On the nonnormality of subiteration for a fluid-structure-interaction problem

Citation for published version (APA):

Brummelen, van, E. H., & Borst, de, R. (2005). On the nonnormality of subiteration for a fluid-structure-interaction problem. *SIAM Journal on Scientific Computing*, 27(2), 599-621. <https://doi.org/10.1137/S1064827503431430>

DOI:

[10.1137/S1064827503431430](https://doi.org/10.1137/S1064827503431430)

Document status and date:

Published: 01/01/2005

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

ON THE NONNORMALITY OF SUBITERATION FOR A FLUID-STRUCTURE-INTERACTION PROBLEM*

E. H. VAN BRUMMELEN[†] AND R. DE BORST[†]

Abstract. Subiteration forms the basic iterative method for solving the aggregated equations in fluid-structure-interaction problems, in which the fluid and structure equations are solved alternately subject to complementary partitions of the interface conditions. In the present work we establish for a prototypical model problem that the subiteration method can be characterized by recursion of a nonnormal operator. This implies that the method typically converges nonmonotonously. Despite formal stability, divergence can occur before asymptotic convergence sets in. It is shown that the transient divergence can amplify the initial error by many orders of magnitude, thus inducing a severe degradation in the robustness and efficiency of the subiteration method. Auxiliary results concern the dependence of the stability and convergence of the subiteration method on the physical parameters in the problem and on the computational time step.

Key words. fluid-structure interaction, subiteration, nonnormality, monolithic methods, convergence and stability

AMS subject classifications. 65N12, 65N22

DOI. 10.1137/S1064827503431430

1. Introduction. Numerical solution methods for fluid-structure interaction are of great importance in a multitude of physical and engineering disciplines, e.g., aerospace engineering [7, 8], offshore engineering [4], biomechanics [2, 13], and civil engineering [24]. Such solution methods are confronted with two fundamental problems. First, the disparity of the initial-boundary-value problems pertaining to the fluid and the structure generally yields severely ill-conditioned systems of equations. Second, the free-boundary character of the interface between the fluid and the structure yields an interdependence between the fluid and structure solutions and their domains of definition, and the computational expenses incurred by implicit treatment of this interdependence are prohibitive in actual applications. Furthermore, several practical impediments emanate from the inherent interconnection between the fluid and structure, such as the loss of software modularity; see [9].

The customary approach to sidestep the aforementioned problems is through *partitioning*. The fluid and structure equations are then solved alternately, subject to complementary partitions of the interface conditions. This iterative procedure is commonly referred to as subiteration, successive approximation, Picard iteration, or block Gauss–Seidel iteration. The conventional implementation of this method imposes the dynamic interface conditions (tractions) on the structure, and the kinematic interface conditions (displacements) on the fluid; see, e.g., [1, 4, 16, 17]. The principal advantage of partitioned solution methods is the affected separation of the fluid-structure-interaction problem into two common subproblems, viz., the independent solution of the initial-boundary-value problems corresponding to the fluid and the structure on prescribed computational domains, and the separate displacement of the sections of the boundaries of the computational domains associated with the interface. The

*Received by the editors July 16, 2003; accepted for publication (in revised form) December 16, 2004; published electronically October 31, 2005.

<http://www.siam.org/journals/sisc/27-2/43143.html>

[†]Delft University of Technology, Faculty of Aerospace Engineering, P.O. Box 5058, 2600 GB Delft, The Netherlands (E.H.vanBrummelen@lr.tudelft.nl, R.deBorst@lr.tudelft.nl).

essential disadvantage of partitioned solution methods pertains to their convergence behavior. Regularly, prohibitory small time steps are required to maintain stability, or convergence is excessively slow.

In the present work we investigate the properties of subiteration for a prototypical fluid-structure-interaction problem. An important new result is that the subiteration method can be characterized by recursion of a nonnormal operator. This implies that the appropriate norm of the operator and its spectral radius are disparate, which reflects in profoundly different transient and asymptotic convergence behavior. In particular, despite formal stability of the operator recursion, convergence can be non-monotonous and transient divergence can precede the asymptotic convergence. This transient divergence can in fact amplify the initial error by many orders of magnitude, thus inducing a severe degradation in the robustness and efficiency of the method. Such transient behavior of one-parameter functions of nonnormal operators has recently also attracted much attention in the analysis of the convergence of numerical methods (see [12] for an overview), and in the field of hydrodynamic stability in [23, 26], where it has been propounded as a fundamental principle in the transition to turbulence.

The analysis in this paper moreover provides useful information on several other relevant properties of the subiteration method, e.g., the dependence of its stability and convergence on physical parameters and on the computational time step.

The contents of this paper are organized as follows: Section 2 states the piston problem. Section 3 describes the subiteration method for fluid-structure-interaction problems. Section 4 derives the characteristic operator associated with the subiteration method on the basis of a linear model problem. Section 5 establishes the nonnormality of the characteristic operator in the absence of reflections and investigates the properties of the operator by approximation on polynomials. Section 6 is concerned with the properties of the general characteristic operator. Finally, section 7 contains concluding remarks.

2. Problem statement. In this paper we restrict ourselves to the analysis of a prototypical fluid-structure model problem, viz., the piston problem. This one-dimensional problem possesses the prominent features of fluid-structure-interaction problems. Below, we present a concise specification of the problem. For further details, see, e.g., [5, 19].

2.1. Piston problem. The piston system comprises the Euler equations of gas dynamics in connection with a simple harmonic oscillator at the fluid/structure interface. We consider the Euler equations in conservative form:

$$(2.1a) \quad \partial_t \mathbf{q} + \partial_x \mathbf{f}(\mathbf{q}) = 0, \quad (x, t) \in \Omega_\alpha,$$

with $\Omega_\alpha := \{(x, t) : 0 < t < \tau; 0 < x < \alpha(t)\}$ an open bounded space/time domain. The function $\mathbf{q}$ maps Ω_α into $\mathbb{R}^3$ and

$$(2.1b) \quad \mathbf{f}(\mathbf{q}) := (q_2, q_2^2/q_1 + p(\mathbf{q}), (p(\mathbf{q}) + q_3) q_2/q_1)^T, \quad p(\mathbf{q}) := (\gamma - 1) \left(q_3 - \frac{1}{2} q_2^2/q_1 \right),$$

with, typically, $\gamma = 7/5$. The boundary $\partial\Omega_\alpha$ consists of the *interface* between the fluid and the structure, $\Gamma_\alpha := \{(x, t) : x = \alpha(t); 0 \leq t \leq \tau\}$, and the fixed boundary $\partial\Omega_\alpha \setminus \Gamma_\alpha$. The function $\mathbf{q}$ in (2.1) is subject to the auxiliary conditions

$$(2.2) \quad q_2(0, t) = 0, \quad 0 < t < \tau, \quad \text{and} \quad \mathbf{q}(x, 0) = \mathbf{q}^0(x), \quad 0 < x < \alpha(0),$$

with $\mathbf{q}^0(x)$ prescribed initial conditions. Moreover, $\mathbf{q}$ must comply with the kinematic interface conditions

$$(2.3a) \quad q_2(\alpha(t), t) = q_1(\alpha(t), t) \dot{\alpha}(t), \quad 0 < t < \tau,$$

$$(2.3b) \quad \alpha(t) = \alpha^0 + z(t), \quad 0 < t < \tau,$$

and the dynamic interface condition

$$(2.3c) \quad p(\mathbf{q}(\alpha(t), t)) = \pi(t), \quad 0 < t < \tau,$$

where $\pi(t)$ and $z(t)$ are interrelated by the harmonic-oscillator equation

$$(2.4) \quad M\ddot{z} + Kz = \pi(t) - p^0 \quad 0 < t < \tau, \quad \text{subject to} \quad z(0) = z^0, \quad \dot{z}(0) = \dot{z}^0,$$

for certain given constants $M, K \in \mathbb{R}_+$ and $z^0, \dot{z}^0 \in \mathbb{R}$.

To establish that the piston problem (2.1)–(2.3) constitutes a *free-boundary* problem, we notice first that the kinematic condition (2.3a) induces an interdependence between the state variables $\mathbf{q}$ and their domain of definition Ω_α . Second, the number of interface conditions is one more than the number of auxiliary conditions required by the initial-boundary-value problems for the fluid and the structure separately. For instance, (2.1)–(2.2) subject to (2.3a) with α given, and (2.4) subject to (2.3c) with p given, are valid problems independently. The above two properties are the distinguishing characteristics of a free-boundary problem.

2.2. Variational formulation. To facilitate the ensuing presentation, we condense the piston problem into the canonical variational form: Find $u \in \mathcal{U}$ such that

$$(2.5a) \quad \mathbf{P}(\bar{u}, u) = \mathbf{p}(\bar{u}) \quad \forall \bar{u} \in \bar{\mathcal{U}}.$$

In (2.5), u represents the quadruple $u := (\mathbf{q}, \alpha, z, \pi)$ in the product space $\mathcal{U} := \mathcal{Q} \times \mathcal{A} \times \mathcal{Z} \times \mathcal{P}$, which consists of the product of the admissible functions for the fluid, the interface displacement, the structure displacement, and the interface traction. The test space $\bar{\mathcal{U}}$ has a similar structure. The functionals $\mathbf{P} : \bar{\mathcal{U}} \times \mathcal{U} \mapsto \mathbb{R}$ and $\mathbf{p} : \bar{\mathcal{U}} \mapsto \mathbb{R}$ can be expanded according to

$$(2.5b) \quad \mathbf{P}((\bar{\mathbf{q}}, \bar{\alpha}, \bar{z}, \bar{\pi}), (\mathbf{q}, \alpha, z, \pi)) := \mathbf{E}(\bar{\mathbf{q}}, \mathbf{q}, \alpha) + \mathbf{S}(\bar{z}, z, \pi) + \mathbf{K}(\bar{\alpha}, \alpha, z) + \mathbf{D}(\bar{\pi}, \pi, \mathbf{q}, \alpha),$$

$$(2.5c) \quad \mathbf{p}((\bar{\mathbf{q}}, \bar{\alpha}, \bar{z}, \bar{\pi})) := \mathbf{e}(\bar{\mathbf{q}}) + \mathbf{s}(\bar{z}) + \mathbf{k}(\bar{\pi}).$$

In (2.5), the functionals $\mathbf{E}, \mathbf{e}$ are associated with the initial-boundary-value problem for the fluid, viz., (2.1) complemented with the kinematic condition (2.3a). The functionals $\mathbf{S}, \mathbf{s}$ pertain to the initial-value problem for the harmonic oscillator (2.4), and $\mathbf{K}, \mathbf{k}$, and $\mathbf{D}$ are associated with the kinematic interface condition (2.3b) and the dynamic interface condition (2.3c), respectively. A specification of the above functionals for the piston problem is provided in [5]. In [6], van Brummelen, Michler, and Borst elaborate upon the formulation of a generic fluid-structure-interaction problem into the canonical variational form (2.5). On account of the canonical variational form, sufficient conditions for the existence of a unique solution to (2.5) can be derived from the generalized nonlinear Lax–Milgram theorem; cf. Appendix A.

3. Subiteration for fluid-structure-interaction problems. The interdependence of the state variables of the fluid and the structure and their domains of definition poses profound complications in the numerical treatment of fluid-structure-interaction problems. The subiteration method which we consider in this paper is the standard approach to bypass these complications. In this section, we first consider the problems pertaining to the application of Newton’s method to the piston problem. We then proceed with a description of the subiteration method.

TABLE 1
Illustration of the connectivity in the fluid-structure-interaction problem.

	α	$\mathbf{q}$	π	z
K	$\times$	0	0	$\times$
E	$\times$	$\times$	0	0
D	$\times$	$\times$	$\times$	0
S	0	0	$\times$	$\times$

3.1. Newton's method. Provided with an initial approximation $u_0 \in \mathcal{U}$, Newton's method for solving the variational problem (2.5) is defined by the following iterative process: for $j = 1, 2, \dots$, find u_j such that

$$(3.1) \quad \mathbf{P}'(\bar{u}, u_{j-1}, u_j - u_{j-1}) = \mathbf{p}(\bar{u}) - \mathbf{P}(\bar{u}, u_{j-1}) \quad \forall \bar{u} \in \bar{\mathcal{U}},$$

with $\mathbf{P}'$ the (assumed existent) Fréchet derivative of the operator $\mathbf{P}$ with respect to its nonlinear argument. If the initial estimate u_0 is sufficiently close to the actual solution, then u_j converges to the solution of the variational problem as $j \rightarrow \infty$.

The above abstract formulation of Newton's method obscures two essential problems, related to the characteristic properties of fluid-structure-interaction problems. The first problem concerns the *shape derivative* $\mathbf{E}'_\alpha$, viz., the derivative of the functional pertaining to the initial-boundary-value problem for the fluid, $\mathbf{E}$, with respect to the interface position, α , induced the free-boundary character. Discrete-approximation methods such as finite elements typically use boundary fitted meshes. As a perturbation of the interface generally yields a deformation of the mesh throughout the entire computational domain, $\mathbf{E}'_\alpha$ behaves as a nonlocal operator. This renders the computational expenses incurred in its evaluation prohibitive in actual applications. The second problem is the inherent interconnection between the fluid and the structure, induced by the interface conditions. The functionals $\mathbf{D}$ and $\mathbf{K}$ depend on $\alpha, \mathbf{q}$, associated with the fluid, as well as on π, z , corresponding to the structure. This is illustrated in Table 1. Consequently, the matrix $\mathbf{P}$ associated with a discrete approximation of the operator $\mathbf{P}'$ in (3.1) is inseparable and the fluid and structure equations must be resolved simultaneously. This has severe practical disadvantages, such as the loss of modularity; see [9]. Moreover, due to the disparate properties of the initial-boundary-value problems associated with the fluid and the structure, the matrix $\mathbf{P}$ is generally severely ill-conditioned.

3.2. Subiteration. The aforementioned complications of Newton's method can be avoided by means of an iterative solution procedure often referred as *subiteration*. Provided with an initial approximation of the structure position $z_0(t)$, the following iterative process defines the subiteration algorithm: for $j = 1, 2, \dots$:

- (S1) Solve the kinematic condition: find $\alpha_j \in \mathcal{A}$: $\mathbf{K}(\bar{\alpha}, \alpha_j, z_{j-1}) = \mathbf{k}(\bar{\alpha}) \quad \forall \bar{\alpha} \in \bar{\mathcal{A}}$.
- (S2) Solve the fluid: find $\mathbf{q}_j \in \mathcal{Q}$: $\mathbf{E}(\bar{\mathbf{q}}, \mathbf{q}_j, \alpha_j) = 0 \quad \forall \bar{\mathbf{q}} \in \bar{\mathcal{Q}}$.
- (S3) Solve the dynamic condition: find $\pi_j \in \mathcal{P}$: $\mathbf{D}(\bar{\pi}, \pi_j, \mathbf{q}_j, \alpha_j) = 0 \quad \forall \bar{\pi} \in \bar{\mathcal{P}}$.
- (S4) Solve the structure: find $z_j \in \mathcal{Z}$: $\mathbf{S}(\bar{z}, z_j, \pi_j) = \mathbf{s}(\bar{z}) \quad \forall \bar{z} \in \bar{\mathcal{Z}}$.

Note that this procedure avoids the computation of the shape derivative as well as the simultaneous treatment of the fluid and the structure.

It is mentionable that the subiteration process forms an instance of the generic iterative method for free-boundary problems: (1) the boundary-value problem(s) is

(are) solved with a subset of the free-boundary conditions imposed, and (2) the free-boundary position is adjusted to relax the remaining free-boundary condition. Different names are used for this iterative procedure, e.g., subiteration, successive approximation, or Picard iteration. For applications to fluid-structure-interaction problems see, for example, [1, 16, 17]. Moreover, we remark that the customary *partitioned* or *staggered* time-integration methods for fluid-structure-interaction problems (see [9, 20, 21, 22]) are essentially identical to the above subiteration method, with the provision that the process is not repeated.

The subiteration method can also be classified as a *defect correction* method [3]. The defect correction approach for a variational problem of the form (2.5a) commences with the determination of an initial estimate $u_0 \in \mathcal{U}$ such that $\tilde{P}(\bar{u}, u_0) = \mathbf{p}(\bar{u})$ for all $\bar{u} \in \bar{\mathcal{U}}$, where the functional $\tilde{P} : \bar{\mathcal{U}} \times \mathcal{U} \mapsto \mathbb{R}$ represents a suitable approximation to P . It then proceeds with finding $u_j \in \mathcal{U}$ such that

$$(3.2) \quad \tilde{P}(\bar{u}, u_j) = \mathbf{p}(\bar{u}) + \tilde{P}(\bar{u}, u_{j-1}) - P(\bar{u}, u_{j-1}) \quad \forall \bar{u} \in \bar{\mathcal{U}},$$

for $j = 1, 2, \dots$. If $\tilde{P}$ is sufficiently close to P , then u_j converges to the actual solution of the variational problem as $j \rightarrow \infty$. To derive the approximate functional for the subiteration method, we note that the approximations generated by (S1)–(S4) satisfy

$$(3.3) \quad E(\bar{\mathbf{q}}, \mathbf{q}_j, \alpha_j) + S(\bar{z}, z_j, \pi_j) + D(\bar{\pi}, \pi_j, \mathbf{q}_j, \alpha_j) + K(\bar{\alpha}, \alpha_j, z_{j-1}) = \mathbf{e}(\bar{\mathbf{q}}) + \mathbf{s}(\bar{z}) + \mathbf{k}(\bar{\alpha}),$$

for all admissible $(\bar{\mathbf{q}}, \bar{\alpha}, \bar{z}, \bar{\pi})$. In general, the functional K is associated with a variational projection, and is separable as $K(\bar{\alpha}, \alpha, z) = K_0(\bar{\alpha}, \alpha) + K_1(\bar{\alpha}, z)$. Upon introducing this additive partition of K into (3.3) and adding suitable partitions of zero, we immediately recover (3.2) with the approximate functional

$$(3.4) \quad \tilde{P}((\bar{\mathbf{q}}, \bar{\alpha}, \bar{z}, \bar{\pi}), (\mathbf{q}, \alpha, z, \pi)) := E(\bar{\mathbf{q}}, \mathbf{q}, \alpha) + S(\bar{z}, z, \pi) + K(\bar{\alpha}, \alpha, \zeta) + D(\bar{\pi}, \pi, \mathbf{q}, \alpha),$$

with ζ any fixed element of $\mathcal{Z}$. Note that the approximation only involves the functional K corresponding to the kinematic interface condition (2.3b). The connectivity table associated with the approximate operator is identical to Table 1, but with the right-upper entry nullified. Thus, the connectivity table is lower triangular, and the subproblems involving the inversion of $\tilde{P}$ can be solved conveniently by the forward substitution process inherent in the subiteration method.

4. Model problem. In this section we develop a linear model problem, which serves as a basis for the ensuing analysis of the properties of the subiteration method in sections 5 and 6.

4.1. Generating solution. To set up the linearized model, we consider the piston problem and we note that if the initial conditions are specified as $\mathbf{q}^0(x) = \mathbf{q}^0$ with $\mathbf{q}^0$ a constant in $\{\mathbf{q} \in \mathbb{R}^3 : 0 < q_1; p(\mathbf{q}) = p^0; q_2 = 0\}$ and $z^0 = \dot{z}^0 = 0$, then the obvious solution to (2.5) is

$$(4.1) \quad u := (\mathbf{q}, \alpha, z, \pi) = (\mathbf{q}^0, \alpha^0, 0, p^0) =: u^0.$$

Hence, if we provide the subiteration method with an initial approximation $z_0(t) = 0$, then the first iteration immediately yields the actual solution. If, instead, the initial approximation is specified as $z_0(t) = \epsilon z'_0(t)$, with z'_0 a suitable function independent of ϵ , then $u_1(\epsilon)$ approaches u^0 as $\epsilon \rightarrow 0$. In this context, u^0 is called a generating solution.

4.2. First-order conditions. The above motivates us to assume that the first approximation generated by the subiteration method can be expanded asymptotically:

$$(4.2) \quad \begin{aligned} \alpha_1(t) &= \alpha^0 + \sum_{l=1}^n \eta_l(\epsilon) \alpha^{(l)}(x, t) + o(\eta_n), & q_{1,j}(x, t) &= q_j^0 + \sum_{l=1}^n \psi_l^j(\epsilon) q_j^{(l)}(x, t) + o(\psi_n^j), \\ \pi_1(t) &= p^0 + \sum_{l=1}^n \xi_l(\epsilon) \pi^{(l)}(x, t) + o(\xi_n), & z_1(t) &= 0 + \sum_{l=1}^n \chi_l(\epsilon) z^{(l)}(x, t) + o(\chi_n), \end{aligned}$$

($j = 1, 2, 3$) for all $n = 1, 2, \dots$, with respect to certain asymptotic sequences $\{\psi_l^j(\epsilon)\}$, $\{\chi_l(\epsilon)\}$, $\{\eta_l(\epsilon)\}$, and $\{\xi_l(\epsilon)\}$, as $\epsilon \rightarrow 0$. For a definition of asymptotic sequences and the Landau symbols, o and O (used below), see, for instance, [15].

By the usual argument that terms of different orders in ϵ must vanish separately as $\epsilon \rightarrow 0$, it follows that terms with subscript l in (4.2) must be of $O(\epsilon^l)$, in order for (S1)–(S4) to provide meaningful relations between successive terms in the expansion. In particular, this implies that $u_1(\epsilon) = u^0 + \epsilon u^{(1)} + O(\epsilon^2)$, i.e., the error in the first-order approximation is $O(\epsilon^2)$.

For transparency of notation, in what follows the subscript 1, indicating the first iteration, is suppressed, and the superscript (1), referring to the infinitesimal perturbation, is replaced by a prime. The first iteration is suppressed and the superscript (1) referring to the first iteration is suppressed. Upon inserting (4.2) in (S1) and collecting terms of $O(\epsilon)$, we obtain the first-order condition $\alpha' = z'_0$. To simplify the analysis, we assume that the first-order perturbation in $\mathbf{q}$ is isentropic, i.e., the ratio p/q_1^γ is constant. Under this assumption, the energy equation becomes redundant and can be discarded. The first-order conditions corresponding to operation (S2) then translate into the initial-boundary-value problem

$$(4.3) \quad \frac{\partial}{\partial t} \begin{pmatrix} q'_1 \\ q'_2 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ c^2 & 0 \end{pmatrix} \cdot \frac{\partial}{\partial x} \begin{pmatrix} q'_1 \\ q'_2 \end{pmatrix} = 0,$$

with $c := \sqrt{\gamma p^0/q_1^0}$ the speed of sound, subject to the initial and boundary conditions

$$(4.4a) \quad q'_1(x, 0) = 0, \quad x \in (0, \alpha^0), \quad q'_2(x, 0) = 0, \quad x \in (0, \alpha^0),$$

$$(4.4b) \quad q'_2(0, t) = 0, \quad t \in (0, \tau), \quad q'_2(\alpha^0, t) = \rho^0 z'_0(t), \quad t \in (0, \tau).$$

The generic solution to (4.3) is

$$(4.5) \quad \begin{pmatrix} q'_1 \\ q'_2 \end{pmatrix}(x, t) = \rho^0 \left(f_+(t + x/c) \begin{pmatrix} 1 \\ -c \end{pmatrix} + f_-(t - x/c) \begin{pmatrix} 1 \\ c \end{pmatrix} \right).$$

The initial and boundary conditions (4.4) imply the following conditions on f_+ , f_- :

$$(4.6a) \quad f_-(-s) + f_+(s) = 0, \quad f_-(-s) - f_+(s) = 0, \quad s \in (0, \alpha^0/c),$$

$$(4.6b) \quad f_-(s) - f_+(s) = 0, \quad f_-(s - \alpha^0/c) - f_+(s + \alpha^0/c) = z'_0(s)/c, \quad s \in (0, \tau).$$

From (4.5)–(4.6) it follows straightforwardly that

$$(4.7) \quad q'_1(\alpha^0, t) = -\frac{\rho^0}{c} \left(z'_0(t) + 2 \sum_{j=1}^{\lfloor ct/2\alpha^0 \rfloor} z'_0(t - 2j\alpha^0/c) \right),$$

where $\lfloor \cdot \rfloor$ denotes the floor (truncation to integer) function. One can infer that the integer $\lfloor ct/2\alpha^0 \rfloor$ represents the number of reflections that have returned from the rigid boundary to the interface in the time interval $[0, t]$.

Under the assumption of isentropy, (S3) immediately yields the first-order perturbation in the aggregated force on the structure: $\pi' = c^2 q_1'(\alpha^0, t)$. From (4.7) it then follows that the structural displacement obtained in step (S4) of the subiteration algorithm complies with the first-order condition

$$(4.8) \quad M\ddot{z}'(t) + Kz'(t) = -c\rho^0 \left(\dot{z}'_0(t) + 2 \sum_{j=1}^{\lfloor ct/2\alpha^0 \rfloor} \dot{z}'_0(t - 2j\alpha^0/c) \right), \quad z'(0) = 0, \quad \dot{z}'(0) = 0.$$

Note that the initial-value problem (4.8) provides a unique relation between successive approximations of the structure displacement in the subiteration process.

To cast (4.8) into a convenient dimensionless form, we introduce characteristic time scales for the structure and the fluid, viz., $\tau_s := 2\pi\sqrt{M/K}$ and $\tau_f := 2\alpha^0/c$, respectively. Moreover, we define characteristic time-scale and mass ratios for the fluid-structure system, $\nu := 2\pi\tau_f/\tau_s$ and $\mu := \rho^0\alpha^0/M$, and a dimensionless time variable $s := 2\pi t/\tau_s$. With these definitions, we obtain

$$(4.9) \quad D^2 z'(s) + z'(s) = -\frac{2\mu}{\nu} \left(Dz'_0(s) + 2 \sum_{j=1}^{\lfloor s/\nu \rfloor} Dz'_0(s - j\nu) \right), \quad z'(0) = 0, \quad Dz'(0) = 0,$$

where D denotes differentiation with respect to the function argument.

4.3. Characteristic operators. The initial-value problem (4.9) associates a $z' \in \mathcal{Z}$ with each $z'_0 \in \mathcal{Z}$. Accordingly, (4.9) generates an automorphic operator on $\mathcal{Z}$. We denote this operator by $\mathbf{C}$. Because the subiteration process essentially reduces to recursion of the operator $\mathbf{C}$, we refer to $\mathbf{C}$ as the *characteristic operator* or *subiteration operator*.

An important simplification of the characteristic operator results from discarding the sum associated with the reflections. We designate the corresponding characteristic operator as the *null-reflection operator* $\mathbf{C}_0$, in particular,

$$(4.10) \quad \mathbf{C}_0 : \{z'_0 \in \mathcal{Z}\} \mapsto \{z' \in \mathcal{Z} : D^2 z' + z' = -2(\mu/\nu)Dz'_0, \quad z'(0) = 0, \quad Dz'(0) = 0\}.$$

By virtue of the linearity of the differential equation and the homogeneity of the initial conditions in (4.9), the following relation between the null-reflection operator and the general characteristic operator holds:

$$(4.11) \quad \mathbf{C} = \left(\mathbf{I} + 2 \sum_j^{\lfloor \tau/\nu \rfloor} \mathbf{H}_\nu^j \right) \mathbf{C}_0,$$

where the shift operator $\mathbf{H}_\nu := \mathbf{H}(\nu)$ on $\mathcal{Z}$ is defined as

$$(4.12) \quad (\mathbf{H}_\nu z)(s) := \begin{cases} 0 & 0 < s \leq \nu, \\ z(s - \nu) & \nu < s < \tau. \end{cases}$$

Note the group property $H(\nu + \nu') = H(\nu)H(\nu')$ for any $\nu, \nu' \in \mathcal{T} =: (0, \tau)$ such that $\nu + \nu' \in \mathcal{T}$.

A second simplification pertains to the limit $\nu \rightarrow 0$. The right member of (4.9) can then be identified as a Riemann sum and it holds that

$$(4.13) \quad Du(s) + 2 \sum_{j=1}^{\lfloor s/\nu \rfloor} Du(s - j\nu) = \frac{2}{\nu} \left(\int_0^s Du(\xi) d\xi + O(\nu) \right) \sim \frac{2}{\nu} (u(s) - u(0)) \quad \text{as } \nu \rightarrow 0.$$

Hence, as $\nu \rightarrow 0$ the characteristic operator approaches the operator C_∞ ,

$$(4.14) \quad C_\infty : \{z'_0 \in \mathcal{Z}\} \mapsto \{z' \in \mathcal{Z} : D^2 z' + z' = -4\mu/\nu^2 (z'_0(s) - z'_0(0)), z'(0) = 0, Dz'(0) = 0\}.$$

We refer to C_∞ according to (4.14) as the *quasi-steady operator*. Notice the different proportionalities $C_0 \propto \mu/\nu$ and $C_\infty \propto \mu/\nu^2$.

5. Nonnormality of the null-reflection operator. In this section we consider the nonnormality of the null-reflection operator. Such nonnormal behavior has important implications for the norm of powers of the operator and, hence, it is essential to the convergence properties of the subiteration method. The extension to the general subiteration operator including reflections is presented in section 6.

5.1. Theory of linear operators. This section reviews the basic theory of linear operators required for the ensuing exposition. For details we refer to the commendable [14].

Consider a Hilbert space $\mathcal{Z}$, equipped with an inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathcal{Z}}$ and an associated norm $\|\cdot\| := \|\cdot\|_{\mathcal{Z}}$, and a bounded linear operator $L : \mathcal{Z} \mapsto \mathcal{Z}$. The adjoint operator L^* is defined by the identity

$$(5.1) \quad \langle L^* w, z \rangle = \langle w, Lz \rangle \quad \forall w, z \in \mathcal{Z}.$$

In fact, it suffices that the identity holds for all w, z in a dense subspace of $\mathcal{Z}$; see [28, p. 263]. The operator L is said to be *normal* if it commutes with its adjoint, i.e., if

$$(5.2) \quad L^* L = L L^*.$$

To present the specific properties pertaining to (5.2), we require some elementary concepts. The resolvent is defined as the operator-valued function

$$(5.3) \quad R(\zeta) := R(L, \zeta) := (L - \zeta I)^{-1}$$

on $\mathbb{C}$. Herein, I is the identity on $\mathcal{Z}$. The resolvent set $\mathcal{R}(L)$ is defined as the subset of $\mathbb{C}$ consisting of all ζ for which $R(\zeta)$ is existent and bounded. The spectrum $\mathcal{S}(L)$ is defined as its complement in $\mathbb{C}$. If L is compact, which includes the case that $\dim \mathcal{Z} < \infty$, then the spectrum consists of at most a countable number of eigenvalues with finite multiplicities, possibly excepting zero if L is singular. The norm of the operator and its spectral radius are defined by

$$(5.4) \quad \|L\| := \sup_{0 \neq z \in \mathcal{Z}} \frac{\|Lz\|}{\|z\|} = \sup_{\substack{0 \neq v \in \mathcal{Z} \\ 0 \neq z \in \mathcal{Z}}} \frac{|\langle v, Lz \rangle|}{\|v\| \|z\|}, \quad \text{spr } L := \lim_{n \rightarrow \infty} \|L^n\|^{1/n} = \inf_{n \in \mathbb{N}} \|L^n\|^{1/n},$$

respectively. It holds that

$$(5.5) \quad \text{spr } \mathbf{L} = \sup_{\zeta \in \mathcal{S}(\mathbf{L})} |\zeta|.$$

A characteristic property of a normal operator is that $\|\mathbf{L}^n\| = \|\mathbf{L}\|^n$ and, accordingly, $\|\mathbf{L}\| = \text{spr } \mathbf{L}$; see [14]. Hence, for a normal operator the spectral radius $\text{spr } \mathbf{L}$ delineates the norm of monomials of the operator $\|\mathbf{L}^n\|$. In particular, if $\text{spr } \mathbf{L} < 1$, then $\|\mathbf{L}^n\|$ converges monotonically to 0 as n increases. For nonnormal operators the spectral radius only represents the asymptotic convergence as $n \rightarrow \infty$. If $\|\mathbf{L}\| > 1 > \text{spr } \mathbf{L}$, then the convergence of $\|\mathbf{L}^n\|$ is generally nonmonotonous and divergence can occur before asymptotic convergence sets in; cf. section 5.4 and, e.g., [12, p. 28]. Relevant in this respect are also the plots of $\|\exp(t\mathbf{L})\|$ (which has similar properties as $\|\mathbf{L}^n\|$) versus t for various nonnormal operators in [25].

Another important property of a normal operator is that it admits a spectral representation on a basis of orthonormal eigenvectors. If $\mathbf{L}$ is compact, we can index the nonzero eigenvalues as $\sigma_1, \sigma_2, \dots$. To each eigenvalue σ_j corresponds an eigenprojection

$$(5.6) \quad \mathbf{P}_j := -\frac{1}{2\pi i} \oint_{\chi_j} \mathbf{R}(\zeta) d\zeta$$

with $i := \sqrt{-1}$ and χ_j any closed curve in $\mathbb{C}$ of which the intersection of its interior and the spectrum contains σ_j only; see [10, p. 5 ff.] or [14, p. 39 ff.]. For compact normal operators $\mathbf{L}$ we have the spectral representation

$$(5.7) \quad \mathbf{L} = \sum_j \sigma_j \mathbf{P}_j$$

and $\mathbf{P}_j^* = \mathbf{P}_j$, $\dim \mathbf{P}_j < \infty$ and $\mathbf{P}_i \mathbf{P}_j = \delta_{ij} \mathbf{P}_i$. Furthermore, together with the orthonormal projection $\mathbf{P}_0$ on the kernel of $\mathbf{L}$ the eigenprojections form a complete orthonormal family and, hence, $\sum \mathbf{P}_j = \mathbf{I}$. Alternatively, we can select orthonormal bases $\{\varphi_j^i\}$ ($i = 1, \dots, \dim \mathbf{P}_j \mathcal{Z}$) for the subspaces $\mathbf{P}_j \mathcal{Z}$ and corresponding conjugate bases $\{\psi_j^i\}$ and rewrite (5.7) as

$$(5.8) \quad \mathbf{L}(\cdot) = \sum_j \sum_{i=1}^{\dim \mathbf{P}_j \mathcal{Z}} \sigma_j \varphi_j^i \langle \psi_j^i, \cdot \rangle.$$

The φ_j^i are the eigenvectors corresponding to the eigenvalue σ_j . The corresponding conjugate vectors $\{\psi_j^i\}$ are defined by $\langle \varphi_j^i, \psi_l^k \rangle = \delta_{jl}^{ik}$, i.e., $\langle \psi_j^i, \cdot \rangle$ represent the characteristic coordinates with respect to the basis $\{\varphi_j^i\}$. An operator which admits a representation conforming to (5.7) with $\mathbf{P}_j \mathbf{P}_k = \delta_{jk} \mathbf{P}_j$ and $\sum \mathbf{P}_j = \mathbf{I}$ is said to be *diagonalizable* or *semisimple*. If, in addition, it holds that the eigenprojections are orthogonal and, accordingly, $\mathbf{P}_j^* = \mathbf{P}_j$, then the normalized eigenvectors form an orthonormal basis of $\mathcal{Z}$ and the operator is said to be *unitarily diagonalizable*. Hence, a compact operator is normal if and only if it is unitarily diagonalizable.

To express the degeneracy of the set of eigenvectors of a nonnormal operator, we define the coercivity coefficient and the continuity coefficient of the projection onto the basis of eigenvectors $\{\varphi_j\} := \{\varphi_j^i\}$, respectively, as the largest number $\gamma(\mathbf{L})$ and the smallest number $\gamma'(\mathbf{L})$ such that

$$(5.9) \quad \inf_{0 \neq \{\hat{z}_j\} \in \hat{\mathcal{Z}}} \sup_{0 \neq v \in \mathcal{Z}} \frac{|\Sigma_j \langle v, \varphi_j \rangle \hat{z}_j|}{\|v\| \|\hat{z}\|} \geq \gamma(\mathbf{L}), \quad \sup_{\substack{0 \neq \{\hat{z}_j\} \in \hat{\mathcal{Z}} \\ 0 \neq v \in \mathcal{Z}}} \frac{|\Sigma_j \langle v, \varphi_j \rangle \hat{z}_j|}{\|v\| \|\hat{z}\|} \leq \gamma'(\mathbf{L}),$$

where $\|\hat{z}\|^2 := \sum \hat{z}_i^2$ and $\hat{\mathcal{Z}}$ represents the space of coefficients $\{\hat{z}_i\}$ such that $\sum \hat{z}_i \varphi_i$ resides in $\mathcal{Z}$. From the spectral representation (5.8) it follows that

$$(5.10) \quad \langle v, \mathbf{L}^n z \rangle = \langle v, \varphi_{i_1} \rangle \sigma_{i_1} \langle \psi_{i_1}, \varphi_{i_2} \rangle \sigma_{i_2} \cdots \langle \psi_{i_{n-1}}, \varphi_{i_n} \rangle \sigma_{i_n} \langle \psi_{i_n}, z \rangle = \sigma_i^n \langle v, \varphi_i \rangle \langle \psi_i, z \rangle,$$

where summation over repeated indices is implied. Furthermore, denoting by $\check{\mathcal{Z}}$ the space of coefficients $\{\check{z}_i\}$ such that $\sum \check{z}_i \psi_i \in \mathcal{Z}$, it holds that

$$(5.11) \quad \sup_{0 \neq z \in \mathcal{Z}} \frac{\|\langle z, \psi \cdot \rangle\|}{\|z\|} = \sup_{\substack{0 \neq \{\check{z}_i\} \in \check{\mathcal{Z}} \\ 0 \neq \{\check{w}_j\} \in \check{\mathcal{Z}}}} \frac{\|\hat{z}_i \langle \varphi_i, \psi_j \rangle \check{w}_j\|}{\|\check{w} \cdot\| \|\hat{z}_i \varphi_i\|} = \sup_{0 \neq \{\check{z}_i\} \in \check{\mathcal{Z}}} \frac{\|\hat{z} \cdot\|}{\|\hat{z}_i \varphi_i\|} \\ = \left(\inf_{0 \neq \{\check{z}_i\} \in \check{\mathcal{Z}}} \frac{\|\hat{z}_i \varphi_i\|}{\|\hat{z} \cdot\|} \right)^{-1} \leq \frac{1}{\gamma(\mathbf{L})},$$

again with summation over repeated indices. Hence, we obtain

$$(5.12) \quad \|\mathbf{L}^n\| = \sup_{\substack{0 \neq z \in \mathcal{Z} \\ 0 \neq v \in \mathcal{Z}}} \frac{|\langle v, \mathbf{L}^n z \rangle|}{\|v\| \|z\|} \leq (\text{spr } \mathbf{L})^n \sup_{\substack{0 \neq v \in \mathcal{Z} \\ 0 \neq z \in \mathcal{Z}}} \frac{\|\langle v, \varphi \cdot \rangle\|}{\|v\|} \frac{\|\langle z, \psi \cdot \rangle\|}{\|z\|} \leq (\text{spr } \mathbf{L})^n \frac{\gamma'(\mathbf{L})}{\gamma(\mathbf{L})}.$$

The ratio $\gamma'(\mathbf{L})/\gamma(\mathbf{L}) =: \kappa(\mathbf{L})$ is called the condition number of the basis of eigenvectors. By (5.4) and (5.12), it holds that $(\text{spr } \mathbf{L})^n \leq \|\mathbf{L}^n\| \leq \kappa(\mathbf{L})(\text{spr } \mathbf{L})^n$. If $\mathbf{L}$ is normal and, accordingly, $\{\varphi_j\}$ forms an orthonormal basis of $\mathcal{Z}$, then $\gamma(\mathbf{L}) = \gamma'(\mathbf{L}) = 1$ by Parseval's identity and, hence, $\kappa(\mathbf{L}) = 1$. Contrastingly, if $\mathbf{L}$ is nonnormal, then $\kappa(\mathbf{L})$ is typically much larger than unity. Moreover, it is mentionable that (5.11) implies that the norm of the characteristic coordinates of any $z \in \mathcal{Z}$ are bounded by $\|\hat{z} \cdot\| \leq \|z\|/\gamma(\mathbf{L})$. This implies that the characteristic coordinates of a typical element of $\mathcal{Z}$ can be enormous if the eigenfunctions are nearly linearly dependent and, accordingly, $\gamma(\mathbf{L})$ is small. Trefethen [25] notes that the properties of an analytic function of the operator, such as $\mathbf{L}^n$ or $\exp(t\mathbf{L})$, are then determined by the interactions of the eigenfunctions rather than their individual behavior.

Finally, we consider the relation between the normality of an operator and its pseudospectra, viz., the one-parameter family of subsets of $\mathbb{C}$ defined by

$$(5.13) \quad \mathcal{S}_\epsilon(\mathbf{L}) := \{\zeta \in \mathbb{C} : \|\mathbf{R}(\mathbf{L}, \zeta)\| \geq \epsilon^{-1}\}$$

for all $\epsilon \in \mathbb{R}_+$. By convention, $\mathcal{S}_0(\mathbf{L})$ is identified with the spectrum $\mathcal{S}(\mathbf{L})$. More precisely, the spectrum consists of all complex numbers ζ for which the resolvent $\mathbf{R}(\mathbf{L}, \zeta)$ is unbounded or its domain is only a proper subset of $\mathcal{Z}$; see also [14]. For any $\zeta \in \mathbb{C}$ it holds that $\mathcal{S}(\mathbf{L} - \zeta \mathbf{I}) = \mathcal{S}(\mathbf{L}) - \zeta$. Hence, by the spectral mapping theorem, $\mathcal{S}(\mathbf{R}(\zeta_0))$ is related to $\mathcal{S}(\mathbf{L})$ by the transformation $\zeta \mapsto (\zeta - \zeta_0)^{-1}$. On account of (5.5) it then follows that

$$(5.14) \quad \text{spr } \mathbf{R}(\zeta_0) = \sup_{\zeta \in \mathcal{S}(\mathbf{R}(\zeta_0))} |\zeta| = \sup_{\zeta \in \mathcal{S}(\mathbf{L})} |\zeta - \zeta_0|^{-1} = \left(\inf_{\zeta \in \mathcal{S}(\mathbf{L})} |\zeta - \zeta_0| \right)^{-1} = \text{dist}(\zeta_0, \mathcal{S}(\mathbf{L}))^{-1},$$

with $\text{dist}(\zeta_0, \mathcal{S}(\mathbf{L})) := \inf_{\zeta \in \mathcal{S}(\mathbf{L})} |\zeta_0 - \zeta|$ the distance of ζ_0 to the spectrum $\mathcal{S}(\mathbf{L})$. As the resolvent of a normal operator is itself a normal operator (see [14]), it holds that $\|\mathbf{R}(\mathbf{L}, \zeta)\| = \text{spr } \mathbf{R}(\mathbf{L}, \zeta)$ if $\mathbf{L}$ is normal. This implies that the ϵ -pseudospectrum of a normal operator is the aggregate of all points in $\mathbb{C}$ at distance at most ϵ from the spectrum. For nonnormal operators, the pseudospectra can be much larger; cf. [25] for examples; see also section 5.4.

5.2. Functional setting. Considering an open interval $\mathcal{T} := (0, \tau)$, an appropriate setting for C_0 is the Sobolev space $H^1(\mathcal{T})$ defined as

$$(5.15) \quad H^1(\mathcal{T}) := \{z \in L^2(\mathcal{T}) : \|z\|_{H^1(\mathcal{T})} < \infty\},$$

where the norm $\|\cdot\|_{H^1(\mathcal{T})}$ is induced by the inner product

$$(5.16) \quad \langle w, z \rangle_{H^1(\mathcal{T})} := \langle w, z \rangle_{L^2(\mathcal{T})} + \langle Dw, Dz \rangle_{L^2(\mathcal{T})}.$$

Whenever the interval is evident from the context, we will suppress it for brevity. The operator C_0 is bounded on H^1 as $\|C_0\|_{H^1} \leq 2\tau$ and unbounded on L^2 ; see Appendix B. Equipped with the norm induced by (5.16), H^1 is a Hilbert space.

As a digression, we note a corollary of the upper bound $\|C_0\|_{H^1} \leq 2\tau$ for the accuracy of the partitioned time-integration methods for fluid-structure-interaction problems (see, e.g., [9, 20, 21, 22]), and for the required number of subiterations for monolithic methods, i.e., methods that actually resolve the aggregated fluid and structure equations. Consider a time interval $\mathcal{T} := (0, T)$, partitioned into contiguous subintervals $\mathcal{T}_i$ of length τ ($i = 1, 2, \dots, T/\tau$). Let z, z_i , and z_i^n denote the actual boundary displacement, its finite-element approximation of polynomial degree m on $\mathcal{T}_i$, and the approximation to z_i generated by n subiterations, respectively. Indicating by the binary relation $\lesssim$ that the left member is at most equal to the right member modulo a positive constant, basic interpolation theory conveys that for sufficiently regular z the interpolation error $d_i := z - z_i$ obeys $\|d_i\|_{H^1(\mathcal{T}_i)} \lesssim \tau^m |z|_{H^m(\mathcal{T}_i)} \lesssim \tau^{m+1/2}$, with $|\cdot|_{H^m}$ the seminorm; see, e.g., [18, Theorem 6.8]. Moreover, the approximation z_{i-1} can then be extended to $\mathcal{T}_i$ such that $\|z_{i-1} - z\|_{H^1(\mathcal{T}_i)} \lesssim \tau^{m+1/2}$. Therefore, if the initial approximation on the interval is generated by extrapolation, i.e., $z_i^0 = z_{i-1}$, then the corresponding evaluation error $e_i^n := z_i^n - z_i$ obeys $\|e_i^0\|_{H^1(\mathcal{T}_i)} \lesssim \tau^{m+1/2}$. The upper bound $\|C_0\|_{H^1} \leq 2\tau$ indicates that $\|e_i^n\|_{H^1} \leq 2\tau \|e_i^{n-1}\|_{H^1}$. This implies that $\|e_i^n\|_{H^1} = O(\tau^{m+3/2}) \ll \|d_i\|_{H^1} = O(\tau^{m+1/2})$ as $\tau \rightarrow 0$. That is, for sufficiently small τ , a single subiteration renders the local evaluation error much smaller than the local interpolation error. However, to ensure that the global evaluation error is smaller than the global interpolation error, two iterations are required. To clarify this, we note that the pressure perturbation is related to the displacement perturbation by $\pi' = \text{const } Dz'$; cf. section 4.2. In the subiteration process the pressure approximation is based on the previous displacement approximation and, hence, the evaluation error in the pressure $\beta_i^n := \pi_i^n - \pi_i$ complies with $\|\beta_i^n\|_{L^2(\mathcal{T}_i)} \lesssim \|e_i^{n-1}\|_{H^1} \lesssim \tau^{m+n-1/2}$. Summation over the intervals imparts that the global evaluation error in the pressure satisfies $\|\beta^n\|_{L^2}^2 = \sum_i \|\beta_i^n\|_{L^2(\mathcal{T}_i)}^2 \lesssim \tau^{2(m+n-1)}$. The global evaluation error in the displacement complies with $D^2 e^n + e^n = \beta^n$ and the homogeneous initial conditions $e^n(0) = D e^n(0) = 0$. A similar derivation as in Appendix B then conveys that $\|e^n\|_{H^1(\mathcal{T})} \leq 2T \|\beta^n\|_{L^2(\mathcal{T})} \lesssim \tau^{m+n-1}$. One subiteration ($n = 1$) thus renders the global evaluation error in the displacement only $O(\tau^m)$ and, hence, it is of the same order as the global interpolation error. The global evaluation error in the displacement after two iterations ($n = 2$) is $O(\tau^{m+1})$ and for sufficiently small τ it is thus negligible compared to the global interpolation error. In conclusion, partitioned methods and monolithic methods possess identical orders of accuracy. However, the proportionality constants can be very different. Furthermore, for appropriately small time steps a monolithic method requires only two subiteration per time step for convergence to sufficient accuracy.

5.3. Nonnormality of C_0 . To establish the nonnormality of the null-reflection operator, we derive its adjoint and demonstrate that there exist admissible functions that violate the identity (5.2).

Consider an open interval $\mathcal{T} := (0, \tau)$. Let $\mathcal{Z} := H^1(\mathcal{T})$, equipped with the standard L^2 inner product. Being a closed linear subspace of the Hilbert space L^2 , the closure $\overline{\mathcal{Z}}$ of $\mathcal{Z}$ in L^2 is a Hilbert space. To avoid the conveyance of constants, we set $2\mu/\nu = 1$ in (4.10) without loss of generality. A useful device in the derivation of the adjoint operator is the translation operator $\mathsf{T} : z \in \mathcal{Z} \mapsto z - z(0) \in \mathcal{Z}$. As $C_0 z$ depends on z through Dz only, C_0 is translation invariant and $C_0 \mathsf{T} = C_0$. To determine the adjoint operator of C_0 on $\overline{\mathcal{Z}}$, we select an arbitrary $w \in \mathcal{Z}$ and associate a function ψ by

$$(5.17) \quad D^2\psi + \psi = w, \quad \psi(\tau) = 0, \quad D\psi(\tau) = 0.$$

Moreover, we consider an arbitrary $z \in \mathcal{Z}$ and denote by v its image under $C_0 \mathsf{T}$. The adjoint operator follows from the following chains of identities:

$$(5.18) \quad \begin{aligned} \langle w, C_0 z \rangle &= \langle w, C_0 \mathsf{T} z \rangle = \langle D^2\psi + \psi, v \rangle = v D\psi|_0^\tau - Dv\psi|_0^\tau + \langle \psi, D^2v + v \rangle \\ &= \langle \psi, D^2v + v \rangle = \langle -\psi, D \mathsf{T} z \rangle = -\mathsf{T} z \psi|_0^\tau + \langle D\psi, \mathsf{T} z \rangle = \langle D\psi, \mathsf{T} z \rangle, \end{aligned}$$

and

$$(5.19) \quad \langle C_0^* w, z \rangle = \langle w, C_0 z \rangle = \langle w, C_0 \mathsf{T} z \rangle = \langle C_0^* w, \mathsf{T} z \rangle.$$

The third and sixth identity in (5.18) follow by integration by parts. The boundary terms in (5.18) vanish in virtue of the homogeneous initial and final conditions on v and ψ , respectively, and the property $(\mathsf{T} z)(0) = 0$ of the translation operator. From (5.18)–(5.19) the adjoint operator is readily identified as

$$(5.20) \quad C_0^* : \{w \in \mathcal{Z}\} \mapsto \{D\psi \in \mathcal{Z} : D^2\psi + \psi = w, \psi(\tau) = 0, D\psi(\tau) = 0\}.$$

As w, z in (5.18)–(5.19) are arbitrary, we have the implication $\langle w, C_0 z \rangle = \langle C_0^* w, z \rangle$ for all $w, z \in \mathcal{Z}$, in accordance with (5.1). Noting that $\mathcal{Z}$ is dense in $\overline{\mathcal{Z}}$, we find that C_0^* according to (5.20) is the adjoint operator of C_0 on $\overline{\mathcal{Z}}$.

The normality of C_0 can now be disproved straightforwardly by establishing that there exists a $z \in \mathcal{Z}$ such that $C_0^* C_0 z \neq C_0 C_0^* z$. For instance, $\sin \in \mathcal{Z}$ and

$$(5.21) \quad \begin{aligned} (C_0^* C_0 - C_0 C_0^*) \sin(s) &= [2(s - \tau) \cos(s - 2\tau) - \sin(s + 2\tau) - 2 \sin(s - 2\tau) \\ &\quad - (4\tau s - 3 - 2\tau^2) \sin(s) - 2s \cos(s)]/16 \neq 0, \end{aligned}$$

which completes the proof.

5.4. Approximation on polynomials. To elaborate further the properties of the subiteration procedure, we consider the null-reflection characteristic operator on polynomials.

Let $\mathcal{Z}_m$ be the space of polynomials of degree m on the open interval $\mathcal{T} := (0, \tau)$ with bounded H^1 norm. By the Weierstrass approximation theorem, one can conceive of $\mathcal{Z}_m$ as an approximation to the continuous functions in $H^1(\mathcal{T})$. Equipped with the inner product (5.16), $\mathcal{Z}_m$ is a Hilbert space. For given m , we associate to any $z \in \mathcal{Z}_m$ a function $v \in \mathcal{Z}_m$ by the variational statement

$$(5.22a) \quad A(w, v) = B(w, z) \quad \forall w \in \mathcal{Z}_m,$$

where

$$(5.22b) \quad A(w, v) := \langle w, D^2 v + v \rangle_{L^2} + w Dv|_0 - v Dw|_0, \quad B(w, z) := -\langle w, Dz \rangle_{L^2}.$$

Equation (5.22) represents a weak formulation of $v = C_0 z$ with weakly imposed initial conditions. To make (5.22) amenable to numerical treatment, we introduce a basis $\{\theta_0, \dots, \theta_m\}$ of $\mathcal{Z}_m$, where θ_i is the normalized Legendre polynomial of degree i on the interval $\mathcal{T}$. Equation (5.22) then translates into the linear algebra problem $\mathbf{A} \cdot \bar{\mathbf{v}} = \mathbf{B} \cdot \bar{\mathbf{z}}$, where $A_{ij} = A(\theta_i, \theta_j)$ and $B_{ij} = B(\theta_i, \theta_j)$, and the vectors $\bar{\mathbf{v}}$ and $\bar{\mathbf{z}}$ collect the coefficients of v and z with respect to the basis of Legendre polynomials, respectively. Hence, with respect to this basis the mapping C_0 is represented by the matrix $\mathbf{C}_0 := \mathbf{A}^{-1} \cdot \mathbf{B}$.

To express the H^1 norm of C_0 in terms of $\mathbf{C}_0$, we note that for any two functions $v, z \in \mathcal{Z}_m$ the H^1 inner product can be expressed as

$$(5.23) \quad \langle v, z \rangle_{H^1} = \sum_{i,j} \bar{v}_i W_{ij} \bar{z}_j \quad \text{with} \quad W_{ij} = \langle \theta_i, \theta_j \rangle_{H^1} = \langle \theta_i, \theta_j \rangle_{L^2} + \langle D\theta_i, D\theta_j \rangle_{L^2}.$$

The Gram matrix $\mathbf{W}$ is symmetric positive definite and, hence, it admits a Cholesky factorization, i.e., there exists an upper-triangular matrix $\mathbf{Q}$ such that $\mathbf{W} = \mathbf{Q}^T \cdot \mathbf{Q}$. Accordingly, we can express the H^1 norm of any $z \in \mathcal{Z}_m$ as

$$(5.24) \quad \|z\|_{H^1}^2 = \langle z, z \rangle_{H^1} = \sum_{i,j} \bar{z}_i \langle \theta_i, \theta_j \rangle_{H^1} \bar{z}_j = \bar{\mathbf{z}}^T \cdot \mathbf{Q}^T \cdot \mathbf{Q} \cdot \bar{\mathbf{z}} = \|\mathbf{Q} \cdot \bar{\mathbf{z}}\|^2.$$

In (5.24), the vector norm $\|\cdot\| : \mathbb{R}^m \mapsto \mathbb{R}$ connotes the standard L^2 -norm. The H^1 norm of C_0 on $\mathcal{Z}_m$ can thus be determined as follows:

$$(5.25) \quad \begin{aligned} \|C_0\|_{H^1} &= \sup_{\substack{0 \neq v \in \mathcal{Z}_m \\ 0 \neq z \in \mathcal{Z}_m}} \frac{|\langle v, C_0 z \rangle_{H^1}|}{\|v\|_{H^1} \|z\|_{H^1}} = \sup_{\substack{0 \neq \bar{\mathbf{v}} \in \mathbb{R}^m \\ 0 \neq \bar{\mathbf{z}} \in \mathbb{R}^m}} \frac{|\bar{\mathbf{v}}^T \cdot \mathbf{W} \cdot \mathbf{C}_0 \cdot \bar{\mathbf{z}}|}{\|\mathbf{Q} \cdot \bar{\mathbf{v}}\| \|\mathbf{Q} \cdot \bar{\mathbf{z}}\|} \\ &= \sup_{\substack{0 \neq \hat{\mathbf{v}} \in \mathbb{R}^m \\ 0 \neq \hat{\mathbf{z}} \in \mathbb{R}^m}} \frac{|\hat{\mathbf{v}}^T \cdot \mathbf{Q} \cdot \mathbf{C}_0 \cdot \mathbf{Q}^{-1} \cdot \hat{\mathbf{z}}|}{\|\hat{\mathbf{v}}\| \|\hat{\mathbf{z}}\|} = \|\mathbf{Q} \cdot \mathbf{C}_0 \cdot \mathbf{Q}^{-1}\|, \end{aligned}$$

where $\hat{\cdot} := \mathbf{Q} \cdot \bar{\cdot}$. Note that the final form in (5.25) is computable. Table 2 lists the norm $\|C_0\|_{H^1(\mathcal{T})}$ for representative values of the interval length τ and the polynomial degree m . A discussion of the results is postponed to the end of this section.

The question pertaining to the relation between the spectrum of C_0 on a function space such as H^1 , and the spectrum of C_0 on a finite-dimensional subspace such as $\mathcal{Z}_m$, is a profoundly difficult one, and we shall not pursue it here. Let us just mention that the spectrum of any bounded operator is nonempty [14, p. 176] and, in particular, the spectrum of C_0 on H^1 is nonempty. Below, we consider the spectrum of C_0 on $\mathcal{Z}_m$ independently. To relate the pseudospectra of C_0 on $\mathcal{Z}_m$ to the representation matrix $\mathbf{C}_0$, we state the following identities:

$$(5.26) \quad \| (C_0 - \zeta \mathbf{I})^{-1} \|_{H^1} = \|\mathbf{Q} \cdot (C_0 - \zeta \mathbf{I})^{-1} \cdot \mathbf{Q}^{-1}\| = \|(\mathbf{Q} \cdot \mathbf{C}_0 \cdot \mathbf{Q}^{-1} - \zeta \mathbf{I})^{-1}\|.$$

The second expression implies that the spectrum of C_0 on $\mathcal{Z}_m$ coincides with the spectrum of $\mathbf{C}_0$. The final expression demonstrates that the H^1 pseudospectra of C_0 on $\mathcal{Z}_m$ coincide with the pseudospectra of $\mathbf{Q} \cdot \mathbf{C}_0 \cdot \mathbf{Q}^{-1}$. This reflects that the spectrum does not depend on the selected norm (H^1 or L^2), but the pseudospectra do. Figure 1

TABLE 2
 H^1 norm and spectral radius of the null-reflection operator C_0 with $2\mu/\nu = 1$ on Z_m .

	$m \setminus \tau$	10^{-2}	10^{-1}	10^0	10^1	10^2
$\ C_0\ _{H^1}$	2^6	6.366×10^{-3}	6.363×10^{-2}	6.029×10^{-1}	3.684×10^0	3.235×10^1
	2^7	6.366×10^{-3}	6.363×10^{-2}	6.029×10^{-1}	3.684×10^0	3.235×10^1
	2^8	6.366×10^{-3}	6.363×10^{-2}	6.029×10^{-1}	3.684×10^0	3.235×10^1
$\text{spr } C_0$	2^6	3.947×10^{-4}	3.947×10^{-3}	3.941×10^{-2}	3.461×10^{-1}	3.872×10^0
	2^7	4.161×10^{-4}	4.161×10^{-3}	4.154×10^{-2}	3.600×10^{-1}	2.037×10^0
	2^8	4.314×10^{-4}	4.314×10^{-3}	4.307×10^{-2}	3.697×10^{-1}	2.099×10^0

FIG. 1. Spectra (●) and H^1 ϵ -pseudospectra of C_0 on Z_m for $m = 2^7$, $\tau = 1$ (left), $\tau = 10^2$ (right), $\epsilon = 10^{-13}$ (—), 10^{-11} (—), 10^{-9} (···), 10^{-7} (— · —), 10^{-5} (—), 10^{-3} (— · —), 10^{-1} (···).

displays the H^1 pseudospectra¹ of C_0 for $m = 2^7$, and $\tau = 1$ and $\tau = 10^2$. Moreover, Table 2 lists the spectral radius of C_0 for representative values of τ and m .

The coercivity coefficient $\gamma(C_0)$ can be determined as follows:

$$\begin{aligned}
 & \inf_{0 \neq \{z_i\} \in \hat{Z}_m} \sup_{0 \neq v \in Z_m} \frac{|\Sigma_i \langle v, \varphi_i \rangle_{H^1} \hat{z}_i|}{\|v\|_{H^1} \|\hat{z}\|} = \inf_{0 \neq \{z_i\} \in \hat{Z}_m} \sup_{0 \neq \bar{v} \in \mathbb{R}^m} \frac{|\Sigma_i \bar{v}^T \cdot \mathbf{W} \cdot \bar{\varphi}_j \| \mathbf{Q} \cdot \bar{\varphi}_i \|^{-1} \hat{z}_i|}{\| \mathbf{Q} \cdot \bar{v} \| \| \hat{z} \|} \\
 (5.27) \quad & = \inf_{0 \neq \hat{z} \in \mathbb{R}^m} \sup_{0 \neq \mathbf{w} \in \mathbb{R}^m} \frac{|\mathbf{w}^T \cdot \mathbf{U} \cdot \hat{z}|}{\| \mathbf{w} \| \| \hat{z} \|} = \inf_{0 \neq \mathbf{z} \in \mathbb{R}^m} \sup_{0 \neq \mathbf{u} \in \mathbb{R}^m} \frac{|\mathbf{u}^T \cdot \mathbf{F}^T \cdot \mathbf{U} \cdot \mathbf{G} \cdot \mathbf{z}|}{\| \mathbf{u} \| \| \mathbf{z} \|} = \rho_{\min}.
 \end{aligned}$$

with $\mathbf{w} := \mathbf{Q} \cdot \bar{v}$, $\mathbf{U} := (\mathbf{Q} \cdot \bar{\varphi}_1 / \|\mathbf{Q} \cdot \bar{\varphi}_1\|, \mathbf{Q} \cdot \bar{\varphi}_2 / \|\mathbf{Q} \cdot \bar{\varphi}_2\|, \dots)$ and $\rho_{\min}$ the smallest singular value of $\mathbf{U}$. One simply infers that $\mathbf{U}$ contains the normalized eigenvectors of $\mathbf{Q} \cdot C_0 \cdot \mathbf{Q}^{-1}$. The matrices $\mathbf{F}$, $\mathbf{R}$, and $\mathbf{G}$ pertain to the singular-value decomposition of $\mathbf{U}$ according to $\mathbf{U} = \mathbf{F} \cdot \mathbf{R} \cdot \mathbf{G}^T$; see, e.g., [11, 12]. As the matrices $\mathbf{F}$ and $\mathbf{G}$ are unitary, the transformations $\mathbf{w} = \mathbf{F} \cdot \mathbf{u}$ and $\hat{z} = \mathbf{G} \cdot \mathbf{z}$ are norm preserving. This leads to the third identity. Similarly, we obtain for the continuity coefficient $\gamma'(C_0) = \rho_{\max}$. For the parameters in Table 2, the coercivity and continuity coefficients thus computed yield condition numbers $\kappa(C_0) > 10^{10}$.

¹We thankfully acknowledge the use of the *EigTool* package by T.G. Wright, M. Embree, and L.N. Trefethen for producing the pseudospectra plots in Figure 1; see [27] for further information.

The computed norms in Table 2 are virtually independent of the polynomial degree m . This is a result of the smoothness of the function that induces the norm and of its image, i.e., the functions $\bar{v}$ and $\bar{z}$ defined by

$$(5.28) \quad \{\bar{v}, \bar{z}\} := \arg \sup_{\substack{0 \neq v \in H^1 \\ 0 \neq z \in H^1}} \frac{|\langle v, C_0 z \rangle_{H^1}|}{\|v\|_{H^1} \|z\|_{H^1}}$$

are sufficiently smooth to render the error in their approximation in $\mathcal{Z}_m$ with $m \in \{2^6, 2^7, 2^8\}$ negligible. The erratic behavior of the computed spectral radius conjecturally emanates from the nonnormality of C_0 . Figure 1 illustrates that the ϵ -pseudospectra of C_0 are much larger than the union of ϵ -discs centered at the spectrum, in accordance with section 5.1. As $\mathcal{S}_\epsilon(C_0)$ is the aggregate of the spectra of all perturbed operators $C_0 + C'_0$ with $\|C'_0\| \leq \epsilon$ (see [14, p. 208] or [10, p. 4]), i.e.,

$$(5.29) \quad \mathcal{S}_\epsilon(C_0) = \{\zeta \in \mathbb{C} : \zeta \in \mathcal{S}(C_0 + C'_0), \|C'_0\| \leq \epsilon\}$$

(in the appropriate norm) this implies that the spectrum is sensitive to perturbations of the operator. An alternative measure for this sensitivity is the condition of the eigenvalues; see, e.g., [11, p. 323]. For the results in Table 2, the condition of the largest eigenvalue exceeds 10^{10} .

The results in Table 2 indicate that $\text{spr } C_0 \propto \|C_0\|_{H^1} \propto \tau$ as $\tau \rightarrow 0$. To verify this hypothesis, consider C_0 on the interval $s \in (0, \tau)$ and define an auxiliary variable $\xi := s/\tau$. For any admissible z with image $v := C_0 z$ it holds that

$$(5.30) \quad \frac{d^2 v}{d\xi^2} \frac{1}{\tau^2} + v = -\frac{dz}{d\xi} \frac{1}{\tau}, \quad \xi \in (0, 1).$$

Because $D^2 v$ is the principal term in the operator, the term v can be ignored as $\tau \rightarrow 0$. Equation (5.30) is then equivalent to $D^2 v = -\tau D z$. Noting that τ appears as a multiplicative factor, the result is immediate. In general, for sufficiently small τ , the spectrum and pseudospectra of C_0 scale with τ . As τ increases, the principal term $D^2 v$ becomes a singular perturbation to v and the spectrum deforms; see also Figure 1. For sufficiently small τ , the proportionality of $\text{spr } C_0$ and $\|C_0\|$ to τ indicates that the convergence of the subiteration process deteriorates linearly with an increase of the time interval. Let us also note that the results in Table 2 comply with the upper bound $\|C_0\|_{H^1} \leq 2\tau$ derived in Appendix B. Moreover, the results indicate that the bound is reasonably sharp.

To illustrate the effect of nonnormality on the convergence of the subiteration method, the left and right plots in Figure 2, respectively, plot $\|C_0^n\|_{H^1}$ on the polynomials $\mathcal{Z}_m (= \|(\mathbf{Q} \cdot C_0 \cdot \mathbf{Q}^{-1})^n\|)$ with $m = 2^8$ versus the iteration counter n for $2\mu/\nu = 1$ and representative lengths of the time interval τ , and for $\tau = 1$ and representative values of the parameter $2\mu/\nu$. In all cases, the spectral radius is less than 1 and the subiteration method is formally stable. However, monotone convergence occurs only if $\|C_0\|_{H^1} \leq 1$. The figures illustrate that if $\|C_0\|$ is large, then the error can increase by many orders of magnitude before the asymptotic convergence sets in. Note that this behavior is in accordance with the indication provided by the condition numbers $\kappa(C_0)$. Furthermore, it is to be noted that the difference between the results in the right plot in Figure 2 emanates exclusively from a rescaling of the spectrum, as $2\mu/\nu$ appears as a multiplicative factor to the operator; cf. (4.9). The nonnormality of the underlying operators (in any suitable measure, e.g., $\kappa(C_0)$ or $\|C_0\|/\text{spr } C_0$) is

FIG. 2. Convergence of subiteration: $\|C_0^n\|_{H^1}$ on $\mathcal{Z}_m$ ($m = 2^8$) with (left) $2\mu/\nu = 1$ and $\tau = 10^0$ (+), $\tau = 10^{1/2}$ (o), $\tau = 10^1$ (□), $\tau = 10^{3/2}$ (Δ), and (right) $\tau = 1$ and $2\mu/\nu = 1$ (+), $2\mu/\nu = 5$ (o), $2\mu/\nu = 10$ (□), $2\mu/\nu = 20$ (Δ).

identical. This exemplifies that if the spectral radius is sufficiently small, then the effect of nonnormality on the convergence of the subiteration method is concealed.

By definition, the operator norm $\|C_0^n\|_{H^1}$ is an upper bound on the convergence behavior pertaining to a particular element of H^1 , i.e., $\|C_0^n\|_{H^1} \geq \|C_0^n z\|_{H^1} \|z\|_{H^1}^{-1}$ for all $z \in H^1$. The upper bound is sharp in the sense that there exists an element in H^1 for which equality holds. However, in practice, the initial error is typically confined to a proper subspace of H^1 . For instance, if the solution is sufficiently regular and the initial approximation on a time interval is obtained by a k -order extrapolation of the solution on the previous time interval, then the initial error e is confined to a regular subspace of H^1 with, moreover, $[D^i e](0) = 0$ for all $i = 0, 1, \dots, k-1$. The norm of C_0^n on this subspace can be much smaller, yielding $\|C_0^n\|_{H^1}$ practically unsharp. To illustrate the effect of confinement of the initial error due to extrapolation, Figure 3 plots $\|C_0^n z\|_{H^1} \|z\|_{H^1}^{-1}$ versus n for $\tau = 1$ and $2\mu/\nu = 10$, and $z(s) = s^k$, $k = 1, 2, 3, 4$. Note that s^k has $k-1$ vanishing derivatives at $s = 0$. Hence, we can conceive of s^k as a k -order-extrapolation error. Figure 3 indicates that the nonnormality-induced divergence is reduced with increasing k . Moreover, for large k , the discrepancy between $\|C_0^n\|_{H^1}$ and $\|C_0^n z\|_{H^1} \|z\|_{H^1}^{-1}$ can increase to several orders of magnitude.

6. Nonnormality of the subiteration operator. In this section we are concerned with the properties of the general characteristic operator, i.e., the operator including reflections.

6.1. Nonnormality of C . The normality of C can be disproved by deriving the associated adjoint operator and establishing that an admissible z exists that contradicts $C^* C z = C C^* z$. Let the space of admissible functions $\mathcal{Z}$ be defined as in section 5.3. Recalling the definition of the shift operator (4.12), from the sequence of identities

$$(6.1) \quad \langle H_\nu^* v, z \rangle_{L^2} = \langle v, H_\nu z \rangle_{L^2} = \int_\nu^\tau v(s) z(s - \nu) ds = \int_0^{\tau - \nu} v(s + \nu) z(s) ds$$

we readily identify the adjoint shift operator as

$$(6.2) \quad (H_\nu^* z)(s) := \begin{cases} 0 & \nu \leq s < \tau, \\ z(s + \nu) & 0 < s < \tau - \nu. \end{cases}$$

FIG. 3. Convergence of subiteration: $\|C_0^n z\|_{H^1}^{-1} \|z\|_{H^1}^{-1}$ on $\mathcal{Z}_m$ ($m = 2^8$) with $\tau = 1$, $2\mu/\nu = 10$ and $z(s) = s^k$, $k = 1$ ($\circ$), $k = 2$ ($\triangle$), $k = 3$ ($+$), and $k = 4$ ($*$). Norm $\|C_0^n\|_{H^1}$ indicated by ($\square$).

Moreover, on account of (4.11) it holds that

$$(6.3) \quad \langle v, Cz \rangle_{L^2} = \langle v, C_0 z + 2 \sum_j H_{j\nu} C_0 z \rangle_{L^2} = \langle C_0^* v + 2 \sum_j C_0^* H_{j\nu}^* v, z \rangle_{L^2}.$$

One can infer that H_ν^* commutes with C_0^* and, hence,

$$(6.4) \quad C^* = \left(I + 2 \sum_i^{\lfloor \tau/\nu \rfloor} H_{i\nu}^* \right) C_0^*.$$

The normality of C can now be disproved straightforwardly by establishing that there exists a $z \in \bar{\mathcal{Z}}$ such that $C^* C z \neq C C^* z$.

6.2. Approximation on polynomials. We consider the characteristic operator C on the polynomials $\mathcal{Z}_m$ on an interval $\mathcal{T} := (0, \tau)$ similarly as in section 5.4. The representation matrix $\mathbf{C}$ can be derived from the null-reflection matrix $\mathbf{C}_0$ and a shift matrix on the basis of (4.11). Recalling that θ_i designates the normalized Legendre polynomial of degree i , we define the representation matrix $\mathbf{H}_\nu$ of the shift operator H_ν according to (4.12) as

$$(6.5) \quad H_{ij}(\nu) := \langle \theta_i, H_\nu \theta_j \rangle_{L^2} = \int_\nu^\tau \theta_i(s) \theta_j(s - \nu) ds.$$

Equation (4.11) leads to

$$(6.6) \quad \mathbf{C} := \left(\mathbf{I} + 2 \sum_{j=1}^{\lfloor \tau/\nu \rfloor} \mathbf{H}^j(\nu) \right) \cdot \mathbf{C}_0.$$

By means of (6.6) we can establish the characteristics of C on $\mathcal{Z}_m$ through analogous expressions to (5.25)–(5.27).

The top and bottom plots in Figure 4, respectively, display the norm and the spectral radius of the subiteration operator C on $\mathcal{Z}_m$ with $m = 2^8$ versus ν^{-1} for $\tau = 10^{-2}, 10^{-1}, \dots, 10^2$. Furthermore, the properly rescaled norm and spectral radius of the quasi-steady operator are displayed, viz., $(2/\nu) \text{spr } C_\infty$ and $(2/\nu) \|C_\infty\|$ with $4\mu/\nu^2$ in (4.14) set to 1. For completeness, we mention that the representation matrix of

FIG. 4. Plots (log-log) of the H^1 norm (top) and the spectral radius (bottom) of $\mathbf{C}$ (—), $\mathbf{C}_0$ (····) and $\mathbf{C}_\infty$ (---, rescaled) on $\mathcal{Z}_m$ ($m = 2^8$) for $\tau = 10^{-2}, 10^{-1}, \dots, 10^2$ (bottom to top).

$\mathbf{C}_\infty$ with respect to $\{\theta_i\}$ is given by $\mathbf{C}_\infty := -\mathbf{A}^{-1} \cdot \mathbf{T}$ with $A_{ij} = A(\theta_i, \theta_j)$ according to (5.22b) and $T_{ij} := \langle \theta_i, \mathbf{T} \theta_j \rangle_{L^2} = \delta_{ij} - \langle \theta_i, \theta_j(0) \rangle_{L^2}$. Note that if $\nu > \tau$, then $\mathbf{C} = \mathbf{C}_0$ and $\text{spr } \mathbf{C}$ and $\|\mathbf{C}\|$ coincide with the corresponding entries in Table 2. The distance between the curves in Figure 4 indicates that $\text{spr } \mathbf{C} \propto \|\mathbf{C}\|_{H^1} \propto \tau^2$ for appropriately small ν and τ ($\nu < \tau$). Noting that for any $z \in \mathcal{Z}$ with image $v := \mathbf{C}_\infty z$ it holds that

$$(6.7) \quad \frac{d^2 v}{d\xi^2} \frac{1}{\tau^2} + v = -(z - z(0)), \quad \xi \in (0, 1),$$

by the same rationale as in section 5.4 we conclude that the norm and spectral radius of $\mathbf{C}_\infty$ are indeed proportional to τ^2 as $\tau \rightarrow 0$. In particular, this implies that for appropriately small ν and τ the convergence of the subiteration method degrades quadratically with increasing τ . Figure 4, moreover, indicates that for small τ the spectral radius of $\mathbf{C}$ is insensitive to a few reflections, but its norm is not: $\text{spr } \mathbf{C}$ remains approximately constant upto $\tau/\nu = 10$, but $\|\mathbf{C}\|$ increases immediately if τ/ν exceeds 1. For large τ , the spectral radius and the norm of $\mathbf{C}$ behave erratically in the transition from $\mathbf{C}_0$ to $\mathbf{C}_\infty$. We shall not further pursue the intricacies of this transition.

To demonstrate the effect of nonnormality on the convergence of the subiteration method in the presence of reflections, the top and bottom plots in Figure 5 plot $\|\mathbf{C}^n\|_{H^1}$ on $\mathcal{Z}_m$ ($= \|(\mathbf{Q} \cdot \mathbf{C} \cdot \mathbf{Q}^{-1})^n\|$) with $m = 2^8$ versus the iteration counter n for $\tau/\nu = 2^2, 2^3, \dots, 2^7$, and for $\tau = 1$ and $2\mu/\nu = 2$, and $\tau = 10^2$ and $2\mu/\nu = 0.2$, respectively. The values of $2\mu/\nu$ have been selected such that $\text{spr } \mathbf{C} < 1$ in all

FIG. 5. Convergence of subiteration: $\|C^n\|_{H^1}$ on $\mathcal{Z}_m$ ($m = 2^8$) with $\tau = 1$ and $2\mu/\nu = 2$ (top), $\tau = 10^2$ and $2\mu/\nu = 0.2$ (bottom), and $\tau/\nu < 1$ (+), $\tau/\nu = 2^2$ (o), $\tau/\nu = 2^3$ ($\square$), $\tau/\nu = 2^4$ ($\triangle$), $\tau/\nu = 2^5$ (*), $\tau/\nu = 2^6$ (∇), and $\tau/\nu = 2^7$ (x).

cases. Figure 5 corroborates that the asymptotic convergence of the subiteration method is insensitive to a few reflections, in conformity with the aforementioned insensitivity of the spectral radius. However, as indicated by the sensitivity of the norm, the initial convergence behavior depends notably on the number of reflections. Furthermore, the figure shows that for $\tau = 1$ both the initial divergence and the asymptotic-convergence rate deteriorate gradually with decreasing ν . The gradual deterioration is in accordance with the smooth transition from $\text{spr } C_0$ to $\text{spr } C_\infty$ and from $\|C_0\|$ to $\|C_\infty\|$ as ν decreases in Figure 4. For $\tau = 10^2$ the convergence behavior behaves erratically with increasing τ/ν , in accordance with the erratic behavior of $\text{spr } C$ and $\|C\|$ in the transition from C_0 to C_∞ . On account of the nonnormality of the subiteration process, the error can again increase by many orders of magnitude before asymptotic convergence sets in.

7. Conclusion. We investigated the basic subiteration method for solving the aggregated equations in fluid-structure-interaction problems, on the basis of a prototypical model problem. An important new result is that the subiteration method is characterizable by recursion of a nonnormal operator. This implies that the norm and spectral radius of the operator are disparate, reflecting in profoundly dissimilar transient and asymptotic convergence behavior of the subiteration method. Convergence can be nonmonotonous and divergence can precede the asymptotic convergence, despite formal stability. By means of approximation on polynomials we established that the transient divergence can in fact amplify the initial error by many orders of

magnitude. Hence, the nonnormality induces a severe degradation in the robustness and efficiency of the subiteration method.

Auxiliary information provided by the analysis concerns the dependence of the stability and convergence of the subiteration method on the physical parameters in the problem and on the computational time step. We established that the convergence rate of the subiteration method is generally proportional to the ratio of the characteristic fluid mass over the structure mass. In the absence of reflections the convergence rate depends linearly on the ratio of the characteristic structure time to the characteristic fluid time. As the number of reflections increases, this dependence becomes quadratic. Moreover, in the absence of reflections the convergence rate is uniformly proportional to the time-step size. In particular, this proportionality implies that for appropriately small time steps, only two iterations are required for convergence to sufficient accuracy.

It is anticipated that the results generalize *mutatis mutandis* to other fluid-structure-interaction problems. In particular, we conjecture that the nonnormality of the subiteration method is independent of the underlying problem.

Appendix A. The generalized nonlinear Lax–Milgram theorem. Let $(\mathcal{U}, \langle \cdot, \cdot \rangle_{\mathcal{U}})$ and $(\mathcal{V}, \langle \cdot, \cdot \rangle_{\mathcal{V}})$ be two real Hilbert spaces with corresponding norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$. Let the functional $P : \mathcal{V} \times \mathcal{U} \mapsto \mathbb{R}$ be linear in its first argument and possibly nonlinear in its second argument, and let $p : \mathcal{V} \mapsto \mathbb{R}$ be a linear functional. Sufficient conditions for the existence of a unique solution to the variational problem

$$(A.1) \quad P(v, u) = p(v) \quad \forall v \in \mathcal{V}$$

are:

- (i) p is a continuous linear functional on $\mathcal{V}$,
- (ii) $P(\cdot, u)$ is a continuous linear functional on $\mathcal{V}$ for all $u \in \mathcal{U}$,
- (iii) $\inf_{u \neq u'} \sup_{0 \neq v \in \mathcal{V}} \frac{|P(v, u') - P(v, u)|}{\|u' - u\|_{\mathcal{U}} \|v\|_{\mathcal{V}}} \geq C_1 > 0$,
- (iv) $|P(v, u') - P(v, u)| \leq C_2 \|u' - u\|_{\mathcal{U}} \|v\|_{\mathcal{V}} \quad \forall u, u' \in \mathcal{U} \text{ and } \forall v \in \mathcal{V}$,
- (v) $\sup_{u \in \mathcal{U}} |P(v, u)| > 0 \quad \forall v \in \mathcal{V} \setminus \{0\}$

for certain positive constants C_1 and C_2 .

The proof proceeds in a similar manner as for the linear case in [18, p. 316 ff.]. By (ii) and the Riesz representation theorem, for each fixed $u \in \mathcal{U}$ there exists a unique element $\Lambda_P u$ of $\mathcal{V}$ such that $P(v, u) = \langle v, \Lambda_P u \rangle_{\mathcal{V}}$ for all $v \in \mathcal{V}$. This correspondence defines an operator $\Lambda_P : \mathcal{U} \mapsto \mathcal{V}$. Similarly, by (i) and the Riesz theorem, there exists a $v_p \in \mathcal{V}$ such that $p(v) = \langle v, v_p \rangle_{\mathcal{V}}$ for all $v \in \mathcal{V}$. Hence, the variational problem is equivalent to the operator equation $\Lambda_P u = v_p$. The existence of a unique solution to the operator equation is ensured if the operator $\Lambda_P : \mathcal{U} \mapsto \mathcal{V}$ is bijective, i.e., if Λ_P is surjective and injective; see [28, pp. 16–17].

To prove the injectivity of Λ_P , we must establish that $\Lambda_P u' = \Lambda_P u$ implies $u' = u$ or, equivalently, that $\|\Lambda_P u' - \Lambda_P u\|_{\mathcal{V}} = 0$ implies $\|u' - u\|_{\mathcal{U}} = 0$. By property (iii) it holds that

$$(A.2) \quad \begin{aligned} \|\Lambda_P u' - \Lambda_P u\|_{\mathcal{V}} &= \sup_{0 \neq v \in \mathcal{V}} \frac{|\langle v, \Lambda_P u' - \Lambda_P u \rangle_{\mathcal{V}}|}{\|v\|_{\mathcal{V}}} = \sup_{0 \neq v \in \mathcal{V}} \frac{|\langle v, \Lambda_P u' \rangle_{\mathcal{V}} - \langle v, \Lambda_P u \rangle_{\mathcal{V}}|}{\|v\|_{\mathcal{V}}} \\ &= \sup_{0 \neq v \in \mathcal{V}} \frac{|P(v, u') - P(v, u)|}{\|v\|_{\mathcal{V}}} \geq C_1 \|u' - u\|_{\mathcal{U}} \end{aligned}$$

for all u' and u in $\mathcal{U}$. Hence, injectivity is ascertained.

To prove the surjectivity of Λ_P , we must verify that $\Lambda_P \mathcal{U} = \mathcal{V}$, i.e., the range $\mathcal{R}(\Lambda_P)$ of the operator Λ_P with domain $\mathcal{U}$ coincides with $\mathcal{V}$. To this end, we show that $\mathcal{R}(\Lambda_P)$ is closed and, moreover, that its orthogonal complement in $\mathcal{V}$ is empty. To establish closedness of $\mathcal{R}(\Lambda_P)$, we note that for any pair u_n and u_m in $\mathcal{U}$ it holds that

$$\begin{aligned} \|\Lambda_P u_n - \Lambda_P u_m\|_{\mathcal{V}} &= \sup_{0 \neq v \in \mathcal{V}} \frac{|\langle v, \Lambda_P u_n \rangle_{\mathcal{V}} - \langle v, \Lambda_P u_m \rangle_{\mathcal{V}}|}{\|v\|_{\mathcal{V}}} = \sup_{0 \neq v \in \mathcal{V}} \frac{|\mathbf{P}(v, u_n) - \mathbf{P}(v, u_m)|}{\|v\|_{\mathcal{V}}} \\ (A.3) \quad &\leq C_2 \|u_n - u_m\|_{\mathcal{U}}. \end{aligned}$$

Hence, any Cauchy sequence $\{u_n\}$ in $\mathcal{U}$ generates a Cauchy sequence $\{\Lambda_P u_n\}$ in $\mathcal{R}(\Lambda_P)$ and $u_n \rightarrow u \in \mathcal{U}$ implies $\Lambda_P u_n \rightarrow \Lambda_P u \in \mathcal{R}(\Lambda_P)$. Therefore, $\mathcal{R}(\Lambda_P)$ is closed. Now suppose that the orthogonal complement $\mathcal{R}(\Lambda_P)^\perp$ of $\mathcal{R}(\Lambda_P)$ in $\mathcal{V}$ is nonempty. Then there exists a $v' \in \mathcal{R}(\Lambda_P)^\perp$ such that $\langle v', \Lambda_P u \rangle_{\mathcal{V}} = 0$ for all $u \in \mathcal{U}$. However, this is equivalent to $|\langle v', \Lambda_P u \rangle_{\mathcal{V}}| = |\mathbf{P}(v', u)| = 0$ for all $u \in \mathcal{U}$, in contradiction to stipulation (v). Hence, such a v' is nonexistent. This implies that $\mathcal{V} = \mathcal{R}(\Lambda_P) \oplus \mathcal{R}(\Lambda_P)^\perp = \mathcal{R}(\Lambda_P) \oplus \emptyset = \mathcal{R}(\Lambda_P)$. Hence, the operator Λ_P is surjective.

Appendix B. Unboundedness of $\|\mathbf{C}_0\|_{L^2}$ and boundedness of $\|\mathbf{C}_0\|_{H^1}$.

Considering a bounded open interval $\mathcal{T} := (0, \tau)$, the objective is to establish that (i) $\|\mathbf{C}_0\|_{L^2} = \infty$ and (ii) $\|\mathbf{C}_0\|_{H^1} \leq 2\tau < \infty$.

(i) To show that $\mathbf{C}_0$ is unbounded on $L^2(\mathcal{T})$ we consider the sequence of continuous functions $\{z_n\}$ with elements

$$(B.1) \quad z_n(s) := \begin{cases} 1 - \sqrt{1 - (ns - 1)^2} & \text{if } 0 \leq s \leq 1/n, \\ 0 & \text{if } s > 1/n. \end{cases}$$

The sequence $\{z_n\}$ is a Cauchy sequence in L^2 and, hence, $z_n \rightarrow z \in L^2$ as $n \rightarrow \infty$. It is important to note that $\|z_n\|_{L^2(\mathcal{T})} = O(1/\sqrt{n})$ as $n \rightarrow \infty$ for any τ independent of n . For the image of z_n under $\mathbf{C}_0$ according to (4.10) we obtain

$$(B.2) \quad v_n := \mathbf{C}_0 z_n = c_1 \sin(s) + c_2 \cos(s) - \sin(s) \int \sin(s) z_n(s) ds - \cos(s) \int \cos(s) z_n(s) ds,$$

the constants c_1 and c_2 being determined by the homogeneous initial conditions on v_n . For $s > 1/n$, v_n can be expanded asymptotically as

$$(B.3) \quad v_n = \left(1 + \left(\frac{\pi}{2} - \frac{5}{6}\right) \frac{1}{n^2} + O\left(\frac{1}{n^4}\right)\right) \sin(s)$$

$$(B.4) \quad + \left(\left(\frac{\pi}{2} - 1\right) \frac{1}{n} + \left(\frac{1}{2} - \frac{5\pi}{16}\right) \frac{1}{n^3} + O\left(\frac{1}{n^5}\right)\right) \cos(s),$$

as $n \rightarrow \infty$. The expression for v_n in the interval $0 \leq s \leq 1/n$ is not required in what follows. From (B.3) it follows that

$$(B.5) \quad \|\mathbf{C}_0\|_{L^2} := \sup_{0 \neq z \in L^2} \frac{\|\mathbf{C}_0 z\|_{L^2}}{\|z\|_{L^2}} \geq \frac{\|v_n\|_{L^2}}{\|z_n\|_{L^2}} \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

which completes the proof.

(ii) Denoting by v the image of an arbitrary function $z \in H^1$ under C_0 and setting $2\mu/\nu := 1$ for convenience, (4.10) implies

$$(B.6) \quad D^2v + v = -Dz, \quad v(0) = 0, \quad Dv(0) = 0.$$

Multiplying the differential equation by Dv and integrating from 0 to s , we obtain

$$(B.7) \quad \begin{aligned} -\int_0^s Dv(\xi) Dz(\xi) d\xi &= \int_0^s Dv(\xi) (D^2v(\xi) + v(\xi)) d\xi \\ &= \int_0^s \frac{d}{d\xi} \left(\frac{1}{2} [Dv(\xi)]^2 + \frac{1}{2} [v(\xi)]^2 \right) d\xi = \frac{1}{2} [Dv(s)]^2 + \frac{1}{2} [v(s)]^2. \end{aligned}$$

The final identity in (B.7) is obtained by invoking the initial conditions. Hence,

$$(B.8) \quad \begin{aligned} \frac{1}{2} \|v\|_{H^1}^2 &= \frac{1}{2} \int_0^\tau [Dv(s)]^2 + [v(s)]^2 ds = \left| \int_0^\tau \int_0^s Dv(\xi) Dz(\xi) d\xi ds \right| \\ &\leq \int_0^\tau \|Dv\|_{L^2(0,s)} \|Dz\|_{L^2(0,s)} ds \leq \tau \|Dv\|_{L^2(\mathcal{T})} \|Dz\|_{L^2(\mathcal{T})} \leq \tau \|v\|_{H^1(\mathcal{T})} \|z\|_{H^1(\mathcal{T})}. \end{aligned}$$

The first upper bound is a consequence of Schwarz' inequality (see, e.g., [28]). Equation (B.8) implies $\|C_0 z\|_{H^1} \leq 2\tau \|z\|_{H^1}$ and, since z is arbitrary, $\|C_0\|_{H^1} \leq 2\tau$. This was to be proved.

REFERENCES

- [1] J. ALONSO AND A. JAMESON, *Fully-implicit time-marching aeroelastic solutions*, AIAA, 56 (1994), pp. 1–13.
- [2] M. BATHE AND R. KAMM, *A fluid-structure interaction finite element analysis of pulsatile blood flow through a compliant stenotic artery*, Journal of Biomechanical Engineering, 121 (1999), pp. 361–369.
- [3] K. BÖHMER, P. HEMKER, AND H. STETTER, *The defect correction approach*, Comput. Suppl., 5 (1984), pp. 1–32.
- [4] L. BRODERICK AND J. LEONARD, *Nonlinear response of membranes to ocean waves using boundary and finite elements*, Ocean Engng., 22 (1995), pp. 731–745.
- [5] E. VAN BRUMMELEN, S. HULSHOFF, AND R. DE BORST, *Energy conservation under incompatibility for fluid-structure interaction problems*, Comput. Methods Appl. Mech. Engng., 192 (2003), pp. 2727–2748.
- [6] E. VAN BRUMMELEN, C. MICHLER, AND R. DE BORST, *Interface-GMRES(R) Acceleration of Subiteration for Fluid-Structure-Interaction Problems*, Tech. report DACS-05-001, Delft University of Technology, Delft, The Netherlands, 2005. Available at <http://www.em.lr.tudelft.nl/downloads/DACS-05-001.pdf>.
- [7] C. FARHAT, P. GEUZAIN, AND G. BROWN, *Application of a three-field nonlinear fluidstructure formulation to the prediction of the aeroelastic parameters of an f-16 fighter*, Comput. & Fluids, 32 (2003), pp. 3–29.
- [8] C. FARHAT, P. GEUZAIN, AND C. GRANDMONT, *The discrete geometric conservation law and the nonlinear stability of ALE schemes for the solution of flow problems on moving grids*, J. Comput. Phys., 174 (2001), pp. 669–694.
- [9] C. FELIPPA, K. PARK, AND C. FARHAT, *Partitioned analysis of coupled mechanical systems*, Comput. Methods Appl. Mech. Engng., 190 (2001), pp. 3247–3270.
- [10] I. GOHBERG AND M. KREĬN, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Translations of Mathematical Monographs 18, AMS, Providence, RI, 1969.
- [11] G. GOLUB AND P. LOAN, *Matrix Computations*, Johns Hopkins, 1996.
- [12] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, Frontiers Appl. Math. 17, SIAM, Philadelphia, 1997.
- [13] M. HEIL, *Stokes flow in an elastic tube—a large-displacement fluid-structure interaction problem*, Internat. J. Numer. Methods Fluids, 28 (1998), pp. 243–265.

- [14] T. KATO, *Perturbation Theory for Linear Operators*, Grundlehren der Mathematischen Wissenschaft 132, Springer, Berlin, corrected print of the 2nd ed., 1984.
- [15] J. KEVORKIAN AND J. COLE, *Perturbation Methods in Applied Mathematics*, Appl. Math. Sci. 34, Springer-Verlag, Berlin, 1981.
- [16] P. LETALLEC AND J. MOURO, *Fluid structure interaction with large structural displacements*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 3039–3067.
- [17] R. B. MELVILLE AND S. A. MORTON, *Fully-implicit aeroelasticity on overset grid systems*, 36th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, 1998.
- [18] J. ODEN AND J. REDDY, *An Introduction to the Mathematical Theory of Finite Elements*, Pure and Applied Mathematics, John Wiley & Sons, New York, 1974.
- [19] S. PIPERNO, *Simulation Numérique de Phénomènes d'Interaction Fluide-Structure*, Ph.D. thesis, école Nationale des Ponts et Chaussées, France, 1995 (partly in French).
- [20] S. PIPERNO, *Explicit/implicit fluid/structure staggered procedures with a structural predictor and fluid subcycling for 2D inviscid aeroelastic simulations*, Internat. J. Numer. Methods Fluids, 25 (1997), pp. 1207–1226.
- [21] S. PIPERNO AND C. FARHAT, *Partitioned procedures for the transient solution of coupled aeroelastic problems. II: Energy transfer analysis and three-dimensional applications*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 3147–3170.
- [22] S. PIPERNO, C. FARHAT, AND B. LARROUTUROU, *Partitioned procedures for the transient solution of coupled aeroelastic problems. I: Model problem, theory and two-dimensional application*, Comput. Methods Appl. Mech. Engrg., 124 (1995), pp. 79–112.
- [23] S. C. REDDY, P. J. SCHMID, AND D. S. HENNINGSON, *Pseudospectra of the Orr–Sommerfeld operator*, SIAM J. Appl. Math., 53 (1993), pp. 15–47.
- [24] I. ROBERTSON, S. SHERWIN, AND P. BEARMAN, *Prediction of flutter instability due to cross winds of the second forth road bridge*, in Computational Fluid Dynamics '01: Proceedings of the ECCOMAS CFD Conference 2001, Swansea, Wales, U.K., 2001, on CD-rom.
- [25] L. TREFETHEN, *Pseudospectra of linear operators*, SIAM Rev., 39 (1997), pp. 383–406.
- [26] L. TREFETHEN, A. TREFETHEN, S. REDDY, AND T. DRISCOLL, *Hydrodynamic stability without eigenvalues*, Science, 261 (1993), pp. 578–584.
- [27] T. WRIGHT, *Matlab Pseudospectra GUI*, Oxford University Computing Laboratory. Available at <http://www.comlab.ox.ac.uk/pseudospectra/eigtool/>.
- [28] E. ZEIDLER, *Applied Functional Analysis: Applications to Mathematical Physics*, Appl. Math. Sci. 108, Springer, Berlin, 1995.