ESAIM: M2AN 49 (2015) 839–854
DOI: 10.1051/m2an/2014058

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

AN EFFECTIVE PRECONDITIONER FOR A PML SYSTEM FOR ELECTROMAGNETIC SCATTERING PROBLEM

Qiya Hu[1,*], Chunmei Liu[2,**], Shi Shu[2,***] and Jun Zou[3,****]

Abstract. In this work we are concerned with an efficient numerical solution of a perfectly matched layer (PML) system for a Maxwell scattering problem. The PML system is discretized by the edge finite element method, resulting in a symmetric but indefinite complex algebraic system. When the real and imaginary parts are considered independently, the complex algebraic system can be further transformed into a real generalized saddle-point system with some special structure. Based on an crucial observation to its Schur complement, we construct a symmetric and positive definite block diagonal preconditioner for the saddle-point system. Numerical experiments are presented to demonstrate the effectiveness and robustness of the new preconditioner.

Mathematics Subject Classification. 35Q60, 65E05, 78A45, 78M10.

Received July 19, 2013. Revised July 11, 2014.
Published online 8 April 2015.

1. Introduce

Perfectly matched layer (PML) is a popular and effective technique for truncating an unbounded domain where an electromagnetic wave scattering problem is defined. It was proposed by Bérenger [2] for the time-dependent Maxwell equations, in an intention to construct a fictitious layer outside the "region of interest" so

Keywords and phrases. Maxwell scattering problem, edge finite elements, PML equations, Schur complement-type preconditioner.

* *This author was supported by the Major Research Plan of Natural Science Foundation of China G91130015, The Key Project of Natural Science Foundation of China G11031006 and National Basic Research Program of China G2011309702.*

** *This author was supported by the Scientific Research Fund of the Hunan Provincial Education Department of China Grant 12A138, Specialized research Fund for the Doctoral Program of Higher Education of China Grant 20124301110003 and NSFC Projects 11201159 and 11426102.*

*** *This author was supported by NSFC Project 91130002 and 11171281.*

**** *This author was substantially supported by Hong Kong RGC grant (Project 404611).*

[1] LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China. hqy@lsec.cc.ac.cn

[2] Department of Mathematics and Computational Science, Hunan University of Science and Engineering, Hunan, 425199, P.R. China. liuchunmei0629@163.com

[3] School of Mathematics and Computational Science, Xiangtan University, Hunan, 411105, P.R. China. shushi@xtu.edu.cn

[4] Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. zou@math.cuhk.edu.hk

that plane waves entering the layer can be well absorbed. This approach was developed in terms of a complex change of variables (or stretching) for the Maxwell problems in frequency domain; see, *e.g.*, [8, 9, 11, 12, 17, 18]. The well-posedness of the PML system was studied, and the convergence of the PML solution to the one of the original Maxwell system in the infinite domain was also verified, provided that the truncated domain is sufficiently large [8,11]. Furthermore, it was also showed in [8,11] that the solution to the original Maxwell system in infinite domain is preserved by the PML system in the computational domain while it decays exponentially in the PML layer. Efficient preconditioned iterative methods have been well developed for the elliptic-type Maxwell's equations, with nearly optimal convergence rate independent of the mesh size; see, *e.g.*, [13, 15, 19] and the references therein. However, not much has been done in literature on fast solvers for a PML system of the electromagnetic wave scattering problem when it is discretized by the edge finite elements. We are aware of only the recent work [21] and two works by Botros and Volakis, who presented the GMRES solver coupled with an approximate inverse preconditioner [6], and proposed some optimal choices of the PML parameters and tested the numerical performance of the GMRES solver with these parameters [7].

In this work we shall study some fast algorithms for solving the PML system of a Maxwell scattering problem. As we see, the stiffness matrix of the discrete system resulting from the edge element discretization of the PML system is complex, and Hermitian but indefinite. When the real and imaginary parts are considered independently, the complex algebraic system can be further transformed into a real generalized saddle-point system with a symmetric but indefinite 2×2 block coefficient matrix. This coefficient matrix is essentially different from the coefficient matrices of the standard saddle-point systems arising from, *e.g.*, the Maxwell system or Navier−Stokes equations: the two diagonal blocks are now both indefinite; the diagonal and off-diagonal blocks are both obtained from some indefinite **curl curl**-type second order differential operators. We shall propose a 2×2 block diagonal positive definite preconditioner, whose two diagonal blocks can be viewed as some regularizations of the diagonal blocks of the coefficient matrix and the Schur complement corresponding to the saddle-point system. This construction of the preconditioner is based on a crucial observation on the Schur complement, which has a very complicated structure. Numerical experiments will be presented to demonstrate the effectiveness and robustness of the new preconditioner.

The rest of the paper is organized as follows. In Section 2, we will introduce the PML equations for a Maxwell scattering problem and present some approximation results about the PML equations. We shall discuss in Section 3 the edge finite element discretization of the PML system, then construct two preconditioners for the discrete edge element system, and provide some analysis on the preconditioners based on a general framework in Section 4. Finally, we present some numerical examples to illustrate the competitive behavior of the preconditioners in Section 5.

2. PML system and its convergence

For the ease of notation, we shall restrict all our discussions from now on in two dimensions, but all the results can be naturally extended to three dimensions. The theoretical results about the PML system to be introduced in this section were demonstrated only for three dimensions in [8], but are also true for two dimensions (see Sect. 7, [8]). So we shall cite these results directly from [8] below.

Let $\Omega_0 \subset \mathbb{R}^2$ be a bounded domain containing the origin with a boundary Γ_0, and Ω_0^c the complement of its closure. Then for a sufficiently large positive constant L, we define three domains Ω_1, Ω_2 and Ω_3 (see Fig. 1):

$$\Omega_1 = (-a,a)^2 \setminus \bar{\Omega}_0, \quad \Omega_2 = (-b,b)^2 \setminus [-a,a]^2, \quad \Omega_3 = (-L,L)^2 \setminus [-b,b]^2,$$

and set $\Omega_L = \Omega_1 \cup \Omega_2 \cup \Omega_3$, with Γ_L being its boundary.

For a vector-valued function $\mathbf{v} = (v_1, v_2)^t$ and a scalar function g, we shall adopt the following conventional definitions:

$$\nabla \times \mathbf{v} = \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}; \quad \boldsymbol{\nabla} \times g = \left(\frac{\partial g}{\partial y}, -\frac{\partial g}{\partial x}\right)^t.$$

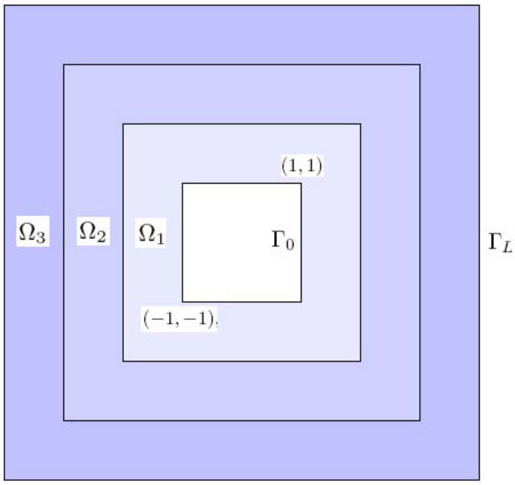


FIGURE 1. The bounded domain Ω_L.

In this work, we shall consider the following electromagnetic wave scattering problem by the impenetrable scatterer Ω_0 [8]:

$$\begin{cases} \boldsymbol{\nabla} \times (\mu^{-1}\nabla \times \boldsymbol{u}) - k^2\varepsilon \boldsymbol{u} = \quad \boldsymbol{0} \quad \text{in } \Omega_0^c, \\ \qquad\qquad\qquad \boldsymbol{u}\cdot\boldsymbol{\tau} = \boldsymbol{g}\cdot\boldsymbol{\tau} \text{ on } \Gamma_0 \end{cases} \tag{2.1}$$

where $\boldsymbol{\tau}$ is unit tangential vector on Γ_0, μ and ε are the magnetic permeability and the electric permittivity, respectively, k is the wavenumber and $\boldsymbol{g}$ is the trace of a function $\widetilde{\boldsymbol{g}} \in \boldsymbol{H}_{loc}(\mathbf{curl}; \overline{\Omega^c})$ on Γ_0, where $\boldsymbol{H}_{loc}(\mathbf{curl}; \overline{\Omega^c})$ denotes the set of functions on Ω^c whose restrictions to $\Omega^c \cap D$ are in $H(\mathbf{curl}; \Omega^c \cap D)$ for any bounded domain D. The system (2.1) is often complemented by the Silver−Müller radiation condition to select the physically interested outgoing waves [8]:

$$\lim_{r\to\infty} r^{1/2}(\nabla \times \boldsymbol{u} - ik\boldsymbol{u}\cdot\tau) = 0, \tag{2.2}$$

where r is the magnitude of the position vector (x, y).

It is easy to see that the system (2.1) reduces to the following one when the medium is homogeneous, *i.e.*, $\mu = \varepsilon = 1$,

$$\begin{cases} \boldsymbol{\nabla} \times (\nabla \times \boldsymbol{u}) - k^2 \boldsymbol{u} = \quad \boldsymbol{0} \quad \text{in } \Omega_0^c, \\ \qquad\qquad \boldsymbol{u}\cdot\boldsymbol{n} = \boldsymbol{g}\cdot\boldsymbol{n} \text{ on } \Gamma_0. \end{cases} \tag{2.3}$$

Next we introduce the PML approximation of the scattering system (2.1) [8]. For this aim, we define an even function $\sigma \in C^0(\mathbb{R})$:

$$\sigma(t) = \begin{cases} 0 & \text{for } |t| \le a, \\ \sigma_0 \frac{(|t|-a)}{b-a} & \text{for } a < |t| < b, \\ \sigma_0 & \text{for } |t| \ge b \end{cases} \tag{2.4}$$

for a positive constant σ_0 (the PML strength), and the two functions

$$d(t) = 1 + z\,\sigma(t), \quad \widetilde{\sigma}(r) = \frac{1}{r}\int_0^r \sigma(t)\mathrm{d}t \tag{2.5}$$

where z is a complex number. For $z = i$ or $1 + i$, we introduce the following complex stretching

$$T(x, y) = ((1 + z\,\widetilde{\sigma}(x))x, (1 + z\,\widetilde{\sigma}(y))y) \equiv (\widetilde{x}, \widetilde{y}), \tag{2.6}$$

which is a transformation between the rectangular coordinate system (x, y) and the complex coordinate system $(\widetilde{x}, \widetilde{y})$.

Noting that Maxwell equations is invariant in different coordinates [16], system (2.1) changes to the following equivalence system after the complex stretching T in (2.6):

$$\begin{cases} \widetilde{\boldsymbol{\nabla}} \times \mu^{-1}(\widetilde{\nabla} \times \widetilde{\boldsymbol{u}}) - k^2\varepsilon\widetilde{\boldsymbol{u}} = \mathbf{0} & \text{in } \Omega_0^c, \\ \widetilde{\boldsymbol{u}} \cdot \boldsymbol{\tau} = \widetilde{\boldsymbol{g}} \cdot \boldsymbol{\tau} & \text{on } \Gamma_0, \\ \lim\limits_{\widetilde{r} \to \infty} \widetilde{r}^{1/2}(\widetilde{\nabla} \times \widetilde{\boldsymbol{u}} - ik\widetilde{\boldsymbol{u}} \cdot \boldsymbol{\tau}) = 0 \end{cases} \tag{2.7}$$

where two operators $\widetilde{\nabla}\times$ and $\widetilde{\boldsymbol{\nabla}}\times$ in the complex coordinate system $(\widetilde{x}, \widetilde{y})$ are connected respectively with two operators $\nabla\times$ and $\boldsymbol{\nabla}\times$ in the rectangular coordinate system (x, y) as follows:

$$\widetilde{\nabla} \times \boldsymbol{w} = J^{-1}\nabla \times (\boldsymbol{B}\boldsymbol{w}), \quad \widetilde{\boldsymbol{\nabla}} \times w = \boldsymbol{A}\boldsymbol{\nabla} \times w, \tag{2.8}$$

where J, $\boldsymbol{A}$ and $\boldsymbol{B}$ are given by

$$J = \mathrm{d}(x)\mathrm{d}(y), \quad \boldsymbol{A} = \mathrm{diag}(1/\mathrm{d}(y), 1/\mathrm{d}(x)), \quad \boldsymbol{B} = \mathrm{diag}\left(\mathrm{d}(x), \mathrm{d}(y)\right).$$

Now we introduce a Hilbert space on the unbounded domain Ω_0^c:

$$\widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl}; \Omega_0^c) \equiv \{\boldsymbol{\theta} : \boldsymbol{B}\boldsymbol{\theta} \in \boldsymbol{H}_{loc}(\mathbf{curl}; \Omega_0^c),\ \boldsymbol{\theta} \cdot \tau = \widetilde{\boldsymbol{g}} \cdot \boldsymbol{\tau} \text{ on } \Gamma_0\},$$

and two bilinear forms:

$$[\boldsymbol{v}, \boldsymbol{w}]_{\Omega_0^c} = \int_{\Omega_0^c} J\boldsymbol{v} \cdot \boldsymbol{w}\mathrm{d}\boldsymbol{x}, \quad [\varphi, \psi]_{\Omega_0^c} = \int_{\Omega_0^c} J\varphi\psi\mathrm{d}\boldsymbol{x}. \tag{2.9}$$

Then for any $\widetilde{w} \in \widehat{H}_{loc}(\mathbf{curl}; \Omega_0^c)$ and $\boldsymbol{v} \in \widehat{\boldsymbol{H}}_0(\mathbf{curl}; \Omega_0^c)$, we have the Green's formula

$$[\widetilde{\boldsymbol{\nabla}} \times \widetilde{w}, \boldsymbol{v}]_{\Omega_0^c} = [\widetilde{w}, \widetilde{\nabla} \times \boldsymbol{v}]_{\Omega_0^c}. \tag{2.10}$$

Using this integration by parts formula, we derive the variational formulation of (2.7):

Find $\widetilde{\boldsymbol{u}} \in \widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl}; \Omega_0^c)$ such that

$$\widetilde{A}(\widetilde{\boldsymbol{u}}, \boldsymbol{\phi}) = \mathbf{0} \quad \forall\ \boldsymbol{\phi} \in \widehat{\boldsymbol{H}}_0(\mathbf{curl}; \Omega_0^c), \tag{2.11}$$

where the bilinear form $\widetilde{A}(\cdot, \cdot)$ is given by

$$\widetilde{A}(\widetilde{\boldsymbol{u}}, \boldsymbol{\phi}) = [\mu^{-1}\widetilde{\nabla} \times \widetilde{\boldsymbol{u}}, \widetilde{\nabla} \times \boldsymbol{\phi}]_{\Omega_0^c} - [k^2\varepsilon\widetilde{\boldsymbol{u}}, \boldsymbol{\phi}]_{\Omega_0^c}. \tag{2.12}$$

Next we will present a few results on the estimates and convergence of the solution to the variational system (2.11), all under the following condition:

$$z = i \ \text{ or } \ z = 1 + i \ \text{ and } \ \arg(1 + i\sigma_0) < \frac{\pi}{3}\cdot \tag{2.13}$$

We have the following well-posedness for the PML scattering problem (2.11) [8].

Corollary 2.1. *Under the condition* (2.13)*, there is a unique* $\widetilde{\boldsymbol{u}} \in \widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl}; \Omega_0^c)$ *to the system* (2.11)*, and the solution* $\widetilde{\boldsymbol{u}}$ *coincides with the solution* $\boldsymbol{u}$ *to* (2.3) *in* Ω_1*. And there are constants* $C > 0$ *and* $\alpha > 0$ *such that the following stability estimate holds for* $L \geq b$*:*

$$\|\widetilde{\boldsymbol{u}}\|_{\boldsymbol{H}^{1/2}(\Gamma_L)} \leq C\mathrm{e}^{-\alpha kL}\|\widetilde{\boldsymbol{g}}\|_{\boldsymbol{H}(\mathbf{curl};\ \Omega_L)}.$$

Note that the model problem (2.11) is defined on an unbounded domain Ω_0^c, which is inconvenient for numerical computations. Now we truncate Ω_0^c by the bounded domain Ω_L, and accordingly introduce the following Sobolev spaces on Ω_L:

$$\widehat{\boldsymbol{H}}(\mathbf{curl};\Omega_L) \equiv \{\boldsymbol{\theta} : \boldsymbol{B\theta} \in \boldsymbol{H}(\mathbf{curl};\Omega_L)\},$$

$$\widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl};\Omega_L) \equiv \{\boldsymbol{\theta} : \boldsymbol{B\theta} \in \boldsymbol{H}(\mathbf{curl};\Omega_L),\ \boldsymbol{\theta}\cdot\boldsymbol{\tau} = \widetilde{\boldsymbol{g}}\cdot\boldsymbol{\tau} \text{ on } \Gamma_0 \text{ and } \boldsymbol{\theta}\cdot\boldsymbol{\tau} = 0 \text{ on } \Gamma_L\}.$$

Then the approximate variational problem of (2.11) on the bounded domain Ω_L can be formulated as follows: Find $\widetilde{\boldsymbol{u}}_L \in \widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl};\Omega_L)$ such that

$$\widetilde{A}_L(\widetilde{\boldsymbol{u}}_L,\boldsymbol{\phi}) = \mathbf{0} \quad \forall\ \boldsymbol{\phi} \in \widehat{\boldsymbol{H}}_{\mathbf{0}}(\mathbf{curl};\Omega_L), \tag{2.14}$$

where $\widetilde{A}_L(\cdot,\cdot)$ is given by

$$\widetilde{A}_L(\widetilde{\boldsymbol{u}}_L,\boldsymbol{\phi}) = [\mu^{-1}\widetilde{\nabla}\times\widetilde{\boldsymbol{u}}_L, \widetilde{\nabla}\times\boldsymbol{\phi}]_{\Omega_L} - [k^2\varepsilon\widetilde{\boldsymbol{u}}_L,\boldsymbol{\phi}]_{\Omega_L}. \tag{2.15}$$

The system (2.14) has a unique solution provided that L is large enough; see ([8], Thm. 6.1), and this solution converges to the PML solution on Ω_L exponentially.

Corollary 2.2 (*cf.* [8], Thm. 6.2)**.** *Under the condition* (2.13)*, there is a positive constant L_0 such that for $L \geq L_0$,*

$$\|\widetilde{\boldsymbol{u}} - \widetilde{\boldsymbol{u}}_L\|_{\widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl};\ \Omega_L)} \leq C(L_0)\mathrm{e}^{-\alpha kL}\|\widetilde{\boldsymbol{g}}\|_{\boldsymbol{H}(\mathbf{curl};\ \Omega_L)}.$$

Using Corollaries 2.1 and 2.2, we have the following approximation theory.

Theorem 2.3. *Under condition* (2.13)*, the solution $\widetilde{\boldsymbol{u}}_L$ to the PML system* (2.14) *approximates the solution $\boldsymbol{u}$ to the system* (2.3) *in Ω_1 exponentially.:*

$$\|\boldsymbol{u} - \widetilde{\boldsymbol{u}}_L\|_{\widehat{\boldsymbol{H}}_{\widetilde{\boldsymbol{g}}}(\mathbf{curl};\ \Omega_1)} \leq C(L_0)\mathrm{e}^{-\alpha kL}\|\widetilde{\boldsymbol{g}}\|_{\boldsymbol{H}(\mathbf{curl};\ \Omega_L)}. \tag{2.16}$$

Noting that the curl operator $\widetilde{\nabla}\times$ in equation (2.14) is inconvenient for numerical computations, we will transform it to the common curl operator $\nabla\times$ on Ω_L. For this purpose, we introduce a complex Sobolev space:

$$\mathbf{H}_{\boldsymbol{g}}(\mathrm{curl};\Omega_L) = \{\boldsymbol{u} = \boldsymbol{u}_r + i\boldsymbol{u}_i\ :\ \boldsymbol{u}_r \in H_{\boldsymbol{g}_r}(\mathrm{curl};\Omega_L), \boldsymbol{u}_i \in H_{\boldsymbol{g}_i}(\mathrm{curl};\Omega_L)\},$$

where $\boldsymbol{g}_r$ and $\boldsymbol{g}_i$ are the real and imaginary part of function $\boldsymbol{g}$, and (with $l = r$ or i)

$$H_{\boldsymbol{g}_l}(\mathrm{curl};\Omega_L) = \{\boldsymbol{v} :\ \boldsymbol{v} \in H(\mathrm{curl};\Omega_L),\ \boldsymbol{v}\cdot\boldsymbol{\tau}|_{\Gamma_L} = 0,\ \boldsymbol{v}\cdot\boldsymbol{\tau}|_{\Gamma_0} = \boldsymbol{g}_l\cdot\boldsymbol{\tau}|_{\Gamma_0}\}.$$

Using the above space and (2.8), we can write (2.14) equivalently as follows:
Find $\boldsymbol{u}_L \in \mathbf{H}_{\boldsymbol{g}}(\mathrm{curl};\Omega_L)$ such that

$$a(\boldsymbol{u}_L,\boldsymbol{\psi}) = \mathbf{0} \quad \forall\ \boldsymbol{\psi} \in \mathbf{H}_0(\mathrm{curl};\Omega_L), \tag{2.17}$$

where $a(\cdot,\cdot)$ is given by

$$a(\boldsymbol{u}_L,\boldsymbol{\psi}) = \int_{\Omega_L} J^{-1}(\nabla\times\boldsymbol{u}_L)(\overline{\nabla\times\boldsymbol{\psi}})\mathrm{d}\boldsymbol{x} - k^2\int_{\Omega_L}((\boldsymbol{AB})^{-1}\boldsymbol{u}_L)\cdot\overline{\boldsymbol{\psi}}\mathrm{d}\boldsymbol{x}. \tag{2.18}$$

Without loss of generality, we shall consider only the case with $z = i$. Then by direct complex arithmetic computings, we can write J^{-1} and $(\boldsymbol{AB})^{-1}$ explicitly as

$$J^{-1} = \alpha + i\beta, \quad (\boldsymbol{AB})^{-1} = D + iE, \tag{2.19}$$

where α, β, D and E are given by

$$D = \mathrm{diag}(d_1, d_2), \quad \alpha = \frac{1-\sigma(x)\sigma(y)}{(1+\sigma^2(x))(1+\sigma^2(y))},$$
$$E = \mathrm{diag}(e_1, e_2), \quad \beta = -\frac{\sigma(x)+\sigma(y)}{(1+\sigma^2(x))(1+\sigma^2(y))},$$

with

$$d_1 = \frac{1+\sigma(x)\sigma(y)}{1+\sigma^2(x)}, \quad e_1 = \frac{\sigma(y)-\sigma(x)}{1+\sigma^2(x)},$$
$$d_2 = \frac{1+\sigma(x)\sigma(y)}{1+\sigma^2(y)}, \quad e_2 = \frac{\sigma(x)-\sigma(y)}{1+\sigma^2(y)}.$$

Now we separate the real and imaginary parts of $\boldsymbol{u}_L$ and write $\boldsymbol{u}_L$ as

$$\boldsymbol{u}_L = \boldsymbol{u}_r + i\boldsymbol{u}_i \tag{2.20}$$

with $\boldsymbol{u}_r \in H_{\boldsymbol{g}_r}(\mathrm{curl}; \Omega_L)$ and $\boldsymbol{u}_i \in H_{\boldsymbol{g}_i}(\mathrm{curl}; \Omega_L)$. Then a straightforward computation gives the equivalent weak formulation of (2.17):

Find $\boldsymbol{u}_r \in H_{\boldsymbol{g}_r}(\mathrm{curl}; \Omega_L)$ and $\boldsymbol{u}_i \in H_{\boldsymbol{g}_i}(\mathrm{curl}; \Omega_L)$ such that

$$a(\boldsymbol{u}_r, \boldsymbol{u}_i; \boldsymbol{\psi}_r, -\boldsymbol{\psi}_i) = \boldsymbol{0}, \ \forall\ \boldsymbol{\psi}_r, \boldsymbol{\psi}_i \in H_0(\mathrm{curl}; \Omega_L), \tag{2.21}$$

where $a(\boldsymbol{u}_r, \boldsymbol{u}_i; \boldsymbol{\psi}_r, -\boldsymbol{\psi}_i)$ is given by

$$\begin{aligned} a(\boldsymbol{u}_r, \boldsymbol{u}_i; \boldsymbol{\psi}_r, -\boldsymbol{\psi}_i) = & \int_{\Omega_L} (\alpha(\nabla\times\boldsymbol{u}_r)(\nabla\times\boldsymbol{\psi}_r) - \alpha(\nabla\times\boldsymbol{u}_i)(\nabla\times\boldsymbol{\psi}_i))\mathrm{d}\boldsymbol{x} \\ & - \int_{\Omega_L} (\beta(\nabla\times\boldsymbol{u}_i)(\nabla\times\boldsymbol{\psi}_r) + \beta(\nabla\times\boldsymbol{u}_r)(\nabla\times\boldsymbol{\psi}_i))\mathrm{d}\boldsymbol{x} \\ & -k^2\int_{\Omega_L} (\boldsymbol{u}_r^T D\boldsymbol{\psi}_r - \boldsymbol{u}_i^T D\boldsymbol{\psi}_i)\mathrm{d}\boldsymbol{x} \\ & +k^2\int_{\Omega_L} (\boldsymbol{u}_i^T E\boldsymbol{\psi}_r + \boldsymbol{u}_r^T E\boldsymbol{\psi}_i)\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{2.22}$$

3. Edge element discretization and preconditioners

In this section we shall discuss the edge element discretization of the PML variational system (2.21) and some effective preconditioners for the discrete system. Assume that Ω_L is covered by a quasi-uniform triangulation $\mathcal{T}_h$ of triangular elements, with h being the maximum diameter among all the triangles in $\mathcal{T}_h$. Let $\mathcal{E}_h$ be the set of all edges in the triangulation $\mathcal{T}_h$. We will use the lowest order edge elements of the first family for the discretization:

$$\mathbb{V}_{h,\boldsymbol{g}}(\Omega_L) = \{\boldsymbol{u} = \boldsymbol{u}_r + i\boldsymbol{u}_i \ : \ \boldsymbol{u}_r \in V_{h,\boldsymbol{g}_r}(\Omega_L), \boldsymbol{u}_i \in V_{h,\boldsymbol{g}_i}(\Omega_L)\}, \tag{3.1}$$

where $V_{h,\boldsymbol{g}_r}(\Omega_L)$ and $V_{h,\boldsymbol{g}_i}(\Omega_L)$ are given by

$$\begin{aligned} V_{h,\boldsymbol{g}_l}(\Omega_L) = \{\boldsymbol{v}:\ & \boldsymbol{v} \in H(\mathrm{curl}; \Omega_L),\ \boldsymbol{v}|_K \in \boldsymbol{\mathcal{R}}_1 \quad \forall\, K \in \mathcal{T}_h; \\ & \boldsymbol{v}\cdot\boldsymbol{\tau}|_e = 0 \ \ \forall e \in \mathcal{E}_h \cap \Gamma_L;\ \boldsymbol{v}\cdot\boldsymbol{\tau}|_e = \boldsymbol{g}_l\cdot\boldsymbol{\tau}|_e \ \forall e \in \mathcal{E}_h \cap \Gamma_0\} \end{aligned}$$

with $l = r$ and i. Here $\boldsymbol{\mathcal{R}}_1$ is the following space of linear polynomials:

$$\boldsymbol{\mathcal{R}}_1 = (\mathcal{P}_0)^2 \oplus \left\{\boldsymbol{p} \in (\tilde{\mathcal{P}}_1)^2 \ : \ \boldsymbol{x}\cdot\boldsymbol{p} = 0\right\},$$

where $\mathcal{P}_0$ and $\tilde{\mathcal{P}}_1$ are respectively the space of constants and the space of homogeneous linear polynomials. Similarly, we can define $V_{h,0}(\Omega_L)$.

Now we can formulate our edge element approximation of the variational problem (2.21): Find $\boldsymbol{u}_{r,h} \in V_{h,\boldsymbol{g}_r}(\Omega_L)$ and $\boldsymbol{u}_{i,h} \in V_{h,\boldsymbol{g}_i}(\Omega_L)$ such that

$$a(\boldsymbol{u}_{r,h}, \boldsymbol{u}_{i,h}; \boldsymbol{\psi}_{r,h}, -\boldsymbol{\psi}_{i,h}) = \mathbf{0} \quad \forall\ \boldsymbol{\psi}_{r,h},\ \boldsymbol{\psi}_{i,h} \in V_{h,0}(\Omega_L). \tag{3.2}$$

The major goal of this work is to propose an effective preconditioner for the use in an iterative method for solving the edge element system (3.2). For the purpose, we first write the system (3.2) in a matrix-vector form. Let m_h^0 be the number of all the edges in $\mathcal{T}_h$ lying inside Ω_L and $(m_h - m_h^0)$ be the number of all the edges in $\mathcal{T}_h$ lying on $\partial\Omega_L$, and $\boldsymbol{\psi}_h^j$ the jth basis function of the space $V_h(\Omega_L)$ $(1 \le j \le m_h)$. Then we can express $\boldsymbol{u}_{r,h}$ and $\boldsymbol{u}_{i,h}$ in (3.2) as

$$\boldsymbol{u}_{r,h} = \sum_{j=1}^{m_h} x_r^j \boldsymbol{\psi}_h^j \quad \text{and} \quad \boldsymbol{u}_{i,h} = \sum_{j=1}^{m_h} x_i^j \boldsymbol{\psi}_h^j,$$

and their coefficients corresponding to the edges of $\mathcal{T}_h$ lying inside Ω_L as the vector

$$X = \left(x_r^1, x_r^2, \ldots, x_r^{m_h^0}, x_i^1, x_i^2, \ldots, x_i^{m_h^0} \right)^T.$$

With these expressions, we can obtain the matrix-vector form of the weak formulation (3.2):

$$\mathcal{M}X = F, \tag{3.3}$$

where F is a vector that absorbs the information of $\boldsymbol{g}_r$ and $\boldsymbol{g}_i$ on Γ_0, and $\mathcal{M}$ is a 2×2 block matrix given by

$$\mathcal{M} = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B} & -\mathcal{A} \end{pmatrix} \tag{3.4}$$

with $\mathcal{A}$ and $\mathcal{B}$ being the stiffness matrices associated respectively with the bilinear forms

$$a_{\mathcal{A}}(\boldsymbol{u}_{r,h}, \boldsymbol{\psi}_{r,h}) = \int_{\Omega_L} \big(\alpha(\nabla \times \boldsymbol{u}_{r,h}) \cdot (\nabla \times \boldsymbol{\psi}_{r,h}) - k^2 \boldsymbol{u}_{r,h}^T D \boldsymbol{\psi}_{r,h} \big) \mathrm{d}\boldsymbol{x} \tag{3.5}$$

and

$$a_{\mathcal{B}}(\boldsymbol{u}_{i,h}, \boldsymbol{\psi}_{r,h}) = -\int_{\Omega_L} \big(\beta(\nabla \times \boldsymbol{u}_{i,h}) \cdot (\nabla \times \boldsymbol{\psi}_{r,h}) - k^2 \boldsymbol{u}_{i,h}^T E \boldsymbol{\psi}_{r,h} \big) \mathrm{d}\boldsymbol{x}. \tag{3.6}$$

The matrix $\mathcal{M}$ in (3.4) is symmetric but indefinite. We can observe the following unfavorable features of matrix $\mathcal{M}$: the coefficient function α in the bilinear form (3.5) may change signs and vanish in some part of Ω_L; both diagonal blocks $\mathcal{A}$ and $-\mathcal{A}$ are indefinite themselves; and the off-diagonal blocks $\mathcal{B}$ are generated from an indefinite **curl curl**-type second order differential operator, similarly to the diagonal block $\mathcal{A}$. So the generalized saddle-point system (3.3) is essentially different in nature from the standard saddle-point systems arising from, *e.g.*, the Maxwell system or Navier−Stokes equations. Hence it is rather challenging to construct an efficient and robust preconditioner for the generalized saddle-point system (3.3).

Next we shall construct a Schur complement-type preconditioner for the saddle-point system. To illustrate our motivation, we shall first present a simple block-diagonal preconditioner.

3.1. A simple preconditioner

As usual, we assume that the original electromagnetic scattering problem (2.1) and (2.2) has a unique solution, and so the matrix $\mathcal{M}$ is nonsingular. Without loss of generality, we assume that both $\mathcal{A}$ and $\mathcal{A} + \mathcal{B}\mathcal{A}^{-1}\mathcal{B}$ are nonsingular in the rest of the paper; otherwise, both $\mathcal{B}$ and $\mathcal{B} + \mathcal{A}\mathcal{B}^{-1}\mathcal{A}$ are nonsingular, and we can then proceed our study in a similar manner.

Let $\hat{\mathcal{A}}$ be the stiffness matrix induced by the following bilinear form

$$\hat{a}_{\mathcal{A}}(\boldsymbol{u},\boldsymbol{v}) = \int_{\Omega_L} |\alpha|(\nabla\times\boldsymbol{u})(\nabla\times\boldsymbol{v})\mathrm{d}\boldsymbol{x} + k^2\int_{\Omega_L} \boldsymbol{u}^T D\boldsymbol{v}\mathrm{d}\boldsymbol{x} \quad \forall\, \boldsymbol{u},\boldsymbol{v}\in \mathbb{V}_{h,0}(\Omega_L). \tag{3.7}$$

It is easy to see that matrix $\hat{\mathcal{A}}$ is symmetric and positive definite. The bilinear form $\hat{a}_{\mathcal{A}}$ can be viewed as a regularization of the bilinear form $a_{\mathcal{A}}$ in (3.5), the matrix $\hat{\mathcal{A}}$ should be an efficient preconditioner for $\mathcal{A}$. Then we may take the following block-diagonal matrix to be a simple preconditioner for $\mathcal{M}$ in (3.4):

$$\mathcal{K}^{-1} = \begin{pmatrix} \hat{\mathcal{A}}^{-1} & \\ & \hat{\mathcal{A}}^{-1} \end{pmatrix}. \tag{3.8}$$

Unfortunately, this simple preconditioner does not work well, as we shall see from the numerical experiments in Section 5. In the next section we will construct a more efficient preconditioner for $\mathcal{M}$.

3.2. A preconditioner based on Schur complement

We can easily see that the Schur complement of the system (3.3) is $\mathcal{A}+\mathcal{B}\mathcal{A}^{-1}\mathcal{B}$. This suggests us a natural more effective preconditioner than (3.8):

$$\mathcal{P} = \begin{pmatrix} \hat{\mathcal{A}}^{-1} & \\ & (\hat{\mathcal{A}}+\mathcal{B}\hat{\mathcal{A}}^{-1}\mathcal{B})^{-1} \end{pmatrix}. \tag{3.9}$$

However, the inverse of the Schur complement $\hat{\mathcal{A}}+\mathcal{B}\hat{\mathcal{A}}^{-1}\mathcal{B}$ is difficult to realize numerically. We shall work out below an approximation of the Schur complement, which can be implemented much more effectively.

Let I be the identity operator. By the definition of the matrix $\hat{\mathcal{A}}$ in (3.7), the differential operator defining $\hat{\mathcal{A}}$ can be written in the form:

$$\nabla\times(|\alpha|\nabla\times) + k^2 D\, I. \tag{3.10}$$

Then the differential operator corresponding to $\hat{\mathcal{A}}^{-1}$ may be formally, and approximately, written as

$$(\nabla\times(|\alpha|\nabla\times))^{-1} + k^{-2}D^{-1}I. \tag{3.11}$$

Similarly, the differential operator defining $\mathcal{B}$ can be written in the form

$$-\nabla\times(\beta\nabla\times) + k^2 E\, I. \tag{3.12}$$

Therefore, the differential operator corresponding to the Schur complement $\hat{\mathcal{A}}+\mathcal{B}\hat{\mathcal{A}}^{-1}\mathcal{B}$ may be formally, and approximately, written in the form

$$\nabla\times(\hat{\alpha}\nabla\times) + k^2\hat{D}\, I \tag{3.13}$$

where $\hat{\alpha}$ and $\hat{D}$ are given by

$$\hat{\alpha} = \begin{cases} \frac{1}{2}(|\alpha|+|\alpha|^{-1}\beta^2) & \text{if } \alpha\neq 0,\\ \beta^2 & \text{if } \alpha = 0,\end{cases} \tag{3.14}$$

$$\hat{D} = \frac{1}{2}\mathrm{diag}\Big(d_1+(d_1)^{-1}(e_1)^2, d_2+(d_2)^{-1}(e_2)^2\Big). \tag{3.15}$$

Here the factor $1/2$ can be viewed as a relaxation parameter, which reflects some balance between two diagonal blocks in the proposed preconditioner (see below).

The bilinear form corresponding to the operator (3.13) can be defined by

$$\int_{\Omega_L} \hat{\alpha}(\nabla \times \boldsymbol{u})(\nabla \times \boldsymbol{\psi})\mathrm{d}\boldsymbol{x} + k^2 \int_{\Omega_L} \boldsymbol{u}^T \hat{D}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}, \quad \forall \boldsymbol{u}, \boldsymbol{\psi} \in \mathbb{V}_{h,0}(\Omega_L). \tag{3.16}$$

Let $\hat{\mathcal{S}}$ be the stiffness matrix associated with this bilinear form, then a preconditioner for $\mathcal{M}$ can be defined by

$$\hat{\mathcal{P}} = \begin{pmatrix} \hat{\mathcal{A}}^{-1} & \\ & \hat{\mathcal{S}}^{-1} \end{pmatrix}. \tag{3.17}$$

Remark 3.1. We can see that both $\hat{\mathcal{A}}$ and $\hat{S}$ are symmetric and positive definite, and are generated by the edge element discretization of the elliptic-type Maxwell's equations. They can be replaced by any other existing preconditioners for the elliptic-type Maxwell's equations, for example, the domain decomposition preconditioner [20] or the HX preconditioner [14]. This is an important advantage of the proposed preconditioner for applications.

4. Analysis

In this section we propose a general framework for the analysis of the efficiency of the preconditions of the type $\hat{\mathcal{P}}$ in (3.17) for the global system $\mathcal{M}$ in (3.3).

4.1. General framework

Let $W_h(\Omega_L)$ be a general two-dimensional vector-valued finite element space associated with the triangulation $\mathcal{T}_h$ of domain Ω_L, and A and B be two symmetric (not necessary positive definite) discrete operators mapping from $W_h(\Omega_L)$ to $W_h(\Omega_L)$. Then for any two given functions $f_h, g_h \in W_h(\Omega_L)$, we consider the following operator equation:

$$\begin{pmatrix} A & B \\ B & -A \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_h \\ \boldsymbol{v}_h \end{pmatrix} = \begin{pmatrix} f_h \\ g_h \end{pmatrix}. \tag{4.18}$$

We assume that the system (4.18) is uniquely solvable, namely the coefficient operator

$$M = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}$$

is nonsingular. Then we can easily check that either both A and $A + BA^{-1}B$ are nonsingular, or both B and $B + AB^{-1}A$ are nonsingular. Without loss of generality, we assume that both A and $S = A + BA^{-1}B$ are nonsingular in the rest of this section. For the case when both B and $B + AB^{-1}A$ are nonsingular, we can proceed our subsequent study in an exact same manner.

Let $\hat{A}, \hat{S} : W_h(\Omega_L) \to W_h(\Omega_L)$ be two symmetric and positive definite operators, which are assumed to be two efficient preconditioners respectively for A and the Schur complement $\hat{A} + B\hat{A}^{-1}B$, that is, the following conditions hold for $\hat{A}$:

$$|(A\boldsymbol{\varphi}, \boldsymbol{\varphi})| \le C_1(\hat{A}\boldsymbol{\varphi}, \boldsymbol{\varphi}) \quad \forall\, \boldsymbol{\varphi} \in W_h(\Omega_L), \tag{4.19}$$

$$\sup_{\boldsymbol{\psi} \in W_h(\Omega_L)} \frac{(A\boldsymbol{\varphi}, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_{\hat{A}}} \ge \gamma_1 \|\boldsymbol{\varphi}\|_{\hat{A}} \quad \forall\, \boldsymbol{\varphi} \in W_h(\Omega_L), \tag{4.20}$$

while the following conditions hold for $\hat{S}$:

$$|(S\boldsymbol{\varphi}, \boldsymbol{\varphi})| \le C_2(\hat{S}\boldsymbol{\varphi}, \boldsymbol{\varphi}) \quad \forall\, \boldsymbol{\varphi} \in W_h(\Omega_L), \tag{4.21}$$

$$\sup_{\boldsymbol{\psi} \in W_h(\Omega_L)} \frac{(S\boldsymbol{\varphi}, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_{\hat{S}}} \ge \gamma_2 \|\boldsymbol{\varphi}\|_{\hat{S}} \quad \forall\, \boldsymbol{\varphi} \in W_h(\Omega_L), \tag{4.22}$$

$$((\hat{A} + B\hat{A}^{-1}B)\boldsymbol{\varphi}, \boldsymbol{\varphi}) \le C_3(\hat{S}\boldsymbol{\varphi}, \boldsymbol{\varphi}) \quad \forall\, \boldsymbol{\varphi} \in W_h(\Omega_L). \tag{4.23}$$

We can easily see from (4.23) and (4.20) that

$$(B\hat{A}^{-1}B\boldsymbol{\varphi},\boldsymbol{\varphi}) \le C_3(\hat{S}\boldsymbol{\varphi},\boldsymbol{\varphi}) \quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L), \tag{4.24}$$

$$(A^{-1}\hat{A}A^{-1}\boldsymbol{\varphi},\boldsymbol{\varphi}) \le \gamma_1^{-1}(\hat{A}^{-1}\boldsymbol{\varphi},\boldsymbol{\varphi}) \quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L). \tag{4.25}$$

Now we define

$$\hat{M} = \begin{pmatrix} \hat{A} & \\ & \hat{S} \end{pmatrix}, \qquad \tilde{M} \equiv \hat{M}^{-\frac{1}{2}} M \hat{M}^{-\frac{1}{2}} = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}^T & -\tilde{\tilde{A}} \end{pmatrix},$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{\tilde{A}}$ are given by

$$\tilde{A} = \hat{A}^{-\frac{1}{2}} A \hat{A}^{-\frac{1}{2}}, \quad \tilde{B} = \hat{A}^{-\frac{1}{2}} B \hat{S}^{-\frac{1}{2}}, \quad \tilde{\tilde{A}} = \hat{S}^{-\frac{1}{2}} A \hat{S}^{-\frac{1}{2}}.$$

Theorem 4.1. *Let γ_1, γ_2 and C_3 be three positive numbers satisfying (4.20), (4.22) and (4.24), and $\gamma = C_3^{-1}\gamma_1\min\{\gamma_1,\gamma_2\}$. Then, it holds that*

$$\sup_{\boldsymbol{v}\in W_h(\Omega_L)\times W_h(\Omega_L)} \frac{(\tilde{M}\boldsymbol{u},\boldsymbol{v})}{\|\boldsymbol{v}\|_{L^2(\Omega_L)}} \ge \gamma\|\boldsymbol{u}\|_{L^2(\Omega_L)} \quad \forall\ \boldsymbol{u}\in W_h(\Omega_L)\times W_h(\Omega_L). \tag{4.26}$$

Proof. First, it can be verified directly that

$$\tilde{\tilde{A}} + \tilde{B}^T\tilde{A}^{-1}\tilde{B} = \hat{S}^{-\frac{1}{2}}(A + BA^{-1}B)\hat{S}^{-\frac{1}{2}}. \tag{4.27}$$

Then we can derive by using (4.20) and (4.22) that

$$\sup_{\boldsymbol{\psi}\in W_h(\Omega_L)} \frac{(\tilde{A}\boldsymbol{\varphi},\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_{L^2(\Omega_L)}} \ge \gamma_1\|\boldsymbol{\varphi}\|_{L^2(\Omega_L)} \quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L), \tag{4.28}$$

$$\sup_{\boldsymbol{\psi}\in W_h(\Omega_L)} \frac{((\tilde{\tilde{A}} + \tilde{B}^T\tilde{A}^{-1}\tilde{B})\boldsymbol{\varphi},\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_{L^2(\Omega_L)}} \ge \gamma_2\|\boldsymbol{\varphi}\|_{L^2(\Omega_L)} \quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L). \tag{4.29}$$

This implies that, for any $u_1, u_2 \in W_h(\Omega_L)$, there are $v_1, v_2 \in W_h(\Omega_L)$ satisfying

$$\left(\begin{pmatrix} \tilde{A} & \\ & -(\tilde{\tilde{A}} + \tilde{B}^T\tilde{A}^{-1}\tilde{B}) \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right) \ge \hat{\gamma}\left\|\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\right\|_{L^2(\Omega_L)}\left\|\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}\right\|_{L^2(\Omega_L)} \tag{4.30}$$

with $\hat{\gamma} = \min\{\gamma_1,\gamma_2\}$.

It is straightforward to check the following factorization

$$\tilde{M} = \begin{pmatrix} I & 0 \\ \tilde{B}\tilde{A}^{-1} & I \end{pmatrix}\begin{pmatrix} \tilde{A} & \\ & -(\tilde{\tilde{A}} + \tilde{B}^T\tilde{A}^{-1}\tilde{B}) \end{pmatrix}\begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix}.$$

This, along with (4.30), suggests us to define

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix}\boldsymbol{u}$$

for any $\boldsymbol{u}\in W_h(\Omega_L)\times W_h(\Omega_L)$, and to choose $\boldsymbol{v}$ by

$$\begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix}\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Then using (4.30) we deduce

$$\begin{aligned}(\tilde{M}\boldsymbol{u},\boldsymbol{v}) &= \left(\begin{pmatrix}\tilde{A} & \\ & -(\tilde{\tilde{A}}+\tilde{B}^T\tilde{A}^{-1}\tilde{B})\end{pmatrix}\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{u},\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{v}\right)\\ &\geq \hat{\gamma}\left\|\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{u}\right\|_{L^2(\Omega_L)}\left\|\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{v}\right\|_{L^2(\Omega_L)}. \end{aligned}\tag{4.31}$$

Now we take $\boldsymbol{u}=(\boldsymbol{\varphi},\boldsymbol{\psi})^T$, then

$$\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\begin{pmatrix}\boldsymbol{\varphi}\\ \boldsymbol{\psi}\end{pmatrix}=\begin{pmatrix}u_1\\ u_2\end{pmatrix},\tag{4.32}$$

which implies that

$$\boldsymbol{\psi}=u_2,\quad \boldsymbol{\varphi}=u_1-\tilde{A}^{-1}\tilde{B}u_2.$$

It follows from (4.24) and (4.25) that

$$(BA^{-1}\hat{A}A^{-1}B\varphi,\varphi)\leq \gamma_1^{-1}C_3(\hat{S}\varphi,\varphi),\quad \forall\varphi\in W_h(\Omega_L),$$

or equivalently,

$$(\hat{S}^{-\frac{1}{2}}BA^{-1}\hat{A}A^{-1}B\hat{S}^{-\frac{1}{2}}\varphi,\varphi)\leq \gamma_1^{-1}C_3(\varphi,\varphi),\quad \forall\varphi\in W_h(\Omega_L),\tag{4.33}$$

which gives

$$\|\tilde{A}^{-1}\tilde{B}u_2\|_{L^2(\Omega_L)}\leq \gamma_1^{-1}C_3\|u_2\|_{L^2(\Omega_L)}.$$

Using this, we obtain

$$\|\boldsymbol{\psi}\|_{L^2(\Omega_L)}=\|u_2\|_{L^2(\Omega_L)}\quad\text{and}\quad \|\boldsymbol{\varphi}\|_{L^2(\Omega_L)}\leq \gamma_1^{-1}C_3(\|u_1\|_{L^2(\Omega_L)}+\|u_2\|_{L^2(\Omega_L)}).$$

Then, we can further derive by (4.32) that

$$\|\boldsymbol{u}\|_{L^2(\Omega_L)}\leq \gamma_1^{-1}C_3(\|u_1\|^2_{L^2(\Omega_L)}+\|u_2\|^2_{L^2(\Omega_L)})^{\frac{1}{2}}=\gamma_1^{-1}C_3\left\|\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{u}\right\|_{L^2(\Omega_L)}.\tag{4.34}$$

Similarly we can obtain

$$\|\boldsymbol{v}\|_{L^2(\Omega_L)}\leq \gamma_1^{-1}C_3\left\|\begin{pmatrix}I & \tilde{A}^{-1}\tilde{B}\\ 0 & I\end{pmatrix}\boldsymbol{v}\right\|_{L^2(\Omega_L)}.\tag{4.35}$$

Now the desired estimate (4.26) is a direct consequence of (4.31), (4.34) and (4.35). $\square$

Theorem 4.2. *Let C_1, γ_1, C_2 and C_3 be the positive numbers satisfying (4.19)–(4.21) and (4.24), and $C=C_3\gamma_1^{-1}\max\{C_1,C_2\}$. Then, the following estimate holds*

$$|(\tilde{M}\boldsymbol{v},\boldsymbol{v})|\leq C\|\boldsymbol{v}\|^2_{L^2(\Omega_L)}\quad \forall\ \boldsymbol{v}\in W_h(\Omega_L)\times W_h(\Omega_L).\tag{4.36}$$

Proof. We can easily see from (4.19) and (4.21) that

$$\begin{aligned}|(\tilde{A}\boldsymbol{\varphi},\boldsymbol{\varphi})|&\leq C_1\|\boldsymbol{\varphi}\|^2_{L^2}\quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L),\\ |((\tilde{\tilde{A}}+\tilde{B}^T\tilde{A}^{-1}\tilde{B})\boldsymbol{\varphi},\boldsymbol{\varphi})|&\leq C_2\|\boldsymbol{\varphi}\|^2_{L^2}\quad \forall\ \boldsymbol{\varphi}\in W_h(\Omega_L).\end{aligned}$$

Using these two estimates and the following factorization of $\tilde{M}$,

$$\tilde{M} = \begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix}^T \begin{pmatrix} \tilde{A} & \\ & -(\tilde{\tilde{A}} + \tilde{B}^T\tilde{A}^{-1}\tilde{B}) \end{pmatrix} \begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix},$$

we can readily derive the desired estimate as follows:

$$|(\tilde{M}\boldsymbol{v}, \boldsymbol{v})| \le \max\{C_1, C_2\} \left\| \begin{pmatrix} I & \tilde{A}^{-1}\tilde{B} \\ 0 & I \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right\|_{L^2}^2 \le C(\|v_1\|^2 + \|v_2\|^2) = C\|\boldsymbol{v}\|_{L^2}^2$$

where we have used (4.33) to estimate the upper bound of the operator $\tilde{A}^{-1}\tilde{B}$. $\square$

4.2. An application

If we choose the space $W_h(\Omega_L)$ to be the edge element space $V_{h,0}(\Omega_L)$ defined in Section 3, and two discrete operators A and B such that

$$(A\boldsymbol{u}_h, \boldsymbol{\varphi}_h) = \int_\Omega (\alpha(\nabla \times \boldsymbol{u}_h)(\nabla \times \boldsymbol{\varphi}_h) - k^2 D\boldsymbol{u}_h \cdot \boldsymbol{\varphi}_h)\mathrm{d}x,$$

$$(B\boldsymbol{v}_h, \boldsymbol{\psi}_h) = -\int_\Omega (\beta(\nabla \times \boldsymbol{v}_h)(\nabla \times \boldsymbol{\psi}_h) - k^2 E\boldsymbol{v}_h \cdot \boldsymbol{\psi}_h)\mathrm{d}x$$

for any $\boldsymbol{u}_h,\ \boldsymbol{v}_h,\ \boldsymbol{\varphi}_h,\ \boldsymbol{\psi}_h \in V_{h,0}(\Omega_L)$. Then the edge element system (3.3) can be written in the operator form (4.18).

Let $\hat{A}$ and $\hat{S}$ be the operator forms of the matrices $\hat{\mathcal{A}}$ and $\hat{\mathcal{S}}$ defined in Section 3.2, respectively. Noting that $V_{h,0}(\Omega_L)$ is a finite dimensional space, and that both A and S are nonsingular, we can easily know that the positive numbers $\gamma_1,\ \gamma_2,\ C_1,\ C_2\ C_3$ satisfying (4.19)−(4.24) always exist, possibly depending on mesh size h or growing slowly when h decreases. In particular, when these positive numbers are independent of the mesh size h or grow very slowly (*e.g.*, logarithmically) when h decreases, the preconditioner $\hat{\mathcal{P}}$ in (3.17) will be a very efficient preconditioner for the matrix $\mathcal{M}$ in (3.4). But we are still unable to establish very satisfactory estimates of these positives numbers for our current choices of $\hat{A}$ and $\hat{S}$, although our numerical experiments in the next section suggest the existence of some nice constants which grow very slowly when h decreases. The main difficulties lie in two facts: (1) the operator A in (4.18) is indefinite itself; (2) the coefficient α in the bilinear form $a_{\mathcal{A}}$ in (3.5) may change signs and vanish in some part of Ω. We are unaware of any existing theories which may help assess the performance of some preconditioning-type methods for the generalized saddle-point system like (3.4). It is the first time to provide in this section a set of criteria for such assessments.

The preconditioners $\hat{\mathcal{A}}$ and $\hat{\mathcal{S}}$ we proposed in Section 3.2 which form the global preconditioner $\hat{\mathcal{P}}$ (*cf.* (3.17)) for matrix $\mathcal{M}$ (*cf.* (3.4)) are just one possible choice. One may apply or construct any other more efficient preconditioners for matrix $\mathcal{M}$, especially those for which the positive numbers $\gamma_1,\ \gamma_2,\ C_1,\ C_2\ C_3$ satisfying (4.19)−(4.24) can be estimated explicitly in terms of h and grow slowly when h decreases. This topic will be studied further in our future work.

5. Numerical experiments

In this section we present some numerical results to illustrate the effectiveness of the preconditioners introduced in Section 3. We shall take a model example from [8], which is the PML equations to approximate the electromagnetic scattering problems in two dimensions.

In our experiments, we take the coefficients in the scattering system (2.1) and its boundary data $\boldsymbol{g}$ to be

$$\mu = 1\,, \quad \varepsilon = 1\,, \quad k = 1\,, \quad \boldsymbol{g} = \nabla \times \left[H_1^{(1)}(r)\mathrm{e}^{i\theta}\right]\big|_{\Gamma_0}\,,$$

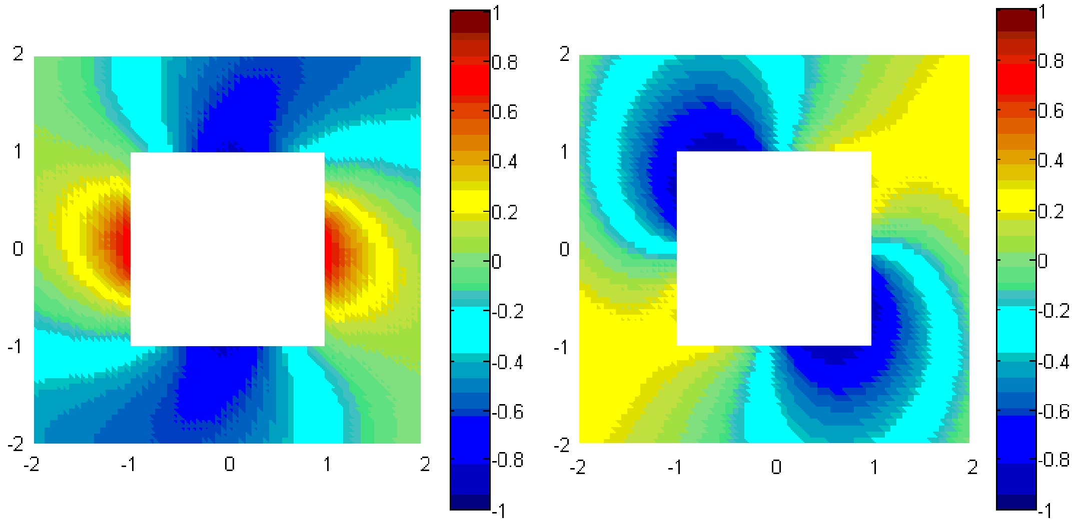


FIGURE 2. First component (left) and second component (right) of $\boldsymbol{u}_r$.

TABLE 1. Number of iterations and CPU times of the MINRES method.

h	Iter	Time
8/16	3845	2.52(s)
8/32	21815	63.39(s)
8/64	88610	1352.11(s)
8/128	317226	22092.86(s)

where $H_1^{(1)}(r)$ is the Hankel function of first kind. Then the analytic solution to the system (2.1) is given by $\boldsymbol{u} = \nabla \times \left[H_1^{(1)}(r)\mathrm{e}^{i\theta}\right]$. Furthermore, we take the PML parameters and the PML domain Ω_L in Section 2 to be

$$a = 2, \quad b = 3, \quad L = 4, \quad \sigma_0 = 4, \quad \Omega_0 = (-1,1)^2, \quad \Omega_L = (-4,4)^2 \setminus [-1,1]^2,$$

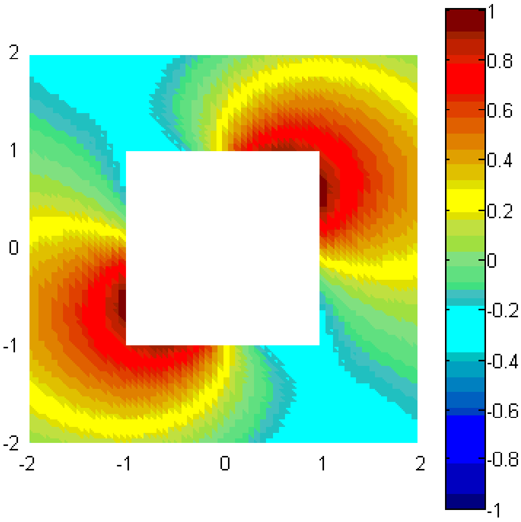
The two components of $\boldsymbol{u}$, *i.e.*, its real and imaginary parts $\boldsymbol{u}_r$ and $\boldsymbol{u}_i$, are showed in Figures 2 and 3, respectively.

We are interested in the approximate solution to the PML system (2.17) only in the computational domain $\Omega_1 = (-2,2)^2 \setminus [-1,1]^2$, so we will show the edge element PML solution in Ω_1, namely the solution to the algebraic system (3.3). For solving the edge element system (3.3), we use the MINRES method without any preconditioner and the preconditioned MINRES method with the preconditioners introduced in Section 3, and compare the performance of these two methods. The iterations are terminated when the relative L^2-norm of the residual falls below 10^{-6}, and the number of iterations and the required CPU times are reported.

In Table 1, we can see the numbers of iterations and the required CPU times with different mesh sizes when the MINRES method is applied without any preconditioner.

We may observe from Table 1 that the condition number of the matrix $\mathcal{M}$ grows quickly as the mesh size h reduces, so the MINRES method (without any preconditioner) is quite expensive and impractical for solving such a PML system like (3.3) when the system is of large size.

Next we shall test the performance of two preconditioners, the simple block-diagonal preconditioner $\mathcal{K}$ in (3.8) and the Schur complement-type preconditioner $\hat{\mathcal{P}}$ in (3.17), when they are used with the preconditioned MINRES method for solving the PML system (3.3). As our aim is to find the effectiveness of the proposed preconditioners and do not study the further preconditioning for the preconditioners $\hat{\mathcal{A}}$ or $\hat{\mathcal{S}}$ used in (3.17), we shall implement the actions of $\hat{\mathcal{A}}^{-1}$ and $\hat{\mathcal{S}}^{-1}$ by the PCG iteration with the HX preconditioner [14], whose action involves the



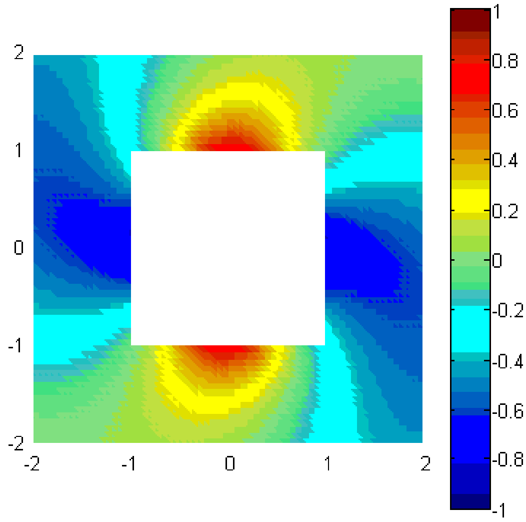


FIGURE 3. First complement (left) and second component (right) of $\boldsymbol{u}_i$.

TABLE 2. Number of iterations and CPU times of the MINRES method with preconditioner $\mathcal{K}$.

h	Iter	Time
8/16	121	2.76(s)
8/32	169	16.02(s)
8/64	342	148.85(s)
8/128	677	2078.92(s)
8/256	$>10^4$	

TABLE 3. Errors of the PML edge element solutions.

h	$\\|\boldsymbol{u}-\boldsymbol{u}_h\\|_{0,\Omega_1}$	Ratio	$\\|\boldsymbol{u}-\boldsymbol{u}_h\\|_{\boldsymbol{H}(\mathbf{curl};\Omega_1)}$	Ratio
8/16	6.0118e-1		6.8581e-1	
8/32	3.1081e-1	1.93	3.4915e-1	1.96
8/64	1.5672e-1	1.98	1.7475e-1	2.00
8/128	7.8556e-2	2.00	8.7496e-2	2.00
8/256	3.9303e-2	2.00	4.3766e-2	2.00
8/512	1.9654e-2	2.00	2.1884e-2	2.00

solutions of three simple elliptic equations by the algebraic multigrid method. This makes the implementation of the preconditioner $\hat{\mathcal{P}}$ very convenient and computationally cheap.

We first solve the PML edge element system (3.3) by the preconditioned MINRES method with the simple block-diagonal preconditioner $\mathcal{K}$ in (3.8). The number of iterations and the CPU times are reported in Table 2, where Iter stands for the iteration counts of the preconditioned MINIRES method for solving the system (3.3).

We can see from Tables 1 and 2 that the preconditioned MINRES method with the simple block-diagonal preconditioner $\mathcal{K}$ outperforms essentially the MINRES method without any preconditioner. But the convergence of the former deteriorates still rapidly when the mesh size reduces.

Now we solve the system (3.3) by the preconditioned MINRES method with the new Schur complement preconditioner $\hat{\mathcal{P}}$ in (3.17). The errors of the PML edge element solutions are listed in Table 3, from which we can see that the PML edge element solution converges optimally (with first order accuracy) in both L^2-norm

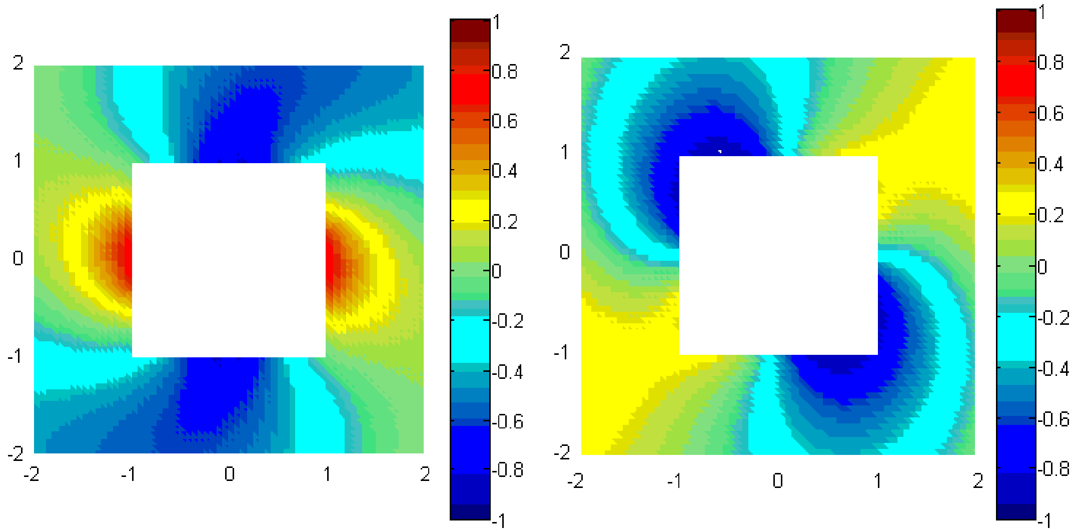


FIGURE 4. First complement (left) and second component (right) of $\boldsymbol{u}_{h,r}$.

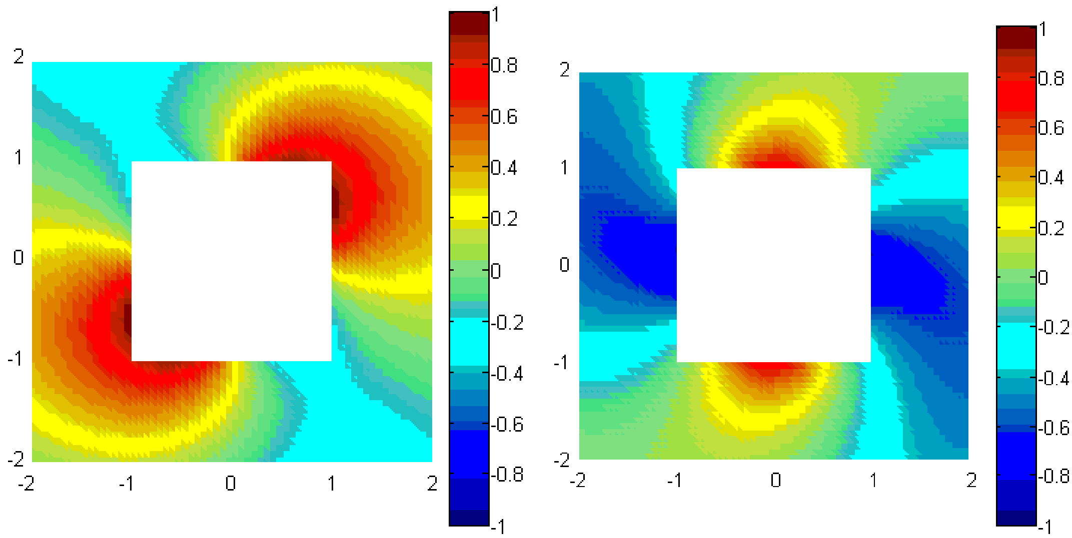


FIGURE 5. First component (left) and second component (right) of $\boldsymbol{u}_{h,i}$.

and $H(\mathbf{curl})$-norm. Furthermore, we have shown the two complements of the real part $\boldsymbol{u}_{r,h}$ and the imaginary part $\boldsymbol{u}_{i,h}$ of the edge element solution $\boldsymbol{u}_h$; see Figures 4 and 5. Comparing these two figures with Figures 2 and 3, we can see that the PML edge element solution $\boldsymbol{u}_h$ approximates the analytic solution $\boldsymbol{u}$ well.

The number of iterations and the CPU times are reported in Table 4. We can see from Table 4 that the new Schur complement-type preconditioner $\hat{\mathcal{P}}$ in (3.17) performs very well, and is much more effective than the simple block-diagonal preconditioner $\mathcal{K}$ in (3.8) when the discrete system becomes large. And most importantly, the number of iterations becomes more stabilized when the meshes are fine enough. Especially, we can see that the number of iterations for the coarse mesh $h = 8/64$ is nearly the same as the number of iterations for the

Table 4. Number of iterations and CPU times of the MINRES method with preconditioner $\mathcal{P}$.

h	Iter	Time
8/16	128	3.65(s)
8/32	166	16.81(s)
8/64	248	112.00(s)
8/128	149	460.69(s)
8/256	164	2705.72(s)
8/512	253	20864.19(s)

much finer mesh $h = 8/512$, which results in a quite large discrete PML system, with a total number of degrees of freedom being 1 472 000. The convergence of this type in terms of mesh size is what we like to see in most applications.

Acknowledgements. The authors would like to thank three anonymous referees for their many insightful and constructive comments and suggestions, which have helped us improve the organisation and quality of the paper essentially.

References

[1] E. Bécache and P. Joly, On the analysis of Bérenger's perfectly matched layers for Maxwell's equations. *ESAIM: M2AN* **36** (2002) 87–119.

[2] J.-P. Bérenger, A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.* **114** (1994) 185-200.

[3] J.-P. Bérenger, Perfectly Matched Layer (PML) for Computational Electromagnetics. Vol. 2 of *Synthesis Lectures on Computational Electromagnetics*. Morgan Claypool (2007).

[4] H. Banks and B. Browning, Time domain electromagnetic scattering using finite elements and perfectly matched layers. *Comput. Methods. Appl. Mech. Eng.* **194** (2005) 149–168.

[5] G. Bao and H. Wu, Convergence analysis of the perfectly matched layer problems for time harmonic Maxwell's equations. *SIAM J. Numer. Anal.* **43** (2005) 2121–2143.

[6] Y. Botros and J. Volakis, Preconditioned generalized minimal residual iterative scheme for perfectly matched layer terminated applications. *IEEE Microw. Guided Wave Letters* **9** (1999) 45–47.

[7] Y.Botros and J.Volakis, Perfectly matched layer termination for finite-element meshed: Implementation and application, Microwave. *Optim. Tech. Lett.* **23** (1999) 166–172.

[8] J. Bramble and J. Pasciak, Analysis of a cartesian PML approximation to the three dimensional electromagnetic wave scattering problem, *Int. J. Numer. Anal. Model.* **9** (2012) 543–561.

[9] J. Chen and Z. Chen, An adaptive perfectly matched layer technique for 3-D time harmonic electromagnetic scattering problems. *Math. Comput.* **77** (2008) 673–698.

[10] J. Chen, Z. Chen, T. Cui and L. Zhang, An Adaptive Finite Element Method for the Eddy Current Model with Circuit/Field Couplings, *SIAM J. Sci. Comput.* **32** (2010) 1020–1042.

[11] Z. Chen, T. Cui and L. Zhang, An adaptive anisotropic perfectly matched layer method for 3-D time harmonic electromagnetic scattering problems, *Numer. Math.* **125** (2013) 639–677.

[12] W. Chew and W. Weedon, A 3d perfectly matched medium for modified Maxwell's equations with streched coordinates. *Microwave Opt. Technol. Lett.* **13** (1994) 599–604.

[13] R. Hiptmair, Multigrid method for Maxwell's equations. *SIAM J. Numer. Anal.* **36** (1998) 204–225.

[14] R. Hiptmair and J. Xu, Nodal auxiliary spaces preconditions in H(curl) and H(div) spaces. *SIAM J. Numer. Anal.* **45** (2007) 2483–2509.

[15] Q. Hu and J. Zou, Substructuring preconditioners for saddle-point problems arising from Maxwell's equations in three dimensions. *Math. Comput.* **73** (2004) 35–61.

[16] E. Post, Formal Structure of Electromagnetics: General Covariance and Electromagnetics. Dover (1997).

[17] Wonseok Shin and Shanhui Fan, Choice of the perfectly matched layer boundary condition for frequency-domain Maxwell's equations solvers. *J. Comput. Phys.* **231** (2012) 3406–3431.

[18] F. L. Teixeira and W. C. Chew, Differential forms, metrics, and the reflectionless absorption of electromagnetic waves. *J. Electromagn. Waves Appl.* **13** (1999) 665–686.

[19] A. Toselli, Overlapping Schwarz methods for Maxwell's equations in three dimensions. *Numer. Math.* **86** (2000) 733–752.

[20] A. Toselli, O. Widlund and B. Wohlmuth, An iterative substructuring method for Maxwell's equations in two dimensions. *Math. Comput.* **70** (2001) 935–947.

[21] P. Tsuji, B. Engquist and L. Ying, A sweeping preconditioner for time-harmonic Maxwell's equations with finite elements. *J. Comput. Phys.* **231** (2012) 3770–3783.