Optimal preconditioners on Solving the Poisson equation with Neumann boundary conditions

Byungjoon Lee and Chohong Min

February 10, 2021

Abstract

MILU preconditioner is well known [16, 3] to be the optimal choice among all the ILU-type preconditioners in solving the Poisson equation with Dirichlet boundary conditions. However, it is less known which is an optimal preconditioner in solving the Poisson equation with Neumann boundary conditions. The condition number of an unpreconditioned matrix is as large as $O\left(h^{-2}\right)$, where h is the step size of grid. Only the optimal preconditioner results in condition number $O\left(h^{-1}\right)$, while the others such as Jacobi and ILU result in $O\left(h^{-2}\right)$.

We review Relaxed ILU and Perturbed MILU preconditioners in the case of Neumann boundary conditions, and present empirical results which indicate that the former is optimal in two dimensions and the latter is optimal in two and three dimensions. To the best of our knowledge, these empirical results have not been rigorously verified yet. We present a formal proof for the optimality of Relaxed ILU in rectangular domains, and discuss its possible extension to general smooth domains and Perturbed MILU.

1 Introduction

The Poisson equation has been intensively studied in sciences and engineering, and applied to a variety of areas such as fluid mechanics, electrical engineering, and image processing. A common problem faced in the applications is to solve a linear system with a large condition number, which delays the convergence of its iterative approximations and generates round-off errors that may result in the loss of many significant digits [8]. The large condition number is innate in the Poisson equation, because the Laplace operator is unbounded. The eigenvalues of the Laplace operator are real and countable, and range from a nonpositive value to the negative infinity [4]. On a certain numerical method for solving the Poisson equation, its associated matrix, approximating the Laplace operator, would have eigenvalues ranging from a nonpositive value to a negative number with very large absolute value. As a result, the condition number of the matrix is large and grows larger as the grid of computation is refined.

For such an ill-conditioned linear system $Ax = b$, preconditioning is required to mitigate the large condition number. A matrix M is called a preconditioner of A, if it is easy to invert M and $M \approx A$. With a preconditioner M of A present, it is advantageous to solve $M^{-1}Ax = M^{-1}b$ instead of $Ax = b$. The Laplace operator is self-adjoint and elliptic [4]. Being its discrete analogue, the matrix A is symmetric and an M-matrix [24]. For such types of matrices, Jacobi, Symmetric Gauss-Seidel (SGS), Incomplete LU (ILU), and Modified ILU (MILU) are well known to be effective [24].

In his seminal paper [16], Gustafsson analyzed aforementioned preconditioners on the standard five-point scheme in rectangular domains with Dirichlet boundary conditions. The matrix A associated with the scheme has a condition number of size $O\left(h^{-2}\right)$ on a grid of step size h. His analysis showed that all the preconditioners except MILU have a condition number of size $O\left(h^{-2}\right)$, and the necessary condition among ILU types to have $O\left(h^{-1}\right)$ is only MILU. He failed to show that MILU is indeed a sufficient condition for $O\left(h^{-1}\right)$. The original conjecture was later proved [3]. Furthermore, the conjecture turns out to hold true even in general smooth domains with Dirichlet boundary conditions. Min et al. [29] proved the conjecture in the case of Shortley-Weller's method [25], and Yoon et al. [28] proved the conjecture in the case of Gibou et al's method [14].

In the case of Neumann boundary conditions, the optimal choice of preconditioner becomes unclear. The solution of the Poisson equation with Neumann boundary conditions is not unique since the addition of any constant to the solution makes another solution. Usually, the sum-zero condition is imposed to make the solution unique, the boundary value problem becomes well posed, and the condition number is measured excluding the zero eigenvalue. While Jacobi, SGS, and ILU preconditioners are still effective for Neumann conditions, MILU fails to be defined in general. MILU takes the form $(L+E)\,E^{-1}\,(E+U)$ where E is a diagonal matrix, and L and U denote the strictly lower and upper triangular parts of A, respectively. E becomes zero at every right-top corner [21], when grid nodes are ordered either in lexicographical or Cuthill-Mckee orderings [13]. When the grid of computation has only one right-top corner, the associated matrix A and the diagonal matrix E both have nullity one, and MILU can be somehow defined [3, 17, 20]. However, when there are more than one right-top corners, there is no known possible way to define MILU, as far as we know.

There are two conventional preconditioners that can resolve the limitation of MILU when it comes to Neumann boundary conditions: one is the relaxed ILU (RILU) [1, 2] and the other is the perturbed MILU (PMILU) [3, 16]. The RILU preconditioner takes an interpolated form of ILU and MILU with a small mixing ratio of ILU. Though small, the presence of ILU makes the RILU nonsingular and well-defined. In earlier works [5, 21], the similar strategies, called MILU-ILU mixture, were introduced but, from our survey, they turn out to be identical to the RILU. As MILU does in the case of the Dirichlet boundary conditions, the authors in [21] empirically found out that the optimal size $O\left(h^{-1}\right)$ is attained by RILU of $1 : O\left(h^2\right)$ in 2D, which is called RILU conjecture.

While the RILU manipulates the existing preconditioners in its construction, the PMILU chooses to perturb the original matrix A to resolve the singularity. By noting that the vector 1 is in the null space of A, the PMILU preconditioner is defined by the MILU preconditioner of the diagonally perturbed matrix $A+\epsilon D$ for sufficiently small $\epsilon > 0$. Although many researchers [17, 3] analyzed the performance of the PMILU for the Dirichlet boundary conditions, concrete analysis for Neumann boundary conditions is unsettled, to the best to our survey.

The aim of this paper is to remind this community of the optimality in RILU and PMILU preconditioners for the Neumann boundary conditions, and to provide a mathematical analysis on the optimality. In section 2, we review Gustafsson's conjecture on Dirichlet boundary conditions and the MILU preconditioner. The RILU and PMILU are reviewed in section 3 and 4, respectively, with conjectures on the optimality supported by empirical results. In section 5, the RILU conjecture in the case of rectangular domain is proved, equipped with a novel rectangular inequality and a non-local decomposition of Heaviside weights. Numerical results are presented in section 6 to validate the optimality conjectured in this work and conclusions and future work are presented in section 7.

2 Preliminaries

In this section, we present brief reviews of the MILU preconditioner for the Neumann boundary conditions. Let Ω be a general smooth domain in $\mathbb{R}^2$ with the boundary $\partial\Omega$. Then, the Poisson equation with Neumann boundary condition reads as follow:

$$\begin{cases} -\triangle u = f\,, & \text{in } \Omega \\ \frac{\partial u}{\partial n} = g\,, & \text{on } \partial\Omega \end{cases}. \tag{1}$$

2.1 Purvis-Burkhalter Discretization

For solving the Poisson equation with Neumann boundary conditions, there are numerous efficient methodologies such as finite volume methods, finite element methods [11, 12] and spectral methods [15, 22]. Each methodology retains its own advantages and disadvantages.

The finite volume method [23, 19] by Purvis and Burkhalter is a simple five-point scheme defined on a uniform grid. In spite of its simplicity, the method can deal with general smooth domains and attains the second-order accuracy, which is guaranteed by a mathematical analysis [30]. The method lacks of the adaptivity to resolve capturing more important details. Since the elliptic nature of the Poisson equation enforces strong coupling between any two locations in the domain, an adaptive grid may not be effective as expected and the disadvantage lacking the adaptivity is alleviated. From these reasons, we select the Purvis-Burkhalter method among the many numerical methods and scrutinize it throughout this paper.

Consider a uniform grid $h\mathbb{Z}^2$ with step size h. For each grid point $(x_i, y_j) \in h\mathbb{Z}^2$, we define the rectangular control volume C_{ij} centered at the grid point with four edges $e_{i\pm\frac{1}{2},j}$, $e_{i,j\pm\frac{1}{2}}$ as

$$\begin{array}{ccccc} C_{ij} & = & \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right] & \times & \left[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}\right] \\ e_{i\pm\frac{1}{2},j} & = & x_{i\pm\frac{1}{2}} & \times & \left[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}\right] \\ e_{i,j\pm\frac{1}{2}} & = & \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right] & \times & y_{j\pm\frac{1}{2}} \end{array}.$$

Let Ω^h be the set of grid points whose control volume intersect with the domain Ω. In order to measure the amount of occupied volume to the domain, we utilize the Heaviside function defined as follows.

Definition 1. (Heaviside function) For each $e_{i+\frac{1}{2},j}$, $e_{i,j+\frac{1}{2}}$, the Heaviside function $H_{i+\frac{1}{2},j}$ and $H_{i-\frac{1}{2},j}$ are defined by

$$H_{i+\frac{1}{2},j} = \frac{\text{length}\left(e_{i+\frac{1}{2},j} \cap \Omega\right)}{\text{length}\left(e_{i+\frac{1}{2},j}\right)},\ H_{i,j+\frac{1}{2}} = \frac{\text{length}\left(e_{i,j+\frac{1}{2}} \cap \Omega\right)}{\text{length}\left(e_{i,j+\frac{1}{2}}\right)},$$

respectively.

The Purvis-Burkhalter discretization takes a finite volume approach to approximate the Poisson equation (1). More precisely, integrating both sides of equation (1) over each control volume and applying the divergence theorem, we can have

$$\int\int_{C_{ij}\cap\Omega} (-\triangle u)\, dA = \int_{\partial(C_{ij}\cap\Omega)} \frac{\partial u}{\partial n} d\Gamma = \int\int_{C_{ij}\cap\Omega} f dA,$$

i.e.,

$$-\int_{\partial C_{ij}\cap\Omega} \frac{\partial u}{\partial n} d\Gamma = \int_{C_{ij}\cap\partial\Omega} g d\Gamma - \int\int_{C_{ij}\cap\Omega} f dA \tag{2}$$

where dA and $d\Gamma$ denote to the area and length elements, respectively.

With the aid of Heaviside function, the difference equation $Au = b$ associated with (2) can be attained by introducing a linear operator $A : \Omega^h \to \mathbb{R}$ defined as

$$\begin{aligned}(Au)_{(i,j)} = & \left[u_{(i,j)} - u_{(i+1,j)}\right] H_{i+\frac{1}{2},j} + \left[u_{(i,j)} - u_{(i,j+1)}\right] H_{i,j+\frac{1}{2}} \\ & + \left[u_{(i,j)} - u_{(i-1,j)}\right] H_{i-\frac{1}{2},j} + \left[u_{(i,j)} - u_{(i,j-1)}\right] H_{i,j-\frac{1}{2}}.\end{aligned} \tag{3}$$

Note that the operator A is a semi-positive definite operator from the definition in (3).

2.2 MILU preconditioners

In this subsection, we review the MILU preconditioners for solving (1). Let A be the matrix as in (3), and L, D, and U are strictly lower triangular, diagonal, and strictly upper triangular matrices of A, respectively. Then, we consider the preconditioner M of the form

$$M = (L + E)\, E^{-1}\, (E + U) = L + E + U + LE^{-1}U$$

where E is a diagonal matrix. Note that M can be treated as an approximation of A, that is, $M = A + R$ where $R = E + LE^{-1}U - D$.

The ILU preconditioner M is obtained by enforcing diagonal elements of R to be zero. Hence, the diagonal matrix E is defined by

$$\left(E + LE^{-1}U\right)_{i,i} = D_{i,i}, \ \ i = 0, 1, \ldots, K - 1. \tag{4}$$

Let $E_{i,j}$ be the diagonal element of E corresponding to the grid point (x_i, y_j). Then, each $E_{i,j}$ can be computed by the following recursive relation, obtained from (4):

Definition 2 (ILU)**.** We define ILU to be the matrix $M = (L + E)\, E^{-1}\, (E + U)$ where E is defined by the formula

$$
\begin{aligned}
E_{0,0} &= a_{0,0} \\
E_{i,j} &= a_{i,j} - \frac{H^2_{i-\frac{1}{2},j}}{E_{i-1,j}} - \frac{H^2_{i,j-\frac{1}{2}}}{E_{i,j-1}},
\end{aligned}
$$

Here $E_{i,j}$ is recursively updated in the lexicographical order [13].

In [16], Gustafsson found that the ILU preconditioner could reduce computational cost but keeps the order of condition number, it remains to be of order $O\left(h^{-2}\right)$ when the matrix A is associated with the standard five-point scheme with Dirichlet boundary conditions. As a matter of fact, for the ILU preconditioner M, one can easily choose $u \in C_0^1(\Omega)$ for which $\langle Ru, u\rangle = O\left(h^{-2}\right)$, which results in $\langle Au, u\rangle / \langle Mu, u\rangle = O\left(h^2\right)$ with the fact that $\langle Au, u\rangle = O(1)$ for any $u \in C_0^1(\Omega)$. Then, $\kappa\left(M^{-1}A\right)$ is of order $O\left(h^{-2}\right)$ since $\langle Ae_1, e_1\rangle / \langle Me_1, e_1\rangle = O(1)$ where $e_1 = (1, 0, \ldots, 0)$.

Based on this observation, he suggested the MILU preconditioner M taking E of the form

$$
\sum_{j=1}^{K} \left(E + LE^{-1}U\right)_{ij} = D_{ii} \; i = 1, \ldots, K. \tag{5}
$$

Note that E in (5) is defined so as to satisfy a necessary condition

$$
-O(1) \leq \langle Ru, u\rangle \leq O\left(h^{-1}\right), \; \forall u \in C_0^1(\Omega)
$$

for $\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right)$. The recursive formula for $E_{i,j}$ corresponding to the grid point (x_i, y_j) reads as follow:

Definition 3 (MILU)**.** We define MILU to be the matrix $M = (L+E)\,E^{-1}(E+U)$ where E is defined by the formula

$$
\begin{aligned}
E_{0,0} &= a_{0,0} \\
E_{i,j} &= a_{i,j} - \frac{H_{i-\frac{1}{2},j}}{E_{i-1,j}} \left(H_{i-\frac{1}{2},j} + H_{i-1,j+\frac{1}{2}}\right) - \frac{H_{i,j-\frac{1}{2}}}{E_{i,j-1}} \left(H_{i,j-\frac{1}{2}} + H_{i+\frac{1}{2},j-1}\right),
\end{aligned}
$$

Here $E_{i,j}$ is recursively updated in the lexicographical order [13].

While it is proven that the MILU preconditioner is the optimal choice among ILU types for Dirichlet boundary conditions, the problem of selecting an optimal preconditioner for Neumann boundary conditions remains unsettled. A strong candidate for the optimal choice might be the MILU, but it fails to be defined for Neumann boundary conditions since it produces zero entries in E at every right-top corner of the domain [21].

3 Relaxed ILU

The relaxed ILU (RILU) was first introduced by [1, 2] to apply an MILU-type preconditioner to singular systems. The main idea of RILU is to mix the ILU and MILU with a specific ratio r so that it avoids the singularity that arises from the Neumann conditions as well as it takes advantage of the optimality inherent in MILU. The RILU preconditioner M can be computed by the following formula for $E_{i,j}$ corresponding to the grid point (x_i, y_j) of the diagonal matrix E.

Definition 4 (RILU)**.** For $r \in (0,1)$, we define RILU(r) to be the matrix $M = (L+E)\,E^{-1}(E+U)$ where E is defined by the formula

$$
\begin{aligned}
E_{0,0} &= a_{0,0} \\
E_{i,j} &= a_{i,j} - \frac{H_{i-\frac{1}{2},j}}{E_{i-1,j}} \left(H_{i-\frac{1}{2},j} + (1-r)\,H_{i-1,j+\frac{1}{2}}\right) - \frac{H_{i,j-\frac{1}{2}}}{E_{i,j-1}} \left(H_{i,j-\frac{1}{2}} + (1-r)\,H_{i+\frac{1}{2},j-1}\right),
\end{aligned}
$$

Here $E_{i,j}$ is recursively updated in the lexicographical order [13].

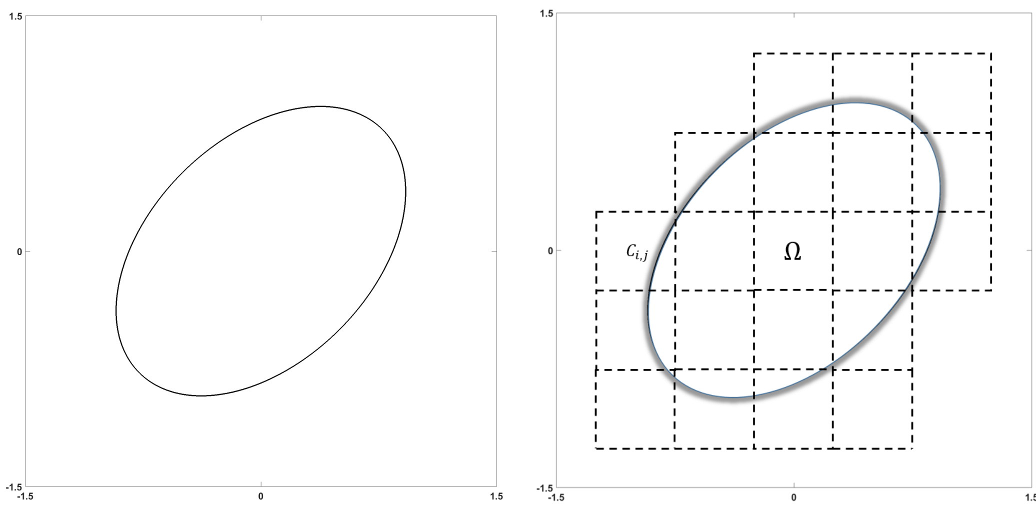


Figure 1: The domain Ω for the numerical test in section 6.1 (left) and its corresponding grid configuration Ω^h (right). Each cell C_{ij} is surrounded by dotted lines.

There have been several researches on choosing the optimal ratio r. Van der Vorst [27] reported empirically that $r = C \cdot h^2$ for some constant $C > 0$ provides the best performance and Chan [6] showed that this ratio achieves the optimal order $O\left(h^{-1}\right)$ in condition number for the Poisson equation with periodic boundary conditions. For all practical purposes, Van der Vorst [26] and Bridson [5] proposed to select r to be 0.05 and 0.03, respectively. In the recent research by Park et al. [21] , the authors verified that the optimality in condition number of the matrix associated with the Neumann problem (1) is obtained by RILU(r) when the ratio is $r = C \cdot h^2$ for some constant C. Based on this observation, they posed the following conjecture, so called the RILU conjecture.

Conjecture. *Let A be the matrix associated with the Purvis-Burkhalter discretization [23, 19] for solving the Poisson equation with Neumann boundary conditions in a smooth domain Ω, and let M be the RILU preconditioner with mixing ratio r. When $r = C \cdot h^2$ for some moderate constant $C > 0$ independent of h, then we have*

$$\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right),$$

for any smooth domain $\Omega \subset \mathbb{R}^2$. In practice, we may set $C = 1$.

This conjecture was supported by numerous empirical tests in [21]. For example, see section 6.1 and section 6.2.

4 Perturbed MILU

Another well-known preconditioner for solving (1) is so called perturbed MILU (PMILU) preconditioner proposed by [3, 16]. The PMILU preconditioner M is simply obtained by adding small perturbation of diagonal matrix D to MILU preconditioner. The recursive formula for each diagonal element $E_{i,j}$ of E in PMILU is defined as follow.

Definition 5 (PMILU)**.** For sufficient small $\epsilon > 0$, we define PMILU(ϵ) to be the matrix $M = (L + E)\, E^{-1}\, (E + U)$ where E is defined by the formula

$$\begin{aligned}
E_{0,0} &= a_{0,0} \\
E_{i,j} &= (1+\epsilon)\, a_{i,j} - \frac{H_{i-\frac{1}{2},j}}{E_{i-1,j}} \left(H_{i-\frac{1}{2},j} + H_{i-1,j+\frac{1}{2}}\right) - \frac{H_{i,j-\frac{1}{2}}}{E_{i,j-1}} \left(H_{i,j-\frac{1}{2}} + H_{i+\frac{1}{2},j-1}\right),
\end{aligned}$$

Here $E_{i,j}$ is recursively updated in the lexicographical order [13].

To the best of our knowledge, the estimation of condition number of PMILU preconditioned matrix for the Neumann boundary conditions is not fully understood. Beauwens [3] analyzed the relation between upper bounds of the condition number of the preconditioned matrix and the selection of perturbation parameter ϵ for nonsingular Stieltjes matrices. As an extension to this work, Notay [20] proposed several strategies to select the parameter ϵ for PMILU so that it achieves the optimal order of $O\left(h^{-1}\right)$ with numerical evidences.

An interesting observation in PMILU preconditioner is that it attains the optimal condition number of order $O\left(h^{-1}\right)$ for any smooth domain in both two and three dimension with a certain choice of ϵ. From numerous empirical tests, we found out that $\epsilon = Ch^2$ is an optimal choice for the parameter ϵ as in RILU conjecture. For example, see the examples in section 6.1 and section 6.2. From this observation, we finalize this section by posing the following conjecture.

Conjecture. *Let A be the matrix associated with the Purvis-Burkhalter discretization [23, 19] for solving the Poisson equation with Neumann boundary conditions in a smooth domain Ω, and let M be the PMILU preconditioner with perturbation parameter ϵ. When $\epsilon = C \cdot h^2$ for some moderate constant $C > 0$ independent of h, then we have*

$$\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right),$$

for any smooth domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. In practice, we may set $C = 1$.

5 Mathematical proof for RILU conjecture

Since the seminal work of Gustafsson [16], MILU preconditioner has been renowned for its optimality among all ILU-type preconditioners in solving the Poisson equation with Dirichlet boundary conditions. In this article, we empirically showed that RILU and PMILU achieve the optimality in solving the Poisson equation with Neumann boundary conditions. We are not the first to report empirical results stating the optimality: e.g. Chan et al. [7] and references therein. However, to the best of our knowledge, there is no known rigorous analysis to support the empirically observed optimality in the case of Neumann boundary conditions. We have endeavored to provide a rigorous analysis to explain the results of RILU and PMILU in section 3 and section 4, and managed to prove the result of RILU in the case of rectangular domains, which we report here.

5.1 Conventional analysis in the case of Dirichlet boundary conditions

Gustafsson in [16] showed that an ILU-type preconditioner is necessarily the MILU preconditioner when it achieves the optimal condition number $O\left(h^{-1}\right)$ in numerically solving the Poisson equation. Precisely speaking, it is necessarily more or less the MILU preconditioner with a margin of error of $O\left(h^2\right)$ allowed. His argument is succinct and easy to understand, but its reverse is not a simple matter at all. He could not prove the reverse argument that the MILU preconditioner, the unperturbed one, results in $O\left(h^{-1}\right)$ condition number. The reverse argument was posted as a conjecture, and was proved in [3, 28, 30] in the case of Dirichlet boundary conditions.

In this section, we review some conventional analysis used in the proofs and point out the obstacles that the analysis faces in the case of Neumann boundary conditions. The first step of the analysis is to convert the estimation of eigenvalues into the estimation of Rayleigh quotients.

$$\lambda_{\min/\max}\left(M^{-1}A\right) = \min/\max_{u\neq 0} \frac{\langle Au, u\rangle}{\langle Mu, u\rangle} = \min/\max_{u\neq 0} \frac{1}{1 + \langle Ru, u\rangle / \langle Au, u\rangle}$$

Here, R denotes the residue of the preconditioning, $R := M - A$. When A is the associated matrix of the standard five-point scheme, the residue value $\langle Ru, u\rangle$ is given as

$$-\langle Ru, u\rangle = \sum_{i,j} \frac{H_{i+\frac{1}{2},j} H_{i,j+\frac{1}{2}}}{E_{i,j}} \left(u_{i+1,j} - u_{i,j+1}\right)^2,$$

where $H_{i+\frac{1}{2},j}$ and $H_{i,j+\frac{1}{2}}$ denote the matrix elements $A_{(i,j),(i+1,j)}$ and $A_{(i,j),(i,j+1)}$, respectively. Utilizing the following **triangular inequality** [16] that holds for any w_1, $w_2 > 0$ and any $u \in \mathbb{R}^{2\times 2}$

$$\frac{w_1 w_2}{w_1 + w_2} \left(u_{1,0} - u_{0,1}\right)^2 \leq w_1 \left(u_{1,0} - u_{0,0}\right)^2 + w_2 \left(u_{0,1} - u_{0,0}\right)^2, \tag{6}$$

the second step of the conventional analysis obtains the following estimate of Rayleigh quotients.

$$
\begin{aligned}
0 \le -\langle Ru,u\rangle &\le \sum_{i,j} \frac{H_{i+\frac{1}{2},j}+H_{i,j+\frac{1}{2}}}{E_{i,j}} \left[H_{i+\frac{1}{2},j}\left(u_{i+1,j}-u_{i,j}\right)^2 + H_{i,j+\frac{1}{2}}\left(u_{i,j+1}-u_{i,j}\right)^2 \right] \\
&\le \left(\max_{i,j} \frac{H_{i+\frac{1}{2},j}+H_{i,j+\frac{1}{2}}}{E_{i,j}} \right) \sum_{i,j} \left[H_{i+\frac{1}{2},j}\left(u_{i+1,j}-u_{i,j}\right)^2 + H_{i,j+\frac{1}{2}}\left(u_{i,j+1}-u_{i,j}\right)^2 \right] \\
&\le \left(\max_{i,j} \frac{H_{i+\frac{1}{2},j}+H_{i,j+\frac{1}{2}}}{E_{i,j}} \right) \langle Au,u\rangle .
\end{aligned}
$$

The inequality (6) bounds the square of diagonal difference by a weighted sum of the squares on the two sides of a triangle. With this regard, we refer to the inequality as a **triangular inequality**.

In the case of Dirichlet boundary conditions, the maximum is bounded above by a fraction $\frac{c}{c+h}$ for some constant $c > 0$ that is independent of step size h. According to the bound, Rayleigh quotients are estimated as

$$
1 \le \frac{1}{1+\frac{\langle Ru,u\rangle}{\langle Au,u\rangle}} \le \frac{1}{1-\left(\max_{i,j}\frac{H_{i+\frac{1}{2},j}+H_{i,j+\frac{1}{2}}}{E_{i,j}}\right)} \le \frac{c+h}{h}, \tag{7}
$$

and the condition number is estimated as the desired optimal order.

$$
\kappa\left(M^{-1}A\right) = \frac{\lambda_{max}\left(M^{-1}A\right)}{\lambda_{min}\left(M^{-1}A\right)} \le \frac{c+h}{h} = O\left(h^{-1}\right)
$$

5.2 Conventional analysis in the case of Neumann boundary conditions

The conventional analysis is very effective in the case of Dirichlet boundary conditions and is valid for arbitrary smooth domains. However the analysis is not that much effective in the case of Neumann boundary conditions. For simplicity, let us consider a rectangular domain with Neumann boundary condition imposed on its every side, and assume a grid $\{(x_i,y_j)\,|\,i=0,\cdots,i_{\max}-1 \text{ and } j=0,\cdots,j_{\max}-1\}$ of uniform step size h placed on the domain. The matrix A associated with the standard five-point scheme has entries as

$$
H_{i+\frac{1}{2},j} := A_{(i,j),(i+1,j)} = \begin{cases} 0 & \text{if } i=-1 \text{ or } i_{\max}-1 \\ 1 & \text{else} \end{cases}, \text{ and}
$$

$$
H_{i,j+\frac{1}{2}} := A_{(i,j),(i,j+1)} = \begin{cases} 0 & \text{if } j=-1 \text{ or } j_{\max}-1 \\ 1 & \text{else} \end{cases}.
$$

It can be shown [21] that the diagonal entries of the MILU preconditioner applied to A is simply

$$
E^{MILU}_{i,j} = H_{i+\frac{1}{2},j} + H_{i,j+\frac{1}{2}}.
$$

Note that $E^{MILU}_{i,j} = 0$ when $i = i_{\max}-1$ and $j = j_{\max}-1$. Thus E^{MILU} is not invertible, and MILU preconditioning is not applicable in the case of Neumann boundary conditions. On the other hand, RILU preconditioner has positive diagonal entries $E^{RILU}_{i,j} > 0$ and its inverse is well defined. For RILU preconditioner M^{RILU}, the Rayleigh quotients are bounded as

$$
\frac{1}{1+\frac{\langle Ru,u\rangle}{\langle Au,u\rangle}} \le \frac{1}{1-\left(\max_{i,j}\frac{H_{i+\frac{1}{2},j}+H_{i,j+\frac{1}{2}}}{E_{i,j}}\right)} \le \frac{1}{1-1} = +\infty,
$$

which leads to a meaningless estimation $\kappa\left(\left(M^{RILU}\right)^{-1}A\right) \le \infty$. However, the empirical results, see section 3, indicate $\kappa\left(\left(M^{RILU}\right)^{-1}A\right) \le O\left(h^{-1}\right)$. With this regard, the conventional analysis fails to properly estimate the condition number of RILU. In order to overcome the failure of the conventional analysis, we introduce a novel estimate in the next section. Based on this experience, we doubt that the conventional analysis would succeed in proving the optimality of either PMILU or RILU in irregular domains.

We scrutinized each part of the conventional analysis to identify the part from which the poor estimation result, and noticed that the triangle inequality was applied to each grid cell **locally**. We concluded that any local application of such triangle inequalities cannot come up with the optimal order $O\left(h^{-1}\right)$, and devised up a novel rectangular-type inequality and a global planning in applying the inequality to grid cells, which we report in the next section.

5.3 Novel analysis in the case of rectangular domains

In the previous section, we pointed out the drawbacks of the conventional analysis in dealing with Neumann boundary conditions. The main feature of the conventional approach is to apply the triangle inequality locally and independently to each grid cell. Contrary to the conventional analysis, we apply instead the following rectangular inequality to each grid cell, and its application is not local, but will be maneuvered globally.

Lemma 1. *(Rectangular inequality) For any $u \in \mathbb{R}^{2\times 2}$ and any $w_1, w_2, r \geq 0$, the following inequality holds.*

$$\frac{w_1 + w_2}{2 - r}\left[\left(u_{1,0} - u_{0,1}\right)^2 - r\left(u_{1,0}^2 + u_{0,1}^2\right)\right] \leq w_1\left[\left(u_{1,0} - u_{0,0}\right)^2 + \left(u_{0,1} - u_{0,0}\right)^2\right] + w_2\left[\left(u_{1,0} - u_{1,1}\right)^2 + \left(u_{0,1} - u_{1,1}\right)^2\right]$$

Proof. From $\left(u_{1,0} - u_{0,1}\right)^2 \leq 2\left(u_{1,0}^2 + u_{0,1}^2\right)$, we can have

$$\begin{aligned}\frac{w_1 + w_2}{2 - r}\left[\left(u_{1,0} - u_{0,1}\right)^2 - r\left(u_{1,0}^2 + u_{0,1}^2\right)\right] &\leq \frac{w_1 + w_2}{2 - r}\left(u_{1,0} - u_{0,1}\right)^2\left[1 - \frac{r}{2}\right] \\ &= \frac{w_1 + w_2}{2}\left(u_{1,0} - u_{0,1}\right)^2.\end{aligned}$$

Comparing the RHS of the rectangular inequality with $\frac{w_1+w_2}{2}\left(u_{1,0} - u_{0,1}\right)^2$ gives the proof of the lemma.

$$\begin{aligned}w_1\left[\left(u_{1,0} - u_{0,0}\right)^2 + \left(u_{0,1} - u_{0,0}\right)^2\right] + w_2\left[\left(u_{1,0} - u_{1,1}\right)^2 + \left(u_{0,1} - u_{1,1}\right)^2\right] - \frac{w_1 + w_2}{2}\left(u_{1,0} - u_{0,1}\right)^2 \\ = 2w_1\left(u_{0,0} - \frac{u_{0,1} + u_{1,0}}{2}\right)^2 + 2w_2\left(u_{1,1} - \frac{u_{0,1} + u_{1,0}}{2}\right)^2 \geq 0.\end{aligned}$$

$\square$

For the preconditioned matrix by RILU($r = Ch^2$), the residual value is given as

$$-\langle Ru, u\rangle = \sum_{i=0}^{i_{\max}-1}\sum_{j=0}^{j_{\max}-1}\frac{1}{E_{i,j}}\left[\left(u_{i+1,j} - u_{i,j+1}\right)^2 - r\left(u_{i+1,j}^2 + u_{i,j+1}^2\right)\right].$$

The above summation visits every grid cell. Lemma 1 is going to be applied to every grid cell. Since the lemma holds for any weight $w_1, w_2 \geq 0$, we have a lot of freedom in the application of the lemma to the summation. We managed to find some choices of the weights to prove the optimality of RILU(h^2). On a grid cell $[x_i, x_{i+1}]\times[y_j, y_{j+1}]$, we choose the weights $w_{i,j}^1$ and $w_{i,j}^2$ as

$$\begin{aligned}w_{i,j}^1 &= 1 - \frac{2 - r\left(i_{\max} + j_{\max}\right)}{2\left(i_{\max} + j_{\max} - 1\right)}\left(i + j\right), \text{ and} \\ w_{i,j}^2 &= \frac{2 - r\left(i_{\max} + j_{\max}\right)}{2\left(i_{\max} + j_{\max} - 1\right)}\left(i + j + 1\right).\end{aligned} \tag{8}$$

Note that $w_{i,j}^1$, $w_{i,j}^2 \in [0, 1]$, since $i + j \leq (i_{\max} - 1) + (j_{\max} - 1) = i_{\max} + j_{\max} - 2$ and $r\left(i_{\max} + j_{\max}\right) = O\left(h\right)$ is smaller than 2 for sufficiently small h. The following theorem shows how the choices lead to our main result.

Theorem 1. *Let A be the matrix associated with (3) and let R be the residue matrix $R = A - M$ where M is the RILU preconditioner with mixing ratio $r = Ch^2$ for some constant $C > 0$ and sufficiently small h. Then, the following inequality holds for any $u \in \mathbb{R}^{i_{\max} \times j_{\max}}$.*

$$\frac{i_{\max} + j_{\max}}{i_{\max} + j_{\max} - 1} \langle -Ru, u \rangle \leq \langle Au, u \rangle$$

Proof. When the domain is rectangular, the diagonal elements E_{ij} of the preconditioner RILU(r) is given as

$$E_{0,0} = 2$$

$$E_{i,j} = \begin{cases} 3 - \frac{2-r}{E_{i,j-1}}, & \text{if } i = 0 \\ 3 - \frac{2-r}{E_{i-1,j}}, & \text{if } j = 0 \\ 4 - \frac{2-r}{E_{i-1,j}} - \frac{2-r}{E_{i,j-1}}, & \text{else .} \end{cases}$$

A mathematical induction on $n = i + j$ easily leads to $E_{ij} \geq 2$, and we have

$$\begin{aligned} \langle -Ru, u \rangle &= \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} \frac{1}{E_{ij}} \left[(u_{i+1,j} - u_{i,j+1})^2 - r \left(u_{i+1,j}^2 + u_{i,j+1}^2 \right) \right] \\ &\leq \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} \frac{1}{2} \left[(u_{i+1,j} - u_{i,j+1})^2 - r \left(u_{i+1,j}^2 + u_{i,j+1}^2 \right) \right] \end{aligned}$$

We apply lemma 1 to each grid cell $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$ with the choices of weights $w_{i,j}^1$ and $w_{i,j}^2$ in (8). Using the fact that $w_{i,j}^1 + w_{i,j}^2 = \frac{2-r}{2} \frac{i_{\max}+j_{\max}}{i_{\max}+j_{\max}-1}$, we have

$$\begin{aligned} \frac{i_{\max} + j_{\max}}{i_{\max} + j_{\max} - 1} \langle -Ru, u \rangle &\leq \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} \frac{i_{\max} + j_{\max}}{i_{\max} + j_{\max} - 1} \frac{1}{2} \left[(u_{i+1,j} - u_{i,j+1})^2 - r \left(u_{i+1,j}^2 + u_{i,j+1}^2 \right) \right] \\ &= \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} \frac{w_{i,j}^1 + w_{i,j}^2}{2 - r} \left[(u_{i+1,j} - u_{i,j+1})^2 - r \left(u_{i+1,j}^2 + u_{i,j+1}^2 \right) \right] \\ &\leq \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} w_{i,j}^1 \left[(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 \right] \\ &\quad + \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}-1} w_{i,j}^2 \left[(u_{i+1,j} - u_{i+1,j+1})^2 + (u_{i,j+1} - u_{i+1,j+1})^2 \right] \end{aligned}$$

Note that $w_{i,j}^1$ is multiplied to the left and bottom sides of the grid cell $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$, and $w_{i,j}^2$ to the right and top sides. Thus, each side, except on the boundary, is multiplied once by w^1 and once again by w^2. Combining the two multiplications, we have

$$\begin{aligned} \frac{i_{\max} + j_{\max}}{i_{\max} + j_{\max} - 1} \langle -Ru, u \rangle \leq & \sum_{i=0}^{i_{\max}-1} \sum_{j=0}^{j_{\max}} \left(w_{i,j}^1 + w_{i,j-1}^2 \right) (u_{i+1,j} - u_{i,j})^2 \\ & + \sum_{i=0}^{i_{\max}} \sum_{j=0}^{j_{\max}-1} \left(w_{i,j}^1 + w_{i-1,j}^2 \right) (u_{i,j+1} - u_{i,j})^2 . \end{aligned}$$

Since $w_{i,j}^1 + w_{i,j-1}^2 = w_{i,j}^1 + w_{i-1,j}^2 = 1$, we proved that

$$\frac{i_{\max}+j_{\max}}{i_{\max}+j_{\max}-1}\langle -Ru,u\rangle \leq \sum_{i=0}^{i_{\max}-1}\sum_{j=0}^{j_{\max}}(u_{i+1,j}-u_{i,j})^2+\sum_{i=0}^{i_{\max}}\sum_{j=0}^{j_{\max}-1}(u_{i,j+1}-u_{i,j})^2$$
$$\frac{i_{\max}+j_{\max}}{i_{\max}+j_{\max}-1}\langle -Ru,u\rangle \leq \langle Au,u\rangle .$$

□

Corollary 1. *(Upper bound) Let A be the matrix associated with (3) and M be the RILU preconditioner with mixing ratio $r = Ch^2$. For any $u \perp 1$ and sufficiently small h, we have an upper bound of the Rayleigh quotient as*

$$\frac{\langle Au,u\rangle}{\langle Mu,u\rangle} \leq i_{\max}+j_{\max} = O(h^{-1}).$$

Proof. We apply theorem 1 to have

$$\frac{\langle Ru,u\rangle}{\langle Au,u\rangle} \geq -\frac{i_{\max}+j_{\max}-1}{i_{\max}+j_{\max}} \text{ and}$$
$$\frac{\langle Au,u\rangle}{\langle Mu,u\rangle} = \frac{1}{1+\frac{\langle Ru,u\rangle}{\langle Au,u\rangle}} \leq \frac{1}{1-\frac{i_{\max}+j_{\max}-1}{i_{\max}+j_{\max}}} = i_{\max}+j_{\max}.$$

□

Now, we turn our attention to seeking a lower bound of Rayleigh quotients. Since the domain Ω is assumed to be rectangular, one can easily show that the matrix A has eigenvalues

$$\lambda^{m,n} = 4 - 2\cos\left(\frac{m\pi}{i_{\max}}\right) - 2\cos\left(\frac{n\pi}{j_{\max}}\right),$$

for $m = 0, \cdots, i_{\max}-1$ and $n = 0, \cdots, j_{\max}-1$, and their corresponding eigenvectors are

$$u_{i,j}^{m,n} = \cos\left(\frac{m\pi}{i_{\max}}\left(i+\frac{1}{2}\right)\right)\cos\left(\frac{n\pi}{j_{\max}}\left(j+\frac{1}{2}\right)\right),$$

for $i = 0, \cdots, i_{\max}-1$ and $j = 0, \cdots, j_{\max}-1$. Using the above facts, we now introduce a lower bound for Rayleigh quotients $\langle Au,u\rangle / \langle Mu,u\rangle$.

Lemma 2. *Let A be the matrix associated with (3) and let R be the residue matrix $R = A - M$ where M is the MILU-ILU preconditioner with mixing ratio $r > 0$. For any $u \perp 1$,*

$$\langle -Ru,u\rangle \geq -\frac{[\max(i_{\max},j_{\max})]^2}{\pi^2} r \langle Au,u\rangle .$$

Proof.

$$\begin{aligned}\langle -Ru,u\rangle &= \sum_{i=0}^{i_{\max}-1}\sum_{j=0}^{j_{\max}-1}\frac{1}{E_{ij}}\left[(u_{i+1,j}-u_{i,j+1})^2 - r\left(u_{i+1,j}^2+u_{i,j+1}^2\right)\right]\\ &\geq \sum_{i=0}^{i_{\max}-1}\sum_{j=0}^{j_{\max}-1} -r\left(u_{i+1,j}^2+u_{i,j+1}^2\right)\\ &\geq -r\langle u,u\rangle\end{aligned}$$

Note that $\lambda^{0,0} = 0 = \min\lambda(A)$ and $u^{0,0} = 1$. Since $u \perp 1$, and we can apply the Min-Max theorem to obtain

$$\begin{aligned}\frac{\langle Au,u\rangle}{\langle u,u\rangle} &\geq \min_{\lambda(A)\neq 0}\lambda(A)\\ &= \min\left(\lambda^{1,0},\lambda^{0,1}\right)\\ &= 2-2\max\left(\cos\frac{\pi}{i_{\max}},\cos\frac{\pi}{j_{\max}}\right).\end{aligned}$$

From $\cos x \geq 1 - \frac{1}{2}x^2$, we easily have

$$\begin{aligned}
\langle u, u\rangle &\leq \frac{\left[\max\left(i_{\max}, j_{\max}\right)\right]^2}{\pi^2} \langle Au, u\rangle \ \text{ and} \\
\langle -Ru, u\rangle &\geq -r\langle u, u\rangle \\
&\geq -\frac{\left[\max\left(i_{\max}, j_{\max}\right)\right]^2}{\pi^2} r\langle Au, u\rangle .
\end{aligned}$$

$\square$

Corollary 2. *(Lower bound) Let A be the matrix associated with (3) and M is the MILU-ILU preconditioner with mixing ratio r. For any $u \perp 1$,*

$$\frac{\langle Au, u\rangle}{\langle Mu, u\rangle} \geq \frac{\pi^2}{\pi^2 + \left[\max\left(i_{\max}, j_{\max}\right)\right]^2 r}.$$

Proof. From the above theorem, we have

$$\begin{aligned}
\frac{\langle Ru, u\rangle}{\langle Au, u\rangle} &\leq \frac{\left[\max\left(i_{\max}, j_{\max}\right)\right]^2}{\pi^2} r \text{ and} \\
\frac{\langle Au, u\rangle}{\langle Mu, u\rangle} = \frac{1}{1 + \frac{\langle Ru, u\rangle}{\langle Au, u\rangle}} &\geq \frac{1}{1 + \frac{\left[\max\left(i_{\max}, j_{\max}\right)\right]^2}{\pi^2} r} = \frac{\pi^2}{\pi^2 + \left[\max\left(i_{\max}, j_{\max}\right)\right]^2 r}.
\end{aligned}$$

$\square$

Now we are ready to prove a theorem, which is the main argument of this paper.

Theorem 2. *Let A be the matrix associated with (3), and let M be the MILU-ILU preconditioner with mixing ratio r. When $r = Ch^2$ for some constant $C > 0$ independent of h, we have*

$$\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right)$$

for a rectangular domain $\Omega = [a_x, b_x] \times [a_y, b_y]$.

Proof. Let $r = Ch^2$ for some constant $C > 0$. Then, from corollary 1, 2 and the fact that $i_{\max} = (b_x - a_x)/h$, $j_{\max} = (b_y - a_y)/h$, we obtain

$$C_1 \leq \left\langle M^{-1}Au, u\right\rangle = \frac{\langle Au, u\rangle}{\langle Mu, u\rangle} \leq \frac{C_2}{h}$$

for constants C_1, C_2 where $C_1 = \frac{\pi^2}{\pi^2 + C[\max(b_x - a_x, b_y - a_y)]^2}$, $C_2 = (b_x - a_x) + (b_y - a_y)$. Hence, by the Min-Max theorem, we have

$$\kappa\left(M^{-1}A\right) = \frac{\lambda_{\max}\left(M^{-1}A\right)}{\lambda_{\min}\left(M^{-1}A\right)} \leq \left(\frac{C_2}{C_1}\right) h^{-1},$$

which proves the theorem. $\square$

6 Numerical results

In this section, we carry out numerical tests to validate the two conjecture in section 3 and 4, and verify the mathematical analysis in section 5. All the computation were implemented in C++, and conducted on a regular personal computer with a 3.60 GHz CPU and 16.0 GB memory. Every linear system is solved by the Preconditioned Conjugate Gradient (PCG) and the stopping criterion for PCG is chosen to be $\|r_n\| / \|r_0\| < 10^{-10}$.

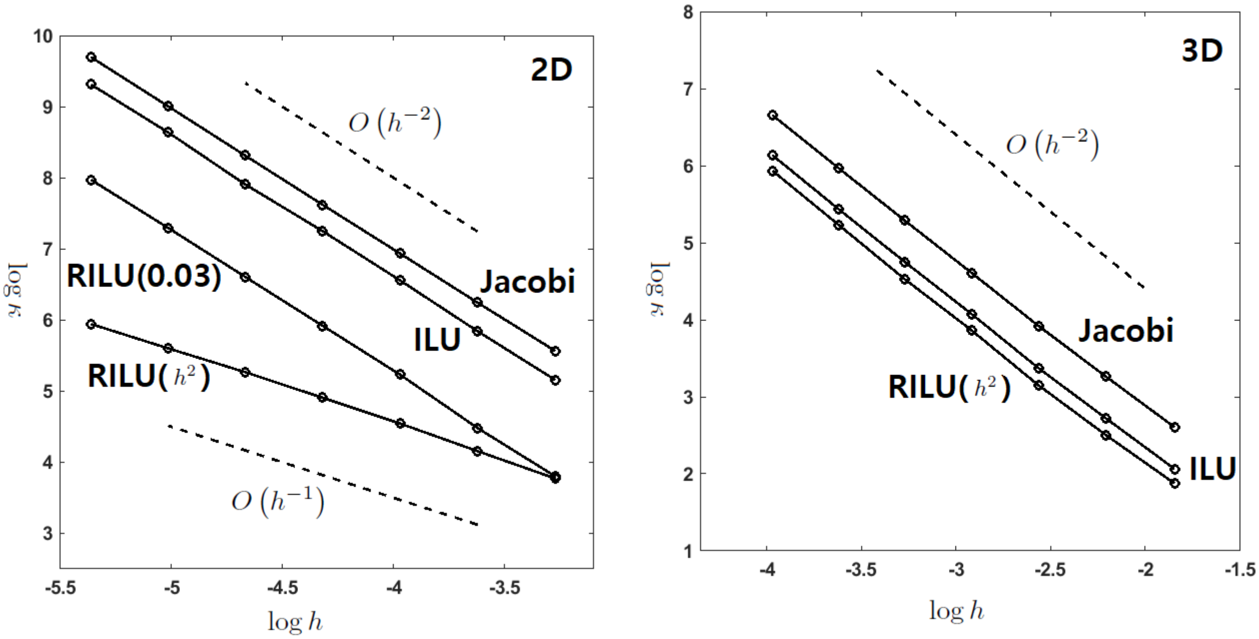


Figure 2: Plots of condition number $\kappa\left(M^{-1}A\right)$ when a preconditioner M is Jacobi, ILU, RILU(0.03), and RILU$\left(h^2\right)$ for the examples in section 6.1 and 6.2. The results show that RILU$\left(h^2\right)$ achieves the optimal condition number of order $O\left(h^{-1}\right)$ for two dimensional case (left) but fails to attain the optimal condition number for three dimensional case (right).

6.1 Elliptical domain

Let us consider a domain Ω of tilted ellipse given by $\Omega = \left\{(x,y) \in \mathbb{R}^2 | 17x^2 - 14xy + 17y^2 \leq 12\right\}$, as in figure 1, and let A be the matrix associated with the Purvis-Burkhalter method (3). The profiles of condition number $\kappa\left(M^{-1}A\right)$ are presented in the left of figure 2 and 4, when a preconditioner M is Jacobi, ILU, RILU(0.03), RILU$\left(h^2\right)$, and PMILU$\left(h^2\right)$. The result shows that RILU$\left(h^2\right)$ and PMILU$\left(h^2\right)$ attain an optimal condition number of order $O\left(h^{-1}\right)$ while the others do not. Figure 3 depicts the performance of PCG on the linear system $Au = b$ for each preconditioner M. In terms of the iteration number to convergence in figure 3, we can see that the iteration numbers of PMILU$\left(h^2\right)$ and RILU$\left(h^2\right)$ are sorely about 14% of that of Jacobi, 34% of that of ILU, and 73% of that of RILU(0.03). This implies that a decrease in condition number $\kappa\left(M^{-1}A\right)$ by one order of magnitude can significantly enhance the performance of linear solver, which clarifies the importance of RILU and PMILU conjectures.

6.2 Three dimensional examples

In this subsection, we test two examples to validate the conjecture of PMILU in section 4. One is an ellipsoidal domain, and the other is a banana-shaped domain from [9]. At first, we consider a tilted ellipsoid domain $\Omega = \left\{(x,y,z) \in \mathbb{R}^3 | 25x^2 - 10xy + 25y^2 + 24z^2 \leq 24\right\}$ and let A be the matrix with the Purvis-Burkahalter discretization. The right plot in figure 2 compares the condition number of RILU$\left(h^2\right)$ with those of Jacobi and ILU, from which we infer that the RILU conjecture does not hold for three dimensional problem with Neumann boundary conditions.

The right plot in figure 4 shows the comparison of the condition number of PMILU$\left(h^2\right)$ with those of Jacobi, ILU, and RILU$\left(h^2\right)$. The result in the plot indicates that PMILU$\left(h^2\right)$ in the only preconditioner that achieves the optimal condition number of order $O\left(h^{-1}\right)$ in three dimension, confirming the conjecture. Furthermore, the plot shows that PMILU$\left(h^2\right)$ has the smallest condition number in magnitude. Therefore, we can expect that PMILU$\left(h^2\right)$ converges faster than the others. This can be verified by the figure 5. The iteration number of PMILU$\left(h^2\right)$ is sorely about 19% of those of RILU$\left(h^2\right)$ and ILU, and 5.8% of that of Jacobi.

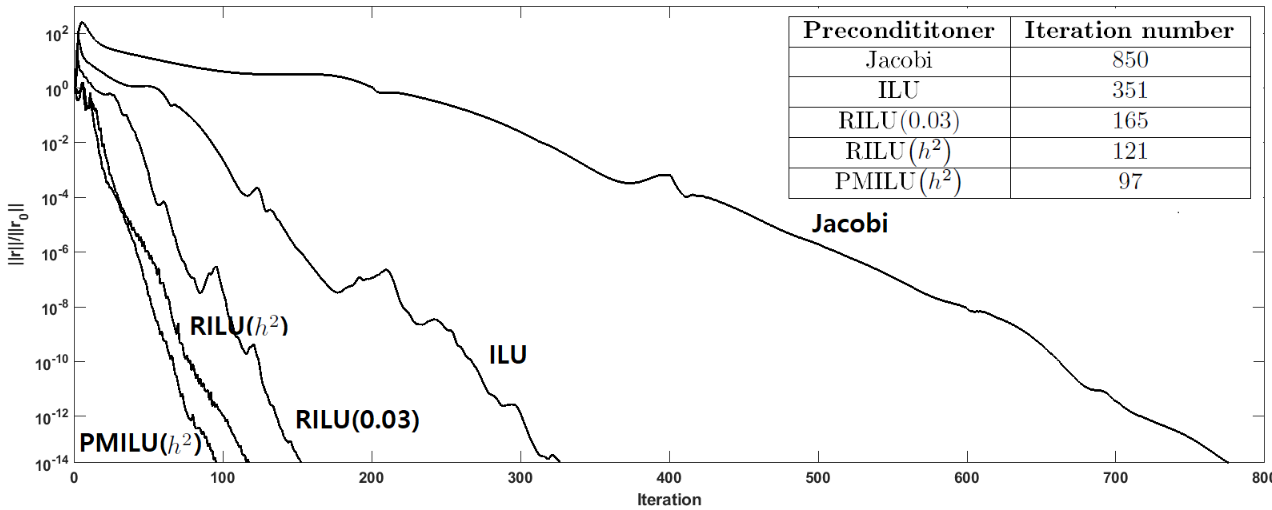


Figure 3: The plot of the norm of residual with respect to the iteration number when the PCG applies to the linear system (3) when $h = 0.005$ for the example in section 6.1.

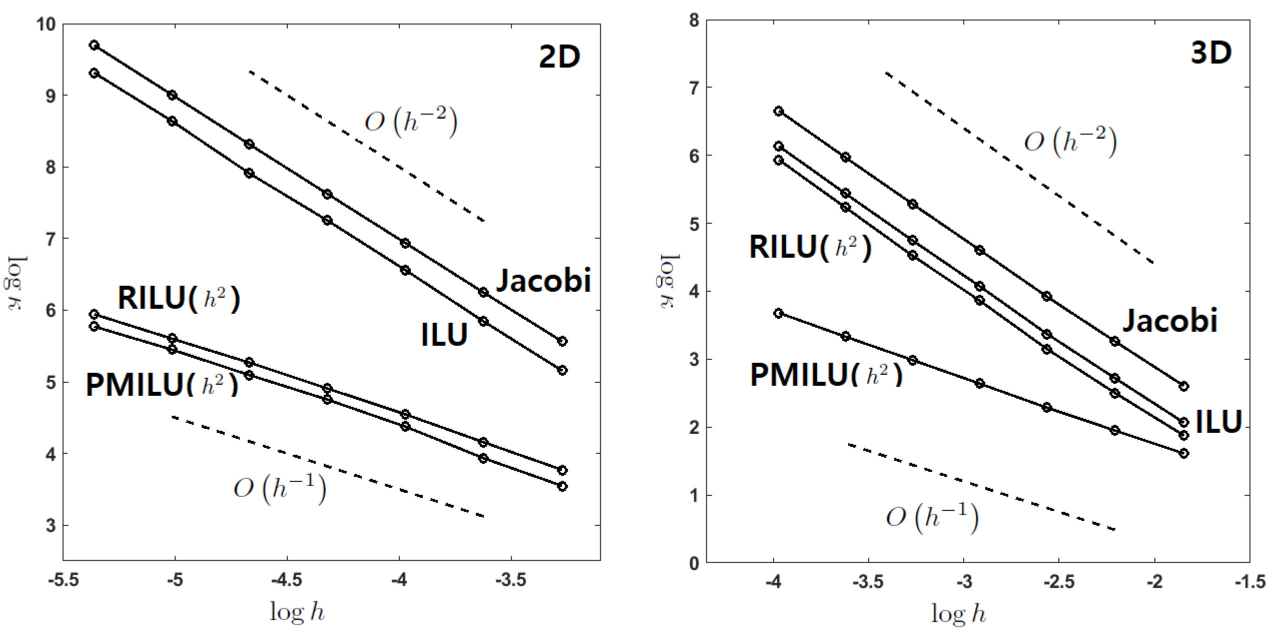


Figure 4: Plots of condition number $\kappa\left(M^{-1}A\right)$ when a preconditioner M is Jacobi, ILU, RILU$\left(h^2\right)$, or PMILU$\left(h^2\right)$ for the examples in section 6.1 and 6.2. The results show that PMILU$\left(h^2\right)$ achieves the optimal condition number of order $O\left(h^{-1}\right)$ for both two and three dimensional cases.

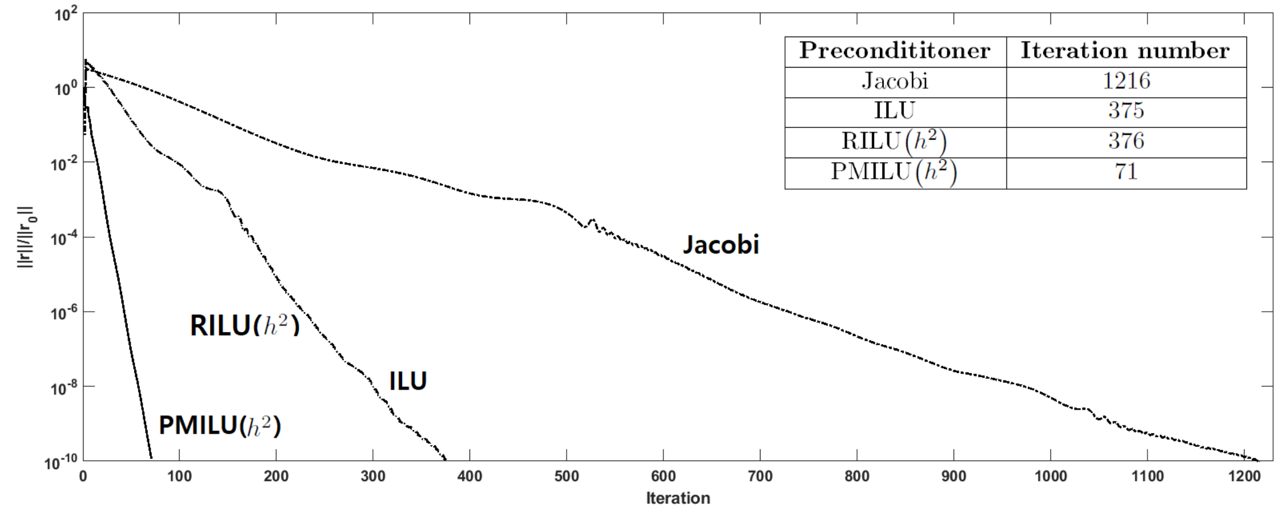


Figure 5: The plot of the norm of residual with respect to the iteration number when the PCG applies to the linear system $Au = b$ when $h = 0.005$ for the example in section 6.2

Secondly, figure 6 shows the comparison of the condition numbers and the norm of residual with respect to the iteration number when the domain Ω is a banana-shaped domain from [9], which contains regions with high curvature. The results are consistent with the fact that PMILU$\left(h^2\right)$ in the optimal preconditioner in the sense that its condition number is of order $O\left(h^{-1}\right)$, still confirming the conjecture. It also implies that the performance of preconditioning is not significantly affected by the shape of domain.

6.3 Validation of the analysis

We presented in section 5 a rigorous analysis for the optimality of RILU(h^2) in rectangular domains. The analysis consists of the upper bound in corollary 1 and the lower bound in corollary 2. On a domain $[0,3]\times[0,2]$, eigenvalues are numerically generated for each step size h, and compared to the estimates of the analysis. Figure 7 confirms the upper-bound estimate $\lambda_{max}\left(M^{-1}A\right) \le i_{max} + j_{max}$, and figure 8 assures the lower-bound estimate $\lambda_{min}\left(M^{-1}A\right) \le \pi^2/\left(\pi^2 + \left[\max\left(i_{\max}, j_{\max}\right)\right]^2 r\right)$ with choice $r = h^2$. Figure 9 shows the profiles of the condition number $\kappa\left(M^{-1}A\right)$ and the number $\left(C_2/C_1\right)h^{-1}$, as in the proof of theorem 2, with respect to the step size h. Combining the two estimates, the condition number $\kappa\left(M^{-1}A\right)$ is bounded above by $\frac{\pi^2+9}{\pi^2}(i_{max} + j_{max}) = O\left(h^{-1}\right)$, which is validated in figure 9. Furthermore, figure 9 implies that the estimation $\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right)$ is tight.

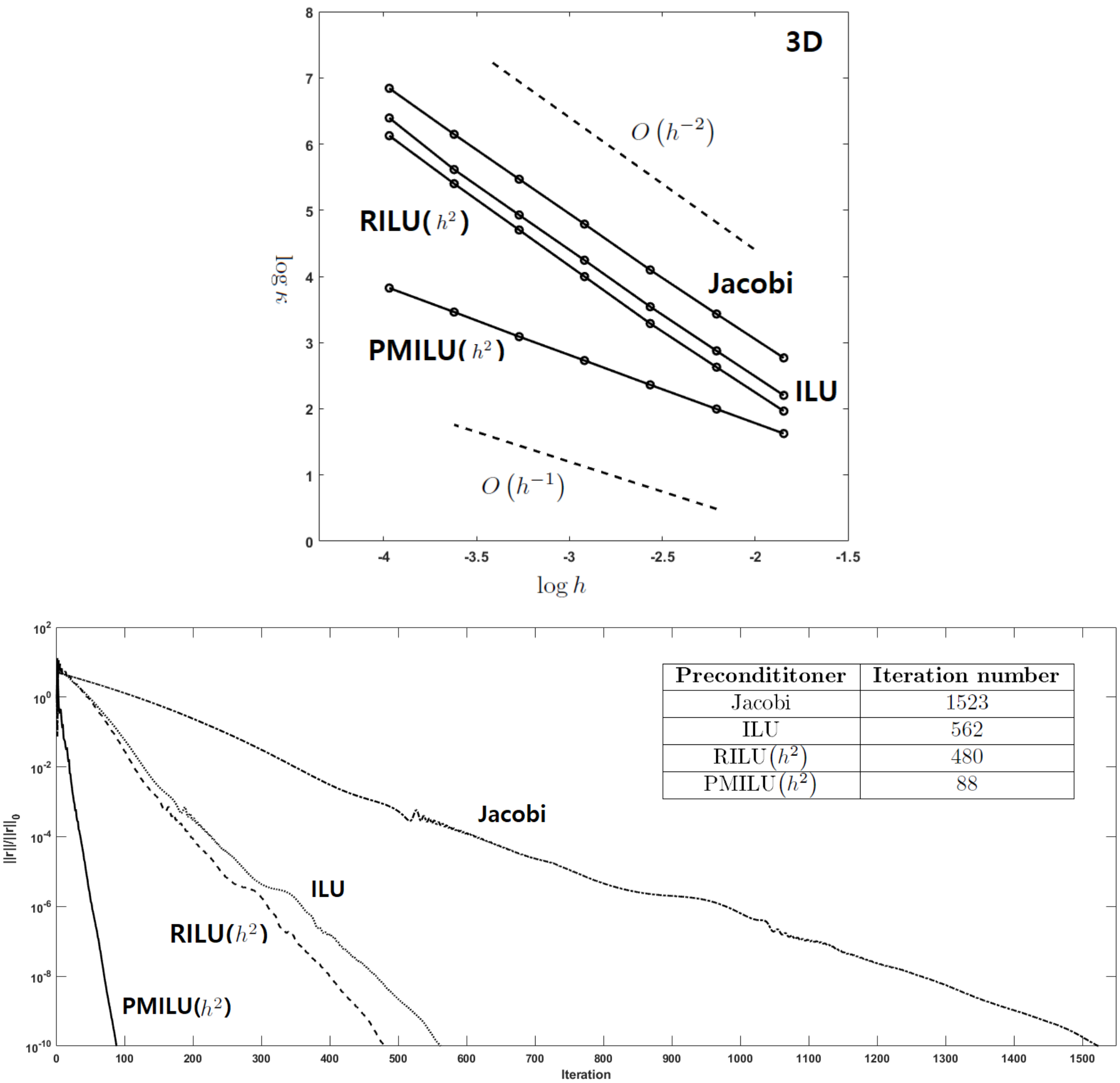
6.4 Application to the incompressible fluid

One of the main applications of the Poisson equation with Neumann boundary conditions is the incompressible fluid, whose motion is governed by the following Navier-Stokes equations

$$\begin{cases} \rho\left(\frac{\partial U}{\partial t} + (U\cdot\nabla U)\,U\right) & = -\nabla p + \nabla\cdot\left(\mu\left(\nabla U + \nabla U^T\right)\right) + \rho F \\ \nabla\cdot U & = 0 \end{cases}. \tag{9}$$

Here, U is the fluid velocity, ρ is the density, μ is the viscosity, p is the pressure, and F is the acceleration by external force. A standard method to solve (9) is the splitting type method, called the projection method [10], based on the Hodge decomposition. The projection method mainly consists of two stages. At the first stage, the method computes an intermediate velocity U^* only with terms with momentum, e.g.

$$\rho\left(\frac{U^* - U^n}{\Delta t}\right) = \mu\nabla\cdot\left(\nabla U^n + (\nabla U^n)\right) + \rho F^n.$$



Precondititoner	Iteration number
Jacobi	1523
ILU	562
RILU(h^2)	480
PMILU(h^2)	88

Figure 6: Plots of condition number $\kappa\left(M^{-1}A\right)$ when a preconditioner M is Jacobi, ILU, RILU$\left(h^2\right)$, or PMILU$\left(h^2\right)$ (top) and the norm of residual with respect to the iteration number when the PCG applies to the linear system $Au = b$ with $h = 0.005$ (bottom), when the domain Ω is a banana-shaped domain from [9].

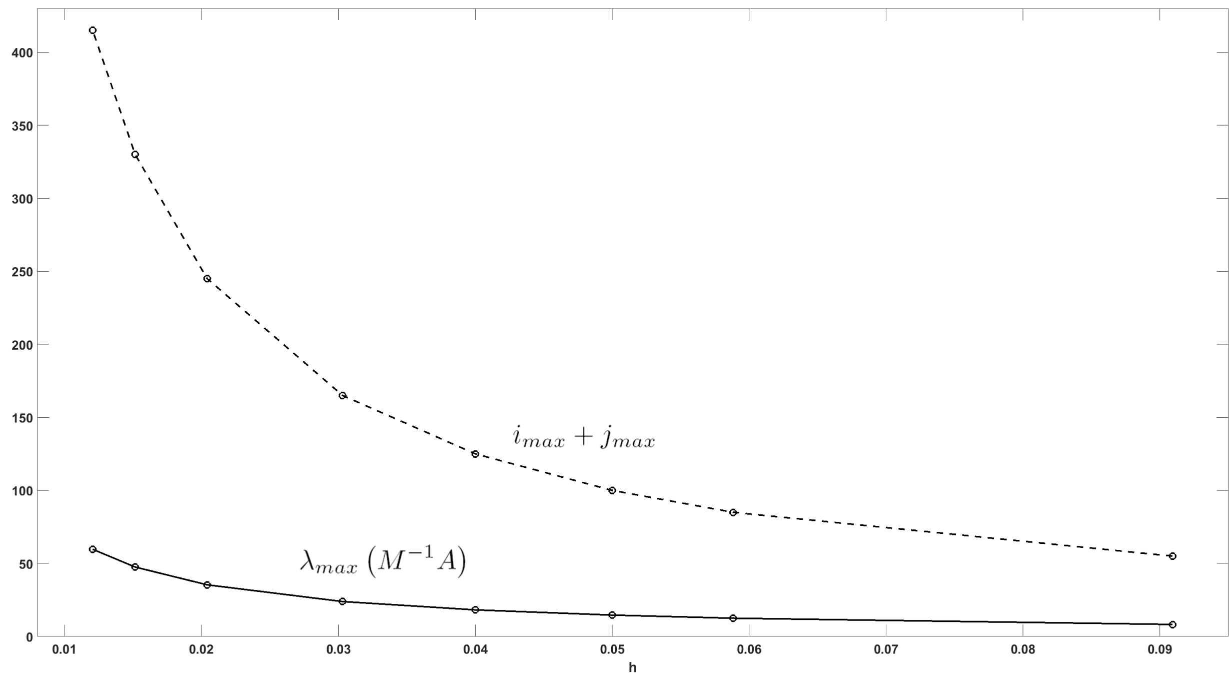


Figure 7: Plots of the maximum eigenvalue $\lambda\left(M^{-1}A\right)$ (solid) and an upper bound $i_{max} + j_{max}$ of the Rayleigh quotient $\langle Au, u\rangle / \langle Mu, u\rangle$ (dotted) for the example in section 6.3. Note that the profile of $\lambda_{max}\left(M^{-1}A\right)$ is bounded above by the upper bound as proved in Corollary 1.

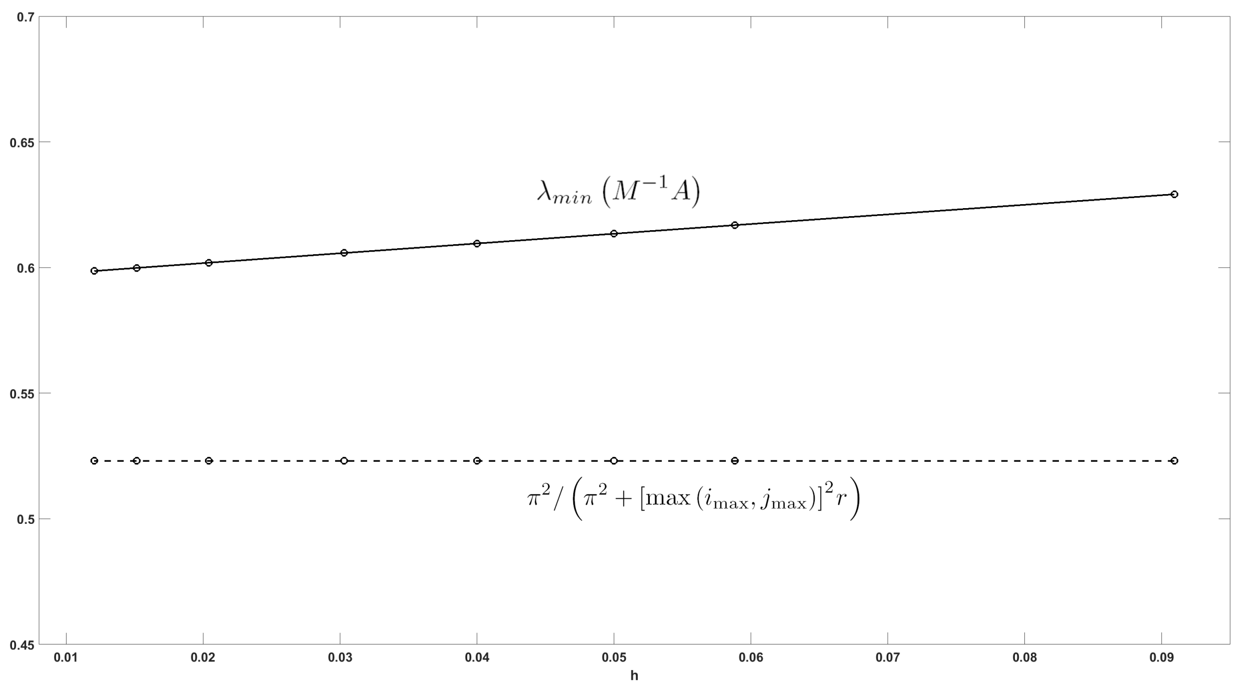


Figure 8: Plots of the maximum eigenvalue $\lambda\left(M^{-1}A\right)$ (solid) and an upper bound $\pi^2/\left(\pi^2 + [\max(i_{max}, j_{max})]^2 r\right)$ of the Rayleigh quotient $\langle Au, u\rangle / \langle Mu, u\rangle$ (dotted) when $r = h^2$ for the example in section 6.3. This result validates Corollary 2, which states $\lambda_{min}\left(M^{-1}A\right)$ are bounded below by the lower bound.

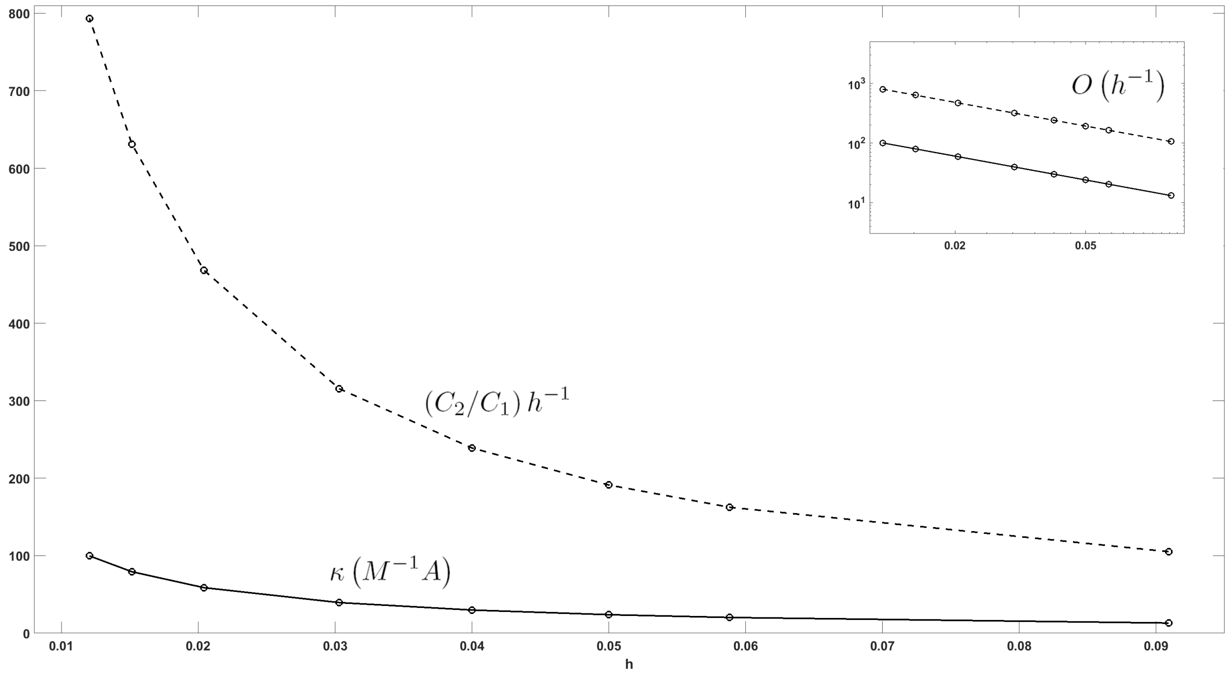


Figure 9: Plots of condition numbers $\kappa\left(M^{-1}A\right)$ and an upper bound $(C_2/C_1)\, h^{-1}$ of $k\left(M^{-1}A\right)$ where C_1, C_2 are constants in the proof of theorem 2 when $r = h^2$ for the example in section 6.3. Numerical results show that $\kappa\left(M^{-1}A\right)$ is bounded above by $(C_2/C_1)\, h^{-1}$, which means $\kappa\left(M^{-1}A\right) = O\left(h^{-1}\right)$. Loglog plot of $\kappa\left(M^{-1}A\right)$ and $(C_2/C_1)\, h^{-1}$ (top right, subfigure) support this argument.

Once U^* is obtained, the incompressibility is enforced with the pressure component at the second stage as follow

$$\begin{cases} \rho\left(\frac{U^{n+1}-U^*}{\Delta t}\right) & = -\nabla p^{n+1} \\ \nabla \cdot U^{n+1} & = 0 \end{cases},$$

which leads to the Poisson equation with Neumann boundary conditions. Since the large linear system appears on the discretization of the second stage, the computational cost for the projection method heavily depends on the efficiency in solving the Poisson equation.

In order to verify the enhancement on the projection method from the choice of optimal preconditioner, let us consider a single vortex flow in a domain $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]^3$ of three dimensions with the initial velocity $U = (u, v, w)$, which is given by

$$\begin{aligned} u\,(x, y, z) &= -2\cos(x)\sin(y)\sin(z) \\ v\,(x, y, z) &= \sin(x)\cos(y)\sin(z) \\ w\,(x, y, z) &= \sin(x)\sin(y)\cos(z)\,. \end{aligned}$$

Among many candidates, we choose the SL-BDF [18] method to discretize the Navier-Stokes equations as follows:

$$\begin{cases} \frac{\frac{3}{2}U^* - 2U_d^n + \frac{1}{2}U_d^{n-1}}{\Delta t} & = \nabla \cdot \left(\nabla U^* + (\nabla U^*)^T\right) + F^{n+1} \\ \frac{\frac{3}{2}U^{n+1} - \frac{3}{2}U^*}{\Delta t} & = -\nabla\left(p^{n+1} - p^n\right) \\ \nabla \cdot U^{n+1} & = 0 \end{cases}.$$

Here, U_d is the velocity obtained by the semi-Lagrangian approximation. The linear systems associated with diffusion and the Poisson equation are solved by PCG. For the comparison purpose, we implement several numerical experiments on a single vortex flow with different preconditioners in solving the Poisson equation.

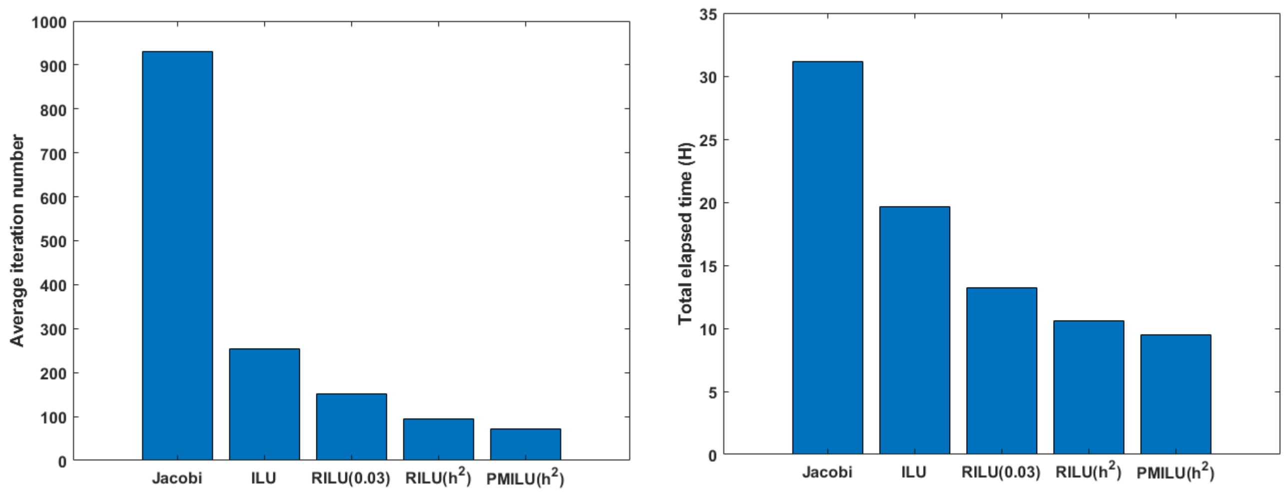


Figure 10: Three dimensional fluid simulation in section 6.4. A standard projection method in a grid 200^3 is utilized to solve the fluid flow. The average iteration number (left) in the projection step and the total elapsed time (right, in hours) are plotted for each preconditioner. PMILU$\left(h^2\right)$ outperforms the other conventional preconditioners. Its average iteration is merely 7.7% of Jacobi, 28% of ILU, 47% of RILU(0.03), and 76% of RILU$\left(h^2\right)$ and its total elapsed time is just 30% of Jacobi, 48% of ILU, 71% of RILU(0.03), and 89% of RILU$\left(h^2\right)$.

	Diffusion (s)	Poisson equation (s)
Jacobi	36	180
ILU	36	88
RILU(0.03)	36	60
RILU$\left(h^2\right)$	36	31
PMILU$\left(h^2\right)$	36	25

Table 1: Computation costs (in seconds) of diffusion and the Poisson equation in one iteration of SL-BDF with the resolution $200 \times 200 \times 200$

Figure 10 shows the performance of SL-BDF for the single vortex flow when a preconditioner is Jacobi, ILU, RILU(0.03), and PMILU$\left(h^2\right)$ with $h = \pi/200$. As we expected, the performance of SL-BDF with PMILU$\left(h^2\right)$ preconditioner in solving the Poisson equation outperforms the other methods both in terms of average iteration numbers (7.7% of Jacobi, 28% of ILU, 47% of RILU(0.03), and 76% of RILU$\left(h^2\right)$) and elapsed times (30% of Jacobi, 48% of ILU, 71% of RILU(0.03), and 89% of RILU$\left(h^2\right)$). Note that there is a discrepancy between the average iteration numbers and the elapsed times in the improvement of performance. This is due to the fact that the computational cost of the linear systems for diffusion in SL-BDF is not negligible when the systems become large. Table 1 shows computational costs of diffusion and the Poisson equation in one iteration for each preconditioner.

7 Conclusion and future work

In this work, we discuss what is an optimal preconditioner in solving the Poisson equation with Neumann boundary conditions. Only the optimal preconditioner results in condition number $O\left(h^{-1}\right)$, while the other preconditioners such as Jacobi and ILU result in $O\left(h^{-2}\right)$. We reviewed RILU [1, 2] and PMILU [3, 16] preconditioners, and presented empirical results which show that RILU is optimal only in two dimensions and PMILU is optimal both in two and three dimensions.

We are not the first to report such empirical results, e.g. Chan et al. [7] and references therein, however, there is no known rigorous analysis that explains the empirically observed optimality, to the best of our knowledge. We have endeavored to provide a rigorous analysis that confirms the optimality and managed to prove the optimality of RILU in the case rectangular domains.

MILU is well known to be optimal in solving the Poisson equation with Dirichlet boundary conditions and there have been well established analysis [3, 28, 30] that confirms the optimality. We reviewed in section 5.1 that the analysis mainly consists of the triangular type inequality (6) and its local and independent application to each grid cell. In section 5.2, we explained why the conventional analysis fails in the case of Neumann boundary conditions.

The main attribute of our success in the proof is the rectangular inequality in lemma 1. When the triangular inequality is applied to each grid cell in the conventional analysis, the triangles in the application are edge disjoint and there is just one way of choosing the weights in the inequality. As explained in section 5.2, the choice is not successful in dealing with Neumann boundary conditions. However, when the rectangular inequality is applied to each grid cell, each edge, except on the boundary, is included twice in the inequalities because an edge is adjacent to two grid cell. With this regard, there are many ways in choosing the weights: any two nonnegative weights are fine if their sum equals the coefficient of the edge in the residual sum. In the case of rectangular domains, the weight on the right and top sided is chosen to gradually increase from zero to a number less than one, and the other weight is then fixed from the constraint on their sum. From these choices, we could prove the optimality of RILU in rectangular domains. In dealing with general smooth domains and PMILU, we believe that there is a certain choice of weights to prove the optimalities. Owing to the limitation of time and resource, we confine this work to RILU in rectangular domains and plan to extend it to RILU and PMILU in smooth domains in near future.

Acknowledgement

The research of Chohong Min was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (Grant No. 2019R1A6A1A11051177). The research of Byungjoon Lee was supported by the Catholic University of Korea, Research Fund, 2019 and POSCO Science Fellowship of POSCO TJ Park Foundation and NRF grant 2020R1A2C4002378.

References

[1] C. Ashcraft and R. Grimes. On vectorizing incomplete factorization and ssor preconditioners. *SIAM Journal on Scientific and Statistical Computing*, 9(1):122–151, 1988.

[2] O. Axelsson and G. Lindskog. On the eigenvalue distribution of a class of preconditioning methods. *Numerische Mathematik*, 48(5):479–498, 1986.

[3] R. Beauwens. Modified incomplete factorization strategies. In *Preconditioned Conjugate Gradient Methods*, pages 1–16. Springer, 1990.

[4] M. S. Birman and M. Z. Solomjak. *Spectral theory of self-adjoint operators in Hilbert space*, volume 5. Springer Science & Business Media, 2012.

[5] R. Bridson. *Fluid simulation for computer graphics*. CRC press, 2015.

[6] T. F. Chan. Fourier analysis of relaxed incomplete factorization preconditioners. *SIAM journal on scientific and statistical computing*, 12(3):668–680, 1991.

[7] T. F. Chan and H. A. Van Der Vorst. Approximate and incomplete factorizations. In *Parallel numerical algorithms*, pages 167–202. Springer, 1997.

[8] E. W. Cheney and D. Kincaid. *Numerical mathematics and computing*. Cengage Learning, 2012.

[9] I.-L. Chern and Y.-C. Shu. A coupling interface method for elliptic interface problems. *Journal of Computational Physics*, 225(2):2138–2174, 2007.

[10] A. J. Chorin. The numerical solution of the navier-stokes equations for an incompressible fluid. *Bulletin of the American Mathematical Society*, 73(6):928–931, 1967.

[11] P. Ciarlet. *The finite element method for elliptic problems*. SIAM, 2002.

[12] B. Cockburn, G. E. Karniadakis, and C.-W. Shu. The development of discontinuous galerkin methods. In *Discontinuous Galerkin Methods*, pages 3–50. Springer, 2000.

[13] I. S. Duff and G. Meurant. The effect of ordering on preconditioned conjugate gradients. *BIT Numerical Mathematics*, 29(4):635–657, 1989.

[14] F. Gibou, R. Fedkiw, L.-T. Cheng, and M. Kang. A second-order-accurate symmetric discretization of the poisson equation on irregular domains. *Journal of Computational Physics*, 176(1):205–227, 2002.

[15] D. Gottlieb and S. A. Orszag. *Numerical analysis of spectral methods: theory and applications*. SIAM, 1977.

[16] I. Gustafsson. A class of first order factorization methods. *BIT Numerical Mathematics*, 18(2):142–156, 1978.

[17] I. Gustafsson. A class of preconditioned conjugate gradient methods applied to finite element equations. In *Preconditioned Conjugate Gradient Methods*, pages 44–57. Springer, 1990.

[18] C. Min and F. Gibou. A second order accurate projection method for the incompressible navier–stokes equations on non-graded adaptive grids. *Journal of Computational Physics*, 219(2):912–929, 2006.

[19] Y. T. Ng, C. Min, and F. Gibou. An efficient fluid–solid coupling algorithm for single-phase flows. *Journal of Computational Physics*, 228(23):8807–8829, 2009.

[20] Y. Notay. Solving positive (semi) definite linear systems by preconditioned iterative methods. In *Preconditioned Conjugate Gradient Methods*, pages 105–125. Springer, 1990.

[21] Y. Park, J. Kim, J. Jung, E. Lee, and C. Min. An efficient milu preconditioning for solving the 2d poisson equation with neumann boundary condition. *Journal of Computational Physics*, 356:115–126, 2018.

[22] H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky. A multidomain spectral method for solving elliptic equations. *Computer physics communications*, 152(3):253–273, 2003.

[23] J. W. Purvis and J. E. Burkhalter. Prediction of critical mach number for store configurations. *AIAA Journal*, 17(11):1170–1177, 1979.

[24] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[25] G. H. Shortley and R. Weller. The numerical solution of laplace's equation. *Journal of Applied Physics*, 9(5):334–348, 1938.

[26] H. A. Van der Vorst. High performance preconditioning. *SIAM Journal on Scientific and Statistical Computing*, 10(6):1174–1185, 1989.

[27] H. A. Van der Vorst. The convergence behaviour of preconditioned cg and cg-s in the presence of rounding errors. In *Preconditioned conjugate gradient methods*, pages 126–136. Springer, 1990.

[28] G. Yoon and C. Min. Analyses on the finite difference method by gibou et al. for poisson equation. *Journal of Computational Physics*, 280:184–194, 2015.

[29] G. Yoon and C. Min. Comparison of eigenvalue ratios in artificial boundary perturbation and jacobi preconditioning for solving poisson equation. *Journal of Computational Physics*, 349:1–10, 2017.

[30] G. Yoon, J.-H. Park, and C. Min. Convergence analysis on the gibou–min method for the hodge projection. *Communications in Mathematical Sciences*, 15(5):1211–1220, 2017.