Block Preconditioners for the Coupled Stokes–Darcy Problem

Michele Benzi
Scuola Normale Superiore, Pisa

SCUOLA
NORMALE
SUPERIORE

Iterative Solution of Large-Scale Saddle-Point Problems
Cortona, 9–20 May, 2022

Gene Howard Golub (1932–2007)

Ghent, Belgium, September 2006 (courtesy of Gérard Meurant)

Acknowledgements

- Joint work with Fatemeh Beik
- Special thanks to Scott Ladenheim for providing the test problems

Note: Full details in

F. P. A. Beik and M. Benzi, *Preconditioning techniques for the coupled Stokes–Darcy problem: spectral and field-of-values analysis*, Numerische Mathematik, 150(2), 2022, pp. 257–298.

Outline

- The continuous problem
- The discrete problem
- Block preconditioners
- Eigenvalue and Field-of-Values analysis
- Numerical experiments
- Conclusions

The coupled Stokes–Darcy system

The coupled Stokes–Darcy model describes the interaction between free flow and porous media flow. It is a fundamental problem in several fields.

In one subregion of the flow domain Ω a free-flowing fluid is governed by the (Navier-)Stokes equations; in another subregion, the fluid follows Darcy's Law.

The equations are coupled by conditions across the interface between the two subregions.

In this talk we will only consider the case of stationary problems and Stokes flow.

The coupled Stokes–Darcy system (cont.)

Let Ω be a computational domain partitioned into two disjoint subdomains Ω_1 and Ω_2, separated by an interface Γ_{12}. We assume that the flow in Ω_1 is governed by the stationary Stokes equations:

$$-\nabla \cdot (2\nu D(\mathbf{u}_1) - p_1 \mathbf{I}) = \mathbf{f}_1 \quad \text{in} \quad \Omega_1,$$

$$\nabla \cdot \mathbf{u}_1 = 0 \quad \text{in} \quad \Omega_1,$$

$$\mathbf{u}_1 = 0 \quad \text{on} \quad \Gamma_1 = \partial\Omega_1 \cap \partial\Omega.$$

Here $\nu > 0$ represents the kinematic viscosity, $\mathbf{u}_1$ and p_1 denote the velocity and pressure in Ω_1, $\mathbf{f}_1$ is an external force acting on the fluid, $\mathbf{I}$ is the identity matrix, and

$$D(\mathbf{u}_1) = \frac{1}{2}\left(\nabla \mathbf{u}_1 + \nabla \mathbf{u}_1^T\right)$$

is the rate of strain tensor.

The coupled Stokes–Darcy system (cont.)

We also assume that the boundary $\Gamma_2 = \partial\Omega \cap \partial\Omega_2$ of the porous medium is partitioned into disjoint Neumann and Dirichlet parts Γ_{2N} and Γ_{2D}, with Γ_{2D} having positive measure.

The flow in Ω_2 is governed by Darcy's Law:

$$-\nabla \cdot \mathbf{K}\,\nabla p_2 = f_2 \quad \text{in} \quad \Omega_2,$$

$$p_2 = g_D \quad \text{on} \quad \Gamma_{2D},$$

$$\mathbf{K}\,\nabla p_2 \cdot \mathbf{n}_2 = g_N \quad \text{on} \quad \Gamma_{2N}.$$

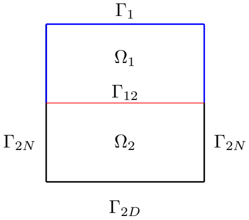
Here p_2 represents the Darcy pressure in Ω_2, and the SPD matrix $\mathbf{K}$ represents the hydraulic conductivity in the porous medium. The Darcy velocity can be obtained from the pressure using

$$\mathbf{u}_2 = -\mathbf{K}\,\nabla p_2 \quad \text{in} \quad \Omega_2.$$

The coupled Stokes–Darcy system (cont.)

Computational domain

- Let Ω be a bounded domain in $\mathbb{R}^2$ subdivided into disjoint subdomains Ω_1 and Ω_2 by an interface Γ_{12}



- The boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2$ with:

$$\Gamma_1 = \partial\Omega_1 \backslash \Gamma_{12} \text{ and } \Gamma_2 = \Gamma_{2N} \cup \Gamma_{2D}$$

1

The coupled Stokes–Darcy system (cont.)

The coupling between the two flows comes from the following interface conditions on the internal boundary Γ_{12}:

Let $\mathbf{n}_{12}$ and $\mathbf{t}_{12}$ denote the unit normal vector directed from Ω_1 to Ω_2 and the unit tangent vector to the interface. Then we impose

$$\mathbf{u}_1 \cdot \mathbf{n}_{12} = -(\mathbf{K}\,\nabla p_2) \cdot \mathbf{n}_{12},$$

$$(-2\nu D(\mathbf{u}_1)\,\mathbf{n}_{12} + p_1 \mathbf{n}_{12}) \cdot \mathbf{n}_{12} = p_2,$$

$$\mathbf{u}_1 \cdot \mathbf{t}_{12} = -2\nu G(D(\mathbf{u}_1)\,\mathbf{n}_{12}) \cdot \mathbf{t}_{12}.$$

The first two conditions enforce mass conservation and the balance of normal forces across the interface; the third condition represents the Beavers–Joseph–Saffman (BJS) Law, in which G is an experimentally determined constant.

The coupled Stokes–Darcy system (cont.)

Let

$$\mathbf{X} = \{\mathbf{v}_1 \in (H^1(\Omega_1))^d \,|\, \mathbf{v}_1 = \mathbf{0} \;\text{ on }\; \Gamma_1\}, \quad Q_1 = L^2(\Omega_1)$$

be the Stokes velocity and pressure spaces and let

$$Q_2 = \{q_2 \in H^1(\Omega_2) \,|\, q_2 = 0 \;\text{ in }\; \Gamma_{2D}\}$$

be the Darcy pressure space.

The weak formulation of the coupled Stokes-Darcy problem is: find $\mathbf{u}_1 \in \mathbf{X}$, $p_1 \in Q_1$ and $p_2 \in Q_2$ such that

$$a(\mathbf{u}_1, p_2; \mathbf{v}_1, q_2) + b(\mathbf{v}_1, p_1) = \mathbf{f}(\mathbf{v}_1, q_2) \quad \forall \mathbf{v}_1 \in \mathbf{X}, \; \forall q_2 \in Q_2,$$

$$b(\mathbf{u}_1, q_1) = 0 \;\; \forall q_1 \in Q_1.$$

The bilinear forms a and b and the functional $\mathbf{f}$ are given in the next slide. Brezzi–Fortin theory ensures the well-posedness of the problem.

The coupled Stokes–Darcy system (cont.)

Here

$$a(\mathbf{u}_1, p_2; \mathbf{v}_1, q_2) = a_{\Omega_1}(\mathbf{u}_1, \mathbf{v}_1) + a_{\Omega_2}(p_2, q_2) + a_{\Gamma_{12}}(\mathbf{u}_1, p_2; \mathbf{v}_1, q_2)$$

where

$$a_{\Omega_1}(\mathbf{u}_1, \mathbf{v}_1) = 2\nu \int_{\Omega_1} D(\mathbf{u}_1) \, : \, D(\mathbf{v}_1) + \frac{1}{G} \int_{\Gamma_{12}} (\mathbf{u}_1 \cdot \mathbf{t}_{12})(\mathbf{v}_1 \cdot \mathbf{t}_{12}),$$

$$a_{\Omega_2}(p_2, q_2) = \int_{\Omega_2} \mathbf{K} \, \nabla p_2 \cdot \nabla q_2,$$

$$a_{\Gamma_{12}}(\mathbf{u}_1, p_2; \mathbf{v}_1, q_2) = \int_{\Gamma_{12}} (p_2 \mathbf{v}_1 - q_2 \mathbf{u}_1) \cdot \mathbf{n}_{12}.$$

Also,

$$b(\mathbf{u}_1, q_1) = -\int_{\Omega_1} (\nabla \cdot \mathbf{u}_1) q_1,$$

and

$$\mathbf{f}(\mathbf{v}_1, q_2) = \int_{\Omega_1} \mathbf{f}_1 \cdot \mathbf{v}_1 + \int_{\Omega_2} f_2 q_2 + \int_{\Gamma_{2N}} g_N q_2.$$

The discrete problem

The weak form is discretized using conforming finite elements spaces $\mathbf{X}^h \subset \mathbf{X}$, $Q_1^h \subset Q_1$ satisfying the inf-sup condition for the Stokes velocity and pressure, such as the MINI and Taylor–Hood elements.

For the Darcy pressure a space of piecewise continuous polynomials $Q_2^h \subset Q_2$ is used (linear in 2D, quadratic in 3D).

The discrete version of the coupled Stokes–Darcy system is of the form

$$\mathcal{A}\,\mathbf{x} = \begin{bmatrix} A_{\Omega_2} & A_{\Gamma_{12}}^T & 0 \\ -A_{\Gamma_{21}} & A_{\Omega_1} & B^T \\ 0 & B & 0 \end{bmatrix} \begin{bmatrix} \hat{p}_2 \\ \hat{\mathbf{u}}_1 \\ \hat{p}_1 \end{bmatrix} = \begin{bmatrix} \hat{f}_2 \\ \hat{\mathbf{f}}_1 \\ 0 \end{bmatrix} = \mathbf{b},$$

where A_{Ω_2}, A_{Ω_1}, $A_{\Gamma_{12}}$ are the matrices of the discrete bilinear forms corresponding to a_{Ω_2}, a_{Ω_1} and $a_{\Gamma_{12}}$, and B is the discrete divergence. Under our assumptions A_{Ω_2} and A_{Ω_1} are SPD, B has full row rank, and $\mathcal{A}$ is nonsingular.

The discrete problem (cont.)

Iterative methods for the solution of this problem have been developed by several authors, for example in

- M. Discacciati and A. Quarteroni, Convergence analysis of a subdomain iterative method for the finite element approximation of the coupling of Stokes and Darcy equations, *Comput. Vis. Sci.*, 6:93–103, 2004.

- M. Discacciati, A. Quarteroni and A. Valli, Robin-Robin domain decomposition methods for the Stokes–Darcy coupling, *SIAM J. Numer. Anal.*, 45:1246–1268, 2007.

- M. Discacciati and A. Quarteroni, Navier–Stokes/Darcy coupling: Modeling, analysis and numerical approximation, *Rev. Mat. Complut.*, 22:315–426, 2009.

- M. Discacciati and L. Gerardo-Giorda, Optimized Schwarz methods for the Stokes–Darcy coupling, *IMA J. Numer. Anal.*, 38:1959–1983, 2018.

These authors focus on domain decomposition/iterative substructuring methods.

The discrete problem (cont.)

Here we are interested in the iterative solution of this block linear system using preconditioned Krylov subspace methods. Our work builds on the following papers:

M. Cai, M. Mu and J. Xu, Preconditioning techniques for a mixed Stokes/Darcy model in porous media applications. *J. Comput. Appl. Math.*, 233:346–355, 2009.

P. Chidyagwai, S. Ladenheim and D. B. Szyld, Constraint preconditioning for the coupled Stokes-Darcy system. *SIAM J. Sci. Comput.*, 38:A668–A690, 2016.

Our contributions include a new block preconditioner based on an augmented Lagrangian formulation, together with spectral and Field-of-Values analysis of the preconditioned matrices. We also perform more extensive experiments with inexact inner solves, especially in 3D.

Block preconditioners

We now introduce a slight change of notation and rewrite the discrete Stokes–Darcy system in the form

$$\mathcal{A}\,\mathbf{x} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & B^T \\ 0 & B & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{b}, \qquad (1)$$

where A_{11}, A_{22} are both SPD, $A_{21} = -A_{12}^T$ and B has full row rank.

We observe in passing that this system is an example of a double saddle point problem, and that similarly structured systems arise in a number of applications, see

F. A. P. Beik and M. Benzi, Iterative methods for double saddle point systems, *SIAM J. Matrix Analysis & Applications*, 39:902–921, 2018.

Block preconditioners (cont.)

Cai et al. (2009) proposed the following block triangular preconditioner:

$$\mathcal{P}_{T_{1,\rho}} := \mathcal{P}_{T_1}(\rho) = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & B & -\rho M_p \end{bmatrix}.$$

where M_p is the mass matrix coming from the Stokes pressure and $\rho > 0$ a parameter.

Chidyagwai et al. (2016) in addition investigated the following constraint preconditioners:

$$\mathcal{P}_{con_D} = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & B^T \\ 0 & B & 0 \end{bmatrix}, \quad \mathcal{P}_{con_T} = \begin{bmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & B^T \\ 0 & B & 0 \end{bmatrix}.$$

An Augmented Lagrangian-type preconditioner

Here we propose and analyze a new efective block preconditioner, defined as

$$\mathcal{P}_{r,\alpha} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ 0 & A_{22} + rB^T Q^{-1} B & B^T \\ 0 & 0 & -\frac{1}{\alpha} Q \end{bmatrix}, \qquad (2)$$

applied to the equivalent augmented system of equations $\hat{\mathcal{A}}\,\mathbf{x} = \hat{\mathbf{b}}$, where

$$\hat{\mathcal{A}} = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} + rB^T Q^{-1} B & B^T \\ 0 & B & 0 \end{bmatrix}, \qquad (3)$$

$\hat{\mathbf{b}} = [b_1; b_2 + rB^T Q^{-1} b_3; b_3]$, the matrix Q is SPD, $r \geq 0$ and $\alpha > 0$ are given user-defined parameters.

Here we set $r = \alpha$ or $r = 0$. In the case that $r = \alpha$, the preconditioner is denoted by $\mathcal{P}_r$.

Eigenvalue bounds for $\mathcal{P}_r^{-1}\hat{\mathcal{A}}$

Theorem: The eigenvalues of $\mathcal{P}_r^{-1}\hat{\mathcal{A}}$ are all real and positive. More precisely, we have for all $r > 0$:

$$\sigma(\mathcal{P}_r^{-1}\hat{\mathcal{A}}) \subset \left[\theta, 2 + \frac{\lambda_{\max}(A_{12}^T A_{11}^{-1} A_{12})}{\lambda_{\min}(A_{22})}\right],$$

where

$$\theta = \frac{\zeta}{2 + \lambda_{\max}(A_{12}^T A_{11}^{-1} A_{12})/\lambda_{\min}(A_{22})}$$

with

$$\zeta = \min\left\{ \frac{r\,\mathbf{y}^* B^T Q^{-1} B\mathbf{y}}{\mathbf{y}^* A_{22}\mathbf{y} + r\,\mathbf{y}^* B^T Q^{-1} B\mathbf{y}} \,\middle|\, \mathbf{y} \notin \mathrm{Ker}(B) \right\}.$$

Moreover, if $A_{22} \succcurlyeq A_{12}^T A_{11}^{-1} A_{12}$, then all the eigenvalues lie in the interval $[1, 2]$ in the limit as $r \to \infty$.

Eigenvalue bounds for $\mathcal{P}_r^{-1}\hat{\mathcal{A}}$ (cont.)

Actually, it turns out that most of the eigenvalues of $\mathcal{P}_r^{-1}\mathcal{A}$ tend to 1 as $r \to \infty$.

Indeed, the proof of the previous theorem reveals that if $\lambda \in \sigma(\mathcal{P}_r^{-1}\hat{\mathcal{A}})$, then there exists a nonzero vector $\mathbf{y}$ such that λ satisfies the following quadratic equation:

$$\lambda^2 - \gamma\lambda + \eta = 0, \tag{4}$$

where

$$\gamma = 1 + \frac{\mathbf{y}^* \left(A_{12}^T A_{11}^{-1} A_{12} + r\, B^T Q^{-1} B\right) \mathbf{y}}{\mathbf{y}^* \hat{A}_{22}\, \mathbf{y}} \quad \text{and} \quad \eta = \frac{r\, \mathbf{y}^* B^T Q^{-1} B\, \mathbf{y}}{\mathbf{y}^* \hat{A}_{22}\, \mathbf{y}},$$

with $\hat{A}_{22} = A_{22} + rB^T Q^{-1} B$. Evidently, $\gamma = 1 + \tilde{\gamma} + \eta$ with

$$\tilde{\gamma} = \frac{\mathbf{y}^* \left(A_{12}^T A_{11}^{-1} A_{12}\right) \mathbf{y}}{\mathbf{y}^* \hat{A}_{22}\, \mathbf{y}}.$$

Eigenvalue bounds for $\mathcal{P}_r^{-1}\hat{A}$ (cont.)

If λ_1 and λ_2 are the roots of (4) then

$$\lambda_1\lambda_2 = \eta \quad \text{and} \quad \lambda_1 + \lambda_2 = \gamma.$$

Since $\eta \to 1$ and $\tilde{\gamma} \to 0$ as $r \to \infty$ we conclude that if $\mathbf{y} \notin \text{Ker}(B)$, then

$$\lambda_1\lambda_2 \to 1 \quad \text{and} \quad \lambda_1 + \lambda_2 \to 2,$$

as $r \to \infty$, i.e., all the eigenvalues tend to 1 for $r \to \infty$.

Using large values of $r > 0$ leads to ill-conditioned subsystems that must be solved at each outer iteration (but see my second talk for this!).

In our numerical tests good results are obtained for moderate values of r, not exceeding $r = 30$. Also, a practical choice of Q is the diagonal of the pressure mass matrix M_p.

Eigenvalue plots

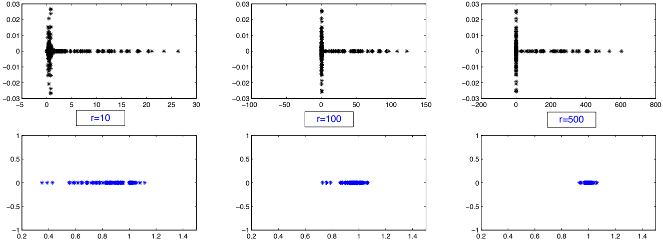


Figure: Eigenvalue distributions of $\hat{\mathcal{A}}$ (top) versus that of the preconditioned matrix $\mathcal{P}_r^{-1}\hat{\mathcal{A}}$ (bottom) for different values of r, with $Q = \mathrm{diag}(M_p)$ for a 3D coupled Stokes-Darcy problem with 1695 dof's.

The problem with eigenvalue bounds

Remark: It is well known that for non-normal problems, the eigenvalues alone are not sufficient to characterize the convergence rate of Krylov subspace methods. In particular, unlike in the SPD case, we cannot conclude that nonsymmetric Krylov iterations will converge with mesh-independent rates from the fact that the eigenvalue spectra of the preconditioned matrices all lie in a fixed interval excluding 0.

Something more than eigenvalue bounds is needed.

Field-of-Values analysis

Recall that two families of symmetric positive definite matrices $\{A_h\}$ and $\{B_h\}$ are said to be spectrally equivalent if there exist h-independent constants α and β with

$$0 < \alpha \leq \lambda_i(B_h^{-1} A_h) \leq \beta, \quad \forall i.$$

Equivalently, $\{A_h\}$ and $\{B_h\}$ are spectrally equivalent if the spectral condition number $\kappa_2(B_h^{-1} A_h)$ is uniformly bounded with respect to h.

Yet another equivalent definition is that the generalized Rayleigh quotients associated with A_h and B_h are uniformly bounded:

$$0 < \alpha \leq \frac{\langle A_h \mathbf{x}, \mathbf{x} \rangle}{\langle B_h \mathbf{x}, \mathbf{x} \rangle} \leq \beta, \quad \forall \mathbf{x} \neq \mathbf{0}.$$

Note that this is an equivalence relation between families of matrices.

Field-of-Values analysis (cont.)

If a discretized PDE leads to a sequence of linear systems $A_h \mathbf{u}_h = \mathbf{b}_h$, a family of spectrally equivalent preconditioners $\{B_h\}$ guarantees that the PCG method will converge in a number of steps that is uniformly bounded with respect to the parameter h.

If h denotes some measure of the mesh size, the resulting PCG iteration exhibits mesh-independent convergence.

If, in addition, the cost of applying the preconditioner B_h is linear in the number of DOFs, we say that the preconditioner is optimal with respect to the mesh size h.

In general, of course, the actual performance of the preconditioner can be affected by other factors.

Field-of-Values analysis (cont.)

When the preconditioned system is not symmetrizable with positive eigenvalues, for example because the preconditioner is indefinite or non-normal, then spectral equivalence is no longer the appropriate tool to analyze the convergence of preconditioned Krylov methods, and PCG cannot be applied.

In this case, the notions of norm equivalence and of Field-of-Values equivalence, proposed by G. Starke and others, often provide the theoretical framework needed to establish optimality of a class of preconditioners for Krylov methods like GMRES.

Note: This work builds on earlier results in the form of convergence bounds for Krylov methods for nonsymmetric problems due to Elman, Eiermann, and others.

Field-of-Values analysis (cont.)

- A. Klawonn and G. Starke, *Block triangular preconditioners for nonsymmetric saddle point problems: field-of-values analysis*, Numer. Math. 81:577–594, 1999.
- D. Loghin and A. J. Wathen, *Analysis of preconditioners for saddle–point problems*, SIAM J. Sci. Comput., 25:2029–2049, 2004.
- M. Benzi and M. A. Olshanskii, *Field-of-values convergence analysis of augmented Lagrangian preconditioners for the linearized Navier-Stokes problem*, SIAM J. Numer. Anal., 49:770–788, 2011.
- E. Aulisa, G. Bornia, V. Howle, and G. Ke, *Field-of-values analysis of preconditioned linearized Rayleigh–Bénard convection problems*, J. Comp. Appl. Math., 369:112582, 2020.
- M. Benzi, *Some uses of the field of values in numerical analysis*, Boll. Unione Matematica Italiana, 14:159–177, 2021.

Field-of-Values analysis (cont.)

Definition: Let $H \in \mathbb{R}^{n\times n}$ be SPD. Two nonsingular matrices $M, N \in \mathbb{R}^{n\times n}$ are H-norm-equivalent, $M \sim_H N$, if there exist positive constants α_0 and β_0 independent of n such that

$$\alpha_0 \le \frac{\|M\mathbf{x}\|_H}{\|N\mathbf{x}\|_H} \le \beta_0, \quad \forall \mathbf{x} \neq \mathbf{0}.$$

Note that $M \sim_H N$ is equivalent to

$$\left\|MN^{-1}\right\|_H \le \beta_0 \tag{5a}$$

$$\left\|NM^{-1}\right\|_H \le \alpha_0^{-1} \tag{5b}$$

We recall that the matrix H-norm $\|A\|_H$ is the operator norm induced by the vector H-norm $\|\mathbf{x}\|_H = \langle H\mathbf{x}, \mathbf{x}\rangle^{\frac{1}{2}}$.

Note: Of course, here M and N should be thought of as families of matrices, parametrized by h.

Field-of-Values analysis (cont.)

We assume that the matrix $\mathcal{A} \in \mathbb{R}^{n\times n}$ satisfies the following stability conditions:

$$\max_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T\mathcal{A}\mathbf{v}}{\|\mathbf{w}\|_H\|\mathbf{v}\|_H} \le c_1, \tag{6a}$$

$$\min_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^T\mathcal{A}v}{\|\mathbf{w}\|_H\|\mathbf{v}\|_H} \ge c_2, \tag{6b}$$

where c_1 and c_2 are positive constants independent of n, and the matrix H is SPD. For the discrete Stokes-Darcy problem we take

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}, \quad H_1 = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad H_2 = M_p,$$

where M_p denotes the mass matrix for the Stokes pressure space.

Field-of-Values analysis (cont.)

Definition: Two nonsingular matrices $M, N \in \mathbb{R}^{n\times n}$ are said to be H-field-of-values-equivalent, $M \approx_H N$, if there exist positive constants α_0 and β_0 independent of n such that the following holds for all nonzero $\mathbf{x} \in \mathbb{R}^n$:

$$\alpha_0 \leq \frac{\left\langle MN^{-1}\mathbf{x}, \mathbf{x}\right\rangle_H}{\langle \mathbf{x}, \mathbf{x}\rangle_H} \quad \text{and} \quad \left\|MN^{-1}\right\|_H \leq \beta_0$$

Remark: If M and N are SPD and $H = I_n$, this reduces to spectral equivalence.

In brief: if we can show that a preconditioner $\mathcal{P}$ is H-FoV-equivalent to $\mathcal{A}$ for a certain choice of H, then the H-fields of values of the matrices $\{\mathcal{A}\mathcal{P}^{-1}\}_n$ are bounded and bounded away from 0 uniformly in n. As a consequence of general convergence results (work by Elman, Eiermann, Beckermann,...) this fact implies that preconditioned GMRES converges at a rate that is independent of n, and therefore of h. For the constraint preconditioners, this equivalence has been established in Chidyagwai et al. (SISC, 2016).

Field-of-Values analysis (cont.)

We now consider the preconditioner $\mathcal{P}_r$. Note that $\mathcal{P}_r$ is an extension of the augmented Lagrangian preconditioner studied, e.g., in B. and Olshanskii (SISC, 2006).

Let us write down the matrix $\hat{\mathcal{A}}$ in the following form:

$$\hat{\mathcal{A}} = \left[\begin{array}{cc|c} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} + rB^T Q^{-1} B & B^T \\ \hline 0 & B & 0 \end{array}\right] = \left[\begin{array}{cc} A_r & C^T \\ C & 0_{n_2 \times n_2} \end{array}\right] \qquad (7)$$

where $A_r \in \mathbb{R}^{n_1 \times n_1}$ and $C = [\,0 \;\; B\,] \in \mathbb{R}^{n_2 \times n_1}$.

Similarly, we write

$$\mathcal{P}_r = \left[\begin{array}{cc|c} A_{11} & A_{12} & 0 \\ 0 & A_{22} + rB^T Q^{-1} B & B^T \\ \hline 0 & 0 & -\frac{1}{r} Q \end{array}\right] = \left[\begin{array}{cc} P_r & C^T \\ 0_{n_2 \times n_1} & -\frac{1}{r} Q \end{array}\right],$$

where $Q \in \mathbb{R}^{n_2 \times n_2}$ is SPD and $r > 0$ is given; in practice, we use $Q = \text{diag}(M_p)$.

Field-of-Values analysis (cont.)

Theorem: Let $\hat{\mathcal{A}}$, $\mathcal{P}_r$ be defined as before and satisfy the usual assumptions. In addition, assume there exists a constant $\gamma > 0$ such that for any $\mathbf{y} \in \mathbb{R}^{n_2} \setminus \{\mathbf{0}\}$, the following inequality holds:

$$\frac{\left\langle SQ^{-1}\mathbf{y}, \mathbf{y}\right\rangle_{H_2^{-1}}}{\left\langle \mathbf{y}, \mathbf{y}\right\rangle_{H_2^{-1}}} \geq \gamma,$$

where $S = CP_r^{-1}C^T$. If $r > 1$ and $A_r \approx_{H^{-1}} P_r$, then there exists $\rho_0 > 0$ such that $\hat{\mathcal{A}} \approx_{H^{-1}} \mathcal{P}_r$ for all $r \geq \rho_0$ provided

$$\|A_r P_r^{-1} - I\|_{H_1^{-1}} \leq r^{-1}.$$

The conditions of our theorem are satisfied by FEM discretizations that satisfy the inf-sup condition. Hence, under our assumptions, $\mathcal{P}_r$ guarantees mesh-independent convergence rates when used with GMRES if r is sufficiently large.

Note: in the paper we show that the restriction $r > 1$ can be removed in many cases.

Numerical experiments

Our test problem is taken from Chidyagwai et al. (SISC, 2016). It consists of a 3D coupled flow problem in the cube $\Omega = \Omega_1 \cup \Omega_2$ with

$$\Omega_1 = [0,2] \times [0,2] \times [1,2] \quad \text{and} \quad \Omega_2 = [0,2] \times [0,2] \times [0,1].$$

The porous medium Ω_2 contains an embedded impermeable cube $[0.75, 1.25] \times [0.75, 1.25] \times [0, 0.50]$. The hydraulic conductivities of the porous medium and embedded impermeable enclosure are $\kappa_1 \mathbf{I}$ an $\kappa_2 \mathbf{I}$, respectively, with $\kappa_1 = 1$ and $\kappa_2 = 10^{-10}$.

The kinematic viscosity is $\nu = 1.0$.

On the horizontal part of Γ_1 we prescribe $\mathbf{u}_1 = (0, 0, -1)^T$ at $z = 2$ and the no-slip condition on the lateral sides of Γ_1.

We prescribe homogeneous Dirichlet boundary conditions on Γ_2 ($z = 0$) and homogeneous Neumann conditions on the rest of the boundary of the porous medium.

Numerical experiments (cont.)

We evaluate different variants of the $\mathcal{P}_{con_D}$, $\mathcal{P}_{con_T}$ and $\mathcal{P}_{T_{1,\rho}}$ preconditioners in conjunction with Flexible GMRES (FGMRES) for solving the problem $\mathcal{A}\,\mathbf{x} = \mathbf{b}$, and the preconditioner $\mathcal{P}_r$ for the augmented Lagrangian formulation $\hat{\mathcal{A}}\,\mathbf{x} = \hat{\mathbf{b}}$.

All of the computations were performed using MATLAB R2020b with an Intel Core i7-10750H CPU @ 2.60GHz processor and 16.0GB RAM.

In all of the experiments, we have used right-hand sides corresponding to random solution vectors and averaged results over 10 test runs.

Numerical experiments (cont.)

At each iteration of FGMRES, we need to solve at least two SPD linear systems as subtasks. These are either solved by (P)CG using loose tolerances.

For the linear systems arising as subtasks, the inner PCG solver for A_{11} (A_{22} and $A_{22} + rB^T Q^{-1} B$) was terminated when the relative residual norm was below 10^{-1} (resp., 10^{-2}) or when the maximum number of 5 (resp., 25) iterations was reached.

The preconditioner $\mathcal{P}_{T_1,\rho}$ requires applying the inverse of M_p. We used PCG with tolerance 10^{-3} or a maximum of 20 iterations.

The preconditioners for PCG are incomplete Cholesky factorizations constructed using the MATLAB function `ichol(.,opts)`, where opts.type $=$'ict', with drop tolerances between 10^{-4} and 10^{-2}.

Numerical experiments (cont.)

In the following tables, outer iteration counts are reported under "Iter".

Under "Iter$_{pcg_i}$" ("Iter$_{cg_i}$"), we report the total number of inner PCG (or CG) iterations performed for solving the linear systems corresponding to block (i, i) of the corresponding preconditioner, where $i = 1, 2$.

In all of the following numerical tests, the initial guess is taken to be the zero vector and the iterations are stopped as soon as $\|\mathcal{A}\mathbf{x}_k - \mathbf{b}\|_2 < 10^{-6}\|\mathbf{b}\|_2$ (or $\|\hat{\mathcal{A}}\mathbf{x}_k - \hat{\mathbf{b}}\|_2 < 10^{-6}\|\mathbf{b}\|_2$), where $\mathbf{x}_k$ is the obtained k-th approximate solution. In all cases tested this criterion produced sufficiently accurate solutions.

Numerical experiments (cont.)

For solving the linear systems corresponding to $A_{22} + rB^TQ^{-1}B$, we tested two strategies:

- Approach I: The matrix is not formed explicitly and the CG method is used without preconditioning.
- Approach II: We formed $A_{22} + rB^TQ^{-1}B$ and used PCG where the preconditioner was constructed by
  - ichol(., struct(‘type‘,‘ict‘,‘droptol‘,1e-3,‘diagcomp‘,0.01)).

Remark: Whereas we could successfully compute the ichol factor using zero shift for the first two problem sizes, we found that adding the shift 0.01 was necessary for larger sizes.

Table 1: Results for FGMRES in conjunction with preconditioner $\mathcal{P}_r$, Approach I.

	$r = 2$				$r = 5$			
size	FGMRES		Inner iterations		FGMRES		Inner iterations	
	Iter	CPU time	Iter_{pcg_1}	Iter_{cg_2}	Iter	CPU time	Iter_{pcg_1}	Iter_{cg_2}
1695	22	0.0521	70	432	15	**0.0404**	52	337
10809	20	0.7018	79	525	15	**0.5293**	63	387
76653	20	7.4126	93	570	15	**5.7076**	71	390
576213	23	71.439	110	595	21	**63.847**	98	536

Table 2: Results for FGMRES in conjunction with preconditioner $\mathcal{P}_r$, Approach II.

size	$r = 5$				$r = 10$			
	FGMRES		Inner iterations		FGMRES		Inner iterations	
	Iter	CPU time	Iter_{pcg_1}	Iter_{pcg_2}	Iter	CPU time	Iter_{pcg_1}	Iter_{pcg_2}
1695	16	0.0864	57	150	13	0.0467	43	80
10809	15	1.3697	63	121	12	0.8918	49	69
76653	13	11.966	64	95	11	8.2454	50	56
576213	14	189.43	64	103	11	155.603	49	58

Note: Now the iteration counts are lower, but the iterations are more expensive because of the cost of forming $A_{22} + rB^TQ^{-1}B$ and computing the corresponding incomplete factorizations. For the largest problem size, solve times are more than doubled!

Table 3: Results for FGMRES in conjunction with preconditioners $\mathcal{P}_{T,0.6}$ and $\tilde{\mathcal{P}}_{T_1,0.6}$.

	$\mathcal{P}_{T_1,0.6}$				$\tilde{\mathcal{P}}_{T_1,0.6}$			
size	FGMRES		Inner iterations		FGMRES		Inner iterations	
	Iter	CPU time	Iter$_{pcg_1}$	Iter$_{pcg_2}$	Iter	CPU time	Iter$_{pcg_1}$	Iter$_{pcg_2}$
1695	21	0.1482	69	437	31	0.2066	97	680
10809	21	2.2306	57	418	37	3.9359	127	836
76653	21	22.175	63	466	38	26.219	117	511
576213	22	214.43	101	516	37	317.45	167	753

In the $\tilde{\mathcal{P}}_{T_1,0.6}$ variant, the pressure mass matrix in the (3,3) block of the preconditioner is replaced by its diagonal.

Table 4: Results for FGMRES in conjunction with preconditioners $\mathcal{P}_{con_D}$ and $\mathcal{P}_{con_T}$.

	$\mathcal{P}_{con_D}$				$\mathcal{P}_{con_T}$			
size	FGMRES		Inner iterations		FGMRES		Inner iterations	
	Iter	CPU time	Iter_{pcg_1}	Iter_{pcg_2}	Iter	CPU time	Iter_{pcg_1}	Iter_{pcg_2}
1695	21	0.1295	77	606	18	0.1001	68	462
10809	20	1.2693	97	697	19	1.2694	89	619
76653	29	18.417	109	975	26	16.920	103	860
576213	61	319.99	167	805	72	380.37	189	974

These results show that constraint preconditioners implemented inexactly (with IC-PCG used for the inexact inner solves) suffer a severe degradation as $h \to 0$ (unlike the other preconditioners), hence they don't appear to be competitive.

Experimens with ARMS

In alternative to IC-CG, we performed some experiments with Saad's ARMS preconditioner for the subsystems associated with the various block preconditioners.

With this approach all block preconditioners (including the constraint ones) appear robust, displaying mesh-independent convergence. The iteration times with all preconditioners tested now show good scalability; the augmented Lagrangian-based preconditioner $\mathcal{P}_r$ still outperforms all others, converging in 16 iterations (about 63.6s) when $r = 5$ on the largest size problem, while the block triangular preconditioner $\mathcal{P}_{T,0.6}$ takes 20 iterations (about 82.8s).

The construction costs for ARMS, however, appear to be prohibitive, at least in Matlab, completely off-setting any gains in performance.

Conclusions and future work

- We have studied several block preconditioners for FEM discretizations of the coupled Stokes-Darcy system
- Preconditioners based on the augmented Lagrangian approach show good performance on the test problems considered
- "Ideal" variants of some of the preconditioners have been shown to be FoV-equivalent to the system matrix, but...
- ... solution times do not scale perfectly when inexact solves with IC-CG are used
- Using ARMS leads to a scalable iterative solver but preconditioner construction is too expensive; a better alternative is needed (see next talk!)
- To do: coupled Navier-Stokes-Darcy system.