Helmholtz Institute for Supercomputational Physics (HISP)
Fourth Helmholtz Summer School 2004

COMPUTATIONAL FLUID MECHANICS

Laurette TUCKERMAN
laurette@limsi.fr

Recursion Relations and Influence Matrices

Recursion Relations and Influence Matrices

1 Elliptic problems: (Navier-)Stokes, Laplace, and Poisson

Although much of the interesting behavior of the Navier-Stokes equations, from multiplicity of steady states to turbulence, results from their nonlinearity, the difficult basic computational issues are those of the associated linear problem, i.e. the Stokes problem.

$$\partial_t \mathbf{u} = -\nabla p + \nu \boldsymbol{\Delta} \mathbf{u} + \mathbf{f} \tag{1a}$$

$$\nabla \cdot \mathbf{u} = 0 \tag{1b}$$

$$\mathbf{u}|_{\mathrm{bdy}} = \mathbf{b} \tag{1c}$$

where $\mathbf{f}$ represents an external force and $\mathbf{b}$ is the imposed value of the velocity at the boundary. In fact, (1) can encompass the Navier-Stokes equations as well, by allowing $\mathbf{f}$ to include the nonlinear terms $-(\mathbf{u} \cdot \nabla)\mathbf{u}$. Many of these issues even arise in the simpler problems of the scalar heat or diffusion equation

$$\partial_t u = \nu \Delta u + f \tag{2}$$

where f can be zero (homogeneous problem) or non-zero (inhomogeneous problem), the Helmholtz problem

$$(\Delta - \kappa^2) u = f \tag{3}$$

the Laplace problem

$$\Delta u = 0 \tag{4}$$

or the Poisson problem

$$\Delta u = f \tag{5}$$

The time-dependent problems (1) and (2) are parabolic, and so require an initial condition in time and a boundary condition in space. Problems (3)-(5) are elliptic, and so require a boundary condition. We solve (2)-(5) subject to either Dirichlet boundary conditions

$$u|_{\mathrm{bdy}} = b \tag{6}$$

or Neumann boundary conditions

$$\partial_n u|_{\mathrm{bdy}} = b \tag{7}$$

where ∂_n is the derivative normal to the boundary.

In this chapter, we consider methods from linear algebra for transforming the matrices resulting from discretizing these problems. We consider only spectral methods [1] in regular domains: Cartesian, cylindrical, or spherical. Spectral methods are easiest to implement for periodic domains. Our goal is to explain how to impose boundary conditions and, for the Navier-Stokes equations, incompressibility, in domains with one or more bounded directions.

2 Review of Boundary Conditions and Spectral Methods

We first recall the usual techniques in spectral methods, using as an illustrative example the one-dimensional version of the Poisson problem (5) with Dirichlet boundary conditions:

$$\partial_x^2 f = g \tag{8a}$$
$$f(\pm 1) = 0 \tag{8b}$$

Here, and throughout, we restrict ourselves to homogeneous boundary conditions for simplicity. Our treatment is nevertheless quite general since inhomogeneous boundary conditions can be incorporated by redefining the problem. For example, when considering the problem

$$\partial_x^2 \hat{f} = g \tag{9a}$$
$$\hat{f}(\pm 1) = b_\pm \tag{9b}$$

then

$$f \equiv \hat{f} - \left(\frac{b_+ + b_-}{2} + \frac{b_+ - b_-}{2} x \right) \tag{10}$$

is a solution to (8). The same technique can be generalized to the Navier-Stokes equations (1) by subtracting a solution to the unforced Stokes problem ($\mathbf{f} = 0$) with inhomogeneous boundary conditions ($\mathbf{b} \neq 0$).

System (8) is the statement of the spatially continuous problem. That is, $f(x)$ and $g(x)$ represent continuous functions, and ∂_x^2 represents the second derivative operator (we do not distinguish here between partial and ordinary derivatives). We now discretize (8) using a spectral representation in space. We represent functions, the unknown left-hand-side f and the known right-hand-side g, by vectors consisting of the coefficients of their truncated expansions in a set of basis functions, typically Chebyshev polynomials $T_0, T_1, ...T_N$.

$$f(x) = \sum_{j=0}^{N} f_j T_j(x) \tag{11}$$

D^2 is a matrix which describes the action of the second derivative operator on the polynomial series:

$$\partial_x^2 \sum_{j=0}^{N} f_j T_j(x) = \sum_{j=0}^{N} f_j T_j''(x) = \sum_{j=0}^{N} f_j \sum_{\substack{k=0 \\ j+k\text{even}}}^{j} D_{jk}^2 T_k(x) = \sum_{k=0}^{N} \sum_{\substack{j=k \\ j+k\text{even}}}^{N} D_{jk}^2 f_j T_k(x) \tag{12}$$

To transform the physical or grid representation $\{f(x_k)\}$ to the spectral or coefficient representation $\{f_j\}$, we define $\mathcal{F}$, the Chebyshev transform whose inverse is

$$[\mathcal{F}^{-1}]_{kj} = T_j(x_k) \tag{13}$$

For the usual Chebyshev grid $x_k \equiv \cos(\pi k/N)$, $k = 0, \ldots N$, the points are clustered near the boundaries, which is desirable in order to resolve steeper gradients near boundaries, and the Chebyshev transform can be accomplished via a cosine transform since $T_j(x) = \cos(j(\text{acos}(x)))$, so that

$$[\mathcal{F}^{-1}]_{kj} = \cos(j(\text{acos}\cos(\pi k/N)) = \cos(jk\pi/N) \tag{14}$$

We define the second derivative matrix in physical space, formally equivalent to D^2, as $\mathcal{F}^{-1}D^2\mathcal{F}$

We now discuss two techniques for imposing the boundary conditions (8b). In the first, called *collocation*, we represent functions in physical space and write (8a) as:

$$\begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & \mathcal{F}^{-1}D^2\mathcal{F} & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \begin{pmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{pmatrix} = \begin{pmatrix} g(x_0) \\ g(x_1) \\ g(x_2) \\ \vdots \\ g(x_N) \end{pmatrix} \tag{15}$$

We then replace the rows corresponding to the boundary points $x_0 = 1$ and $x_N = -1$ by the boundary conditions:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ . & . & . & . & . \\ . & . & (\mathcal{F}^{-1}D^2\mathcal{F})_{\text{int}} & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{pmatrix} = \begin{pmatrix} 0 \\ g(x_1) \\ g(x_2) \\ \vdots \\ 0 \end{pmatrix} \tag{16}$$

where $(\mathcal{F}^{-1}D^2\mathcal{F})_{\text{int}}$ refers to the rectangular matrix containing the indicated rows of $\mathcal{F}^{-1}D^2\mathcal{F}$, i.e. all but the two boundary rows.

The *tau* method is another technique for imposing boundary conditions. We represent f and g in spectral space and write (8a) as:

$$\begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & D^2 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_N \end{pmatrix} \tag{17}$$

We then replace the rows corresponding to the highest-frequency basis functions T_{N-1}, T_N by the boundary conditions, using the fact that $T_j(\pm 1) = (\pm 1)^j$.

$$\begin{pmatrix} . & . & . & . & . \\ . & . & D^2_{\text{lo}} & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & \dots \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} = \begin{pmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{18}$$

where D^2_{lo} refers to the indicated rows of the matrix D^2, i.e. all but the last two rows. We will denote by $\hat{D}^2$ the matrix D^2 with the substitution of boundary rows and by $\hat{g}$ the vector g with the substitution of boundary values as in (18).

Although the physical representation has greater intuitive appeal, more efficient methods can be formulated for solving equations in spectral representations such as (17) and (18), as we will see below.

We now examine more closely the second-derivative matrix D^2. For the Chebyshev polynomial basis, the elements of D^2 are:

$$D^2_{jk} = \begin{cases} \frac{1}{c_j}k(k^2 - j^2) & j < k \text{ and } j + k \text{ even} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where $c_j = 1 + \delta_{j0}$. Thus D^2 has only $\sim N^2/4$ non-zero elements.

Because Chebyshev polynomials are either even or odd (as are members of all other defined polynomial families), D^2 can be decoupled into two blocks D^2_{even} and D^2_{odd}.

$$D^2_{\text{even}} = \begin{bmatrix} 0 & 4 & 32 & 108 & \dots \\ 0 & 0 & 48 & 192 & \dots \\ 0 & 0 & 0 & 120 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \dots \end{bmatrix} \qquad D^2_{\text{odd}} = \begin{bmatrix} 0 & 24 & 120 & 336 & \dots \\ 0 & 0 & 80 & 280 & \dots \\ 0 & 0 & 0 & 168 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \dots \end{bmatrix} \tag{20}$$

We can decouple the boundary conditions (8b) as well by rewriting them as:

$$0 = \frac{1}{2}(f(+1) + f(-1)) = \frac{1}{2}\sum_{j=0}^{N}((+1)^j + (-1)^j)f_j = \sum_{\substack{j=0 \\ j \text{ even}}}^{N} f_j \tag{21a}$$

$$0 = \frac{1}{2}(f(+1) - f(-1)) = \frac{1}{2}\sum_{j=0}^{N}((+1)^j - (-1)^j)f_j = \sum_{\substack{j=0 \\ j \text{ odd}}}^{N} f_j \tag{21b}$$

This leads to the matrices:

$$\hat{D}^2_{\text{even}} = \begin{bmatrix} 0 & 4 & 32 & 108 & \dots \\ 0 & 0 & 48 & 192 & \dots \\ 0 & 0 & 0 & 120 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & 1 & 1 & 1 & \dots \end{bmatrix} \qquad \hat{D}^2_{\text{odd}} = \begin{bmatrix} 0 & 24 & 120 & 336 & \dots \\ 0 & 0 & 80 & 280 & \dots \\ 0 & 0 & 0 & 168 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & 1 & 1 & 1 & \dots \end{bmatrix} \tag{22}$$

We will assume that N is odd, so that D^2_{even} and D^2_{odd} are of the same size, $(N+1)/2$.

3 Eigenvalues of the Laplacian

D^2_{even} and D^2_{odd} are singular, reflecting the fact that two boundary conditions (8b) are necessary to specify a solution to (8a). An equivalent way to state this is that D^2 has a two-dimensional null space consisting of multiples of $1 = T_0$ (a null vector for D^2_{even}) and of $x = T_1$ (a null vector for D^2_{odd}), as seen by the presence of rows of zeros. Another equivalent statement is that the range or image space of D^2 is only $((N+1)-2)$-dimensional, meaning here that a polynomial g of degree N or $N-1$ cannot be obtained by twice differentiating a polynomial f of degree N.

In fact D^2_{even} and D^2_{odd} are not only singular, they are also non-diagonalizable. The *only* eigenvectors of D^2_{even} and D^2_{odd} are the previously mentioned null vectors. This reflects the fact that the second derivative of a polynomial is a polynomial of lower degree and so cannot be a (non-zero) multiple of the original polynomial. D^2_{even} and D^2_{odd} are each large *Jordan blocks*, with *generalized eigenvectors* replacing eigenvectors. An eigenvalue-eigenvector pair (λ, ϕ) of a matrix A obeys:

$$(A - \lambda I)\phi = 0 \tag{23}$$

while a Jordan block of, for example, size four, obeys:

$$\begin{aligned}(A-\lambda I)\phi_1 &= \phi_2\\ (A-\lambda I)\phi_2 &= \phi_3\\ (A-\lambda I)\phi_3 &= \phi_4\\ (A-\lambda I)\phi_4 &= 0\end{aligned} \tag{24}$$

where ϕ_1, ϕ_2, and ϕ_3 are called generalized eigenvectors. For D^2_{even} and D^2_{odd}, $\lambda = 0$ and (24) can be seen as stating that successive application of D^2 lowers the degree of polynomials, until finally the polynomials x and 1 are attained, and these vanish under a final application of D^2.

The inclusion of boundary conditions alters these properties. We must now consider the problem, e.g. for D^2_{even}:

$$\begin{bmatrix} 0 & 4 & 32 & 108 & \dots \\ 0 & 0 & 48 & 192 & \dots \\ 0 & 0 & 0 & 120 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & 1 & 1 & 1 & \dots \end{bmatrix} \begin{pmatrix} f_0 \\ f_2 \\ f_4 \\ \vdots \\ f_{N-1} \end{pmatrix} = \begin{pmatrix} g_0 \\ g_2 \\ g_4 \\ \vdots \\ 0 \end{pmatrix} \tag{25}$$

We rewrite (25) by partitioning the matrix and vectors as follows:

$$\left(\begin{array}{ccccc|c} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & D^2_{\text{lo,lo}} & \cdot & \cdot & D^2_{\text{lo,hi}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & 1^T & \cdot & \cdot & 1 \end{array}\right) \left(\begin{array}{c} \cdot \\ f_{\text{lo}} \\ \cdot \\ \hline \cdot \\ f_{\text{hi}} \end{array}\right) = \left(\begin{array}{c} \cdot \\ g_{\text{lo}} \\ \cdot \\ \hline \cdot \\ 0 \end{array}\right) \tag{26}$$

where $D^2_{\text{lo,lo}}$ and $D^2_{\text{lo,hi}}$ are the indicated portions of D^2 and 1^T is the row vector of ones. We may then carry out a *Schur decomposition*, which is merely a matrix generalization of the usual solution of a 2×2 system via elimination:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \left[\begin{array}{c} x \\ \hline y \end{array}\right] = \left[\begin{array}{c} z \\ \hline w \end{array}\right] \tag{27}$$

$$Ax + By = z \tag{28a}$$
$$Cx + Dy = w \tag{28b}$$

which leads to

$$y = D^{-1}(w - Cx) \tag{29a}$$
$$(A - BD^{-1}C)x = z - BD^{-1}w \tag{29b}$$

For (26), (29) becomes

$$(D^2_{\text{lo,lo}} - D^2_{\text{lo,hi}}1^T)f_{\text{lo}} = g_{\text{lo}} \tag{30}$$

which we shall write as

$$\tilde{D}^2 \hat{f} = \tilde{g} \tag{31}$$

where $\tilde{D}^2$ is of size $(N+1)/2 - 1$.

Information can be obtained about the behavior of equation(26) from the eigenvalues of the matrix in (30). The eigenvalues of this matrix as a function of the number N of Chebyshev polynomials are shown in figure 1. The extremal eigenvalues are $-2.467 = (\pi/2)^2$ and $-0.047 \times N^{4.44}$. The corresponding eigenvectors are shown on the right.

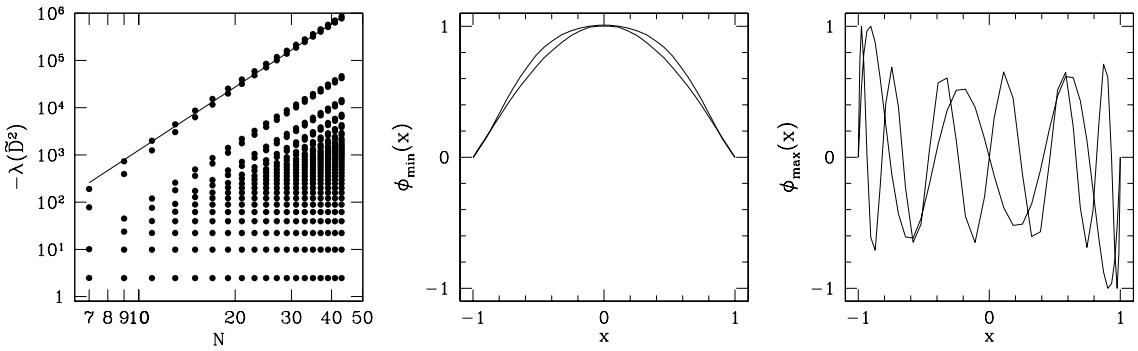

Figure 1: Diagram on the left shows the eigenvalues of $\tilde{D}^2$, the second derivative operator in the Chebyshev polynomial basis with homogeneous Dirichlet boundary conditions, as a function of N, the highest degree of Chebyshev polynomials. All the eigenvalues are negative. The eigenvalue closest to zero is $\lambda_{\min} = -2.467 = -(\pi/2)^2$. The corresponding eigenvector $\phi_{\min}(x)$ is shown in the middle for $N = 7$ and $N = 13$, for which it is indistinguishable from the exact eigenfunction $\cos(\pi x/2)$ of the continuous problem. The eigenvalue largest in absolute value increases like $\lambda_{\max} \sim N^{4.44}$. Its corresponding eigenvector $\phi_{\max}(x)$ is shown on the right for $N = 7$ and $N = 13$. This eigenvector has the highest spatial frequency supported by the grid.

We may also study (26) by considering the problem:

$$\left(\begin{array}{ccccc|c} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & D^2_{\mathrm{lo,lo}} & \cdot & \cdot & D^2_{\mathrm{lo,hi}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & 1^T & \cdot & \cdot & 1 \end{array}\right) \left(\begin{array}{c} \cdot \\ f_{\mathrm{lo}} \\ \cdot \\ \cdot \\ \hline f_{\mathrm{hi}} \end{array}\right) = \lambda \left(\begin{array}{ccccc|c} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & I & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & 0^T & \cdot & \cdot & 0 \end{array}\right) \left(\begin{array}{c} \cdot \\ f_{\mathrm{lo}} \\ \cdot \\ \cdot \\ \hline f_{hi} \end{array}\right) \tag{32}$$

where 0^T is the row vector of zeros. Equation (32) is called a *generalized eigenproblem*, even though this usage is completely unrelated to that of (24). Its solutions are also solutions to (30).

4 The Heat Equation

We turn to the time-dependent heat or diffusion equation

$$\partial_t u = \nu \Delta u + f \tag{33a}$$

$$u|_{\mathrm{bdy}} = 0 \tag{33b}$$

where f may be time-dependent, or even a nonlinear function of u. We use the same Chebyshev spatial discretization as in the previous sections, replacing u, f by vectors of Chebyshev coefficients and the continuous Δ by the matrix D^2. We must now consider temporal discretization. Recall that explicit time

integration formulas for diffusive terms require prohibitively small timesteps for stability. Implicit formulas must be used, the simplest of which are backward Euler (first order accurate) and Crank-Nicolson (second order accurate). Using backward Euler time integration for the diffusive term in (33a) and forward Euler time integration for f, we obtain

$$\begin{aligned} u^{n+1} &= u^n + \Delta t \left(\nu D^2 u^{n+1} + f^n\right) \\ (I - \nu \Delta t D^2) u^{n+1} &= u^n + \Delta t f^n \end{aligned} \tag{34}$$

while Crank-Nicolson time integration for the diffusive term and Adams-Bashforth time integration for the remaining term leads to

$$\begin{aligned} u^{n+1} &= u^n + \frac{\Delta t}{2}\left(D^2(u^{n+1} + u^n) + 3f^n - f^{n-1}\right) \\ (I - \frac{\nu \Delta t}{2} D^2) u^{n+1} &= (I + \frac{\nu \Delta t}{2} D^2) u^n + \frac{\Delta t}{2}(3f^n - f^{n-1}) \end{aligned} \tag{35}$$

We abbreviate either of (34) or (35) as

$$(I - \epsilon D^2) u = s \tag{36}$$

which is $(-\epsilon)$ times the matrix corresponding to the one-dimensional Helmholtz operator (3).

To impose boundary conditions on (36) using the tau method, we again replace the row of the matrix corresponding to the highest frequencies by the boundary condition:

$$\begin{pmatrix} . & . & . & . & . \\ . & . & I - \epsilon D^2 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & 1 & 1 & \dots & \dots \\ 1 & -1 & 1 & \dots & \dots \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-1} \\ u_N \end{pmatrix} = \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{37}$$

Equation (37) also decouples into even and odd Chebyshev modes. Although the spatially continuous differential operators ∂_x^2 and $(I - \epsilon \partial_x^2)$ are of the same nature, their discretized versions are not. The null space of $(I - \epsilon \partial_x^2)$ is, like that of ∂_x^2, two-dimensional, but consists of the exponential functions $\exp(\pm x/\sqrt{\epsilon})$, which cannot be represented in a polynomial basis. Hence, $I - \epsilon D^2$ has no null space. Similarly, the range of $(I - \epsilon D^2)$ is of full dimension $N + 1$: any polynomial can serve as the right-hand-side of $(I - \epsilon D^2)\ u = s$.

5 Recursion Relations

We now turn to the question of solving (18) or (37) economically. Solutions to $D^2 f = g$ obey the well-known recursion relation

$$4j(j^2 - 1) f_j = (j+1) c_{j-2} g_{j-2} - 2j g_j + (j-1) g_{j+2} \qquad \text{for} j \geq 2 \tag{38}$$

The recursion relation (38) can be verified by calculating

$$
\begin{aligned}
& c_{j-2}(j+1)D^2_{j-2,k} - 2jD^2_{j,k} + (j-1)D^2_{j+2,k} \qquad (39)\\
= & \begin{cases} 0 & \text{if } j > k \\ (j+1)k(k^2-(j-2)^2) & \text{if } j = k \\ (j+1)k(k^2-(j-2)^2) - 2jk(k^2-j^2) & \text{if } j = k-2 \\ (j+1)k(k^2-(j-2)^2) - 2jk(k^2-j^2) + (j-1)k(k^2-(j+2)^2) & \text{if } j \le k-4 \end{cases}\\
= & \begin{cases} 0 & \text{if } j > k \\ (j+1)4j(j-1) & \text{if } j = k \\ (j+1)(j+2)8j - 2j(j+1)4(j+1) & \text{if } j = k-2 \\ k^3\left[(j+1) - 2j + (j-1)\right] + k\left[-(j+1)(j-2)^2 + 2j^3 - (j-1)(j+2)^2\right] & \text{if } j \le k-4 \end{cases}\\
= & \begin{cases} 0 & \text{if } j > k \\ (j+1)4j(j-1) & \text{if } j = k \\ 0 & \text{if } j = k-2 \\ 0 & \text{if } j \le k-4 \end{cases}
\end{aligned}
$$

In order to reduce the discrete Poisson equation (25) using the recursion relation and to write the result in matrix form, we multiply both sides by:

$$
B_{\text{even}} \equiv \begin{bmatrix} 1.5 & -4 & 1 & 0 & \dots & 0 & 0 \\ 0 & 5 & -8 & 3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \qquad (40)
$$

which encodes the recursion relation (38). This leads to:

$$
\begin{bmatrix} 0 & 24 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 144 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 4N(N-1)(N-2) \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ \vdots \\ f_{N-3} \\ f_{N-1} \end{bmatrix} = \begin{bmatrix} 1.5 & -4 & 1 & 0 & \dots & 0 & 0 \\ 0 & 5 & -8 & 3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & N & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} g_0 \\ g_2 \\ \vdots \\ g_{N-3} \\ 0 \end{bmatrix} \qquad (41)
$$

Finally, the rows of equation (41) are permuted to place the boundary row at the top, making the matrices diagonally dominant.

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 \\ 0 & 24 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 144 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 4N(N-1)(N-2) \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_4 \\ \vdots \\ f_{N-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 1.5 & -4 & 1 & 0 & \dots & 0 & 0 \\ 0 & 5 & -8 & 3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & N & 0 \end{bmatrix} \begin{bmatrix} g_0 \\ g_2 \\ g_4 \\ \vdots \\ 0 \end{bmatrix} \qquad (42)
$$

Equation (42) can be solved very easily and rapidly. The same operations can also be carried out for the odd modes, D^2_{odd} with a banded matrix B_{odd}.

We would like to generalize (42) in order to solve other elliptic or parabolic problems. Applying the

operator of (40) to $(I - \epsilon D^2)$ leads to:

$$
\begin{aligned}
& c_{j-2}(j+1)(\delta_{j-2,k} - \epsilon D^2_{j-2,k}) - 2j(\delta_{jk} - \epsilon D^2_{j,k}) + (j-1)(\delta_{j+2,k} - \epsilon D^2_{j+2,k}) \qquad (43) \\
= & \begin{cases} 0 & \text{if } j \geq k+4 \\ c_{j-2}(j+1) & \text{if } j = k+2 \\ -2j - \epsilon(j+1)4j(j-1) & \text{if } j = k \\ j-1 & \text{if } j = k-2 \\ 0 & \text{if } j \leq k-4 \end{cases}
\end{aligned}
$$

Written in terms of matrices, the equation $(I - \epsilon D^2) f = g$ for the even polynomials becomes

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 1.5 & -(4+24\epsilon) & 1 & 0 & 0 & \dots \\ 0 & 5 & -(8+144\epsilon) & 3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 1.5 & -4 & 1 & 0 & 0 & \dots \\ 0 & 5 & -8 & 3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} g_0 \\ g_2 \\ g_4 \\ \vdots \end{bmatrix} \qquad (44)
$$

Although (44) is not quite as easily solved as (42), it still can be solved very rapidly by Gaussian elimination. We review the steps of Gaussian elimination in order to emphasize points specific to solution of equations like (44). Gaussian elimination involves two steps. The first is a preprocessing step (meaning that it is particular to the matrix A on the left, but not to the right-hand-side) in which the matrix A is factored as the product of two triangular matrices, usually called an LU decomposition, with factoring matrices which are lower triangular (L) and upper triangular (U).

$$A = LU \qquad (45a)$$

The second is the solve, which must be carried out for each right-hand-side, i.e. for each time step in (34), (35), or (36), in which the factorization is used to calculate the solution.

$$LUf = g \Longleftrightarrow \begin{cases} Lh = g \\ Uf = h \end{cases} \qquad (45b)$$

For a general matrix, factoring requires time $O(N^3)$ and solving requires time $O(N^2)$. For a banded matrix, with J_L non-zero subdiagonal bands and J_U non-zero superdiagonal bands, factoring requires time $O(J_L J_U N)$: L will have J_L non-zero subdiagonals and U will have J_U non-zero superdiagonals. The time required for solving is $O((J_L + J_U)N)$, a considerable economy when $J_L, J_U \ll N$. It is, in fact, this economy that motivates the choice of the tau method (18) rather than the collocation method (16), for imposing boundary conditions.

In general, deviations from bandedness cause "fill-in": the triangular factorizing matrices will be full up to the bandwidth of the extra elements, rather than being as sparse as the original matrix. The matrix on the left-hand-side of (44) is not banded, since it contains an additional full row corresponding to the boundary condition. But the original structure of a banded matrix is preserved by LU decomposition when there are full rows at the bottom and/or columns on the right, and by UL decomposition when there are full rows at the top and/or columns on the left, as shown in figure 2. Such matrices are sometimes called *bordered matrices*. Equation (44) should therefore be solved by Gaussian elimination using a banded UL decomposition.

$$A = UL \qquad (46a)$$

$$ULf = g \Longleftrightarrow \begin{cases} Uh = g \\ Lf = h \end{cases} \qquad (46b)$$

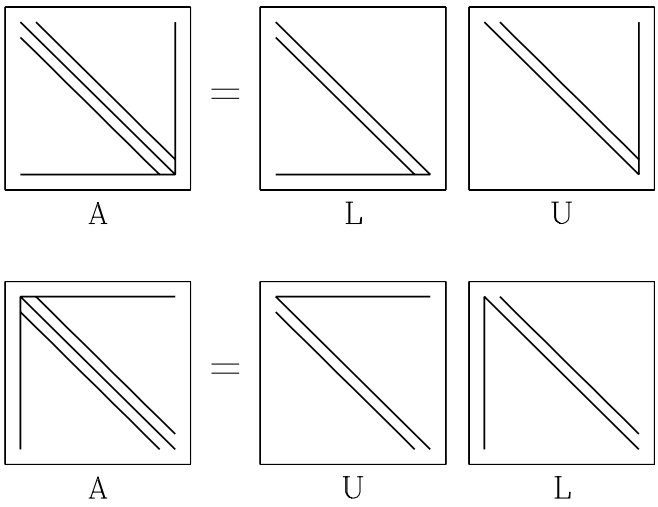

Figure 2: Bandedness is preserved by LU decomposition of a banded matrix with extra rows at the bottom and/or columns at the right, and by UL decomposition of a banded matrix with extra rows at the top and/or columns on the left. Such matrices are sometimes called *bordered* matrices.

In general, pivoting, i.e. re-ordering rows (partial pivoting) or rows and columns (full pivoting), is required in order for the solve to be numerically stable. Full pivoting leads to complete fill-in, whereas partial pivoting, which is all that is usually required, leads only to a small amount of fill-in: if $A = UL$, then U has J_U non-zero superdiagonal bands and L has $J_L + J_U$ subdiagonal bands. However, pivoting is performed in order to make a matrix *diagonally dominant* (meaning that the element of largest absolute value in a row is the diagonal element). The matrix on the left-hand-side of (44) is already diagonally dominant and requires no pivoting. It is for this reason that the boundary condition row was permuted at the bottom of (41) was permuted to the top of (42).

Recursion relations such as (38) occur quite generally for spectral representations of differential operators. We give two more examples. In polar coordinates (r, θ), Poisson's equation reads

$$\Delta f = \frac{1}{r}\partial_r r \partial_r f + \frac{1}{r^2}\partial_\theta^2 f = g \tag{47}$$

If f and g are decomposed into Fourier series in θ, then the equation for each Fourier component m is decoupled from the others:

$$\frac{1}{r}\partial_r r \partial_r f - \frac{m^2}{r^2} f = g \tag{48}$$

We multiply through by r^2 to obtain

$$r\partial_r r \partial_r f - m^2 f = r^2 g \tag{49}$$

The operator

$$r\partial_r r \partial_r \tag{50}$$

acting on series of Chebyshev polynomials has the matrix form

$$D^2_{jk} = \begin{cases} \frac{1}{c_j} k(k^2 - j^2) & j < k \text{ and } j + k \text{ even} \\ j^2 & j = k \\ 0 & \text{otherwise} \end{cases} \tag{51}$$

which is very similar to (19). The three-term recursion relation corresponding to

$$r\partial_r r\partial_r f = h \tag{52}$$

is similar to (38):

$$\begin{aligned}&(j+1)(j-2)^2 c_{j-2} f_{j-2} - 2j(2-j^2) f_j + (j-1)(j+1)^2 f_{j+2} \\ &= (j+1) c_{j-2} h_{j-2} - 2j h_j + (j-1) h_{j+2}\end{aligned} \tag{53}$$

The five-term recursion relation corresponding to (49) is

$$\begin{aligned}&(j+1)c_{j-2}((j-2)^2 - m^2) f_{j-2} - 2j((2-j^2) - m^2) f_j + (j-1)((j+1)^2 - m^2) f_{j+2} \\ &= (j+1) c_{j-2} c_{j-4} g_{j-4}/4 + ((j+1) c_{j-2}(c_{j-2} + c_{j-3}) - 2j c_{j-2}) g_{j-2}/4 \\ &+((j+1)c_{j-2} - 2j(c_j + c_{j-1}) + (j-1)c_j) g_j/4 \\ &+(-2j + (j-1)(c_{j+2} + c_{j+3}) g_{j+2}/4 + (j-1) g_{j+4}/4\end{aligned}$$

where we have used for $h = r^2 g$ the three-term recursion relation

$$h_j = (c_{j-2} g_{j-2} + (c_j + c_{j-1}) g_j + g_{j+2})/4 \tag{54}$$

However, as explained in **Polar Coordinates**, the m^{th} Fourier component of an analytic function $f(r,\theta)$ should contain only powers of r which are greater or equal to m and of the same parity. There, we discuss expansions which are appropriate for polar coordinates.

The third example we give arises from spherical coordinates, where the Laplacian is

$$\Delta f = \frac{1}{\rho^2}\left(\partial_\rho \rho^2 \partial_\rho + \frac{1}{\sin\xi}\partial_\xi \sin\xi\ \partial_\xi + \frac{1}{\sin^2\xi}\partial_\phi^2\right) f \tag{55}$$

The operator $\frac{1}{\sin\xi}\,\partial_\xi \sin\xi\ \partial_\xi$, when acting on the sine series $\sum_j f_j \sin(j\xi)$, is represented by the matrix

$$D^2_{jk} = \begin{cases} -2j & j < k \text{ and } j+k \text{ even} \\ -j(j+1) & j = k \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

The recursion relation equivalent to $D^2 f = g$ is

$$j(j+1)(j+2)\,(f_{j+2} - f_j) = (j+2)\,g_j - j\,g_{j+2} \tag{57}$$

The algebraic property that the matrices (19), (51), and (56) have in common is that all are upper triangular, with their off-diagonal elements taken from a rank-J matrix. This guarantees the existence of a banded matrix B such that BD^2 is also banded. The prototypical example of B is the matrix of (40), with the top row permuted to the bottom row.

THEOREM. Let D^2 be an upper triangular matrix of the form

$$D^2_{ik} = \begin{cases} \sum_{j=1}^{J} S_{ij} T_{jk} & j < k \\ S_{ik} & j = k \\ 0 & \text{otherwise} \end{cases} \tag{58}$$

Define the $J \times J$ matrices S^k and the J-length vectors s^k by

$$S^k_{ij} \equiv S_{j,k+i} \tag{59a}$$

$$s^k_j \equiv S_{kj} \tag{59b}$$

and suppose that s^k is in the range of S^k. (This will always be true if S^k is non-singular.) Then there exists an invertible banded matrix B, depending only on S, with J nonzero superdiagonals, such that BD^2 is also banded with J nonzero superdiagonals.

The proof is constructive and straightforward [8]. We write the matrix BD^2 and show that, in order for the required elements to vanish, elements of B must satisfy equations whose solution is guaranteed by the conditions of the theorem.

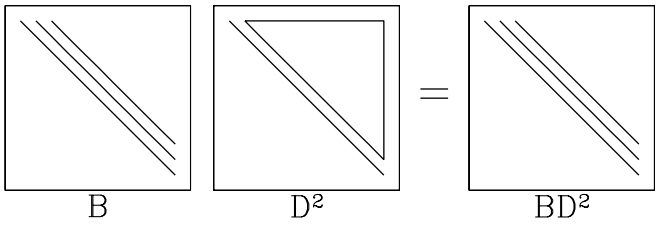

Figure 3: Illustration of the theorem. If D^2 is upper triangular with its superdiagonal portion taken from a rank-J (here, $J = 2$) matrix, then there exists an invertible upper triangular banded matrix B with J superdiagonals such that BD^2 has the same structure as B. (An additional technical condition is required of D^2.) This is equivalent to the existence of recursion relations for acting with or inverting D^2.

6 Several dimensions

Most problems of interest are two- or three- dimensional. In addition, the solution of one-dimensional problems is so inexpensive that there is no need to develop specialized economical solution methods. The recursion relations described above are useful because they can be extended to different types of two- and three- dimensional domains.

The simplest possibility is for the additional direction to be periodic of length 2π. We represent functions in a Chebyshev(x)-Fourier(y) basis:

$$f(x,y) = \sum_{j=0}^{N_x} \sum_{m=0}^{N_y} T_j(x) \left(f_{jm}\, e^{imy} + f^*_{jm}\, e^{-imy} \right) \qquad \text{or} \tag{60a}$$

$$f(x,y) = \sum_{j=0}^{N_x} \sum_{m=0}^{N_y} T_j(x) \left(f^c_{jm} \cos(my) + f^s_{jm} \sin(my) \right) \tag{60b}$$

The spectrally discretized Laplacian is block diagonal and $\Delta f = g$ becomes:

$$\left[\begin{array}{c|c|c|c|c} D_x^2 & 0 & 0 & 0 & 0 \\ \hline 0 & D_x^2 - I & 0 & 0 & 0 \\ \hline 0 & 0 & D_x^2 - 4I & 0 & 0 \\ \hline 0 & 0 & 0 & \ddots & 0 \\ \hline 0 & 0 & 0 & 0 & D_x^2 - N_y^2 I \end{array}\right] \left[\begin{array}{c} f_{.0} \\ \hline f_{.1} \\ \hline f_{.2} \\ \hline \vdots \\ \hline f_{.N_y} \end{array}\right] = \left[\begin{array}{c} g_{.0} \\ \hline g_{.1} \\ \hline g_{.2} \\ \hline \vdots \\ \hline g_{.N_y} \end{array}\right] \tag{61}$$

In (61), each of the submatrices should be considered as further divided into two blocks according to the parity of the Chebyshev polynomials in x. In addition, boundary conditions are applied to each block as described previously for the tau method. Here we include for simplicity in the notation B_x the replacement of high-frequency rows by boundary conditions as in (18), multiplication by the block diagonal matrix whose blocks are B_x defined in (40) which leads to recursion relations in the x direction, and finally the permutation of the boundary condition to the first row. The resulting equation for the m^{th} Fourier component is:

$$B_x(D_x^2 - m^2 I) f_{.m} = B_x g_{.m} \tag{62}$$

If the additional direction is bounded, we may use a Chebyshev(x)-Chebyshev(y) basis:

$$f(x,y) = \sum_{j=0}^{N_x} \sum_{k=0}^{N_y} T_j(x) T_k(y/L_y) \tag{63}$$

We make the following choices for clarity of notation: We use the notation $A(k,k')$ rather than $A_{k,k'}$ for matrix elements. We assume that the matrices have been subdivided into the four blocks which are decoupled by parity, corresponding to even(x)-even(y), even(x)-odd(y), odd(x)-even(y), and odd(x)-odd(y), we will consider only one such block, whose size is $(N_x+1)(N_y+1)/4$. Lengths are normalized so that the x direction is of length 2 and the y direction is of length $L_y \leq 2$, and correspondingly we take $N_y \leq N_x$.

We write the spectrally discretized Laplacian matrix as

$$L^2(j,k,j',k') = D_x^2(j,j')\delta_{k,k'} + \delta_{j,j'} D_y^2(k,k') \tag{64}$$

Matrices corresponding to the x and y directions commute, since they are multiples of the identity in the other direction. We may multiply successively by matrices $B_x(j,j')\delta_{k,k'}$ and $\delta_{j,j'} B_y(k,k')$ as follows:

$$\begin{aligned} \sum_{j'',k''} \quad & B_x(j,j'')\delta_{k,k''} \left[D_x^2(j'',j')\delta_{k'',k'} + \delta_{j,j''} D_y^2(k'',k')\right] \\ & = (B_x D_x^2)(j,j')\delta_{k,k'} + B_x(j,j') D_y^2(k,k') \\ \sum_{j'',k''} \quad & \delta_{j,j''} B_y(k,k'') \left[(B_x D_x^2)(j'',j')\delta_{k'',k'} + B_x(j'',j') D_y^2(k'',k')\right] \\ & = B_y(k,k')(B_x D_x^2)(j,j') + B_x(j,j')(B_y D_y^2)(k,k') \end{aligned} \tag{65}$$

But the matrix of (65) cannot be treated by the recursion relation methods described previously.

A better alternative is to use what is called a Haidvogel-Zang decomposition [2]. This method calls for using an eigenvector-eigenvalue decomposition of the Schur matrix (30) in one direction (preferably the shorter direction, here y) and a recursion relation in the other direction. We write

$$\tilde{D}_y^2 = V_y \Lambda_y V_y^{-1} \tag{66}$$

where V_y is the $(N_y - 1) \times (N_y - 1)$ matrix of eigenvectors and Λ_y is the diagonal matrix of eigenvalues (possibly with 2×2 blocks corresponding to complex eigenvectors). The diagonalization procedure in y corresponds to multiplying

$$\begin{aligned} (D_x^2 + D_y^2) f &= g \\ V_y^{-1}(D_x^2 + \tilde{D}_y^2) V_y V_y^{-1} f &= V_y^{-1} g \\ (D_x^2 + V_y^{-1} \tilde{D}_y^2 V_y) V_y^{-1} f &= V_y^{-1} g \\ (D_x^2 + \Lambda_y) V_y^{-1} f &= V_y^{-1} g \end{aligned} \tag{67}$$

The matrix is now identical in structure to that in the Chebyshev(x)-Fourier(y) representation, in which ∂_{yy} was already diagonal and no boundary conditions were required. Boundary conditions are imposed in the x direction and the standard recursion relation is used:

$$B_x (D_x^2 + \Lambda_y) V_y^{-1} f = B_x V_y^{-1} g \tag{68}$$

The matrices $B_x(D_x^2 + \Lambda_y)$ are banded (with an extra row) in the x direction and diagonal in the y direction by construction. The time required for solution of Poisson's equation via this method is

$$\begin{array}{lll} g & \longrightarrow V_y^{-1} g & O(N_x N_y^2) \\ V_y^{-1} g & \longrightarrow B_x V_y^{-1} g & O(N_x N_y) \\ B_x V_y^{-1} g & \longrightarrow [B_x (D_x^2 + \Lambda_y)]^{-1} B_x V_y^{-1} g = V_y^{-1} f & O(N_x N_y) \\ V_y^{-1} f & \longrightarrow f & O(N_x N_y^2) \end{array} \tag{69}$$

The preprocessing time for diagonalization of D_y^2 and for UL decomposition of $B_x(D_x^2 + \Lambda_y I)$ is

$$\begin{array}{lll} \tilde{D}_y^2 & \longrightarrow V_y \Lambda_y V_y^{-1} & O(N_y^2) \\ B_x (D_x^2 + \Lambda_y) & \longrightarrow UL & O(N_x N_y) \end{array} \tag{70}$$

Adding another periodic direction is not difficult. Thus, the methods described above can be used in cylindrical domains, since the θ direction is periodic. If periodicity is imposed in the z direction, then there are two periodic directions (θ, z). If the z direction is finite, then recursion relations can be used in either the r or the z direction, and diagonalization in the remaining direction. For any periodic direction, the time required to solve the decoupled linear problems is directly proportional to the number of points or Fourier modes in that direction. We do not address here techniques for adding a third non-periodic direction.

7 Influence matrix: Streamfunction-vorticity formulation and Sherman-Morrison-Woodbury formula

The influence matrix, or Green's function, method is used to solve linear problems which are partly, but not fully, coupled. To describe the technique, we use the important example of the two-dimensional Navier-Stokes equations in a rectangular domain,

$$\begin{aligned} \partial_t \mathbf{u} &= -\nabla p + \nu \boldsymbol{\Delta} \mathbf{u} + \mathbf{f} && \text{(71a)} \\ \nabla \cdot \mathbf{u} &= 0 && \text{(71b)} \\ \mathbf{u}|_{\text{bdy}} &= 0 && \text{(71c)} \end{aligned}$$

and carry out the standard manipulations that lead to the streamfunction-vorticity formulation. We write

$$\begin{aligned} \mathbf{u} &= \nabla \times \psi \mathbf{e_z} && \text{(72a)} \\ \omega &= \mathbf{e_z} \cdot \nabla \times \mathbf{u} = \mathbf{e_z} \cdot \nabla \times \nabla \times \psi \mathbf{e_z} = -\Delta \psi && \text{(72b)} \end{aligned}$$

Thus

$$\begin{aligned} u &= -\partial_y \psi && \text{(73a)} \\ v &= \partial_x \psi && \text{(73b)} \end{aligned}$$

We consider two cases for the y direction, periodic and finite. We begin with the periodic case.

$$\begin{aligned} u &= -\partial_y \psi = 0 \text{ at } x = \pm 1 && \text{(74a)} \\ v &= \partial_x \psi = 0 \text{ at } x = \pm 1 && \text{(74b)} \end{aligned}$$

Then (74a) implies that $\psi(x = \pm 1, y) = \psi_\pm$. Since ψ is defined up to an arbitrary constant, we may take $\psi_- = 0$ and define $\psi]_-^+ = \psi_+ - \psi_- = \psi_+$. The *flux* in the y direction is

$$\int dx\ v(x,y) = \int dx\ \partial_x \psi(x,y) = \psi(1,y) - \psi(-1,y) = \psi]_-^+ \qquad (75)$$

Fixing the flux in the y direction completes the specification of the Dirichlet boundary conditions for ψ at $x = \pm 1$. As before, we will take these to be homogeneous for convenience, (though non-zero flux cases are often treated as well). One can instead complete the specification of the problem by imposing the pressure drop in y rather than the flux, but this leads to a more complicated condition. The boundary condition on v leads to homogeneous Neumann boundary conditions on ψ at $x = \pm 1$.

The second possibility we consider is a finite y direction, which we shall take for convenience to be bounded by $y = \pm 1$. In addition to (74) we have

$$\begin{aligned} u &= -\partial_y \psi = 0 \text{ at } y = \pm 1 && \text{(76a)} \\ v &= \partial_x \psi = 0 \text{ at } y = \pm 1 && \text{(76b)} \end{aligned}$$

Then (76b) leads to the conclusion that $\psi]_-^+ = 0$ and, moreover, that $\psi = 0$ everywhere on the boundary. Finally, (76a) leads to homogeneous Neumann boundary conditions on ψ at $y = \pm 1$. We thus have

$$\begin{aligned} \psi|_{\text{bdy}} &= 0 && \text{(77a)} \\ \partial_n \psi|_{\text{bdy}} &= 0 && \text{(77b)} \end{aligned}$$

where ∂_n refers to the partial derivative in the direction normal to the boundary, which may consist of the entire rectangular boundary or merely of $x = \pm 1$.

Taking $\mathbf{e_z} \cdot \nabla \times$ of the Navier-Stokes equations (71a), we obtain

$$\partial_t \omega = \nu \Delta \omega + f \tag{78}$$

where $f \equiv \mathbf{e_z} \cdot \nabla \times \mathbf{f}$. We no longer require (71b) since (72a) is divergence-free by construction.

As explained previously, we use implicit time integration for the diffusive term and explicit time integration for the remaining term which representing the nonlinear term and sources. As before, we obtain

$$(I - \epsilon \Delta)\omega = s \tag{79}$$

where $\omega = \omega^{n+1}$, $\epsilon = \nu \Delta t$ and $s = \omega^n + \Delta t f^n$ for first-order time-stepping as in (34)-(36).

We now discuss the imposition of boundary conditions (77) on system (79). One possibility would be to write (80) as a single fourth-order system

$$(I - \epsilon \Delta)\Delta \psi = -s \tag{80}$$

and to replace the appropriate number of rows with boundary conditions. If the y direction is periodic, then four boundary conditions ($\psi(\pm 1, y) = \partial_x \psi(\pm 1, y) = 0$) would be imposed by replacing the four highest Chebyshev(x) modes $N_x - 3, \ldots N_x$ for each Fourier(y) mode. If the y direction is finite, then Dirichlet and Neumann boundary conditions are to be imposed at each boundary point, requiring the substitution of $2(N_x + N_y)$ rows corresponding to either $j = N_x - 1, N_x$ or $m = N_y - 1, N_y$. However, the resulting time-integration scheme is known to be unstable, or, equivalently, the discretized matrix has spurious eigenvalues which do not correspond to growth rates of the continuous problem (78)-(77). In addition, if there is more than one non-periodic dimension, then $\Delta\Delta$ contains cross terms (like $\partial_x^2 \partial_x^2$) which may be difficult to represent economically.

We now present an approach which, while complicated, leads to correct results. In the *influence matrix*, or *Green's function* technique, we solve two nested problems, the Helmholtz problem (79 and the Poisson problem (72a).

$$(I - \epsilon \Delta)\omega = s \tag{81a}$$

$$\Delta \psi = -\omega \tag{81b}$$

The difficulty is that there are too many boundary conditions for ψ and none for ω.

We seek a solution (ψ, ω) as a superposition of a particular ($s \neq 0$) and a homogeneous ($s = 0$) solution. The homogeneous solution in turn is constructed as a superposition of *Green's functions* which are calculated in a preprocessing step as follows. Corresponding to each point $\mathbf{x}_i$ on the boundary, we construct a homogeneous solution $(\psi_i^{\hom}, \omega_i^{\hom})$ by solving the problem

$$(I - \epsilon \Delta)\omega_i^{\hom} = 0 \tag{82a}$$

$$\omega_i^{\hom}(\mathbf{x}_j) = \delta_{ij} \quad \mathbf{x}_j \in \text{boundary} \tag{82b}$$

$$\Delta \psi_i^{\hom} = -\omega_i^{\hom} \tag{82c}$$

$$\psi_i^{\hom}(\mathbf{x}_j) = 0 \quad \mathbf{x}_j \in \text{boundary} \tag{82d}$$

We then calculate

$$C_{ji} \equiv \partial_n \psi_i^{\hom}(\mathbf{x}_j) \quad \mathbf{x}_j \in \text{boundary} \tag{83}$$

C is called the *influence matrix* or the *capacitance matrix*. It describes the influence of applying a boundary condition $\omega_i^{\rm hom}(\mathbf{x}_j) = \delta_{ij}$ on the value of $\partial_n \psi_i^{\rm hom}(\mathbf{x}_j)$. If the x and y directions are both bounded, then we calculate a homogeneous solution for each boundary point $\mathbf{x} = (x_i, y_i)$. If, instead, the y direction is periodic, then we calculate a homogeneous solution for each x boundary and each Fourier mode in y. C would then be block diagonal, with separate blocks corresponding to each Fourier mode.

Prepared with these results, we solve at each timestep

$$(I - \epsilon\Delta)\omega^{\rm part} = s \tag{84a}$$
$$\omega^{\rm part}(\mathbf{x}_j) = 0 \quad \mathbf{x}_j \in \text{boundary} \tag{84b}$$
$$\Delta\psi^{\rm part} = -\omega^{\rm part} \tag{84c}$$
$$\psi^{\rm part}(\mathbf{x}_j) = 0 \quad \mathbf{x}_j \in \text{boundary} \tag{84d}$$

Superposing the particular and homogeneous solutions, we obtain

$$\omega = \omega^{\rm part} + \sum_i \alpha_i \omega_i^{\rm hom} \tag{85a}$$
$$\psi = \psi^{\rm part} + \sum_i \alpha_i \psi_i^{\rm hom} \tag{85b}$$

By construction, ψ satisfies homogeneous Dirichlet, but not Neumann, boundary conditions:

$$\psi(\mathbf{x}_j) = 0 \tag{86a}$$
$$\partial_n\psi(\mathbf{x}_j) = \partial_n\psi^{\rm part}(\mathbf{x}_j) + \sum_i \alpha_i \partial_n \psi_i^{\rm hom}(\mathbf{x}_j)$$
$$= \partial_n\psi^{\rm part}(\mathbf{x}_j) + \alpha_i C_{ji} \tag{86b}$$

The coefficients α_i of each homogeneous solution at a given timestep are determined by solving the linear system:

$$\sum_i C_{ji}\alpha_i = -\partial_n\psi^{\rm part}(\mathbf{x}_j) \tag{87}$$

The requirements for this algorithm are as follows. Preprocessing requires generation of the homogeneous solutions via repeated solution of (82), generation of C via (83), and the inversion (or LU decomposition) of C.

Geometry	Number of homog. solns.	Time for single homog. soln.	Size of C	Time for C^{-1}
y bounded	$2(N_x + N_y)$	$N_x N_y^2$	$2(N_x + N_y) \times 2(N_x + N_y)$	$(2(N_x + N_y))^3$
y periodic	$2N_y$	$N_x N_y$	$N_y(2 \times 2)$	$N_y(2)^3$

Each timestep then requires solving the particular problem (84), acting with C^{-1} and then adding the appropriate combination of homogeneous solutions. Storing the homogeneous solutions is possible, but an alternative is to solve (84) twice at each timestep, once with homogeneous Dirichlet boundary conditions, and a second time with the corrected Dirichlet boundary values specified by inversion of C.

Geometry	Time for particular soln.	Time for action by C^{-1}	Storage for homog. solns.
y bounded	$2(N_x + N_y)N_x N_y$	$N_x N_y^2$	$(2(N_x + N_y))^2$
y periodic	$2N_x N_y^2$	$N_x N_y$	$N_y(2)^2$

We can interpret this procedure, like all algorithms involving Green's functions, in terms of the *Sherman-Morrison-Woodbury formula* [6, 9] of matrix algebra. In order to explain this formula, we review the concept of *rank* for matrices. Consider forming an $N \times N$ matrix by multiplying a column vector v by a row vector w^T:

$$A^1_{ik} = v_i w^T_k \tag{88}$$

This matrix is extremely singular. By construction, all of its rows (and all of its columns) are linearly dependent. For example, the first column is the scalar w^T_1 times the vector v, the second column is w^T_2 times the vector v, and so on. All singular matrices have at least one zero eigenvalue, but the eigenvalues of A^1 are all zero except for a single one (assuming that neither v nor w^T is the zero vector). A^1 is a matrix of rank 1. The sum of two rank-1 matrices is in general a rank-2 matrix:

$$A^2_{ik} = v_i w^T_k + \tilde{v}_i \tilde{w}^T_k \tag{89}$$

(In (88) and (89), the superscript does not designate a power of a matrix, but merely labels its rank.) v and $\tilde{v}$ are column vectors, and w^T and $\tilde{w}^T$ are row vectors. We can write (89) more compactly as

$$A^2_{ik} = \sum_{j=1}^{2} V_{ij} W^T_{jk} \tag{90}$$

where V is an $N \times 2$ matrix whose first column is v and whose second column is $\tilde{v}$, and W^T is a $2 \times N$ matrix whose first row is w^T and whose second row is $\tilde{w}^T$. A rank-2 matrix has two non-zero eigenvalues. Its rows (columns) are linear combinations of two row (column) vectors In general, a rank-J matrix, defined by

$$A^J_{ik} = \sum_{j=1}^{J} V_{ij} W^T_{jk} \tag{91}$$

has $N - J$ zero eigenvalues and J linearly independent rows or columns.

The Sherman-Morrison formula relates the inverses of two matrices which differ by a rank-1 matrix. This formula was generalized by Woodbury to relate the inverses of two matrices which differ by a *low-rank* matrix. The formula is

$$(H + VW^T)^{-1} = H^{-1} - H^{-1} V C^{-1} W^T H^{-1} \tag{92a}$$

$$C \equiv I + W^T H^{-1} V \tag{92b}$$

Formula (92) is easily verified by substitution:

$$\begin{aligned} (H + VW^T) & \left[H^{-1} - H^{-1} V C^{-1} W^T H^{-1}\right] \\ &= (H + VW^T) H^{-1} - (H + VW^T) H^{-1} V C^{-1} W^T H^{-1} \\ &= I + VW^T H^{-1} - V(I + W^T H^{-1} V) C^{-1} W^T H^{-1} \\ &= I + VW^T H^{-1} - VCC^{-1} W^T H^{-1} \\ &= I \end{aligned} \tag{93}$$

Note that changing one row of a matrix corresponds to a rank-1 change of that matrix, by setting v in (88) to be zero in all entries except for the row which changes, i.e. $v = (0, 0 \ldots 1, 0, \ldots 0)$, and by setting w^T to be the difference between the unchanged and changed row. Similarly, changing J rows of a matrix corresponds to a rank-J change of that matrix, by setting V to be zero except for a $J \times J$

identity block corresponding to the changed rows (where we assume for convenience that the J changed rows are contiguous) and W^T to be the difference between the changed and unchanged rows. Changing J columns of a matrix is equivalent to a rank-J change where W^T is zero except for a $J \times J$ identity block and V contains the diffference between the changed and unchanged columns.

We return to the problem of imposition of boundary conditions (77) on (79) and re-interpret the Green's function procedure set out in (82)-(87) in terms of the Sherman-Morrison-Woodbury formula (92). We write the differential system (79) as

$$\begin{bmatrix} I - \epsilon\Delta & 0 \\ I & \Delta \end{bmatrix} \begin{bmatrix} \omega \\ \psi \end{bmatrix} = \begin{bmatrix} s \\ 0 \end{bmatrix} \tag{94}$$

We substitute the boundary conditions (77) for the high-frequency rows of (94):

$$\left[\begin{array}{cc} (I - \epsilon\Delta)_{\mathrm{lo}} & 0_{\mathrm{lo}} \\ \hline 0 & \partial_n|_{\mathrm{bdy}} \\ \hline I_{\mathrm{lo}} & \Delta_{\mathrm{lo}} \\ \hline 0 & |_{\mathrm{bdy}} \end{array}\right] \begin{bmatrix} \omega \\ \\ \psi \\ \\ \end{bmatrix} = \begin{bmatrix} s_{\mathrm{lo}} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{95}$$

The problem (84) for the particular solution, which has distinct uncoupled (and false) boundary conditions for ω and ψ, is written:

$$\left[\begin{array}{cc} (I - \epsilon\Delta)_{\mathrm{lo}} & 0_{\mathrm{lo}} \\ \hline |_{\mathrm{bdy}} & 0 \\ \hline I_{\mathrm{lo}} & \Delta_{\mathrm{lo}} \\ \hline 0 & |_{\mathrm{bdy}} \end{array}\right] \begin{bmatrix} \omega \\ \\ \psi \\ \\ \end{bmatrix} = \begin{bmatrix} s_{\mathrm{lo}} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{96}$$

We call the matrix of the particular problem H, in the nomenclature of the Sherman-Morrison-Woodbury formula, because the linear problem (96) is easy to solve. We write H^{-1} as shorthand for a *procedure* for solving a system involving this matrix, without ever actually constructing an inverse matrix. The difference between H and the matrix of the exact problem (95) is

$$\left[\begin{array}{cc} 0 & 0 \\ \hline -|_{\mathrm{bdy}} & \partial_n|_{\mathrm{bdy}} \\ \hline 0 & 0 \\ \hline 0 & 0 \end{array}\right] = \begin{bmatrix} 0 \\ I \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -|_{\mathrm{bdy}} & \partial_n|_{\mathrm{bdy}} \end{bmatrix} \equiv VW^T \tag{97}$$

The preprocessing step we have described for constructing the influence matrix C can be written as follows.

$$\begin{array}{ccccccc} \partial_n\psi^{\mathrm{hom}}|_{\mathrm{bdy}} & \leftarrow (-\omega + \partial_n\psi)^{\mathrm{hom}}|_{\mathrm{bdy}} & \leftarrow \begin{bmatrix} \psi^{\mathrm{hom}} \\ \\ \omega^{\mathrm{hom}} \end{bmatrix} & \leftarrow \begin{bmatrix} 0 \\ \omega^{\mathrm{hom}}|_{\mathrm{bdy}} \\ 0 \\ 0 \end{bmatrix} & \leftarrow & \omega^{\mathrm{hom}}|_{\mathrm{bdy}} \\ & I+ & W^T & H^{-1} & & V & \end{array} \tag{98}$$

The action of V is to insert successive boundary values for ω^{hom}. The action of H^{-1} is to solve the homogeneous problem. The action of W^T is to evaluate $[-\omega^{\mathrm{hom}} + \partial_n\psi^{\mathrm{hom}}]|_{\mathrm{bdy}}$. The final addition of I adds $\omega^{\mathrm{hom}}|_{\mathrm{bdy}}$, to yield $\partial_n\psi^{\mathrm{hom}}|_{\mathrm{bdy}}$.

The timestep can be related to the Sherman-Morrison-Woodbury formula by reading from right to left as follows.

$$\begin{array}{ccccccccc} \begin{bmatrix} \psi^{\mathrm{hom}} \\ \\ \omega^{\mathrm{hom}} \\ \\ \end{bmatrix} & \leftarrow & \begin{bmatrix} 0 \\ \omega^{\mathrm{hom}}|_{\mathrm{bdy}} \\ 0 \\ 0 \end{bmatrix} & \leftarrow \omega^{\mathrm{hom}}|_{\mathrm{bdy}} & \leftarrow (\partial_n\psi - \omega)^{\mathrm{part}}|_{\mathrm{bdy}} & \leftarrow & \begin{bmatrix} \psi^{\mathrm{part}} \\ \\ \omega^{\mathrm{part}} \\ \\ \end{bmatrix} & \leftarrow & \begin{bmatrix} s_{\mathrm{lo}} \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ & & H^{-1} & V & C^{-1} & & -W^T & & H^{-1} \end{array} \tag{99}$$

H^{-1} on the inhomogeneous right-hand-side solves the particular problem. $-W^T$ calculates $\partial_n\psi^{\mathrm{part}}|_{\mathrm{bdy}}$ ($\omega^{\mathrm{part}}|_{\mathrm{bdy}}$ is zero by construction). C^{-1} calculates the correct value of $\omega^{\mathrm{hom}}|_{\mathrm{bdy}}$ to compensate. V inserts these boundary values into the right-hand-side. H^{-1} solves the homogeneous problem with the newly calculated boundary values. This homogeneous solution is added to the particular solution.

8 Influence matrix: the velocity-pressure formulation

The issue of pressure in the incompressible Navier-Stokes equations arises in many different forms and is impossible to escape entirely, though it may be concealed (either intentionally or unintentionally). In three dimensions, the streamfunction-vorticity formulation cannot be used. Returning to the original Navier-Stokes formulation

$$\partial_t \mathbf{u} = -\nabla p + \nu \boldsymbol{\Delta} \mathbf{u} + \mathbf{f} \tag{100a}$$

$$\nabla \cdot \mathbf{u} = 0 \tag{100b}$$

$$\mathbf{u}|_{\mathrm{bdy}} = 0 \tag{100c}$$

one often takes the divergence of (100a) to replace it by a Poisson equation for the pressure:

$$\Delta p = \nabla \cdot \mathbf{f} \tag{101}$$

where (100b) has been used and we recall that $\mathbf{f}$ contains any nonlinear terms and/or external forces. The same problem arises for the pressure p as for the vorticity ω: there are no boundary conditions for this elliptic equation. Erroneous reasoning is often used to justify the use of a boundary condition derived from (100a), for example evaluating the normal component $\mathbf{n}\cdot$ of (100a) at the boundary:

$$\partial_n p = \mathbf{n} \cdot (\nu \boldsymbol{\Delta} \mathbf{u} + \mathbf{f}) \tag{102}$$

But (102) is not a legitimate boundary condition for (101) when (100a) is still in use as a differential equation to evolve $\mathbf{u}$: differential equations cannot also be used as boundary conditions. Consider, for example, the simple ordinary differential equation:

$$u' = 1 \tag{103}$$

whose solution is

$$u = x + c \tag{104}$$

where the unknown constant c is to be determined by a boundary condition. If we try to use as a boundary condition the evaluation of (103) at a boundary, for example

$$u'(x = 0) = 1 \tag{105}$$

we obtain from (104) the trivial equation

$$1 = 1 \tag{106}$$

Or consider $D^2 f = g$ in one dimension. Many boundary conditions will determine f uniquely, i.e. determine the coefficients of the null vectors 1 and x of D^2. But a boundary condition which is *not* adequate is:

$$D^2 f(\pm 1) = a_\pm \tag{107}$$

Condition (107) will yield an infinite number of solutions if $a_\pm = g(\pm 1)$ and no solutions otherwise. This situation is well known from linear algebra. Let A be a singular matrix, with an adjoint A^T which has null vector ϕ. Then

$$Af = g \text{ has a solution if and only if } \langle \phi, g \rangle = 0 \tag{108}$$

The rightwards implication is obvious, since

$$\langle \phi, g \rangle = \langle \phi, Af \rangle = \langle A^T \phi, f \rangle = 0 \tag{109}$$

but the leftwards implication is also true. In addition, the solution is not unique, since any multiple of a null vector of A can be added to f. The same reasoning holds when a differential operator is singular. When the boundary condition is derived from the differential operator, the resulting operator remains singular and has either no solution or an infinite number.

It turns out that replacing (100b) by (101) is legitimate, but that a proper complementary boundary condition to insure equivalence is

$$(\nabla \cdot \mathbf{u})|_{\text{bdy}} = 0 \tag{110}$$

This can be understood by reasoning that

$$\left.\begin{array}{c} (\partial_t - \nu\Delta) f = 0 \\ f|_{\text{bdy}} = 0 \\ f(t=0) = 0 \end{array}\right\} \Longrightarrow f = 0 \tag{111}$$

Hence

$$\left.\begin{array}{c} (\partial_t - \nu\Delta)(\nabla \cdot u) = \nabla \cdot (-\nabla p + \mathbf{f}) = 0 \\ (\nabla \cdot u)|_{\text{bdy}} = 0 \\ (\nabla \cdot u)(t=0) = 0 \end{array}\right\} \Longrightarrow \nabla \cdot u = 0 \tag{112}$$

Extensive discussions of the appropriate boundary conditions for the pressure Poisson equation can be found elsewhere, e.g. [5, 7].

We are thus faced again with differential equations (100a) and (101) which are coupled via a boundary condition (110). We may solve this coupled system by calculating in a preprocessing step homogeneous solutions $(p_i^{\text{hom}}, \mathbf{u}_i^{\text{hom}})$ which satisfy:

$$\Delta p_i^{\text{hom}} = 0 \tag{113a}$$

$$p_i^{\text{hom}}(\mathbf{x}_j) = \delta_{ij} \quad \mathbf{x}_j \in \text{boundary} \tag{113b}$$

$$(I - \epsilon\Delta)\mathbf{u}_i^{\text{hom}} = -\nabla p_i^{\text{hom}} \tag{113c}$$

$$\mathbf{u}_i^{\text{hom}}(\mathbf{x}_j) = \mathbf{0} \quad \mathbf{x}_j \in \text{boundary} \tag{113d}$$

and the influence matrix

$$C_{ji} \equiv \nabla \cdot \mathbf{u}_i^{\text{hom}}(\mathbf{x}_j) \quad \mathbf{x}_j \in \text{boundary} \tag{114}$$

Then, at each timestep, we calculate a particular solution

$$\Delta p^{\rm part} = \nabla \cdot \mathbf{f} \tag{115a}$$

$$p^{\rm part}(\mathbf{x}_j) = 0 \quad \mathbf{x}_j \in \text{boundary} \tag{115b}$$

$$(I - \epsilon\Delta)\mathbf{u}^{\rm part} = -\nabla p^{\rm part} + \mathbf{f} \tag{115c}$$

$$\mathbf{u}_i^{\rm hom}(\mathbf{x}_j) = \mathbf{0} \quad \mathbf{x}_j \in \text{boundary} \tag{115d}$$

and construct the exact solution as a superposition

$$\mathbf{u} = \mathbf{u}^{\rm part} + \sum_i \alpha_i \mathbf{u}_i^{\rm hom} \tag{116}$$

with the coefficients α_i determined by inverting the influence matrix C.

As in the streamfunction-vorticity case, the influence matrix C is composed of decoupled blocks corresponding to each Fourier mode and to each parity. The scaling of the algorithm, similar to that for the streamfunction-vorticity formulation, shows that the influence matrix method is not feasible for domains with more than two non-periodic directions, since the number of boundary points increases from 2 (line) to $2(N_x + N_y)$ (rectangle) to $2(N_xN_y + N_yN_z + N_xN_z)$ (three-dimensional rectangle).

We mention that a large number of CFD codes use erroneous boundary conditions such as (102). These codes do not insure that the flow at each step is incompressible. Typically, a timestep in a *fractional step* algorithm consists of three stages. First, the nonlinear terms are integrated explicitly. Secondly, a pressure-Poisson equation (101) is solved with boundary conditions like (102) and the resulting pressure gradient subtracted to impose incompressibility. Thirdly, the diffusive terms are integrated implicitly, with imposition of boundary conditions on the velocity. The velocity field will be divergence-free at the end of the second stage, and will satisfy the boundary conditions at the end of the third, but will then have a divergence which is of the order of a power of the timestep Δt. However, many such codes achieve satisfactory results nonetheless, in the sense that they agree empirically to some acceptable degree with experiments, with other codes, and/or with exact known solutions. One might argue that it is not necessary to achieve higher accuracy in the imposition of incompressibility (or the boundary conditions) than in the solution of the differential equation and that erroneous procedures are permissible if the error incurred does not exceed that of truncation. However, it is important to understand this source of error and to guard against pathological results resulting from it.

Even the influence matrix treatment we have described here does not yield exactly divergence-free velocity fields. This is because in deriving the pressure Poisson equation (101) from the Navier-Stokes equations, we have used the commutation of the vector Laplacian and the divergence to conclude that for an incompressible $\mathbf{u}$,

$$\nabla \cdot \boldsymbol{\Delta}\mathbf{u} = \Delta(\nabla \cdot \mathbf{u}) = 0 \tag{117}$$

This is true for the continuous operators or in periodic domains. But this property is violated when rows of the Laplacian have been replaced by boundary conditions. Instead, (101) should be replaced by

$$\Delta p = \nabla \cdot (\mathbf{f} + \boldsymbol{\tau}) \tag{118}$$

where $\boldsymbol{\tau}$ consists of coefficients of high-frequency modes related to the discarding of equations as in (18). Exact imposition of incompressibility to machine accuracy requires taking this *tau error* into account. This too can be done with Green's functions, but is too complicated to explain here.

Another method of treating the pressure is to use the vorticity formulation, a vector generalization of the two-dimensional streamfunction-vorticity formulation. By taking the full curl of the Navier-Stokes equations, one obtains an equation for the evolution of the vector vorticity:

$$\partial_t \boldsymbol{\omega} = \nu \Delta \boldsymbol{\omega} + \nabla \times \mathbf{f} \tag{119}$$

However, as before, the problem is that no boundary conditions exist for $\boldsymbol{\omega}$. A correct formulation exists, using Green's functions and the influence matrix, but incorrect formulations are often used in which conditions imposed on the vorticity are convenient but incorrect, or in which the vorticity is not equal to the curl of the velocity. We shall not further discuss this formulation.

9 Implication vs. Equivalence

Often a reformulation of the Navier-Stokes equations is arrived at by differentiation, for example taking the curl of the equations to eliminate the pressure, or the divergence to derive a pressure-Poisson equation. However, insufficient attention is paid to the question of whether these steps are reversible. That is, once a solution is obtained to the reformulated problem, is it the unique solution to the original problem? Here, we will state (and, in some cases, prove) the equivalence of two sets of equations which have general applications to hydrodynamics and magnetohydrodynamics.

The first result, equivalent to Stoke's Theorem and which we shall not demonstrate, is the following. In either two dimensions or three dimensions,

$$\mathbf{f} = \nabla p \quad \Leftrightarrow \quad \nabla \times \mathbf{f} = 0 \tag{120}$$

The rightwards implication is obvious, but the leftwards implication is also true.

The second result [3] is:

$$\mathbf{g} = 0 \quad \Leftrightarrow \quad \begin{cases} \mathbf{e} \cdot \mathbf{g} = 0 & \text{in } \Omega \\ \mathbf{e} \cdot \nabla \times \mathbf{g} = 0 & \text{in } \Omega \\ \nabla \cdot \mathbf{g} = 0 & \text{in } \Omega \\ \mathbf{n} \cdot \mathbf{g} = 0 & \text{on } \partial\Omega_h \end{cases} \tag{121}$$

where $\mathbf{e}$ is a unit vector, $\partial\Omega_h$ is the boundary of slices Ω_h of Ω perpendicular to $\mathbf{e}$, and $\mathbf{n}$ is the normal vector to the boundary, The rightwards implication of (121) is obvious. The leftwards implication of (121) is proved as follows:

$$\left.\begin{array}{r} 0 = \mathbf{e} \cdot \mathbf{g} \\ 0 = \mathbf{e} \cdot \nabla \times \mathbf{g} \end{array}\right\} \Rightarrow \left.\begin{array}{r} \mathbf{g} = \nabla_h p \\ \nabla \cdot \mathbf{g} = 0 \end{array}\right\} \Rightarrow \left.\begin{array}{l} \Delta_h p = 0 \\ \frac{\partial p}{\partial n} = 0 \text{ on } \partial\Omega_h \end{array}\right\} \Rightarrow p = p_0(e) \Rightarrow \mathbf{g} = 0$$

where the subscript h restricts differential operators to the directions perpendicular to $\mathbf{e}$, which we shall call horizontal, e denotes the coordinate corresponding to $\mathbf{e}$, which we call vertical, and the two-dimensional version of (120) has been used.

A third result is the existence of the vector potential and of the poloidal-toroidal decomposition for solenoidal fields:

$$\nabla \cdot \mathbf{B} = 0 \quad \Leftrightarrow \quad \mathbf{B} = \nabla \times \mathbf{A} = \nabla \times (\psi \mathbf{e} + \nabla \times \phi \mathbf{e}) \tag{122}$$

The leftward implication is obvious, but the rightward implication is also true. These results are used in the next section.

10 Influence Matrix: Poloidal-Toroidal Decomposition

The poloidal-toroidal decomposition generalizes to three dimensions the two-dimensional streamfunction-vorticity formulation We select a unit vector $\mathbf{e}$ in a distinguished direction which we term vertical and write:

$$\mathbf{u} = \nabla \times \psi\mathbf{e} + \nabla \times \nabla \times \phi\mathbf{e} \tag{123}$$

The advantage of (123) is that $\mathbf{u}$ is divergence-free by construction and involves only two scalar fields.

We take $\mathbf{e} \cdot \nabla\times$ and $\mathbf{e} \cdot \nabla \times \nabla\times$ of the Navier-Stokes equations, written as

$$(\partial_t - \nu\mathbf{\Delta})\mathbf{u} = -\nabla p + \mathbf{f} \tag{124}$$

to obtain evolution equations for the two scalar potentials:

$$(\partial_t - \nu\Delta)\Delta_h\psi = -\mathbf{e} \cdot \nabla \times \mathbf{f} \tag{125a}$$

$$(\partial_t - \nu\Delta)\Delta\Delta_h\phi = \mathbf{e} \cdot \nabla \times \nabla \times \mathbf{f} \tag{125b}$$

where Δ_h is the two-dimensional Laplacian in the directions perpendicular to $\mathbf{e}$, called the horizontal directions.

Equation (125a) for ψ is $2^{\rm nd}$ order in the vertical directions and $4^{\rm th}$ order in the horizontal directions. while equation (125b) for ϕ is $4^{\rm th}$ order in the vertical directions and $6^{\rm th}$ order in the horizontal directions. A corresponding number of boundary conditions are required: (2+4)/2=3 at each vertical boundary and (4+6)/2=5 at each horizontal boundary.

The three boundary conditions required at the vertical boundaries are merely those corresponding to the three velocity components. It is desirable for the horizontal directions to be periodic so that Δ_h takes a relatively simple form and so that there exist no horizontal boundaries at which conditions must be imposed. The poloidal-toroidal decomposition is commonly used in two such geometries: a spherical geometry with $\mathbf{e} = \mathbf{e}_\rho$, and a three-dimensional Cartesian geometry with one bounded direction along $\mathbf{e}$ and two periodic directions. It may also be used in a cylindrical geometry with periodic z and θ directions, with $\mathbf{e} = \mathbf{e_r}$.

Here we describe a more complicated application [4], that of a finite cylinder, with $\mathbf{e}$ chosen to be $\mathbf{e_z}$. One of the horizontal directions, θ, is periodic, but the other, r, is not. In this geometry, the horizontal boundary consists of $r = R$; the requirement of analyticity can be thought of as supplying an additional boundary condition at the origin (see **Polar coordinates**). Two major complications arise. Of the five boundary conditions required at horizontal boundaries, three are those corresponding to the three velocity components. However, two more need to be specified. These come from the operations of differentiation embodied in (122) and (125). The second complication is that most of the boundary conditions couple ψ and ϕ. After the coupling has been reduced, these conditions can be imposed via the influence matrix method.

For $\mathbf{e} = \mathbf{e_z}$, the decomposition (122) reads

$$u_r = \frac{1}{r}\partial_\theta\psi + \partial_z\partial_r\phi \tag{126a}$$

$$u_\theta = -\partial_r\psi + \frac{1}{r}\partial_z\partial_\theta\phi \tag{126b}$$

$$u_z = -\Delta_h\phi \tag{126c}$$

Equations (126) show that the scalar potentials ψ, ϕ are defined up to a *gauge*. That is, different choices of ψ, ϕ will result in the same $\mathbf{u}$. In particular, if

$$\begin{aligned} \Delta_h \phi_g &= 0 && (127a) \\ \nabla_h \psi_g &= \mathbf{e_z} \times \nabla(\partial_z \phi_g) && (127b) \end{aligned}$$

then $\psi + \psi_g$, $\phi + \phi_g$ will yield the same $\mathbf{u}$ as ϕ, ψ. In order to specify ϕ uniquely, we need to choose a solution to (127a). This can be done by specifying boundary conditions for ϕ, an obvious choice being homogeneous Dirichlet boundary conditions. Similarly, we need to choose a solution to (127b). A single value of ψ must be specified for each value of z, another obvious choice being $\psi(r = 0, \theta, z) = 0$. The *gauge conditions* we choose are thus:

$$\phi = 0 \text{ at } r = R \quad (128a)$$

$$\psi = 0 \text{ at } r = 0 \quad (128b)$$

Equation (128a) provides one of the two additional horizontal boundary conditions required; the role of (128b) will be clarified later.

The second additional horizontal boundary condition comes from (120)-(121). That is, we need to supplement (125), arrived at by differentiating (124), with a condition which plays the role of specifying a constant of integration. We have

$$\begin{aligned} (\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f} = -\nabla p \iff \\ \nabla \times ((\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f}) = 0 \end{aligned} \quad (129)$$

As stated in (121), this is in turn equivalent to the four conditions

$$\begin{aligned} \mathbf{e_z} \cdot (\nabla \times ((\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f}) &= 0 && (130a) \\ \mathbf{e_z} \cdot \nabla \times (\nabla \times ((\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f}) &= 0 && (130b) \\ \nabla \cdot (\nabla \times ((\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f}) &= 0 && (130c) \\ \mathbf{e_r} \cdot (\nabla \times ((\partial_t - \nu \mathbf{\Delta})\mathbf{u} - \mathbf{f}) &= 0 \qquad \text{at } r = R && (130d) \end{aligned}$$

Equations (130a)-(130b) are those we impose, (125a)-(125b), while (130c) is always satisfied. Equation (130d) is the remaining radial boundary condition required. The most problematic term in (130d) $\mathbf{e_r} \cdot \nabla \times \partial_t \mathbf{u}$, contains a time-derivative but involves only tangential derivatives of $\mathbf{u}$ at $r = R$, which do not evolve (in fact they are zero) when $\mathbf{u}(r = R) = 0$. The remaining terms are calculated to yield (when $\mathbf{f} = -(\mathbf{u} \cdot \nabla)\mathbf{u}$)

$$r \partial_r \partial_z \Delta_h \psi - \partial_\theta \Delta \Delta_h \phi = 0 \text{ at } r = R \quad (131)$$

Corresponding to (126), the velocity boundary conditions are

$$\begin{aligned} 0 = \frac{1}{r}\partial_\theta \psi + \partial_z \partial_r \phi \qquad &\text{at } r = R && (132a) \\ 0 = -\partial_r \psi + \frac{1}{r}\partial_z \partial_\theta \phi \qquad &\text{at } r = R && (132b) \\ 0 = -\Delta_h \phi \qquad &\text{at } r = R && (132c) \\ 0 = \frac{1}{r}\partial_\theta \psi + \partial_z \partial_r \phi \qquad &\text{at } z = \pm h && (132d) \\ 0 = -\partial_r \psi + \frac{1}{r}\partial_z \partial_\theta \phi \qquad &\text{at } z = \pm h && (132e) \\ 0 = -\Delta_h \phi \qquad &\text{at } z = \pm h && (132f) \end{aligned}$$

These conditions couple ψ and ϕ, but we will see that they can be decoupled to some extent.

Equations (128a),(130d), and (132a)-(132f) are the eight boundary conditions required. We now turn to the task of simplifying these as much as possible.

The gauge condition (128a) implies

$$\phi = 0 \Longrightarrow \partial_\theta \phi = \partial_z \phi = 0 \text{ at } r = R \tag{133a}$$

so (132b) becomes

$$0 = \partial_r \psi \text{ at } r = R \tag{133b}$$

This Neumann boundary condition is supplemented, as required, by the specification of a single value via the gauge condition (128b).

Condition (132d) at $z = \pm h$ can be rewritten

$$0 = r\mathbf{e_z} \cdot \nabla \times (\frac{1}{r}(-\psi \mathbf{e_r} + \partial_z \phi \mathbf{e}_\theta)) \Longrightarrow \frac{1}{r}(-\psi \mathbf{e_r} + \partial_z \phi \mathbf{e}_\theta) = \nabla_h f \Longrightarrow \left\{ \begin{array}{c} -\frac{1}{r}\psi = \partial_r f \\ \partial_z \phi = \partial_\theta f \end{array} \right. \tag{134a}$$

Condition (132e) then implies that at $z = \pm h$

$$0 = -\partial_r \psi + \frac{1}{r}\partial_z \partial_\theta \phi = \partial_r r \partial_r f + \frac{1}{r}\partial_\theta^2 \phi = r\Delta_h f \tag{134b}$$

Equations (133a) and (134a) imply

$$0 = \partial_z \phi = \partial_\theta f \text{ at } r = R \Longrightarrow f(R, \theta) = f_0 \tag{134c}$$

The solution to Laplace's equation (134b) with boundary condition (134c) is

$$f(r, \theta) = f_0 \tag{134d}$$

which, with (134a), implies

$$\psi = \partial_z \phi = 0 \text{ at } z = \pm h \tag{134e}$$

The velocity boundary conditions then become:

$$0 = \frac{1}{r}\partial_\theta \psi + \partial_z \partial_r \phi \qquad \text{at } r = R \tag{135a}$$

$$0 = \partial_r \psi \qquad \text{at } r = R \tag{135b}$$

$$0 = \Delta_h \phi \qquad \text{at } r = R \tag{135c}$$

$$0 = \psi \qquad \text{at } z = \pm h \tag{135d}$$

$$0 = \partial_z \phi \qquad \text{at } z = \pm h \tag{135e}$$

$$0 = \Delta_h \phi \qquad \text{at } z = \pm h \tag{135f}$$

There are still two coupled boundary conditions, (135a) and (131). Even without coupled boundary conditions, we have seen, for the 2D streamfunction-vorticity formulation, that the boundary conditions provided by the physics of the problem may not match the requirements for easy solution of the nested elliptic problems. When we solve (125a) and (125b) as nested Helmholtz, Poisson, and two-dimensional Poisson problems, ideally, for

$$(\partial_t - \nu\Delta)\Delta_h \psi = -\mathbf{e} \cdot \nabla \times \mathbf{f}$$

we would have boundary conditions on

$$\psi \qquad \text{at } r = R \tag{136a}$$

$$\Delta_h \psi \qquad \text{at } r = R, z = \pm h \tag{136b}$$

while for

$$(\partial_t - \nu\Delta)\Delta\Delta_h\phi = \mathbf{e} \cdot \nabla \times \nabla \times \mathbf{f}$$

we would have conditions on

$$\phi \qquad \text{at } r = R \tag{136c}$$

$$\Delta_h \phi \qquad \text{at } r = R, z = \pm h \tag{136d}$$

$$\Delta\Delta_h \phi \qquad \text{at } r = R, z = \pm h \tag{136e}$$

Although (135b) and (128b), (128a), and (135c)-(135f) provide boundary conditions for (136a), (136c), and (136d), the remaining equations (135a), (135d, (135e), and (131) do not fit so nicely. Instead, we formulate homogeneous and particular problems corresponding to tractable boundary conditions on (136a)-(136e). We calculate the influence matrix relating convenient but false Dirichlet boundary conditions on (136b) and (136e) to the actual unsatisfied boundary conditions (135a), (135d), (135e), and (131). The solution at each timestep is the appropriate superposition of a particular solution with the homogeneous solutions, or Green's functions.

References

[1] D. Gottlieb & S.A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, SIAM, 1977.

[2] D.B. Haidvogel & T.A. Zang, *Accurate solutions of Poisson's equation by expansion in Chebyshev polynomials.* J. Comp. Phys. **30**, 167–180 (1979).

[3] F. Marqués, *On boundary conditions for velocity potentials in confined flows: Application to Couette flow.* Phys. Fluids A **2** 729–737, (1990).

[4] F. Marqués, M. Net, J.M. Massaguer & I. Mercader, *Thermal convection in vertical cylinders. A method based on potentials of velocity.* Comput. Methods Appl. Mech. Engrg. **110**, 157–169 (1993).

[5] D. Rempfer, *On boundary conditions for incompressible Navier-Stokes problems.* Applied Mechanics Reviews, in press.

[6] J. Sherman & W.J. Morrison, Ann. Math. Statist. **20**, 621 (1949); **21**, 124 (1950).

[7] L.S. Tuckerman, *Divergence-free velocity fields in nonperiodic geometries.* J. Comput. Phys. **80**, 403–441 (1989).

[8] L.S. Tuckerman, *Transformations of matrices into banded form.* J. Comput. Phys., **84**, 360–376 (1989).

[9] M. Woodbury, Memorandum Report 42, Statistical Research Group, Princeton NJ (1950).