SIAM J. SCI. COMPUT.
Vol. 33, No. 5, pp. 2785–2804

ITERATIVE METHODS FOR NEUTRON TRANSPORT EIGENVALUE PROBLEMS*

FYNN SCHEBEN† AND IVAN G. GRAHAM†

Abstract. We discuss iterative methods for computing criticality in nuclear reactors. In general this requires the solution of a generalized eigenvalue problem for an unsymmetric integro-differential operator in six independent variables, modeling transport, scattering, and fission, where the dependent variable is the neutron angular flux. In engineering practice this problem is often solved iteratively, using some variant of the inverse power method. Because of the high dimension, matrix representations for the operators are often not available and the inner solves needed for the eigenvalue iteration are implemented by matrix-free inner iterations. This leads to technically complicated inexact iterative methods, for which there appears to be no published rigorous convergence theory. For the monoenergetic homogeneous model case with isotropic scattering and vacuum boundary conditions, we show that, before discretization, the general nonsymmetric eigenproblem for the angular flux is equivalent to a certain related eigenproblem for the scalar flux, involving a symmetric positive definite weakly singular integral operator (in space only). This correspondence to a symmetric problem (in a space of reduced dimension) permits us to give a convergence theory for inexact inverse iteration and related methods. In particular this theory provides rather precise criteria on how accurate the inner solves need to be in order for the whole iterative method to converge. We also give examples of discretizations which have a corresponding symmetric finite-dimensional reduced form. The theory is illustrated with numerical examples for several test problems of physical relevance, using GMRES as the inner solver.

Key words. neutron transport, criticality, generalized eigenvalue problem, symmetry, inexact inverse iteration

AMS subject classifications. 45C05, 65F15, 65N25, 65R20, 65Z05, 82D75

DOI. 10.1137/100799022

1. Reactor criticality problems.

Climate change is a challenging problem of great contemporary interest. It is still open to debate whether nuclear power is a solution to this problem or not, but certainly ensuring the safety and optimal performance of existing nuclear reactors is an important task of great environmental significance. When operating a nuclear reactor, the engineer seeks to achieve a sustainable chain reaction where the neutrons produced balance the neutrons that either are absorbed or leave the system through the outer boundary. The chain reaction depends on the material composition and geometry of the reactor and can be controlled by inserting or removing control rods.

Mathematically the problem of modeling this balance may be written as

$$\mathcal{T}\Psi - \mathcal{S}\Psi = \lambda \mathcal{F}\Psi, \tag{1.1}$$

where $\Psi(\mathbf{r}, E, \mathbf{\Omega})$ is the flux of neutrons per unit volume with energy $E \in \mathbb{R}^+$ at position $\mathbf{r} \in \mathbb{R}^3$ in direction $\mathbf{\Omega} \in \mathbb{S}^2$ (the unit sphere in $\mathbb{R}^3$) and the operators $\mathcal{T}$, $\mathcal{S}$, and $\mathcal{F}$ describing, respectively, transport, scattering, and fission in the reactor are

*Received by the editors June 16, 2010; accepted for publication (in revised form) February 25, 2011; published electronically October 27, 2011.
http://www.siam.org/journals/sisc/33-5/79902.html
†Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom (F.Scheben@bath.ac.uk, I.G.Graham@bath.ac.uk). The first author's research was supported by Serco Technical and Assurance Services, the UK Engineering and Physical Sciences Research Council, and the University of Bath.

given by

$$\mathcal{T}\Psi = \boldsymbol{\Omega}\cdot\nabla\Psi(\mathbf{r},E,\boldsymbol{\Omega}) + \sigma(\mathbf{r},E)\Psi(\mathbf{r},E,\boldsymbol{\Omega}),$$
$$\mathcal{S}\Psi = \frac{1}{4\pi}\int_{\mathbb{R}^+}\int_{\mathbb{S}^2}\sigma_s(\mathbf{r},E',E,\boldsymbol{\Omega}',\boldsymbol{\Omega})\Psi(\mathbf{r},E',\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'\,\mathrm{d}E',$$
$$\mathcal{F}\Psi = \frac{\chi(E)}{4\pi}\int_{\mathbb{R}^+}\nu(\mathbf{r},E')\sigma_f(\mathbf{r},E')\int_{\mathbb{S}^2}\Psi(\mathbf{r},E',\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'\,\mathrm{d}E'$$

(see, e.g., [14, equations (1–16), (1–104)]). Here σ_s and σ_f are the (macroscopic) *scattering* and *fission cross-sections*, ν is the *neutron yield*, and χ is the *fission neutron distribution*. The *total cross-section* σ is defined by

$$\sigma(\mathbf{r},E) = \sigma_c(\mathbf{r},E) + \frac{1}{4\pi}\int_{\mathbb{R}^+}\int_{\mathbb{S}^2}\sigma_s(\mathbf{r},E,E',\boldsymbol{\Omega},\boldsymbol{\Omega}')\,\mathrm{d}\boldsymbol{\Omega}'\,\mathrm{d}E' + \sigma_f(\mathbf{r},E), \tag{1.2}$$

where σ_c denotes the *capture cross-section*.

Equation (1.1) is to be solved for $\mathbf{r}$ in some bounded domain $V\subset\mathbb{R}^3$ (the reactor), subject to suitable boundary conditions (see below for an example). The eigenvalue λ with the smallest modulus has direct physical meaning: Using Krein–Rutman arguments, under quite general assumptions, this can be shown to be real, positive, and simple [18]. The value of λ describes the balance between transport and scattering on one hand and fission on the other. The reactor is called *subcritical* if $\lambda>1$, *supercritical* if $\lambda<1$, and *critical* if $\lambda=1$. Designing a reactor so that λ is close to 1 is a key inverse problem in nuclear engineering. To do this we need efficient methods to compute λ for any given reactor (the forward problem), and that is the focus of this paper.

We note that there is a large amount of background literature on neutron transport theory and nuclear engineering (e.g., [3, 14, 24]). There has also been widespread interest from numerical analysts (e.g., [2, 12, 15, 20]), but this activity is related mainly to the solution of source problems where a unique solution Ψ to (1.1) is to be found for given λ and with the addition of a forcing term on the right-hand side. Some discretization error estimates for computed eigenvalues are presented, for example, in [2, 20], and a brief discussion of the inverse power method in the context of neutron transport is in [1]. The eigenvalue problem considered here and related problems have a large classical literature (e.g., [16, 17]). However, we do not know a systematic convergence analysis for eigenvalue iterative methods of shift-invert type in the case when only inexact solves are available at each step of the iteration, as is often done in engineering practice.

Advanced iterative methods such as (variants of) Arnoldi's method have been applied to the criticality problem quite recently (e.g., [25]). However, not many of the recent advances in the theory of inexact iterative methods for eigenvalue problems seem to have been applied to the solution of the reactor criticality problem, even though this problem has remained an active area of interest in nuclear engineering over many years. In the present paper we exploit numerical analysis results from the 1980's, combined with recent advances on inexact eigenvalue iterative methods, to obtain a new analysis of eigenvalue iterations for neutron criticality problems (and indeed new enhancements of existing algorithms). The essential ingredient of this paper is the observation that, while (1.1) is an unsymmetric equation in six independent variables, it has, for a class of model problems, an underlying reduced form which is self-adjoint, and this structure allows us to give a rather simple analysis of eigenvalue iterations and identify useful new methods.

Model problems. Now let us consider the homogeneous model problem of isotropic scattering in the monoenergetic case subject to vacuum boundary conditions. Then $\chi = 1$ and all the cross-sections are constant with $\sigma = \sigma_c + \sigma_s + \sigma_f$. In this reactor the neutrons travel with the same constant speed, and no neutrons enter the reactor from the outside.

Three-dimensional (3D) model. In the 3D case, (1.1) takes the form

$$\boldsymbol{\Omega} \cdot \nabla \Psi(\mathbf{r}, \boldsymbol{\Omega}) + \sigma \Psi(\mathbf{r}, \boldsymbol{\Omega}) - \frac{\sigma_s}{4\pi} \int_{\mathbb{S}^2} \Psi(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}' = \lambda \, \frac{\nu\sigma_f}{4\pi} \int_{\mathbb{S}^2} \Psi(\mathbf{r}, \boldsymbol{\Omega}') \, \mathrm{d}\boldsymbol{\Omega}', \tag{1.3}$$

and this is to be solved for all $(\mathbf{r}, \boldsymbol{\Omega}) \in V \times \mathbb{S}^2$, subject to

$$\Psi(\mathbf{r}, \boldsymbol{\Omega}) = 0 \quad \text{when} \quad \mathbf{n}(\mathbf{r}) \cdot \boldsymbol{\Omega} < 0, \quad \mathbf{r} \in \partial V, \tag{1.4}$$

where $\mathbf{n}(\mathbf{r})$ denotes the outward unit normal at $\mathbf{r} \in \partial V$, the boundary of V.

Two subcases of this have been considered in the literature. To describe them, let $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$ denote the usual spherical polar coordinates on $\mathbb{S}^2$.

Two-dimensional (2D) model. Here it is assumed that $\Psi(\mathbf{r}, \boldsymbol{\Omega}) = \Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}})$, where $\widetilde{\mathbf{r}} \in \widetilde{V} \subset \mathbb{R}^2$ and $\widetilde{\boldsymbol{\Omega}} = (\cos\varphi, \sin\varphi)$ lies on the unit circle $\mathbb{S}^1$. The resulting model problem

$$\widetilde{\boldsymbol{\Omega}} \cdot \widetilde{\nabla} \Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}}) + \sigma \Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}}) - \frac{\sigma_s}{2\pi} \int_{\mathbb{S}^1} \Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}}') \, \mathrm{d}\widetilde{\boldsymbol{\Omega}}' = \lambda \, \frac{\nu\sigma_f}{2\pi} \int_{\mathbb{S}^1} \Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}}') \, \mathrm{d}\widetilde{\boldsymbol{\Omega}}'$$

is to be solved on $\widetilde{V} \times \mathbb{S}^1$ (where $\widetilde{\nabla}$ denotes the 2D gradient), subject to

$$\Psi(\widetilde{\mathbf{r}}, \widetilde{\boldsymbol{\Omega}}) = 0 \quad \text{when} \quad \widetilde{\mathbf{n}}(\widetilde{\mathbf{r}}) \cdot \widetilde{\boldsymbol{\Omega}} < 0, \quad \widetilde{\mathbf{r}} \in \partial\widetilde{V},$$

where $\widetilde{\mathbf{n}}(\widetilde{\mathbf{r}})$ again denotes the outward unit normal at $\widetilde{\mathbf{r}} \in \partial\widetilde{V}$ (see, e.g., [2] and the references therein).

One-dimensional (1D) model. Here it is assumed that $\Psi(\mathbf{r}, \boldsymbol{\Omega}) = \Psi(z, \mu)$, where $z \in [0, 1]$ and $\mu = \cos\theta \in [-1, 1]$, and (1.3), (1.4) reduce to

$$\mu \frac{\partial}{\partial z} \Psi(z, \mu) + \sigma \Psi(z, \mu) - \frac{\sigma_s}{2} \int_{-1}^{1} \Psi(z, \mu') \, \mathrm{d}\mu' = \lambda \, \frac{\nu\sigma_f}{2} \int_{-1}^{1} \Psi(z, \mu') \, \mathrm{d}\mu' \tag{1.5}$$

to be solved on $[0, 1] \times [-1, 1]$, subject to

$$\Psi(0, \mu) = 0 \quad \text{when} \quad \mu > 0 \quad \text{and} \quad \Psi(1, \mu) = 0 \quad \text{when} \quad \mu < 0. \tag{1.6}$$

This ("1D slab geometry") model has received a lot of attention in the literature (e.g., in [15, 20]).

In section 2 we establish the relation of all three model problems to a symmetric positive definite integral operator eigenvalue problem. We then discuss iterative methods for the computation of the eigenvalue (section 3) and study their convergence. Section 4 gives examples of discretizations that preserve the underlying symmetry in the case of the 1D model problem, and in section 5 we give numerical results which illustrate the presented theory. This and related work are described in detail in the Ph.D. thesis [22].

2. Relation to a symmetric problem. Throughout this section we will work exclusively with the 3D model problem (1.3), (1.4), but all the methods we develop will also be applicable to the 2D and 1D model problems, as we shall remark below. We will show that the generalized eigenvalue problem (1.3), (1.4) for (λ, Ψ) is equivalent to a corresponding "reduced" generalized eigenvalue problem for (λ, ϕ), where ϕ is the scalar flux

$$\phi(\mathbf{r}) = \mathcal{P}\Psi := \frac{1}{4\pi} \int_{\mathbb{S}^2} \Psi(\mathbf{r}, \mathbf{\Omega}') \, \mathrm{d}\mathbf{\Omega}'. \tag{2.1}$$

As we shall see below, the reduced problem involves a certain self-adjoint compact integral operator, and this allows us to show that the eigenvalues of the original problem are real and positive and that the eigenfunctions comprise a complete orthonormal sequence. This is key to the eigenvalue convergence analysis given later in the paper.

Integral equation reformulations of the neutron transport source problem are well known (see, e.g., [14] or [2, 20] for 2D and 1D analogues), and they are often used as a tool in the design and/or analysis of iterative schemes; see, e.g., [5, 9]. However, we have not seen the reduction of the eigenvalue problem written explicitly in the form we present below and do not know of any literature which exploits this structure for a convergence analysis of eigenvalue iterative methods.

To make the reduction mathematically precise, we consider the usual Lebesgue space $L^2(V)$ with norm $\|\cdot\|_{L^2(V)}$. Also for any $1 \le p \le \infty$ we introduce the space

$$L^2(V, L^p(\mathbb{S}^2)) := \left\{ \Psi : V \times \mathbb{S}^2 \to \mathbb{R} : \int_V \|\Psi(\mathbf{r}, \cdot)\|^2_{L^p(\mathbb{S}^2)} \, \mathrm{d}\mathbf{r} < \infty \right\}, \tag{2.2}$$

with norm $\|\Psi\|^2_{L^2(V,L^p(\mathbb{S}^2))} := \int_V \|\Psi(\mathbf{r}, \cdot)\|^2_{L^p(\mathbb{S}^2)} \, \mathrm{d}\mathbf{r}$. Clearly, the operator $\mathcal{P}$ defined in (2.1) is a bounded linear operator from $L^2(V, L^1(\mathbb{S}^2))$ to $L^2(V)$. In the following lemma, we make use of the notation

$$d(\mathbf{r}, \mathbf{\Omega}) := \inf\{s > 0 : \mathbf{r} - s\mathbf{\Omega} \notin V\}.$$

Throughout we assume that V is a convex domain in $\mathbb{R}^3$, and for convenience we assume its boundary ∂V is C^1, so that the normal direction $\mathbf{n}$ is a continuous function on ∂V. It then follows that $\mathbf{\Omega}$ is an inward pointing direction at the boundary point $\mathbf{r} - d(\mathbf{r}, \mathbf{\Omega})\mathbf{\Omega} \in \partial V$, and so by (1.4),

$$\Psi(\mathbf{r} - d(\mathbf{r}, \mathbf{\Omega})\mathbf{\Omega}, \mathbf{\Omega}) = 0. \tag{2.3}$$

LEMMA 2.1. *Suppose $g \in L^2(V, L^\infty(\mathbb{S}^2))$, and consider the problem of solving*

$$\mathcal{T}\Psi(\mathbf{r}, \mathbf{\Omega}) = g(\mathbf{r}, \mathbf{\Omega}) \tag{2.4}$$

on $V \times \mathbb{S}^2$, together with the boundary condition (1.4). *This problem has a unique solution $\Psi \in L^2(V, L^1(\mathbb{S}^2))$ given by*

$$\Psi(\mathbf{r}, \mathbf{\Omega}) = \int_0^{d(\mathbf{r},\mathbf{\Omega})} \exp(-\sigma s) g(\mathbf{r} - s\mathbf{\Omega}, \mathbf{\Omega}) \, \mathrm{d}s \quad \textit{for} \quad (\mathbf{r}, \mathbf{\Omega}) \in V \times \mathbb{S}^2. \tag{2.5}$$

Proof. First observe that (2.4) is equivalent to the statement

$$-\frac{\mathrm{d}}{\mathrm{d}s} \left[\Psi(\mathbf{r} - s\mathbf{\Omega}, \mathbf{\Omega}) \exp(-\sigma s) \right] = g(\mathbf{r} - s\mathbf{\Omega}, \mathbf{\Omega}) \exp(-\sigma s), \tag{2.6}$$

provided $\mathbf{r} \in V$, $\mathbf{\Omega} \in \mathbb{S}^2$, and $s > 0$ are such that $\mathbf{r} - s\mathbf{\Omega} \in V$.

To show that the formula (2.5) yields a solution of (2.4), observe that if (2.5) holds, then, provided $\mathbf{r} - s\boldsymbol{\Omega} \in V$,

$$\Psi(\mathbf{r} - s\boldsymbol{\Omega}, \boldsymbol{\Omega}) = \int_0^{d(\mathbf{r}-s\boldsymbol{\Omega},\boldsymbol{\Omega})} \exp(-\sigma s')g(\mathbf{r} - (s+s')\boldsymbol{\Omega}, \boldsymbol{\Omega})\,\mathrm{d}s'.$$

Now making the change of variables $s'' = s' + s$ and observing that

$$d(\mathbf{r} - s\boldsymbol{\Omega}, \boldsymbol{\Omega}) + s = d(\mathbf{r}, \boldsymbol{\Omega}),$$

we obtain

$$\Psi(\mathbf{r} - s\boldsymbol{\Omega}, \boldsymbol{\Omega}) \exp(-\sigma s) = \int_s^{d(\mathbf{r},\boldsymbol{\Omega})} \exp(-\sigma s'')g(\mathbf{r} - s''\boldsymbol{\Omega}, \boldsymbol{\Omega})\,\mathrm{d}s'', \tag{2.7}$$

which can easily be seen to imply (2.6).

Uniqueness of the solution to (2.4) is trivial since with $g = 0$, integrating (2.6) from $s = 0$ to $s = d(\mathbf{r}, \boldsymbol{\Omega})$ easily shows that Ψ vanishes. The proof that $\Psi \in L^2(V, L^1(\mathbb{S}^2))$ is deferred to Remark 2.3. ∎

LEMMA 2.2. *Consider* (2.4) *in the special case* $g(\mathbf{r}, \boldsymbol{\Omega}) = g(\mathbf{r})$ *with* $g \in L^2(V)$, *and using the solution* Ψ, *define* $\phi = \mathcal{P}\Psi$. *Then* $\phi \in L^2(V)$ *and*

$$\left.\begin{aligned} \phi(\mathbf{r}) = (\mathcal{K}_\sigma g)(\mathbf{r}) &:= \int_V k_\sigma(\mathbf{r} - \mathbf{r}')g(\mathbf{r}')\,\mathrm{d}\mathbf{r}', \\ \textit{where} \quad k_\sigma(\mathbf{x}) &:= \frac{1}{4\pi}\,\frac{\exp(-\sigma\|\mathbf{x}\|_2)}{\|\mathbf{x}\|_2^2}\,, \quad \mathbf{x} \in \mathbb{R}^3. \end{aligned}\right\} \tag{2.8}$$

Proof. Using (2.5) and applying $\mathcal{P}$ yield

$$\phi(\mathbf{r}) = \int_{\mathbb{S}^2} \int_0^{d(\mathbf{r},\boldsymbol{\Omega})} \frac{\exp(-\sigma s)}{4\pi s^2}\, g(\mathbf{r} - s\boldsymbol{\Omega})\, s^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\Omega}.$$

Now, using spherical coordinates centered at $\mathbf{r}$ with $\mathbf{r}' = \mathbf{r} - s\boldsymbol{\Omega}$, we obtain (2.8). Finally, it follows from [13, p. 324] that $\mathcal{K}_\sigma$ is bounded on $L^2(V)$, and thus it follows that $\phi \in L^2(V)$. ∎

Remark 2.3. The fact that Ψ given by (2.5) lies in $L^2(V, L^1(\mathbb{S}^2))$ when $g \in L^2(V, L^\infty(\mathbb{S}^2))$ can now be easily proved by using (2.5) to obtain

$$\begin{aligned} \|\Psi(\mathbf{r}, \cdot)\|_{L^1(\mathbb{S}^2)} &\le 4\pi \int_{\mathbb{S}^2} \int_0^{d(\mathbf{r},\boldsymbol{\Omega})} \frac{\exp(-\sigma s)}{4\pi s^2} \|g(\mathbf{r} - s\boldsymbol{\Omega}, \cdot)\|_{L^\infty(\mathbb{S}^2)}\, s^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\Omega} \\ &= 4\pi\,(\mathcal{K}_\sigma f)(\mathbf{r}), \end{aligned}$$

where $f(\mathbf{r}) = \|g(\mathbf{r}, \cdot)\|_{L^\infty(\mathbb{S}^2)} \in L^2(V)$. The result follows since $\mathcal{K}_\sigma$ is a bounded linear operator on $L^2(V)$.

COROLLARY 2.4. *If* (λ, Ψ) *is an eigenpair in* $L^2(V, L^1(\mathbb{S}^2))$ *for* (1.3), (1.4), *then* (λ, ϕ), *with* $\phi = \mathcal{P}\Psi$, *is an eigenpair in* $L^2(V)$ *of the reduced generalized eigenvalue problem*

$$\phi(\mathbf{r}) - \sigma_s \mathcal{K}_\sigma \phi(\mathbf{r}) = \lambda\,\nu\sigma_f \mathcal{K}_\sigma \phi(\mathbf{r}), \quad \mathbf{r} \in V. \tag{2.9}$$

Conversely, if (λ, ϕ) *is an eigenpair of problem* (2.9) *in* $L^2(V)$, *and if we define* Ψ *by solving* $\mathcal{T}\Psi = \sigma_s\phi + \lambda\,\nu\sigma_f\phi$, *subject to the boundary condition* (1.4), *then* (λ, Ψ) *is an eigenpair for* (1.3) *in* $L^2(V, L^1(\mathbb{S}^2))$.

Proof. Suppose (λ, Ψ) is an eigenpair of (1.3) in $L^2(V, L^1(\mathbb{S}^2))$; then

$$\mathcal{T}\Psi(\mathbf{r}, \mathbf{\Omega}) - \sigma_s \phi(\mathbf{r}) = \lambda \nu \sigma_f \phi(\mathbf{r}), \quad \text{where} \quad \phi(\mathbf{r}) = \mathcal{P}\Psi(\mathbf{r}, \mathbf{\Omega}).$$

Now it follows from Lemma 2.2 and the linearity of $\mathcal{K}_\sigma$, that (2.9) holds in $L^2(V)$.

To prove the converse statement, let $\Psi \in L^2(V, L^1(\mathbb{S}^2))$ be the unique solution of $\mathcal{T}\Psi = \sigma_s \phi + \lambda \nu \sigma_f \phi$ and set $\tilde{\phi} := \mathcal{P}\Psi$. Lemma 2.2 and then (2.9) imply

$$\tilde{\phi}(\mathbf{r}) = \sigma_s \mathcal{K}_\sigma \phi(\mathbf{r}) + \lambda \nu \sigma_f \mathcal{K}_\sigma \phi(\mathbf{r}) = \phi(\mathbf{r}).$$

Hence $\mathcal{T}\Psi = \sigma_s \tilde{\phi} + \lambda \nu \sigma_f \tilde{\phi} = \mathcal{S}\Psi + \lambda \mathcal{F}\Psi$, as required. ∎

The operator $\mathcal{K}_\sigma$ is clearly self-adjoint. Furthermore, by [13, p. 332], $\mathcal{K}_\sigma$ is compact on $L^2(V)$. In addition we have the following result.

LEMMA 2.5. *All the eigenvalues of the operator $\mathcal{K}_\sigma$ are positive, and so $\mathcal{K}_\sigma$ is positive definite. Moreover, $\|\mathcal{K}_\sigma\|_{\mathscr{L}(L^2(V))} \le 1/\sigma$.*

Proof. Suppose $\mathcal{K}_\sigma f = \omega f$ for some $f \in L^2(V)$ and some eigenvalue ω which must be real. Let $\Psi \in L^2(V, L^1(\mathbb{S}^2))$ be the solution of $\mathcal{T}\Psi = f$, satisfying (1.4). Then, defining $\phi := \mathcal{P}\Psi$ and using Lemma 2.1, we have $\phi = \mathcal{K}_\sigma f = \omega f$. Thus

$$\begin{aligned}\omega f^2(\mathbf{r}) = \phi(\mathbf{r}) f(\mathbf{r}) &= \frac{1}{4\pi} \int_{\mathbb{S}^2} \Psi(\mathbf{r}, \mathbf{\Omega}) f(\mathbf{r}) \, \mathrm{d}\mathbf{\Omega} \\ &= \frac{1}{4\pi} \int_{\mathbb{S}^2} \mathbf{\Omega} \cdot [\Psi(\mathbf{r}, \mathbf{\Omega}) \nabla \Psi(\mathbf{r}, \mathbf{\Omega})] \, \mathrm{d}\mathbf{\Omega} + \frac{\sigma}{4\pi} \int_{\mathbb{S}^2} \Psi^2(\mathbf{r}, \mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega}.\end{aligned}$$

Integrating over V and applying the divergence theorem, the first term on the right-hand side becomes

$$\begin{aligned}\frac{1}{4\pi} \int_{\mathbb{S}^2} \mathbf{\Omega} \cdot \left[\int_V \Psi(\mathbf{r}, \mathbf{\Omega}) \nabla \Psi(\mathbf{r}, \mathbf{\Omega}) \, \mathrm{d}\mathbf{r} \right] \mathrm{d}\mathbf{\Omega} &= \frac{1}{8\pi} \int_{\mathbb{S}^2} \mathbf{\Omega} \cdot \left[\int_V \nabla \left[\Psi^2(\mathbf{r}, \mathbf{\Omega}) \right] \, \mathrm{d}\mathbf{r} \right] \mathrm{d}\mathbf{\Omega} \\ &= \frac{1}{8\pi} \int_{\mathbb{S}^2} \int_V \nabla \cdot \left[\Psi^2(\mathbf{r}, \mathbf{\Omega}) \mathbf{\Omega} \right] \, \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{\Omega} \\ &= \frac{1}{8\pi} \int_{\mathbb{S}^2} \int_{\partial V} \Psi^2(\mathbf{r}, \mathbf{\Omega}) \, [\mathbf{\Omega} \cdot \mathbf{n}(\mathbf{r})] \, \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{\Omega} \ge 0,\end{aligned}$$

where we used (1.4) for the final estimate. Hence

$$\omega \int_V f^2(\mathbf{r}) \, \mathrm{d}\mathbf{r} \ge \frac{\sigma}{4\pi} \int_V \int_{\mathbb{S}^2} \Psi^2(\mathbf{r}, \mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega} \, \mathrm{d}\mathbf{r},$$

and finally, as $f \neq 0$, the integrals on both sides are positive and it follows that $\omega > 0$. The positive-definiteness of $\mathcal{K}_\sigma$ then follows from, e.g., [21, p. 193] or [11, section 3.5].

To prove the bound on $\|\mathcal{K}_\sigma\|_{\mathscr{L}(L^2(V))}$, note that $\mathcal{K}_\sigma \phi = k_\sigma * \phi^e$, where ϕ^e is the trivial extension of ϕ to all of $\mathbb{R}^3$ by choosing ϕ^e to be zero outside of V, and $*$ denotes convolution on $\mathbb{R}^3$. Then, applying Young's inequality for convolutions (see, e.g., [10, p. 296]), we have

$$\|\mathcal{K}_\sigma \phi\|_{L^2(V)} \le \|k_\sigma * \phi^e\|_{L^2(\mathbb{R}^3)} \le \|k_\sigma\|_{L^1(\mathbb{R}^3)} \|\phi^e\|_{L^2(\mathbb{R}^3)} = \|k_\sigma\|_{L^1(\mathbb{R}^3)} \|\phi\|_{L^2(V)}.$$

Now, using spherical coordinates,

$$\|k_\sigma\|_{L^1(\mathbb{R}^3)} = \int_0^\infty \int_{\mathbb{S}^2} \frac{1}{4\pi} \frac{\exp(-\sigma \|s\mathbf{\Omega}\|_2)}{\|s\mathbf{\Omega}\|_2^2} s^2 \, \mathrm{d}\mathbf{\Omega} \, \mathrm{d}s = \int_0^\infty \exp(-\sigma s) \, \mathrm{d}s = \frac{1}{\sigma},$$

and hence $\|\mathcal{K}_\sigma\|_{\mathscr{L}(L^2(V))} \le \sigma^{-1}$. ∎

Now, by the spectral theorem for self-adjoint compact operators (see, e.g., [11, section 3.3]) and Lemma 2.5 for $\omega_j > 0$, $\mathcal{K}_\sigma$ has a sequence of eigenpairs $\{(\omega_j, e_j)\}_{j=1}^{\infty}$, where the sequence $\{\omega_j\}$ is positive and monotone nonincreasing and converges to zero as $j \to \infty$, and $\{e_j\}$ is a complete orthonormal sequence in $L^2(V)$. Moreover, from Lemma 2.5 and the fact that $\sigma = \sigma_c + \sigma_s + \sigma_f$ (and all cross-sections are positive), we have

$$\sigma_s \omega_j \leq \sigma_s \sigma^{-1} < 1.$$

Combining this with Corollary 2.4 gives the following.

LEMMA 2.6. *The eigenvalues λ in Corollary 2.4 are*

$$\lambda_j = \frac{1 - \sigma_s \omega_j}{\nu \sigma_f \omega_j}. \tag{2.10}$$

The sequence $\{\lambda_j\}$ is positive and nondecreasing and tends to infinity as $j \to \infty$.

Also crucial to the physical meaning of the eigenvalue problem (1.1) is the fact that λ_1 (the eigenvalue of physical interest) is simple, which we assume from now on. This can be proved by an application of the Krein–Rutman theorem, but we do not pursue this further here (see [18] for a classical reference on this topic).

In the following section we will be interested in convergence of iterative methods for finding λ_1. Exploiting the complete orthonormal sequence $\{e_j\}$, for any $\phi \in L^2(V)$, we write $\phi = \sum_{j=1}^{\infty} \xi_j(\phi) e_j$, where $\xi_j(\phi) = (\phi, e_j)_{L^2(V)}$, and

$$\|\phi\|_{L^2(V)}^2 = \sum_{j=1}^{\infty} |\xi_j(\phi)|^2 = c(\phi)^2 + s(\phi)^2, \tag{2.11}$$

where $c(\phi) = |\xi_1(\phi)|$ and $s(\phi)^2 = \sum_{j=2}^{\infty} |\xi_j(\phi)|^2$. Then the proximity of a normalized ϕ to e_1 may be characterised by how close the "tangent" $t(\phi) := s(\phi)/c(\phi)$ is to zero. We use this orthogonal decomposition as a tool for obtaining estimates for the rate of convergence of inexact inverse iteration algorithms in the following section. The procedure is analogous to that in [4] (see also [6, 8] for more sophisticated applications). While these references considered the matrix generalized eigenvalue problem, a novel feature of our analysis here is that we apply analogous arguments adapted to the infinite-dimensional generalized operator eigenvalue problem (2.9).

Remark 2.7. Before leaving this section, we remark that an analogous analysis can be obtained for the integral operator forms of the 2D and 1D model problems introduced in section 1. For the 2D case the reduced problem is

$$\phi(\widetilde{\mathbf{r}}) - \sigma_s \mathcal{K}_\sigma \phi(\widetilde{\mathbf{r}}) = \lambda \nu \sigma_f \mathcal{K}_\sigma \phi(\widetilde{\mathbf{r}}), \quad \text{where} \quad \phi(\widetilde{\mathbf{r}}) := \frac{1}{2\pi} \int_{\mathbb{S}^1} \Psi(\widetilde{\mathbf{r}}, \widetilde{\mathbf{\Omega}}') \, \mathrm{d}\widetilde{\mathbf{\Omega}}',$$

$$(\mathcal{K}_\sigma g)(\widetilde{\mathbf{r}}) := \int_V k_\sigma(\widetilde{\mathbf{r}} - \widetilde{\mathbf{r}}') g(\widetilde{\mathbf{r}}') \, \mathrm{d}\widetilde{\mathbf{r}}', \quad \text{and} \quad k_\sigma(\mathbf{x}) := \frac{1}{2\pi} \frac{\exp(-\sigma \|\mathbf{x}\|_2)}{\|\mathbf{x}\|_2}, \quad \mathbf{x} \in \mathbb{R}^2.$$

For the 1D problem the equivalent to (2.9) is

$$\phi(z) - \sigma_s \mathcal{K}_\sigma \phi(z) = \lambda \nu \sigma_f \mathcal{K}_\sigma \phi(z), \quad \text{where} \quad \phi(z) := \frac{1}{2} \int_{-1}^{1} \Psi(z, \mu') \, \mathrm{d}\mu',$$

$$(\mathcal{K}_\sigma g)(z) := \int_0^1 k_\sigma(z - z') g(z') \, \mathrm{d}z', \quad \text{and} \quad k_\sigma(x) := \frac{1}{2} \int_0^1 \exp\left(\frac{-\sigma|x|}{\mu}\right) \frac{\mathrm{d}\mu}{\mu}.$$

3. Iterative methods for reactor criticality. In Algorithm 1 we present inexact inverse iteration for (1.3). When approximately solving the linear system for the next iterate (step (†)), we measure the residual using the following scalar quantity. For $v \in L^2(V, L^\infty(\mathbb{S}^2))$ we set

$$\|v\|_* = \|\mathcal{P}\mathcal{T}^{-1}v\|_{L^2(V)}, \tag{3.1}$$

which is well defined by Lemma 2.1. Moreover, Lemma 2.2 tells us that if $v(\mathbf{r}, \mathbf{\Omega}) = v(\mathbf{r})$, for all $(\mathbf{r}, \mathbf{\Omega}) \in V \times \mathbb{S}^2$, we have $\|v\|_* = \|\mathcal{K}_\sigma v\|_{L^2(V)}$, so that $\|\cdot\|_*$ acts as a norm on the subspace of all functions in $L^2(V, L^\infty(\mathbb{S}^2))$ which are constant with respect to their second argument.

ALGORITHM 1. Inexact inverse iteration with shift

Require: Starting guess $\Psi^{(0)}$.
for $i = 0, 1, 2, \ldots$ **do**
Choose a shift $\alpha^{(i)}$ and an inner tolerance $\tau^{(i)} \geq 0$.
Compute $\tilde{\Psi}^{(i+1)}$ so that $\|(\mathcal{T} - \mathcal{S} - \alpha^{(i)}\mathcal{F})\tilde{\Psi}^{(i+1)} - \mathcal{F}\Psi^{(i)}\|_* \leq \tau^{(i)}$. (†)
Normalize $\Psi^{(i+1)} = \tilde{\Psi}^{(i+1)}/\|\mathcal{P}\tilde{\Psi}^{(i+1)}\|_{L^2(V)}$.
end for

In this algorithm we implicitly require $\Psi^{(i)}, \tilde{\Psi}^{(i)} \in L^2(V, L^1(\mathbb{S}^2))$. We typically stop the algorithm if the eigenvalue residual

$$\mathrm{res}^{(i)} := (\mathcal{T} - \mathcal{S} - \rho^{(i)}\mathcal{F})\Psi^{(i)}$$

is sufficiently small in some norm, where $\rho^{(i)}$ is a suitable eigenvalue approximation (e.g., a Rayleigh quotient) derived from $\Psi^{(i)}$, provided $\Psi^{(i)}$ is rich enough in a certain eigendirection. We discuss a particular choice of $\rho^{(i)}$ below. A simple application of Lemma 2.2 proves the following result.

LEMMA 3.1. *If $\tilde{\Psi}^{(i)}$ and $\Psi^{(i)}$ are computed by Algorithm* 1*, and if we introduce the corresponding scalar fluxes $\tilde{\phi}^{(i)} := \mathcal{P}\tilde{\Psi}^{(i)}$ and $\phi^{(i)} := \mathcal{P}\Psi^{(i)}$, then*

$$\|(I - (\sigma_s + \alpha^{(i)}\nu\sigma_f)\mathcal{K}_\sigma)\tilde{\phi}^{(i+1)} - \nu\sigma_f\mathcal{K}_\sigma\phi^{(i)}\|_{L^2(V)} \leq \tau^{(i)} \tag{3.2}$$

and

$$\phi^{(i+1)} = \frac{\tilde{\phi}^{(i+1)}}{\|\tilde{\phi}^{(i+1)}\|_{L^2(V)}}. \tag{3.3}$$

Thus, when $\Psi^{(i)}$ is close to an eigenvector corresponding to the minimal eigenvalue of (1.1), then $\phi^{(i)}$ is predominantly in the direction e_1 and $t(\phi^{(i)})$ will be close to 0. The following theorem gives a mechanism for bounding $t(\phi^{(i+1)})$. This theorem will be used in Corollaries 3.4 and 3.5 to obtain the convergence properties of several variants of Algorithm 1. For convenience we will discuss an abstract version of (3.2), (3.3) where the superscripts are suppressed.

THEOREM 3.2. *Suppose $s(\phi) \neq 0$ and*

$$\|(I - (\sigma_s + \alpha\,\nu\sigma_f)\mathcal{K}_\sigma)\tilde{\phi} - \nu\sigma_f\mathcal{K}_\sigma\phi\|_{L^2(V)} \leq \tau, \tag{3.4}$$

and set

$$\phi' = \frac{\tilde{\phi}}{\|\tilde{\phi}\|_{L^2(V)}}. \tag{3.5}$$

Then, if $\tau < \nu\sigma_f\omega_1 c(\phi)$, *we have with constant* $C_1 = 1/(\nu\sigma_f\omega_2)$

$$t(\phi') \le \left(\frac{s(\phi) + C_1\,\tau}{c(\phi) - C_1\,\tau}\right)\left|\frac{\lambda_1 - \alpha}{\lambda_2 - \alpha}\right|. \tag{3.6}$$

Proof. To make the notation simpler, without loss of generality we set $\nu = 1$ in the proof. First observe that if $\tilde{\phi} = 0$ in (3.4), then, since $s(\phi) \neq 0$, we have

$$\tau \ge \sigma_f \|\mathcal{K}_\sigma \phi\|_{L^2(V)} = \sigma_f \left\{\sum_{j=1}^{\infty} \omega_j^2 |\xi_j(\phi)|^2\right\}^{1/2} > \sigma_f \omega_1 c(\phi),$$

which contradicts the assumption. So $\tilde{\phi} \neq 0$ and the normalization (3.5) is well defined.

To obtain the bound on $t(\phi')$, set $R := (I - (\sigma_s + \alpha\sigma_f)\mathcal{K}_\sigma)\tilde{\phi} - \sigma_f \mathcal{K}_\sigma \phi$. Because the (ω_j, e_j) are eigenpairs of $\mathcal{K}_\sigma$, we have (using (3.5) and (2.10)), for all $j \ge 1$,

$$\begin{aligned} \xi_j(R) &= (1 - (\sigma_s + \alpha\sigma_f)\omega_j)\xi_j(\tilde{\phi}) - \sigma_f\omega_j\xi_j(\phi) \\ &= \sigma_f\omega_j \left[\|\tilde{\phi}\|_{L^2(V)}(\lambda_j - \alpha)\xi_j(\phi') - \xi_j(\phi)\right]. \end{aligned} \tag{3.7}$$

Now, using (2.11) and (3.4), we have

$$\tau \ge \|R\|_{L^2(V)} \ge |\xi_1(R)| \ge \sigma_f\omega_1 \left[c(\phi) - \|\tilde{\phi}\|_{L^2(V)}\,|\lambda_1 - \alpha|\,c(\phi')\right],$$

and a rearrangement of this yields

$$\frac{1}{c(\phi')} \le \left(\frac{\sigma_f\omega_1}{\sigma_f\omega_1 c(\phi) - \tau}\right)\,|\lambda_1 - \alpha|\,\|\tilde{\phi}\|_{L^2(V)}. \tag{3.8}$$

On the other hand, rearranging (3.7) gives for $j \ge 2$

$$\xi_j(\phi')\|\tilde{\phi}\|_{L^2(V)} = \left(\frac{1}{\lambda_j - \alpha}\right)\left[\frac{\xi_j(R)}{\sigma_f\omega_j} + \xi_j(\phi)\right]. \tag{3.9}$$

Now recall that λ_j increases and that (via (2.10) with $\nu = 1$)

$$\sigma_f(\lambda_j - \alpha)\omega_j = 1 - (\sigma_s + \sigma_f\alpha)\omega_j,$$

which increases as well. Hence, squaring (3.9), summing over $j = 2, \ldots, \infty$, and recalling (3.4), we obtain

$$s(\phi')\|\tilde{\phi}\|_{L^2(V)} \le \frac{1}{|\lambda_2 - \alpha|}\left(\frac{\tau}{\sigma_f\omega_2} + s(\phi)\right). \tag{3.10}$$

Finally, by rearranging the product of (3.8) and (3.10) and using the definition of C_1, we obtain the result. ∎

The estimate (3.6) contains a great deal of information about the convergence of Algorithm 1. For example, if $\alpha^{(i)}$ converges quadratically to λ_1, then, with a fixed choice of $\tau^{(i)} = \tau_0$ (satisfying the assumption of Theorem 3.2), the algorithm will converge quadratically. A possible candidate for $\alpha^{(i)}$ is given in the following lemma.

LEMMA 3.3. *Given $\Psi^{(i)}$, consider the "scalar flux Rayleigh quotient"*

$$\widetilde{\rho}^{(i)} := \frac{(\mathcal{P}\Psi^{(i)}, \mathcal{P}\mathcal{T}^{-1}(\mathcal{T}-\mathcal{S})\Psi^{(i)})_{L^2(V)}}{(\mathcal{P}\Psi^{(i)}, \mathcal{P}\mathcal{T}^{-1}\mathcal{F}\Psi^{(i)})_{L^2(V)}} = \frac{(\phi^{(i)}, (I-\sigma_s\mathcal{K}_\sigma)\phi^{(i)})_{L^2(V)}}{(\phi^{(i)}, \nu\sigma_f\mathcal{K}_\sigma\phi^{(i)})_{L^2(V)}}. \tag{3.11}$$

This enjoys the estimate

$$|\widetilde{\rho}^{(i)} - \lambda_1| = \mathcal{O}(s(\phi^{(i)})^2).$$

We call (3.11) *the scalar flux Rayleigh quotient because it uses the formulation of the eigenvalue problem* (2.9) *for the scalar flux.*

Proof. By writing $\phi^{(i)} = c(\phi^{(i)})e_1 + s(\phi^{(i)})u^{(i)}$, where $\|u^{(i)}\|_{L^2(V)} = 1$ and $(u^{(i)}, e_1)_{L^2(V)} = 0$, and using (2.10), we see that

$$\widetilde{\rho}^{(i)} = \frac{(1-\sigma_s\omega_1)\, c(\phi^{(i)})^2 + \mathcal{O}(s(\phi^{(i)})^2)}{\nu\sigma_f\omega_1\, c(\phi^{(i)})^2 + \mathcal{O}(s(\phi^{(i)})^2)} = \lambda_1 + \mathcal{O}(s(\phi^{(i)})^2). \qquad \square$$

Using this lemma and Theorem 3.2, we now obtain the following convergence rate for Algorithm 1.

COROLLARY 3.4. *Suppose that for every step in Algorithm* 1 *the conditions of Theorem* 3.2 *are satisfied and the shift $\alpha^{(i)} = \widetilde{\rho}^{(i)}$ (Rayleigh quotient iteration) is applied. Then*

$$t(\phi^{(i+1)}) \le \left(\frac{s(\phi^{(i)}) + C_1\,\tau^{(i)}}{c(\phi^{(i)}) - C_1\,\tau^{(i)}}\right)\left|\frac{C_2}{\lambda_2-\lambda_1}\right|\; t(\phi^{(i)})^2, \quad C_2 \text{ constant.}$$

Hence Algorithm 1 *converges quadratically. The convergence rate is even cubic if the tolerances decrease with rate*

$$\tau^{(i)} \le C_3 s(\phi^{(i)}), \quad C_3 \text{ constant.} \tag{3.12}$$

On the other hand, using (3.6) for a fixed shift and decreasing tolerances, we get the following result.

COROLLARY 3.5. *If in every iteration of Algorithm* 1 *the conditions of Theorem* 3.2 *are met and if fixed shifts $\alpha^{(i)} = \alpha_0$, as well as tolerances satisfying* (3.12) *are used, then for small enough C_3 in* (3.12)

$$t(\phi^{(i+1)}) \le \left(\frac{1 + C_1C_3}{1 - C_1C_3t(\phi^{(i)})}\right)\left|\frac{\lambda_1-\alpha_0}{\lambda_2-\alpha_0}\right|\; t(\phi^{(i)}).$$

Hence, provided the shift α_0 is close enough to λ_1, we obtain linear convergence of the algorithm.

Note that this analysis gives no guarantee that Algorithm 1 converges when we use a fixed shift and constant tolerances. In the final section of the paper we investigate this question numerically.

This type of analysis will also extend to other iterative methods such as Jacobi–Davidson (e.g., as is done in a different context in [6, 8]). The analysis here is given only for the continuous problem (1.1), but it provides a guide to how iterations will behave in discrete cases, as we see in the final section, where we investigate two 1D model problems of different complexities.

4. Symmetry under discretization. Retaining the underlying symmetry of the scalar flux problem (2.9) in a discretization is a delicate matter. Applying, for example, a standard discrete ordinates approach to the 1D problem (1.5) using Gauss–Legendre quadrature for the angular variable and a Crank–Nicolson scheme to approximate the spatial derivative, as discussed in [20], does not preserve the underlying symmetry in the discretization [22]. As a result of this, the discrete equivalent of the scalar flux Rayleigh quotient $\widetilde{\rho}^{(i)}$ in (3.11) does not approximate the desired eigenvalue up to second order, as is proved for the continuous problem in Lemma 3.3. Hence inexact inverse iteration loses an order in the convergence rate for such a shift strategy, as we will see in the numerical results in the next section.

While a full study of symmetry-preserving discretizations is beyond the scope of this paper, we show here (by several examples) that natural symmetry-preserving discretizations do exist. First we consider the semidiscrete case of (1.5) and (1.6), where we discretize only with respect to the spatial variable z (and leave the angular variable μ continuous). The discrete approximation to the operator $\mathcal{K}_\sigma = \mathcal{P}\mathcal{T}^{-1}$ is obtained by applying the inverse of the discrete version of $\mathcal{T}$ and then integrating over μ. This turns out to be symmetric in the discrete spatial variable when certain conditions are met. In addition we describe how to preserve the symmetry under further discretization with respect to the angular variable μ. In both examples below, analogously to Lemma 2.1, we consider for any $g \in L^2[0,1]$ discrete versions of the problem

$$\mathcal{T}\Psi := \mu\frac{\partial}{\partial z}\Psi(z,\mu) + \sigma\Psi(z,\mu) = g(z), \quad z \in [0,1], \quad \mu \in [-1,1], \tag{4.1}$$

subject to vacuum boundary conditions (1.6). Different questions related to symmetric forms of the transport equation are studied in [5, 19].

4.1. Finite difference methods on uniform meshes. Here we introduce a uniform spatial mesh $z_j := jh$, $j = 0, \dots, M$ with $h = 1/M$, and an Euler-type method (i.e., a first-order finite difference approximation of the derivative), integrating from left to right for $\mu > 0$ and from right to left for $\mu < 0$, i.e.,

$$\begin{aligned}
&\mu\,\frac{\Psi(z_j,\mu) - \Psi(z_{j-1},\mu)}{h} + \sigma\Psi(z_{j-1},\mu) = g(z_{j-1}) \quad \text{for} \quad \mu > 0,\ j = 1,\dots,M,\\
&\mu\,\frac{\Psi(z_j,\mu) - \Psi(z_{j-1},\mu)}{h} + \sigma\Psi(z_j,\mu) = g(z_j) \quad \text{for} \quad \mu < 0,\ j = 1,\dots,M,\\
&\text{with} \quad \Psi(z_0,\mu) = 0 \quad \text{when} \quad \mu > 0 \quad \text{and} \quad \Psi(z_M,\mu) = 0 \quad \text{when} \quad \mu < 0.
\end{aligned} \tag{4.2}$$

These equations can be written as the two linear systems

$$A^+(\mu)\mathbf{\Psi}^+(\mu) = \mathbf{g}^+ \quad \text{for} \quad \mu > 0 \quad \text{and} \quad A^-(\mu)\mathbf{\Psi}^-(\mu) = \mathbf{g}^- \quad \text{for} \quad \mu < 0,$$

where $\mathbf{\Psi}^+(\mu) = (\Psi(z_1,\mu),\dots,\Psi(z_M,\mu))^T$, $\mathbf{g}^+ = (g(z_0),\dots,g(z_{M-1}))^T$, $\mathbf{\Psi}^-(\mu) = (\Psi(z_0,\mu),\dots,\Psi(z_{M-1},\mu))^T$, and $\mathbf{g}^- = (g(z_1),\dots,g(z_M))^T$, and where $A^+(\mu)$ is the lower bidiagonal matrix and $A^-(\mu)$ is the upper bidiagonal matrix given, respectively, by

$$A^+(\mu) = \begin{bmatrix} \frac{\mu}{h} & & & \\ \left(\sigma - \frac{\mu}{h}\right) & \frac{\mu}{h} & & \\ & \ddots & \ddots & \\ & & \left(\sigma - \frac{\mu}{h}\right) & \frac{\mu}{h} \end{bmatrix}, \quad A^-(\mu) = \begin{bmatrix} -\frac{\mu}{h} & \left(\sigma + \frac{\mu}{h}\right) & & \\ & \ddots & \ddots & \\ & & -\frac{\mu}{h} & \left(\sigma + \frac{\mu}{h}\right) \\ & & & -\frac{\mu}{h} \end{bmatrix}.$$

Note that $A^-(\mu)$ and $A^+(\mu)$ are both nonsingular and that

$$A^-(-\mu) = A^+(\mu)^T \quad \text{and so} \quad A^-(-\mu)^{-1} = A^+(\mu)^{-T}. \tag{4.3}$$

This condition plays a crucial role in our proof that this difference scheme retains the underlying symmetry. With $\mathbf{\Psi}(\mu) = (\Psi(z_0,\mu),\dots,\Psi(z_M,\mu))^T$ and by reducing the problem now to the space of scalar fluxes, we have for $\boldsymbol{\phi} = (\phi(z_0),\dots,\phi(z_M))^T$ and $\mathbf{g} = (g(z_0),\dots,g(z_M))^T \in \mathbb{R}^{M+1}$

$$\begin{aligned}\boldsymbol{\phi} &= \int_{-1}^{1} \mathbf{\Psi}(\mu)\,\mathrm{d}\mu = \int_0^1 \begin{pmatrix} \mathbf{0} & (A^-(-\mu))^{-1} \\ 0 & \mathbf{0}^T \end{pmatrix} \mathbf{g}\,\mathrm{d}\mu + \int_0^1 \begin{pmatrix} \mathbf{0}^T & 0 \\ (A^+(\mu))^{-1} & \mathbf{0} \end{pmatrix} \mathbf{g}\,\mathrm{d}\mu \\ &= \widetilde{K}\mathbf{g}, \quad \text{where} \quad \widetilde{K} := \left[\int_0^1 A(\mu)\,\mathrm{d}\mu\right],\end{aligned}$$

and

$$A(\mu) = \begin{pmatrix} \mathbf{0} & (A^-(-\mu))^{-1} \\ 0 & \mathbf{0}^T \end{pmatrix} + \begin{pmatrix} \mathbf{0}^T & 0 \\ (A^+(\mu))^{-1} & \mathbf{0} \end{pmatrix}.$$

The symmetry of $\widetilde{K}$ then follows easily from (4.3).

Finally, let us now consider the full discretization. Suppose that we choose a quadrature rule with points $\{\mu_k\} \subset [-1,1]\backslash\{0\}$ and weights $\{w_k\}$ indexed from $k = -N,\dots,-1$ and $k = 1,\dots,N$ with the property

$$\mu_{-k} = -\mu_k \quad \text{and} \quad w_{-k} = w_k \quad \text{for all} \quad k = 1,\dots,N.$$

(An example would be the $2N$ point Gauss–Legendre rule.) Then we obtain $\boldsymbol{\phi} = K\mathbf{g}$ with a symmetric $K = \sum_{k=1}^N w_k A(\mu_k)$.

4.2. Finite element methods. As an alternative to the finite difference methods of the previous section, consider an arbitrary mesh $0 = z_0 < z_1 < \cdots < z_M = 1$. For $i = 0,\dots,M$, let φ_i denote the usual piecewise linear "hat" function on $[0,1]$, satisfying $\varphi_i(z_j) = \delta_{i,j}$, set $h_j = z_j - z_{j-1}$, and define $(\cdot,\cdot)$ to be the $L^2([0,1])$ inner product.

Now use the following approximation for (4.1). If $\mu > 0$, we approximate $\Psi(z,\mu)$ by $\Psi^+(z,\mu) := \sum_{j=1}^M \Psi_j^+(\mu)\varphi_j(z)$ and determine the coefficients $\Psi_j^+(\mu)$ by requiring

$$\left(\mu\frac{\partial}{\partial z}\Psi^+(\cdot,\mu) + \sigma\Psi^+(\cdot,\mu), \varphi_{i-1}\right) = (g,\varphi_{i-1}), \quad i = 1,\dots,M.$$

This is easily seen to be equivalent to the $M \times M$ system

$$A^+(\mu)\mathbf{\Psi}^+(\mu) = \mathbf{g}^+,$$

where $\mathbf{\Psi}^+(\mu) = (\Psi_1^+(\mu),\dots,\Psi_M^+(\mu))^T$, $\mathbf{g}^+ = ((g,\varphi_0),\dots,(g,\varphi_{M-1}))^T$, and

$$(A^+(\mu))_{i,j} = \mu(\varphi_j',\varphi_{i-1}) + \sigma(\varphi_j,\varphi_{i-1}).$$

Similarly, for $\mu < 0$ approximate $\Psi(z,\mu)$ by $\Psi^-(z,\mu) := \sum_{j=0}^{M-1} \Psi_j^-(\mu)\varphi_j(z)$ defined by

$$\left(\mu\frac{\partial}{\partial z}\Psi^-(\cdot,\mu) + \sigma\Psi^-(\cdot,\mu), \varphi_i\right) = (g,\varphi_i), \quad i = 1,\dots,M.$$

This is equivalent to

$$A^-(\mu)\mathbf{\Psi}^-(\mu) = \mathbf{g}^-,$$

where $\mathbf{\Psi}^-(\mu) = (\Psi_0^-(\mu), \ldots, \Psi_{M-1}^-(\mu))^T$, $\mathbf{g}^- = ((g, \varphi_1), \ldots, (g, \varphi_M))^T$, and

$$(A^-(\mu))_{i,j} = \mu(\varphi'_{j-1}, \varphi_i) + \sigma(\varphi_{j-1}, \varphi_i).$$

The matrices $A^+(\mu)$ are lower tridiagonal with positive diagonal, while the matrices $A^-(\mu)$ are upper tridiagonal with positive diagonal, and hence $A^{\pm}(\mu)$ are nonsingular. Now we notice that by integration by parts $(\varphi'_{i-1}, \varphi_j) = -(\varphi'_j, \varphi_{i-1})$, and so the crucial condition (4.3) is also satisfied by the matrices $A^{\pm}(\mu)$. Therefore, these finite element methods also have the property that the symmetry of the discrete version of the operator $\mathcal{P}\mathcal{T}^{-1}$ is conserved.

5. Numerical results. We now consider numerical results for two 1D model problems of different complexities to illustrate the theory.

5.1. Los Alamos benchmark test set problem. This model problem is taken from a collection of benchmark tests produced at Los Alamos National Laboratory [23]. Problem number 2 of the test set corresponds to the 1D problem (1.5), (1.6). Specifications of the problem are given in Table 5.1.

TABLE 5.1
Data for the problem from the Los Alamos benchmark test set (problem number 2).

σ	σ_s	σ_f	ν
0.32640	0.225216	0.081600	3.24
Slab length: L = 3.707444cm			

For our discretization we approximate the integrals in (1.5) by Gauss quadrature with an even number of quadrature points on $[-1, 1]$. The spatial discretization (with respect to z in (1.5)) is done by the upwind Euler scheme discussed in section 4.1, which preserves symmetry under reduction to the scalar flux. We also apply a Crank–Nicolson scheme for the spatial approximation, and further details of this, together with bounds on the discretization error, can be found in [20].

We use 128 equally sized spatial intervals and 128 angular Gauss points, leading to a nonsymmetric generalized matrix eigenvalue problem of dimension 16384×16384. The eigenvalues nearest zero of the discrete problems are $\lambda_1^{\mathrm{Eul}} \approx 0.99570$, $\lambda_2^{\mathrm{Eul}} \approx 2.60907$ and $\lambda_1^{\mathrm{CN}} \approx 1.00003$, $\lambda_2^{\mathrm{CN}} \approx 2.60530$. Our stopping criterion for the outer iteration is $\|\mathbf{res}^{(i)}\|_2 < 10^{-14}$, where $\mathbf{res}^{(i)} := (T - S - \rho^{(i)}F)\mathbf{\Psi}^{(i)}$ with T, S, F, and $\mathbf{\Psi}$ being the discrete versions of $\mathcal{T}$, $\mathcal{S}$, $\mathcal{F}$, and Ψ, respectively, and where $\rho^{(i)} = \rho(\mathbf{\Psi}^{(i)})$ and

$$\rho(\mathbf{\Psi}) := \frac{(\mathbf{\Psi}, (T - S)\mathbf{\Psi})}{(\mathbf{\Psi}, F\mathbf{\Psi})} \tag{5.1}$$

is here called the "angular flux Rayleigh quotient," and is distinct from the scalar flux Rayleigh quotient in (3.11), where $(\cdot, \cdot)$ represents the ℓ_2 inner product over all spatial and angular discrete variables. Note that we compute this eigenproblem residual $\mathbf{res}^{(i)}$ in the full spatially and angular dependent space.

Problem (†) in Algorithm 1 is solved using the GMRES function in MATLAB 2009b with an LU factorization of T as preconditioner which proves essential for

ensuring convergence of GMRES. As starting guess $\mathbf{\Psi}^{(0)}$ we use a normalized vector with equal positive entries. To measure the convergence rate of Algorithm 1, we consider the eigenvalue error $\Delta^{(i)} := |\lambda_1 - \rho^{(i)}|$, where λ_1 is the computed eigenvalue when the iteration terminates.

Table 5.2 shows the numerical results for fixed shifts $\alpha_0 = 0.9$ and $\alpha_0 = 0.99$. We used decreasing tolerances $\tau^{(i)} \leq 0.1 \, \|PT^{-1} \, \mathbf{res}^{(i)}\|_2$ for the inner solves, where P denotes the discrete version of the projection operator $\mathcal{P}$. The results clearly show linear and not quadratic convergence in both cases with a faster linear rate when $\alpha = 0.99$, agreeing with Corollary 3.5.

TABLE 5.2

Numerical results for Algorithm 1 *with decreasing tolerances* $\tau^{(i)} \leq 0.1 \, \|PT^{-1} \, \mathbf{res}^{(i)}\|_2$ *when using the symmetry-preserving upwind scheme.*

	$\alpha_0 = 0.9$			$\alpha_0 = 0.99$		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	3.3E-02			3.3E-02		
1	2.6E-04	8.0E-03	2.5E-01	1.4E-05	4.2E-04	1.3E-02
2	5.4E-06	2.0E-02	7.7E+01	2.0E-08	1.4E-03	1.0E+02
3	1.3E-07	2.4E-02	4.6E+03	3.0E-11	1.5E-03	7.7E+04
4	3.3E-09	2.6E-02	2.0E+05	4.5E-14	1.5E-03	5.1E+07
5	8.6E-11	2.6E-02	7.7E+06	0.0E+00	0.0E+00	0.0E+00
6	2.2E-12	2.6E-02	3.0E+08			
7	5.8E-14	2.6E-02	1.2E+10			
8	0.0E+00	0.0E+00	0.0E+00			

When replacing the symmetry-preserving Euler scheme with the Crank–Nicolson discretization, which does not preserve the symmetry of the reduction, we obtain results very similar to those in Table 5.2, suggesting that the convergence for fixed shift iteration is not influenced by retaining the underlying symmetry if the tolerances decrease sufficiently fast and the fixed shift is close enough to the desired eigenvalue. Such convergence when using a similar inexact inverse iteration method for the non-symmetric matrix eigenvalue problem is discussed in [7].

Surprisingly, even for a fixed shift $\alpha_0 = 0.3$ and a fixed inner tolerance $\tau_0 = 0.1$, we still obtained linear convergence. This appears to be due to the fact that for sufficiently large i, GMRES is observed to converge after one iteration and the accuracy of the GMRES solves for the inner systems (†) increases. This then results in a (slowly) decreasing (effective) inner tolerance, leading to linear convergence of the method.

Tables 5.3 and 5.4 concern the variable shift case, comparing the convergence for $\alpha^{(i)} = \rho^{(i)}$, the angular flux Rayleigh quotient in (5.1), and $\alpha^{(i)} = \widetilde{\rho}^{(i)}$, the scalar flux Rayleigh quotient from (3.11).

In Table 5.3 we obtain only linear convergence for the symmetry-preserving Euler scheme and fixed inner tolerances when using the angular flux Rayleigh quotient ρ as shift, but the numerical results suggest quadratic convergence for the scalar flux Rayleigh quotient $\widetilde{\rho}$. This agrees with our theory, and so we recommend the use of the scalar flux Rayleigh quotient $\widetilde{\rho}$ as shift.

When applying the Crank–Nicolson scheme for the spatial approximation, the underlying symmetry gets lost in the discretization and neither of the variable shifts achieves quadratic convergence. We used twice as many angular and spatial discretization points than in Table 5.3 to produce Table 5.4. In this case the convergence rates are clearer to establish from the numerical results. Both shifts give only linear convergence, emphasizing the benefits of using a symmetry-preserving discretization.

TABLE 5.3
Results for Algorithm 1 *with constant tolerance* $\tau_0 = 0.1$ *and two different Rayleigh quotient shifts for matrices arising from the application of the Euler scheme described in section* 4.1.

	$\alpha^{(i)} = \rho^{(i)}$			$\alpha^{(i)} = \widetilde{\rho}^{(i)}$		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	3.3E-02			3.3E-02		
1	9.7E-05	3.0E-03	9.1E-02	9.4E-05	2.9E-03	8.8E-02
2	2.4E-08	2.4E-04	2.5E+00	2.4E-11	2.5E-07	2.7E-03
3	1.6E-11	6.8E-04	2.9E+04	0.0E+00	0.0E+00	0.0E+00
4	1.1E-15	7.0E-05	4.4E+06			
5	0.0E+00	0.0E+00	0.0E+00			

TABLE 5.4
Numerical results for Algorithm 1 *with constant tolerance* $\tau_0 = 0.1$ *and Rayleigh quotient shifts using a Crank–Nicolson scheme for the spatial approximation.*

	$\alpha^{(i)} = \rho^{(i)}$			$\alpha^{(i)} = \widetilde{\rho}^{(i)}$		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	3.3E-02			3.3E-02		
1	7.8E-05	2.4E-03	7.3E-02	7.4E-05	2.3E-03	6.9E-02
2	9.7E-09	1.3E-04	1.6E+00	8.3E-10	1.1E-05	1.5E-01
3	3.1E-12	3.1E-04	3.2E+04	1.3E-14	1.6E-05	1.9E+04
4	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00

However, using the scalar flux Rayleigh quotient $\widetilde{\rho}$ in the nonsymmetric case is not disadvantageous but actually leads to slightly faster (but still linear) convergence.

Due to reaching machine precision so quickly, we were not able to clearly establish the predicted cubic convergence for a Rayleigh quotient shift and decreasing tolerances when using the symmetric Euler discretization and our scalar flux Rayleigh quotient shift $\widetilde{\rho}$. When applying decreasing tolerances to the other three variable shift cases that we considered in Tables 5.3 and 5.4, the numerical results suggest the gain of an additional order in the convergence rate, leading to quadratic convergence for those problems, as Table 5.5 indicates. One of the future tasks could be to redo these calculations using variable precision arithmetic.

TABLE 5.5
Numerical results as in Table 5.4 *but with decreasing tolerances* $\tau^{(i)} \leq 0.1 \, \|PT^{-1} \, \mathbf{res}^{(i)}\|_2$.

	$\alpha^{(i)} = \rho^{(i)}$			$\alpha^{(i)} = \widetilde{\rho}^{(i)}$		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	3.3E-02			3.3E-02		
1	8.0E-05	2.4E-03	7.2E-02	7.2E-05	2.2E-03	6.5E-02
2	1.7E-09	2.1E-05	2.7E-01	1.3E-10	1.8E-06	2.5E-02
3	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00

The following numerical results are now for a more realistic problem with two different material regions and neutrons of two energy levels.

5.2. Control rod insertion model problem. This model problem describes the core of a nuclear reactor for different insertion depths of a control rod. Identical cells are usually arranged in a lattice structure or in rings surrounding a central pin. The resulting geometrical symmetry can be exploited by modeling only half of

a cell and enforcing reflective boundary conditions (as indicated in Figure 5.1). In one dimension, with $V = [0, L]$ and E denoting the energy, the reflective boundary conditions are

$$\Psi(0, E, \mu) = \Psi(0, E, -\mu) \quad \text{when} \quad \mu > 0 \quad \text{and}$$
$$\Psi(L, E, \mu) = \Psi(L, E, -\mu) \quad \text{when} \quad \mu < 0.$$

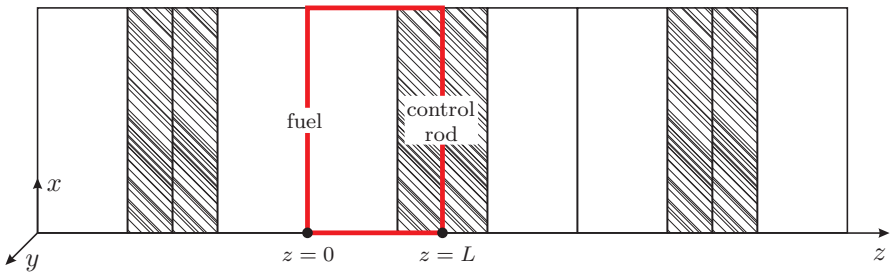

FIG. 5.1. *Criticality computations on large lattice structures can be done approximately by modeling a part of them (highlighted) and using reflective boundary conditions at the sides.*

We model the problem as a slab reactor (constant flux in the x and y dimensions) with two regions in the z direction, the fuel and the absorber, as shown in Figure 5.1. The latter region consists of a homogenized, nonfissile mix of control rod material and remaining water if the rod is not fully inserted. Depending on the insertion depth of the control rod, the material properties in the absorber region change. Within each region we assume the material cross-sections to be constant.

The energy spectrum of this model problem is constrained to neutrons of high and low energies (denoted by subscripts h and l), with angular fluxes Ψ_h and Ψ_l, which are linked by the fission and scatter operators. The latter now includes, in addition to self-scatter within the same energy groups, scatter from high to low energies ($\sigma_{s,h\to l}$), and vice versa ($\sigma_{s,l\to h}$). It is assumed that all fission product neutrons are of high energy, i.e., $(\chi_h, \chi_l) = (1, 0)$. The problem is then analogous to (1.5), but with two energy groups and spatially dependent cross-sections, and takes the form

$$\begin{aligned} &\mu \frac{\partial}{\partial z} \begin{pmatrix} \Psi_h(z,\mu) \\ \Psi_l(z,\mu) \end{pmatrix} + \begin{pmatrix} \sigma_h(z) & 0 \\ 0 & \sigma_l(z) \end{pmatrix} \begin{pmatrix} \Psi_h(z,\mu) \\ \Psi_l(z,\mu) \end{pmatrix} \\ &\quad - \begin{pmatrix} \sigma_{s,h\to h}(z) & \sigma_{s,l\to h}(z) \\ \sigma_{s,h\to l}(z) & \sigma_{s,l\to l}(z) \end{pmatrix} \frac{1}{2} \int_{-1}^{1} \begin{pmatrix} \Psi_h(z,\mu') \\ \Psi_l(z,\mu') \end{pmatrix} \mathrm{d}\mu' \\ &= \lambda \begin{pmatrix} \nu_h(z)\sigma_{f,h}(z) & \nu_l(z)\sigma_{f,l}(z) \\ 0 & 0 \end{pmatrix} \frac{1}{2} \int_{-1}^{1} \begin{pmatrix} \Psi_h(z,\mu') \\ \Psi_l(z,\mu') \end{pmatrix} \mathrm{d}\mu'. \end{aligned} \tag{5.2}$$

In Tables 5.7–5.10 we apply a Gauss quadrature and Crank–Nicolson scheme with 128 uniform spatial intervals in the fuel region and 8 equally sized intervals in the absorber part of the problem (resolving the material boundary), as well as 128 angles, leading to a system of size 34816×34816. The convergence behavior of Algorithm 1 is investigated with respect to three different material compositions in the absorber region: (i) the pure absorber case; (ii) a mix of 10% absorber and 90% water; and (iii) a homogeneous case, where the absorber and fuel region have the same cross-sections. The principal eigenvalues in cases (i)–(iii) are $\lambda_1^{\mathrm{CN}} \approx 1.18$, 0.92,

TABLE 5.6
Data for the control rod problem; the scatter cross-sections are arranged as in (5.2).

	Properties of the fuel in (i)–(iii) and absorber in (iii)				
	σ	σ_s		σ_f	ν
h	2.11228E-01	1.90001E-01	1.16636E-05	3.01008E-04	2.48225
l	7.23458E-01	1.85926E-02	7.04384E-01	1.01367E-02	2.43832

	Absorber properties for (i)			Absorber properties for (ii)		
	σ	σ_s		σ	σ_s	
h	3.96908E-02	1.76684E-02	1.75847E-06	1.78882E-01	1.39293E-01	9.30325E-06
l	1.74551E-01	1.12667E-05	1.60722E-02	1.03217E+00	3.37989E-02	1.00381E+00
Problem length: L = 5.25cm (fuel region: 5.0cm, absorber region: 0.25cm)						

and 0.85 and $\lambda_2^{\mathrm{CN}} \approx 99.31$, 87.70, and 82.39, respectively, and the problem details are given in Table 5.6.

The theory does not apply directly to (5.2), and even the homogeneous problem (iii) does not have an obvious symmetric reduction. Moreover, we assumed vacuum boundary conditions for our analysis above, while this model problem has reflective boundary conditions. But the numerical results are nevertheless interesting and give an indication for possible extensions of our analysis.

For our first test we used the same starting vector and stopping criterion as in the Los Alamos problem but changed the fixed shift to $\alpha_0 = 0.5$. With this and a constant inner tolerance $\tau_0 = 0.1$, we failed to converge to our demanded accuracy in cases (i)–(iii). The first five columns in Table 5.7 show that the norm of the residual and the error in the eigenvalue do not decrease any further between 200 and 2000 iterations. The increasing accuracy of the inner GMRES solves, which we saw for the Los Alamos problem, was not observed here.

TABLE 5.7
Fixed shift $\alpha_0 = 0.5$; for $\tau_0 = 10^{-12}$ the problems converge within i iterations.

	$\tau_0 = 0.1$				$\tau_0 = 10^{-12}$		
	$\Delta^{(200)}$	$\|\mathbf{res}^{(200)}\|_2$	$\Delta^{(2000)}$	$\|\mathbf{res}^{(2000)}\|_2$	i	$\Delta^{(i)}$	$\|\mathbf{res}^{(i)}\|_2$
Pure absorber	5.5E-02	2.0E-04	5.5E-02	2.0E-04	6	0.0E+00	3.2E-15
Absorber & water mix	9.3E-03	2.4E-04	9.3E-03	2.4E-04	5	0.0E+00	3.9E-16
Homogeneous case	3.9E-03	1.5E-04	3.9E-03	1.5E-04	1	0.0E+00	1.4E-15

We recovered convergence only by decreasing the fixed tolerance τ_0 to 10^{-12} and less, as the final columns in Table 5.7 show. These small tolerances resulted in almost exact solves of the linear system so that the convergence is not greatly surprising. The statement that the homogeneous problem was solved in only one iteration (last row in Table 5.7) is no typing error but is due to the fact that our starting vector with equal entries is almost an eigenvector in this case. So in order not to give problem (iii) an advantage for the remaining numerical tests, we changed our starting vector to one whose entries were chosen randomly in $(0, 1)$. Repeating the previous test for the homogeneous case with a random starting vector increased the number of iterations needed to converge to five.

Table 5.8 gives numerical results for cases (i)–(iii) using a fixed shift and decreasing tolerances. We obtain, as in the Los Alamos problem, linear but not quadratic convergence. Apart from the first iterate, the convergence for all three cases appears to be similar. This suggests that the heterogeneity does not impair the convergence in this case.

TABLE 5.8
Control rod problem using a fixed shift $\alpha_0 = 0.5$ and $\tau^{(i)} \leq 0.1 \|PT^{-1}\,\mathbf{res}^{(i)}\|_2$.

	Pure absorber			10% absorber, 90% water			Homogeneous material		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	9.0E-01			8.9E-01			8.5E-01		
1	1.7E-04	1.9E-04	2.1E-04	1.1E-05	1.2E-05	1.4E-05	4.2E-06	5.0E-06	5.9E-06
2	1.0E-06	6.0E-03	3.6E+01	8.8E-08	8.2E-03	7.7E+02	4.6E-09	1.1E-03	2.6E+02
3	4.7E-09	4.7E-03	4.6E+03	2.6E-10	2.9E-03	3.3E+04	3.7E-11	8.2E-03	1.8E+06
4	2.4E-11	5.2E-03	1.1E+06	1.0E-12	4.0E-03	1.5E+07	1.7E-13	4.4E-03	1.2E+08
5	2.3E-13	9.5E-03	3.9E+08	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
6	0.0E+00	0.0E+00	0.0E+00						

Table 5.9 illustrates the convergence properties using a constant tolerance $\tau_0 = 0.1$ and the variable shift $\alpha^{(i)}$ chosen to be the scalar flux Rayleigh quotient $\widetilde{\rho}^{(i)}$. As in the Los Alamos problem for the Crank–Nicolson discretization, we obtain linear but not quadratic convergence. The numerical results suggest that for the use of a Rayleigh quotient shift, the heterogeneity in the first two problems may influence the speed of the linear convergence.

TABLE 5.9
Results for the control rod problem with $\alpha^{(i)} = \widetilde{\rho}^{(i)}$ and constant tolerance $\tau_0 = 0.1$.

	Pure absorber			10% absorber, 90% water			Homogeneous material		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	9.0E-01			8.9E-01			8.5E-01		
1	1.1E-02	1.3E-02	1.4E-02	1.7E-05	1.9E-05	2.2E-05	1.9E-05	2.3E-05	2.7E-05
2	1.6E-04	1.4E-02	1.2E+00	1.6E-06	9.4E-02	5.5E+03	8.4E-08	4.4E-03	2.3E+02
3	1.4E-05	8.9E-02	5.6E+02	9.4E-09	5.8E-03	3.6E+03	2.4E-10	2.8E-03	3.3E+04
4	4.7E-08	3.3E-03	2.3E+02	3.8E-09	4.0E-01	4.3E+07	7.6E-13	3.3E-03	1.4E+07
5	5.2E-10	1.1E-02	2.4E+05	5.8E-11	1.6E-02	4.1E+06	0.0E+00	0.0E+00	0.0E+00
6	1.0E-11	1.9E-02	3.6E+07	0.0E+00	0.0E+00	0.0E+00			
7	0.0E+00	0.0E+00	0.0E+00						

Solving the same fixed tolerance problems with the angular flux Rayleigh quotient $\rho^{(i)}$ gave convergence results similar to those in Table 5.9 without indicating superiority of one Rayleigh quotient over the other.

Now, using Rayleigh quotient shifts and decreasing tolerances, the obtained convergence rates for the two variable shift cases improve (see, as an example, Table 5.10), but due to the few iterations needed, we are not able to clearly establish whether quadratic convergence is obtained.

TABLE 5.10
Numerical results for the control rod problem using the angular flux Rayleigh quotient shift $\rho^{(i)}$ and decreasing tolerances $\tau^{(i)} \leq 0.1 \|PT^{-1}\,\mathbf{res}^{(i)}\|_2$ for different materials in the absorber region.

	Pure absorber			10% absorber, 90% water			Homogeneous material		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	9.0E-01			8.9E-01			8.5E-01		
1	1.3E-05	1.5E-05	1.6E-05	2.7E-05	3.0E-05	3.4E-05	1.0E-05	1.2E-05	1.4E-05
2	1.3E-11	1.0E-06	7.5E-02	2.2E-11	8.2E-07	3.0E-02	2.4E-12	2.3E-07	2.3E-02
3	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00

Finally, Table 5.11 contains numerical results for the control rod problem with

TABLE 5.11
Results for the control rod problem with $\alpha^{(i)} = \widetilde{\rho}^{(i)}$ and constant tolerance $\tau_0 = 0.1$ when using the Euler scheme discussed in section 4 with 272 spatial intervals and 128 angular directions.

	Pure absorber			10% absorber, 90% water			Homogeneous material		
i	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$	$\Delta^{(i)}$	$\frac{\Delta^{(i)}}{\Delta^{(i-1)}}$	$\frac{\Delta^{(i)}}{(\Delta^{(i-1)})^2}$
0	4.6E+00			4.7E+00			4.5E+00		
1	1.2E-02	2.6E-03	5.5E-04	2.1E-05	4.5E-06	9.6E-07	2.1E-05	4.6E-06	1.0E-06
2	1.8E-04	1.5E-02	1.3E+00	5.8E-06	2.7E-01	1.3E+04	1.2E-07	5.6E-03	2.7E+02
3	1.8E-05	1.0E-01	5.7E+02	3.0E-07	5.1E-02	8.9E+03	4.8E-10	4.2E-03	3.6E+04
4	5.8E-08	3.2E-03	1.8E+02	2.5E-09	8.4E-03	2.9E+04	6.9E-11	1.4E-01	3.0E+08
5	1.0E-09	1.8E-02	3.0E+05	8.9E-11	3.6E-02	1.4E+07	2.9E-13	4.3E-03	6.2E+07
6	2.9E-12	2.8E-03	2.8E+06	1.9E-12	2.1E-02	2.4E+08	0.0E+00	0.0E+00	0.0E+00
7	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00			

$\alpha^{(i)} = \widetilde{\rho}^{(i)}$ and constant tolerance $\tau_0 = 0.1$ when using the Euler scheme discussed in section 4. The tests use 272 spatial intervals and 128 angular directions, and the principal eigenvalues in cases (i)–(iii), respectively, are $\lambda_1^{\mathrm{Eul}} \approx 1.24$, 0.92, and 0.87. The results clearly show linear and not quadratic convergence. (The latter was observed, for example, in Table 5.3.)

6. Conclusion. We provided a convergence analysis for inexact inverse iteration to solve the criticality problem in neutron transport theory for monoenergetic homogeneous model problems with isotropic scattering and vacuum boundary conditions. Numerical experiments on model problems with one space and one angular dimension were presented. A homogeneous monoenergetic test problem was considered as well as a more realistic heterogeneous physical problem which also has two energy levels. The numerical results showed to be in good agreement with the theory and emphasized the advantage of using a symmetry-preserving discretization. The theory provides guidelines for the choice of shift and inner tolerance strategies in eigenvalue iterative methods and also helps us to identify scalar flux Rayleigh quotients which can give more accurate eigenvalue approximations than angular flux Rayleigh quotients.

Acknowledgments. We thank Melina Freitag and Alastair Spence (University of Bath), as well as Paul Smith (Serco Technical and Assurance Services), for useful discussions.

REFERENCES

[1] E. J. ALLEN AND R. M. BERRY, *The inverse power method for calculation of multiplication factors*, Ann. Nuclear Energy, 29 (2002), pp. 929–935.
[2] M. ASADZADEH, *L_p and eigenvalue error estimates for the discrete ordinates method for two-dimensional neutron transport*, SIAM J. Numer. Anal., 26 (1989), pp. 66–87.
[3] G. I. BELL AND S. GLASSTONE, *Nuclear Reactor Theory*, Reinhold, New York, 1970.
[4] J. BERNS-MÜLLER, I. G. GRAHAM, AND A. SPENCE, *Inexact inverse iteration for symmetric matrices*, Linear Algebra Appl., 416 (2006), pp. 389–413.
[5] B. CHANG, *The conjugate gradient method solves the neutron transport equation h-optimally*, Numer. Linear Algebra Appl., 14 (2007), pp. 751–769.
[6] M. A. FREITAG, *Inner-outer Iterative Methods for Eigenvalue Problems - Convergence and Preconditioning*, Ph.D. thesis, Department of Mathematical Sciences, University of Bath, Bath, UK, 2007.
[7] M. A. FREITAG AND A. SPENCE, *Convergence theory for inexact inverse iteration applied to the generalised nonsymmetric eigenproblem*, Electron. Trans. Numer. Anal., 28 (2007), pp. 40–64.

[8] M. A. Freitag and A. Spence, *Rayleigh quotient iteration and simplified Jacobi-Davidson method with preconditioned iterative solves*, Linear Algebra Appl., 428 (2008), pp. 2049–2060.
[9] S. Hamilton, M. Benzi, and J. Warsa, *Negative flux fixups in discontinuous finite element S_N transport*, in Proceedings of the International Conference on Mathematics, Computational Methods and Reactor Physics (M&C 2009), American Nuclear Society, LaGrange Park, IL, 2009.
[10] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis*, Springer-Verlag, Berlin, New York, 1979.
[11] H. Hochstadt, *Integral Equations*, Wiley-Interscience, New York, London, Sydney, 1973.
[12] C. Johnson and J. Pitkäranta, *Convergence of a fully discrete scheme for two-dimensional neutron transport*, SIAM J. Numer. Anal., 20 (1983), pp. 951–966.
[13] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, 2nd ed., Pergamon Press, Oxford, UK, 1982.
[14] E. E. Lewis and W. F. Miller, Jr., *Computational Methods of Neutron Transport*, John Wiley & Sons, New York, 1984.
[15] T. A. Manteuffel and K. J. Ressel, *Least-squares finite-element solution of the neutron transport equation in diffusive regimes*, SIAM J. Numer. Anal., 35 (1998), pp. 806–835.
[16] G. Marchuk and V. Lebedev, *Numerical Methods in the Theory of Neutron Transport*, Harwood Academic, New York, 1986.
[17] I. Marek, *On a problem of mathematical physics*, Apl. Mat., 11 (1966), pp. 89–112.
[18] J. Mika, *Existence and uniqueness of the solution to the critical problem in neutron transport theory*, Studia Math., 37 (1970/71), pp. 213–225.
[19] J. E. Morel, B. T. Adams, T. Noh, J. M. McGhee, T. M. Evans, and T. J. Urbatsch, *Spatial discretizations for self-adjoint forms of the radiative transfer equations*, J. Comput. Phys., 214 (2006), pp. 12–40.
[20] J. Pitkäranta and L. R. Scott, *Error estimates for the combined spatial and angular approximations of the transport equation for slab geometry*, SIAM J. Numer. Anal., 20 (1983), pp. 922–950.
[21] B. P. Rynne and M. A. Youngson, *Linear Functional Analysis*, Springer Undergrad. Math. Ser., Springer-Verlag, London, 2008.
[22] F. Scheben, *Iterative Methods for Criticality Computations in Neutron Transport Theory*, Ph.D. thesis, Department of Mathematics, University of Bath, Bath, UK, 2011.
[23] A. Sood, R. A. Forster, and D. K. Parsons, *Analytical benchmark test set for criticality code verification*, Prog. Nucl. Energy, 42 (2003), pp. 55–106.
[24] W. M. Stacey, *Nuclear Reactor Physics*, Wiley-VCH, Berlin, 2007.
[25] J. S. Warsa, T. A. Wareing, J. E. Morel, J. M. McGhee, and R. B. Lehoucq, *Krylov subspace iterations for deterministic k-eigenvalue calculations*, Nucl. Sci. Eng., 147 (2004), pp. 26–42.