6.5 The Saddle Point Stokes Problem

So far the matrix C has been diagonal—no trouble to invert. This section jumps to a fluid flow problem that is still linear (simpler than Navier-Stokes). But now C^{-1} represents the positive definite Laplacian $(-\Delta)$ in the continuous problem, and it is a finite difference or finite element discretization in the matrix problem. The matrix C^{-1} is sparse, but C itself involves a serious boundary value problem.

In this case we don't reduce the computation all the way to $A^{\mathrm{T}}CA$! Instead we stay with the sparse matrix whose blocks are C^{-1}, A, A^{T}, and 0. This is an *indefinite* matrix, with that zero block on its diagonal. The solution v, p is a *saddle point*.

This section also discusses the preconditioning of indefinite systems (not only Stokes).

The Stokes Problem

The unknowns in the two-dimensional Stokes problem are the velocity components in $v = (v_1(x,y), v_2(x,y))$ and the pressure $p(x,y)$. The flow is *incompressible* so the velocity vector is constrained by $\operatorname{div} v = 0$. Therefore A^{T} is (minus) the divergence and A must be the gradient. Notice the different letters in our framework (u is changed to p, w is changed to v, and b is changed to f):

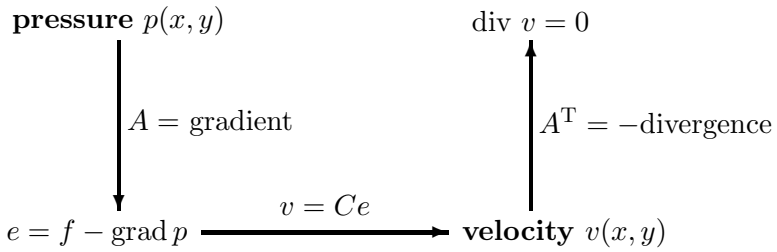


Figure 6.15: The three-step $A^{\mathrm{T}}CA$ framework for the Stokes problem.

The real novelty is in $e = C^{-1}v = (-\Delta v_1, -\Delta v_2)$. The Stokes problem asks us to find a velocity vector v and a pressure p that solve $-\Delta v = f - \operatorname{grad} p$ and $\operatorname{div} v = 0$:

$$-\left(\frac{\partial^2 v_1}{\partial x^2} + \frac{\partial^2 v_1}{\partial y^2}\right) + \frac{\partial p}{\partial x} = f_1(x,y) \tag{1a}$$

$$-\left(\frac{\partial^2 v_2}{\partial x^2} + \frac{\partial^2 v_2}{\partial y^2}\right) + \frac{\partial p}{\partial y} = f_2(x,y) \tag{1b}$$

$$\frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} = 0 \tag{1c}$$

This describes **slow viscous flow**. It is not for aeronautics! The full Navier-Stokes equations have extra nonlinear terms from the motion of the underlying fluid. The Stokes problem could come into its own in biological applications, not for large-scale blood flow in the heart but for small-scale movements in capillaries (or even in cells). This linear problem is simpler than Navier-Stokes, and needs to be understood first. We keep it simple by omitting discussion of the boundary conditions.

In the language of optimization, the pressure $p(x,y)$ is the Lagrange multiplier that imposes the incompressibility constraint div $v = 0$ when the energy is minimized:

$$\text{Minimize} \iint \left(|\operatorname{grad} v_1|^2 + |\operatorname{grad} v_2|^2 - 2f \cdot \operatorname{grad} v \right) dx\, dy \ \text{ subject to } \ \operatorname{div} v = 0\,.$$

The Lagrangian $L(v,p)$ adds $\int\int \delta L/\delta v = 0$ and $\delta L/\delta p = 0$ are exactly the Stokes equations (1). Chapter 6 will develop this primal-dual block form to find the saddle point of L—here we only need to know A and C.

The key point is that we do not eliminate v to reach $K = A^{\mathrm{T}}CA$, because C is *the inverse of the Laplacian*. It is diagonal only in frequency space, not in physical x, y space. The Stokes problem stays in its block form (1) with two unknowns v and p:

$$\textbf{Block form (continuous)} \quad \begin{bmatrix} C^{-1} & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} v \\ p \end{bmatrix} = \begin{bmatrix} -\Delta & \operatorname{grad} \\ -\operatorname{div} & 0 \end{bmatrix} \begin{bmatrix} v \\ p \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}. \tag{2}$$

In the discrete case, that block matrix will be symmetric but ***indefinite*** (as usual). It will have positive and also negative eigenvalues and pivots (positive pivots from C^{-1}, negative pivots from $-A^{\mathrm{T}}CA$). Finite elements for velocity v and pressure p have to be carefully matched to ensure that the discrete problem has a good solution. The approximating polynomials are often *one degree higher* for v than for p.

The Inf-Sup Condition

Restricting a positive definite operator K to a subspace (by a projection P) never destroys positive definiteness. All eigenvalues of PKP^{T} lie safely between $\lambda_{\min}(K)$ and $\lambda_{\max}(K)$. But an indefinite operator has a *negative* $\lambda_{\min}$, so this interval includes zero: not safe! We must take particular care that the block matrix is safely invertible.

These **mixed problems**, or **saddle point problems**, are partly positive and partly negative. The new requirement for a bounded inverse is an *inf-sup condition* (also known as a Babuska-Brezzi condition): **For every p there must be a v so that**

$$v^{\mathrm{T}}Ap \ge \beta \sqrt{v^{\mathrm{T}}C^{-1}v}\ \sqrt{p^{\mathrm{T}}p} \ \text{ for a fixed } \beta > 0\,. \tag{3}$$

This "compatibility" of velocities and pressures produces $\|(A^{\mathrm{T}}CA)^{-1}\| \le 1/\beta^2$.

Condition (3) will immediately fail if there is a nonzero pressure p with $Ap = 0$. In the discrete case, this means that A must have full column rank. Otherwise the

last columns of the block matrix in (2) are not independent, and the block matrix is not invertible. In particular, the rectangular A cannot be short and wide.

More exactly, the lower bound (3) on $v^{\mathrm{T}}Ap$ requires that a nonzero pressure p may not be orthogonal to every $A^{\mathrm{T}}v$. The space of all $A^{\mathrm{T}}v$'s must have *at least* the dimension of the space of p's. Thinking of A^{T} as divergence, this indicates why the velocity finite elements are often one polynomial degree above the pressure finite elements. But dimensions of spaces are not sufficient to confirm that (3) is actually true. Each choice of finite elements for v and p requires its own analysis.

Now we explain the highlighted statement after equation (3): $A^{\mathrm{T}}CA$ is bounded below by β^2 and $\|(A^{\mathrm{T}}CA)^{-1}\| \le 1/\beta^2$ when the inf-sup condition holds. Begin by introducing $w = C^{-1/2}v$. The continuous and discrete Laplacians, both called C^{-1}, have positive definite square roots $C^{-1/2}$ (never to be computed). Then (3) reads:

$$\text{For every } p \text{ there must be a } w \text{ so that } \quad \frac{w^{\mathrm{T}}C^{1/2}Ap}{\|w\|} \ge \beta\|p\|\,. \tag{4}$$

This is where we take the maximum over all w (the *sup*). On the left, the maximum of $w^{\mathrm{T}}z/\|w\|$ is exactly $\|z\|$. This is attained when $w = z$ (then the cosine is one). Since (4) has $z = C^{1/2}Ap$, this best choice of w reduces (3) to

$$\text{For every } p,\ \|z\| = \|C^{1/2}Ap\| \ge \beta\|p\|\,. \tag{5}$$

Square both sides, and minimize their ratio over all p (the *inf*). The inequality

$$\|C^{1/2}Ap\|^2 \ge \beta^2\|p\|^2 \quad \text{or} \quad p^{\mathrm{T}}A^{\mathrm{T}}CAp \ge \beta^2 p^{\mathrm{T}}p \tag{6}$$

is tightest at the lowest eigenvector and eigenvalue, where $A^{\mathrm{T}}CAp = \beta^2 p$. The largest eigenvalue of $(A^{\mathrm{T}}CA)^{-1}$ is $1/\beta^2$. The matrix is symmetric so its norm is also $1/\beta^2$.

A key goal in replacing $\|(A^{\mathrm{T}}CA)^{-1}\| \le 1/\beta^2$ by the inf-sup condition is to work with C^{-1} and not C. We can apply (3) to the continuous problem (it holds, so Stokes is well-posed). We can also test (3) on subspaces V_h and P_h of finite element trial functions. It is a *compatibility condition*: for every pressure p_h in P_h, there must be a velocity v_h in V_h satisfying (3). If and when this is established, with a bound β_h that stays away from zero, the **mixed finite element method** will be stable.

Correction to $p^{\mathrm{T}}p$: Finite elements are functions $p_h(x,y)$, not just vectors v and p. So inner products and norms of v_h and p_h come from integrals and not sums. The L^2 norm of a pressure trial function $p_h = \sum p_j\phi_j(x,y)$ is connected to the discrete coefficient vector p through a positive definite "pressure mass matrix" Q:

$$\|p_h\|^2 = \iint (p_h)^2\,dx\,dy = \sum\sum p_i p_j \iint \phi_i\phi_j\,dx\,dy = p^{\mathrm{T}}Qp\,. \tag{7}$$

The correct inf-sup condition (3) changes $p^{\mathrm{T}}p$ to $p^{\mathrm{T}}Qp$. When satisfied for the finite elements $\mathbf{3}, \mathbf{5}, \mathbf{8}$ in the list below, the inequality (6) involves $p^{\mathrm{T}}Qp$:

Correction to (6) $$p^{\mathrm{T}}A^{\mathrm{T}}CAp \ge \beta^2 p^{\mathrm{T}}Qp \tag{8}$$

This makes Q (or even its diagonal part) a simple and useful candidate as preconditioner. Not the best, but it is directly available and it scales the problem correctly.

Testing the Inf-Sup Condition

The inf-sup condition can be delicate. The books by Brezzi-Fortin and Elman-Silvester-Wathen are excellent references. From the latter book we report on typical results, writing P_0, P_1, P_2 for constant, linear, and quadratic polynomials on triangles, and Q_0, Q_1, Q_2 for constant, bilinear, and biquadratic elements on quadrilaterals:

1. Velocities in P_1, pressures in P_0: failure

2. Velocities in P_1, pressures in P_1: failure

3. Velocities in P_2, pressures in P_1: **success**

4. Velocities in Q_1, pressures in Q_1: failure

5. Velocities in Q_2, pressures in Q_1: **success** (Figure 6.17)

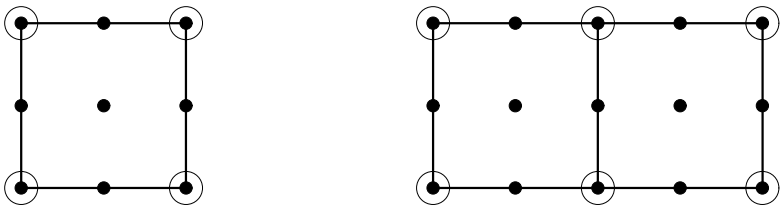


Figure 6.16: Q_2 velocities • and Q_1 pressures ◯: failure on one square, success on two.

For $Q_2 - Q_1$ on one square, A is 2 by 4 (four pressures, two equations $\operatorname{div} v = 0$ at the center). There will surely be non-constant solutions to $Ap = 0$: failure.

For two squares, A is 6 by 6 (six pressures, $\operatorname{div} v = 0$ at three internal nodes). Only a constant pressure solves $Ap = 0$, and this *hydrostatic solution* is removed by the boundary conditions (like grounding a node, here a better choice would be $\iint p\, dx\, dy = 0$). A related "serendipity" element is also successful, by removing the node inside the square and the x^2y^2 term in the velocities. That leaves 8 nodes matching 8 terms in the polynomial.

The list of element pairs also includes pressures P_{-1} and Q_{-1} that are *not continuous* between elements (this is permitted since delta functions don't appear):

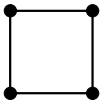
6. Velocities in P_2, pressures in P_{-1}: failure

7. Velocities in Q_2, pressures in Q_{-1}: failure

8. Velocities in Q_2, pressures in P_{-1}: **success**

9. Velocities in Q_1, pressures in Q_0: failure (Figure 6.18)

In each square
$v = a + bx + cy + dxy$ in Q_1
$p =$ *constant* in Q_0



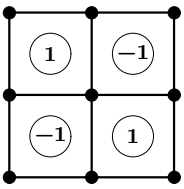


Figure 6.17: Bilinear velocities •, constant pressures ◯ (unstable checkerboard mode).

The failure of $Q_1 - Q_0$ is especially important. Q_0 means a constant pressure in each square. This is the simplest pair of elements that are *conforming* (no delta functions in grad v because Q_1 elements $a + bx + cy + dxy$ are continuous across edges).

Four squares are not stable because that pressure vector $p = (1, -1, 1, -1)$ satisfies $Ap = 0$. More squares don't help—this checkerboard pattern with "white squares $= 1$" and "black squares $= -1$" continues to be a spurious mode satisfying $Ap = 0$. The columns of A, even after $p =$ constant is removed, are still not independent.

This $Q_1 - Q_0$ element is so simple and potentially useful that it is often *stabilized* by relaxing the incompressibility equation $A^{\mathrm{T}}v = 0$. We insert a new anti-checkerboard matrix $-\alpha E$, exactly chosen to eliminate $(1, -1, 1, -1)$ from the nullspace!

Stabilized The block matrix becomes $\begin{bmatrix} C^{-1} & A \\ A^{\mathrm{T}} & -\alpha E \end{bmatrix}$ with $E = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$.

Our zero block is replaced by the negative semidefinite matrix $-\alpha E$. Elimination leaves $-A^{\mathrm{T}}CA - \alpha E$ in that block, which is more strongly negative. The $Q_1 - Q_0$ construction is now uniformly invertible, since E times the checkerboard p is not zero.

In this special case we know the guilty p. In other cases E is constructed with only the constant pressure in its nullspace (which isn't harmful). The new αE is taken small enough so that its positive eigenvalues roughly match those of $A^{\mathrm{T}}CA$. The zero eigenvalue of $A^{\mathrm{T}}CA$ (produced by the checkerboard pressure) becomes positive.

Solving the Discrete Saddle Point Problem

Don't forget that this is the fundamental problem of scientific computing!

$$\begin{bmatrix} C^{-1} & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} b \\ f \end{bmatrix} \quad \text{or} \quad A^{\mathrm{T}}CAu = f - A^{\mathrm{T}}Cb. \tag{9}$$

In this section, w has become v and u is p. In earlier sections, when C was a diagonal matrix (small-block diagonal, in a multidimensional problem), the reduction to $A^{\mathrm{T}}CA$ was very reasonable. Now the Stokes problem illustrates that C^{-1} can be sparse (coming from finite differences or finite elements) while C is full. In that case

we may prefer to compute with the sparse block form, even though its matrix is not at all positive definite like $A^{\mathrm{T}}CA$.

A direct method (elimination) is appropriate up to $n = 10^4$ or 10^5. It will be used in most problems! Of course an indefinite matrix may require row exchanges to avoid small pivots. A good code will decide that without our intervention.

The pivoting can be organized using 2 by 2 blocks, or we may reorder the unknowns as in Section ____ to keep the L and U factors as sparse as possible. If we eliminate in the original order, subtracting $A^{\mathrm{T}}C$ times one block row from the other, then $-K = -A^{\mathrm{T}}CA$ will appear as the **Schur complement** in the lower right block.

For very large systems, a direct method demands too much computation (and memory). We'll look briefly at iterative methods, remembering that the conjugate gradient method expects positive definiteness. It could be used with $A^{\mathrm{T}}CA$, but it is not safe for the indefinite problem. Each CG step applies $A^{\mathrm{T}}CA$ (possibly preconditioned) to a vector. The matrix-vector multiplications Ap and $A^{\mathrm{T}}v$ are fast. In between, multiplying by C amounts to solving the system $C^{-1}v = Ap$. In practice this will be two linear systems for the velocity components v_1 and v_2, with the same Laplace matrix—two "Poisson solves" at each step. This is not impossible but some inner iteration like multigrid is likely to be used.

A number of iterative methods compute both v and p, based on the block form:

1. The **Uzawa method** has $C^{-1}v^{k+1} = f - Ap^k$ and $p^{k+1} = p^k + \alpha A^{\mathrm{T}}v^{k+1}$
2. **Penalty methods** are described in Chapter ____ on optimization
3. The **augmented Lagrangian method** is presented in detail by Glowinski-Le Tallec.

These and more are discussed by Quarteroni-Valli for the Stokes problem. They all involve the "Poisson solve" associated with C. We turn instead to a closer look at *preconditioners for the general saddle point problem*, whether it arises in fluid flow or optimization or network analysis or elsewhere.

Preconditioning Saddle Point Problems

The basic methods for large sparse indefinite problems are MINRES and GMRES (symmetric and nonsymmetric). Our focus here is on constructing a good preconditioner.

When we only know the matrix entries in C^{-1} and A, the code will have to construct a good preconditioner "blindly" from that information. Recall the fundamental dividing line mentioned in the preface, to be *given a matrix or given a problem*. When we know the underlying problem (like Stokes), we can devise a preconditioner by hand. There is a similar separation between algebraic multigrid and geometric multigrid (the computer thinks or we do).

Look first at the saddle point matrix M and its block factorization $M = LDL^{\mathrm{T}}$:

$$M = \begin{bmatrix} C^{-1} & A \\ A^{\mathrm{T}} & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ A^{\mathrm{T}}C & I \end{bmatrix} \begin{bmatrix} C^{-1} & 0 \\ 0 & -A^{\mathrm{T}}CA \end{bmatrix} \begin{bmatrix} I & CA \\ 0 & I \end{bmatrix}. \tag{10}$$

$m \quad n \qquad m \quad n \qquad m \quad n \qquad m \quad n$

M^{-1} clearly involves C and $(A^{\mathrm{T}}CA)^{-1}$ from the inverse of that middle matrix. Our preconditioners will keep this m, n block structure. To come close to M^{-1}, as preconditioners are intended to do, we could approximate the blocks of M and invert or we could directly approximate C and $(A^{\mathrm{T}}CA)^{-1}$.

Three approximations to M are suggested by Elman-Silvester-Wathen:

Preconditioners $P_1 = \begin{bmatrix} \widehat{C}^{-1} & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}$ $P_2 = \begin{bmatrix} \widehat{C}^{-1} & 0 \\ 0 & \widehat{K} \end{bmatrix}$ $P_3 = \begin{bmatrix} \widehat{C}^{-1} & A \\ 0 & \widehat{K} \end{bmatrix}$

The matrix A (often relatively simple) is kept unchanged. $\widehat{C}$ approximates the Poisson solver C, and $\widehat{K}$ approximates $K = A^{\mathrm{T}}CA$. We must choose $\widehat{C}$ and $\widehat{K}$.

A key point about P_2 and P_3: If $\widehat{C} = C$ and $\widehat{K} = K$, then $P_2^{-1}M$ has only *three* different eigenvalues and $P_3^{-1}M$ has only the eigenvalue $\lambda = 1$. In this case. In that case MINRES and GMRES converge in three or two steps. So those changes from P_1 add very few iterations. Wathen begins with simple suggestions for the preconditioner P_2^{-1}:

1. Replace the Poisson solver C by a single multigrid cycle $\widehat{C}$

2. Replace K^{-1} by four conjugate gradient steps preconditioned by Q or $\mathsf{diag}(Q)$.

The multigrid cycle applies a basic iteration on fine and also coarse meshes, to reduce both high and low frequencies in the error residual (Section ____). The matrix Q that preconditions K is the pressure mass matrix in (7). Its energy $p^{\mathrm{T}}Qp$ gives a lower (and also upper) bound to $p^{\mathrm{T}}Kp$. Usually $\mathsf{diag}(Q)$ also has this "equispectral property", scaling the system correctly even if the approximation to K is not really close.

$\widehat{C}$ and $\widehat{K}$ are defined algorithmically by multigrid and preconditioned conjugate gradient cycles. This is simpler than trying to express them by a matrix.

Nonsymmetric Problems and Model Reduction

We move now toward problems more general than Stokes. They might be associated with the Navier-Stokes equations (div v is still zero) and its linearizations. In that case C will be unsymmetric, from the first derivative terms in the linearization. *Convection joins diffusion* (also in many problems from chemical engineering).

A key point is the relative strength of those two terms, inertia versus viscosity. It is measured by the **Reynolds number** Re = (density)(velocity)(distance)/(viscosity).

Stokes flow is the limiting case Re = 0 for high viscosity and low velocity. The high speed Euler equation $dv/dt = -\operatorname{grad} p + F$ is at the other extreme Re $\to \infty$.

Beyond computational fluid dynamics are other differential equations with their own constraints, and then all the algebraic problems of constrained optimization. Our discussion is moving from "given a problem" toward "given a matrix." The constraints make it a saddle point matrix—possibly unsymmetric, with A^{T} changed to B. We look to preconditioners of the same general forms, as a guide.

Always the key is an approximate inverse to the *Schur complement* $-K = -BCA$:

$$\text{Elimination reduces } \begin{bmatrix} C^{-1} & A \\ B & 0 \end{bmatrix} \text{ to } \begin{bmatrix} C^{-1} & A \\ 0 & -BCA \end{bmatrix} = \begin{bmatrix} C^{-1} & A \\ 0 & -K \end{bmatrix}. \tag{11}$$

Normally K will be dense. If it is also large, we look at iterative methods. Multigrid for C will become *algebraic multigrid* [Briggs-McCormick-van Hensen]. The approximation is constructed directly from the matrix, not using the differential equation that may be in the background.

For $K = BCA$, a "black box" approach to preconditioning is suggested in [____]: Reduce C to a smaller model matrix $\widehat{C}$, satisfying

$$A\tilde{C} \approx CA \quad \text{and then} \quad \widehat{K} = BA\widehat{C} \approx BCA. \tag{12}$$

Notice that C is m by m (large) while $\widehat{C}$ is n by n. The model reduction $CA = A\widehat{C}$ will not be exactly solvable (more equations than unknowns). The meaning of $\approx$ is to be decided by the user! We may apply weighted least squares to compute $\widehat{C}^{-1}$ from $A\widehat{C}^{-1} \approx C^{-1}A$, since the saddle point problem gives us C^{-1} and not C.

I expect new ideas will keep coming, for this fundamental indefinite problem.

Advection-Diffusion Problems

The Stokes problem involves the Laplacian $-\Delta v$. The corresponding matrix C^{-1} is "essentially symmetric." Perfect symmetry holds in the interior and could only be lost at the boundary (it shouldn't be). Now, linearizing the $v \cdot \operatorname{grad} v$ term in the Navier-Stokes equation produces a first derivative "advection" term, and symmetry is definitely lost.