \lambda \leq \min(\mathbb{L}_p)\} \\ \lambda_M &= \min\{\lambda : \lambda \in \mathbb{L}\backslash\mathbb{L}_p, \lambda \geq \max(\mathbb{L}_p)\}\end{aligned}$$

and

$$\gamma = \frac{\lambda_M - \min(\Omega^2)}{\lambda_M - \sigma} \Big/ \frac{\max(\Omega^2) - \lambda_m}{\sigma - \lambda_m} .$$

If $\sigma \in \mathcal{I}_p$ and $\omega^2 \in \Omega^2$, then

$$\kappa_M(B) \leq \max(\gamma, \gamma^{-1}) .$$

If there are no eigenvalues on the left of $\min(\mathbb{L}_p)$*, then*

$$\gamma = \frac{\lambda_M - \min(\Omega^2)}{\lambda_M - \sigma} \ .$$

A similar conclusion holds when there are no eigenvalues to the right of $\max(\mathbb{L}_p)$. When $\mathbb{L}_p$ includes more and more eigenvalues so that λ_M and λ_m move further away from σ, $\min(\Omega^2)$ and $\max(\Omega^2)$, γ gets closer to one. We can thus include eigenvalues outside Ω^2 into $\mathbb{L}_p$. This makes γ closer to one.

3.4. Quasi deflation. In this section, we perform the spectral analysis for the case that the exact eigenvalues U_p are not available, but Ritz vectors $\widehat{U}_p$, see (2.12). (We will discuss in §4 how $\widehat{U}_p$ and $\widehat{\Theta}_p$ can be computed by the Lanczos method.) The matrix B from the deflated equation from (2.8) becomes

$$\widehat{B} = (I - \widehat{U}_p\widehat{U}_p^T M)K_\sigma^{-1}A(I - \widehat{U}_p\widehat{U}_p^T M) \ .$$

The algorithm in §2.4.2 can be seen as a Krylov method for $\widehat{B}$ augmented with the columns of $\widehat{U}_p$. In this section, we make a spectral analysis of $\widehat{B}$.

Traditional perturbation analysis allows us to prove properties about the spectrum of $\widehat{B}$; see [27, 28, 9]. We now derive a bound on the eigenvalues of $\widehat{B}$.

LEMMA 3.4. *Let*

$$\widehat{C} = (I - \widehat{U}_p\widehat{U}_p^T M)K_\sigma^{-1}M(I - \widehat{U}_p\widehat{U}_p^T M) \ .$$

If $\widehat{\theta}_j$ *is a nonzero eigenvalue of* $\widehat{C}$*, then there is an eigenvalue* θ_j *of* $K_\sigma^{-1}M$ *such that*

$$|\widehat{\theta}_j - \theta_j| \leq \|\sqrt{M}R_p\|_2 \ .$$

Proof. First let

$$\widehat{\Theta}_p = \widehat{U}_p^T M K_\sigma^{-1} M \widehat{U}_p \ .$$

Then, note from

$$\begin{aligned} K_\sigma^{-1}M\widehat{U}_p &= \widehat{U}_p\widehat{\Theta}_p + R_p \\ \widehat{U}^T M K_\sigma^{-1} M &= \widehat{\Theta}_p\widehat{U}_p^T M + R_p^T M \end{aligned}$$

that

$$\widehat{C} = K_\sigma^{-1}M - \widehat{U}_p\widehat{\Theta}_p\widehat{U}_p^T M + \widehat{U}_p R_p^T M + R_p\widehat{U}_p^T M \ . \tag{3.3}$$

Multiplying (3.3) on the left with $\sqrt{M}$ produces

$$\sqrt{M}\widehat{C} = \sqrt{M}K_\sigma^{-1}M - \sqrt{M}\widehat{U}_p\widehat{\Theta}_p\widehat{U}_p^T M + \sqrt{M}\widehat{U}_p R_p^T M + \sqrt{M}R_p\widehat{U}_p^T M \ .$$

Further note that if $\widehat{u} \neq 0$ is an eigenvector associated with $\widehat{\theta}_j \neq 0$, then $\widehat{U}_p^T M\widehat{u} = 0$. Defining $\widehat{w} = \sqrt{M}\widehat{u}$, we have

$$\sqrt{M}\widehat{C}\widehat{u} = \sqrt{M}K_\sigma^{-1}\sqrt{M}\widehat{w} + \sqrt{M}\widehat{U}_p R_p^T\sqrt{M}\widehat{w} \ .$$

By the Bauer-Fike Theorem [28], we have that there is a θ_j, eigenvalue of $\sqrt{M}K_\sigma^{-1}\sqrt{M}$, so that

$$\begin{aligned}|\widehat{\theta}_j - \theta_j| &\leq \|\sqrt{M}U_p R_p^T \sqrt{M}\|_2 \\ &\leq \|R_p^T\sqrt{M}\|_2\end{aligned}$$

This proves the lemma. ∎

THEOREM 3.5. *The spectral properties of $\widehat{B}$ can be summarized as follows :*

1. *$\widehat{B}\widehat{U}_p = 0$.*
2. *$\widehat{B}$ has $n-p$ nonzero eigenvalues $\widehat{\phi}_j$ with eigenvectors M orthogonal to the columns of $\widehat{U}_p$. For each nonzero eigenvalue $\widehat{\phi}_j$, there is an eigenvalue ϕ_j of B so that*

$$|\widehat{\phi}_j - \phi_j| \leq |\omega^2 - \sigma| \|\sqrt{M}R_p\|_2 \ .$$

Proof. The proof follows from Lemma 3.4. Note that

$$\widehat{B} = (I - \widehat{U}_p\widehat{U}_p^T M) + (\sigma - \omega^2)\widehat{C} \ .$$

∎

If the eigenvalues of $\widehat{B}$ are close to the eigenvalues of B, the condition number $\kappa(\widehat{B}) \simeq \kappa(B)$. Let $\phi_j = 1 + (\omega^2 - \sigma)\theta_j$ for $j = 1, \ldots, n$ be the eigenvalues of B, and $\widehat{\phi}_j$ be the eigenvalues of $\widehat{B}$. Recall that the spectral condition number $\kappa(B) = \max|\phi| / \min_{\phi\neq 0}|\phi|$. With $\xi = |\omega^2 - \sigma| \|\sqrt{M}R_p^T\|_2$, and assuming that $|\phi_j| > \xi$, we have that

$$\begin{aligned}\kappa(\widehat{B}) &\leq \frac{\max|\phi| + \xi}{\min_{\phi\neq 0}|\phi| - \xi} \\ &\leq \kappa(B)\frac{\max|\phi| + \xi}{\max|\phi|}\frac{\min_{\phi\neq 0}|\phi|}{\min_{\phi\neq 0}|\phi| - \xi}\end{aligned}$$

If $\mathbb{L}_p$ contains all eigenvalues in $[\min(\Omega^2), \max(\Omega^2]$, the eigenvalues of B are never significantly smaller than one. This implies that the $\widehat{\theta}_j$'s need not be very close to eigenvalues of $K_\sigma^{-1}M$ in order to have $\kappa(\widehat{B}) \simeq \kappa(B)$.

3.5. Residual terms. The residual in (2.13) and (2.16) have two terms. The second term is driven to zero by selecting k high enough. The first term is constant in (2.13) and depends on k in (2.16). In the residual term in (2.13), we have the term

$$(\sigma - \omega^2)R_p e_j \frac{\widehat{u}_j^T M b}{1 + (\omega^2 - \sigma)\widehat{\theta}_j} \simeq R_p e_j \frac{\sigma - \omega^2}{\widehat{\lambda}_j - \omega^2}\widehat{u}_j^T f$$

which is small if $\widehat{\lambda}_j$ lies far away from ω^2. If we want $\|R_p z^{(1)}\|$ smaller than some tolerance, we do not have to put a strong condition on the columns of R_p associated with Ritz values far away from Ω^2.

This also holds to some extent for (2.16). Indeed, by multiplying (2.15) on the left with e_j^T with $1 \leq j \leq p$, we see that

$$\left| e_j^T z^{(1)} - \frac{\widehat{u}_j^T M b}{1 + (\omega^2 - \sigma)\widehat{\theta}_j} \right| \leq \left| \frac{\omega^2 - \sigma}{1 + (\omega^2 - \sigma)\widehat{\theta}_j} \right| \|R_p^T M V_k z^{(2)}\|_2$$

If $\|R_p^T M V_k\|$ is not too large, $e_j^T z_j^{(1)}$ does not vary much in a relative sense and lies around

$$\frac{\widehat{u}_j^T M b}{1+(\omega^2-\sigma)\widehat{\theta}_j} \simeq \frac{\widehat{u}_j^T f}{\widehat{\lambda}_j - \omega^2} \ .$$

In (2.13) and (2.16), the residual has a term of the order

$$\|R_p e_j\| \left| \frac{\widehat{u}_j^T f}{\widehat{\lambda}_j - \omega^2} \right| \ ,$$

which is small if $\widehat{\lambda}_j$ lies far away from ω^2. If $|e_j^T z_j^{(1)}|$ is small, $\|R_p e_j\|$ does not have to be tiny to obtain a small residual norm.

4. Lanczos Ritz values. At this stage, we have not explained how we compute U_p and solve the parameterized linear system (2.8).

Let

$$T_k Y = Y \widehat{\Theta}_k \tag{4.1}$$

be the eigendecomposition of T_k, where $Y = [y_1, \ldots, y_k]$ and $\widehat{\Theta}_k = \mathrm{diag}(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$. The $\widehat{\theta}_j$'s are called Ritz values and the columns of $\widehat{U}_k = V_k Y = [\hat{u}_1, \ldots, \hat{u}_k]$ Ritz vectors. From (2.10), we find that

$$\begin{aligned} \rho_j &= \|K_\sigma^{-1} M \widehat{u}_j - \widehat{\theta}_j \widehat{u}_j\|_M \\ &= \|v_{k+1} \beta_k e_k^T y_j\|_M \\ &= \beta_k |e_k^T y_j| \ . \end{aligned}$$

Then define $\widehat{\lambda}_j = \sigma + \widehat{\theta}_j^{-1}$. If ρ_j is small, we have that

$$K \widehat{u}_j \simeq \widehat{\lambda}_j M \widehat{u}_j \ .$$

The Ritz vectors form a basis of the Krylov space. Following (2.11), the solution of (1.1) can thus be expressed in terms of Ritz vectors as follows :

$$x = \sum_{j=1}^{k} \hat{u}_j \frac{\hat{u}_j^T M b}{1-(\omega^2-\sigma)\hat{\theta}_j} = \sum_{j=1}^{k} \hat{u}_j \frac{\hat{w}_j^T f}{\hat{\lambda}_j - \omega^2} \ , \tag{4.2}$$

with

$$\hat{w}_j = \hat{u}_j + v_{k+1} \frac{\beta_k e_k^T y_j}{\hat{\theta}_j}$$

the purified Ritz vector [25].

In this section, we analyze how close these Ritz values are to the eigenvalues of B when the Lanczos method is applied to solve (1.1).

Usually, the stopping criterion for solving (1.1) takes the form

$$\|r\|/(\|b\| + \|B\|\|x\|) \leq \tau \ ,$$

where τ is a prescribed tolerance. The following theorem shows that the residual norms of the Ritz pairs corresponding to $\hat{\lambda}_j$ in $[\min(\Omega^2), \max(\Omega^2)]$ is proportional to the residual tolerance for the linear system.

THEOREM 4.1. *Let $(\hat{\theta}_j, \hat{u}_j)$, $j = 1, \ldots, k$ be the Ritz pairs from the Lanczos method. If*

$$\|r\|_M \leq \tau(\|b\|_M + \|B\|\|x\|_M) \tag{4.3}$$

for all $\omega^2 \in \Omega^2$, then

$$\|K_\sigma^{-1} M \hat{u}_j - \hat{\theta}_j \hat{u}_j\|_M \leq 6\tau |\hat{\theta}_j| \|B\|_M$$

when $\hat{\lambda}_j = \sigma + \hat{\theta}_j^{-1} \in \Omega^2$.

Proof. The proof is similar to the proof of Lemma 4.1 in [22]. Let $\alpha = \omega^2 - \sigma$. From

$$\begin{aligned} r &= b - K_\sigma^{-1} A \widetilde{x} \\ K_\sigma^{-1} A &= I - \alpha K_\sigma^{-1} M \ , \end{aligned}$$

$b = v_1 \|b\|_M$, and (2.11), we have

$$\begin{aligned} r &= \|b\|_M v_1 - V_k(I - \alpha T_k) z + \alpha \beta_k v_{k+1} e_k^T z \\ &= \alpha \beta_k v_{k+1} e_k^T z \ . \end{aligned}$$

Next, from (4.1) and (4.2), we have that

$$z = \sum_{j=1}^{k} y_j \frac{y_j^T e_1 \|b\|_M}{1 - \alpha \hat{\theta}_j} = \sum_{j=1}^{k} y_j \frac{\hat{u}_j^T M b}{1 - \alpha \hat{\theta}_j} = \sum_{j=1}^{k} y_j \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2} (\hat{u}_j^T M b) \ .$$

With $\rho_j = \beta_k e_k^T y_j$, we have

$$\begin{aligned} r &= \sum_{j=1}^{k} \alpha \beta_k v_{k+1} e_k^T y_j \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2} (\hat{u}_j^T M b) \\ &= \alpha v_{k+1} \sum_{j=1}^{k} \rho_j \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2} (\hat{u}_j^T M b) \ . \end{aligned}$$

For each $i = 1, \ldots, k$, for which $\hat{\lambda}_j \in \mathcal{I}_p$, we can determine $\omega^2 \in \mathcal{I}_p$ so that

$$\begin{aligned} \left| \rho_i \frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2} (\hat{u}_i^T M b) \right| &\geq \left| \sum_{j \neq i} \rho_j \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2} (\hat{u}_j^T M b) \right| \\ \left| \frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2} (\hat{u}_i^T M b) \right| &\geq \left(\sum_{j \neq i} \left| \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2} (\hat{u}_j^T M b) \right|^2 \right)^{1/2} \\ \left| \frac{\hat{\lambda}_i - \omega^2}{(\hat{\lambda}_i - \sigma)(\hat{u}_i^T M b)} \right| \|b\|_M &\leq \|B\| \\ |\omega^2 - \sigma| &\geq |\hat{\lambda}_i - \sigma| \end{aligned}$$

since $|\hat{\lambda}_i - \omega^2|$ can be made arbitrarily small, by picking ω^2 close to $\hat{\lambda}_i$. Note that in the Lanczos process all Ritz values are simple, so the terms in the summation for x and r with $j \neq i$ remain small.

We then have

$$\|r\|_M/|\alpha| \geq \left|\rho_i \frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2}(\hat{u}_i^T M b)\right| - \left|\sum_{j\neq i} \rho_j \frac{\hat{\lambda}_j - \sigma}{\hat{\lambda}_j - \omega^2}(\hat{u}_j^T M b)\right|$$

$$\left|\rho_i \frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2}(\hat{u}_i^T M b)\right| \leq 2\|r\|_M/|\alpha| \tag{4.4}$$

and

$$\|x\|_M \leq 2\left|\frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2}(\hat{u}_i^T M b)\right| .$$

From (4.3) and (4.4), we have that

$$\left|\rho_i \frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2}(\hat{u}_i^T M b)\right| \leq 2\frac{\tau}{|\alpha|}\left(\|b\|_M + 2\|B\|\left|\frac{\hat{\lambda}_i - \sigma}{\hat{\lambda}_i - \omega^2}(\hat{u}_i^T M b)\right|\right)$$

$$|\rho_i| \leq 2|\alpha|\frac{\tau}{|\alpha|}\left(\frac{\hat{\lambda}_i - \omega^2}{(\hat{\lambda}_i - \sigma)(\hat{u}_i^T M b)}\|b\|_M + 2\|B\|\right)$$

Note that

$$\|K_\sigma^{-1} M \hat{u}_j - \hat{\theta}_j \hat{u}_j\|_M = |\rho_j| .$$

This proves the theorem. ∎

So, the backward error on the linear solves determines the backward error on the Ritz pairs. A reasonable precision of x for all $\omega \in \mathcal{I}_p$ should does provide accurate enough Ritz values near $\omega^2 \in \Omega^2$.

5. Multiple right-hand sides. The goal is to solve

$$(K - \omega^2 M)[x_1, \ldots, x_s] = [f_1, \ldots, f_s] . \tag{5.1}$$

We could use a block Krylov method for solving all right-hand sides at once. The alternative is to solve each system independently, which can be useful for saving memory (less vectors to store), or is the only option when the right-hand sides are not available at once. The idea is to solve (5.1) column by column by separate Lanczos processes.

We use the following algorithms:

ALGORITHM 5.1 (Multiple RHS solution using DRHSL).

1. Solve $K_\sigma^{-1}(K - \alpha M)x_1 = K_\sigma^{-1} f_1$ using the parameterized Lanczos method.
2. Compute Ritz pairs $(\widehat{\Theta}_p, \widehat{U}_p)$ of $K_\sigma^{-1} M$ and let $\widehat{\Lambda}_p = \sigma I + \widehat{\Theta}_p^{-1}$.
3. For $j = 2, 3, \ldots, s$:
   - 3.1. Solve $K_\sigma b_j = f_j$.
   - 3.2. Solve the diagonal system $(\widehat{\Lambda}_p - \sigma I)^{-1}(\widehat{\Lambda}_p - \omega^2 I)z_j = \widehat{U}_p^T M b_j$.
   - 3.3. Solve $K_\sigma^{-1}(K - \alpha M)\widetilde{x}_j = \widehat{P}_\perp b_j$ using the parameterized Lanczos method.
   - 3.4. Let the solution be $x_j = U_p z_j + \widetilde{x}_j$.

ALGORITHM 5.2 (Multiple RHS solution using DML).
1. Solve $K_\sigma^{-1}(K-\alpha M)x_1 = K_\sigma^{-1}f_1$ using the parameterized Lanczos method.
2. Compute Ritz pairs $(\widehat{\Theta}_p, \widehat{U}_p)$ of $K_\sigma^{-1}M$ and let $\widehat{\Lambda}_p = \sigma I + \widehat{\Theta}_p^{-1}$.
3. For $j = 2, 3, \ldots, s$:
   3.1. Solve $K_\sigma b_j = f_j$.
   3.2. Compute the Krylov space for $\hat{P}_\perp K_\sigma^{-1} M \hat{P}_\perp$ with starting vector $\widehat{P}_\perp K_\sigma^{-1} f_j$.
   3.3. Solve $[\widehat{U}_p\ V_k]^T M K_\sigma^{-1} A [\widehat{U}_p\ V_k] z_j = [\widehat{U}_p\ V_k]^T M b_j$ for z_j.
   3.4. Let the solution be $x_j = [\widehat{U}_p\ V_k] z_j$.

There are only few implementation differences between the two algorithms. Algorithm 5.1 does not require the storage of the Lanczos vectors since the solution can be updated at each iteration step. However, reorthogonalization cannot be performed which may lead to loss of precision, see [22]. Algorithm 5.2 requires the storage of V_k, which may be hard when k is large.

We have mentioned in the previous sections that the deflated Ritz pairs do not have to have small residual norms since the condition number does not change much. However, a small residual term $R_p z^{(1)}$ can generally only be achieved when $\|R_p\|$ is not large.

If the p deflated Ritz pairs are computed by another method than the Lanczos linear solver itself, we may assume that $\|R_p\|$ in (2.16) is small. In this case, we have exact deflation.

If the Ritz pairs are computed by the Lanczos method for the first right-hand side as in Algorithms 5.1 and 5.2, the residual terms related to Ritz values in Ω^2 are small following the analysis in §4. All the eigenvalues in Ω are computed, otherwise the residual norm cannot be small for all ω's. This suggests that all linear systems in Step 2.3 are positive definite and condition numbers are most likely good. Only when $u_j^T M b = 0$, the corresponding eigenvalue is not computed.

We have shown that k usually is not large, since the condition number $\kappa(B)$ is small. There is one situation where $\kappa(B)$ can be large, i.e. when $\widehat{\Lambda}_p$ does not contain approximations to all eigenvalues in Ω^2. Although this is impossible to happen in theory when $u_j^T M b \neq 0$, it may happen in practice when eigenvalues are clustered or multiple, since the Ritz approximations may not be present for all eigenvalues.

A problem may arise when also Ritz values outside Ω^2 are deflated, since their residual norms are not bounded by Theorem 4.1. However, from [4], the error on x usually increases more or less monotonically when ω^2 goes away from σ. From §4, we may conclude that the Ritz residual norms also increase more or less monotonically. We now have a closer look at (2.16) and §3.5. If $\widehat{\lambda}_j$ lies outside Ω^2,

$$|e_j^T z^{(1)}| \simeq \left| \frac{u_j^T f}{\widehat{\lambda}_j - \omega^2} \right|$$

is small. The corresponding term in term $R_p z^{(1)}$ is likely to be small as well. Therefore, we do not expect large residual terms for Ritz values selected outside Ω^2.

6. Numerical Example. In this section, we show the numerical performance of Algorithms 5.1 and 5.2 for a test problem arising from a a structural model of car windscreen. This is a 3D problem discretized with 7564 nodes and 5400 linear hexahedral elements (3 layers of 60×30 elements). The mesh is shown in Figure 6.1. The material is glass with the following properties: the Young modulus is $7\,10^{10}\mathrm{N/m^2}$, the density is $2490\mathrm{kg/m^3}$, and the Poisson ratio is 0.23. The structural boundaries are

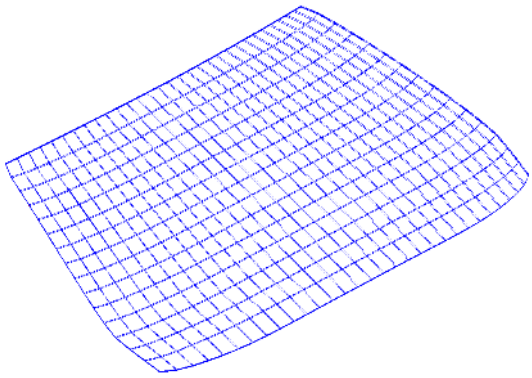

FIG. 6.1. *Mesh of the car windscreen*

free (free-free boundary conditions). The plate is subjected to a point force applied on a corner node [22] [1].

The discretized problem has dimension $n = 22,692$. The goal is to compute $x(\omega)$ with $\omega \in [0, 100]$. In order to generate the plots the frequency range was discretized as $\{\omega_1, \ldots, \omega_m\} = \{0.5j, j = 1, \ldots, m\}$ with $m = 400$. We used shift $\sigma = 1$. The MUMPS direct solver [24] was employed for the solution of the linear system of the coefficient K_σ in the Lanczos procedure and Algorithms 5.1 and 5.2.

We performed a first run with right-hand side f with $f_j = 0$ for $j \neq 5673$ and $f_{5673} = 1$, which corresponds to a point load on a corner of the windscreen. We used $k = 20$ Lanczos iterations. Then we kept the Ritz values below 2×100^2. There are $p = 14$ Ritz values in this interval. The residual norms of the kept Ritz pairs are smaller than 3×10^{-7}. The largest residual norm is for the Ritz value corresponding to $\omega = 104$, which is just outside the interval Ω.

Next, we performed a second run with the right-hand side f with $f_j = 0$ for $j > 1$ and $f_1 = 1$. We used 6 (additional) Lanczos iterations to make a total of $p + k = 20$ vectors. For given ω,

$$\kappa = \max_{j>p} \frac{\lambda_j - \omega^2}{\lambda_j - \sigma} / \min_{j>p} \frac{\lambda_j - \omega^2}{\lambda_j - \sigma} < \widetilde{\kappa} := 1 \cdot \max_{j>p} \frac{\lambda_j - \sigma}{\lambda_j - \omega^2} = \frac{\lambda_{p+1} - \sigma}{\lambda_{p+1} - \omega^2}$$

The largest $\widetilde{\kappa}$ is for $\omega = \max(\Omega)$. In this example, the maximum $\widetilde{\kappa}$ is $(142.089^2 - 1)/(142.089^2 - 100^2) = 1.9813$, so the convergence ratio in (3.2) is 0.1693. After six iterations, the error norm is reduced by approximately 2×10^{-5}. Figure 6.2 shows the results. Both Algorithms 5.1 and 5.2 produce the same results: solution and error curves cannot be distinguished in the figures. The additional iterations (only six) is low so that loss of orthogonality in the Lanczos vectors is most likely not having an impact, and the Ritz vectors have small residual norms so that the deflation is practically perfect as in Lemma 2.1.

We performed a third run with f being 1 everywhere. The conclusions are similar as for the previous situation. Figure 6.3 shows the results.

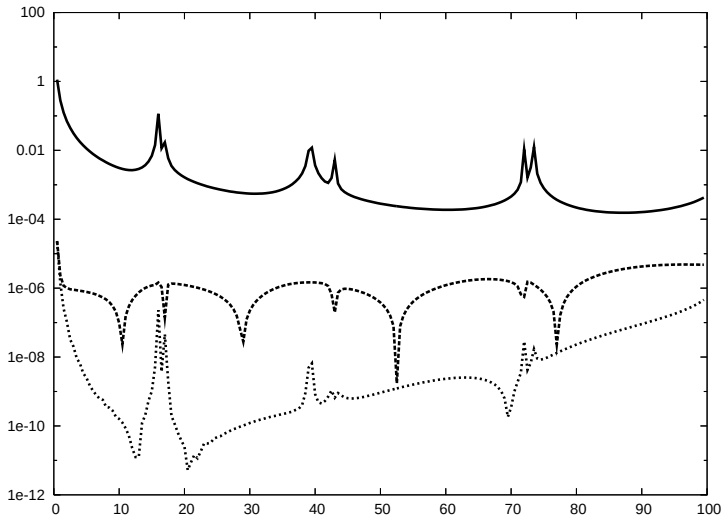

FIG. 6.2. *Solution norm $\|x\|_2$ in function of ω in solid line, error on $\|x\|_2$ for superposition on 14 modes as a dashed line and error on on $\|x\|_2$ for superposition on 14 modes and 6 additional Lanczos iteration as a dotted line.*

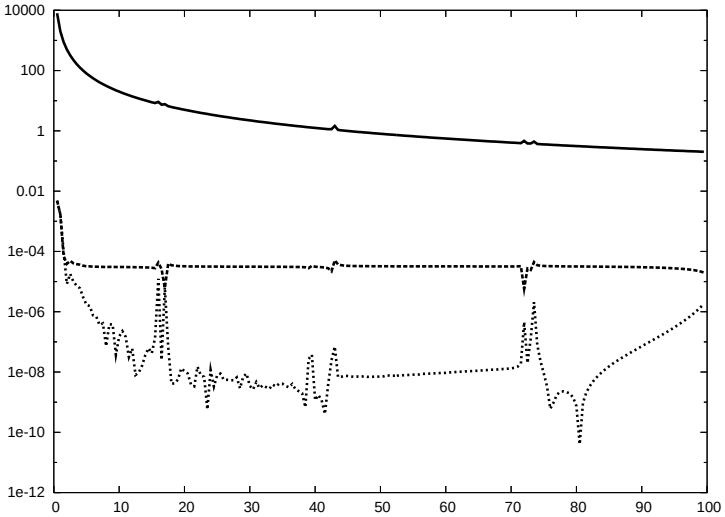

FIG. 6.3. *Solution norm $\|x\|_2$ in function of ω in solid line, error on $\|x\|_2$ for superposition on 14 modes as a dashed line and error on on $\|x\|_2$ for superposition on 14 modes and 6 additional Lanczos iteration as a dotted line.*

7. Conclusions. We have applied the notion of recycling Ritz vectors to linear systems with a parameter in frequency sweeping. We have presented an algorithm and theory for the symmetric case. We have given various arguments and a numerical example that show that recycling may significantly reduce the number of iterations in the Lanczos method.

Further extensions of this work lie in the application to proportional damping as in [23], and using the Lanczos method as iterative method in the AMLS frequency

sweeping method [18].

Acknowledgements. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

Bai was supported in part by the National Science Foundation grant DMS-0611548.

REFERENCES

[1] Oberwolfach model reduction benchmark collection, 2004. http://www.imtek.uni-freiburg.de/simulation/benchmark/.
[2] Z. Bai and R. Freund. A symmetric band Lanczos process based on coupled recurrences and some applications. Numerical Analysis Manuscript 00-8-04, Bell Laboratories, Murray Hill, New Jersey, 2000.
[3] Z. Bai and R. Freund. A partial Padé-via-Lanczos method for reduced-order modeling. *Linear Alg. Appl.*, 332–334:141–166, 2001.
[4] Z. Bai and Q. Ye. Error estimation of the Padé approximation of transfer functions via the Lanczos process. *ETNA*, 7:1–17, 1998.
[5] J. K. Bennighof and M. F. Kaplan. Frequency sweep analysis using multi-level substructuring, global modes and iteration. In *Proceedings of 39th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference*, 1998.
[6] J. K. Bennighof and C. K. Kim. An addaptive multi-level substururing method for efficient modeling of complex structures. In *Proceedings of the AIAA 33rd SDM Conference, Dallas, Texas*, pages 1631–1639, 1992.
[7] J. K. Bennighof and R. B. Lehoucq. An automated multilevel substructuring method for eigenspace computation in linear elastodynamics. *SIAM J. Sci. Comput.*, 25(6):2084–2106, 2004.
[8] Tony F. Chan and Michael K. Ng. Galerkin projection methods for solving multiple linear systems. *SIAM Journal on Scientific Computing*, 21(3):836–850, 1999.
[9] F. Chatelin. *Eigenvalues of matrices*. John Wiley and Sons, 1993.
[10] R. K. Clough and J. Penzien. *Dynamics of Structures*. Mac Graw Hill.
[11] D. Darnell, R.B. Morgan, and W. Wilcox. Deflated GMRES for systems with multiple shifts and multiple right-hand sides. *???*, ?, 2007.
[12] Kilmer M. E. and E. de Sturler. Recycling subspace information for diffuse optical tomography. *SIAM Journal on Scientific Computing*, 27(6):2140–2166, 2006.
[13] P. Feldman and R. W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Trans. Computer-Aided Design*, CAD-14:639–649, 1995.
[14] A. Feriani, F. Perotti, and V. Simoncini. Iterative system solvers for the frequency analysis of linear mechanical systems. Technical Report 1077, Instituto di Analysi Numerica - CNR, Via Addiategrasso, 20, 27100 Pavia, Italy, 1998.
[15] K. Gallivan, E. Grimme, and P. Van Dooren. A rational Lanczos algorithm for model reduction. *Numerical Algorithms*, 12:33–63, 1996.
[16] A. Greenbaum. *Iterative methods for solving linear systems*. SIAM, Philadelphia, PA, USA, 1997.
[17] R.G. Grimes, J.G. Lewis, and H.D. Simon. A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems. *SIAM J. Matrix Anal. Applic.*, 15:228–272, 1994.
[18] J. H. Ko and Z. Bai. High-frequency response analysis via algebraic substructuring, 2007.
[19] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand.*, 45:255–282, 1950.
[20] C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bur. of Stand.*, 49:33–53, 1952.
[21] K. Meerbergen. The rational Lanczos method for Hermitian eigenvalue problems. *Numerical Linear Algebra with Applications*, 8:33–52, 2001.
[22] K. Meerbergen. The solution of parametrized symmetric linear systems. *SIAM J. Matrix Anal. Appl.*, 24(4):1038–1059, 2003.
[23] K. Meerbergen. Fast frequency response computation for Rayleigh damping. *International Journal of Numerical Methods in Engineering*, 73(1):96–106, 2008.
[24] MUMPS. MUlfrontal Massively Parallel Solver, 2001. `http://graal.ens-lyon.fr/MUMPS/`.

[25] B. Nour-Omid, B.N. Parlett, T. Ericsson, and P.S. Jensen. How to implement the spectral transformation. *Mathematics of Computation*, 48:663–673, 1987.

[26] M.L. Parks, E. De Sturler, G. Mackey, D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. Technical Report UIUCDS-R-2004-2421 (CS), UILU-ENG-2004-1722 (ENGR), University of Illinois at Urbana-Champaign, 2004.

[27] B.N. Parlett. *The symmetric eigenvalue problem*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, USA, 1998.

[28] Y. Saad. *Numerical methods for large eigenvalue problems*. Algorithms and Architectures for Advanced Scientific Computing. Manchester University Press, Manchester, UK, 1992.

[29] Y. Saad, M. Yeung, J. Erhel, and F. Guyomarc'h. A deflated version of the conjugate gradient algorithm. *SIAM Journal on Scientific Computing*, 21(5):1909–1926, 2000.

[30] V. Simoncini. Linear systems with a quadratic parameter and application to structural dynamics. In D. R. Kincaid and A. Elster, editor, *Iterative Methods in Scientific Computation II*, volume 5 of *IMACS Series in Computational and Applied Mathematics*, pages 451–461, 1999.

[31] V. Simoncini and F. Perotti. On the numerical solution of $(\lambda^2 A+\lambda B+C)x = b$ and application to structural dynamics. *SIAM J. Sci. Comput.*, 23(6):1876–1898, 2002.

[32] S. Wang, E. de Sturler, and G. H. Paulino. Large-scale topology optimization using preconditioned Krylov subspace methods with recycling. *Intern. J. Numerical Methods in Engineering*, 69(12):2441–2468, 2007.

[33] E. L. Wilson, M.-W. Yuan, and J. M. Dickens. Dynamic analysis by direct superposition of Ritz vectors. *Earthquake Engineering and Structural Dynamics*, 10:813–821, 1982.