Numerical Algorithms (2021) 86:281–302
https://doi.org/10.1007/s11075-020-00889-7

ORIGINAL PAPER

Preconditioners and their analyses for edge element saddle-point systems arising from time-harmonic Maxwell's equations

Ying Liang[1] · Hua Xiang[2] · Shiyang Zhang[2] · Jun Zou[1]

Received: 23 July 2018 / Accepted: 21 January 2020 / Published online: 16 March 2020

Abstract

We derive and propose a family of new preconditioners for the saddle-point systems arising from the edge element discretization of the time-harmonic Maxwell's equations in three dimensions. With the new preconditioners, we show that the preconditioned conjugate gradient method can apply for the saddle-point systems when wave numbers are smaller than a positive critical number, while the iterative methods like the preconditioned MINRES may apply when wave numbers are larger than the critical number. The spectral behaviors of the resulting preconditioned systems for some existing and new preconditioners are analyzed and compared, and several two-dimensional numerical experiments are presented to demonstrate and compare the efficiencies of these preconditioners.

Keywords Time-harmonic Maxwell's equations · Saddle-point system · Preconditioners

Mathematics Subject Classification (2010) 65F10 · 65N22 · 65N30

✉ Jun Zou
zou@math.cuhk.edu.hk

Ying Liang
yliang@math.cuhk.edu.hk

Hua Xiang
hxiang@whu.edu.cn

Shiyang Zhang
hydzhang@whu.edu.cn

[1] Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, People's Republic of China

[2] School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, People's Republic of China

1 Introduction

In this work, we investigate and compare some effective preconditioning solvers for the following saddle-point system:

$$\mathcal{K}\begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A - k^2 M & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{1.1}$$

where $u \in \mathbb{R}^n$, $p \in \mathbb{R}^m$, A, $M \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{m\times n}$, with $m \leq n$. We assume that $\mathcal{K}$ is non-singular, so B must be of full row rank. We are particularly interested in the case where A is symmetric semi-positive definite, and $\dim(ker(A)) = m$, that is, A is maximally rank deficient [7, 8]. The matrix M is assumed to be symmetric positive definite, and k is a given real number.

The saddle-point system of form (1.1) with a maximal rank deficient A arises from many applications, including the numerical solution of time-harmonic Maxwell's equations [8, 9, 19] where k represents the wave number, the underdetermined norm-minimization problems [2], and geophysical inverse problems; see more details in the very recent paper [7]. This reference is a very inspiring and innovative work and has developed a class of indefinite block preconditioners for the use with the conjugate gradient (CG) method, which may converge rapidly under certain conditions when it is applied for solving the general saddle-point system of form (1.1) with a vanishing wave number (i.e., $k = 0$). It was also pointed out in [7] that the saddle-point system under the aforementioned particular setting has not received as much attention as other situations, for example, the case of a symmetric positive definite A. But as it was demonstrated in [7] for the special case $k = 0$, when A is maximally rank deficient, some nice mathematical structures may be revealed and adopted to help construct efficient solution methods. The current work is initiated and motivated by [7] and intends to develop further in this direction. We show that new efficient numerical methods can be equally constructed for more general case with non-vanishing wave numbers, i.e., $k \neq 0$.

Though most results of this work apply also to the general saddle-point system of form (1.1) with a maximal rank deficient A, we focus mainly on the saddle-point system (1.1) that arises from time-harmonic Maxwell's equations [4, 6, 11, 19]:

$$\begin{cases} \nabla \times \nabla \times u - k^2 u + \nabla p = J \ \text{ in } \ \Omega, \\ \nabla \cdot u = \rho \ \text{ in } \ \Omega, \\ u \times n = 0 \ \text{ on } \ \partial\Omega, \\ p = 0 \ \text{ on } \ \partial\Omega \end{cases} \tag{1.2}$$

where u is a vector field, p is the scalar multiplier, J is the given external source, and ρ is the density of charge. Ω is a simply connected domain in $\mathbb{R}^3$ with a connected boundary $\partial\Omega$, with n being its outward unit normal. The wave number k is given by $k^2 = \omega^2\varepsilon\mu$, where ω, ε, and μ are positive frequency, permittivity, and permeability of the medium, respectively. We assume that k^2 is not an interior Maxwell eigenvalue, but is allowed to be zero, and know the cases with appropriately small and large frequencies are physically relevant in magnetostatics, wave propagation, and other applications [8]. We refer to [3, chapter 11] for a survey on this topic. The introduction of the Lagrange multiplier p in (1.2) may not be absolutely necessary

for the general case $k \neq 0$, for which the divergence constraint does not need to be enforced directly and explicitly. That is, it is possible to solve directly for u using the first equation in (1.2) with $p = 0$ mathematically [9], although it is still challenging to design an efficient numerical solver for this indefinite system. The saddle-point formulation (1.2) with the Lagrange multiplier p is stable and well-posed [6], and especially it ensures the stability and Gauss's law directly when k is small and may better handle the singularity of the solution at the boundary of the domain [4, 6, 19]. More importantly, the mixed form (1.2) provides some extra flexibility on the computational aspect [12–15] and leads to better numerical stability and more efficient numerical solvers than the single system (1.2) without Lagrange multiplier (i.e., $p = 0$), as it was shown in [7, 8]. And this is also the main motivation and focus of the current work.

After discretizing (1.2) by using the Nédélec elements of the first kind [16, 17] for the approximation of the vector field u and the standard nodal elements for the multiplier p, we derive the saddle-point system (1.1) of our interest, where $A \in \mathbb{R}^{n\times n}$ corresponds now to the discrete version of the curl-curl operator, $B \in \mathbb{R}^{m\times n}$ is a discrete divergence operator, and $M \in \mathbb{R}^{n\times n}$ is the vector mass matrix. We assume that the coefficient matrix $\mathcal{K}$ in (1.1) and its leading block $A - k^2M$ ($k \neq 0$) are both non-singular, which is true when meshes are sufficiently fine [8].

For the special case of the saddle-point system (1.1), namely the system with the vanishing wave number ($k = 0$):

$$\mathcal{A}\begin{pmatrix} u \\ p \end{pmatrix} \equiv \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1.3}$$

a very effective preconditioner of the form

$$\mathcal{P}_0^{-1} = \begin{pmatrix} (A+M)^{-1}(I - B^TL^{-1}C^T) & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix} \tag{1.4}$$

was proposed in [7] for solving the saddle-point system (1.3). Here the matrix $L \in \mathbb{R}^{m\times m}$ is the discrete Laplacian, while $C \in \mathbb{R}^{n\times m}$ is a sparse matrix, whose columns span $ker(A)$ and can be formed easily and explicitly using the gradients of the standard nodal bases [7, 8]. It is important to verify that the preconditioned system $\mathcal{P}_0^{-1}\mathcal{A}$ is block diagonal [7]:

$$\mathcal{P}_0^{-1}\mathcal{A} = \begin{pmatrix} (A+M)^{-1}(A + B^TL^{-1}B) & 0 \\ 0 & I \end{pmatrix}.$$

Since both matrices $A + M$ and $A + B^TL^{-1}B$ are symmetric positive definite, we can apply a CG-like method for the preconditioned system $\mathcal{P}_0^{-1}\mathcal{A}$ in a non-standard inner product, even though both $\mathcal{A}$ and $\mathcal{P}_0$ are indefinite.

For the more general case $k \neq 0$, the block triangular preconditioners

$$\mathcal{M}_{\eta,\varepsilon} = \begin{bmatrix} A + (\eta - k^2)M & (1-\eta\varepsilon)B^T \\ 0 & \varepsilon L \end{bmatrix} \tag{1.5}$$

with double variable relaxation parameters $\eta > k^2$ and $\varepsilon \neq 0$ were studied in [5, 8, 23, 24].

In this work, we construct some new preconditioners for the general saddle-point system (1.1). As it was shown in [7] that the aforementioned preconditioner $\mathcal{P}_0^{-1}$ in (1.4) works very effectively for the special and simple case with vanishing wave number ($k = 0$), we demonstrate that similar preconditioners can be constructed and generalized also for the saddle-point linear system (1.1) with more general cases, i.e., the non-vanishing wave numbers $k \neq 0$. And we will see analytically the spectral distributions of these new preconditioners are quite similar to the ones of the existing effective preconditioners (1.5). But the new preconditioners can be applied with the CG iteration under a non-standard inner product although both the coefficient matrix $\mathcal{K}$ and the new preconditioner are indefinite, and numerically they perform mostly better and stabler than the existing preconditioners (1.5).

The rest of the paper is arranged as follows. We develop in Section 2 an important formula for computing the inverse of $\mathcal{K}$, based on which we propose in Section 3 a family of new preconditioners and compare their performance with some existing preconditioners for the saddle-point system (1.1) with general wave numbers, then study and compare the spectral properties of the preconditioned matrices. Several two-dimensional numerical experiments are presented in Section 4 to demonstrate the performance of the new preconditoners and their comparisons with some existing preconditioners. Finally, some concluding remarks are included in Section 5 to summarize the main results of the paper and some possible future directions.

2 Computing the inverse of $\mathcal{K}$

We derive in this section some formulas for computing the inverse of the matrix $\mathcal{K}$ in (1.1). To do so, we first recall some useful properties of the matrices A, B, M, L, and C, which are introduced in Section 1.

Proposition 2.1 *The matrices A, B, M, L, and C have the following properties* [7, 8]*:*

(i) $\mathbb{R}^n = ker(A) \oplus ker(B)$.

(ii) *There exists a constant $\bar{\alpha} > 0$ independent of mesh size such that $u^T Au \geq \bar{\alpha}\, u^T Mu, \quad \forall\, u \in ker(B)$.*

(iii) $C = M^{-1}B^T$, $L = BM^{-1}B^T$, $AC = 0$.

(iv) *The inverse of $\mathcal{A}$ can be represented by*

$$\mathcal{A}^{-1} = \begin{pmatrix} V & CL^{-1} \\ L^{-1}C^T & 0 \end{pmatrix}, \tag{2.1}$$

where the diagonal block V is given by

$$V = (A + B^T L^{-1} B)^{-1}(I - B^T L^{-1} C^T) = (A + B^T L^{-1} B)^{-1} - CL^{-1}C^T. \tag{2.2}$$

Now we are ready to derive an explicit formula for computing the inverse of the general saddle-point matrix $\mathcal{K}$ in (1.1) with non-vanishing wave numbers $k \neq 0$.

Theorem 2.2 *The inverse of $\mathcal{K}$ in (1.1) has the representation*

$$\mathcal{K}^{-1} = \begin{pmatrix} T & CL^{-1} \\ L^{-1}C^T & k^2L^{-1} \end{pmatrix}, \tag{2.3}$$

where T satisfies

$$(A - k^2M)T = I - B^TL^{-1}C^T, \quad BT = 0. \tag{2.4}$$

Proof We write $\mathcal{K}^{-1}$ as a perturbation of $\mathcal{A}^{-1}$ in the form

$$\mathcal{K}^{-1} = \mathcal{A}^{-1} + \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}, \tag{2.5}$$

then using the fact that $\mathcal{K}\mathcal{K}^{-1} = I$, namely

$$\left[\mathcal{A} + \begin{pmatrix} -k^2M & 0 \\ 0 & 0 \end{pmatrix}\right] \cdot \left[\mathcal{A}^{-1} + \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}\right] = I,$$

we obtain by a direct computation that

$$-k^2M(V + X_1) + AX_1 + B^TX_3 = 0, \tag{2.6}$$

$$-k^2(B^TL^{-1} + MX_2) + AX_2 + B^TX_4 = 0, \tag{2.7}$$

$$BX_1 = 0, \quad BX_2 = 0. \tag{2.8}$$

On the other hand, we can see directly from (2.1) and the identity $\mathcal{A}\mathcal{A}^{-1} = I$ that

$$AV = I - B^TL^{-1}C^T, \quad BV = 0. \tag{2.9}$$

Then noting that $V + X_1$ is the (1,1) block of $\mathcal{K}^{-1}$ from (2.5), we get from (2.8) and (2.9) that

$$BT = B(V + X_1) = 0.$$

Multiplying (2.6) by C^T, we derive

$$-k^2B(V + X_1) + LX_3 = 0,$$

which gives

$$X_3 = k^2L^{-1}B(V + X_1) = 0. \tag{2.10}$$

Similarly, multiplying (2.7) by C^T, we obtain

$$-k^2(I + BX_2) + LX_4 = 0.$$

Combining this equality with the second relation in (2.8), we come to

$$X_4 = k^2L^{-1}. \tag{2.11}$$

Now we can substitute (2.11) into (2.7) to get

$$(A - k^2M)X_2 = 0, \tag{2.12}$$

which proves $X_2 = 0$.

Noting that we have proved $X_3 = 0$, then (2.6) reduces to $-k^2M(V + X_1) + AX_1 = 0$, or $(A - k^2M)(V + X_1) = AV$, which completes the desired proof. □

The following result can help us understand the leading block T of the inverse of $\mathcal{K}$ in (2.3).

Theorem 2.3 *The matrix $A + \eta B^T L^{-1} B - k^2 M$ is non-singular for any $\eta \neq k^2$, and its null space is exactly the same as that of A for $\eta = k^2$.*

Proof By means of the result (i) in Proportion 2.1, we can write for any $u \in \mathbb{R}^n$ that

$$u = u_A + u_B\,, \quad u_A \in ker(A), \;\; u_B \in ker(B). \tag{2.13}$$

Using this decomposition, it is easy to see that if

$$(A + \eta B^T L^{-1} B - k^2 M)u = 0,$$

then we have

$$(A - k^2 M)u_B + \eta B^T L^{-1} B u_A - k^2 M u_A = 0.$$

As the columns of C span the null space of A, there exists $p \in \mathbb{R}^m$ such that $u_A = Cp$. Using this and Proposition 2.1(iii), we can readily reduce the above identity to

$$(A - k^2 M)u_B + (\eta - k^2) B^T p = 0.$$

Multiplying its both sides by C^T and using Proposition 2.1(iii) again, we derive $BM^{-1}B^T p = 0$, so is $p = 0$, leading to $(A - k^2 M)u_B = 0$, or $u_B = 0$. Hence we have proved

$$u = u_A + u_B = Cp + u_B = 0,$$

and also the non-singularity of the desired matrix $A + \eta B^T L^{-1} B - k^2 M$.

Next, we consider the case with $\eta = k^2$ and show two matrices $A + k^2 B^T L^{-1} B - k^2 M$ and A have the same null space. First, we assume $u \in ker(A)$ and write $u = u_A + u_B$ as in (2.13), then the proof of the first part above shows that $u = u_A = Cp$. Using this and Proposition 2.1(iii), we can derive

$$(A + \eta B^T L^{-1} B - k^2 M)u = (\eta - k^2) B^T p = 0.$$

Now if u is in the null space of $A + k^2 B^T L^{-1} B - k^2 M$, and we can still write $u = u_A + u_B$ as in (2.13), and follow the proof of the first part above for the non-singularity of the matrix, but with $\eta = k^2$ now, we can deduce $(A - k^2 M)u_B = 0$. This implies $u_B = 0$, hence we know $u = u_A \in ker(A)$. □

The following result introduces a very crucial parameter η to the expression of the leading block T of the inverse of $\mathcal{K}$ in (2.3), and it can take an arbitrary value except for $\eta = k^2$.

Corollary 2.1 *For any $\eta \neq k^2$, it holds that*

$$\begin{aligned} T &= (A + \eta B^T L^{-1} B - k^2 M)^{-1} \left(I - B^T L^{-1} C^T \right) \\ &= (A + \eta B^T L^{-1} B - k^2 M)^{-1} - \frac{1}{\eta - k^2} C L^{-1} C^T. \end{aligned} \tag{2.14}$$

Proof Using the second relation, and then the first relation in (2.4), we readily see that

$$(A + \eta B^T L^{-1} B - k^2 M)T = (A - k^2 M)T = I - B^T L^{-1} C^T,$$

which implies the first relation in (2.14). To see the second relation, we first use the fact that $AC = 0$ from Proposition 2.1(iii), then use the first two relations in Proposition 2.1(iii) to deduce

$$\begin{aligned}(A + \eta B^T L^{-1} B - k^2 M)C &= \eta B^T L^{-1} BC - k^2 MC \\ &= \eta B^T L^{-1}(BM^{-1} B^T) - k^2 B^T \\ &= (\eta - k^2) B^T .\end{aligned}$$

This implies

$$(A + \eta B^T L^{-1} B - k^2 M)^{-1} B^T = \frac{1}{\eta - k^2} C .$$

Using this and the first relation in (2.14), we readily see the second relation. □

It is very interesting to see from the above relation that the matrix T that is independent of the parameter η looks closely to depend on η. In conclusion, we obtain from Theorem 2.2 and Corollary 2.1 the following formula for computing the inverse of the matrix $\mathcal{K}$ in (1.1):

$$\mathcal{K}^{-1} = \begin{pmatrix} (A + \eta B^T L^{-1} B - k^2 M)^{-1}(I - B^T L^{-1} C^T) & CL^{-1} \\ L^{-1} C^T & k^2 L^{-1} \end{pmatrix} . \tag{2.15}$$

This important explicit representation forms the basis in our construction of some new preconditioners in the next section.

3 New preconditioners and their spectral properties

The formula (2.15) suggests us some natural preconditioners for the saddle-point matrix $\mathcal{K}$ in (1.1) with the general non-vanishing wave numbers $k \neq 0$. However, the action of the (1,1) block of (2.15) can be very expensive to evaluate, since the matrix $B^T L^{-1} B$ is dense. To overcome the difficulty, we choose to replace the dense matrix $B^T L^{-1} B$ by the vector mass matrix M, as they are spectrally equivalent on the null space of A [8]. This leads to the following simplified preconditioner for the matrix $\mathcal{K}$:

$$\mathcal{P}^{-1} \equiv \begin{pmatrix} (A + \eta M - k^2 M)^{-1}(I - B^T L^{-1} C^T) & CL^{-1} \\ L^{-1} C^T & k^2 L^{-1} \end{pmatrix} . \tag{3.1}$$

For the simple case with vanishing wave number ($k = 0$) and $\eta = 1$, the preconditioner (3.1) reduces to the existing one $\mathcal{P}_0^{-1}$ in (1.4). We remark that the matrix $I - B^T L^{-1} C^T$ in (3.1) is an oblique projector, which can be very helpful in the analysis of Maxwell-type problems (see, e.g., [20, 21]). To ensure the non-singularity of the matrix $A + \eta M - k^2 M$ involved in (3.1), we can simply set the parameter $\eta > k^2$ so that it becomes symmetric positive definite. And it is important to note that this choice also guarantees the non-singularity of the entire matrix on the right-hand side of (3.1), as discussed below.

Theorem 3.1 *For any $\eta > k^2$, the matrix on the right-hand side of (3.1) is non-singular.*

Proof It is direct to check that the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ is given by

$$\mathcal{P}^{-1}\mathcal{K} = \begin{pmatrix} (A+\eta M - k^2 M)^{-1}(A + k^2 B^T L^{-1} B - k^2 M) + CL^{-1}B & 0 \\ 0 & I \end{pmatrix}.$$

Using Proposition 2.1 (i), we can further write the (1, 1) block of the above matrix as

$$\begin{aligned} & (A+\eta M - k^2 M)^{-1}(A + k^2 B^T L^{-1} B - k^2 M) + CL^{-1}B \\ = \; & (A+\eta M - k^2 M)^{-1}(A + k^2 B^T L^{-1} B - k^2 M) \\ & +(A+\eta M - k^2 M)^{-1}(\eta MCL^{-1}B - k^2 MCL^{-1}B) \\ = \; & (A+\eta M - k^2 M)^{-1}(A + \eta B^T L^{-1} B - k^2 M)\,, \end{aligned}$$

so the preconditioned system $\mathcal{P}^{-1}\mathcal{K}$ reads as

$$\mathcal{P}^{-1}\mathcal{K} = \begin{pmatrix} (A+\eta M - k^2 M)^{-1}(A + \eta B^T L^{-1} B - k^2 M) & 0 \\ 0 & I \end{pmatrix}. \tag{3.2}$$

We know that the leading block of $\mathcal{P}^{-1}\mathcal{K}$ in (3.2) is non-singular by Theorem 2.3, hence the desired conclusion follows. □

Note that $A+\eta M - k^2 M$ and its inverse are always symmetric positive definite for $\eta > k^2$. Actually, the original matrix $A + \eta B^T L^{-1} B - k^2 M$ can be also symmetric positive definite as shown below.

Theorem 3.2 *For any $\eta > k^2$ and $k^2 < \bar{\alpha}$, the matrix $A + \eta B^T L^{-1} B - k^2 M$ is symmetric positive definite.*

Proof For any $u \in \mathbb{R}^n$, we can write $u = u_A + u_B$ with $u_A \in ker(A)$ and $u_B \in ker(B)$. By Proposition 2.1, we know $u_A^T M u_B = 0$ and $u_A^T B^T L^{-1} B u_A = u_A^T M u_A$. Therefore, we can derive

$$\begin{aligned} & u^T(A + \eta B^T L^{-1} B - k^2 M)u \\ = \; & u_A^T(A + \eta B^T L^{-1} B - k^2 M)u_A + u_B^T(A + \eta B^T L^{-1} B - k^2 M)u_B \\ = \; & u_A^T(\eta B^T L^{-1} B - k^2 M)u_A + u_B^T(A - k^2 M)u_B \\ = \; & u_B^T(A - k^2 M)u_B + (\eta - k^2)u_A^T M u_A\,. \end{aligned} \tag{3.3}$$

Noting that $u_B^T A u_B \geq \bar{\alpha} u_B^T M u_B$ by Proposition 2.1(ii), we further deduce

$$u^T(A + \eta B^T L^{-1} B - k^2 M)u \geq (\bar{\alpha} - k^2)u_B^T M u_B + (\eta - k^2)u_A^T M u_A > 0\,,$$

which proves our desired result. □

Using the representation (3.2) and Theorem 3.2, we know for any $\eta > k^2$ and $k^2 < \bar{\alpha}$ that the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ is self-adjoint and positive definite with respect to the inner product

$$\langle x, y\rangle = x^T \begin{pmatrix} A + \eta M - k^2 M & 0 \\ 0 & I \end{pmatrix} y. \tag{3.4}$$

Therefore, we can apply the CG iteration [1] in this special inner product for solving the preconditioned system $\mathcal{P}^{-1}\mathcal{K}$. But for larger wave numbers, namely $k^2 \geq \bar{\alpha}$,

Theorem 3.2 does not ensure the positive definiteness of the preconditioned system, so the CG iteration may fail theoretically. However, as we see in Section 4, this is not the case numerically. Even if it fails, we may still apply the preconditioned MINRES with the above non-standard inner product.

We know the convergence rates of the CG and MINRES can be reflected often by the spectrum of the preconditioned system. For this purpose, we are going to study the spectral properties of the preconditioned system $\mathcal{P}^{-1}\mathcal{K}$. First, we present an interesting observation that the symmetric positive definiteness of the matrix $A + \eta B^T L^{-1} B - k^2 M$ depends in some sense only on the wave number k, not on η.

Theorem 3.3 *For any two numbers $\eta_1, \eta_2 > k^2$, $A+\eta_1 B^T L^{-1} B-k^2 M$ is symmetric positive definite if and only if $A + \eta_2 B^T L^{-1} B - k^2 M$ is symmetric positive definite.*

Proof For any $\eta_1 > k^2$, suppose that $A+\eta_1 B^T L^{-1} B-k^2 M$ is not symmetric positive definite. As this matrix is non-singular by Theorem 2.3, hence it is not symmetric semi-positive definite. Therefore, there exists $u \in \mathbb{R}^n$ satisfying $u^T(A + \eta_1 B^T L^{-1} B - k^2 M)u < 0$. But we can write $u = u_A + u_B$ with $u_A \in ker(A)$ and $u_B \in ker(B)$. Then we can see from (3.3) that $u_B \neq 0$ and $u_B^T(A - k^2 M)u_B < 0$. Now for any $\eta_2 > k^2$, we can easily check $u_B^T(A+\eta_2 B^T L^{-1} B-k^2 M)u_B = u_B^T(A-k^2 M)u_B < 0$, hence $A + \eta_2 B^T L^{-1} B - k^2 M$ is not symmetric positive definite. □

Next we present several results about the eigenvalues of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$.

Lemma 3.4 *For any $\eta > k^2$, $\lambda = 1$ is an eigenvalue of $(A + \eta M - k^2 M)^{-1}(A + \eta B^T L^{-1} B - k^2 M)$ with its algebraic multiplicity being m, and the rest of the eigenvalues are bounded by*

$$\frac{\bar{\alpha} - k^2}{\bar{\alpha} + \eta - k^2} < \lambda < 1. \tag{3.5}$$

Proof The result was proved in [8, Theorem 5.1] for $\eta = 1$ and $k^2 < 1$. But our desired results for an arbitrary positive η can be done similarly. □

The following result is a direct consequence of Lemma 3.4 by using the formula (3.2).

Theorem 3.5 *For any $\eta > k^2$, $\lambda = 1$ is an eigenvalue of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ with its algebraic multiplicity being $2m$, and the rest of the eigenvalues are bounded as in (3.5).*

Now we like to make some spectral comparisons between the two preconditioned systems generated by our new preconditioner $\mathcal{P}$ and the existing block triangular one $\mathcal{M}_{\eta,\varepsilon}$ in (1.5) for the saddle-point matrix $\mathcal{K}$. We first recall the following results from [24, Theorem 2.6].

Theorem 3.6 *For any $\eta > k^2$, both $\lambda_1 = 1$ and $\lambda_2 = -\frac{1}{\varepsilon(\eta-k^2)}$ are the eigenvalues of $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$, each with its algebraic multiplicity m. And the rest of the eigenvalues are bounded as in (3.5).*

We see from Theorems 3.5 and 3.6 that the spectra of $\mathcal{P}^{-1}\mathcal{K}$ and $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ are quite similar, except that the latter has an extra eigenvalue λ_2, with its algebraic multiplicity being m. This will be also confirmed numerically in the next section.

The block triangular preconditioners $\mathcal{M}_{\eta,\varepsilon}$ reduce to symmetric if we set $\varepsilon = \frac{1}{\eta}$:

$$\mathcal{M}_{\eta,1/\eta} = \begin{bmatrix} A + (\eta - k^2)M & 0 \\ 0 & \frac{1}{\eta}L \end{bmatrix}. \tag{3.6}$$

This preconditioner was analyzed and applied in [8, 23] along with the minimal residual (MINRES) iteration. We may observe from Theorems 3.5 and 3.6 that the eigenvalues of our preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ are a little better clustered than those of $\mathcal{M}_{\eta,1/\eta}^{-1}\mathcal{K}$ as its eigenvalue λ_2 is smaller than $\frac{\bar{\alpha}-k^2}{\bar{\alpha}+\eta-k^2}$. But our new preconditioner $\mathcal{P}$ can be applied with CG for $k^2 < \bar{\alpha}$, and MINRES for $k^2 \geq \bar{\alpha}$. And more importantly, as we see from our numerical experiments in the next section, we can also apply the new preconditioner $\mathcal{P}$ with CG even for $k^2 \geq \bar{\alpha}$ and the convergence is still rather stable, while CG with preconditioner $\mathcal{M}_{\eta,\varepsilon}$ in (3.6) breaks down most of the time.

On the other hand, if we choose $\varepsilon \neq 1/\eta$, the preconditioner $\mathcal{M}_{\eta,\varepsilon}$ is non-symmetric, and the methods like the generalized minimal residual method should be used, which are less economical than methods like CG or MINRES. Note that for $\varepsilon = -\frac{1}{\eta-k^2}$, we have $\lambda_2 = \lambda_1$, so $\lambda = 1$ is an eigenvalue of $\mathcal{M}_{\eta,\varepsilon}^{-1}\mathcal{K}$ with its algebraic multiplicity being $2m$, the same as for $\mathcal{P}^{-1}\mathcal{K}$.

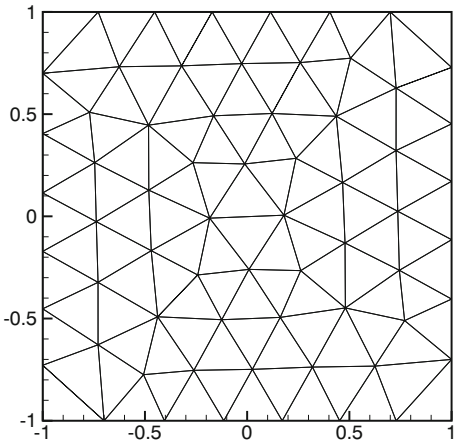
Now we consider the inner iterations associated with the new preconditioner $\mathcal{P}$. For any two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we can write

$$\begin{aligned} \mathcal{P}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} (A + \eta M - k^2 M)^{-1}(x - B^T L^{-1} C^T x) + C L^{-1} y \\ L^{-1} C^T x + k^2 L^{-1} y \end{pmatrix} \\ &= \begin{pmatrix} (A + \eta M - k^2 M)^{-1} x - \frac{1}{\eta - k^2} C L^{-1} C^T x + C L^{-1} y \\ L^{-1} C^T x + k^2 L^{-1} y \end{pmatrix}. \end{aligned}$$

So we need to solve two linear systems associated with the discrete Laplacian L and one with $A + (\eta - k^2)M$ at each evaluation of the action of $\mathcal{P}^{-1}$. Many fast solvers are available for solving these two symmetric and positive definite systems [10, 15]. We use the first form to implement the action of $\mathcal{P}^{-1}$. We know from Theorem 3.5 that a small difference $\eta - k^2$ may result in a better convergence of the preconditioned Krylov subspace methods. But if $\eta - k^2$ is too small, the matrix $A + (\eta - k^2)M$ would become nearly singular (we refer to [7, 8] for the discussion about the approximation of the matrix $(A + (\eta - k^2)M)^{-1}$ to the pseudo-inverse of A when $(\eta - k^2) \to 0$).

We know that the parameter $\bar{\alpha}$ depends only on the shape regularity of the mesh and the approximation order of the finite elements used, and it is independent of the mesh size [9, Theorem 4.7]. Numerically we may expect an upper bound for k^2 that

Fig. 1 Mesh G1: n+m=187



ensures the positive definiteness of $A + \eta B^T L^{-1} B - k^2 M$, and this bound should be independent of the mesh size. We shall check this numerically in the next section.

4 Numerical experiments

In this section, we present several two-dimensional numerical experiments to demonstrate and compare the spectral distributions of the preconditioned systems of the saddle-point problem (1.1) with the existing preconditioner $\mathcal{M}_{\eta,1/\eta}$ in (1.5) and the new one $\mathcal{P}$ in (3.1), as well as to compare the performance of these preconditioners. The edge elements of lowest order are used for the discretization of the system (1.2) in a square domain $\bar{\Omega} =: \{(x, y); -1 \le x \le 1, -1 \le y \le 1\}$ or an L-shaped domain (see Figs. 1 and 2). The square domain is partitioned using unstructured simplicial meshes generated by EasyMesh [18], where the desired side lengths of the

Fig. 2 Mesh L1: n+m=185

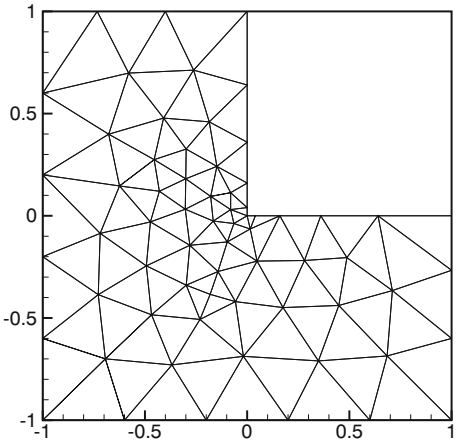
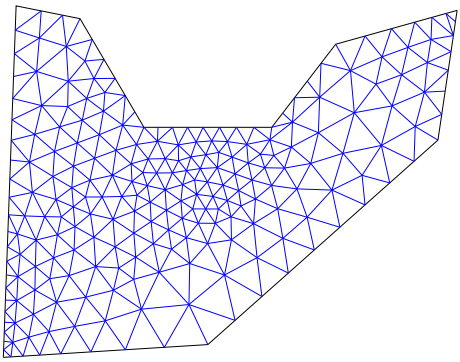


Fig. 3 Mesh U1: n+m=711



triangles that contain one of the vertices of the domain are set to be the same, resulting in a sequence of recursively refined meshes G1 through G5, and a corresponding sequence of the saddle-point systems (1.1) of size $m + n$= 187, 437, 1777, 7217, 23,769 respectively. Similarly for the L-shaped domain, we apply the EasyMesh to generate a sequence of recursively refined meshes L1 through L5, where the desired side lengths of the triangles that contain the origin are one-tenth of the desired side lengths of the triangles that contain other vertices of the domain, and a corresponding sequence of the saddle-point systems (1.1) of size $m + n$= 185, 409, 1177, 5325, 29,277 respectively. To test the effect of the geometry of the domain and the irregular meshes on the algorithms, we further generate a series of less spatially regular meshes U1 through U5 on an irregular polygonal domain Ω (see Figs. 3 and 4 for the meshes U1 and U4), resulting in a sequence of the saddle-point systems (1.1) with the total degrees of freedom being $m + n$= 711, 1712, 4355, 9544, and 24,665 respectively.

We use MATLAB on a laptop (inter(R) Core(TM) i7-4510U CPU @ 2.00 GHz, 2.60 GHz, 4-GB RAM) to implement all numerical iterative solvers. The solver with $A + (\eta - k^2)M$ is achieved by PCG with the Hiptmair-Xu preconditioner [10], while the solver with the discrete Laplacian L is realized by PCG with an incomplete Cholesky factorization as a preconditioner.

The right-hand side of (1.1), denoted by b, is set to be a vector with all components being ones unless specified otherwise, and the zero vector is used as the initial guess

Fig. 4 Mesh U4: n+m=9544

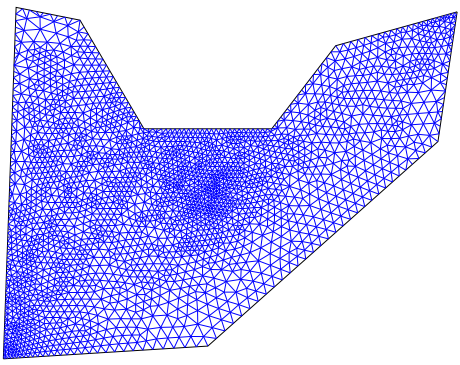


Table 1 Smallest eigenvalue of matrix A_η in magnitude with $\eta = k^2 + 1$

k	G2	G3	G4	k	L2	L3	L4	k	U2	U3	U4
0	0.4738	0.4776	0.4769	0	0.4582	0.4758	0.4654	0	0.3461	0.3381	0.3409
1	0.4738	0.4776	0.4769	1	0.2575	0.2704	0.2753	0.75	0.0177	0.0178	0.0182
1.55	0.0360	0.0369	0.0373	1.2	0.0128	0.0175	0.0200	0.8	−0.0279	−0.0264	−0.0265
1.6	−0.0543	−0.0536	−0.0533	1.25	−0.0556	−0.0530	−0.0512	1	−0.2400	−0.2321	−0.2345
2	−0.8988	−0.8875	−0.8823	2	−1.4249	−1.4580	−1.4674	2	−2.0132	−1.9485	−1.9736
4	−7.9434	−7.8425	−7.7907	4	−8.3664	−8.3974	−8.4450	4	−9.1374	−8.8463	−8.9761

Eigenvalues above the dotted line are positive: A_η is positive definite; eigenvalues under the dotted line are negative: A_η is not positive definite

$x^{(0)}$ for all iterations. We run PCG [22] with our new preconditioner $\mathcal{P}$ for solving the saddle-point system (1.1), and the preconditioned MINRES with the block triangular preconditioner $\mathcal{M}_{\eta,1/\eta}$, and write these two methods as $\mathcal{P}$-CG and $\mathcal{M}$-MINRES respectively in all tables. In all our numerical examples, the outer iterations are terminated based on the criterion $\|b - \mathcal{K}x^{(k)}\|_2 \le 10^{-6}\|b\|_2$, where $x^{(k)}$ is the kth iterate. We take the parameter $\eta = k^2 + 1$ and set the stopping criteria for all Laplacian solvers (including both L-solvers and those Laplacian solvers inside the Hiptmair-Xu preconditioner) to be a relative l_2-norm error of the residual less than the same tolerance, unless otherwise stated. The computing times (in seconds) may also be listed, which include the times spent by the incomplete Cholesky factorizations for all Laplacian solvers.

4.1 Numerical spectral analysis

We know from (3.2) and Theorem 3.2 that PCG can be applied with our new preconditioner $\mathcal{P}$ under the special inner product defined in (3.4) if the matrix

$$A_\eta = \begin{pmatrix} A + \eta B^T L^{-1} B - k^2 M & 0 \\ 0 & I_m \end{pmatrix}$$

is symmetric positive definite. We shall conduct some experiments below to check the positive definiteness of this matrix. For this purpose, we compute its smallest eigenvalues corresponding to different wave numbers. The results are shown in Tables 1 (with $\eta = k^2 + 1$) and 2 (with various ηs).

Table 2 Smallest eigenvalues of matrix A_η in magnitude: all well bounded from zero

Mesh	G1	G2	G3	G4	L1	L2	L3	L4
$k = 4, \eta = 17$	0.4677	0.4738	0.4776	0.4769	0.4787	0.4582	0.4758	0.4654
$k = 4, \eta = 24$	1.8963	−2.0601	−2.0846	−2.0966	−1.9134	−1.9650	−1.9311	−1.9705
$k = 2, \eta = 5$	0.4677	0.4738	0.4776	0.4769	−0.2541	−0.2640	−0.2705	−0.2692
$k = 2, \eta = 6$	0.5061	0.5310	0.5311	0.5338	−0.2540	−0.2640	−0.2705	−0.2692

As we may see from Table 1 that on the meshes G2 through G4, the matrices A_η are symmetric and positive definite with $k = 0, 1, 1.55$, but not positive definite for $k \geq 1.6$. Similarly, on the meshes L2 through L4, the matrices A_η are symmetric and positive definite with $k = 0, 1, 1.2$, but not positive definite for $k \geq 1.25$. Similar observations can be made on the meshes U2 through U4. As predicted by Theorem 3.2, the definiteness of A_η is independent of mesh size. This is indeed confirmed by the results in Tables 1 and 2: once these smallest eigenvalues get stabilized on a rather coarse mesh, they do not change much with further mesh refinements.

As we know, A_η is symmetric and positive definite for smaller k, i.e., $k^2 < \bar{\alpha}$, thus PCG can be applied with our new preconditioner $\mathcal{P}$ instead of MINRES, although the original system $\mathcal{K}$ is indefinite. When k is larger, i.e., $k^2 \geq \bar{\alpha}$, the corresponding preconditioned matrices are no longer positive definite, then MINRES should be used theoretically. However, our numerical experiments show that PCG does not fail even for $k^2 \geq \bar{\alpha}$, in fact, PCG converges very well and stably for our numerical examples; see some examples in Section 4.2. But this is not the case when PCG is applied with the preconditioner $\mathcal{M}_{\eta,1/\eta}$. To see this, we have re-run all the experiments in Table 6, but with the CG iteration now, instead of MINRES. In each of the 30 numerical experiments, we have always experienced the case that one dividend becomes too small, which causes the breakdown of the iterative process. The reasons behind are simple: we need to divide by $p_k^T \mathcal{K} p_k$ (with p_k being the kth search direction) at the kth CG iteration with preconditioner $\mathcal{M}_{\eta,1/\eta}$, and to divide by $p_k^T A_\eta p_k$ at the kth CG iteration with the new preconditioner $\mathcal{P}$, due to the existence of a special inner product (3.4). Figure 5 shows the distributions of all the small eigenvalues (smaller than 0.3) of the two matrices $\mathcal{K}$ and A_η for $k = 4$ (noting that most eigenvalues are larger than 0.3, but not shown in the figure). We know these smaller and negative eigenvalues contribute mainly to the breakdown of the iterations. As one can see from the figure, $\mathcal{K}$ has many more small eigenvalues (the red part) than A_η (the blue part), which explains clearly the strong instability of PCG with the preconditioner $\mathcal{M}_{\eta,1/\eta}$ and the good stability of PCG with the new preconditioner $\mathcal{P}$.

To check if we can apply PCG with the new preconditioner $\mathcal{P}$ on general meshes, we further investigate the influence of the mesh size on the smallest eigenvalue of A_η in magnitude. As confirmed by the results in Table 1, Table 2 verifies again that the smallest eigenvalues of A_η in magnitude are basically independent of the mesh size,

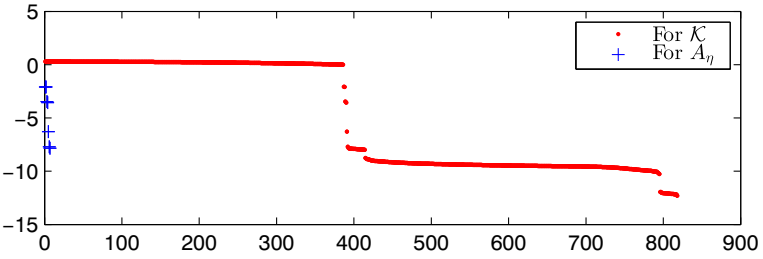


Fig. 5 Distributions of small eigenvalues (smaller than 0.3) of the coefficient matrix $\mathcal{K}$ (red part) and the matrix A_η (blue part) on grid G3 for $k = 4$ and $\eta = k^2 + 1$. There are many more eigenvalues close to zero for $\mathcal{K}$ (red part) than for A_η (blue part)

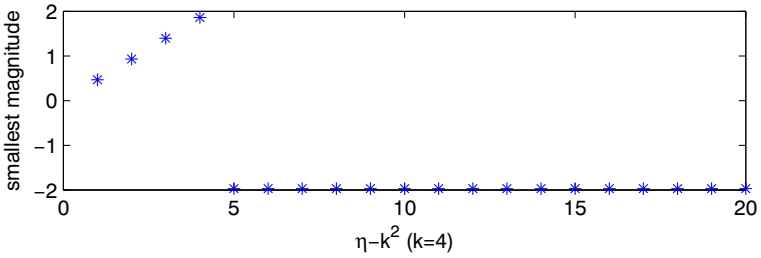


Fig. 6 Smallest eigenvalue of the matrix A_η in magnitude on grid L4 for $k = 4$ and $\eta - k^2 = 1, 2 \cdots 20$

so the mesh size can be very fine in order to resolve the highly oscillatory waves in the high-frequency cases. And more importantly, these smallest eigenvalues are all well separated from the origin. These observations suggest that we may also apply PCG with the new preconditioner $\mathcal{P}$ for large wave numbers, and all our numerical experiments in this section have demonstrated very good stability and convergence of PCG with the new preconditioner $\mathcal{P}$.

We end this section with some more numerical results to show the influence of the difference $\eta - k^2$ on the smallest eigenvalue of A_η and the clustering of the eigenvalues of the preconditioned system $\mathcal{P}^{-1}\mathcal{K}$. Figure 6 shows the influence of the difference $\eta - k^2$ on the smallest eigenvalue of A_η in magnitude. A larger $\eta - k^2$ makes the smallest eigenvalue of A_η bigger in magnitude. Figure 7 plots the eigen-distribution of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ on mesh G3 with a different wave number k. We can see from the figure that the eigenvalues for $k = 0$ and 1 are well bounded, and there are only a few eigenvalues that lie between 0.22 and 0.8, while all the remaining eigenvalues stay in the range 0.8 and 1. These results are consistent with our theoretical prediction (see Theorem 3.5). For $k = 2, 4$, we see negative eigenvalues: the higher the wave number is, the less clustered the eigenvalues are.

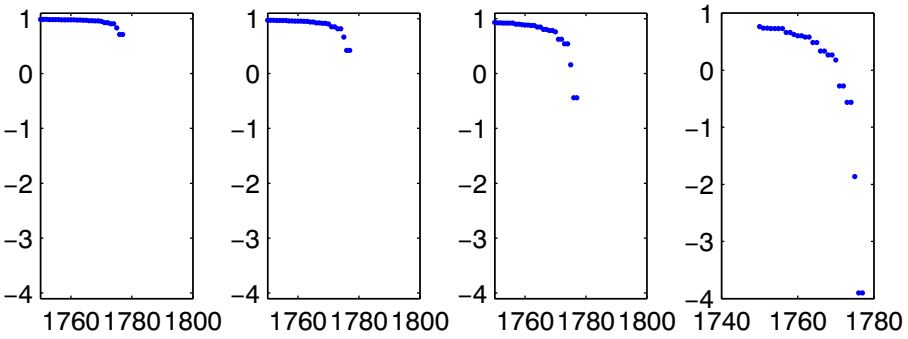


Fig. 7 Distributions of smallest (27) eigenvalues of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{K}$ on grid G3 (from left to right: $k = 0, 1, 2, 4$, with $\eta = k^2 + 1$)

4.2 Basic numerical performance

We have shown in Tables 3 and 4 the numbers of iteration and the overall execution times for the two methods, $\mathcal{P}$-CG and $\mathcal{M}_{\eta,1/\eta}$-MINRES, with different meshes and wave numbers, using the same tolerance 10^{-6} for all inner solvers associated with both $A+(\eta-k^2)M$ and the discrete Laplacian L. As we may observe from the table, the required numbers of iteration for the new method $\mathcal{P}$-CG are generally smaller than those for $\mathcal{M}_{\eta,1/\eta}$-MINRES, and this is consistent with our theoretical prediction in Section 3. But $\mathcal{M}_{\eta,1/\eta}$-MINRES takes mostly about $19 \sim 50\%$ more times than $\mathcal{P}$-CG. We also observe that the required numbers of iteration are basically independent of mesh size, and this is a very desired property in application. These experiments are done to make a general comparison between these two methods, and more efficient Laplacian solvers should be used in applications.

Next, we conduct some numerical experiments to find the effects of inexact inner solvers with $A+(\eta-k^2)M$ and the discrete Laplacian L on the performance of the preconditioners, when the tolerance for the Laplacian solvers inside the Hiptmair-Xu preconditioners is set to be fixed at 10^{-1}. The results are shown in Tables 5

Table 3 Meshes G1 to G5 and $\eta = k^2 + 1$

k	0	1.0	1.55	1.6	2	4
Mesh G1						
$\mathcal{P}$-CG	5 (0.6802)	6 (0.7485)	11 (1.2736)	11 (1.268)	11(1.2619)	25 (2.6965)
$\mathcal{M}$-MINRES	7 (0.8907)	9 (1.1619)	15 (1.6414)	15 (1.6309)	14 (1.5287)	31 (3.2299)
Ratio	1.3095	1.5523	1.2888	1.2863	1.2114	1.1978
Mesh G2						
$\mathcal{P}$-CG	5 (0.949)	7 (1.1789)	12 (1.9092)	12 (1.8898)	11 (1.7518)	28 (4.1755)
$\mathcal{M}$-MINRES	7 (1.2051)	9 (1.4634)	15 (2.3171)	15 (2.261)	15 (2.2492)	31 (4.4716)
Ratio	1.2698	1.2414	1.2137	1.1965	1.2839	1.0709
Mesh G3						
$\mathcal{P}$-CG	5 (1.9158)	6 (2.1352)	11 (3.5818)	11 (3.6182)	11 (3.5805)	25 (7.5431)
$\mathcal{M}$-MINRES	8 (2.7483)	9 (2.9572)	15 (4.6596)	15 (4.6461)	14 (4.337)	31 (9.0741)
Ratio	1.4346	1.385	1.3009	1.2841	1.2113	1.203
Mesh G4						
$\mathcal{P}$-CG	5 (6.2827)	6 (7.2006)	9 (10.3151)	9 (10.2293)	11 (12.1859)	24 (25.2278)
$\mathcal{M}$-MINRES	7 (8.1305)	9 (10.0955)	12 (13.2987)	12 (13.0848)	14 (15.0183)	29 (29.4057)
Ratio	1.2941	1.402	1.2892	1.2791	1.2324	1.1656
Mesh G5						
$\mathcal{P}$-CG	5 (32.9871)	6 (37.8760)	9 (53.7789)	9 (53.6557)	11 (63.9640)	23 (127.2520)
$\mathcal{M}$-MINRES	7 (43.0383)	8 (48.4078)	12 (69.4350)	12 (69.3690)	14 (79.6365)	29 (156.8750)
Ratio	1.3047	1.2781	1.2911	1.2929	1.2450	1.2328

Rows for $\mathcal{P}$-CG and $\mathcal{M}$-MINRES: numbers of iteration (execution times); rows for *Ratio*: ratios between the times spent by two methods

Table 4 Meshes L1 to L5 and $\eta = k^2 + 1$

k	0	1.0	1.2	1.25	2	4
Mesh L1						
$\mathcal{P}$-CG	5 (0.8006)	7 (1.0285)	9 (1.2749)	8 (1.1593)	10 (1.3678)	25 (3.3425)
$\mathcal{M}$-MINRES	8 (1.1821)	9 (1.3191)	12 (1.7786)	10 (1.3660)	15 (1.9633)	31 (3.8042)
Ratio	1.4765	1.2826	1.3951	1.1783	1.4353	1.1382
Mesh L2						
$\mathcal{P}$-CG	6 (1.3000)	7 (1.3541)	9 (1.7064)	8 (1.5272)	12 (2.2195)	28 (4.8256)
$\mathcal{M}$-MINRES	8 (1.5563)	9 (1.6863)	12 (2.1860)	10 (1.8770)	15 (2.7282)	32 (5.4724)
Ratio	1.1972	1.2454	1.2810	1.2290	1.2292	1.1340
Mesh L3						
$\mathcal{P}$-CG	5 (1.8053)	7 (2.1953)	9 (2.7333)	8 (2.5120)	12 (3.5724)	25 (7.0046)
$\mathcal{M}$-MINRES	8 (2.5005)	9 (2.7071)	11 (3.2302)	11 (3.3899)	15 (4.2672)	31 (8.4044)
Ratio	1.3851	1.2331	1.1818	1.3495	1.1945	1.1998
Mesh L4						
$\mathcal{P}$-CG	5 (5.3455)	7 (6.9149)	8 (7.7637)	8 (7.7648)	12 (11.0665)	24 (21.0020)
$\mathcal{M}$-MINRES	8 (7.7230)	9 (8.5107)	12 (11.0335)	12 (11.0477)	15 (13.4594)	31 (26.6709)
Ratio	1.4448	1.2308	1.4212	1.4228	1.2162	1.2699
Mesh L5						
$\mathcal{P}$-CG	5 (53.0033)	7 (69.6419)	8 (78.3754)	8 (78.5196)	10 (95.8931)	26 (231.7050)
$\mathcal{M}$-MINRES	8 (78.9650)	9 (87.1453)	10 (95.7683)	10 (95.6296)	13 (121.6840)	30 (261.2080)
Ratio	1.4898	1.2513	1.2219	1.2179	1.2690	1.1273

Rows for $\mathcal{P}$-CG and $\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES: numbers of iteration (execution times); rows for *Ratio*: ratios between the times spent by two methods

and 6, from which we may observe that the execution times on the lower triangular part are smaller than the times on the upper triangular part, which shows that a good pair of tolerance strategies for the two inner solvers would be a looser tolerance for the solver with $A + (\eta - k^2)M$ and a tighter tolerance for the solver with L. Indeed, we know the solver with $A + (\eta - k^2)M$ is much more expensive than the solver with L. This good pair of tolerance strategies may be taken later for more comparisons.

Table 5 Numbers of iteration and execution times of $\mathcal{P}$-CG on grid L4 with $k = 0$, $\eta = 1$ and different tolerances for the inner solvers with $A + (\eta - k^2)M$ (listed on the left side) and L (listed on the top)

	1e−5	1e−4	1e−3	1e−2	1e−1
1e−5	5 (2.0808)	5 (2.1045)	5 (2.1666)	6 (2.3585)	13 (4.5092)
1e−4	5 (1.6821)	5 (1.6615)	5 (1.6706)	6 (2.0957)	13 (3.9118)
1e−3	5 (1.3027)	5 (1.3421)	5 (1.3220)	6 (1.5743)	13 (3.0280)
1e−2	7 (1.1731)	6 (1.1331)	6 (1.1461)	6 (1.1881)	13 (2.3217)
1e−1	8(0.7564)	9 (0.7813)	8 (0.7640)	7 (0.8225)	13 (1.5585)

Tolerance of Laplacian solvers inside the Hitmair-Xu preconditioner is set to 1e−1

Table 6 Numbers of iteration and execution times of $\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES on grid L4 with $k = 0$, $\eta = 1$, and different tolerances for the inner solvers with $A + (\eta - k^2)M$ (listed on the left) and L (listed on the top)

	1e−5	1e−4	1e−3	1e−2	1e−1
1e−5	8 (2.9245)	8 (3.0198)	9 (3.1409)	10 (3.5018)	21 (7.5691)
1e−4	8 (2.4044)	8 (2.4364)	9 (2.6340)	10 (2.9974)	21 (5.8159)
1e−3	8 (1.7549)	8 (1.7433)	9 (2.0404)	10 (2.3071)	21 (4.5918)
1e−2	10 (1.4443)	10 (1.4053)	9 (1.3736)	10 (1.6640)	21 (3.5088)
1e−1	18 (1.2268)	18 (1.2463)	15 (0.9492)	17 (1.1446)	21 (2.0830)

Tolerance of Laplacian solvers inside the Hitmair-Xu preconditioner is set to 1e−1

Table 7 lists the ratios between the times spent by $\mathcal{M}_{\eta,1/\eta}$-MINRES and $\mathcal{P}$-CG from Tables 5 and 6, and confirms that the new method $\mathcal{P}$-CG has a clear advantage over the method $\mathcal{M}_{\eta,1/\eta}$-MINRES.

We may recall from Tables 5 and 6 that both the $\mathcal{P}$-CG and $\mathcal{M}_{\eta,1/\eta}$-MINRES methods perform best with the tolerance pair (1e−1, 1e−5) for the inner solvers with $A + (\eta - k^2)M$ and L respectively for the parameters $k = 0$ and $\eta = 1$. In Table 8, we show some more experiments to further investigate this phenomenon. We can see from this table that the methods perform best with the tolerance pair (1e−1,1e−5) for $k = 0$, 1, 1.2, or 1.25. But for $k = 4$, one may need a relatively tighter inner solver. And again, this table confirms that the new method $\mathcal{P}$-CG has a clear advantage over the method $\mathcal{M}_{\eta,1/\eta}$-MINRES.

4.3 Experiments for non-trivial geometries and less spatially regular meshes

In this subsection, we present some numerical results on irregular meshes U1 through U5 (see Figs. 3 and 4 for U1 and U4 respectively). We first recall from the numerical spectral results shown in Table 1 for U2 through U4, similarly to the regular meshes L2 through L4 or G2 through G4, we observe that the positive definiteness of the matrix A_η can be determined by the comparison of k^2 with a mesh-independent critical value, as suggested by Theorem 3.2.

Then we have conducted the experiments to compare the performance of the two methods, $\mathcal{P}$-CG and $\mathcal{M}_{\eta,1/\eta}$-MINRES, on the irregular meshes U1 through U5. The numbers of iteration and the overall execution times of the two methods are shown

Table 7 Ratios between times taken by $\mathcal{M}_{\eta,1/\eta}$-MINRES and $\mathcal{P}$-CG as listed in Tables 6 and 5

	1e−5	1e−4	1e−3	1e−2	1e−1
1e−5	1.4054	1.4349	1.4497	1.4848	1.6786
1e−4	1.4294	1.4664	1.5767	1.4303	1.4868
1e−3	1.3471	1.2990	1.5435	1.4655	1.5164
1e−2	1.2311	1.2402	1.1985	1.4005	1.5113
1e−1	1.6219	1.5952	1.2425	1.3915	1.3365

Table 8 First column: tolerance pairs for the inner solvers with $A + (\eta - k^2)M$ and L; first line at each row: times spent by $\mathcal{P}$-CG ($\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES) on grid L4 with different wave numbers k; second line at each row: ratios between the times spent by $\mathcal{M}_{\eta,\frac{1}{\eta}}$-MINRES and $\mathcal{P}$-CG

k	0	1	1.2	1.25	2	4
1e−4/1e−5	3.4148 (4.5445)	3.7516 (5.6651)	4.6408 (5.9792)	4.5615 (5.9787)	5.9447 (7.5844)	12.3785 (13.2688)
	1.3308	1.5100	1.2884	1.3107	1.2758	1.0719
1e−3/1e−5	2.3986 (3.1542)	3.2000 (4.4446)	3.4226 (4.7965)	3.7015 (4.8802)	4.4020 (5.4902)	10.1951 (10.8177)
	1.3150	1.3889	1.4014	1.3185	1.2472	1.0611
1e−2/1e−5	1.5173 (2.0689)	1.9599 (2.6223)	2.6071 (3.6012)	2.6189 (3.2503)	3.0453 (3.4254)	6.8830 (8.1561)
	1.3635	1.3380	1.3813	1.2411	1.1248	1.1850
1e−1/1e−5	0.8272 (1.3258)	1.2529 (1.5368)	2.7211 (4.9149)	1.8538 (3.4134)	3.7716 (8.3334)	7.5735 (9.7269)
	1.6027	1.2266	1.8062	1.8413	2.2095	1.2843

Table 9 Meshes U1 to U5 and $\eta = k^2 + 1$

k	0	0.75	0.8	1	2	4
Mesh U1						
$\mathcal{P}$-CG	7 (0.7492)	10 (1.018)	10 (1.0288)	10 (1.0335)	20 (1.9001)	59 (5.2988)
$\mathcal{M}$-MINRES	9 (0.9153)	12 (1.3417)	12 (1.1973)	12 (1.2099)	20 (1.8728)	70 (6.3153)
Ratio	1.2218	1.318	1.1638	1.1707	0.9857	1.1918
Mesh U2						
$\mathcal{P}$-CG	7 (1.054)	10 (1.3704)	10 (1.411)	10 (1.3915)	19 (2.4406)	66 (7.9437)
$\mathcal{M}$-MINRES	9 (1.2544)	12 (1.5561)	12 (1.5747)	12 (1.5871)	20 (2.5195)	67 (7.9125)
Ratio	1.1901	1.1355	1.116	1.1406	1.0323	0.9961
Mesh U3						
$\mathcal{P}$-CG	6 (2.18)	9 (3.0904)	10 (3.3467)	10 (3.3281)	19 (5.9447)	65 (18.3948)
$\mathcal{M}$-MINRES	9 (3.0537)	12 (3.7377)	12 (3.7001)	12 (3.6424)	22 (6.052)	66 (17.5027)
Ratio	1.4008	1.2095	1.1056	1.0945	1.0181	0.9515
Mesh U4						
$\mathcal{P}$-CG	7 (5.5643)	9 (6.7966)	9 (6.6811)	9 (6.7945)	17 (11.6895)	59 (38.1788)
$\mathcal{M}$-MINRES	9 (7.146)	11 (8.0544)	12 (8.3088)	12 (8.8649)	21 (14.2766)	67 (40.6807)
Ratio	1.2842	1.1851	1.2436	1.3047	1.2213	1.0655
Mesh U5						
$\mathcal{P}$-CG	7 (19.8287)	9 (25.7299)	9 (23.9108)	13 (33.5189)	23 (57.343)	59 (146.2211)
$\mathcal{M}$-MINRES	9 (26.6412)	12 (32.3105)	12 (39.6709)	17 (41.3345)	29 (69.131)	66 (156.1844)
Ratio	1.3436	1.2558	1.6591	1.2332	1.2056	1.0681

Rows for $\mathcal{P}$-CG and $\mathcal{M}$-MINRES: numbers of iteration (execution times); rows for *Ratio*: ratios between the times spent by two methods

Table 10 Meshes U1 to U5, with $\eta = k^2 + 1$ and randomly generated right-hand sides

k	0	0.75	0.8	1	2	4
Mesh U1						
$\mathcal{P}$-CG	7 (0.7605)	10 (1.0301)	10 (1.0354)	10 (1.0468)	20 (2.0831)	59 (5.4389)
$\mathcal{M}$-MINRES	9 (0.8946)	12 (1.1807)	12 (1.265)	12 (1.1841)	20 (2.0013)	69 (6.0542)
Ratio	1.1762	1.1463	1.2218	1.1312	0.9608	1.1131
Mesh U2						
$\mathcal{P}$-CG	7 (1.0199)	10 (1.3646)	10 (1.4556)	10 (1.3907)	18 (2.3448)	66 (8.0441)
$\mathcal{M}$-MINRES	9 (1.2591)	12 (1.6478)	12 (1.5936)	12 (1.5427)	21 (2.5964)	68 (7.9414)
Ratio	1.2345	1.2075	1.0948	1.1093	1.1073	0.9872
Mesh U3						
$\mathcal{P}$-CG	7 (2.3734)	9 (2.859)	10 (3.0765)	10 (3.117)	19 (5.7769)	62 (17.538)
$\mathcal{M}$-MINRES	9 (2.7132)	12 (3.7343)	12 (4.0946)	12 (3.4991)	22 (7.1118)	72 (19.8068)
Ratio	1.1432	1.3061	1.3309	1.1226	1.2311	1.1294
Mesh U4						
$\mathcal{P}$-CG	7 (5.1162)	9 (6.3297)	9 (6.5258)	9 (6.2415)	18 (11.6957)	62 (38.9867)
$\mathcal{M}$-MINRES	9 (6.5882)	11 (8.0213)	11 (8.2975)	12 (8.1823)	18 (11.5939)	67 (42.5028)
Ratio	1.2877	1.2672	1.2715	1.311	0.9913	1.0902
Mesh U5						
$\mathcal{P}$-CG	7 (21.9416)	9 (29.1238)	9 (27.5198)	9 (28.6648)	17 (48.6905)	59 (157.6150)
$\mathcal{M}$-MINRES	9 (25.8734)	11 (30.1301)	11 (28.5676)	11 (29.8763)	21 (51.3032)	66 (164.8363)
Ratio	1.1792	1.0346	1.0381	1.0423	1.0537	1.0458

Rows for $\mathcal{P}$-CG and $\mathcal{M}$-MINRES: numbers of iteration (execution times); rows for *Ratio*: ratios between the times spent by two methods

in Table 9 when the right-hand side functions are chosen to be all ones and Table 10 when the the right-hand side functions are generated randomly. We observe again that the required numbers of iteration for the proposed new method $\mathcal{P}$-CG are generally smaller than those for $\mathcal{M}_{\eta,1/\eta}$-MINRES, which justifies our theoretical results in Section 3, and the required numbers of iteration are basically independent of the mesh size. We can also find that when k^2 is smaller than or near the critical value, the iteration performance of the $\mathcal{P}$-CG is always better than $\mathcal{M}_{\eta,1/\eta}$-MINRES, and when k^2 is relatively bigger than the critical value, $\mathcal{P}$-CG still performs better than $\mathcal{M}_{\eta,1/\eta}$-MINRES in most cases.

5 Concluding remarks

Based on some special properties and structures of the saddle-point system arising from the edge element discretization of the time-harmonic Maxwell's equations, we have derived a family of new preconditioners for general non-vanishing wave numbers. Several nice properties of the new preconditioners are presented and the spectral estimates of the corresponding preconditioned systems are established. It is important to note that the preconditioned systems are symmetric positive definite with respect to a special inner product for wave numbers that are smaller than a positive critical value, so the PCG iteration can be applied with the new preconditioners for solving the preconditioned systems. When wave numbers are larger than the critical value, our theory does not ensure the convergence of the PCG, but MINRES can then be applied with the new preconditioners; moreover, our various numerical experiments have shown that the PCG also works very effectively and stably. The performances of the new preconditioners and some existing effective preconditioners are compared through several numerical examples, and the results indicate a clear advantage of the new preconditioners in terms of the stability and computing times when the exact or inexact inner solvers are considered.

We may recall that all the analyses and results in this work have been conducted for the time-harmonic Maxwell's equation (1.2) that can be viewed as a simplified Maxwell system in a vacuum. But the new preconditioners may be extended to the time-harmonic Maxwell's equations in a homogeneous medium, where the corresponding electric permittivity and magnetic permeability are smooth variable functions. This extension can be realized by combining the analyses in this work with the corresponding results of Proposition 2.1 to the homogeneous medium, which were established in [15].

Furthermore, we emphasize that all the analyses and results in this work are based on the use of the exact inner solvers involved in the new preconditioners. But it is more realistic in applications to replace the exact inner solvers by inexact ones. Then most of our results in this work may not be true. It is an important topic to explore whether it is possible to extend our results to the cases with inexact inner solvers.

Acknowledgments The authors would like to thank the anonymous referees for their many insightful and constructive comments and suggestions that have helped us improve the structure and results of the paper essentially.

Funding information The research of this project was financially supported by the National Natural Science Foundation of China under grants 11571265 and 11471253. The work of J. Zou was substantially supported by Hong Kong RGC General Research Fund (Project 14304517) and NSFC/Hong Kong RGC Joint Research Scheme 2016/17 (Project N_CUHK437/16).

References

1. Ashby, S.F., Manteuffel, T.A., Saylor, P.E.: A taxonomy for conjugate gradient methods. SIAM J. Numer. Anal. **27**, 1542–1568 (1990)
2. Benzi, M., Golub, G.H., Liesen, J.: Numerical solutions of saddle point problems . Acta Numerica **14**, 1–137 (2005)
3. Boffi, D., Brezzi, F., Fortin, M.: Mixed finite element methods and applications, Vol. 44 of Springer series in computational mathematics. Springer, Berlin (2013)
4. Chen, Z., Du, Q., Zou, J.: Finite element methods with matching and nonmatching meshes for Maxwell equations with discontinuous coefficients. SIAM J. Numer. Anal. **37**, 1542–1570 (2000)
5. Cheng, G.-H., Huang, T.-Z., Shen, S.-Q.: Block triangular preconditioners for the discretized time-harmonic Maxwell equations in mixed form. Comput. Phys. Commun. **180**, 192–196 (2009)
6. Demkowicz, L., Vardapetyan, L.: Modeling of electromagnetic absorption/scattering problems using hp-adaptive finite elements. Comput. Methods Appl. Mech. Engrg. **152**, 103–124 (1998)
7. Estrin, R., Greif, C.: On nonsingular saddle-point systems with a maximally rank deficient leading block. SIAM J. Matrix Anal. Appl. **36**, 367–384 (2015)
8. Greif, C., Schötzau, D.: Preconditioners for the discretized time-harmonic Maxwell equations in mixed form. Numer. Lin. Algebra Appl. **14**, 281–297 (2007)
9. Hiptmair, R.: Finite elements in computational electromagnetism. Acta Numerica **11**, 237–339 (2002)
10. Hiptmair, R., Xu, J.: Nodal auxiliary space preconditioning in H(curl) and H(div) spaces. SIAM J. Numer. Anal. **45**, 2483–2509 (2007)
11. Houston, P., Perugia, I., Schötzau, D.: Mixed discontinuous Galerkin approximation of the Maxwell operator: Non-stabilized formulation. J. Sci. Comput. **22-23**, 315–346 (2005)
12. Hu, Q., Zou, J.: Nonlinear inexact Uzawa algorithms for linear and nonlinear saddle-point problems. SIAM J. Optimiz. **16**, 798–825 (2006)
13. Hu, Q., Zou, J.: Two new variants of nonlinear inexact Uzawa algorithms for saddle-point problems. Numer. Math. **93**, 333–359 (2002)
14. Kolev, T., Vassilevski, P.: Some experience with a H^1-based auxiliary space AMG for H(curl) problems, Report UCRL-TR-221841, LLNL, Livermore CA (2006)
15. Li, D., Greif, C., Schötzau, D.: Parallel numerical solution of the time-harmonic Maxwell equations in mixed form. Numer. Lin. Algebra Appl. **19**, 525–539 (2012)
16. Monk, P.: Analysis of a finite element method for Maxwell's equations. SIAM J. Numer. Anal. **29**, 714–729 (1992)
17. Nédélec, J.C.: Mixed finite elements in $\mathbb{R}^3$. Numer. Math. **35**, 315–341 (1980)
18. Niceno, B.: EasyMesh. http://web.mit.edu/easymesh_v1.4/www/easymesh.html
19. Perugia, I., Schötzau, D., Monk, P.: Stabilized interior penalty methods for the time-harmonic Maxwell equations. Comput. Methods Appl. Mech. Eng. **191**, 4675–4697 (2002)
20. Perugia, I., Simoncini, V.: Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations. Numer. Lin. Alg. Appl. **7**, 585–616 (2000)
21. Perugia, I., Simoncini, V., Arioli, M.: Linear algebra methods in a mixed approximation of magneto-static problems. SIAM J. Sci. Comput. **21**, 1085–1101 (1999)
22. Pestana, J., Wathen, A.J.: Combination preconditioning of saddle point systems for positive definiteness. Numer. Lin. Algebra Appl. **20**, 785–808 (2013)
23. Wu, S.L., Huang, T.Z., Li, C.X.: Modified block preconditioners for the discretized time-harmonic Maxwell equations in mixed form. J. Comput. Appl. Math. **237**, 419–431 (2013)
24. Zeng, Y., Li, C.: New preconditioners with two variable relaxation parameters for the discretized time-harmonic Maxwell wquations in mixed form. Math. Comput. Probl. Eng. **2012**, 1–13 (2012)