PRECONDITIONING A CLASS OF FOURTH ORDER PROBLEMS BY OPERATOR SPLITTING

EBERHARD BÄNSCH*, PEDRO MORIN†, AND RICARDO H. NOCHETTO‡

Abstract. We develop preconditioners for systems arising from finite element discretizations of parabolic problems which are fourth order in space. We consider boundary conditions which yield a natural splitting of the discretized fourth order operator into two (discrete) linear second order elliptic operators, and exploit this property in designing the preconditioners. The underlying idea is that efficient methods and software to solve second order problems with optimal computational effort are widely available. We propose symmetric and non-symmetric preconditioners, along with theory and numerical experiments. They both document crucial properties of the preconditioners as well as their practical performance. It is important to note that we neither need H^s-regularity, $s > 1$, of the continuous problem nor quasi-uniform grids.

Keywords: parabolic problems, fourth order problems, preconditioning, finite elements, operator splitting, condition number, iterative methods.

1. Introduction. For suitable boundary conditions, combining semi-implicit time discretization with time-step Δt and operator splitting for evolution equations governed by (nonlinear) fourth order operators leads to linear elliptic systems of the form

$$\begin{aligned} u - \Delta t \operatorname{div}(a \operatorname{grad} v) &= f, \\ \operatorname{div}(b \operatorname{grad} u) + v &= g, \end{aligned} \tag{1.1}$$

on a bounded polyhedral domain $\Omega \subseteq \mathbb{R}^d$, $d \geq 1$. Hereafter, a and b are measurable functions with values on symmetric positive definite (s.p.d.) matrices and satisfying

$$\lambda_a(x)|\xi|^2 \leq \xi^T a(x)\xi \leq \Lambda_a(x)|\xi|^2, \quad \lambda_b(x)|\xi|^2 \leq \xi^T b(x)\xi \leq \Lambda_b(x)|\xi|^2, \tag{1.2}$$

for all $\xi \in \mathbb{R}^d$. In this paper we develop preconditioners for spatially discretized versions of this system, for instance via finite element methods, thereby relying on the existence of efficient methods and software for each component of the system [18, 20, 21, 23, 37, 40, 48, 49, 50, 52, 53, 54], especially on graded meshes [1, 26, 49, 51]. Similar ideas have been developed for the bi-harmonic operator with Dirichlet boundary conditions $u = \partial_\nu u = 0$, typical in structural mechanics, and quasi-uniform meshes in [15, 16, 19, 44, 45]. Materials science and fluid dynamics problems come with different boundary conditions which yield the operator splitting (1.1). We are interested in the latter case, graded shape-regular meshes, and condition numbers which are insensitive to Δt and the mesh size.

In § 1.1 we motivate the importance of studying (1.1) for geometric PDE, discuss boundary conditions in § 1.2 giving rise to (1.1) along with its weak formulation, and close the Introduction in § 1.3 with a presentation of the preconditioners for (1.1).

*Angewandte Mathematik III, Friedrich-Alexander-Universität Erlangen-Nürnberg, Haberstr. 2, 91058 Erlangen, Germany

†Instituto de Matemática Aplicada del Litoral, CONICET - Universidad Nacional del Litoral, Güemes 3450, S3000GLN Santa Fe, Argentina. Partially supported by CONICET through Grant PIP 5478 and PIP 5811, and Universidad Nacional del Litoral through Grant CAI+D PI-62-312.

‡Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742. Partially supported by NSF grants DMS-0505454, DMS-0807811 and INT-0126272.

1.1. Motivation: Geometric PDE. Evolution equations governed by (nonlinear) fourth order operators arise in a number of fields from materials science and fluid dynamics to geometry. *Surface diffusion* is a geometric flow governed by the following equation on an evolving hypersurface (or curve) $\Gamma \subset \mathbb{R}^d$ of co-dimension 1

$$V = -\Delta_\Gamma H, \tag{1.3}$$

where V is the normal velocity, H is the total curvature of Γ (sum of principal curvatures), and Δ_Γ is the Laplace-Beltrami operator on Γ [3, 4, 5, 6, 8, 11, 24]. It is important to realize that, once written parametrically, Δ_Γ is like a second order elliptic operator with variable coefficients, and so is its finite element counterpart [6, 7, 8, 25, 27, 32, 41]. Equation (1.3) may be viewed as a gradient flow for the surface area $\int_\Gamma 1$ of Γ with the H^{-1}-metric on Γ. For graphs Γ, described by the function u over $\Omega \subset \mathbb{R}^{n-1}$, we have [5, 27]

$$H = \operatorname{div} \frac{\nabla u}{q(u)}, \qquad V = \frac{\partial_t u}{q(u)},$$

where $q(u) = \sqrt{1 + |\nabla u|^2}$. This leads to the natural operator splitting

$$\frac{\partial_t u}{q(u)} = \Delta_\Gamma v, \qquad v = -\operatorname{div} \frac{\nabla u}{q(u)},$$

with $v = -H$. Once supplemented with, for instance, periodic boundary conditions, this system can be rewritten in Ω as follows:

$$\partial_t u - \operatorname{div}\Big((q(u)\Big(I - \frac{\nabla u \otimes \nabla u}{q(u)^2}\Big)\nabla v\Big) = 0, \qquad v + \operatorname{div} \frac{\nabla u}{q(u)} = 0. \tag{1.4}$$

A semi-implicit Euler discretization in time with time-step Δt leads to (1.1) for unknowns u_n, v_n with $a = q(u_{n-1})(I - q(u_{n-1})^{-2}\nabla u_{n-1} \otimes \nabla u_{n-1})$ and $b = q(u_{n-1})^{-1}$.

For closed surfaces, we mention the parametric method of [5], which leads to a system essentially of the form (1.1). The following fundamental geometric relationship between position $\boldsymbol{X}$ and vector-valued curvature $\boldsymbol{H} = H\boldsymbol{\nu}$, with $\boldsymbol{\nu}$ being the unit outer normal to Γ, plays a key role [31]:

$$\boldsymbol{H} = -\Delta_\Gamma \boldsymbol{X}. \tag{1.5}$$

In particular, integration by parts on Γ, along with $V = \partial_t \boldsymbol{X} \cdot \boldsymbol{\nu}$, transforms (1.3) and (1.5) into the following system [6]:

$$\partial_t \boldsymbol{X} \cdot \boldsymbol{\nu} - \Delta_\Gamma H = 0, \qquad H\boldsymbol{\nu} + \Delta_\Gamma \boldsymbol{X} = 0. \tag{1.6}$$

Employing a semi-implicit Euler discretization $\partial_t \boldsymbol{X}_n \approx \Delta t^{-1}(\boldsymbol{X}_n - \boldsymbol{X}_{n-1})$, which keeps the geometry $\Gamma, \boldsymbol{\nu}$ and operator ∇_Γ explicitly evaluated on Γ_{n-1}, converts (1.6) into a system of the form (1.1), with vector-valued $u = \boldsymbol{X}_n$ and scalar $v = H_n$.

In contrast to surface diffusion, the L^2-gradient flow for the bending energy $\int_\Gamma H^2$ gives rise to the so-called *Willmore flow* in 3D

$$V = \Delta_\Gamma H + \frac{H}{2}\big(2H^2 - \kappa\big), \tag{1.7}$$

where κ stands for the Gauss curvature. For graphs Γ described by a function u, this geometric PDE becomes [27]

$$\frac{\partial_t u}{q} - \operatorname{div}\Big(\frac{1}{q}\Big(I - \frac{\nabla u \otimes \nabla u}{q^2}\Big)\nabla v\Big) - \frac{1}{2}\operatorname{div}\Big(\frac{H^2}{q}\nabla u\Big) = 0, \qquad \frac{v}{q} + \operatorname{div}\frac{\nabla u}{q}, \tag{1.8}$$

where $v = qH$. Again a semi-implicit time discretization leads to a system similar to (1.1). For the parametric case, we mention the vector-valued formulation of [17], which replaces (1.7) by

$$\partial_t \boldsymbol{X} - \Delta_\Gamma \boldsymbol{H} + \operatorname{div}_\Gamma \left((\nabla_\Gamma \boldsymbol{X} + \nabla_\Gamma \boldsymbol{X}^T) \nabla_\Gamma \boldsymbol{H} \right) - \frac{1}{2} \nabla_\Gamma \operatorname{div}_\Gamma \boldsymbol{H} = 0. \tag{1.9}$$

and couples it with (1.5); see the related methods in [32, 41]. In contrast, the method of [8] is based on the following equivalent system:

$$\partial_t \boldsymbol{X} \cdot \boldsymbol{\nu} - \Delta_\Gamma H + \frac{1}{2} H^3 - H |\nabla_\Gamma \boldsymbol{\nu}|^2 = 0, \qquad H\boldsymbol{\nu} + \Delta_\Gamma \boldsymbol{X} = 0. \tag{1.10}$$

A semi-implicit time discretization $\boldsymbol{X}_n = \boldsymbol{X}_{n-1} + \Delta t \boldsymbol{V}_n$, as with (1.6), converts (1.9) and (1.5) (resp. (1.10) and (1.5)) into a system with basic structure similar to (1.1) for the vector-valued functions $u = \boldsymbol{V}_n, v = \boldsymbol{H}_n$ (resp. $u = \boldsymbol{X}_n, v = H_n$).

The celebrated *Cahn-Hilliard* equation, describing the evolution of a binary mixture with concentration u and chemical potential v, reads [2, 13, 24, 27]

$$\partial_t u - \operatorname{div} \Big(M(u) \nabla v \Big) = 0, \qquad v + \Delta u = \psi(u), \tag{1.11}$$

where the mobility $M(u) > 0$ may depend on concentration, and ψ is the derivative of a potential with double-wells of equal heights at ± 1; for instance $\psi(u) = u(u^2 - 1)$. Semi-implicit time discretization leads again to (1.1).

The mobility $M(u)$ may be degenerate in some applications. For instance, a surface Γ evolving by surface diffusion can be recovered as the limit $\epsilon \downarrow 0$ of the zero-level set of the solution u, with singularly perturbed chemical potential

$$v + \epsilon \Delta u = \frac{1}{\epsilon} \psi(u),$$

and degenerate mobility $M(u) = 1 - u^2$ [9, 24, 27]. Lubrication theory for *thin films* yields similar equations with degenerate mobility [12, 14, 35, 36]. Semi-implicit time discretization in turn gives rise to (1.1) with degenerate a and/or b. Our theory below does not cover these degeneracies, but we explore computationally the preconditioners robustness with respect to coefficient degeneracies in §7.

Similar diffuse interface approaches are available to approximate Willmore flow, as well as several variants, in the limit of interface thickness $\epsilon \downarrow 0$ [27, 28, 29, 30]. We finally mention the coupling of membrane bending with orientational order of bilipds via director fields [10]. These models lead to PDE in Ω similar to (1.1).

1.2. Weak Formulation and Boundary Conditions. We now discuss the *weak formulation* of (1.1) along with *boundary conditions* which allow operator splitting. Periodic boundary conditions are customary for (1.3) and (1.7) when the surface Γ is without boundary, as well as for (1.11); in this case, we add the constraint

$$\int_\Omega u = \int_\Omega v = 0, \tag{1.12}$$

to gain uniqueness of (1.1). Alternatively, we assume that the boundary Γ of Ω is split into two disjoint parts Γ_D and Γ_N. We impose homogeneous Dirichlet and Neumann boundary conditions on Γ_D, Γ_N, respectively:

$$u = v = 0 \quad \text{on } \Gamma_D \times (0, T), \tag{1.13}$$

$$\boldsymbol{\nu} \cdot a \nabla v = \boldsymbol{\nu} \cdot b \nabla u = 0 \quad \text{on } \Gamma_N \times (0, T), \tag{1.14}$$

with $\boldsymbol{\nu}$ the outer normal to Γ. In case the Dirichlet part Γ_D of Γ has zero Haussdorff measure $H^{d-1}(\Gamma_D) = 0$, we supplement (1.14) with the vanishing conditions (1.12). Then problem (1.1) takes the weak form: find $u, v \in \mathbb{V}$ such that

$$\begin{aligned}(u,\phi) + \Delta t(a \operatorname{grad} v, \operatorname{grad}\phi) &= (f,\phi), \quad \forall\ \phi \in \mathbb{V}\\ -(b \operatorname{grad} u, \operatorname{grad}\psi) + (v,\psi) &= (g,\psi), \quad \forall\ \psi \in \mathbb{V},\end{aligned}$$

where the subspace $\mathbb{V}$ of $H^1(\Omega)$ incorporates the boundary condition (1.13) or the vanishing condition (1.12). The key property of $\mathbb{V}$ used below is that the seminorm

$$\sqrt{(\nabla v, \nabla v)} \text{ be actually a norm in } \mathbb{V}. \tag{1.15}$$

We stress that homogeneous Dirichlet boundary conditions for the 4th order operator $\Delta^2 u$, namely $u = \partial_{\boldsymbol{\nu}} u = 0$, do not yield the factorization (1.1). We refer to [38] as well as [15, 16, 19, 44, 45] for this case, which is important in plate theory but is excluded from the present analysis.

1.3. Discrete Systems and Preconditioners. Let $\mathcal{T}$ be a shape-regular but possibly graded mesh of Ω. If $\mathbb{V}_\mathcal{T} \subseteq \mathbb{V}$ denotes a finite element space over $\mathcal{T}$ (see Section 3 for details), we consider the following finite element discretization:

$$u_\mathcal{T}, v_\mathcal{T} \in \mathbb{V}_\mathcal{T}: \qquad \begin{cases} (u_\mathcal{T},\phi) + \Delta t\,(a \operatorname{grad} v_\mathcal{T}, \operatorname{grad}\phi) = (f,\phi) \\ -(b \operatorname{grad} u_\mathcal{T}, \operatorname{grad}\psi) + (v_\mathcal{T},\psi) = (g,\phi) \end{cases} \qquad \forall \phi,\psi \in \mathbb{V}_\mathcal{T}. \tag{1.16}$$

We define the discrete operators $A_\mathcal{T} : \mathbb{V}_\mathcal{T} \to \mathbb{V}_\mathcal{T}$ and $B_\mathcal{T} : \mathbb{V}_\mathcal{T} \to \mathbb{V}_\mathcal{T}$ by

$$(A_\mathcal{T}\phi,\psi) = (a \operatorname{grad}\phi, \operatorname{grad}\psi), \quad (B_\mathcal{T}\phi,\psi) = (b \operatorname{grad}\phi, \operatorname{grad}\psi) \quad \forall \phi,\psi \in \mathbb{V}_\mathcal{T}, \tag{1.17}$$

and note that both $A_\mathcal{T}$, $B_\mathcal{T}$ are s.p.d., that is, they are symmetric (w.r.t. the L^2 inner product) and positive definite. The discretized system in $\mathbb{V}_\mathcal{T}$ now reads

$$\begin{aligned} u_\mathcal{T} + \Delta t\, A_\mathcal{T} v_\mathcal{T} &= f_\mathcal{T} \\ -B_\mathcal{T} u_\mathcal{T} + v_\mathcal{T} &= g_\mathcal{T}, \end{aligned}$$

which can be equivalently written as

$$(I + \Delta t\, A_\mathcal{T} B_\mathcal{T}) u_\mathcal{T} = f_\mathcal{T} - \Delta t A_\mathcal{T} g_\mathcal{T} = F_\mathcal{T}. \tag{1.18}$$

Notice that operator $D = (I + \Delta t\, A_\mathcal{T} B_\mathcal{T})$ is symmetric in the inner product induced by $A_\mathcal{T}^{-1}$. We set

$$\tau = \sqrt{\Delta t} \qquad \text{and} \qquad S := \tau A_\mathcal{T}, \quad T := \tau B_\mathcal{T}. \tag{1.19}$$

Therefore $D = (I + ST)$, and (1.18) now reads

$$D u_\mathcal{T} = (I + ST) u_\mathcal{T} = F_\mathcal{T}. \tag{1.20}$$

In this paper we study left, right and left-right preconditioners for this system:

Left Preconditioner: We use $P_L^2 := (I + S)^2$ as a left preconditioner for D, leading to the following preconditioned operator

$$P_L^{-2} D = (I + S)^{-2}(I + ST) = (I + S)^{-2} S(S^{-1} + T).$$

This can also be interpreted as a left symmetric preconditioner $\widetilde{P}_L := (I+S)^2 S^{-1}$ applied to the symmetric operator $\widetilde{D} = (S^{-1}+T)$. We implement this preconditioned system within the algorithm *Preconditioned Conjugate Gradient* (PCG), and recall that within this solver, using $\widetilde{P}_L$ as a left preconditioner for $\widetilde{D}$ is equivalent to solving for the symmetric operator $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$. In § 4 we study the spectral radius of $\widetilde{P}_L^{-1}\widetilde{D}$ which coincides with that of $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$, and conclude that P_L^2 (resp. $\widetilde{P}_L$) is a good preconditioner for D (resp. $\widetilde{D}$).

Right Preconditioner: We use $P_R^2 = (I+T)^2$ as a right preconditioner for D, leading to the following preconditioned operator

$$DP_R^{-2} = (I+ST)(I+T)^{-2} = (T^{-1}+S)T(I+T)^{-2}.$$

This is again the composition of two symmetric operators $\bar{D} = T^{-1}+S$ and $\widetilde{P}_R = T(I+T)^{-2}$. The implementation of this within PCG leads to solving for the symmetric operator $\widetilde{P}_R^{-1/2}\bar{D}\widetilde{P}_R^{-1/2}$, and the situation is analogous to the *Left Preconditioner*, after interchanging the roles of S and T. Due to this analogy we conclude that the same properties of the Left Preconditioner carry over to the Right Preconditioner, and the analysis of the latter is omitted.

Left-Right Preconditioner: We use $P_L = (I+S)$ and $P_R = (I+T)$ as left and right preconditioners for D, respectively, leading to the following preconditioned operator

$$P_L^{-1}DP_R^{-1} = (I+S)^{-1}(I+ST)(I+T)^{-1}$$

It turns out that this preconditioner exhibits better balance, performance, and robustness than the others but at the expense of symmetry. In § 5 we analyze the location of the spectrum of $P_L^{-1}DP_R^{-1}$ and use our findings to study convergence rates for Richardson and GMRes methods in §§ 5.2 and 5.3, respectively.

In order to apply these preconditioners, efficient methods for solving systems

$$(I+\tau C_{\mathcal{T}})v_{\mathcal{T}} = z_{\mathcal{T}}, \qquad C_{\mathcal{T}}v_{\mathcal{T}} = z_{\mathcal{T}} \tag{1.21}$$

are required, where $C_{\mathcal{T}}$ is a s.p.d. operator associated to a second order elliptic problem and $\tau > 0$ is arbitrary. Systems (1.21) can be solved with optimal computational effort thanks to multilevel solvers such as multigrid and BPX preconditioners. We refer to [18, 20, 21, 23, 37, 40, 48, 49, 50, 52, 53, 54] for the general methodology mostly on quasi-uniform meshes $\mathcal{T}$ and [1, 26, 49, 51] for graded meshes. The symmetric operator $\widetilde{D} = S^{-1}+T$ is related to Tikhonov regularization, and multilevel preconditioners were proposed in [22] for quasi-uniform meshes of size h and $\tau \geq h^2$, $\Delta t = \tau^2$ being the regularization parameter (see Corollary 2.1 and conditions (2.7) and (2.8) as well as the discussion in page 476 of [22]).

Therefore, we focus our attention on the study of the proposed preconditioned operators assuming that optimal solution techniques for systems (1.21) are available. We start presenting remarkable computational results for the three preconditioned operators in § 2. After a brief discussion of preliminary results in § 3, we estimate the spectral radius of the symmetric preconditioned operator $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$ in § 4 and we study the location of the spectrum of the nonsymmetric preconditioned operator $P_L^{-1}DP_R^{-1}$ in § 5. In the symmetric case, convergence rates for preconditioned conjugate gradient (PCG) and GMRes (S-GMRes) directly follow. In the non-symmetric

case, instead, we study convergence rates for Richardson and GMRes methods in § 5.2 and § 5.3, respectively, and their application to finite element discretizations in § 5.4. It is worth stressing that our results are valid for any polynomial degree and we neither need H^s-regularity, $s > 1$, of the continuous problem (1.1) nor quasi-uniform grids $\mathcal{T}$. In § 6 we apply the proposed preconditioners to a semi-implicit time stepping of the governing system (1.4) for surface diffusion. We conclude in § 7 with computational experiments revealing the relative merits of each preconditioner.

2. Numerical Study of Preconditioners. In this section we present a preliminary study of the computational performance of the preconditioners presented in § 1.3. As the domain of interest we choose the L-shaped domain

$$\Omega := (-1,1) \times (-1,1) \setminus [0,1] \times [0,1],$$

to verify that the results do not require full H^2-regularity of (1.1). We consider the following four examples, in the sequel referred to as *nice*, *semi*, *nasty*, and *degenerate*, which reflect situations of increasing difficulty.

Example 2.1 (Nice): We consider

$$a(x_1,x_2) = 1, \qquad b(x_1,x_2) = \begin{cases} 0.6, & \text{if } x_2 < x_1, \\ 1.2, & \text{otherwise.} \end{cases}$$

Example 2.2 (Semi): We consider

$$a(x_1,x_2) = 1 + 0.1|x_1| + |x_2|, \qquad b(x_1,x_2) = 1.5 + 0.5\sin(5\pi x_1)\sin(8\pi x_2).$$

Example 2.3 (Nasty): We consider

$$a(x_1,x_2) = 0.3 + 0.1|x_1| + |x_2|, \qquad b(x_1,x_2) = 10 + 3\sin(5\pi x_1)\sin(8\pi x_2).$$

Example 2.4 (Degenerate): The goal of the last example is to investigate the behavior of the preconditioners, when coefficient a degenerates, a case that is not covered by our theory:

$$a(x_1,x_2) = 0.1|x_1| + |x_2| \qquad b(x_1,x_2) = 10 + 3\sin(5\pi x_1)\sin(8\pi x_2).$$

Notice that coefficient a vanishes for $x = (x_1, x_2) = 0$.

We deal with both uniform and graded meshes $\mathcal{T}$, and we let $\mathbb{V}_\mathcal{T}$ be the finite element space of piecewise linear functions with homogeneous Dirichlet boundary conditions. In the rest of this section we report on the behavior of the different preconditioned systems for the examples 2.1–3. Example 2.4 falls out of our theory, and we discuss it in § 7.

For the computations below we observe that there is a canonical homeomorphism $m : L(\mathbb{V}_\mathcal{T}) \to \mathbb{R}^{N\times N}$, with $N = \dim \mathbb{V}_\mathcal{T}$, between the algebra of linear operators $L(\mathbb{V}_\mathcal{T})$ on $\mathbb{V}_\mathcal{T}$ and the algebra of matrices $\mathbb{R}^{N\times N}$ representing the operator for the nodal basis $(\phi_i)_{i=1}^N$ of $\mathbb{V}_\mathcal{T}$. More precisely, operators $A = A_\mathcal{T}, B = B_\mathcal{T}$ have the matrix representation

$$m(A) = \boldsymbol{M}^{-1}\boldsymbol{A}, \quad m(B) = \boldsymbol{M}^{-1}\boldsymbol{B},$$

where $\boldsymbol{M}$ is the mass matrix and $\boldsymbol{A}, \boldsymbol{B}$ are the stiffness matrices, respectively, defined by

$$\boldsymbol{M}_{i,j} = (\phi_j, \phi_i), \qquad \boldsymbol{A}_{i,j} = (a\nabla\phi_j, \nabla\phi_i), \qquad \boldsymbol{B}_{i,j} = (b\nabla\phi_j, \nabla\phi_i).$$

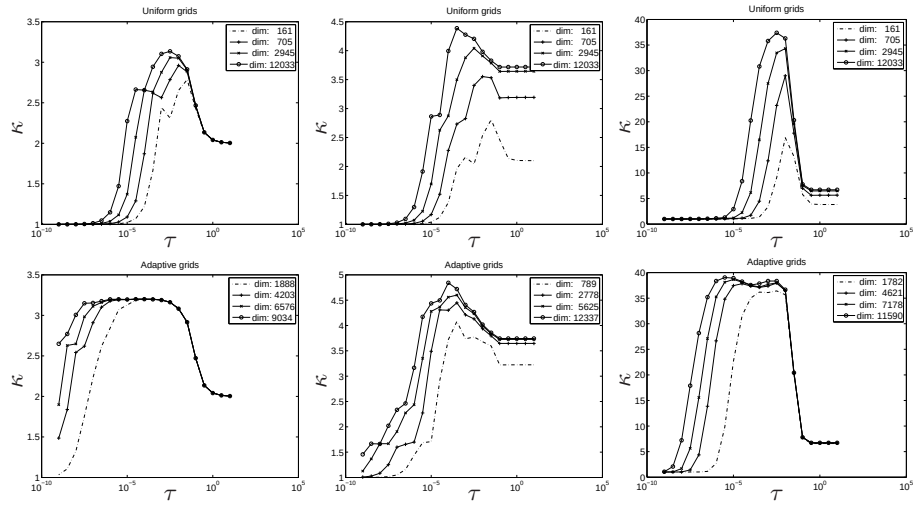

FIG. 2.1. *Condition numbers versus τ for the symmetrically preconditioned systems. The columns correspond to Examples 2.1–2.3 whereas the rows to uniform grids (top) and graded grids (bottom).*

Thus $(\lambda, e) \in \mathbb{R} \times \mathbb{V}_{\mathcal{T}}$ is an eigenpair of A, if and only if $(\lambda, \mathbf{e}) \in \mathbb{R} \times \mathbb{R}^N$ is an eigenpair of $\boldsymbol{M}^{-1}\boldsymbol{A}$, where $\mathbf{e} = (e_j)_{j=1}^N$ is the nodal vector of e: $e = \sum_{j=1}^N e_j \phi_j$. The adjoint operator A^* (with respect to the L^2 inner product) has the matrix representation $m(A^*) = \boldsymbol{M}^{-1}\boldsymbol{A}^T$, whence

$$m(A^*A) = \boldsymbol{M}^{-1}\boldsymbol{A}^T\boldsymbol{M}^{-1}\boldsymbol{A}.$$

We used the finite element toolbox ALBERTA [43] to compute the matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{M}$ and employed the sparse tools of MATLAB to provide matrix-vector products involving matrices associated with operators P, D and R, as well as spectra and condition numbers of these matrices. Moreover, both CG and GMRes were executed within MATLAB and iteration numbers for them were obtained by the procedures provided by MATLAB. In Figures 2.1 and 2.2 below, "dim" stands for $\dim \mathbb{V}_{\mathcal{T}}$. The results reported in these figures are in excellent agreement with the theory of § 4 and § 5.

2.1. Left Preconditioner. Figure 2.1 shows the condition numbers κ versus τ for the symmetric preconditioned system $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$ with

$$\widetilde{D} = S^{-1} + T, \qquad \widetilde{P}_L = (I+S)^2 S^{-1},$$

for both uniform meshes (top row) and graded meshes (bottom row). Different regimes with respect to τ and meshsize can be clearly observed. For large values of τ, we have $\widetilde{P}_L \approx S$ and $\widetilde{D} \approx T$, whence the preconditioned system satisfies

$$\widetilde{P}_L^{-1}\widetilde{D} \approx S^{-1}T = A_{\mathcal{T}}^{-1}B_{\mathcal{T}};$$

since $A_{\mathcal{T}}^{-1}B_{\mathcal{T}}$ is an operator of order zero, its condition number is essentially independent of the meshsize. On the other hand, for very small values of τ, $\widetilde{P}_L \approx S^{-1} \approx \widetilde{D}$, and thus the preconditioned system behaves like

$$\widetilde{P}_L^{-1}\widetilde{D} \approx I;$$

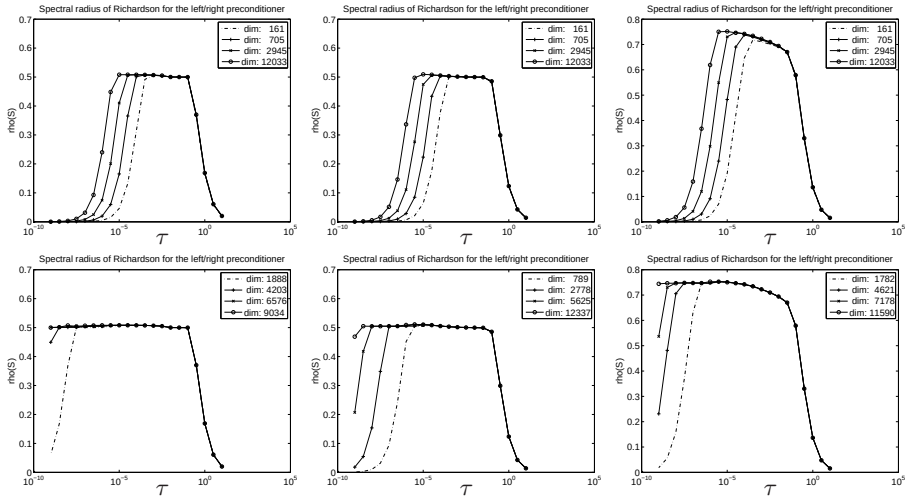

FIG. 2.2. *Spectral radius $\rho(Q)$ for Richardson's iteration operator Q versus τ for the left/right non-symmetrically preconditioned system. The columns correspond to Examples 2.1–3 whereas the rows to uniform grids (top) and graded grids (bottom).*

consequently the condition number tends to 1 as $\tau \to 0$. However, for intermediate values of τ, κ depends on both τ and meshsize, but it is uniformly bounded. The different behavior of κ for the considered examples confirms the dependence of an upper bound for κ on the relationship between a and b, which is further explained in § 4. On the other hand, we observe no substantial difference between uniform and graded grids.

2.2. Left-Right Preconditioner. We consider the (non-symmetric) Left-Right preconditioned operator $P_L^{-1}DP_R^{-1} = (I+S)^{-1}(I+ST)(I+T)^{-1}$. We report in Figure 2.2 the spectral radius $\rho(Q)$ of the Richardson's iteration operator $Q = I - P_L^{-1}DP_R^{-1}$ as a function of τ, for uniform meshes (top row) and graded meshes (bottom row). The value of $\rho(Q)$ appears to be uniformly away from 1, irrespective of τ and meshsize, but to depend on the relationship between matrices a and b. This is further explained in § 5.

3. Preliminary Results. For two complex numbers $z_1, z_2 \in \mathbb{C}$, whenever we write $z_1 \geq z_2$, we understand that $z_1 - z_2 \in \mathbb{R}$ and that $z_1 - z_2 \geq 0$, and likewise for "$\leq$", "$>$", "$<$".

In the following $\mathbb{H}$ denotes a finite dimensional Hilbert space with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$. For any operator $M : \mathbb{H} \to \mathbb{H}$ we denote its spectrum by $\sigma(M)$, and the operator norm induced by the norm of $\mathbb{H}$ with $\|M\|$. We use the fact that if $A : \mathbb{H} \to \mathbb{H}$ is s.p.d., then there exists a s.p.d. operator $A^{1/2} : \mathbb{H} \to \mathbb{H}$ with $A^{1/2}A^{1/2} = A$. In the sequel, we will also use the following elementary lemmas, whose proofs are straightforward, and thus omitted here.

LEMMA 3.1. *Let $M, N : \mathbb{H} \to \mathbb{H}$ be linear operators. Then $\sigma(MN) = \sigma(NM)$.*

LEMMA 3.2. *Let A, B be s.p.d. in $\mathbb{H}$. If there exist two positive constants C_1, C_2 such that*

$$C_1(Au,u) \leq (Bu,u) \leq C_2(Au,u) \quad \forall\, u \in \mathbb{H},$$

then the condition number $\mathrm{cond}(A^{-1/2}BA^{-1/2})$ *with respect to the $\mathbb{H}$-norm is bounded*

by C_2/C_1.

4. Analysis of the Left Preconditioner. In this section we show that

$$\widetilde{P}_L := (I+S)^2 S^{-1}$$

is a "good" preconditioner for the symmetric operator $\widetilde{D} = S^{-1} + T$. More precisely, we prove that the condition number of the preconditioned system $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$ is bounded by a constant depending on the coefficient matrices a, b, but is otherwise independent of the meshsize and τ (see Corollary 4.2 below).

The idea behind the preconditioner $\widetilde{P}_L$ can readily be seen in the particular case of $A_{\mathcal{T}} = B_{\mathcal{T}}$, that yields $S = T = \tau A_{\mathcal{T}}$. In this case, since all operators commute, we can write

$$\sigma(\widetilde{P}_L^{-1}\widetilde{D}) = \sigma\Big((I+\tau A_{\mathcal{T}})^{-2}(I+\tau^2 A_{\mathcal{T}}^2)\Big) = \Big\{(1+\tau^2\lambda^2)/(1+\tau\lambda)^2 |\ \lambda \in \sigma(A_{\mathcal{T}})\ \Big\}$$

so that

$$\sigma(\widetilde{P}_L^{-1}\widetilde{D}) = \sigma(\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}) \subset [1/2, 1],$$

whence, the condition number of the preconditioned operator satisfies

$$\operatorname{cond}\big(\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}\big) \le 2.$$

For the general case $A_{\mathcal{T}} \neq B_{\mathcal{T}}$, the lack of commutativity leads to the following slightly weaker results.

THEOREM 4.1. *Let S and T be s.p.d. operators on $\mathbb{H}$, and $0 < \lambda \le \Lambda$ be constants such that*

$$\big((S^{-1}w, w) + (Sw, w)\big) \ge \lambda\big((S^{-1}w, w) + (Tw, w)\big), \tag{4.1}$$

$$\big((S^{-1}w, w) + (Sw, w)\big) \le \Lambda\big((S^{-1}w, w) + (Tw, w)\big), \tag{4.2}$$

for all $w \in \mathbb{H}$. If $P = (I+S)^2 S^{-1}$, then

$$\operatorname{cond}(P^{-1/2}(S^{-1}+T)P^{-1/2}) \le \frac{2\Lambda}{\lambda}.$$

We postpone the proof of the theorem, and first show some implications related to the efficient solution of (1.20).

COROLLARY 4.2 (Symmetric preconditioner). *If the ellipticity assumptions* (1.2) *hold, then the condition number of the preconditioned operator $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$ is uniformly bounded with a constant solely dependent on the eigenvalues λ_a, Λ_a, λ_b, and Λ_b in* (1.2)*, namely,*

$$\operatorname{cond}\big(\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}\big) \le 2\frac{\max\{\Lambda^+, 1\}}{\min\{\lambda^-, 1\}}, \tag{4.3}$$

with $\Lambda^+ := \sup_{x\in\Omega} \Lambda_a(x)/\lambda_b(x), \lambda^- := \inf_{x\in\Omega} \lambda_a(x)/\Lambda_b(x)$.

Proof. Since for all $w \in \mathbb{H}$

$$(A_{\mathcal{T}}w, w) \ge \lambda^-(B_{\mathcal{T}}w, w), \quad (A_{\mathcal{T}}w, w) \le \Lambda^+(B_{\mathcal{T}}w, w),$$

the assertion follows from Theorem 4.1 upon choosing the L^2–inner product $(\cdot,\cdot)$ in $\mathbb{H} = \mathbb{V}_{\mathcal{T}}$ and $\Lambda = \max\{\Lambda^+, 1\}, \lambda = \min\{\lambda^-, 1\}$. ☐

A crucial feature of $\widetilde{P}_L$ is that the resulting preconditioned operator $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$ turns out to be symmetric, which allows the use of preconditioned CG to solve (1.20). To this end, we only need the evaluation of $\widetilde{P}_L^{-1}$ rather than $\widetilde{P}_L^{-1/2}$. As an immediate consequence of Corollary 4.2, and well known results on the convergence rate of the Conjugate Gradient method [34], we obtain the following corollary:

COROLLARY 4.3. *Let the ellipticity assumptions* (1.2) *hold, let u denote the exact solution to $\widetilde{D}u = \widetilde{F}$ with $\widetilde{D} = S^{-1} + T$. Then, the Conjugate Gradient method for solving this system, preconditioned with $\widetilde{P}_L = (I+S)^2 S^{-1}$, generates a sequence of iterates u_k satisfying*

$$|||u - u_k||| \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k |||u - u_0|||$$

with $|||v||| := \big(v, \widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}v\big)^{1/2}$ and

$$\kappa = \operatorname{cond}\big(\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}\big) \leq 2\frac{\max\{\Lambda^+,1\}}{\min\{\lambda^-,1\}},$$

where $\Lambda^+ := \sup_{x\in\Omega}\Lambda_a(x)/\lambda_b(x), \lambda^- := \inf_{x\in\Omega}\lambda_a(x)/\Lambda_b(x)$. That is, the convergence rate in the $\widetilde{P}_L^{-1/2}\widetilde{D}\widetilde{P}_L^{-1/2}$-norm is dictated by the condition number κ, which solely depends on the eigenvalues λ_a, Λ_a, λ_b, and Λ_b in (1.2)*, but is otherwise independent of the discretization parameters $\mathcal{T}$, τ.*

Having established the implications of Theorem 4.1 we now proceed to its proof.

Proof of Theorem 4.1. Observe first that

$$\begin{aligned} P^{-1/2}(S^{-1}+T)P^{-1/2} &= (I+S)^{-1}S^{1/2}(S^{-1}+T)S^{1/2}(I+S)^{-1} \\ &= (I+S)^{-1}(I+S^{1/2}TS^{1/2})(I+S)^{-1} \\ &= \big[(I+S)^2\big]^{-1/2}(I+S^{1/2}TS^{1/2})\big[(I+S)^2\big]^{-1/2}. \end{aligned}$$

Due to Lemma 3.2 it is thus sufficient to show that

$$\frac{1}{2\Lambda}\big((I+S)^2v, v\big) \;\leq\; (v,v) + (S^{1/2}TS^{1/2}v, v) \;\leq\; \frac{1}{\lambda}\big((I+S)^2v, v\big)$$

for all $v \in \mathbb{H}$. To this end we observe that, since S is s.p.d. we have

$$\big((I+S)^2v, v\big) = (v,v) + 2(Sv,v) + (Sv,Sv) \;\geq\; (v,v) + (Sv,Sv) \tag{4.4}$$

and also, by Cauchy-Schwarz's inequality,

$$\big((I+S)^2v, v\big) = (v,v) + 2(Sv,v) + (Sv,Sv) \;\leq\; 2(v,v) + 2(Sv,Sv). \tag{4.5}$$

Setting $w := S^{1/2}v$ and using (4.1) and (4.4) we get

$$\begin{aligned} \big((I+S)^2v, v\big) &\geq (v,v) + (Sv,Sv) = (S^{-1}w, w) + (Sw, w) \\ &\geq \lambda\big((S^{-1}w,w) + (Tw,w)\big) = \lambda\big((v,v) + (S^{1/2}TS^{1/2}v, v)\big). \end{aligned}$$

Similarly, using (4.2) and (4.5), we obtain

$$\begin{aligned} \big((I+S)^2v, v\big) &\leq 2\big((v,v) + (Sv,Sv)\big) = 2\big((S^{-1}w,w) + (Sw,w)\big) \\ &\leq 2\Lambda\big((S^{-1}w,w) + (Tw,w)\big) = 2\Lambda\big((v,v) + (S^{1/2}TS^{1/2}v, v)\big), \end{aligned}$$

and the assertion is proved. ☐

5. Analysis of the Left-Right Preconditioner. Giving up symmetry might seem questionable because preconditioned CG is a very effective method. However, on the one hand our symmetric preconditioners only take into account information arising from one operator $S = \tau A_{\mathcal{T}}$ or $T = \tau B_{\mathcal{T}}$. On the other hand we have found that in many practical examples a non-symmetric preconditioner using both operators S and T is superior; this is the case when $a(x)$ and $b(x)$ are quite different. In this section we study the non-symmetric preconditioned system

$$P_L^{-1} D P_R^{-1} = (I+S)^{-1}(I+ST)(I+T)^{-1}.$$

We first prove that all the eigenvalues of the preconditioned system $(I+S)^{-1}(I+ST)(I+T)^{-1}$ belong to a ball separated from the origin: we show the existence of $\delta_0 \in (0,1)$, independent of space and time discretization parameters, such that all the eigenvalues of the preconditioned system belong to $\{z \in \mathbb{C} : |z - (1+\delta_0)/2| < (1-\delta_0)/2\}$; see Figure 5.1 below. This result hinges on the *structural* assumptions (5.5) and (5.6) on S and T, which are later proved to hold for the operators arising from finite element discretizations in § 5.4. We use this spectral analysis to study the convergence rates of Richardson's method in § 5.2 and GMRes in § 5.3.

5.1. Spectral Analysis. Since the spectrum of $(I+S)^{-1}(I+ST)(I+T)^{-1}$ coincides with those of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ and $(I+ST)(I+T)^{-1}(I+S)^{-1}$ due to Lemma 3.1, the analysis of the spectrum of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ that follows applies to any of these three preconditioned systems.

We start with the following simple observation.

LEMMA 5.1. *If S, T are s.p.d., then*

$$\sigma\big((I+T)^{-1}(I+S)^{-1}(I+ST)\big) = \mathcal{M}\big((\sigma((S+T)^{-1}(I+ST))\big),$$

where $\mathcal{M}$ is the Möbius transformation defined by $\mathcal{M}(z) := \frac{z}{1+z}$.

Proof. We first observe that (e,λ) is an eigenpair of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ if and only if $e \neq 0$ and

$$(I+ST)e = \lambda(I+S)(I+T)e = \lambda(I+ST+S+T)e,$$

which is equivalent to

$$(1-\lambda)(I+ST)e = \lambda(S+T)e.$$

Clearly $\lambda \neq 1$ for otherwise $S+T$ would be singular. Thus (e,λ) is an eigenpair of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ if and only if

$$(I+ST)e = \frac{\lambda}{1-\lambda}(S+T)e,$$

and therefore λ is an eigenvalue of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ if and only if $\mu = \frac{\lambda}{1-\lambda}$ is an eigenvalue of $(S+T)^{-1}(I+ST)$. This holds if and only if $\mathcal{M}(\mu) = \lambda$. ∎

With the help of Lemma 5.1, we will study the eigenvalues of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ through exploring the location of the eigenvalues of $(S+T)^{-1}(I+ST)$.

LEMMA 5.2. *If S, T are s.p.d., then*

$$\operatorname{Re}\sigma((S+T)^{-1}(I+ST)) > 0.$$

Proof. Let (e,μ) be an eigenpair of $(S+T)^{-1}(I+ST)$, then $(I+ST)e = \mu(S+T)e$ and

$$(\mu S - I)e = (S - \mu I)Te;$$

notice that $\mu = x + iy$ with $x, y \in \mathbb{R}$ may not be real. Therefore, for any $f \in \mathbb{H}$,

$$((\mu S - I)e, f) = (Te, (S - \bar{\mu} I)f). \tag{5.1}$$

We now want to choose f fulfilling $(S - \bar{\mu} I)f = e$. If this were not possible, then $\bar{\mu} = \mu > 0$ would be an eigenvalue of S because S is s.p.d.. Thanks to Lemma 5.3 below, this would imply that $\mu = \operatorname{Re}\mu = 1$ and the claim follows.

Otherwise, let $f \in \mathbb{H}$ satisfy

$$(S - \bar{\mu} I)f = e,$$

and rewrite (5.1) in the form

$$\mu(Sf, Sf) - (Sf, f) - \mu\bar{\mu}(Sf, f) + \bar{\mu}(f, f) = (Te, e). \tag{5.2}$$

Taking the real and imaginary parts of (5.2) we get

$$-x^2(Sf, f) + x\big((Sf, Sf) + (f, f)\big) - y^2(Sf, f) - (Sf, f) = (Te, e), \tag{5.3}$$

$$y\big((Sf, Sf) - (f, f)\big) = 0. \tag{5.4}$$

Equation (5.3) can be written as

$$x^2 - x\frac{(Sf, Sf) + (f, f)}{(Sf, f)} + y^2 = -\frac{(Sf, f) + (Te, e)}{(Sf, f)}$$

whence

$$(x - a)^2 + y^2 = r^2$$

with

$$a = \frac{(Sf, Sf) + (f, f)}{2(Sf, f)}, \qquad r^2 = a^2 - \frac{(Sf, f) + (Te, e)}{(Sf, f)}.$$

This shows that μ lies on the boundary of a ball centered at $a > 0$ with radius r, $0 \le r < a$, and thus μ is located in the right complex half plane. ∎

LEMMA 5.3. *Let S, T be s.p.d. and (e,μ) an eigenpair of $(S+T)^{-1}(I+ST)$ such that μ is also an eigenvalue of S. Then, either the equation*

$$(S - \mu I)f = e$$

admits a solution f, or $\mu = 1$.

Proof. Observe first that $\mu \in \sigma(S)$ implies $\mu \in \mathbb{R}_+$, and $e \in \operatorname{rg}(S - \mu I)$ if and only if $e \perp \ker(S - \mu I)^H = \ker(S - \mu I)$ because $\mu \in \mathbb{R}$. Moreover, the fact that (e,μ) is an eigenpair of $(S+T)^{-1}(I+ST)$ implies

$$\frac{1}{\mu}(I + ST)e = (S + T)e.$$

Let now $h \in \ker(S-\mu I)\setminus\{0\}$, i.e. $Sh=\mu h$. Then

$$(h,e)=\frac{1}{\mu}(Sh,e)=\frac{1}{\mu}(h,Se)=\frac{1}{\mu}(h,\frac{1}{\mu}(I+ST)e-Te),$$

whence

$$(h,e)=\frac{1}{\mu^2}(h,e)-\frac{1}{\mu}(h,Te)+\frac{1}{\mu^2}(h,STe).$$

Upon multiplying by μ^2 and rearranging terms we arrive at

$$(\mu^2-1)(h,e)=(h,STe)-\mu(h,Te)=(Sh,Te)-\mu(h,Te)=\mu(h,Te)-\mu(h,Te)=0.$$

This statement is the assertion in disguise. □

To get more uniform bounds on the eigenvalues we have to make two *structural* assumptions. They will turn out to be quite natural in the applications with FEM of § 5.4.

Assumption 1: *There exists a constant $\alpha>0$ such that*

$$(Tg,g)\geq\alpha(Sg,g)\quad \textit{for all } g\in\mathbb{H}. \tag{5.5}$$

Assumption 2: *There exists a constant $K>0$ such that*

$$\|S^{-1}T\|\leq K. \tag{5.6}$$

THEOREM 5.4. *Let S, T be s.p.d. and let Assumptions 1, 2 hold. Let*

$$b:=\frac{2\sqrt{\alpha}}{1+K},\quad c:=\frac{2\min\{\alpha,1\}}{1+\alpha}. \tag{5.7}$$

There exists a constant $a\geq c$ such that if $r=a\sqrt{1-c/a}$ then

$$\sigma((S+T)^{-1}(I+ST))\ \subseteq\ \overline{B(a,r)}\cup\{x\in\mathbb{R}\mid x\geq b\}\cup\{1\}\subset\mathbb{C}_+.$$

Proof. We will refine the argument of proof of Lemma 5.2. We adopt the notation of such proof and estimate the right-hand side of (5.2) using (5.5) as follows:

$$\begin{aligned}(Te,e)\geq\alpha(Se,e)&=\alpha\big(S(S-\bar\mu I)f,(S-\bar\mu I)f\big)\\&=\alpha\big[(S^2f,Sf)-(S^2f,\bar\mu f)-(\bar\mu Sf,Sf)+(\bar\mu Sf,\bar\mu f)\big]\\&=\alpha\big[(S^2f,Sf)-2x(Sf,Sf)+(x^2+y^2)(Sf,f)\big],\end{aligned}$$

where $\mu=x+iy$. Combining this estimate with (5.3) we get

$$\begin{aligned}&-\big[(Sf,f)+\alpha(S^2f,Sf)\big]\\&\qquad\geq x^2(1+\alpha)(Sf,f)-2x\big[\alpha(Sf,Sf)+\frac{(f,f)+(Sf,Sf)}{2}\big]+y^2(1+\alpha)(Sf,f).\end{aligned}$$

We first examine the case that $y\neq 0$. Using (5.4) and normalizing f, namely taking $(f,f)=(Sf,Sf)=1$, which does not restrict generality, the above estimate reads

$$x^2(1+\alpha)(Sf,f)-2x(1+\alpha)+y^2(1+\alpha)(Sf,f)\leq-\big[(Sf,f)+\alpha(S^2f,Sf)\big].$$

Reordering and completing the square, we thus infer

$$(x-\tilde{a}_\mu)^2+y^2 \le \tilde{a}_\mu^2 - \tilde{a}_\mu \frac{(Sf,f)+\alpha(S^2f,Sf)}{1+\alpha} =: \tilde{r}^2, \tag{5.8}$$

with $\tilde{a}_\mu = 1/(Sf,f) > 0$. Note that by construction the right-hand side of (5.8) is non-negative. To proceed further, we need to estimate $\tilde{r}$ in terms of $\tilde{a}_\mu$.

Since S is s.p.d. we can write

$$\begin{aligned} 1 = (Sf,Sf) = (S^{3/2}f, S^{1/2}f) &\le \frac{1}{2}\big[(S^{3/2}f,S^{3/2}f)+(S^{1/2}f,S^{1/2}f)\big] \\ = \frac{1}{2}\big[(Sf,f)+(S^2f,Sf)\big] &\le \frac{1}{2\min\{\alpha,1\}}\big[(Sf,f)+\alpha(S^2f,Sf)\big], \end{aligned} \tag{5.9}$$

whence

$$\frac{(Sf,f)+\alpha(S^2f,Sf)}{1+\alpha} \ge \frac{2\min\{\alpha,1\}}{1+\alpha} = c$$

with $c > 0$ depending only on α, but not on μ, and $c \le \tilde{a}_\mu$. We thus conclude that if $\mu \in \sigma\big((S+T)^{-1}(I+ST)\big)$ and $\mu \notin \mathbb{R}$, then there exists $\tilde{a}_\mu > 0$ such that $\mu \in \overline{B(\tilde{a}_\mu, \tilde{a}_\mu\sqrt{1-c/\tilde{a}_\mu})}$. Therefore, if $a := \max\{\tilde{a}_\mu \mid \mu \in \sigma\big((S+T)^{-1}(I+ST)\big)\setminus\mathbb{R}\}$, Lemma 5.5 below yields

$$\sigma\big((S+T)^{-1}(I+ST)\big)\setminus\mathbb{R} \subset \overline{B(a, a\sqrt{1-c/a})},$$

and proves the theorem provided $y \neq 0$. It remains to examine the case $y = 0$.

We thus consider pure real eigenvalues $\mu \in \mathbb{R}$. We already know from Lemma 5.2 that $\mu > 0$. Let us re-start the analysis for this case from the equality

$$\mu(S+T)e = (I+ST)e.$$

Multiplying by $g := S^{-1}e$ we get

$$\mu = \frac{((I+ST)e,g)}{(Se,g)+(Te,g)} = \frac{(S^{-1}e,e)+(Te,e)}{(e,e)+(S^{-1}Te,e)}.$$

Making use of Assumption 1 and arguing as in (5.9), the numerator becomes

$$(S^{-1}e,e)+(Te,e) \ge (S^{-1}e,e)+\alpha(Se,e) \ge 2\big(S^{-1/2}e, \sqrt{\alpha}S^{1/2}e\big) = 2\sqrt{\alpha}\,(e,e).$$

On the other hand, employing Assumption 2, the denominator is bounded from above by

$$(e,e)+(S^{-1}Te,e) \le (1+K)\,(e,e).$$

Therefore, we finally conclude $\mu \ge \frac{2\sqrt{\alpha}}{1+K} = b$, which is the asserted estimate. □

LEMMA 5.5. *Let $0 < c < a_1 < a_2$, and let $r_i = a_i\sqrt{1-c/a_i}$ for $i = 1,2$. Then*

$$B(a_1,r_1) \subset B(a_2,r_2).$$

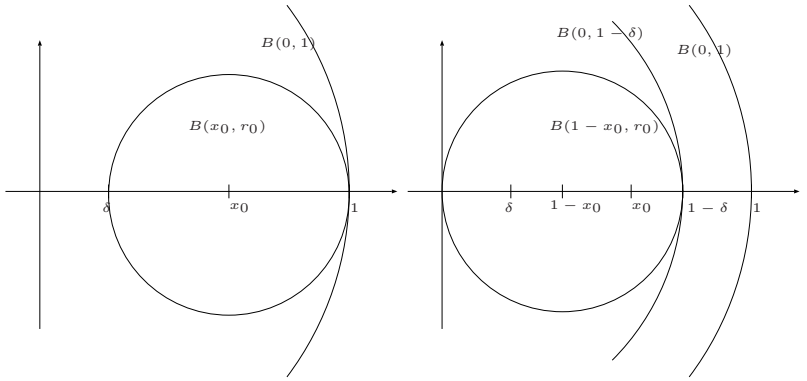

FIG. 5.1. *Location of the spectrum of* $(I+T)^{-1}(I+S)^{-1}(I+ST)$ *(left) and Richardson iteration operator (right).*

Proof. Observe first that for $i = 1, 2$

$$(x,y) \in B(a_i, r_i) \quad \Longleftrightarrow \quad (x-a_i)^2 + y^2 \le a_i^2 - ca_i \quad \Longleftrightarrow \quad x^2 + y^2 \le (2x-c)a_i.$$

This readily implies $2x \ge c$, whence

$$(x,y) \in B(a_1, r_1) \quad \Rightarrow \quad x^2 + y^2 \le (2x-c)a_1 \le (2x-c)a_2 \quad \Rightarrow \quad (x,y) \in B(a_2, r_2),$$

and the proof is complete. ∎

Theorem 5.4 enables us to bound uniformly the spectrum of $(I+T)^{-1}(I+S)^{-1}(I+ST)$.

THEOREM 5.6 (Uniform Spectral Bound). *Let S, T be s.p.d. and Assumptions 1, 2 hold. Let*

$$\delta_0 := \min\Big(\frac{2\sqrt{\alpha}}{1+K+2\sqrt{\alpha}}, \frac{\min(1,\alpha)}{1+\alpha+\min(1,\alpha)}, \frac{1}{2}\Big),$$

and $0 < x_0 := (1+\delta_0)/2$ *and* $r_0 := (1-\delta_0)/2$ *(see Figure 5.1). Then*

$$\sigma\big((I+T)^{-1}(I+S)^{-1}(I+ST)\big) \;\subseteq\; \overline{B(x_0,r_0)} \;\subset \mathbb{C}_+.$$

Proof. According to Lemma 5.1 and Theorem 5.4 the spectrum of $(I+T)^{-1}(I+S)^{-1}(I+ST)$ satisfies

$$\begin{aligned}\sigma\big((I+T)^{-1}(I+S)^{-1}(I+ST)\big) &= \mathcal{M}\big((\sigma((S+T)^{-1}(I+ST))\big) \\ &\subseteq \mathcal{M}\big(\overline{B(a,r)} \cup \{x \in \mathbb{R} \mid x \ge b\} \cup \{1\}\big),\end{aligned}$$

where $\mathcal{M}(z) = \frac{z}{1+z}$ is a Möbius transformation. We first observe that $\mathcal{M}(1) = 1/2$ and $\mathcal{M}(\{x \mid x \ge b\}) \subseteq [\frac{b}{1+b}, 1]$, whence

$$\mathcal{M}\big(\{x \in \mathbb{R} \mid x \ge b\} \cup \{1\}\big) \subset \big[\min\big\{\frac{b}{1+b}, \frac{1}{2}\big\}, 1\big], \qquad \text{with } b = \frac{2\sqrt{\alpha}}{1+K}. \tag{5.10}$$

It remains to find a ball $B(x_0, r_0)$ containing $\mathcal{M}\big(B(a,r)\big)$. We notice that $\mathcal{M}$ maps $B(a,r)$ onto a ball in the complex plane with center in the real axis, the latter due

to symmetry. Thus $\mathcal{M}(B(a,r))$ is determined by the extremal points $\zeta := \mathcal{M}(a-r)$ and $\eta := \mathcal{M}(a+r)$. Since $a-r \geq \inf_{a\geq c}\big(a-a\sqrt{1-c/a}\big) = c/2$ we deduce $\zeta \geq \mathcal{M}(\frac{c}{2})$ whereas $\eta \leq 1$. Therefore,

$$\mathcal{M}\big(\overline{B(a,r)}\big) \subset B\big(\frac{1+\mathcal{M}(\frac{c}{2})}{2}, \frac{1-\mathcal{M}(\frac{c}{2})}{2}\big) \qquad \text{with } c = \frac{2\min\{\alpha,1\}}{1+\alpha}. \tag{5.11}$$

Choosing δ_0, x_0 and r_0 as in the statement of the theorem, the assertion follows from (5.10) and (5.11). □

By Lemma 3.1 the assertion of Theorem 5.6 for the preconditioned operator $(I+T)^{-1}(I+S)^{-1}(I+ST)$ also holds for the other two non-symmetric preconditioned systems $(I+S)^{-1}(I+ST)(I+T)^{-1}$ and $(I+ST)(I+T)^{-1}(I+S)^{-1}$. In what follows we comment on two popular iterative methods for linear systems that greatly benefit from these preconditioners, focusing on the Left-Right preconditioned system $(I+S)^{-1}(I+ST)(I+T)^{-1}$, but emphasizing that the same results hold for the other two options that we have mentioned.

5.2. Richardson's Method. One of the simplest iterative methods to solve equation

$$(I+ST)u = f$$

is Richardson's method, which in our case takes the following form: given an initial guess u_0, define $v_0 = P_R u_0$ and compute the iterates by the recurrence

$$v^{k+1} = v^k - P_L^{-1}\big((I+ST)P_R^{-1}v^k - f\big), \qquad u^{k+1} = P_R^{-1}v^{k+1}, \tag{5.12}$$

where $P_L = (I+T)$ and $P_R = (I+S)$ are the left and right preconditioners, respectively. Notice that the computation $u^{k+1} = P_R^{-1}v^{k+1}$ should only be performed upon convergence, and not in every iteration.

COROLLARY 5.7. *Let S, T be s.p.d. and Assumptions 1, 2 hold. If R is the matrix representation of the preconditioned system $P_L^{-1}(I+ST)P_R^{-1}$, then Richardson's iteration* (5.12) *converges in any norm with an* asymptotic *linear convergence rate of $1-\delta_0$, where δ_0 is given in Theorem 5.6.*

Proof. If $Q = I-R$ denotes the Richardson iteration matrix, its spectrum satisfies $\sigma(Q) = 1-\sigma(R)$. Therefore, the spectral radius $\rho(Q)$ satisfies

$$\rho(Q) = \max\big\{|\lambda| : \lambda \in \sigma(Q)\big\} \leq 1-\delta_0 < 1,$$

which follows immediately from Theorem 5.6 (see Figure 5.1). Given any vector norm $\|\cdot\|$, the corresponding subordinate matrix norm verifies $\rho(Q) = \lim_{k\to\infty}\|Q^k\|^{1/k}$. On the other hand, since $v - v^k = Q^k(v-v^0)$ with $v = P_R^{-1}u$, we see that

$$\lim_{k\to\infty}\frac{\|v-v^k\|}{\rho(Q)^k\|v-v^0\|} \leq \lim_{k\to\infty}\frac{\|Q^k\|}{\rho(Q)^k} = 1.$$

This implies the asserted asymptotic behavior in any norm. □

REMARK 5.8. Since the iteration matrix Q is not symmetric (or normal), the above *asymptotic* convergence rate does not necessarily yield an *error reduction rate* between consecutive iterates.

5.3. GMRes Method. A standard and widely used method for non-symmetric systems is the Generalized Minimum Residual (GMRes) method [42]. Solving the system $(I + ST)u = f$ via left-right preconditioned GMRes consists of solving the equivalent system $\left[P_L^{-1}(I + ST)P_R^{-1}\right]v = P_L^{-1}f$ and $P_R u = v$, by performing the following steps:

1. `Modify the right-hand side: Let` $b = P_L^{-1}f$.
2. `Solve with GMRes the system` $\left[P_L^{-1}(I + ST)P_R^{-1}\right]v = b$.
3. `Update the solution, solving` $P_R u = v$.

Throughout this section, $\|\cdot\|$ will denote the 2-norm for vectors as well as its subordinate matrix norm.

To solve $R\,v = b$, GMRes constructs the unique sequence $\{v_k\}$ with

$$v_k \in v_0 + \text{span}\{r_0, Rr_0, R^2r_0, \ldots, R^{k-1}r_0\}$$

and $\|r_k\|$ minimal, where $r_k := R\,v_k - b$ is the residual. As a consequence of this [39], at each step of the iteration we have

$$r_k = p_k(R)r_0,$$

with p_k a polynomial of degree k with $p_k(0) = 1$. This in turn implies that

$$\frac{\|r_k\|}{\|r_0\|} = \min_{p_k} \|p_k(R)\|, \tag{5.13}$$

where the minimum is taken over all polynomials p_k of degree k that satisfy $p_k(0) = 1$. There are many ways to bound the right-hand side of (5.13), none of them is sharp, and they depend on further properties of R such as R being diagonalizable or normal. Since we do not know that those properties hold for our case, we resort to the general concept of pseudospectrum [47]: *Given $\epsilon > 0$, the ϵ-pseudospectrum $\sigma_\epsilon(R)$ of R is the set of ϵ-eigenvalues of R, namely those $z \in \mathbb{C}$ that are eigenvalues of some matrix $R + E$ with $\|E\| \le \epsilon$.* We will make use of the following result [47, Theorem 26.2].

LEMMA 5.9 (Pseudospectrum estimate). *Let Σ_ϵ be a union of closed curves enclosing the ϵ-pseudospectrum $\sigma_\epsilon(R)$ of R. Then for any polynomial p_k of degree k with $p_k(0) = 1$ we have*

$$\max_{z\in\sigma(R)} |p_k(z)| \le \|p_k(R)\| \le \frac{L_\epsilon}{2\pi\epsilon} \max_{z\in\Sigma_\epsilon(R)} |p_k(z)|, \tag{5.14}$$

where L_ϵ is the arclength of Σ_ϵ.

In order to apply this result we need to control the location of the ϵ-pseudospectrum of R. A result in this direction is given by the following

PROPOSITION 5.10 (Bound on the ϵ-pseudospectrum). *If R is a square matrix of order n and $0 < \epsilon \le 1$, then*

$$\sigma_\epsilon(R) \subset \bigcup_{\lambda\in\sigma(R)} B(\lambda, C_R\,\epsilon^{1/m}),$$

where $C_R := n(1 + \sqrt{n-p})\|V^{-1}\|\|V\|$, with V a nonsingular matrix transforming R into its Jordan canonical form J, i.e. $V^{-1}RV = J$, p is the number of Jordan blocks, and m is the size of the largest Jordan block of R.

Proof. Let $0 < \epsilon \le 1$ be given, and let $z \in \sigma_\epsilon(R)$. By definition there exists a matrix E with $\|E\| \le \epsilon$ such that $z \in \sigma(R + E)$. Corollary 2.2 in [46] states a bound

for the distance between the eigenvalues of two matrices in terms of the norm of the matrix difference. Applied to our case it reads

$$\min_{\lambda\in\sigma(R)} |\lambda - z| \le \sqrt{n}(1+\sqrt{n-p}) \max\left\{ \sqrt{n}\big\|V^{-1}EV\big\|,\ \sqrt[m]{\sqrt{n}\|V^{-1}EV\|} \right\}, \quad (5.15)$$

where V, p, m are as in the statement of the proposition. Since $\|E\| \le \epsilon \le 1$, (5.15) implies the existence of an eigenvalue λ of R such that

$$|z-\lambda| \le n(1+\sqrt{n-p})\|V^{-1}\|\|V\|\epsilon^{1/m},$$

and the claim follows. □

COROLLARY 5.11. *Let S, T be s.p.d. and Assumptions 1, 2 hold. If R is the matrix representation of the preconditioned system $P_L^{-1}(I+ST)P_R^{-1}$, then GMRes' iteration converges with an* asymptotic *linear convergence rate bounded by*

$$\rho := \frac{1-\frac{1}{2}\delta_0}{1+\delta_0},$$

where δ_0 is given in Theorem 5.6. Moreover,

$$\frac{\|r_k\|}{\|r_0\|} \le C_0\rho^k,$$

with $C_0 := C_R^m 2^{2m-1}\dim\mathbb{H}/\delta_0^{m-1}$ and C_R as in Proposition 5.10.

Proof. Let $\epsilon_0 = \frac{1}{2}\delta_0 > 0$ and let C_R and m denote the constants from Proposition 5.10. Then choosing $\epsilon = \left(\frac{\epsilon_0}{2C_R}\right)^m$ so that $C_R\epsilon^{1/m} = \epsilon_0/2$ we have

$$\sigma_\epsilon(R) \subset \bigcup_{\lambda\in\sigma(R)} B\big(\lambda, \frac{\epsilon_0}{2}\big) \subset \overline{B\big(x_0, \frac{1-\delta_0}{2} + \frac{\epsilon_0}{2}\big)}$$

by Theorem 5.6, with $x_0 = \frac{1}{2}(1+\delta_0)$. If we consider $p_k(z) = x_0^{-k}(x_0 - z)^k$ and observe that

$$|p_k(z)| = \frac{|z-x_0|^k}{x_0^k} \le \left(\frac{\frac{1-\delta_0}{2} + \frac{\epsilon_0}{2}}{\frac{1+\delta_0}{2}}\right)^k \le \left(\frac{1-\frac{1}{2}\delta_0}{1+\delta_0}\right)^k, \qquad \forall z\in\sigma_\epsilon(R),$$

then equality (5.13) and bound (5.14) with L_ϵ the arclength of $\Sigma_\epsilon = \bigcup_{\lambda\in\sigma(R)} \partial B(\lambda,\epsilon_0)$ imply the assertion. □

REMARK 5.12. Corollary 5.11 is not fully satisfactory because the constant C_0 still depends on the matrix R, which in turn depends on space and time discretization parameters. A finer analysis leading to uniform bounds would be desirable but requires tools from matrix theory for non-normal matrices to characterize the pseudospectrum of R, that do not seem to be available.

5.4. Application to Finite Element Discretizations. We now discuss Assumptions 1 and 2 within the finite element context and restate Corollaries 5.7 and 5.11. Let $\mathcal{T}$ be a shape regular triangulation of Ω, let $\mathbb{H} = \mathbb{V}_\mathcal{T}$ be the C^0 Lagrange finite element space of fixed degree (not necessarily one), with inner product $(v,w) = \int_\Omega vw$ and norm $\|v\| = \left(\int_\Omega |v|^2\right)^{1/2}$, which are well defined for all $v, w \in L^2(\Omega)$. Recall that $S = \tau A_\mathcal{T}$ and $T = \tau B_\mathcal{T}$ with $A_\mathcal{T}$ and $B_\mathcal{T}$ as in (1.17).

Assumption 1 is a direct consequence of the ellipticity condition (1.2) and is fulfilled with

$$\alpha = \inf_{x\in\Omega} \frac{\lambda_b(x)}{\Lambda_a(x)}.$$

To verify Assumption 2 we adopt the procedure in [33], and so compare the operators $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$. To this end, we need the following *compatibility condition* between the matrices a and b: there exists a scalar function η, piecewise C^1 over the mesh $\mathcal{T}$ of Ω such that

$$b(x) = \eta(x)a(x) \quad \text{a.e. } x \in \Omega. \tag{5.16}$$

It is worth observing that in the case of the coefficients a and b being scalar valued and piecewise smooth, the compatibility condition holds.

For any $u, v \in \mathbb{V}_{\mathcal{T}}$, if $I = I_{\mathcal{T}}$ denotes the Lagrange interpolation operator in $\mathbb{V}_{\mathcal{T}}$, by definition (1.17) we have

$$(Bu, v) = (b\nabla u, \nabla v) = (a\nabla I(\eta u), \nabla v) + d(u, v) = (AI(\eta u), v) + d(u, v),$$

where

$$d(u, v) := (a\big\{\eta\nabla u - \nabla I(\eta\, u)\big\}, \nabla v).$$

Assumption 2 holds as a consequence of the following simple lemmas.

LEMMA 5.13. *Let the ellipticity conditions (1.2) and the compatibility condition (5.16) hold. Then there exists a constant $C > 0$ depending on the shape regularity of $\mathcal{T}$ and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$ such that*

$$|d(u, v)| \le C\Lambda_a \|\nabla\eta\|_{L^\infty(\Omega,\mathcal{T})} \|u\| \, \|\nabla v\| \quad \forall u, v \in \mathbb{V}_{\mathcal{T}},$$

where $\|\nabla\eta\|_{L^\infty(\Omega,\mathcal{T})}$ denotes the broken *norm*

$$\|\nabla\eta\|_{L^\infty(\Omega,\mathcal{T})} := \max_{T\in\mathcal{T}} \|\nabla\eta\|_{L^\infty(T)}.$$

Proof. Let $E : \Omega \to \mathbb{R}$ be defined as

$$E(x) := \eta(x)\nabla u(x) - \nabla I(\eta u)(x) = \sum_i \big(\eta(x) - \eta(x_i)\big) u(x_i)\nabla\phi_i(x).$$

Recall that $\{\phi_i\}_i$ denotes the nodal basis of the C^0 Lagrange finite element space $\mathbb{V}_{\mathcal{T}}$ (see Section 2). Since

$$|d(u, v)| = |(a\, E, \nabla v)| \le \Lambda_a \|\nabla v\| \, \|E\|,$$

it remains to bound $\|E\|$. For any $T \in \mathcal{T}$, denoting with h_T the diameter of T, standard scaling arguments lead to

$$\begin{aligned}
\|E\|^2_{L^2(T)} &\le \sup_{x,y\in T} |\eta(x) - \eta(y)|^2 C_1 \int_T \sum_i \big|u(x_i)\nabla\phi_i(x)\big|^2 \\
&\le h_T^2 \|\nabla\eta\|^2_{L^\infty(T)} \, C_1C_2 \|\nabla u\|^2_{L^2(T)} \\
&\le h_T^2 \|\nabla\eta\|^2_{L^\infty(T)} \frac{C_1C_2C_3}{h_T^2} \|u\|^2_{L^2(T)}, = C_1C_2C_3 \|\nabla\eta\|^2_{L^\infty(T)} \|u\|^2_{L^2(T)},
\end{aligned}$$

where C_1, C_2, C_3 are constants depending only on shape regularity and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$. Adding over all $T \in \mathcal{T}$ we obtain the claim of the lemma. □

LEMMA 5.14. *There exists a constant $C > 0$, depending only on the shape regularity of $\mathcal{T}$ and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$ such that*

$$\|I(\eta\, u)\| \le C\|\eta\|_{L^\infty(\Omega)}\|u\|, \qquad \forall u \in \mathbb{V}_{\mathcal{T}}.$$

Proof. Standard scaling arguments allow us to conclude that for each element $T \in \mathcal{T}$

$$\|I(\eta u)\|^2_{L^2(T)} \le C^2\|\eta\|^2_{L^\infty(T)}\|u\|^2_{L^2(T)},$$

with C depending only on shape regularity of $\mathcal{T}$ and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$. The claim follows by adding over $T \in \mathcal{T}$. □

We are now in a position to prove Assumption 2. Let $\Lambda_P = \Lambda_P(\Omega)$ be the Poincaré constant of Ω, namely

$$\|v\|_{L^2(\Omega)} \le \Lambda_P\|\nabla v\|_{L^2(\Omega)}, \qquad \forall v \in \mathbb{V}, \tag{5.17}$$

where $\mathbb{V}$ is the subset of $H^1(\Omega)$ that incorporates the essential boundary conditions (1.13)–(1.14) provided $H^{d-1}(\Gamma_D) > 0$, or imposes a vanishing meanvalue (1.12) otherwise.

PROPOSITION 5.15. *Under the conditions of ellipticity (1.2) and compatibility (5.16) Assumption 2 holds, i.e.,*

$$\|S^{-1}T\| = \|A_{\mathcal{T}}^{-1}B_{\mathcal{T}}\| \le K := C_0 \max\left\{\|\eta\|_{L^\infty(\Omega)}, \|\operatorname{grad}\eta\|_{L^\infty(\Omega,\mathcal{T})}\right\}\left(1 + \frac{\Lambda_a}{\lambda_a}\Lambda_P\right).$$

with a constant $C_0 > 0$ depending on the shape regularity of $\mathcal{T}$ and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$, and Λ_P defined in (5.17).

Proof. By definition $S^{-1}T = A_{\mathcal{T}}^{-1}B_{\mathcal{T}}$ and

$$\|A_{\mathcal{T}}^{-1}B_{\mathcal{T}}u\| = \sup_{0\neq v\in\mathbb{V}_{\mathcal{T}}}\frac{(A_{\mathcal{T}}^{-1}B_{\mathcal{T}}u, v)}{\|v\|} = \sup_{0\neq v\in\mathbb{V}_{\mathcal{T}}}\frac{(B_{\mathcal{T}}u, A_{\mathcal{T}}^{-1}v)}{\|v\|}.$$

Let $w = A_{\mathcal{T}}^{-1}v$ and use Lemma 5.13 to obtain

$$(B_{\mathcal{T}}u, A_{\mathcal{T}}^{-1}v) = (B_{\mathcal{T}}u, w) = (b\nabla u, \nabla w) = (a\nabla I(\eta u), \nabla w) + d(u, w).$$

Invoking definition (1.17) of A implies

$$(a\nabla I(\eta u), \nabla w) = (A_{\mathcal{T}}w, I(\eta u)) = (I(\eta u), v). \tag{5.18}$$

In view of Lemmas 5.13 and 5.14 we infer that

$$\begin{aligned}|(B_{\mathcal{T}}u, A_{\mathcal{T}}^{-1}v)| &\le |(I(\eta u), v)| + |d(u, w)|\\ &\le C_0\|u\|\big(\|\eta\|_{L^\infty(\Omega)}\|v\| + \Lambda_a\|\operatorname{grad}\eta\|_{L^\infty(\Omega)}\|\nabla w\|\big),\end{aligned} \tag{5.19}$$

with the constant C depending on the shape regularity of $\mathcal{T}$ and the polynomial degree of the finite element space $\mathbb{V}_{\mathcal{T}}$. We now claim that $\|\nabla w\| \le \frac{\Lambda_P}{\lambda_a}\|v\|$. To see this, note that

$$\begin{aligned}\lambda_a\|\nabla w\|^2 \le (a\nabla w, \nabla w) &= (A_{\mathcal{T}}w, w) = (A_{\mathcal{T}}A_{\mathcal{T}}^{-1}v, w)\\ &= (v, w) \le \|v\|\|w\| \le \Lambda_P\|v\|\|\operatorname{grad} w\|,\end{aligned}$$

by Poincaré inequality (5.17) using that $w \in \mathbb{V}$. Now (5.19) reads

$$|(B_{\mathcal{T}} u, A_{\mathcal{T}}^{-1} v)| \leq C_0 \|u\| \|v\| \big(\|\eta\|_{L^\infty(\Omega)} + \Lambda_a \| \operatorname{grad} \eta \|_{L^\infty(\Omega)} \frac{\Lambda_P}{\lambda_a} \big),$$

and the claim of the proposition follows due to (5.18). ☐

COROLLARY 5.16. *Let the ellipticity and compatibility conditions (1.2) and (5.16) be fulfilled, let $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$ denote the discrete operators from (1.17) and $S = \tau A_{\mathcal{T}}$ and $T = \tau B_{\mathcal{T}}$, and let $R = P_L^{-1}(I + ST)P_R^{-1}$ be the Left-Right preconditioned system. Then both Richardson and GMRes iterations converge with an* asymptotic *linear convergence rate. If δ_0 is as in Theorem 5.6, the rate is $1 - \delta_0$ in any norm for the former whereas it is $\frac{1-\frac{1}{2}\delta_0}{1+\delta_0}$ in the 2-norm for the latter.*

Proof. The result is now a simple consequence of Corollaries 5.7 and 5.11 and the above considerations concerning Assumptions 1 and 2. ☐

6. Application to Surface Diffusion. We now discuss the performance of the proposed preconditioners when applied to the nonlinear system (1.4), first linearized via semi-implicit time stepping. We consider the evolution by surface diffusion of the graph plotted in Figure 6.1 (left), with periodic boundary conditions. We use linear finite elements on three nested uniform meshes:

$$\begin{aligned} &8192 \text{ DOFs } (h = 2^{-5} \approx 0.031), \\ &32768 \text{ DOFs } (h = 2^{-6} \approx 0.016), \\ &131072 \text{ DOFs } (h = 2^{-7} \approx 0.0078). \end{aligned}$$

In Figure 6.1 we plot the evolving surface to illustrate that it is rough at the beginning, and slowly regularizes, thereby leading to very high values of $q(u)$ in (1.4), which in turn imply very small values for the coefficient functions a and b in this example. The situation depicted is very similar to the *nasty* example.

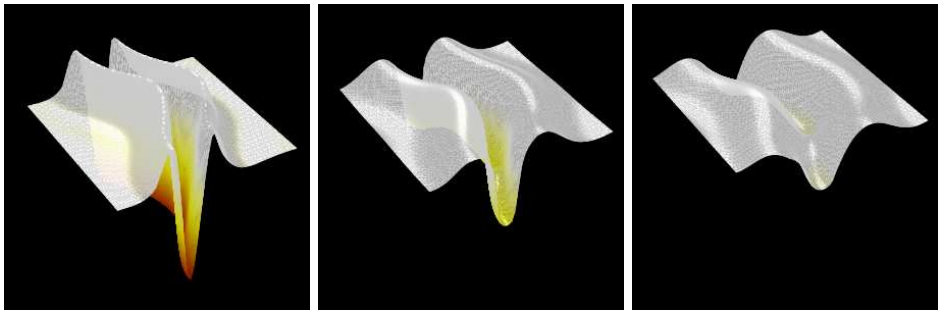

FIG. 6.1. *Evolution of an initially rough graph by surface diffusion governed by the system* (1.4). *Solution at $t = 0$ (left), $t = 0.0005$ (middle), $t = 0.0010$ (right)*

In Figure 6.2 we display the number of iterations versus time (for three different time step sizes, respectively) needed for each preconditioner. The performance of the Left Preconditioner, at least in the initial phase of the evolution, where the surface is still rough, is not great. This is caused by the rather big discrepancy of operators S and T due to (locally) high values of $q(u)$. Interestingly, the Right Preconditioner behaves much better in this situation. The behavior of the Left-Right Preconditioner is again striking. The number of iterations is always between 1 and 4, even in the rough initial phase of the evolution.

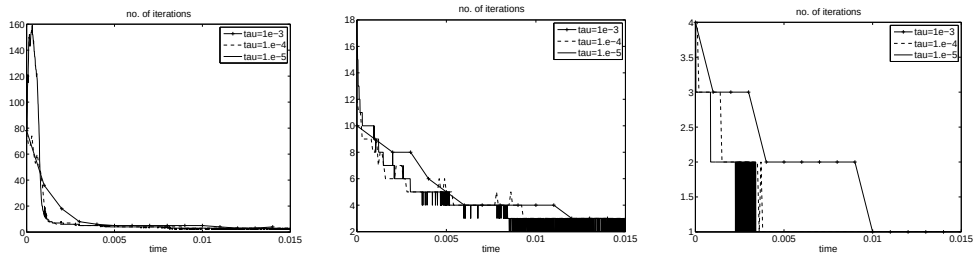

FIG. 6.2. *Number of iterations versus time for the example of surface diffusion of Fig. 6.1: Left Precondtioner (left), Right Preconditioner (center) and Left-Right Preconditioner (right).*

7. Comparisons and Conclusions. The discussion in § 4 and § 5 about the Left and the Left-Right preconditioners indicates that both work well but stops short of displaying which method works best. This is not obvious in terms of condition numbers and spectral radii.

Departure from normality prevents us from estimating theoretically the error reduction rate $\sqrt{\rho(Q^*Q)}$ in the L^2-norm for Richardson's iteration; see Remark 5.8. Nonetheless, we investigate this issue computationally and find the results displayed in Figure 7.1: $\sqrt{\rho(Q^*Q)}$ is approximately the same as $\rho(Q)$, an amazing fact that deserves further research.

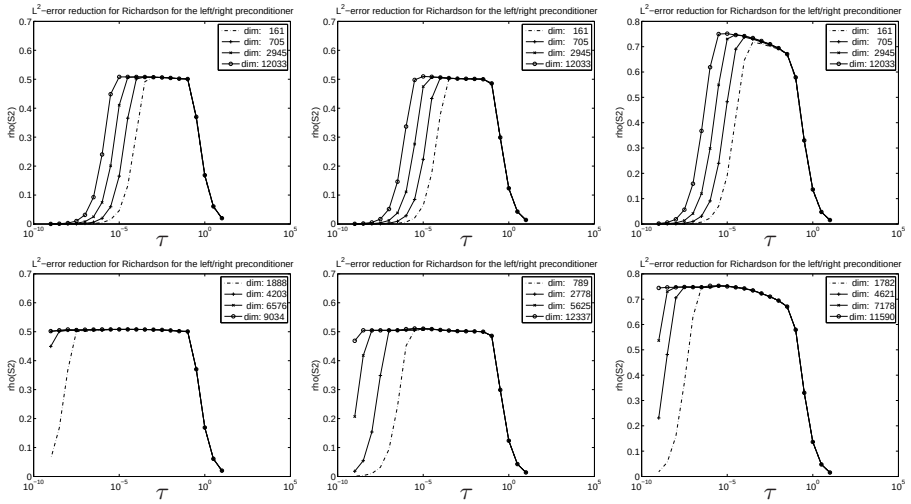

FIG. 7.1. *L^2–error reduction rate $\sqrt{\rho(Q^*Q)}$ for Richardson's iteration versus τ for the left/right non-symmetrically preconditioned system. The plots correspond to Examples 2.1–3 with uniform meshes (top) and adaptive meshes (bottom).*

We embark now on a more systematic comparison of the performance of the two methods. We examine preconditioned CG (PCG) and GMRes (S-GMRes) for the (symmetric) Left Preconditioner as well as Richardson and GMRes for the (non-symmetric) Left-Right Preconditioner. Since the overall computational cost is by far dominated by the evaluation of the preconditioned operator, we report on the number of iterations as an indicator of performance. We hereby assume that the cost per evaluation is comparable for both variants of the preconditioner.

To this end we chose the forcing functions $f = 1, g = 0$ in (1.16) and started all iterations with $u_{\mathcal{T}}^0 := 0$. For a fair comparison we avoid dealing with the stopping tests within MATLAB because they are based on different residual norms for different methods. Instead, we devise an outer loop that terminates at iteration k provided

$$||u_{\mathcal{T}}^k - u_{\mathcal{T}}||_{L^2(\Omega)} \leq 1e\text{-}7,$$

where $u_{\mathcal{T}}$ is a discrete solution obtained by imposing a very sharp stopping criterion to CG. The computational results for the finest partition are displayed in Figure 7.2 (uniform mesh) and Figure 7.3 (graded mesh). For completeness we point out that for coarser meshes the results are essentially the same.

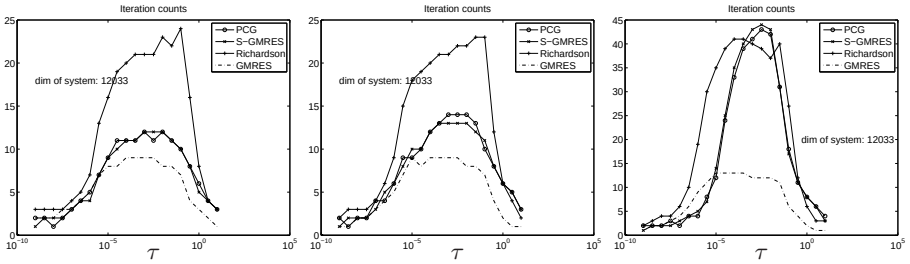

FIG. 7.2. *Number of iterations vs τ for the symmetric and non-symmetric preconditioners on uniform meshes for Examples 2.1–3. Iterative methods: PCG (preconditioned CG) and S-GMRes (GMRes) for the symmetrically preconditioned system; GMRes and Richardson for the non-symmetrically preconditioned system.*

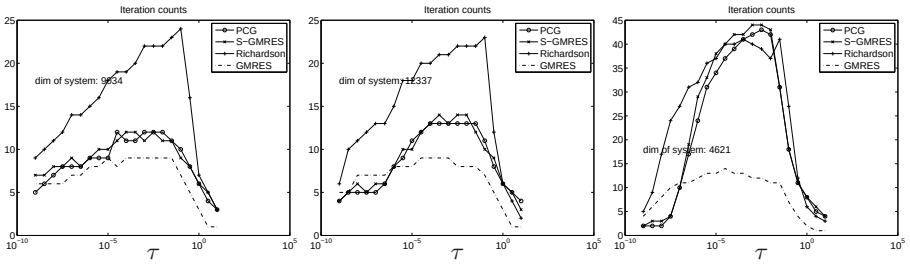

FIG. 7.3. *Number of iterations vs τ for the symmetric and non-symmetric preconditioners on graded meshes for Examples 2.1–3. Iterative methods: PCG (preconditioned CG) and S-GMRes (GMRes) for the symmetrically preconditioned system; GMRes and Richardson for the non-symmetrically preconditioned system.*

The simulations refer to Examples 2.1-3 of § 2. Our findings are as follows:

- For the Left Preconditioner the agreement between computational results of Figure 2.1 and the theoretical upper bound (4.3) is excellent, even though the latter is a bit pessimistic (by a factor 2 in the *nasty* example): according to (4.3) we have

$$(a)\ \ \kappa \leq 4, \qquad (b)\ \ \kappa \leq 8.4, \qquad (c)\ \ \kappa \leq 86.666.$$

- Although the three non-symmetric preconditioned systems mentioned at the beginning of § 5.1 possess the same spectra, their performance within, say, a Krylov space method might be different. Our experiments indicate that it is generally better for the Left-Right preconditioned matrix $R = P_L^{-1} D P_R$. The difference in performance

increases as the coefficient matrices a and b are more dissimilar, as in Examples 2.3 (*nasty*) and 2.4 (*degenerate*).

- The simple Richardson method is worst in most cases, but only by a small factor of 3–5 in terms of number of iterations.
- CG and GMRes for the Left Preconditioner behave very similarly.
- GMRes for the Left-Right preconditioned system achieves the best performance in almost all cases. The comparison is most favorable for the most difficult Examples 2.3 (*nasty*) and 2.4 (*degenerate*).
- The performance of GMRes for the Left-Right preconditioned system is rather robust with respect to the difficulty of the underlying problem.
- For *degenerate* operators, the behavior of conditions numbers, spectral radii and thus iteration counts for most of the iterative methods deteriorate; see Figs. 7.4 and 7.5. This is an indication that our assumption on uniform ellipticity is somewhat sharp. Amazingly, GMRes for the Left-Right preconditioned system performs still reasonably well also in this case.

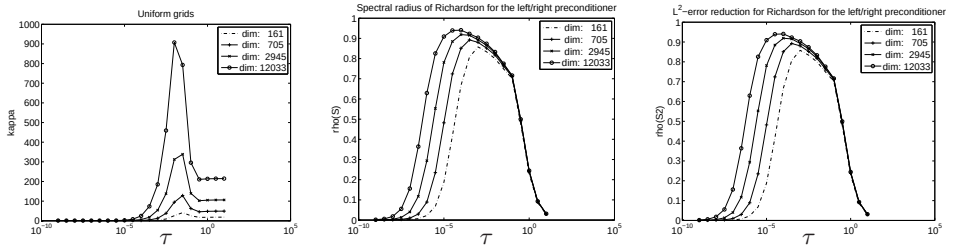

FIG. 7.4. *Example 2.4, degenerate case: condition numbers versus τ for the symmetrically preconditioned systems (left); spectral radius $\rho(Q)$ for Richardson's iteration operator Q for the left/right non-symmetrically preconditioned system (middle); L^2–error reduction rate $\sqrt{\rho(Q^*Q)}$ (right); uniform grids.*

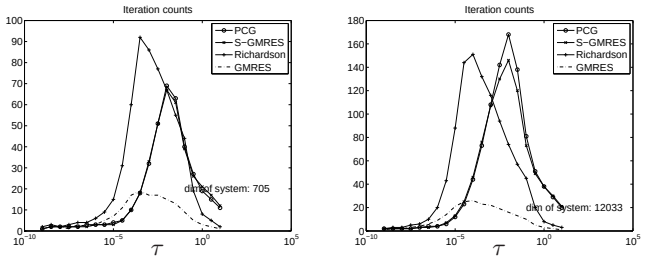

FIG. 7.5. *Number of iterations vs τ for the symmetric and non-symmetric preconditioners on two different uniform meshes for the degenerate Examples 2.4. Iterative methods: PCG (preconditioned CG) and S-GMRes (GMRes) for the symmetrically preconditioned system; GMRes and Richardson for the non-symmetrically preconditioned system.*

- For *surface diffusion* of graphs, our motivating geometric PDE, the Left-Right preconditioner outperforms the other two; see § 6.

On the basis of these experiments, we conclude that it is advisable to use the Left-Right Preconditioner with GMRes, especially when both operators $A_{\mathcal{T}}$ and $B_{\mathcal{T}}$ differ considerably. This makes use of both operators and exhibits the best performance overall.

REFERENCES

[1] B. Aksoylu and M. Holst, *Optimality of multilevel preconditioners for local mesh refinement in three dimensions*, SIAM J. Numer. Anal., 44 (2006), pp. 1005–1025.

[2] F. Bai, C.M. Elliott, A. Gardiner, A. Spence, and A.M. Stuart, *The viscous Cahn-Hilliard equation. I. Computations*, Nonlinearity 8 (1995), 131–160.

[3] E. Bänsch, P. Morin, and R.H. Nochetto, *Finite element methods for surface diffusion*, In: P. Colli, C. Verdi, A. Visintin (eds.), Free Boundary Problems. International Series of Num. Math., vol. 147, 53–63, Birkhuser (2003).

[4] E. Bänsch, P. Morin, R. H. Nochetto, *Surface Diffusion of Graphs: Variational Formulation, Error Analysis, and Simulation.* SIAM J. Numer. Anal. 42 (2004), 773–799.

[5] E. Bänsch, P. Morin, R. H. Nochetto, *Finite element methods for surface diffusion: the parametric case.* J. Comp. Physics 203 (2005), 321–343.

[6] J. Barrett, H. Garcke, and R. Nürnberg, *A parametric finite element method for fourth order geometric evolution equations*, J. Comput. Phys. 222 (2007), 441–467.

[7] J. Barrett, H. Garcke, and R. Nürnberg, *On the variational approximation of combined second and fourth order geometric evolution equations*, SIAM J. Sci. Comput. 29 (2007), 1006–1041.

[8] J.W. Barrett, H. Garcke and R. Nürnberg, *Parametric approximation of Willmore flow and related geometric evolution equations*, SIAM J. Sci. Comput. 31 (2008), 225–253.

[9] J. W. Barrett, R. Nurnberg and V. Styles *Finite element approximation of a phase field model for void electromigration*, SIAM J. Numer. Anal. 42, (2004), 738–772.

[10] S. Bartels, G. Dolzmann, and R.H. Nochetto, *A finite element scheme for the evolution of orientational order in fluid membranes* Model. Math. Anal. Numer. 44 (2010), 1–32.

[11] A.J. Bernoff, A.L. Bertozzi, and T.P. Witelski, *Axisymmetric surface diffusion: dynamics and stability of self-similar pinchoff*, J. Statist. Phys. 93 (1998), 725–776.

[12] A.L. Bertozzi and M.G. Pugh, *Long-wave instabilities and saturation in thin film equations*, Comm. Pure Appl. Math. 51 (1998), 625–661.

[13] J.F. Blowey and C.M. Elliott, *The Cahn-Hilliard gradient theory for phase separation with nonsmooth free energy. II. Numerical analysis*, European J. Appl. Math. 3 (1992), 147–179.

[14] J. Becker, G. Grün, M. Lenz, and M. Rumpf, *Numerical methods for fourth order nonlinear degenerate diffusion problems*, Mathematical theory in fluid mechanics (Paseky, 2001). Appl. Math. 47 (2002), no. 6, 517–543.

[15] P. Bjørstad, *Fast numerical solution of the biharmonic Dirichlet problem on rectangles*, SIAM J. Numer. Anal. 20 (1983), 59–71.

[16] P. Bjørstad and B.P. Tjøstheim, *Efficient algorithms for solving a fourth-order equation with the spectral-Galerkin method*, SIAM J. Sci. Comput. 18 (1997), 621–632.

[17] A. Bonito, R.H. Nochetto, and M.S. Pauletti *Parametric FEM for geometric biomembranes*, J. Comp. Phys. 229 (2010), 3171–3188.

[18] F. Bornemann, *An adaptive multilevel approach to parabolic problems. III. 2D error estimation and multilevel preconditioning,* IMPACT of Computing in Science and Engineering 4 (1992), no. 1, 1–45.

[19] D. Braess and P. Peisker, *On the numerical solution of the biharmonic equation and the role of squaring matrices for preconditioning*, IMA J. Numer. Anal. 6 (1986), 393–404.

[20] J. H. Bramble, *Multigrid Methods*, vol. 294 of Pitman Research Notes in Mathematical Sciences, Longman Scientific & Technical, Essex, England, 1993.

[21] J. H. Bramble, J. E. Pasciak, and J. Xu, *Parallel multilevel preconditioners*, Math. Comp., 55 (1990), pp. 1–22.

[22] J. Bramble, J. Pasciak, and P.S. Vassilevski, *Computational scales of Sobolev norms with application to preconditioning*, Math. Comp. 69 (2000), 463–480.

[23] J. H. Bramble and X. Zhang, *The analysis of multigrid methods*, in Handbook of numerical analysis, Vol. VII, North-Holland, Amsterdam, 2000, pp. 173–415.

[24] J.W. Cahn, C.M. Elliott, and A. Novick-Cohen, *The Cahn-Hilliard equation with a concentration dependent mobility: motion by minus the Laplacian of the mean curvature*, European J. Appl. Math. 7 (1996), 287–301.

[25] U. Clarenz, U. Diewald, G. Dziuk, M. Rumpf, and R. Rusu, *A finite element method for surface restoration with smooth boundary conditions*, Comput. Aided Geom. Design 21 (2004), 427-445.

[26] L. Chen and C.-S. Zhang, *AFEM@matlab: a Matlab package of adaptive finite element methods*, Technique Report, Department of Mathematics, University of Maryland at College Park, (2006).

[27] K. Deckelnick, G. Dziuk, and C.M. Elliott, *Computation of geometric partial differential*

equations and mean curvature flow, Acta Numer. 14 (2005), 139–232.
[28] Q. Du, *A phase field formulation of the Willmore problem*, Nonlinearity 18 (2005), 1249-1267.
[29] Q. Du and X. Wang, *Modelling and simulations of multi-component lipid membranes and open membranes via diffuse interface approaches* J. Math. Bio., 2007.
[30] Q. Du, M. Li, and C. Liu, *Analysis of a phase field Navier-Stokes vesicle-fluid interaction model*, Disc. Cont. Dyn. Sys. B, 8 (2007), 539-556.
[31] G. Dziuk, *An algorithm for evolutionary surfaces*, Numer. Math. 58, (1991) 603-611.
[32] G. Dziuk, *Computational parametric Willmore flow*, Numer. Math. 111 (2008), 55–80.
[33] C.I. Goldstein, T.A. Manteuffel, S.V. Parter, *Preconditioning and Boundary Condtions without H_2 Estimates: L_2 Condition Numbers and the Distribution of the Singular Values,* SIAM J. Numer. Anal. **30**, no. 2 (1993), 343–376.
[34] G. H. Golub and Ch. van Loan, *Matrix computations*, 2nd edition, J. Hopkins University Press, (1989).
[35] G. Grün, *On the convergence of entropy consistent schemes for lubrication type equations in multiple space dimensions*, Math. Comp. 72 (2003), 1251–1279.
[36] G. Grün and M. Rumpf, *Nonnegativity preserving convergent schemes for the thin film equation*, Numer. Math. 87 (2000), 113–152.
[37] W. Hackbusch, *Multigrid Methods and Applications*, vol. 4 of Computational Mathematics, Springer–Verlag, Berlin, (1985).
[38] M. R. Hanisch, *Multigrid preconditioning for mixed finite element methods*. PhD thesis. Cornell University (1991).
[39] N. L. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl. 13 (1992), 778–795.
[40] P. Oswald, *Multilevel Finite Element Approximation, Theory and Applications*, Teubner Skripten zur Numerik, Teubner Verlag, Stuttgart, (1994).
[41] R.E. Rusu, *An algorithm for the elastic flow of surfaces*, Interfaces Free Bound. 7 (2005), 229-239.
[42] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second Edition, SIAM, 2003.
[43] A. Schmidt and K. G. Siebert, *Design of Adaptive Finite Element Software: The Finite Element Toolbox ALBERTA*, Springer LNCSE Series 42 (2005).
[44] D. Silvester and M.D. Mihajlović, *A black-box multigrid preconditioner for the biharmonic equation*, BIT 44 (2004), 151–163.
[45] D. Silvester and M.D. Mihajlović, *Efficient parallel solvers for the biharmonic equation*, Parallel Comput. 30 (2004), 35–55.
[46] Y. Song, *A note on the variation of the spectrum of an arbitrary matrix*, Linear Algebra Appl. 342 (2002), 41–46.
[47] L. N. Trefethen and M. Embree, *Spectra and pseudospectra. The behavior of nonnormal matrices and operators*, Princeton University Press (2005).
[48] O. B. Widlund, *Some Schwarz methods for symmetric and nonsymmetric elliptic problems*, in Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations, D. E. Keyes, T. F. Chan, G. A. Meurant, J. S. Scroggs, and R. G. Voigt, eds., Philadelphia, 1992, SIAM, 19–36.
[49] H. Wu and Z. Chen, *Uniform convergence of multigrid v-cycle on adaptively refined finite element meshes for second order elliptic problems*, Science in China: Series A Mathematics, 49 (2006), 1–28.
[50] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Review, 34 (1992), 581–613.
[51] J. Xu, L. Chen, and R. H. Nochetto, *Optimal multilevel methods for $H(grad)$, $H(curl)$, and $H(div)$ systems on graded and unstructured grids*, in *Multiscale, Nonlinear and Adaptive Approximation*, R. DeVore and A. Kunoth eds, Springer (2009),599-659.
[52] J. Xu and L. Zikatanov, *The method of alternating projections and the method of subspace corrections in Hilbert space*, J. Amer. Math. Soc., 15 (2002), 573–597.
[53] H. Yserentant, *Old and new convergence proofs for multigrid methods*, Acta Numer., (1993), 285–326.
[54] X. Zhang, *Multilevel Schwarz methods*, Numer. Math. 63, (1992), 521–539.