Les Cahiers du GERAD ISSN: 0711–2440

On GSOR, the generalized successive overrelaxation method for double saddle-point problems

N. Huang, Y.-H. Dai, D. Orban, M. A. Saunders

G–2022–35

August 2022

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2022 – Bibliothèque et Archives Canada, 2022

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2022 – Library and Archives Canada, 2022

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

On GSOR, the generalized successive overrelaxation method for double saddle-point problems

Na Huang [a]

Yu-Hong Dai [b]

Dominique Orban [c]

Michael A. Saunders [d]

[a] *Department of Applied Mathematics, College of Science, China Agricultural University, Beijing, China*

[b] *LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*

[c] *GERAD & Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Qc), Canada*

[d] *Systems Optimization Laboratory, Department of Management Science and Engineering, Stanford University, Stanford, CA, USA*

hna@cau.edu.cn
dyh@lsec.cc.ac.cn
dominique.orban@gerad.ca
saunders@stanford.edu

August 2022
Les Cahiers du GERAD
G–2022–35

Abstract : We consider the generalized successive overrelaxation (GSOR) method for solving a class of block three-by-three saddle-point problems. Based on the necessary and sufficient conditions for all roots of a real cubic polynomial to have modulus less than one, we derive convergence results under reasonable assumptions. We also analyze a class of block lower triangular preconditioners induced from GSOR and derive explicit and sharp spectral bounds for the preconditioned matrices. We report numerical experiments on test problems from the liquid crystal director model and the coupled Stokes-Darcy flow, demonstrating the usefulness of GSOR.

Keywords : Iterative methods, double saddle-point systems, saddle-point problems, matrix splitting, successive overrelaxation, preconditioning

Résumé : Nous considérons la méthode de surrelaxation successive généralisée (GSOR) pour la résolution d'une classe de systèmes de points de selle à trois par trois blocs. Sur base des conditions nécessaires et suffisantes pour que les racines d'un polynôme de degré trois soient de module inférieur à l'unité, nous établissons la convergence de la méthode sous des hypothèses raisonnables. Nous analysons également une calsse de préconditionneurs triangulaires inférieurs par blocs induits par GSOR et dérivons des bornes spectrales explicites et fines pour les matrices préconditionnées. Nous présentons des résultats numériques sur des problèmes de crystaux liquides et du flux couplé de Stokes-Darcy, et illustrons l'utilité de GSOR.

Acknowledgements: This research began with the work of our colleague and friend Dr Oleg Burdakov. In 2019, Oleg focused on Barzilai-Borwein-type methods to solve quasi-definite linear systems, and he conducted preliminary tests on double saddle-point problems. While testing the BB-type methods, we found that GSOR performs well on the double saddle-point problems and were motivated to start this work. We are grateful for Oleg's insight and foresight.

We are also grateful to Mingchao Cai, Alison Ramage, and Zhaozheng Liang for providing the test problems used in our numerical experiments.

Research of N. Huang is partially supported by National Natural Science Foundation of China (No. 12001531). Research of D. Orban is partially supported by an NSERC Discovery Grant.

1 Introduction

We consider the double saddle-point problem

$$\mathcal{A}\,w := \begin{pmatrix} A & B^T & C^T \\ B & 0 & 0 \\ C & 0 & -D \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f \\ g \\ h \end{pmatrix} =: b, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{p\times p}$ are symmetric positive definite (SPD) matrices, $B \in \mathbb{R}^{m\times n}$ has full row rank, and $C \in \mathbb{R}^{p\times n}$. Linear systems like (1) arise from many practical applications, such as mixed and mixed-hybrid finite element approximation of the liquid crystal director model [17] and coupled Stokes-Darcy flow [6, 8, 12, 13], and interior methods for quadratic programming problems [9, 10, 19]. We emphasize that (1) is importantly different from the block 3×3 systems considered by Huang et al. in [14, 15].

In principle, (1) can be treated as the block 2×2 saddle-point problem

$$\begin{pmatrix} H & E^T \\ E & -W \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{2}$$

which has been studied for decades [5]. We focus here on splitting iterative methods for (1) by fully utilizing the special structure of $\mathcal{A}$. The generalized successive overrelaxation (GSOR) method of Bai et al. [1] is for (2) with $W = 0$. We extend GSOR to (1) by introducing three parameters. The convergence analysis of this new GSOR method is quite different from that of stationary methods; we derive convergence conditions based on the necessary and sufficient conditions for all roots of a real cubic polynomial to have modulus less than one. We also analyze a class of block lower triangular preconditioners induced from GSOR and show that all eigenvalues of the preconditioned matrices are positive real and can be clustered by appropriate selections of parameters.

For linear systems discretized from a mixed Stokes-Darcy model, Cai et al. [6] proposed preconditioning techniques by treating (1) as system (2) with $W = 0$. Ramage and Gartland Jr. [17] studied a preconditioned nullspace method for solving systems (1) that arise from discretizations of continuum models for the orientational properties of liquid crystals, in which they also partitioned $\mathcal{A}$ into a block 2×2 form. Recently, based on the special structure of $\mathcal{A}$ in (1), several preconditioners were proposed to accelerate Krylov subspace methods. Beik and Benzi [2, 3] analyzed several block diagonal and block triangular preconditioners and derived bounds for the eigenvalues of the preconditioned matrices. An alternating positive semidefinite splitting (APSS) preconditioner and its relaxed variant were proposed by Liang and Zhang [16] to solve double saddle-point problems arising from liquid crystal director models. The improved APSS preconditioner of Ren et al. [18] and the two-parameter block triangular preconditioner of Zhu et al. [21] were also constructed to deal with the same saddle-point problem. However, the latter preconditioners either do not fully exploit the special structure of $\mathcal{A}$ or need to solve several complicated and dense linear systems at each iteration.

It is generally difficult to analyze the spectral properties of a "full" block three-by-three matrix; i.e., one that cannot be reduced to a block 2×2 matrix. Little literature exists on iterative schemes for (1). Uzawa-like methods based on the splitting

$$\mathcal{A} = \begin{pmatrix} A & 0 & 0 \\ B & -\frac{1}{\alpha}Q & 0 \\ C & 0 & M \end{pmatrix} - \begin{pmatrix} 0 & -B^T & -C^T \\ 0 & -\frac{1}{\alpha}Q & 0 \\ 0 & 0 & N \end{pmatrix} \tag{3}$$

were studied by Benzi and Beik [4], where $\alpha > 0$ and the SPD matrix Q are given, and $D = N - M$ with M negative definite. In addition, given a parameter $\omega \neq 0$, they split $\mathcal{A}$ into

$$\mathcal{A} = \frac{1}{\omega}\begin{pmatrix} A & B^T & 0 \\ B & 0 & 0 \\ \omega C & 0 & -D \end{pmatrix} - \frac{1}{\omega}\begin{pmatrix} (1-\omega)A & (1-\omega)B^T & -\omega C^T \\ (1-\omega)B & 0 & 0 \\ 0 & 0 & -(1-\omega)D \end{pmatrix}$$

and proposed a generalized block successive overrelaxation (GBSOR) method. The convergence analysis of these two methods is similar to that of stationary iterative schemes for block 2×2 linear systems, where convergence conditions are derived from a quadratic polynomial equation of the eigenvalues of the iteration matrix. Moreover, GBSOR needs to solve four linear systems at each step: two of the form $Ax = r_1$, one $BA^{-1}B^Ty = r_2$, and one $Dz = r_3$. By partitioning $\mathcal{A}$ into system (2) with $H = A$, Dou and Liang [7] construct a class of block alternating splitting implicit (BASI) iteration methods. At each step, BASI needs to solve several linear systems of the form $\alpha I + A + aB^TB + bC^TC$ and $\alpha I + D + cCC^T$, where I is the identity and α, a, b, c are real scalar constants.

The paper is organized as follows. In Section 2, we present the generalized successive overrelaxation method. In Section 3, convergence of GSOR is established under reasonable assumptions. Preconditioners are analyzed in Section 4. Numerical experiments are reported in Section 5. Conclusions are summarized in Section 6.

Notation

For any $S \in \mathbb{R}^{r\times r}$, its spectral radius, inverse and transpose are denoted $\rho(S)$, S^{-1} and S^T, respectively. For any $s \in \mathbb{C}^r$, its conjugate transpose is denoted s^*.

2 The generalized successive overrelaxation (GSOR) method

In this section, we present GSOR for solving the double saddle-point problem (1). We consider the equivalent unsymmetric system

$$\widehat{\mathcal{A}}\,w := \begin{pmatrix} A & B^T & C^T \\ -B & 0 & 0 \\ -C & 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f \\ -g \\ -h \end{pmatrix} =: \widehat{b}. \tag{4}$$

Although $\widehat{\mathcal{A}}$ is unsymmetric, it has certain desirable properties:

1. $\widehat{\mathcal{A}}$ is semipositive real: $v^T\widehat{\mathcal{A}}v \geq 0$ for all $v \in \mathbb{R}^{n+m+p}$.
2. $\widehat{\mathcal{A}}$ is positive semistable; i.e., its eigenvalues have nonnegative real part.

These properties enable convergence of the classical successive overrelaxation (SOR) method [20]. To improve efficiency, we modify the classical SOR method and propose a generalized version that extends the GSOR method considered in [1].

By introducing the three matrices

$$\mathcal{D} = \begin{pmatrix} A & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & D \end{pmatrix}, \quad \mathcal{L} = \begin{pmatrix} 0 & 0 & 0 \\ B & 0 & 0 \\ C & 0 & 0 \end{pmatrix}, \quad \mathcal{U} = \begin{pmatrix} 0 & -B^T & -C^T \\ 0 & P & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5}$$

where $P \in \mathbb{R}^{m\times m}$ is SPD, we can split $\widehat{\mathcal{A}}$ as

$$\widehat{\mathcal{A}} = \mathcal{D} - \mathcal{L} - \mathcal{U}.$$

Let ω, τ, and θ be three nonzero reals, I_n, I_m, and I_p be identity matrices of appropriate order, and

$$\Omega = \begin{pmatrix} \omega I_n & 0 & 0 \\ 0 & \tau I_m & 0 \\ 0 & 0 & \theta I_p \end{pmatrix}.$$

Consider the following iteration for (1):

$$w_{k+1} = (\mathcal{D} - \Omega\mathcal{L})^{-1}[(I - \Omega)\mathcal{D} + \Omega\mathcal{U}]w_k + (\mathcal{D} - \Omega\mathcal{L})^{-1}\Omega\widehat{b} \tag{6}$$

$$=: \mathcal{T} w_k + (\mathcal{D} - \Omega\mathcal{L})^{-1}\Omega\widehat{b}.$$

From (5),

$$\mathcal{D} - \Omega\mathcal{L} = \begin{pmatrix} A & 0 & 0 \\ -\tau B & P & 0 \\ -\theta C & 0 & D \end{pmatrix}, \tag{7}$$

$$(I - \Omega)\mathcal{D} + \Omega\mathcal{U} = \begin{pmatrix} (1-\omega)A & -\omega B^T & -\omega C^T \\ 0 & P & 0 \\ 0 & 0 & (1-\theta)D \end{pmatrix}. \tag{8}$$

Substituting (7) and (8) into (6), we obtain GSOR as stated in Algorithm 1.

Algorithm 1 The GSOR method

1: Choose $(x_0, y_0, z_0) \in \mathbb{R}^{n+m+p}$, $P \in \mathbb{R}^{m\times m}$ SPD, and ω, τ, $\theta > 0$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Compute $(x_{k+1}, y_{k+1}, z_{k+1})$ according to the iteration

$$\begin{cases} x_{k+1} = x_k + \omega A^{-1}(f - Ax_k - B^T y_k - C^T z_k), \\ y_{k+1} = y_k + \tau P^{-1}(Bx_{k+1} - g), \\ z_{k+1} = z_k + \theta D^{-1}(Cx_{k+1} - Dz_k - h). \end{cases} \tag{2.7}$$

4: **end for**

At each step, GSOR needs to solve only three SPD systems (of order n, m, and p). This is easier than in GBSOR [4], which solves four linear systems involving A, A, $BA^{-1}B^T$, and D.

Iteration scheme (2.7) can also be deduced from the splitting

$$\mathcal{A} = \mathcal{M} - \mathcal{N} := \begin{pmatrix} \frac{1}{\omega}A & 0 & 0 \\ B & -\frac{1}{\tau}P & 0 \\ C & 0 & -\frac{1}{\theta}D \end{pmatrix} - \begin{pmatrix} (\frac{1}{\omega} - 1)A & -B^T & -C^T \\ 0 & -\frac{1}{\tau}P & 0 \\ 0 & 0 & (1 - \frac{1}{\theta})D \end{pmatrix}. \tag{2.6}$$

Therefore, GSOR is a splitting method. In particular if $\omega = 1$, GSOR reduces to the Uzawa-like schemes studied in [4], where $M = -\frac{1}{\theta}D$ and $N = \left(1 - \frac{1}{\theta}\right) D$ in (3).

3 Convergence analysis for GSOR

First, the following two lemmas give readily verifiable necessary and sufficient conditions for all roots of a real polynomial of degree two or three, respectively, to have modulus less than one.

Lemma 3.1. [11, Theorem 1.3] Consider the second-degree polynomial equation

$$\lambda^2 + a_1\lambda + a_0 = 0, \tag{9}$$

where a_0 and a_1 are real numbers. A necessary and sufficient condition for both roots of (9) to lie in the open disk $|\lambda| < 1$ is

$$|a_1| < 1 + a_0 < 2.$$

.....

Lemma 3.2. [11, Theorem 1.4] Consider the third-degree polynomial equation

$$\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0, \tag{10}$$

where a_0, a_1, and a_2 are real numbers. A necessary and sufficient condition for all roots of (10) to lie in the open disk $|\lambda| < 1$ is

$$|a_2 + a_0| < 1 + a_1, \qquad |a_2 - 3a_0| < 3 - a_1, \qquad a_0^2 + a_1 - a_0 a_2 < 1.$$

GSOR is convergent if and only if the spectral radius of $\mathcal{T}$ is less than 1; i.e., $\rho(\mathcal{T}) = \rho((\mathcal{D} - \Omega\mathcal{L})^{-1}[(I-\Omega)\mathcal{D} + \Omega\mathcal{U}]) < 1$. From this point of view, we can now study the convergence properties of GSOR.

Theorem 3.1. Assume that A and D are SPD, and that B has full row rank. Let the maximum eigenvalues of $A^{-1}B^TP^{-1}B$ and $A^{-1}C^TD^{-1}C$ be $\mu_{\max}$ and $\nu_{\max}$, respectively. Then GSOR is convergent if $0 < \theta < 2$ and

$$0 < \omega < \frac{4(2-\theta)}{(2-\theta)(2+\tau\mu_{\max}) + 2\theta\nu_{\max}}, \qquad 0 < \tau < \frac{4(\omega+\theta-\omega\theta)}{\omega\theta\mu_{\max}}.$$

Proof. Let λ and $v = (x, y, z)$ be an eigenvalue and eigenvector of $\mathcal{T}$. Then

$$[(I-\Omega)\mathcal{D} + \Omega\mathcal{U}]v = \lambda(\mathcal{D} - \Omega\mathcal{L})v.$$

Substituting (7) and (8) gives

$$(1-\omega)Ax - \omega B^Ty - \omega C^Tz = \lambda Ax, \tag{11}$$

$$Py = -\lambda\tau Bx + \lambda Py, \tag{12}$$

$$(1-\theta)Dz = -\lambda\theta Cx + \lambda Dz. \tag{13}$$

To continue the proof, we consider sufficient conditions to guarantee $|\lambda| < 1$.

If $\lambda = 1-\theta$, we have $|\lambda| < 1$ if and only if $0 < \theta < 2$. In the following, we assume that $\lambda \neq 1-\theta$ and $0 < \theta < 2$.

Note that we must have $\lambda \neq 1$; otherwise, (11)–(13) give

$$Ax + B^Ty + C^Tz = 0, \qquad Bx = 0, \qquad Dz = Cx, \tag{14}$$

which imply that $0 = x^TAx + x^TB^Ty + x^TC^Tz = x^TAx + z^TDz$. Given that both A and D are SPD, we have $x^TAx = 0$ and $z^TDz = 0$, which further means $x = 0$ and $z = 0$. With B having full row rank, the first equality in (14) gives $y = 0$. This contradicts the fact that v is an eigenvector.

It follows from $\lambda \neq 1$, $\lambda \neq 1-\theta$, and (12)–(13) that

$$y = \frac{\lambda\tau}{\lambda-1}P^{-1}Bx, \qquad z = \frac{\lambda\theta}{\lambda+\theta-1}D^{-1}Cx.$$

Substituting these into (11), we obtain

$$(1-\omega)Ax - \frac{\lambda\omega\tau}{\lambda-1}B^TP^{-1}Bx - \frac{\lambda\omega\theta}{\lambda+\theta-1}C^TD^{-1}Cx = \lambda Ax.$$

Note that $x \neq 0$; otherwise, we have $Py = 0$ and $Dz = 0$, which implies $v = 0$ because P and D are SPD. This contradicts the fact that v is an eigenvector. Therefore, premultiplying both sides by $x^*/(x^*Ax)$ gives

$$1 - \omega - \frac{\lambda\omega\tau}{\lambda-1}\phi(x) - \frac{\lambda\omega\theta}{\lambda+\theta-1}\varphi(x) = \lambda, \tag{15}$$

where

$$\phi(x) = \frac{x^*B^TP^{-1}Bx}{x^*Ax}, \qquad \varphi(x) = \frac{x^*C^TD^{-1}Cx}{x^*Ax}.$$

First, we consider the case $x \in \text{null}(B)$. Clearly, $\phi(x) = 0$. If $x \in \text{null}(C)$ as well, it follows from (15) that $\lambda = 1-\omega$. To guarantee $|\lambda| < 1$, we can assume that $0 < \omega < 2$. If $x \notin \text{null}(C)$, with $\phi(x) = 0$, (15) reduces to the quadratic polynomial equation

$$\lambda^2 + (\omega + \theta + \omega\theta\varphi(x) - 2)\,\lambda + 1 + \omega\theta - \omega - \theta = 0.$$

By Lemma 3.1, both roots λ of this real quadratic equation satisfy $|\lambda| < 1$ if and only if

$$|\omega + \theta + \omega\theta\varphi(x) - 2| < 2 + \omega\theta - \omega - \theta < 2.$$

If $0 < \omega < 2$ and $0 < \theta < 2$, we have $\omega + \theta \geq 2\sqrt{\omega\theta} > \omega\theta \Rightarrow \omega + \theta - \omega\theta > 0$. This, along with $\varphi(x) \geq 0$ yields

$$0 < \omega < \frac{2(2-\theta)}{2 - \theta + \theta\varphi(x)} \leq 2\,. \tag{16}$$

Next, we consider the case $x \notin \text{null}(B)$. Then $\phi(x) > 0$ and (15) can be rewritten as a cubic polynomial equation $\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0$, where

$$\begin{aligned} a_2 &= \theta + \omega + \omega\tau\phi(x) + \omega\theta\varphi(x) - 3, \\ a_1 &= 3 + \omega\theta - 2\omega - 2\theta - \omega\tau(1-\theta)\phi(x) - \omega\theta\varphi(x), \\ a_0 &= \omega + \theta - \omega\theta - 1. \end{aligned}$$

By Lemma 3.2, all roots λ of the above real cubic equation satisfy $|\lambda| < 1$ if and only if

$$\left|\, 2\omega + 2\theta - \omega\theta + \omega\tau\phi(x) + \omega\theta\varphi(x) - 4 \,\right| < 4 + \omega\theta - 2\omega - 2\theta - \omega\tau(1-\theta)\phi(x) - \omega\theta\varphi(x), \tag{17}$$

$$\left|\, \omega\tau\phi(x) + \omega\theta\varphi(x) - 2\omega - 2\theta + 3\omega\theta \,\right| < 2\omega + 2\theta - \omega\theta + \omega\tau(1-\theta)\phi(x) + \omega\theta\varphi(x), \tag{18}$$

$$\theta(\omega\theta - \omega - \theta)\,(1 + \varphi(x)) - \omega\tau(1-\theta)\phi(x) < 0. \tag{19}$$

Note that with $\phi(x) > 0$, $\varphi(x) \geq 0$, and $0 < \theta < 2$, (17) leads to $\tau > 0$ and

$$0 < \omega < \frac{4(2-\theta)}{4 - 2\theta + \tau(2-\theta)\phi(x) + 2\theta\varphi(x)} \leq 2. \tag{20}$$

It follows from (18) that

$$\omega\tau\theta\phi(x) < 4(\omega + \theta - \omega\theta), \tag{21}$$

and

$$\omega\tau(2-\theta)\phi(x) + 2\omega\theta\varphi(x) + 2\omega\theta > 0. \tag{22}$$

If $\omega > 0$, $\tau > 0$ and $0 < \theta < 2$, as $\phi(x) > 0$ and $\varphi(x) \geq 0$, clearly (22) holds. This together with (21) implies that (18) holds if

$$0 < \tau < \frac{4(\omega + \theta - \omega\theta)}{\omega\theta\phi(x)}. \tag{23}$$

Inequality (19) holds if $0 < \theta \leq 1$ because ω, $\tau > 0$. If $1 < \theta < 2$ and $0 < \omega < 2$, solving (19) leads to

$$\tau < \frac{\theta(\omega + \theta - \omega\theta)\,(1 + \varphi(x))}{\omega(\theta - 1)\phi(x)}\,.$$

Note that $\omega > 0$, $\phi(x) > 0$, $\omega + \theta - \omega\theta > 0$ and $4(\theta - 1) < \theta^2(1 + \varphi(x))$, giving

$$\frac{4(\omega + \theta - \omega\theta)}{\omega\theta\phi(x)} < \frac{\theta(\omega + \theta - \omega\theta)\,(1 + \varphi(x))}{\omega(\theta - 1)\phi(x)}.$$

This implies that (19) holds under condition (23).

To sum up, by combining (16), (20), (23) and the fact that $\tau(2-\theta)\phi(x) \geq 0$, we know that $|\lambda| < 1$ if

$$0 < \theta < 2, \tag{24}$$

$$0 < \omega < \frac{4(2-\theta)}{4 - 2\theta + \tau(2-\theta)\phi(x) + 2\theta\varphi(x)}, \tag{25}$$

$$0 < \tau < \frac{4(\omega + \theta - \omega\theta)}{\omega\theta\phi(x)}. \tag{26}$$

For any $x \neq 0$, we have $0 \leq \phi(x) \leq \mu_{\max}$ and $0 \leq \varphi(x) \leq \nu_{\max}$. Combining with (24)–(26) completes the proof. □

Remark 1. We emphasize that parameters ω, τ and θ can be chosen to satisfy the conditions derived in Theorem 3.1. Indeed, as $0 < \omega < 2$ and $0 < \theta < 2$, we get

$$\frac{4(\omega + \theta - \omega\theta)}{\omega\theta\mu_{\max}} = \frac{4}{\theta\mu_{\max}}\left(1 + \frac{\theta}{\omega} - \theta\right) > \frac{4}{\theta\mu_{\max}}\left(1 + \frac{\theta}{2} - \theta\right) = \frac{2(2-\theta)}{\theta\mu_{\max}}.$$

Thus, we can first choose θ satisfying $0 < \theta < 2$, and then choose τ in the open interval $\left(0, \frac{2(2-\theta)}{\theta\mu_{\max}}\right)$. Finally, we choose ω satisfying

$$0 < \omega < \frac{4(2-\theta)}{(2-\theta)(2+\tau\mu_{\max}) + 2\theta\nu_{\max}}.$$

Remark 2. If $\omega = 1$, (15) can be simplified as

$$\lambda^2 + (\theta - 2 + \tau\phi(x) + \theta\varphi(x))\lambda + 1 - \theta - \tau\phi(x) + \tau\theta\phi(x) - \theta\varphi(x) = 0.$$

It follows from Lemma 3.1 that $|\lambda| < 1$ holds if and only if

$$|\theta - 2 + \tau\phi(x) + \theta\varphi(x)| < 2 - \theta - \tau\phi(x) + \tau\theta\phi(x) - \theta\varphi(x) < 2.$$

After some algebra, we see that GSOR with $\omega = 1$ is convergent if

$$0 < \theta < \frac{2}{1+\nu_{\max}}, \qquad 0 < \tau < \frac{2(2-\theta-\theta\nu_{\max})}{(2-\theta)\mu_{\max}}.$$

Remark 3. If $\omega = 1$ and $\theta = 1$, GSOR is the same as the Uzawa-like method studied in [4, section 2.2]. In this case, (15) reduces to

$$\lambda^2 + (\tau\phi(x) + \varphi(x) - 1)\,\lambda - \varphi(x) = 0.$$

By Lemma 3.1, we know that $|\lambda| < 1$ holds if and only if

$$\left|\tau\phi(x) + \varphi(x) - 1\right| < 1 - \varphi(x) < 2.$$

This implies that, for any τ, GSOR diverges when $\nu_{\max} \geq 1$. Therefore, for this special case, GSOR is convergent provided

$$\nu_{\max} < 1, \qquad 0 < \tau < \frac{2(1-\nu_{\max})}{\mu_{\max}}.$$

This result is the same as [4, Theorem 3], which is the convergence theorem of the Uzawa-like method. However, we emphasize that the condition $\nu_{\max} < 1$ is strong. In fact, as shown in Section 5.2 below, the saddle-point problems from the mixed Stokes-Darcy model in porous media applications do not satisfy this condition. With Remark 2, this shows that it is necessary to introduce another parameter.

4 The GSOR preconditioner

We develop and analyze a class of block lower triangular preconditioners to accelerate Krylov methods for (1).

The splitting in (2.6) can induce a preconditioner $\mathcal{M}$ for (1). The corresponding preconditioned matrix $\mathcal{M}^{-1}\mathcal{A}$ has the form

$$\begin{pmatrix} \omega I & \omega A^{-1}B^T & \omega A^{-1}C^T \\ (\omega-1)\tau P^{-1}B & \omega\tau P^{-1}BA^{-1}B^T & \omega\tau P^{-1}BA^{-1}C^T \\ (\omega-1)\theta D^{-1}C & \omega\theta D^{-1}CA^{-1}B^T & \theta I + \omega\theta D^{-1}CA^{-1}C^T \end{pmatrix}.$$

When $\omega = 1$, $\mathcal{M}^{-1}\mathcal{A}$ has at least n eigenvalues equal to 1. As clustered eigenvalues are desirable, we consider the block lower triangular preconditioner

$$\mathcal{P} = \begin{pmatrix} A & 0 & 0 \\ B & -\frac{1}{\tau}P & 0 \\ C & 0 & -\frac{1}{\theta}D \end{pmatrix}.$$

When $\mathcal{P}$ is used to precondition Krylov subspace methods, each step needs to solve three linear systems involving A, P, and D. This is more practical than the block preconditioners of [2, 3] and the (relaxed) APSS preconditioners of [16], which need to solve several dense linear systems involving matrices like $BA^{-1}B^T$, $D + CA^{-1}C^T$, $A + B^TB/\alpha$, and $D + CC^T/\alpha$, where α is a positive number.

To illustrate further the efficiency of our preconditioner $\mathcal{P}$, we derive explicit and sharp bounds on the spectrum of the preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$. By direct calculations, we have

$$\mathcal{P}^{-1}\mathcal{A} = \begin{pmatrix} I & A^{-1}B^T & A^{-1}C^T \\ 0 & \tau P^{-1}BA^{-1}B^T & \tau P^{-1}BA^{-1}C^T \\ 0 & \theta D^{-1}CA^{-1}B^T & \theta I + \theta D^{-1}CA^{-1}C^T \end{pmatrix},$$

which is similar to

$$\begin{pmatrix} I & \hat{B}^T & \hat{C}^T \\ 0 & \tau\hat{B}\hat{B}^T & \tau\hat{B}\hat{C}^T \\ 0 & \theta\hat{C}\hat{B}^T & \theta I + \theta\hat{C}\hat{C}^T \end{pmatrix},$$

where $\hat{B} = P^{-1/2}BA^{-1/2}$ and $\hat{C} = D^{-1/2}CA^{-1/2}$. Thus $\mathcal{P}^{-1}\mathcal{A}$ has eigenvalue 1 with multiplicity n, and the remaining eigenvalues are the same as those of

$$K = \begin{pmatrix} \tau\hat{B}\hat{B}^T & \tau\hat{B}\hat{C}^T \\ \theta\hat{C}\hat{B}^T & \theta I + \theta\hat{C}\hat{C}^T \end{pmatrix}.$$

We can now establish the following theorem.

Theorem 4.1. Assume that A and D are SPD, and B has full row rank. Let the minimum and maximum eigenvalues of $P^{-1}BA^{-1}B^T$ be $\mu_{\min}$ and $\mu_{\max}$. Let the maximum eigenvalue of $D^{-1}CA^{-1}C^T$ be $\nu_{\max}$. Then $\mathcal{P}^{-1}\mathcal{A}$ has eigenvalue 1 with multiplicity at least n, and the remaining eigenvalues lie in the interval

$$\left[\frac{\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}}{2}, \ \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2} \right],$$

where

$$\underline{\Lambda} = \theta(1 + \nu_{\max}) + \tau\mu_{\min}, \qquad \overline{\Lambda} = \theta(1 + \nu_{\max}) + \tau\mu_{\max}. \tag{27}$$

Proof. We need to estimate spectral bounds for K. Let λ be an eigenvalue of K and $(y^T, z^T)^T$ be a corresponding eigenvector. With $\tau > 0$ and $\theta > 0$ we see that K is similar to a symmetric matrix, and hence λ is real. Also,

$$\tau\hat{B}\hat{B}^Ty + \tau\hat{B}\hat{C}^Tz = \lambda y, \tag{28}$$

$$\theta\hat{C}\hat{B}^Ty + \theta z + \theta\hat{C}\hat{C}^Tz = \lambda z. \tag{29}$$

We obtain estimates of λ by considering two cases separately.

Case I: $z \in \text{null}(\hat{C}^T)$. Clearly, $\tau\hat{B}\hat{B}^Ty = \lambda y$ and $\theta\hat{C}\hat{B}^Ty = (\lambda - \theta)z$. This implies that $\lambda = \theta$ or λ is an eigenvalue of $\hat{B}\hat{B}^T$. Note that $\hat{B}\hat{B}^T = P^{-1/2}BA^{-1}B^TP^{-1/2}$ is similar to $P^{-1}BA^{-1}B^T$, so that $\lambda = \theta$ or $\tau\mu_{\min} \le \lambda \le \tau\mu_{\max}$.

Case II: $z \notin \text{null}(\hat{C}^T)$. We only consider the case $\lambda \notin [\tau\mu_{\min}, \tau\mu_{\max}]$, so that $\lambda I - \tau\hat{B}\hat{B}^T$ is nonsingular. With (28), this leads to $y = \tau(\lambda I - \tau\hat{B}\hat{B}^T)^{-1}\hat{B}\hat{C}^T z$. Substituting into (29) gives

$$\tau\theta\hat{C}\hat{B}^T(\lambda I - \tau\hat{B}\hat{B}^T)^{-1}\hat{B}\hat{C}^T z + \theta z + \theta\hat{C}\hat{C}^T z = \lambda z. \tag{30}$$

As $\left(I - \frac{\tau}{\lambda}\hat{B}^T\hat{B}\right)^{-1} = I + \tau\hat{B}^T(\lambda I - \tau\hat{B}\hat{B}^T)^{-1}\hat{B}$, (30) yields

$$\theta\hat{C}\left(I - \frac{\tau}{\lambda}\hat{B}^T\hat{B}\right)^{-1}\hat{C}^T z = (\lambda - \theta)z. \tag{31}$$

We assert that $\lambda > 0$. Otherwise, we have

$$(\lambda - \theta)z^T z < 0 \qquad \text{and} \qquad z^T\hat{C}\left(I - \frac{\tau}{\lambda}\hat{B}^T\hat{B}\right)^{-1}\hat{C}^T z \geq 0,$$

which contradicts (31).

If $\lambda > \tau\mu_{\max}$, as the matrices $\hat{B}^T\hat{B}$ and $\hat{B}\hat{B}^T$ have the same nonzero eigenvalues, it holds for any $0 \neq u \in \mathbb{R}^n$ that

$$u^T\left(I - \frac{\tau}{\lambda}\hat{B}^T\hat{B}\right)u \geq \left(1 - \frac{\tau\mu_{\max}}{\lambda}\right)u^T u > 0.$$

With (31) and the fact that $\hat{C}\hat{C}^T = D^{-1/2}CA^{-1}C^T D^{-1/2}$ is similar to $D^{-1}CA^{-1}C^T$, this leads to

$$\lambda - \theta \leq \theta\left(1 - \frac{\tau\mu_{\max}}{\lambda}\right)^{-1}\frac{z^T\hat{C}\hat{C}^T z}{z^T z} \leq \theta\left(1 - \frac{\tau\mu_{\max}}{\lambda}\right)^{-1}\nu_{\max}.$$

Solving this inequality for λ gives

$$\frac{\overline{\Lambda} - \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2} \leq \lambda \leq \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2}.$$

We can directly check that

$$\overline{\Lambda} - \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}} \leq 2\max\{\theta, \tau\mu_{\max}\} \leq \overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}.$$

Therefore, λ admits the upper bound

$$\lambda \leq \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2}. \tag{32}$$

If $\lambda < \tau\mu_{\min}$, it can be verified that

$$z^T\hat{C}\left(I - \frac{\tau}{\lambda}\hat{B}^T\hat{B}\right)^{-1}\hat{C}^T z \geq \left(1 - \frac{\tau\mu_{\min}}{\lambda}\right)^{-1}z^T\hat{C}\hat{C}^T z \geq \left(1 - \frac{\tau\mu_{\min}}{\lambda}\right)^{-1}\nu_{\max}z^T z.$$

Combining with (31) gives

$$\lambda - \theta \geq \left(1 - \frac{\tau\mu_{\min}}{\lambda}\right)^{-1}\theta\nu_{\max} = \frac{\lambda\theta\nu_{\max}}{\lambda - \tau\mu_{\min}}.$$

Note that $\lambda < \tau\mu_{\min}$ and the inequality can be simplified as

$$\lambda^2 - (\theta + \theta\nu_{\max} + \tau\mu_{\min})\lambda + \tau\theta\mu_{\min} \leq 0.$$

By the definition of $\underline{\Lambda}$, we have

$$\frac{\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}}{2} \leq \lambda \leq \frac{\underline{\Lambda} + \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}}{2}.$$

Similarly, we can check that

$$\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}} \le 2\min\{\theta, \tau\mu_{\min}\} \le \underline{\Lambda} + \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}.$$

This implies that λ admits the lower bound

$$\tau\mu_{\min} > \lambda \ge \frac{\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}}{2}.$$

Combining with (32) and the bounds derived in Case I completes the proof. □

Remark 4. To precondition the equivalent unsymmetric system (4), we can use the block lower triangular preconditioner

$$\widehat{\mathcal{P}} = \begin{pmatrix} A & 0 & 0 \\ -B & \frac{1}{\tau}P & 0 \\ -C & 0 & \frac{1}{\theta}D \end{pmatrix}.$$

Because $\widehat{\mathcal{P}}^{-1}\widehat{\mathcal{A}} = \mathcal{P}^{-1}\mathcal{A}$, the preconditioned matrix $\widehat{\mathcal{P}}^{-1}\widehat{\mathcal{A}}$ possesses the same spectral bounds as in Theorem 4.1.

Remark 5. Theorem 4.1 shows that the preconditioned matrices $\mathcal{P}^{-1}\mathcal{A} = \widehat{\mathcal{P}}^{-1}\widehat{\mathcal{A}}$ are positive stable. Moreover, their condition number is bounded by

$$\max\left\{ \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2}, \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}} \right\}.$$

Using (27), we obtain

$$\frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{2} \le \overline{\Lambda} = \theta(1+\nu_{\max}) + \tau\mu_{\min}$$

and

$$\begin{aligned} \frac{\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}}{\underline{\Lambda} - \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}} &= \frac{\left(\overline{\Lambda} + \sqrt{\overline{\Lambda}^2 - 4\tau\theta\mu_{\max}}\right)\left(\underline{\Lambda} + \sqrt{\underline{\Lambda}^2 - 4\tau\theta\mu_{\min}}\right)}{4\tau\theta\mu_{\min}} \le \frac{\underline{\Lambda}\overline{\Lambda}}{\tau\theta\mu_{\min}} \\ &= \frac{\theta^2(1+\nu_{\max})^2 + \tau^2\mu_{\min}\mu_{\max} + \tau\theta(1+\nu_{\max})(\mu_{\max}+\mu_{\min})}{\tau\theta\mu_{\min}} \\ &= \frac{\theta}{\tau}\frac{(1+\nu_{\max})^2}{\mu_{\min}} + \frac{\tau}{\theta}\mu_{\max} + (1+\nu_{\max})\left(1 + \frac{\mu_{\max}}{\mu_{\min}}\right). \end{aligned}$$

This shows that the matrices $\mathcal{P}^{-1}\mathcal{A}$ and $\widehat{\mathcal{P}}^{-1}\widehat{\mathcal{A}}$ will be well-conditioned given appropriate selections of parameters τ, θ and matrix P when $\nu_{\max}$ is not too large.[1]

5 Numerical experiments

We present the results of numerical tests to examine the feasibility and effectiveness of GSOR. All experiments were run using MATLAB R2015b on a PC with an Intel(R) Core(TM) i7-8550U CPU @ 1.8GHz and 16GB of RAM. The initial guess is taken to be the zero vector, and the algorithms are terminated when the number of iterations exceeds 10^5 or

$$\text{Res} := \|b - \mathcal{A}w_k\|_2/\|b\| \le 10^{-8},$$

[1] This is a reasonable request. As shown in Section 5 below, $\nu_{\max}$ of the saddle-point systems from the liquid crystal directors model and the mixed Stokes-Darcy model in porous media applications is 0.1750 and 1.0057, respectively.

where w_k is the current approximate solution. We report the number of iterations, the CPU time, and the final value of the relative residual, denoted by "Iter", "CPU" and "Res", respectively.

For our GSOR method, we tried just a few values of the parameters ω, τ and θ. We compared our method with the Uzawa-like method (denoted "Uzawa") and the generalization of the block SOR method (denoted "GBSOR") studied in [4, Section 2.2 and Section 3], respectively. We emphasize that the Uzawa method is a special case of our GSOR method with $P = Q$, $\omega = 1$, and $\theta = 1$. For GBSOR, based on [4, Theorem 5], we chose $\omega = s/4$, $s/2$, $3s/4$ (denoted "GBSORa", "GBSORb", "GBSORc", respectively), where $s = 2/(1 + \sqrt{\nu_{\max}})$ is the upper bound of the convergence interval for the parameter ω. We used the function "eigs" to compute $\nu_{\max}$.

We also tested Krylov methods for (1) or (4), such as MINRES, GMRES, and BICGSTAB. For preconditioned MINRES (denoted "BPMINRES"), we use the block diagonal preconditioner

$$\begin{pmatrix} A & 0 & 0 \\ 0 & BA^{-1}B^T & 0 \\ 0 & 0 & D + CA^{-1}C^T \end{pmatrix}.$$

For $D = 0$, this block diagonal preconditioner has been studied in [3]. For preconditioned GMRES, we test the GSOR preconditioner $\mathcal{P}$ with $\tau = \theta = 1$ (denoted "GPGMRES") and the block triangular preconditioner [3] (denoted "BPGMRES")

$$\begin{pmatrix} A & B^T & C^T \\ 0 & -BA^{-1}B^T & 0 \\ 0 & 0 & -(D + CA^{-1}C^T) \end{pmatrix}.$$

5.1 Saddle-point systems from the liquid crystal directors model

Continuum models for the orientational properties of liquid crystals require minimization of free energy functionals of the form

$$\mathcal{F}[u, v, w, U] = \frac{1}{2} \int_0^1 [(u_z^2 + v_z^2 + w_z^2) - \eta^2(\beta + w^2)U_z^2] \mathrm{d}z, \tag{33}$$

where u, v, w, and U are functions of $z \in [0, 1]$ subject to suitable end-point conditions, u_z, v_z, w_z, and U_z denote the first derivatives of the corresponding functions with respect to z, and η and β are prescribed positive parameters. By discretizing with a uniform piecewise-linear finite element scheme with $N + 1$ cells using nodal quadrature and the prescribed boundary conditions, we minimize the free energy (33) under the unit vector constraint. We apply the Lagrange multiplier method to solve this discretized minimization model, and Newton's method to solve the nonlinear equations from the first-order conditions of the Lagrangian. Each step involves the solution of a linear system of the form (1) with $n = 3N$ and $m = p = N$. For more details, we refer to [17].

In our numerical experiments we set $\eta = \sqrt{3}\pi/4$ and $\beta = 0.5$, which is known as the critical switching value. The discretized matrix A is tridiagonal, so in all algorithms we solve systems $Ax = r$ directly by the function "\", which uses a tridiagonal solver. We set $P = BA^{-1}B^T$ and solve systems $Py = r$ using Cholesky factorization. Numerical results are listed in Tables 1 and 2 with $N = 1023, 2047, 4095, 8191, 16383$, where the parameter choices for GSOR and the corresponding notation are as follows:

Method	GSORa	GSORb	GSORc	GSORd
(ω, τ, θ)	$(1, 1, 1)$	$(0.95, 0.95, 0.95)$	$(0.9, 0.8, 1)$	$(0.95, 1, 0.95)$

For this problem, $A - C^T D^{-1} C$ is SPD, which guarantees convergence of the Uzawa-like method [4, Theorem 3]. We set $Q = BA^{-1}B^T$ and $\alpha = 1 - \nu_{\max}$, where $\nu_{\max} = 0.1750$ is the maximum eigenvalue of $A^{-1}C^T D^{-1} C$. MINRES, GMRES and BICGSTAB without preconditioning failed to solve this problem. (For GMRES, we set the restart frequency to 100.) BICGSTAB hit an error condition. Therefore in Table 2 we only report results from preconditioned MINRES and preconditioned GMRES.

Table 1: CPU time for Cholesky factorization of $P = BA^{-1}B^T$.

N	1023	2047	4095	8191	16383
n	3069	6141	12285	24573	49149
m	1023	2047	4095	8191	16383
p	1023	2047	4095	8191	16383
$n+m+p$	5115	10235	20475	40955	81915
$BA^{-1}B^T$	0.086	0.43	1.95	8.93	145.2

To see the role of the parameters in the convergence behavior of GSOR, Figure 1 shows the region of the parameters where GSOR satisfies Res $\leq 10^{-8}$ within $5,000$ iterations, and the characteristic curves of the number of iterations versus the parameters for $N = 1,023$. In Figure 2, we plot the eigenvalue distributions of the original matrix and the GSOR preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$ with different τ and ω.

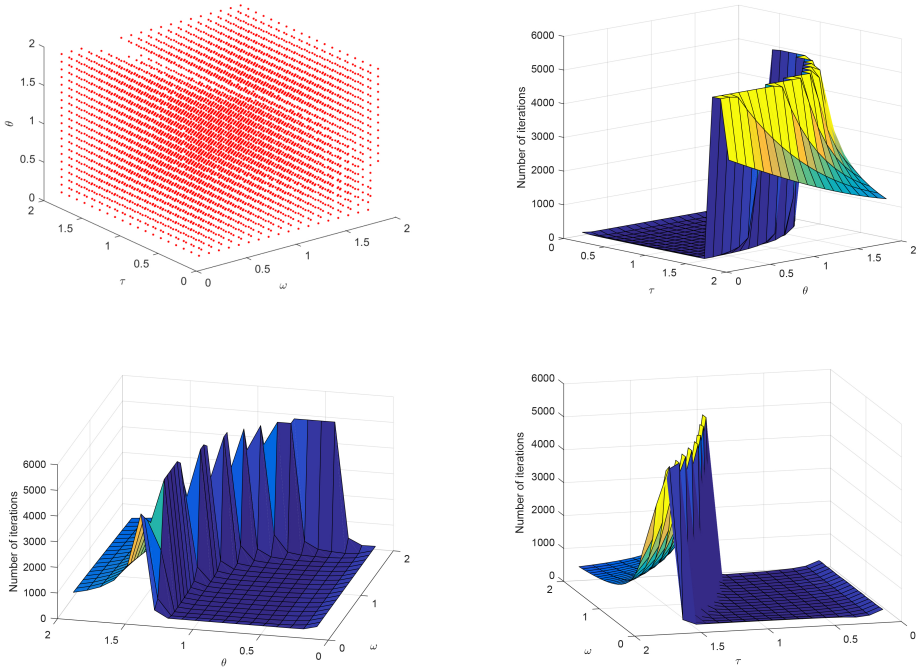


Figure 1: Top left: The region of parameter values for which GSOR satisfies Res $\leq 10^{-8}$ **within** $5,000$ **iterations. Other plots: Characteristic curves for the number of iterations versus parameters ω, τ and θ for GSOR with $\omega = 1$ (top right), $\tau = 1$ (bottom left), and $\theta = 1$ (bottom right). All plots are for saddle-point systems from the liquid crystal directors model with $n = 3069$, $m = p = 1023$.**

5.2 Saddle-point systems from the mixed Stokes-Darcy model in porous media applications

Fluid flow in $\Omega_f \subset \mathbb{R}^2$ coupled with porous media flow in $\Omega_p \subset \mathbb{R}^2$ is governed by the static Stokes equations

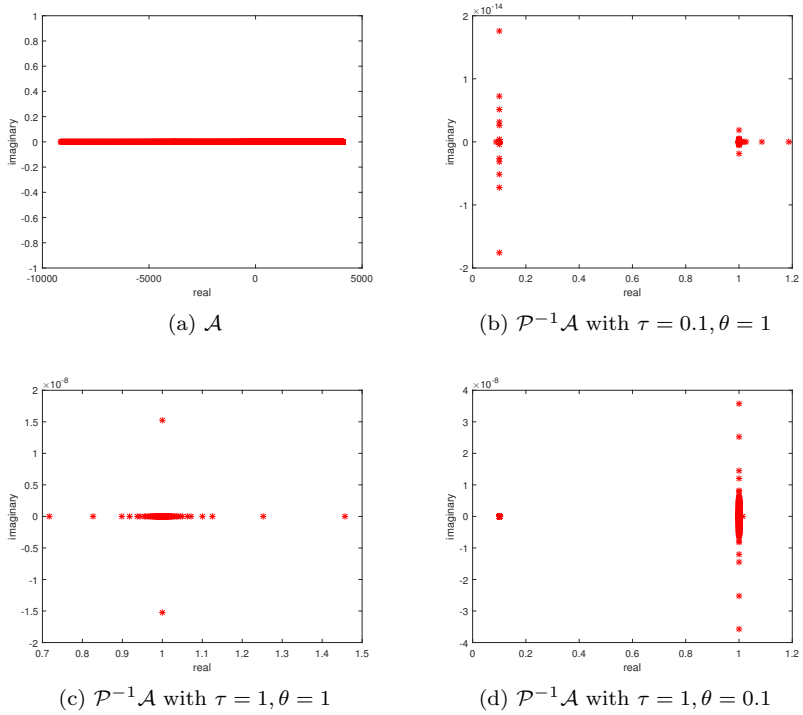
$$-\upsilon\Delta\,\boldsymbol{u}_f + \nabla\, p_f = \boldsymbol{f}, \quad \text{and} \quad \operatorname{div}\boldsymbol{u}_f = 0, \quad \boldsymbol{x} \in \Omega_f, \tag{34}$$

where $\Omega_f \cap \Omega_p = \emptyset$ and $\overline{\Omega}_f \cap \overline{\Omega}_p = \Gamma$ with Γ being an interface, $\upsilon > 0$ is the kinematic viscosity, and $\boldsymbol{f}$ is the external force.

In the porous media region, the governing variable is $\phi = \frac{p_p}{\rho_f g}$, where p_p is the pressure in Ω_p, ρ_f is the fluid density, and g is the gravity acceleration. The velocity $\boldsymbol{u}_p$ of the porous media flow is related to ϕ by Darcy's law and is also divergence free:

$$\boldsymbol{u}_p = -\frac{\epsilon^2}{r\upsilon}\nabla\phi \quad \text{and} \quad -\operatorname{div}\boldsymbol{u}_p = 0, \quad \boldsymbol{x} \in \Omega_p, \tag{35}$$

where r is the volumetric porosity, and ϵ the characteristic length of the porous media.



(a) $\mathcal{A}$ (b) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 0.1, \theta = 1$

(c) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 1, \theta = 1$ (d) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 1, \theta = 0.1$

Figure 2: Eigenvalue distributions of the original matrix and the GSOR preconditioned matrices for saddle-point systems from the liquid crystal directors model with $n = 3069, m = p = 1023$.

Applying finite element discretization to the mixed Stokes-Darcy model (34)–(35) with the Dirichlet boundary conditions leads to linear systems of form (1) [6].

In our numerical experiments, we set $\upsilon = 1$, $r = 1$, and $\epsilon = \sqrt{0.1}$. The computational domain is $\Omega_f = (0,1) \times (1,2)$, $\Omega_p = (0,1) \times (0,1)$ and the interface is $\Gamma = (0,1) \times \{1\}$. We use a uniform mesh with grid parameters $h = 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}$ to decompose Ω_f, P2–P1 elements in the fluid region, and P2 Lagrange elements in the porous media region.

For this problem, P is the pressure mass matrix discretized from the decoupled problem of (34)–(35) [6]. In all algorithms we use Cholesky factorization to solve the systems $Ax = r$, $Py = r$ and $BA^{-1}B^T y = r$. Numerical results for saddle-point systems from the mixed Stokes-Darcy model (34)–(35) are listed in Tables 3 and 4, where the parameters choices for GSOR and the corresponding notation are as follows.

Method	GSORa	GSORb	GSORc	GSORd
(ω, τ, θ)	$(0.5, 1.5, 1.0)$	$(0.5, 1.7, 0.8)$	$(0.5, 1.6, 1.2)$	$(0.6, 1.5, 1.0)$

For this problem, $\nu_{\max} = 1.0057$. The matrix $A - C^T D^{-1} C$ is no longer SPD, so convergence of the Uzawa-like method cannot be guaranteed [4, Theorem 3]. We tested several α ranging from 0.005 to 0.5 for $h = 2^{-3}$. Uzawa failed in all cases. Thus, we do not report results for Uzawa in Table 4. As MINRES and GMRES worked only for systems with $h \geq 2^{-5}$, we again do not report their results.

To see the role of the parameters in the convergence behavior of GSOR, Figure 3 shows the region of parameters for which GSOR satisfies Res $\leq 10^{-8}$ within 5,000 steps, and the characteristic curves of iteration numbers versus parameters for $h = 2^{-3}$. In Figure 4, we plot the eigenvalue distributions of the original matrix and the GSOR preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}$ with different τ and ω.

Table 2: Numerical results for saddle-point systems from the liquid crystal directors model.

	N	1023	2047	4095	8191	16383
GSORa	Iter	24	24	25	25	26
	CPU	0.20	1.08	4.81	23.86	226.04
	Res	5.95e-09	8.42e-09	5.44e-09	7.70e-09	4.96e-09
GSORb	Iter	15	15	16	16	16
	CPU	0.12	0.64	3.03	15.35	152.79
	Res	5.02e-09	7.09e-09	2.53e-09	3.57e-09	5.05e-09
GSORc	Iter	16	17	17	17	17
	CPU	0.14	0.71	3.29	15.28	160.58
	Res	8.41e-09	1.82e-09	2.01e-09	2.34e-09	2.88e-09
GSORd	Iter	14	14	14	14	14
	CPU	0.11	0.59	2.83	12.64	132.68
	Res	7.30e-10	9.62e-10	1.31e-09	1.81e-09	2.53e-09
UZAWA	Iter	18	18	18	20	20
	CPU	0.14	0.76	3.51	18.35	187.64
	Res	4.01e-09	5.67e-09	8.02e-09	1.56e-09	2.22e-09
GBSORa	Iter	72	73	75	76	77
	CPU	0.56	3.12	14.80	70.90	773.60
	Res	8.38e-09	9.23e-09	7.90e-09	8.69e-09	9.57e-09
GBSORb	Iter	29	30	30	31	31
	CPU	0.23	1.31	5.98	28.03	294.17
	Res	7.91e-09	5.95e-09	8.42e-09	6.32e-09	8.95e-09
GBSORc	Iter	35	36	36	37	38
	CPU	0.27	1.59	7.13	34.63	356.66
	Res	7.59e-09	6.34e-09	8.97e-09	7.51e-09	6.30e-09
BPMINRES	Iter	13	13	13	14	14
	CPU	3.50	18.10	96.25	837.49	10736.63
	Res	2.67e-09	4.75e-09	9.35e-09	1.41e-09	1.51e-09
BPGMRES	Iter	8	8	8	8	8
	CPU	3.35	18.63	91.74	543.05	10609.10
	Res	6.42e-09	1.34e-08	9.35e-09	7.74e-08	1.86e-07
GPGMRES	Iter	8	8	8	8	8
	CPU	1.80	8.17	40.53	251.52	3548.65
	Res	9.00e-10	1.92e-09	4.99e-09	1.66e-08	7.79e-08

Table 3: The CPU time of the Cholesky factorization.

h	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}
n	578	2178	8450	33282	132098
m	81	289	1089	4225	16641
p	289	1089	4225	16641	66049
$n+m+p$	948	3556	13764	54148	214788
A	0.0008	0.0048	0.029	0.23	1.75
P	0.0003	0.0004	0.0011	0.02	0.10
$BA^{-1}B^T$	0.0064	0.18	7.18	555.59	31368.15

Tables 1 to 4 and Figures 1 to 4 illustrate that GSOR is a practical method, and its advantages increase with the problem size. We see from Tables 1 to 4 that BPMINRES and BPGMRES are not practical in terms of CPU times. Figures 1 and 3 indicate that the convergence rate of GSOR depends strongly on ω, τ and θ. Figures 2 and 4 show that $\mathcal{P}$ greatly improves the eigenvalue distribution of the original $\mathcal{A}$.

Table 4: Numerical results for saddle-point systems from mixed the Stokes-Darcy model.

	h	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}
GSORa	Iter	50	49	49	47	47
	CPU	0.05	0.15	0.89	10.19	54.22
	Res	5.99e-09	9.80e-09	8.41e-09	9.53e-09	6.85e-09
GSORb	Iter	50	50	50	50	50
	CPU	0.03	0.15	0.91	10.13	64.10
	Res	6.54e-09	5.93e-09	5.51e-09	5.82e-09	5.74e-09
GSORc	Iter	50	50	50	50	49
	CPU	0.04	0.15	0.90	10.02	62.91
	Res	5.98e-09	5.19e-09	6.92e-09	7.00e-09	9.53e-09
GSORd	Iter	48	45	42	39	38
	CPU	0.04	0.14	0.80	7.70	49.98
	Res	8.39e-09	8.53e-09	8.52e-09	9.63e-09	5.26e-09
GBSORa	Iter	158	150	141	132	124
	CPU	0.21	0.84	6.01	90.33	948.46
	Res	9.79e-09	9.16e-09	9.46e-09	9.96e-09	9.77e-09
GBSORb	Iter	73	69	65	61	58
	CPU	0.08	0.36	2.74	41.47	441.51
	Res	9.17e-09	9.18e-09	9.28e-09	9.57e-09	8.28e-09
GBSORc	Iter	44	42	40	37	35
	CPU	0.04	0.19	1.73	25.59	265.67
	Res	9.33e-09	8.25e-09	7.38e-09	9.57e-09	8.61e-09
BICGSTAB	Iter	767.5	1491	2997.5	5912.5	13826.5
	CPU	0.09	0.41	2.23	21.51	288.07
	Res	7.96e-09	7.64e-09	9.57e-09	4.76e-09	9.48e-09
BPMINRES	Iter	18	18	18	17	18
	CPU	0.13	0.71	6.11	216.39	3479.83
	Res	8.34e-09	3.66e-09	1.49e-09	5.70e-09	2.85e-09
BPGMRES	Iter	10	10	10	10	10
	CPU	0.14	0.81	10.00	323.95	3992.25
	Res	1.60e-09	1.56e-09	1.05e-09	7.06e-10	1.39e-09
GPGMRES	Iter	24	24	25	26	27
	CPU	0.14	0.84	3.53	17.34	92.48
	Res	1.06e-09	4.98e-09	6.87e-09	1.53e-09	6.15e-10

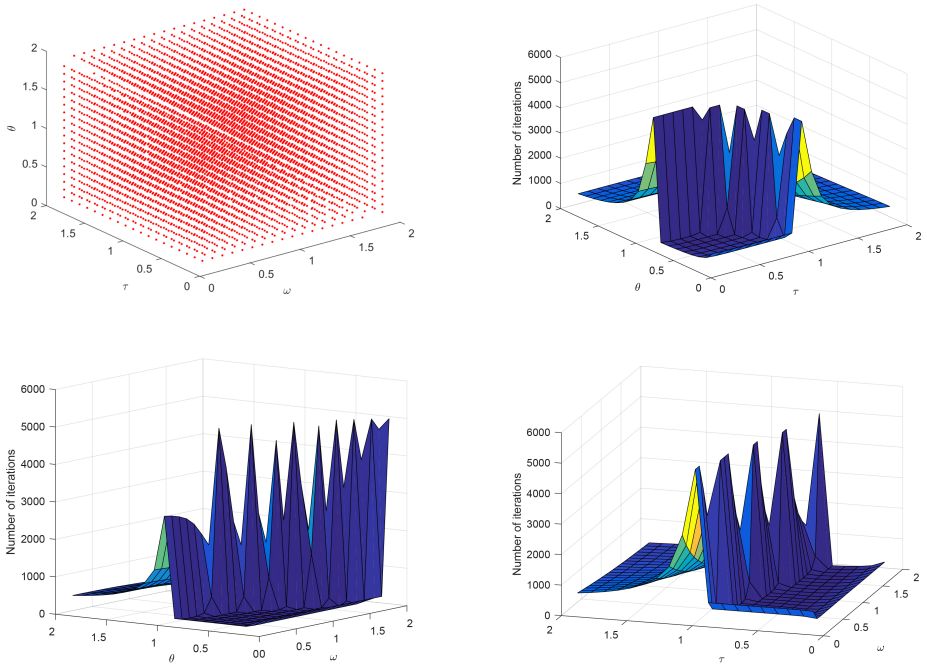
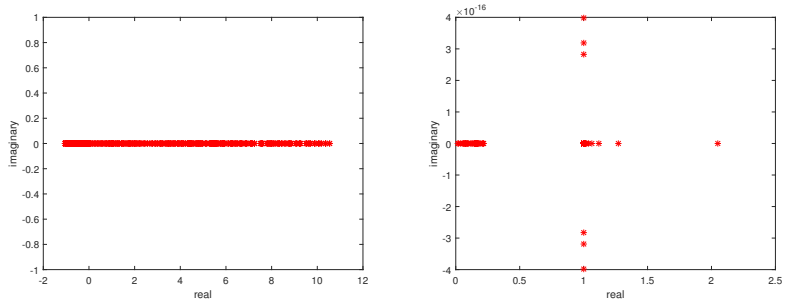
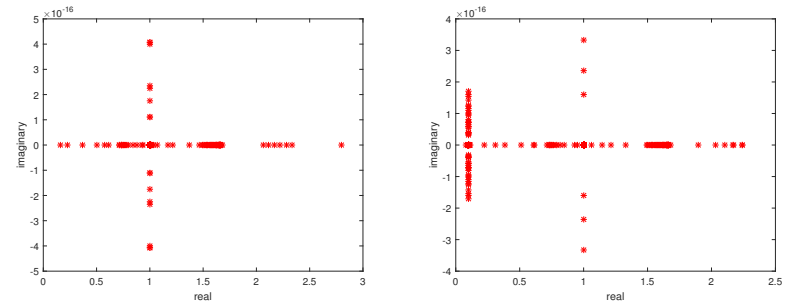


Figure 3: Top left: The region of parameter values for which GSOR satisfies $\mathrm{Res} \leq 10^{-8}$ within $5,000$ iterations. Other plots: Characteristic curves for the number of iterations versus parameters ω, τ and θ for GSOR with $\omega = 1$ (top right), $\tau = 1$ (bottom left), and $\theta = 1$ (bottom right). All plots are for saddle-point systems from the mixed Stokes-Darcy model with $n = 578$, $m = 81$, $p = 289$.



(a) $\mathcal{A}$ (b) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 0.1, \theta = 1$



(c) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 1, \theta = 1$ (d) $\mathcal{P}^{-1}\mathcal{A}$ with $\tau = 1, \theta = 0.1$

Figure 4: Eigenvalue distributions of the original matrix and the GSOR preconditioned matrices for saddle-point systems from the mixed Stokes-Darcy model with $n = 578, m = 81, p = 289$.

6 Conclusions

We presented a theoretical and numerical study of the GSOR method for solving the double saddle-point problem (1). GSOR is convergent with suitable parameters ω, τ, and θ. Unlike existing work, our proof is based on the necessary and sufficient conditions for all roots of a real cubic polynomial to have modulus less than one. We analyzed a class of block lower triangular preconditioners $\mathcal{P}$ induced from GSOR and derived explicit and sharp bounds for the eigenvalues of preconditioned matrices. The numerical results presented are highly encouraging. GSOR requires the least CPU time, and especially for larger problems, its advantages are clear. A shortcoming is the need to choose the three parameters. A practical method to choose them is a topic for future research. 00000000

References

[1] Z. Z. Bai, B. N. Parlett, and Z. Q. Wang, On generalized successive overrelaxation methods for augmented linear systems, Numer. Math., 102 (2005), 1–38.

[2] F. P. A. Beik and M. Benzi, Block preconditioners for saddle point systems arising from liquid crystal directors modeling, Calcolo, 55 (2018), 1–16.

[3] F. P. A. Beik and M. Benzi, Iterative methods for double saddle point systems, SIAM J. Matrix Anal. Appl., 39 (2018), 902–921.

[4] M. Benzi and F. P. A. Beik, Uzawa-type and augmented Lagrangian methods for double saddle point systems, in Structured Matrices in Numerical Linear Algebra, Springer, 2019, 215–236.

[5] M. Benzi, G. H. Golub, and J. Liesen, Numerical solution of saddle point problems, Acta Numerica, 14 (2005), 1–137.

[6] M. C. Cai, M. Mu, and J. C. Xu, Preconditioning techniques for a mixed Stokes/Darcy model in porous media applications, J. Comput. Appl. Math., 233 (2009), 346–355.

[7] Y. Dou and Z. Z. Liang, A class of block alternating splitting implicit iteration methods for double saddle point linear systems, Numer. Linear Algebra Appl., (2022), p. e2455.

[8] A. J. Ellingsrud, Preconditioning unified mixed discretizations of coupled Darcy-Stokes flow, master's thesis, University of Oslo, 2015.

[9] A. Ghannad, D. Orban, and M. A. Saunders, Linear systems arising in interior methods for convex optimization: a symmetric formulation with bounded condition number, Optim. Method Softw., (2021), 1–26, https://doi.org/10.1080/10556788.2021.1965599.

[10] C. Greif, E. Moulding, and D. Orban, Bounds on eigenvalues of matrices arising from interior-point methods, SIAM J. Optim., 24 (2014), 49–83.

[11] E. A. Grove and G. Ladas, Periodicities in nonlinear difference equations, Chapman and Hall/CRC, 2004.

[12] K. E. Holter, M. Kuchta, and K. A. Mardal, Robust preconditioning of monolithically coupled multiphysics problems, arXiv preprint arXiv:2001.05527, (2020).

[13] K. E. Holter, M. Kuchta, and K. A. Mardal, Robust preconditioning for coupled Stokes-Darcy problems with the Darcy problem in primal form, Comput. Math. Appl., 91 (2021), 53–66.

[14] N. Huang, Variable parameter Uzawa method for solving a class of block three-by-three saddle point problems, Numer. Algor., 85 (2020), 1233—1254.

[15] N. Huang, Y. H. Dai, and Q. Y. Hu, Uzawa methods for a class of block three-by-three saddle-point problems, Numer. Linear Algebra Appl., 26 (2019), p. e2265.

[16] Z. Z. Liang and G. F. Zhang, Alternating positive semidefinite splitting preconditioners for double saddle point problems, Calcolo, 56 (2019), 1–17.

[17] A. Ramage and E. C. Gartland Jr, A preconditioned nullspace method for liquid crystal director modeling, SIAM J. Sci. Comput., 35 (2013), B226–B247.

[18] B.-C. Ren, F. Chen, and X.-L. Wang, Improved splitting preconditioner for double saddle point problems arising from liquid crystal director modeling, Numer. Algor., (2022), 1–17.

[19] S. J. Wright, Primal-dual Interior-point Methods, SIAM, Philadelphia, 1997.

[20] D. M. Young, Iterative Solution of Large Linear Systems, Elsevier, 2014.

[21] J.-L. Zhu, Y.-J. Wu, and A.-L. Yang, A two-parameter block triangular preconditioner for double saddle point problem arising from liquid crystal directors modeling, Numer. Algor., 89 (2022), 987–1006.