on the first and second blocks, respectively.

$$
\begin{aligned}
&|(V, (\mathbf{I}-\mathbf{M})\mathbf{N}^T W) + (W, \mathbf{N}(I-\mathbf{M})V)| = 2|((\mathbf{I}-\mathbf{M})V, \mathbf{N}^T W)| \\
&\leq \alpha(V, \mathbf{M}V) + \frac{1}{\alpha}(\mathbf{M}^{-1/2}\mathbf{N}^T W, (\mathbf{I}-\mathbf{M})^2\mathbf{M}^{-1/2}\mathbf{N}^T W) \\
&= \alpha(V, \mathbf{M}V) + \frac{1}{\alpha}(W, \mathbf{N}\mathbf{M}^{-1/2}(\mathbf{I}-\mathbf{M})^2\mathbf{M}^{-1/2}\mathbf{N}^T W).
\end{aligned}
$$

Thus, we obtain

$$
\tilde{\mathcal{R}}_{\mathrm{S},1} \leq \begin{bmatrix} \mathbf{I}-(1-\alpha)\mathbf{M} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}+\mathbf{N}\mathbf{M}^{-1/2}(2\mathbf{M}-\mathbf{M}^2+\frac{1}{\alpha}(\mathbf{I}-\mathbf{M})^2)\mathbf{M}^{-1/2}\mathbf{N}^T \end{bmatrix}.
$$

By substituting $\alpha = \overline{n_m}/2$, the first diagonal block is bounded the same as in Eq. (34), while the second diagonal block is bounded as

$$
\begin{aligned}
&-\mathbf{I}+\mathbf{N}\mathbf{M}^{-1/2}(2\mathbf{M}-\mathbf{M}^2+\frac{1}{\alpha}(\mathbf{I}-\mathbf{M})^2)\mathbf{M}^{-1/2}\mathbf{N}^T \\
&= -\mathbf{I}+\frac{1}{\alpha}\mathbf{N}\mathbf{M}^{-1/2}((1-\alpha)(\mathbf{I}-\mathbf{M})^2+\alpha\mathbf{I})\mathbf{M}^{-1/2}\mathbf{N}^T \\
&\leq -\mathbf{I}+\frac{1}{\alpha}\left((1-\alpha)\sup_{m\in[\underline{m},\overline{m}]}(1-m)^2+\alpha\right)\overline{n_m}\mathbf{I} \\
&\leq -\mathbf{I}+2\left(\left(1-\frac{\overline{n_m}}{2}\right)\sup_{m\in[\underline{m},\overline{m}]}(1-m)^2+\frac{\overline{n_m}}{2}\right)\mathbf{I} \\
&= \mathbf{I}-2\left(1-\frac{\overline{n_m}}{2}\right)(1-\sup_{m\in[\underline{m},\overline{m}]}(1-m)^2)\mathbf{I} \leq \mathbf{I}-(2-\overline{n_m})\min(\underline{m}, 2-\overline{m})\mathbf{I}.
\end{aligned}
$$

Thus, we obtain the upper bound. □

Note For the inexact block LU algorithm, it is not clear whether or not the restriction in Eq. (36) subject to the spectrum of $\mathbf{M}$ is really necessary to bound $\lambda(\tilde{\mathcal{R}}_{\mathrm{S},1})$ larger than -1. We can see from the following that it is really unavoidable. Let us consider the following two-dimensional case.

$$
\tilde{\mathcal{R}}_{\mathrm{S},1} = \begin{bmatrix} 1-m & -(1-m)n \\ -n(1-m) & (2-m)n^2-1 \end{bmatrix}.
$$

The lower eigenvalue is given with $n_m = \frac{n^2}{m}$ as

$$
\begin{aligned}
\underline{\lambda} &= \frac{1}{2}\left(-m+(2-m)n^2-\sqrt{(2-m)^2(1-n^2)^2+4(1-m)^2n^2}\right) \\
&= \frac{1}{2}\left(-m+(2-m)m\,n_m-\sqrt{(2-m)^2(1-m\,n_m)^2+4(1-m)^2m\,n_m}\right)
\end{aligned}
$$

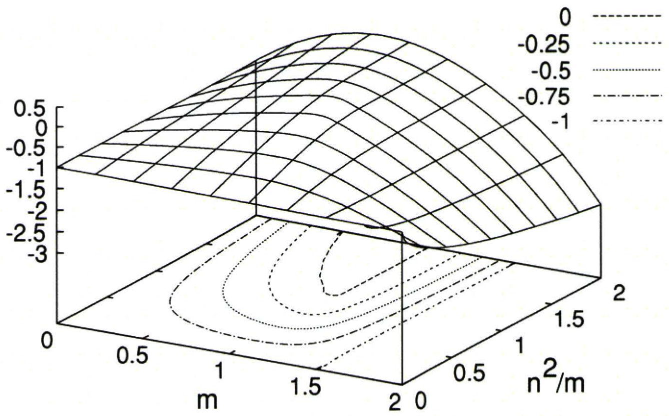


Fig. 1 Landscape of the lower eigenvalue of the 2 × 2 matrix $\widetilde{\mathcal{R}}_{\mathrm{S},1}$ with respect to m (x-axis) and n^2/m (y-axis)

As can be seen in Fig. 1, $\underline{\lambda} < -1$ for $m > 3/2$. Thus, the condition in Eq. (36) cannot be relaxed.

The bounds obtained thus far provide the reduction rate of the error vectors that are given after the transformation in Eq. (27). Here, we give the reduction rate of the error vectors before the transformation, as was done in [3]. We introduce a norm by using the transformation matrix as follows.

$$\left[\left|\begin{bmatrix}V\\W\end{bmatrix}\right|\right] \equiv \left(\mathcal{T}\begin{bmatrix}V\\W\end{bmatrix}, \mathcal{T}\begin{bmatrix}V\\W\end{bmatrix}\right)^{1/2} = \left(\begin{bmatrix}V\\W\end{bmatrix}, \mathcal{T}^T\mathcal{T}\begin{bmatrix}V\\W\end{bmatrix}\right)^{1/2}$$

The positive matrices that determine the norms are given by

$$\mathcal{G}_{\mathrm{T}} \equiv \mathcal{T}_{\mathrm{T}}^T\mathcal{T}_{\mathrm{T}} = \begin{bmatrix}\mathbf{Q}_A - \mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_B\end{bmatrix} \tag{39}$$

for the inexact block triangular algorithm, and

$$\mathcal{G}_{\mathrm{S}} \equiv \mathcal{T}_{\mathrm{S}}^T\mathcal{T}_{\mathrm{S}} = \begin{bmatrix}\mathbf{Q}_A & \mathbf{B}^T\\ \mathbf{B} & \mathbf{Q}_B + \mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T\end{bmatrix} \tag{40}$$

for the inexact block LU algorithm. Note that the condition in Eq. (22) is necessary to define the norm as $\mathcal{G}_{\mathrm{T}}$ in the inexact block triangular case, whereas the condition in Eq. (36) is not required to ensure the positiveness of $\mathcal{G}_{\mathrm{S}}$ since the Schur complement on the second block is equal to $\mathbf{Q}_B$. Using the norms defined above, we can restate the previous lemmas as follows.

Theorem 3 *For the inexact block triangular algorithm:*
Suppose the following conditions are satisfied:

$$\begin{aligned}&0 < \delta_a\mathbf{Q}_A \le \mathbf{A} < \mathbf{Q}_A \text{ with some } \delta_a > 0,\\ &0 < \delta_b\mathbf{Q}_B \le \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T \le (2-\delta_b)\mathbf{Q}_B \text{ with some } \delta_b > 0.\end{aligned} \tag{41}$$

Then, the iterative errors in Eq. (15) *satisfy*

$$\left\| \begin{bmatrix} E^X_{i+1} \\ E^Y_{i+1} \end{bmatrix} \right\| \leq \rho \left\| \begin{bmatrix} E^X_i \\ E^Y_i \end{bmatrix} \right\| \quad \text{with } \rho = 1 - \frac{1}{2}\delta_a \delta_b.$$

Proof This is immediately obtained from Lemma 1. □

Note that in [3] the condition, $\mathbf{BA}^{-1}\mathbf{B} \leq \mathbf{Q}_B$, is required to obtain the upper bound of $\lambda(\widetilde{\mathcal{R}}_{\mathrm{T},1})$, whereas the condition in Eq. (41) is relaxed.

Theorem 4 *For the inexact block LU case:*
Suppose the following conditions are satisfied:

$$\begin{aligned} &0 < \delta_a \mathbf{Q}_A \leq \mathbf{A} \leq \frac{3}{2}(1-\delta_a)\mathbf{Q}_A, \quad \text{with some } \delta_a > 0, \\ &0 < \delta_b \mathbf{Q}_B \leq \mathbf{BA}^{-1}\mathbf{B}^T \leq (2-\delta_b)\mathbf{Q}_B \quad \text{with some } \delta_b > 0. \end{aligned} \tag{42}$$

Then, the iterative errors in Eq. (15) *satisfy*

$$\left\| \begin{bmatrix} E^X_{i+1} \\ E^Y_{i+1} \end{bmatrix} \right\| \leq \rho \left\| \begin{bmatrix} E^X_i \\ E^Y_i \end{bmatrix} \right\| \quad \text{with } \rho = 1 - \frac{1}{2}\delta_a \delta_b.$$

Proof This is immediately obtained from Lemma 2. □

When we apply the iterative methods above, it is desirable to use $\mathbf{Q}_A$ and $\mathbf{Q}_B$ for which δ_a and δ_b are defined independently of the mesh size. Otherwise, an extraordinarily large number of iterations is required. Our target is to propose an efficient, yet simply implementable, preconditioner for applications in nonlinear continuum analysis on an unstructured grid. In this area, it seems to be very difficult to construct good $\mathbf{Q}_A$ and $\mathbf{Q}_B$ matrices. This provides the motivation to apply $\mathcal{P}$ as a preconditioner and explore the convergence properties of the preconditioned Krylov subspace with simply implementable matrices $\mathbf{Q}_A$ and $\mathbf{Q}_B$, such as the ILU or symmetric Gauss–Seidel matrix. Note also that the transformation $\mathcal{T}_{\mathrm{S}}$ in Eq. (29) and therefore the norm matrix (40) can be defined for any positive matrices $\mathbf{Q}_A$ and $\mathbf{Q}_B$. Thus, in the next section, we focus on the use of $\mathcal{P}_{\mathrm{S}}$ as a preconditioner for the Krylov subspace methods.

3 Inexact block LU matrix as a preconditioner

In this section, we analyze the eigenvalue distribution of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$ for the preconditioner $\mathcal{P} = \mathcal{P}_{\mathrm{S}}$ associated with the inexact block LU algorithm. Note that $\mathcal{P}_{\mathrm{S}}$ can be represented using the transformation matrix $\mathcal{T}_{\mathrm{S}}$ in Eq. (29) as

$$\mathcal{P}_{\mathrm{S}} = \mathcal{T}_{\mathrm{S}}^T \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \mathcal{T}_{\mathrm{S}}.$$

Thus, a computation similar to Eq. (24) leads to

$$\mathcal{T}_{\mathrm{S}}\mathcal{P}_{\mathrm{S}}^{-1}\mathcal{A}\mathcal{T}_{\mathrm{S}}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \mathcal{T}_{\mathrm{S}}^{-T}\mathcal{A}\mathcal{T}_{\mathrm{S}}^{-1} = \begin{bmatrix} \mathbf{M} & (\mathbf{I}-\mathbf{M})\mathbf{N}^T \\ -\mathbf{N}(\mathbf{I}-\mathbf{M}) & \mathbf{N}(2\mathbf{I}-\mathbf{M})\mathbf{N}^T \end{bmatrix}. \tag{43}$$

Though the coefficient matrix $\mathcal{A}$ and the preconditioner $\mathcal{P}_{\mathrm{S}}$ are symmetric, it is not possible to define a "real" symmetric matrix similar to the preconditioned matrix $\mathcal{P}_{\mathrm{S}}^{-1}\mathcal{A}$ since both matrices are indefinite. Hence, applications of the Krylov subspace methods for nonsymmetric matrices are unavoidable, and we cannot make any exact estimation of convergence speed of the preconditioned Krylov subspace iteration based solely on eigenvalue distribution for the preconditioned matrix. Nevertheless, information on the eigenvalue distribution of the preconditioned system provides a good indication of the performance of the preconditioner.

Let V and W be the first and second block vectors, respectively, of the eigenvector with eigenvalue λ for the matrix in Eq. (43). Then, we obtain

$$\begin{bmatrix} \frac{(V,\mathbf{M}V)}{(V,V)} & \frac{(V,(\mathbf{I}-\mathbf{M})\mathbf{N}^T W)}{(V,V)} \\ -\frac{(W,\mathbf{N}(\mathbf{I}-\mathbf{M})V)}{(W,W)} & \frac{(W,\mathbf{N}(2\mathbf{I}-\mathbf{M})\mathbf{N}^T W)}{(W,W)} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{44}$$

Here, the symbol (,) means the natural Hermitian dot product of the complex vector space that satisfies

$$(V, \mathbf{N}(\mathbf{I}-\mathbf{M})^T W) = \overline{(W, \mathbf{N}(\mathbf{I}-\mathbf{M})^T V)}, \quad \forall V, W.$$

We assume that V and W are not zero, since the eigenvalue analysis is trivial if one of them is zero. If we define the real numbers as

$$\alpha = \frac{(V,\mathbf{M}V)}{(V,V)}, \quad \beta = \frac{|(V,(\mathbf{I}-\mathbf{M})\mathbf{N}^T W)|}{\|V\|\|W\|}, \quad \gamma = \frac{(W,\mathbf{N}(2\mathbf{I}-\mathbf{M})\mathbf{N}^T W)}{(W,W)}, \tag{45}$$

then we obtain from Eq. (44)

$$(\alpha-\lambda)(\gamma-\lambda)+\beta^2 = 0.$$

Thus, the eigenvalue is given as

$$\lambda = \frac{\alpha+\gamma \pm \sqrt{(\alpha-\gamma)^2-4\beta^2}}{2}. \tag{46}$$

Lemma 5 *Assume that, in addition to the condition on* $\mathbf{M}$ *in Eq.* (36), *the following condition holds.*

$$\underline{n}\mathbf{I} \le \mathbf{N}\mathbf{N}^T \le \overline{n}\mathbf{I}, \quad i.e.\ \underline{n}\mathbf{Q}_B \le \mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T \le \overline{n}\mathbf{Q}_B.$$

If the eigenvalue λ *is a real number, then it is bounded as*

$$\min\left(\underline{m}, \frac{\underline{n}}{2}\right) \leq \lambda \leq \max(\overline{m}, 2\overline{n}). \tag{47}$$

If the imaginary part of λ *is not zero, then*

$$\frac{1}{2}\left(\underline{m} + \frac{\underline{n}}{2}\right) \leq \mathrm{Re}(\lambda) \leq \frac{1}{2}(\overline{m} + 2\overline{n}), \tag{48}$$

$$|\mathrm{Im}(\lambda)| \leq \min(\sqrt{\mathrm{Re}(\lambda)}, \sqrt{\overline{n}}), \tag{49}$$

and the relation $\|V\| = \|W\|$ *holds.*

Proof If λ is a real number, we see from Eq. (46) that

$$\min(\alpha, \gamma) = \frac{\alpha + \gamma - |\alpha - \gamma|}{2} \leq \lambda \leq \frac{\alpha + \gamma + |\alpha - \gamma|}{2} = \max(\alpha, \gamma).$$

From the condition in Eq. (36), we also obtain

$$\frac{1}{2}\underline{n} \leq \gamma \leq 2\overline{n}. \tag{50}$$

This leads to the estimation in Eq. (47).

Next, we consider the case where the imaginary part of λ is not zero. In this case, we immediately see that

$$\mathrm{Re}(\lambda) = \frac{\alpha + \gamma}{2}.$$

Thus, from Eq. (50) we obtain Eq. (48). From the condition in Eq. (36), we see that

$$2(\mathbf{I} - \mathbf{M})^2 \leq 2\mathbf{I} - \mathbf{M}.$$

Thus, the imaginary part of λ can be bounded as follows.

$$\begin{aligned}\mathrm{Im}(\lambda) \leq \beta = \frac{|(V, (\mathbf{I} - \mathbf{M})\mathbf{N}^T W)|}{\|V\|\|W\|} &\leq \left(\frac{(W, \mathbf{N}(\mathbf{I} - \mathbf{M})^2\mathbf{N}^T W)|}{\|W\|^2}\right)^{1/2} \\ \leq \left(\frac{(W, \mathbf{N}(2\mathbf{I} - \mathbf{M})\mathbf{N}^T W)|}{2\|W\|^2}\right)^{1/2} &= \sqrt{\frac{\gamma}{2}} \leq \min(\sqrt{\mathrm{Re}(\lambda)}, \sqrt{\overline{n}}).\end{aligned} \tag{51}$$

As for the norm of the block vectors, from the first and second rows in Eq. (44), we see that

$$\mathrm{Im}(\lambda) = \mathrm{Im}\left(\frac{(V, (\mathbf{I} - \mathbf{M})\mathbf{N}^T W)}{(V, V)}\right) = \mathrm{Im}\left(\frac{-(W, \mathbf{N}(\mathbf{I} - \mathbf{M})V)}{(W, W)}\right).$$

Since the numerator on both sides is the same, we obtain $\|V\| = \|W\|$. □

It may also be valuable to consider the possibility of using a Krylov subspace method for real symmetric matrices like MINRES [11] which is applicable to indefinite systems. A similar study on the Stokes equations was performed by Silvester and Wathen for block diagonal preconditioners [13,15,16]. Here, we consider the following block LU-type positive symmetric matrix.

$$\mathcal{P}_{+\mathrm{S}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{B}\mathbf{Q}_A^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_A & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_B \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{Q}_A^{-1}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

It differs from $\mathcal{P}_{\mathrm{S}}$ in (11) only with respect to the sigh in the second diagonal block in the central block diagonal matrix. Note that $\mathcal{P}_{+\mathrm{S}}$ can be rewritten using $\mathcal{T}_{\mathrm{S}}$ in Eq. (29) as

$$\mathcal{P}_{+\mathrm{S}} = \mathcal{T}_{\mathrm{S}}^T \mathcal{T}_{\mathrm{S}}.$$

Thus, we see that the similarity relation given below holds for the preconditioned matrix.

$$\begin{aligned} &\mathcal{P}_{+\mathrm{S}}^{-1}\mathcal{A} \sim \mathcal{T}_{\mathrm{S}}^{-T}\mathcal{A}\mathcal{T}_{\mathrm{S}}^{-1} \\ &= \begin{bmatrix} \mathbf{Q}_A^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_B^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{B}\mathbf{Q}_A^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{Q}_A^{-1}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_A^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_B^{-1/2} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{M} & (\mathbf{I}-\mathbf{M})\mathbf{N}^T \\ \mathbf{N}(\mathbf{I}-\mathbf{M}) & -\mathbf{N}(2\mathbf{I}-\mathbf{M})\mathbf{N}^T \end{bmatrix} \end{aligned}$$

Note that if $\mathbf{Q}_A = \mathbf{A}$ and $\mathbf{Q}_B = \mathbf{S}$ hold, the preconditioned matrix is similar to $\widetilde{\mathcal{E}}$ defined in Eq. (30). Thus, the MINRES iterative process converges in two iterations. This observation indicates that the above preconditioner performs well if $\mathbf{Q}_A$ and $\mathbf{Q}_B$ are chosen appropriately. From similar computations for $\mathcal{P}_{\mathrm{S}}^{-1}\mathcal{A}$, we see that any eigenvalue of the preconditioned matrix $\mathcal{P}_{+\mathrm{S}}^{-1}\mathcal{A}$ satisfies

$$(\alpha - \lambda)(-\gamma - \lambda) - \beta^2 = 0.$$

Thus, we obtain

$$\lambda = \frac{\alpha - \gamma \pm \sqrt{(\alpha+\gamma)^2 + 4\beta^2}}{2}. \qquad (52)$$

Lemma 6 *Based on the same assumption as Lemma 5, the positive eigenvalues of $\mathcal{P}_{+\mathrm{S}}^{-1}\mathcal{A}$ are bounded as*

$$\underline{m} \le \lambda \le \overline{m} + \sqrt{\overline{n}},$$

while the negative eigenvalues are bounded as

$$-2\overline{n} - \sqrt{\overline{n}} \le \lambda \le -\frac{n}{2}.$$

Proof Note that we obtain the following bounds for β from the same estimation in Eq. (51).

$$\beta^2 \le \frac{\gamma}{2} \le \overline{n}. \tag{53}$$

If $\lambda > 0$, then λ is bounded as

$$\begin{aligned}\lambda &\le \frac{(\alpha-\gamma)+(\alpha+\gamma)+2\beta}{2} \le \alpha+\beta \le \overline{m}+\sqrt{\overline{n}},\\ \lambda &\ge \frac{(\alpha-\gamma)+(\alpha+\gamma)}{2} = \alpha \ge \underline{m}.\end{aligned}$$

If $\lambda < 0$, then λ is bounded as

$$\begin{aligned}\lambda &\le \frac{(\alpha-\gamma)-(\alpha+\gamma)}{2} = -\gamma \le -\frac{1}{2}\underline{n},\\ \lambda &\ge \frac{(\alpha-\gamma)-(\alpha+\gamma)-2\beta}{2} = -\gamma-\beta \ge -2\overline{n}-\sqrt{\overline{n}}.\end{aligned}$$

□

In [7], the following estimation for the convergence of MIRES is given, based on the assumption that the eigenvalues are contained in $[-a, -b] \cup [c, d]$ for $\exists a > \exists b > 0$, $\exists d > \exists c > 0$.

$$\frac{\|r_k\|}{\|r_0\|} \le \min_{P_k, P_k(0)=1} \max_{\lambda_i} |P_k(\lambda_i)| \le 2\left(\frac{\sqrt{\frac{ad}{bc}}-1}{\sqrt{\frac{ad}{bc}}+1}\right)^{k/2}, \tag{54}$$

where r_0 and r_k are residual vectors at the initial and the kth iterations, respectively, and P_k denotes a kth order polynomial. From Eq. (54) and Lemma 6, we make the following observation. Assume that we are dealing with a finite element problem in which the order of the element size is h. It is reasonable to assume that we can construct matrices $\mathbf{Q}_A$ and $\mathbf{Q}_B$, of which $\overline{m}$ and $\overline{n}$ are bounded independently by the element size h. That is,

$$\overline{m} = O(1), \quad \overline{n} = O(1).$$

Then, the condition

$$\left(\frac{\sqrt{\frac{ad}{bc}}-1}{\sqrt{\frac{ad}{bc}}+1}\right)^{k/2} \le \epsilon$$

implies

$$k \ge |\log \epsilon| O\left(\left(\underline{m}\,\underline{n}\right)^{-1/2}\right).$$

If both $\underline{m}$ and $\underline{n}$ are $O(h^2)$, the above inequality implies

$$k \ge |\log \epsilon| O\left(\frac{1}{h^2}\right).$$

However, if one of them is $O(1)$, then

$$k \ge |\log \epsilon| O\left(\frac{1}{h}\right).$$

In the numerical experiments, we see how the bounds $\underline{m}$ and $\underline{n}$ behave with respect to changes in h and how accurate the above estimates are. We also compare the convergence of the above approach with the preconditioned GMRES by $\mathcal{P}_{\mathrm{S}}$.

4 Fill-controlled ILU matrix as a preconditioner

The iterative methods analyzed in Sects. 2 and 3 require construction of $\mathbf{Q}_B$ which should be an approximation of the Schur complement $\mathbf{S} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ or its approximation $\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T$. However, as these matrices are generally dense, the computation thereof should be avoided in practice. In this section, we analyze the property of the ILU preconditioners, proposed in [14], that are not affected by this difficulty. The ILU preconditioner is represented as

$$\mathcal{P}_{\mathrm{I}} = \begin{bmatrix} \mathbf{L}_A + \mathbf{D}_A & \mathbf{0} \\ \widehat{\mathbf{B}} & \mathbf{L}_B + \mathbf{D}_B \end{bmatrix} \begin{bmatrix} \mathbf{D}_A^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D}_B^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{D}_A + \mathbf{L}_A^T & \widehat{\mathbf{B}}^T \\ \mathbf{0} & \mathbf{D}_B + \mathbf{L}_B^T \end{bmatrix}, \tag{55}$$

where $\mathbf{L}_A$ and $\mathbf{L}_B$ are strictly lower triangular matrices, and $\mathbf{D}_A$ and $\mathbf{D}_B$ are diagonal matrices. The nonzero pattern of $\mathbf{L}_A$ is the same as that in the lower triangular part of $\mathbf{A}$, and the nonzero pattern of $\mathbf{L}_B$ is the same as that of the lower triangular part of $\widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T$, where the nonzero pattern of $\widehat{\mathbf{B}}$ is determined from any fill control strategy. Thus, the factorization in the second diagonal block is none other than the ILU factorization of $\widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T$ without fill-ins, since the right hand side of Eq. (55) expands to

$$\mathcal{P}_\mathrm{I} = \begin{bmatrix} (\mathbf{L}_A+\mathbf{D}_A)\mathbf{D}_A^{-1}(\mathbf{D}_A+\mathbf{L}_A^T) & (\mathbf{L}_A+\mathbf{D}_A)\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T \\ \widehat{\mathbf{B}}\mathbf{D}_A^{-1}(\mathbf{D}_A+\mathbf{L}_A^T) & \widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T - (\mathbf{L}_B+\mathbf{D}_B)\mathbf{D}_B^{-1}(\mathbf{D}_B+\mathbf{L}_B^T) \end{bmatrix}. \tag{56}$$

We assume that all the diagonal components of $\mathbf{D}_A$ and $\mathbf{D}_B$ are positive. Then, the preconditioner can be rewritten as

$$\mathcal{P}_\mathrm{I} = \mathcal{T}_\mathrm{I}^T \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \mathcal{T}_\mathrm{I}, \text{ with } \mathcal{T}_\mathrm{I} = \begin{bmatrix} \mathbf{D}_A^{-1/2}(\mathbf{D}_A+\mathbf{L}_A^T) & \mathbf{D}_A^{-1/2}\widehat{\mathbf{B}}^T \\ \mathbf{0} & \mathbf{D}_B^{-1/2}(\mathbf{D}_B+\mathbf{L}_B^T) \end{bmatrix}.$$

Hence, through the similarity transformation of the preconditioned matrix with $\mathcal{T}_\mathrm{I}$, we obtain

$$\begin{aligned} \mathcal{T}_\mathrm{I}\mathcal{P}_\mathrm{I}^{-1}\mathcal{A}\mathcal{T}_\mathrm{I}^{-1} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \mathcal{T}_\mathrm{I}^{-T}\mathcal{A}\mathcal{T}_\mathrm{I}^{-1} \\ &= \begin{bmatrix} \mathbf{M}_\mathrm{I} & \mathbf{N}_\mathrm{I}^T - \mathbf{M}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T \\ -(\mathbf{N}_\mathrm{I} - \widehat{\mathbf{N}}_\mathrm{I}^T\mathbf{M}_\mathrm{I}) & \mathbf{N}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T + \widehat{\mathbf{N}}_\mathrm{I}\mathbf{N}_\mathrm{I}^T - \widehat{\mathbf{N}}_\mathrm{I}\mathbf{M}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T \end{bmatrix}. \end{aligned} \tag{57}$$

Here, the following notations are used.

$$\begin{aligned} \mathbf{M}_\mathrm{I} &= \left((\mathbf{L}_A+\mathbf{D}_A)\mathbf{D}_A^{-1/2}\right)^{-1}\mathbf{A}\left(\mathbf{D}_A^{-1/2}(\mathbf{D}_A+\mathbf{L}_A^T)\right)^{-1}, \\ \mathbf{N}_\mathrm{I} &= \left((\mathbf{L}_B+\mathbf{D}_B)\mathbf{D}_B^{-1/2}\right)^{-1}\mathbf{B}\left(\mathbf{D}_A^{-1/2}(\mathbf{D}_A+\mathbf{L}_A^T)\right)^{-1}, \\ \widehat{\mathbf{N}}_\mathrm{I} &= \left((\mathbf{L}_B+\mathbf{D}_B)\mathbf{D}_B^{-1/2}\right)^{-1}\widehat{\mathbf{B}}\mathbf{D}_A^{-1/2}. \end{aligned}$$

From Eq. (56), we see that $\widehat{\mathbf{N}}_\mathrm{I} = \mathbf{N}_\mathrm{I}$ holds if all the fill-ins in the off-diagonal blocks are taken into account in the construction of $\widehat{\mathbf{B}}$. In this case, Eq. (57) has the same form as Eq. (43). Thus, we obtain similar eigenvalue estimations. However, considering all the fill-ins in the off-diagonal blocks is impractical. Thus, in general, we cannot ensure the positiveness of the second diagonal block in Eq. (57). Hence, we cannot guarantee that all the eigenvalues of the preconditioned matrix are contained in the right-half plane.

As in Sect. 3, we can also construct a positive version of Eq. (55) as

$$\mathcal{P}_{+\mathrm{I}} = \begin{bmatrix} \mathbf{L}_A+\mathbf{D}_A & \mathbf{0} \\ \widehat{\mathbf{B}} & \mathbf{L}_B+\mathbf{D}_B \end{bmatrix} \begin{bmatrix} \mathbf{D}_A^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_B^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{D}_A+\mathbf{L}_A^T & \widehat{\mathbf{B}}^T \\ \mathbf{0} & \mathbf{D}_B+\mathbf{L}_B^T \end{bmatrix} = \mathcal{T}_\mathrm{I}^T\mathcal{T}_\mathrm{I}.$$

In this case, the similarity transformation gives

$$\mathcal{P}_{+\mathrm{I}}^{-1}\mathcal{A} \sim \mathcal{T}_\mathrm{I}^{-T}\mathcal{A}\mathcal{T}_\mathrm{I}^{-1} = \begin{bmatrix} \mathbf{M}_\mathrm{I} & \mathbf{N}_\mathrm{I}^T - \mathbf{M}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T \\ \mathbf{N}_\mathrm{I} - \widehat{\mathbf{N}}_\mathrm{I}^T\mathbf{M}_\mathrm{I} & -(\mathbf{N}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T + \widehat{\mathbf{N}}_\mathrm{I}\mathbf{N}_\mathrm{I}^T - \widehat{\mathbf{N}}_\mathrm{I}\mathbf{M}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T) \end{bmatrix}.$$

Once again, we cannot theoretically bound the eigenvalue distribution of the second diagonal block in the negative region.

5 Numerical experiments

We now examine the performance of the preconditioners introduced in Sects. 3 and 4. The problem tested is a deformation problem of the so-called Mooney–Rivlin body, for which the deformation potential per unit volume is given by

$$W(\mathbf{C}) = c_1\widehat{\mathrm{I}}_C + c_2\widehat{\mathrm{II}}_C. \tag{58}$$

Here, **C** is the right Cauchy Green deformation tensor defined as

$$\mathbf{C} = \left(\mathbf{I} + \frac{\partial \mathbf{u}}{\partial \mathbf{X}}\right)^T \left(\mathbf{I} + \frac{\partial \mathbf{u}}{\partial \mathbf{X}}\right),$$

where **X** is the position vector in the undeformed configuration, and $\mathbf{u} = \mathbf{u}(\mathbf{X})$ is the displacement vector of the material point **X** after the deformation. The invariants are defined as

$$\begin{aligned}\widehat{\mathrm{I}}_C &= \det(\mathbf{C})^{-1/3}\mathrm{tr}(\mathbf{C}),\\ \widehat{\mathrm{II}}_C &= \det(\mathbf{C})^{-2/3}\left(\mathrm{tr}(\mathbf{C})^2 - \mathrm{tr}(\mathbf{C}^2)\right).\end{aligned}$$

In the above case, we solve the variational problem for the energy functional:

$$E = \int_\Omega W(\mathbf{C})d\Omega$$

with the incompressibility constraint, $\det(\mathbf{C}) - 1 = 0$. Here, Ω is the domain of material in the undeformed configuration. The constraint variational problem described in Eqs. (5) and (6) is represented by the following weak formulation:

$$\int_\Omega \left(\frac{\partial W}{\partial \mathbf{C}} + \lambda \det(\mathbf{C})\mathbf{C}^{-1}\right) : \frac{\partial \mathbf{C}}{\partial \mathbf{u}}\delta\mathbf{u}d\Omega = 0, \quad \forall\delta\mathbf{u} \tag{59}$$

$$\int_\Omega \delta\lambda(\det(\mathbf{C}) - 1)d\Omega = 0, \ \forall\delta\lambda. \tag{60}$$

The notation ':' denotes the standard dot product of the tensors, and we apply the formula:

$$\frac{\partial \det(\mathbf{C})}{\partial \mathbf{C}} = \det(\mathbf{C})\mathbf{C}^{-1}.$$

As described in Sect. 1, Eqs. (59) and (60) are further linearized to perform the Newton–Raphson iterations, and we have to solve the system of linear equations given in Eqs. (8) and (9).

When applying a finite element discretization to constraint problems, we have to be mindful of the stability of the discretized linear system with respect to the applied finite element interpolations to the displacement $\mathbf{u}$ and the Lagrange multiplier λ [4]. To ensure stability, we adopt the so-called mini element [1], in which a central displacement node is added to each triangular or tetrahedral element and the Lagrange multiplier nodes are placed on the vertices.

In the inexact block LU preconditioner defined in Eq. (11), we need to determine the matrices $\mathbf{Q}_A$ and $\mathbf{Q}_B$, where $\mathbf{Q}_A$ should be an approximation of $\mathbf{A}$. Thus, we adopt the ILU factorization without fill-in:

$$\mathbf{Q}_A = (\mathbf{L}_A + \mathbf{D}_A)\mathbf{D}_A^{-1}(\mathbf{D}_A + \mathbf{L}_A^T). \tag{61}$$

Note that $\mathbf{L}_A$ and $\mathbf{D}_A$ are the same matrices that appeared in the first block of the ILU factorization in Eq. (55) of the whole matrix $\mathcal{A}$. According to Lemma 5, $\mathbf{Q}_B$ is preferably a good approximation of $\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T$. Though it is easy to perform the matrix multiplication by $\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T$ of any given vector, the direct computation of the components of $\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T$ leads to a dense matrix. Hence, to work with only sparse matrices, we again adopt the factorization in Eq. (55) as

$$\mathbf{Q}_B = (\mathbf{L}_B + \mathbf{D}_B)\mathbf{D}_B^{-1}(\mathbf{D}_B + \mathbf{L}_B^T), \tag{62}$$

based on the following consideration. From Eq. (56), $\mathbf{Q}_B$ in Eq. (62) is expected to be an approximation of $\widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T$. And again from the off-diagonal part in the same equation, $(\mathbf{L}_A + \mathbf{D}_A)\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T$ is expected to be an approximation of $\mathbf{B}^T$. Thus, the following chain of approximative relations holds.

$$\mathbf{B}\left\{(\mathbf{L}_A + \mathbf{D}_A)\mathbf{D}_A^{-1}(\mathbf{D}_A + \mathbf{L}_A^T)\right\}^{-1}\mathbf{B}^T \approx \widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T \approx (\mathbf{L}_B + \mathbf{D}_B)\mathbf{D}_B^{-1}(\mathbf{D}_B + \mathbf{L}_B^T)$$

As for the fill control to construct $\mathbf{Q}_A$ by factorizing $\mathbf{A}$, all fill-ins are ignored. Also, in the factorization of $\widehat{\mathbf{B}}\mathbf{D}_A^{-1}\widehat{\mathbf{B}}^T$ to construct $\mathbf{Q}_B$, we do not allow any fill-ins. However, in the factorization at the off-diagonal block to compute $\widehat{\mathbf{B}}$, we test both level 1 fill-ins and the no fill-in case. We denote these preconditioners by $\mathcal{P}_{\mathrm{S}}(L)$ and $\mathcal{P}_{\mathrm{I}}(L)$, where L is the fill-in level allowed for $\widehat{\mathbf{B}}$.

We now examine the performance of the preconditioners $\mathcal{P}_{\mathrm{S}}(L)$ and $\mathcal{P}_{\mathrm{I}}(L)$ with linear systems arising from a deformation analysis by a simple stretch, in which the undeformed configuration is either a unit square or a unit cube as depicted in Fig. 2. Due to the symmetry of geometry, we compute only one half of the actual material in each direction. For example, the following boundary conditions are imposed on the displacement vector function $\mathbf{u} = (u_1, u_2, u_3)^T$ in the case of the cube.

$$\begin{aligned}&u_1 = u_2 = 0, \quad u_3 = D_3 \text{ on } \{X_3 = 1\} : \text{(the top)}, \quad u_3 = 0 \text{ on } \{X_3 = 0\} : \text{(the bottom)},\\ &u_1 = 0 \text{ on } \{X_2 = 0\}, \quad u_2 = 0 \text{ on } \{X_1 = 0\} : \text{(the side walls)}.\end{aligned}$$

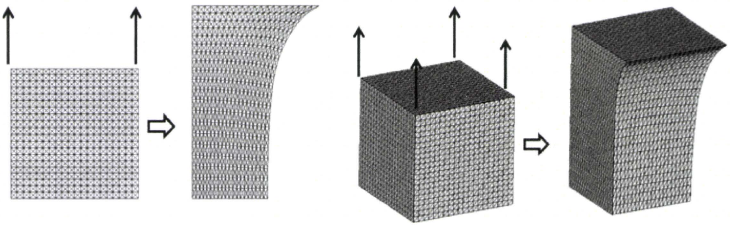


Fig. 2 The initial and final configurations of the two-dimensional 20^2 mesh (*left*) and three-dimensional 20^3 mesh (*right*)

Table 1 The degrees of freedom (D.O.F.) and the number of nonzero components in the matrix (nz($\mathcal{A}$))

Mesh size	D.O.F.	nz($\mathcal{A}$)
Two-dimensional problem		
20^2	5,640	111,686
40^2	22,480	452,166
80^2	89,760	1,819,526
Three-dimensional problem		
10^3	26,510	2,324,764
20^3	206,020	19,201,294

Fig. 3 Arnoldi procedure for $\mathcal{AP}^{-1}$

```
$v_1 = r_0 (= b - \mathcal{A}u_0)$; $v_1 = v_1/\|v_1\|$
for $i = 1, \ldots, m$;
    $w_i = \mathcal{AP}^{-1}v_i$;
    for $j = 1, \ldots, i$;
        $h_{j,i} = (w_i, v_j)$;$w_i = w_i - h_{j,i}v_j$;
    next $j$;
    $h_{i+1,i} = \|w_i\|$; $v_{i+1} = w_i/h_{i+1,i}$;
next $i$;
```

The magnitude of stretch D_3 is increased by 0.1 in each step up to $D_3 = 0.5$. Thus, five incremental steps are required to reach the final configuration. As for the material parameters in Eq. (58), $c_1 = 0.3$, $c_2 = 0.15$ are adopted. The number of elements adopted in each direction is either 20, 40 or 80 in the two-dimensional case, and 10 or 20 in the three-dimensional case. The total degrees of freedom, except for the imposed boundary condition and the number of nonzero components in the coefficient matrix, are given in Table 1. Note that even though the matrices are sparse, there are nearly twenty and a hundred nonzeros in each row in the two- and three-dimensional cases, respectively.

In the linear solutions, we adopted the right-preconditioned full GMRES. The preconditioned GMRES algorithm is based on the Arnoldi procedure [12], in which the orthonormal basis $\{v_i\}_{i=1,\ldots,m}$ is constructed for the preconditioned matrix $\mathcal{AP}^{-1}$ as described in Fig. 3.

By inspecting the Hessenberg matrix defined by

$$H(m) = (h_{j,i})_{i,j=1,m} = ((\mathcal{A}\mathcal{P}^{-1}v_j, v_i))_{i,j=1,m}, \quad (h_{j,i} = 0 \quad \text{for } i+1 < j),$$

we can see how the matrix $\mathcal{A}\mathcal{P}^{-1}$ acts on the Krylov subspace spanned by the basis vectors $v_1, \dots, v_m$. Thus, at the final solution stage, we extracted the Hessenberg matrix $H(m)$ (where m is the number of iterations) and examined its eigenvalue distribution. Table 2 gives the bounds of the eigenvalue distribution, the average number of GMRES iterations in one solution process, and the average CPU time in one solution process. In the experiment, each GMRES iteration was stopped when the relative L2-norm of the residual was smaller than 10^{-8}. For each increment of the displacement, on average four Newton–Raphson iterations were required before the nonlinear equation converged. The computation was performed with a single CPU (Intel(R) Pentium(R) model 4, Clock 3.4 GHz, Cache size 16 kB, Main memory 2 GB). To validate the estimates in Eq. (48) of the eigenvalues with nonzero imaginary parts for $\mathcal{P}_S$, we picked an eigenvalue $\widehat{\lambda}$ that satisfies

$$\mathrm{Re}(\widehat{\lambda}) = \min\{\mathrm{Re}(\lambda) \mid \mathrm{Im}(\lambda) \neq 0\}.$$

The following observations together with their interpretations associated with the theory in Sects. 3 and 4 were made.

1. Minimal real part of the eigenvalues
   In all cases, these minimums are given on the real axis. The values are the same for all preconditioners for a common mesh size. The order of these is obviously h^2, where h indicates the element size. For $\mathcal{P}_S$, these minimums can possibly be determined only from $\underline{m}$ in Eq. (47), since the minimums are the same for the different fill levels of $\widehat{\mathbf{B}}$ which causes the change in $\mathbf{N}$, and thus also the change in $\underline{n}$.
2. Maximal real parts of the eigenvalues
   These are almost independent of the element size. In the case of $\mathcal{P}_S$, they are larger than 2, but decrease with an increase in the fill-in level. This indicates that they are determined from $\overline{n}$ in Eq. (47). It also means that using $\mathcal{P}_S$ in the matrix splitting iterations in Sect. 2 leads to the divergence. We can apply an appropriate scaling factor to $\mathbf{Q}_B$ so that the condition in Eq. (42) in Theorem 4 is satisfied. However, this does not bring about any meaningful convergence improvement in the use as a preconditioner in these examples. The decrease in the maximums with the associated increase in the fill-in level in $\widehat{\mathbf{B}}$ is brought about by the improvement in the approximation of $\widehat{\mathbf{B}}\mathbf{Q}_A\mathbf{B}^T$ by $\mathbf{Q}_B$ that leads to the decrease in $\overline{n}$. For $\mathcal{P}_\mathrm{I}$, the maximums are well bounded, although they increased slightly when the fill-ins were applied.
3. Bounds for the imaginary parts of the eigenvalues
   For $\mathcal{P}_\mathrm{S}$, the bounds are much smaller than one and almost independent of the element size and the fill-in level in $\widehat{\mathbf{B}}$. We cannot give a good explanation for this insensitivity to the fill-in level based on the current theory. On the other hand, the bounds are greater than one for $\mathcal{P}_\mathrm{I}(0)$. But, the bounds are improved in $\mathcal{P}_\mathrm{I}(1)$.

4. Eigenvalues with negative real parts
   As predicted in Theorem 4, all the eigenvalues are included in the right half plane for $\mathcal{P}_\mathrm{S}$. However, for $\mathcal{P}_\mathrm{I}$ there are a few eigenvalues outside this plane, and these disappear if the fill-ins in $\widehat{\mathbf{B}}$ are applied in some cases. This reflects the inherent difficulty in judging the positiveness of the matrix $\mathbf{N}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T + \widehat{\mathbf{N}}_\mathrm{I}\mathbf{N}_\mathrm{I}^T - \widehat{\mathbf{N}}_\mathrm{I}\mathbf{M}_\mathrm{I}\widehat{\mathbf{N}}_\mathrm{I}^T$ in Eq. (57). Note that the imaginary parts of these eigenvalues are not that small and as such, they may not affect the convergence rate too severely.
5. Convergence and execution time
   The dependence of the number of iterations on element size h is slightly larger than $O(1/h)$ for both preconditioners. However, it is much smaller than $O(1/h^2)$. There is no complete theory to explain this convergence behavior. However, this tendency of $\mathcal{P}_\mathrm{S}$ seems to be closely related to the distribution of eigenvalues with

Table 2 The eigenvalue distribution bounds in the right-half plane at the final equilibrium, eigenvalues in the left-half plane if they exist in column 'negative Re', and the average number of GMRES iterations and solution times

Prec.	Mesh	min Re(λ)	max Re(λ)	max Im(λ)	$\widehat{\lambda}$	Negative Re
Two-dimensional triangular mesh (eigenvalue distribution)						
$\mathcal{P}_\mathrm{S}(0)$	20	0.016	10.6	0.44	(0.069, ±0.22)	None
	40	0.0041	10.6	0.45	(0.018, ±0.11)	None
	80	0.0010	10.6	0.47	(0.0047, ±0.057)	None
$\mathcal{P}_\mathrm{I}(0)$	20	0.016	1.68	1.26	–	(−0.011, ±0.51)
	40	0.0041	1.70	1.24	–	(−0.083, ±0.26)
	80	0.0010	1.71	1.24	–	(−0.044, ±0.13) (−0.046, ±0.13) (−0.0099, ±0.56)
$\mathcal{P}_\mathrm{S}(1)$	20	0.016	4.7	0.45	(0.10, ±0.27)	None
	40	0.0041	4.6	0.47	(0.026, ±0.14)	None
	80	0.0010	4.6	0.47	(0.0066, ±0.071)	None
$\mathcal{P}_\mathrm{I}(1)$	20	0.016	1.82	0.77	–	None
	40	0.0041	1.84	0.74	–	(−0.0096, ±0.24)
	80	0.0010	1.84	0.74	–	(−0.014, ±0.12)
Three-dimensional tetrahedral mesh (eigenvalue distribution)						
$\mathcal{P}_\mathrm{S}(0)$	10	0.010	5.52	0.44	(0.11, ±0.0017)	None
	20	0.0025	5.53	0.46	(0.082, ±0.25)	None
$\mathcal{P}_\mathrm{I}(0)$	10	0.010	1.51	1.06	–	(−0.076, ±0.73)
	20	0.0025	1.53	0.97	–	(−0.099, ±0.33) (−0.0052, ±0.49)
$\mathcal{P}_\mathrm{S}(1)$	10	0.010	3.09	0.35	(0.74, ±0.36)	None
	20	0.0025	3.17	0.46	(0.21, ±0.37)	None
$\mathcal{P}_\mathrm{I}(1)$	10	0.010	1.60	0.76	–	None
	20	0.0025	1.64	0.65	–	None

Table 2 Continued

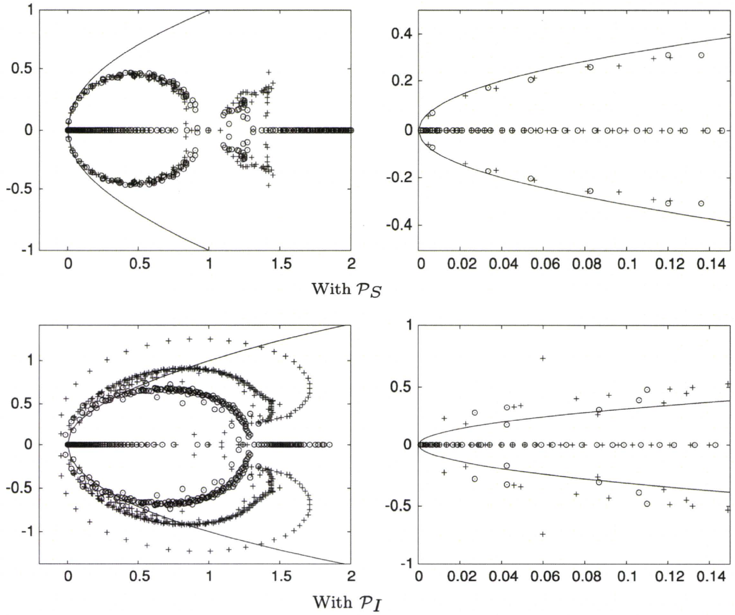
Iterations and solution times

Prec.	Mesh	♯IT	Time (s)
Two-dimensional triangular mesh			
$\mathcal{P}_\mathrm{S}(0)$	20	78	0.31
	40	165	3.39
	80	374	48.5
$\mathcal{P}_\mathrm{I}(0)$	20	84	0.26
	40	192	3.37
	80	459	58.2
$\mathcal{P}_\mathrm{S}(1)$	20	62	0.27
	40	129	2.52
	80	287	33.0
$\mathcal{P}_\mathrm{I}(1)$	20	59	0.23
	40	134	2.45
	80	315	34.7
Three-dimensional tetrahedral mesh			
$\mathcal{P}_\mathrm{S}(0)$	10	100	4.55
	20	227	120
$\mathcal{P}_\mathrm{I}(0)$	10	113	3.79
	20	275	130
$\mathcal{P}_\mathrm{S}(1)$	10	78	4.30
	20	169	90.0
$\mathcal{P}_\mathrm{I}(1)$	10	76	5.29
	20	170	105

The fill level for $\widehat{B}$ is indicated in parentheses in column 'Prec'

nonzero imaginary parts. This is discussed further when we compare the convergence with the positive preconditioner $\mathcal{P}_{+\mathrm{S}}$. In general, the convergence with $\mathcal{P}_\mathrm{S}$ is substantially better than with $\mathcal{P}_\mathrm{I}$, and the difference becomes larger with an increase in mesh size $O(1/h)$. The convergence deterioration with $\mathcal{P}_\mathrm{I}$ may be as a result of the presence of those eigenvalues not included in the right-half plane, since a fairly similar convergence is observed when they disappear with the fill-ins.

Next, in addition to observing the simple bounds of the eigenvalues in the table, we examine the distribution in greater detail, in particular, focusing the eigenvalues around the origin. The eigenvalue distributions for $\mathcal{P}_\mathrm{S}(0)$, $\mathcal{P}_\mathrm{S}(1)$ and $\mathcal{P}_\mathrm{I}(0)$, $\mathcal{P}_\mathrm{I}(1)$ in the 80^2 square problem are depicted in Fig. 4. The plots on the left show the global distributions, while those on the right show the local distributions around the origin in the right-half plane. For comparison with the estimation of the imaginary part in Eq. (49), the curve defined by $x = y^2$ is also depicted. The results for $\mathcal{P}_\mathrm{S}$ show that



With $\mathcal{P}_S$

With $\mathcal{P}_I$

Fig. 4 Eigenvalue distributions for the two-dimensional problem (80^2). *Plus symbol* and *open circle* indicate fill levels of 0 and 1, respectively. The *solid line* shows the curve $y^2 = x$

Eq. (49) provides an exact estimate of the magnitude of the imaginary part around the origin. On the other hand, most of the eigenvalues are out of the bounds around the origin for $\mathcal{P}_\mathrm{I}$. This indicates difficulties of theoretical prediction for the convergence performance of $\mathcal{P}_\mathrm{I}$. However, $\mathcal{P}_\mathrm{I}$ realizes comparable convergence speed to $\mathcal{P}_\mathrm{S}$ as we will see later in this section.

Since there does not appear to be any theory that can explain the almost $\mathrm{O}(1/h)$ convergence rate for $\mathcal{P}_\mathrm{S}$ from the eigenvalue distribution, we validate the applicability of Eq. (54) for the positive symmetric preconditioner $\mathcal{P}_{+\mathrm{S}}$. Table 3 gives the eigenvalue bounds $[-a, -b] \cup [c, d]$ for the Hessenberg matrix for $\mathcal{A}\mathcal{P}_{+\mathrm{S}}(1)^{-1}$ and the average number of GMRES iterations for the two-dimensional case. Note that here we can basically apply MINRES instead of GMRES to reduce the computational cost of the dot products. However, as we are interested in the comparison of the convergence rate with the GMRES iteration preconditioned by $\mathcal{P}_\mathrm{S}(1)$, GMRES is applied to $\mathcal{A}\mathcal{P}_{+\mathrm{S}}(1)^{-1}$. The upper bound $-b$ in the negative region seems to decrease with order h, while the lower bound c in the positive region decreases with order h^2. According to Eq. (54), this results in an order for the number of iterations of $(1/h)^{3/2}$. A similar estimate was obtained by Wathen for a block diagonal preconditioner for the Stokes equation [16]. However, a simple calculation from the number of iterations in the table,

Table 3 Comparison of the eigenvalue bounds and the convergence for $P_{+S}(1)$ and $P_S(1)$ in the two-dimensional case

Mesh size	$-a$	$-b$	c	d	♯IT
$\mathcal{P}_{+S}$					
20	−5.3	−0.37	0.016	1.7	150
40	−5.3	−0.16	0.0041	1.8	333
80	−5.3	−0.075	0.0010	1.9	691
Mesh size	min Re(λ)	max Re(λ)	$\widehat{\lambda}$	♯IT	
$\mathcal{P}_S$					
20	0.016	4.7	(0.10, ±0.27)	62	
40	0.0041	4.6	(0.026, ±0.14)	129	
80	0.0010	4.6	(0.0066, ±0.071)	287	

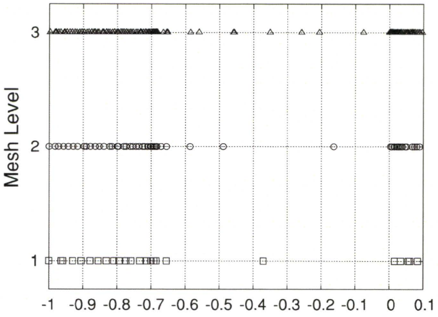


Fig. 5 Eigenvalue distributions on the real axis for the three mesh levels (1:20^2, 2:40^2, 3:80^2)

such as

$$150 \cdot 2^{3/2} \approx 424 \gg 333, \quad 333 \cdot 2^{3/2} \approx 942 \gg 691,$$

indicates faster convergence than that predicted from the theory. The reason for the discrepancy in this convergence estimate may be better understood if we look more closely at the eigenvalue distribution depicted in Fig. 5. The interesting thing here is that there is a common border over all the mesh levels where the dense and sparse distributions in the negative region are separated. Although the number of eigenvalues in the sparse part increases with an increase in the mesh level, their distribution is still much sparser compared with the dense part. This certainly brings about a faster convergence than the previous prediction. The reason of this tendency in the eigenvalue distributions is not entirely clear at the moment. According to the theory, it seems to be related to the eigenvalue distribution of $\mathbf{NN}^T \sim \mathbf{Q}_B^{-1}\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T$.

It is also interesting to compare the convergence rate with $\mathcal{P}_S$. The numbers of iterations with $\mathcal{P}_{+S}$ are more than double those with $\mathcal{P}_S$. In the extreme case where the

Table 4 Comparison of the convergence of the first diagonal block matrix **A** and the whole matrix $\mathcal{A}$

Mesh size	$\mathbf{Q}_A$ for **A**		$\mathcal{P}_{\mathrm{S}}(1)$ for $\mathcal{A}$		$\mathcal{P}_{\mathrm{I}}(1)$ for $\mathcal{A}$	
	♮IT	Time (s)	♮IT	Time (s)	♮IT	Time (s)
20^2	53	0.091	62	0.27	59	0.23
40^2	104	0.82	129	2.52	134	2.45
80^2	201	9.55	287	33.0	315	34.7
10^3	84	3.96	78	5.29	76	4.30
20^3	162	49.2	169	90.0	170	105

exact block LU factorization (i.e. $\mathbf{Q}_A = \mathbf{A},\ \mathbf{Q}_B = \mathbf{S}$) is applied, the preconditioned matrix for the former is similar to $\widetilde{\mathcal{E}}$ defined in Eq. (30), whereas for the latter it is the identity matrix. This may explain the tendency above. We cannot, however, give any reasonable explanation for this relation in the inexact block LU factorization case, because the Krylov subspace, produced by each of these, is different.

Finally, we compare the convergence of the whole matrix with that of only the first diagonal block **A**. In this comparison, we extracted the first diagonal block **A** and F from the whole system in Eq. (1) and solved $\mathbf{A}X = F$ using the GMRES iteration with the preconditioner $\mathbf{Q}_A$ in each Newton–Raphson step. Note that we applied GMRES to the positive symmetric problems, since we are interested in the comparison of the GMRES iteration for the whole system. In Table 4, the average number of iterations and the CPU time are compared. These results imply reasonable performance of the proposed preconditioners for the whole system. In particular, the number of iterations is almost the same for the three-dimensional problem.

6 Conclusions

The main objective of this study was to explore the inexact LU-type preconditioners for saddle point problems arising in nonlinear continuum analysis. Many of the previous studies on saddle point problems assumed that good preconditioners, $\mathbf{Q}_A$ for **A** or $\mathbf{Q}_B$ for **S**, were available. However, in real-life applications of finite element analysis for hyper-elastic materials, this requirement is too strict since we do not have a good understanding of the properties of the resultant matrices **A** and **B** and they usually do not have any hierarchical structure which makes the application of multilevel methods easier. Thus, it is important to construct the best preconditioner for the whole system with the available approximative matrices $\mathbf{Q}_A$ and $\mathbf{Q}_B$. From this point of view, we have tried to make exact estimates of the eigenvalue distribution and to understand the convergence behavior using an approximation $\mathbf{Q}_A$ which is not spectrally equivalent to **A**. We have shown that we can achieve almost order $1/h$ for the number of iterations before convergence by constructing $\mathbf{Q}_A$ and $\mathbf{Q}_B$ from the simple ILU factorizations. Although our theoretical explanation for this convergence rate is still incomplete, we have provided a good starting point for further development of the theory. Furthermore, such a theoretical consideration will certainly bring about further improvements in preconditioning techniques for use in the area of highly nonlinear incompressible continuum analysis.

References

1. Arnold, D.N., Brezzi, F., Fortin, M.: A stable finite element for stokes equations. Calcolo **21**, 337–344 (1984)
2. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numerica **14**, 1–137 (2005)
3. Bramble, J.H., Pasciak, J.E., Vassilev, A.T.: Analysis of the inexact Uzawa algorithm for saddle point problems. SIAM Numer. Anal. **34**, 1072–1092 (1997)
4. Brezzi, F., Bathe, K.J.: A discourse on the stability conditions for mixed finite element formulations. Comput. Methods Appl. Mech. Eng. **82**, 27–57 (1990)
5. Elman, H.C., Silvester, D.J., Wathen, A.J.: Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations. Numer. Math. **90**, 641–664 (2002)
6. Elman, H.C., Silvester, D.J., Wathen, A.J.: Finite Elements and Fast Iterative Solvers. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford (2005)
7. Greenbaum, A.: Iterative Methods for Solving Linear Systems. Frontiers in Applied Mathematics, vol. 17. SIAM, Philadelphia (1997)
8. Klawonn, A., Starke, G.: Block triangular preconditioners for nonsymmetric saddle point problems: field-of-values analysis. Numer. Math. **81**, 577–594 (1999)
9. Little, L., Saad, Y., Smoch, L.: Block preconditioners for symmetric and nonsymmetric saddle point problems. SIAM J. Sci. Comput. **25**, 729–748 (2003)
10. Mardal, K.A., Winther, R.: Uniform preconditioners for the time dependent Stokes problem. Numer. Math. **98**, 305–327 (2004)
11. Paige, C.C., Saunders, M.A.: Solution of sparse indefinite systems of linear equations. SIAM J. Numer. Anal. **12**, 617–629 (1975)
12. Saad, Y., Shultz, M.H.: GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Comput. **7**, 856–869 (1986)
13. Silvester, D.J., Wathen, A.J.: Fast iterative solution of stabilised Stokes systems part II: using general block preconditioners. SIAM J. Numer. Anal. **31**, 1352–1367 (1994)
14. Washio, T., Hisada, T., Watanabe, H., Tezduyar, T.E.: A robust and efficient iterative linear solver for strongly coupled fluid-structure interaction problems. Comput. Methods Appl. Mech. Eng. **194**, 4027–4047 (2005)
15. Wathen, A.J., Silvester, D.J.: Fast iterative solution of stabilised Stokes systems part I: using simple diagonal preconditioners. SIAM J. Numer. Anal. **30**, 630–649 (1993)
16. Wathen, A.J., Fischer, B., Silvester, D.J.: The convergence rate of the minimal residual method for the Stokes problem. Numer. Math. **71**, 121–134 (1995)

Baroreflex Sensitivity Might Predict Responders to Milrinone in Patients With Heart Failure

Takuya KISHI,[1] MD, *and* Kenji SUNAGAWA,[1] MD

SUMMARY

The phosphodiesterase III inhibitor milrinone (MIL) is considered to be effective for "wet and cold" heart failure. In some cases, however, the inotropic effects of milrinone are insufficient. A previous study suggested that baroreflex sensitivity (BRS) predicts the cases in which MIL increases left ventricular *dp/dt*. The aim of this study was to determine whether BRS measured using the spontaneous sequence method predicts the MIL responders. Twenty-four patients with "wet and cold" heart failure whose systolic blood pressure > 100 mmHg were enrolled. At 2 hours MIL improved dyspnea, general fatigue, urine volume, and tricuspid regurgitant pressure gradient in 13 patients (responders; R group), whereas it failed to improve in 11 patients (nonresponders; NR group). BRS in the R group was significantly higher than that in the NR group prior to the MIL infusion. At 2 hours after the MIL infusion, BRS was further increased in the R group, but did not increase in the NR group. The sensitivity and specificity of BRS at a cut-off level of 5 ms/mmHg for the prediction of R group were 0.94 and 0.93, respectively. BRS might be useful for identifying potential responders to milrinone in patients with blood pressure-preserved "wet and cold" heart failure. (Int Heart J 2010; 51: 411-415)

Key words: Heart failure, Baroreflex sensitivity, Milrinone

Milrinone, a phosphodiesterase-III inhibitor (PDEIII-I), has an inotropic and vasodilator effect for "wet and cold" heart failure,[1,2] which is determined as heart failure with congestion and hypoperfusion.[3] In some cases, however, the inotropic effects of milrinone are insufficient and combined treatment with dobutamine is necessary.[1] A previous study suggested that baroreflex sensitivity (BRS) can predict the cases in which milrinone increases left ventricular *dp/dt*.[4] However, whether arterial baroreflex function is related to the inotropic responsiveness to milrinone has not been clarified in human heart failure.

Baroreflex control is one of the key mechanisms responsible for the short-term control of blood pressure.[5-7] Impairment of this reflex has been found in a number of conditions, such as aging,[8] post myocardial infarction,[9,10] hypertension,[7] and heart failure.[11] Baroreflex sensitivity was originally assessed by intra-arterial measurement of the change in pulse interval following a pharmacologically induced change in blood pressure. However, for some time now, noninvasive monitoring of blood pressure using finger plethysmography has been available, and further methods for measuring baroreflex sensitivity have been developed, which assess spontaneous changes in blood pressure and pulse interval, and do not require pharmacological manipulation of blood pressure-spectral analysis.[12-16]

The aim of this study was to determine whether the baroreflex sensitivity measured using the spontaneous sequence method can identify potential milrinone responders or not in patients with sinus rhythm and blood pressure-preserved "wet and cold" heart failure.

METHODS

The present study was approved by the Ethics Committee for Human Research of Kyushu University Graduate School of Medical Sciences. Data collected retrospectively were fully de-identified.

Patient populations: We retrospectively studied patients with symptomatic acute heart failure admitted to Kyushu University Hospital from January 2006 to December 2007 who were treated with intravenous infusion of milrinone. The criteria for enrollment in the study were clinical evidence of acute heart failure diagnosed by Framingham criteria[17] and low cardiac output, which is called "wet and cold" heart failure.[3] We defined low cardiac output from the clinical state of "cold and wet". In those patients, the New York Heart Association (NYHA) functional classification on admission ranged between III and IV. We excluded patients whose systolic blood pressure was < 100 mmHg or who had atrial fibrillation, chronic obstructive pulmonary disease, dehydration, right ventricular myocardial infarction, or right heart failure. Prior medication by intravenous injection of diuretics, nitrates, and morphine was permitted. The dose of milrinone was adjusted according to the condition of each individual patient, and if symptoms of heart failure were not adequately improved by milrinone, concomitant use of or replacement with other agents indicated for the treatment of acute heart failure was

From the [1] Department of Advanced Therapeutics for Cardiovascular Diseases, Cardiovascular Medicine, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan.

Address for correspondence: Takuya Kishi, MD, Department of Advanced Therapeutics for Cardiovascular Diseases, Department of Cardiovascular Medicine, Kyushu University Graduate School of Medical Sciences, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.

Received for publication June 17, 2010.

Revised and accepted September 9, 2010.