Some analytic expressions for symmetric tridiagonal and pentadiagonal matrices

In the course of learning a bit about compact finite differences, I discovered a couple of papers that provide analytic expressions for the determinants and inverses of a few specialized tridiagonal matrices. I wanted to understand this and see if I could reproduce it for symmetric, pentadiagonal matrices as well as generalized versions.

Consider the following $k \times k$, symmetric, tridiagonal matrix

$$M = \begin{pmatrix} D & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & D & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & D & 1 & \cdots & 0 & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 1 & D & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & D & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & D \end{pmatrix}$$

where D is a constant and sits only on the diagonal. We will refer to the determinant of this matrix as M_k. It can be shown that M_k satisfies the three term recursion relation

$$M_k = D\, M_{k-1} - M_{k-2}$$

with “initial conditions” $M_0 = 1$ and $M_1 = D$. This recursion relation can be solved by setting $M_k = r^k$ and finding the roots of

$$0 = r^k - D\, r^{k-1} + r^{k-2}$$

or

$$0 = r^2 - Dr + 1$$

which has solutions

$$r_\pm = \frac{1}{2}\, D \pm \frac{1}{2}\sqrt{D^2 - 4}$$

Thus, we can write for M_k

$$M_k = A\, r_+^k + B\, r_-^k$$

where, from the initial conditions,

$$A = \frac{D + \sqrt{D^2 - 4}}{2\sqrt{D^2 - 4}}$$

$$B = \frac{-D + \sqrt{D^2 - 4}}{2\sqrt{D^2 - 4}}$$

so

$$M_k = \frac{1}{2^{k+1}}\, \frac{1}{\sqrt{D^2 - 4}} \left[\left(D + \sqrt{D^2 - 4}\right)^{k+1} - \left(D - \sqrt{D^2 - 4}\right)^{k+1} \right]$$

Reparameterizing D, we set

$$D = \begin{cases} -2\cosh\lambda & D \le -2 \\ 2\,\cos\lambda & |D| < 2 \\ 2\cosh\lambda & D \ge 2 \end{cases}$$

It can then be shown that

$$M_k = \begin{cases} \dfrac{(-1)^k \sinh\big((k+1)\lambda\big)}{\sinh\lambda} & D \le -2 \\ \dfrac{\sin\big((k+1)\lambda\big)}{\sin\lambda} & |D| < 2 \\ \dfrac{\sinh\big((k+1)\lambda\big)}{\sinh\lambda} & D \ge 2 \end{cases}$$

With the determinant in hand, we can now try to find the exact form of the inverse of this matrix as well. Recall that the inverse can be given in terms of the cofactor expansion. More particularly, if we have a (nonsingular) $k \times k$ matrix, A, the inverse, A^{-1}, is given by $\mathrm{adj(A)}/\det(\mathrm{A})$ where adj(A) is the adjoint or adjugate matrix. The adjoint matrix, in turn, is the transpose of the cofactor matrix, C:

$$\mathrm{adj}(A) = C^T$$

and the cofactor matrix, C, is made from the cofactors of A. If the elements of A are given by a_{ij}, the elements, c_{ij}, of C are found by calculating the determinant of the $(k-1) \times (k-1)$ matrix formed by taking A and striking out the i^{th} row and j^{th} column (and multiplying by $(-1)^{i+j}$).

In our current case, one can show that the cofactor matrix is symmetric and that its elements can be written as

$$c_{ij} = (-1)^{i+j} M_{i-1} M_{k-j} \quad i \le j$$

Using the above relation for M_k, we have for the inverse of M

$$\begin{aligned}
(M^{-1})_{ij}\Big|_{i\le j} &= \frac{c_{ij}}{M_k} \\
&= (-1)^{i+j}\,\frac{M_{i-1}M_{k-j}}{M_k} \\
&= \begin{cases} -\dfrac{\sinh(i\lambda)\,\sinh[(k+1-j)\lambda]}{\sinh\lambda\,\sinh[(k+1)\lambda]} & D \le -2 \\ (-1)^{i+j}\,\dfrac{\sin(i\lambda)\,\sin[(k+1-j)\lambda]}{\sin\lambda\,\sin[(k+1)\lambda]} & |D| < 2 \\ (-1)^{i+j}\,\dfrac{\sinh(i\lambda)\,\sinh[(k+1-j)\lambda]}{\sinh\lambda\,\sinh[(k+1)\lambda]} & D \ge 2 \end{cases} \\
&= \begin{cases} -\dfrac{\cosh[(k+1-j+i)\lambda] - \cosh[(k+1-j-i)\lambda]}{2\,\sinh\lambda\,\sinh[(k+1)\lambda]} & D \le -2 \\ (-1)^{i+j+1}\,\dfrac{\cos[(k+1-j+i)\lambda] - \cos[(k+1-j-i)\lambda]}{2\,\sin\lambda\,\sin[(k+1)\lambda]} & |D| < 2 \\ (-1)^{i+j}\,\dfrac{\cosh[(k+1-j+i)\lambda] - \cosh[(k+1-j-i)\lambda]}{2\,\sinh\lambda\,\sinh[(k+1)\lambda]} & D \ge 2 \end{cases}
\end{aligned}$$

If we want the case that $i > j$, then we can simply swap i and j in the above expressions. However, to write a general expression, we have to recognize a subtlety associated with using finite ranges in i and j by invoking an absolute value, namely

$$(M^{-1})_{ij} = \begin{cases} -\dfrac{\cosh[(k+1-|j-i|)\lambda] - \cosh[(k+1-j-i)\lambda]}{2\,\sinh\lambda\,\sinh[(k+1)\lambda]} & D \le -2 \\ (-1)^{i+j+1}\,\dfrac{\cos[(k+1-|j-i|)\lambda] - \cos[(k+1-j-i)\lambda]}{2\,\sin\lambda\,\sin[(k+1)\lambda]} & |D| < 2 \\ (-1)^{i+j}\,\dfrac{\cosh[(k+1-|j-i|)\lambda] - \cosh[(k+1-j-i)\lambda]}{2\,\sinh\lambda\,\sinh[(k+1)\lambda]} & D \ge 2 \end{cases}$$

This, then, is the exact inverse of our tridiagonal matrix for any real value of D.

"Periodic" boundaries

Of course, the last matrix can be understood as the difference approximation of a derivative of a function. But the first and last rows will not be the same (or even correct) difference approximations. However, if we were to assume a function with periodic boundary conditions, our matrix would become

$$\tilde{M} = \begin{pmatrix} D & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \\ 1 & D & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & D & 1 & \cdots & 0 & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 1 & D & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & D & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & D \end{pmatrix}$$

This becomes a new matrix whose determinant is clearly related to M_k. Indeed, one can show that $\tilde{M}_k \equiv \det(\tilde{M})$ can be written as

$$\tilde{M}_k = DM_{k-1} - 2M_{k-2} + 2(-1)^{k+1}$$

Using our previous results, it can then be shown that

$$\tilde{M}_k = \begin{cases} 2(-1)^k\big[1 - \cosh(k\lambda)\big] & D \le -2 \\ 2\big[(-1)^{k+1} + \cos(k\lambda)\big] & |D| < 2 \\ 2\big[(-1)^{k+1} + \cosh(k\lambda)\big] & D \ge 2 \end{cases}$$

The cofactors needed for the inverse take the form

$$c_{ij}\Big|_{i \le j} = (-1)^{i+j}\Big[M_{i-1}M_{k-j} - M_{i-2}M_{k-1-j} + (-1)^k M_{j-i-1}\Big]$$

and the corresponding inverse can now be shown to be

$$(\tilde{M}^{-1})_{ij} = \begin{cases} \dfrac{(-1)^{k+1}}{2\,\sinh\lambda}\,\dfrac{\sinh[(k-|j-i|)\lambda] + \sinh[|j-i|\lambda]}{1-\cosh(k\lambda)} & D \le -2 \\ \dfrac{(-1)^{i+j}}{2\,\sin\lambda}\,\dfrac{\sin[(k-|j-i|)\lambda] + (-1)^k\,\sin[|j-i|\lambda]}{(-1)^{k+1}+\cos(k\lambda)} & |D| < 2 \\ \dfrac{(-1)^{i+j}}{2\,\sinh\lambda}\,\dfrac{\sinh[(k-|j-i|)\lambda] + (-1)^k\,\sinh[|j-i|\lambda]}{(-1)^{k+1}+\cosh(k\lambda)} & D \ge 2 \end{cases}$$

A pentadiagonal system

We want to extend the previous analysis to systems more complicated than a tridiagonal system. To begin, let us try with a $k \times k$, symmetric, pentadiagonal matrix of the form

$$N = \begin{pmatrix} D & a & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ a & D & a & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & a & D & a & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & a & D & a & 1 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & a & D & a & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & a & D & a & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & a & D & a \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & a & D \end{pmatrix}$$

where a and D are real numbers. Unsurprisingly, the determinant, $\det N \equiv N_k$, is more complicated. Doing the straightforward evaluation requires evaluating a new determinant, S_k, of a (secondary?) matrix given by

$$S = \begin{pmatrix} a & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ a & D & a & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & a & D & a & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & a & D & a & 1 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ & \vdots & & & \vdots & & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & a & D & a & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & a & D & a & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & a & D & a \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & a & D \end{pmatrix}$$

where, clearly, the first row and first column are different from the matrix N. The remainder of S, of course, is otherwise identical and the corresponding determinant, S_k, will be related to N_k. Indeed, with a little work, one can show that the recursion relation now takes the form of a coupled system

$$\begin{aligned} N_k &= D(N_{k-1} - N_{k-3}) - a(S_{k-1} - S_{k-2}) + N_{k-4} \\ S_k &= aN_{k-1} - S_{k-1} \end{aligned}$$

These can be uncoupled to yield

$$0 = N_k - (D-1)N_{k-1} - (D-a^2)N_{k-2} + (D-a^2)N_{k-3} + (D-1)N_{k-4} - N_{k-5}$$

and, surprisingly (at least to me), S_k satisfies exactly the same equation. We can solve this by substituting $N_k = r^k$ and rewriting the resulting fifth order polynomial in r:

$$0 = r^5 - (D-1)r^4 - (D-a^2)r^3 + (D-a^2)r^2 + (D-1)r - 1$$

There is, of course, no "quintic" formula, but we nonetheless would like to solve this. To that end, we first note that $r = 1$ is a solution of the equation. Dividing out this solution, we reduce the polynomial by one degree and get

$$0 = r^4 + (2-D)r^3 + (2-2D+a^2)r^2 + (2-D)r + 1$$

This looks daunting, but Mathematica is able to find solutions, namely four simple roots given by

$$\begin{aligned}
r_1 &= \frac{1}{4}\left[\alpha - \beta - \sqrt{\gamma - \delta}\right] \\
r_2 &= \frac{1}{4}\left[\alpha - \beta + \sqrt{\gamma - \delta}\right] \\
r_3 &= \frac{1}{4}\left[\alpha + \beta - \sqrt{\gamma + \delta}\right] \\
r_4 &= \frac{1}{4}\left[\alpha + \beta + \sqrt{\gamma + \delta}\right]
\end{aligned}$$

where

$$\begin{aligned}
\alpha &= D - 2 \\
\beta &= \sqrt{(D+2)^2 - 4a^2} \\
\gamma &= 2(D^2-4) - 4a^2 = \alpha^2 + \beta^2 - 16 \\
\delta &= 2(D-2)\sqrt{(D+2)^2 - 4a^2} = 2\alpha\beta
\end{aligned}$$

Curiously, one can show that $r_1 r_2 = 1$ and $r_3 r_4 = 1$, so (in an abuse of notation) we will set $r_1 = r$ and $r_3 = s$. In addition, we have the following observation. We have

$$\sqrt{\gamma \pm \delta} = \sqrt{(\alpha \pm \beta)^2 - 16}$$

which leads to an idea for a reparameterization of the constants as

$$\begin{aligned}
\alpha - \beta &= 4\cosh\sigma \\
\alpha + \beta &= 4\cosh\mu
\end{aligned}$$

which leads to

$$\begin{aligned}
\alpha &= 2(\cosh\sigma + \cosh\mu) \\
\beta &= 2(\cosh\mu - \cosh\sigma)
\end{aligned}$$

and

$$\begin{aligned}
r &= \cosh\sigma - \sinh\sigma = e^{-\sigma} \\
s &= \cosh\mu - \sinh\mu = e^{-\mu}
\end{aligned}$$

With this form, we can write the general solution for the determinant as

$$N_k = A + Br^k + Cr^{-k} + D's^k + Es^{-k}$$

where A, B, C, D' and E are constants that will depend on D and a and can be found from our "initial conditions" as a 5 dimensional system of equations:

$$
\begin{aligned}
A + B + C + D' + E &= N_0 = 1 \\
A + Br + Cr^{-1} + D's + Es^{-1} &= N_1 = D \\
A + Br^2 + Cr^{-2} + D's^2 + Es^{-2} &= N_2 = \begin{vmatrix} D & a \\ a & D \end{vmatrix} \\
A + Br^3 + Cr^{-3} + D's^3 + Es^{-3} &= N_3 = \begin{vmatrix} D & a & 1 \\ a & D & a \\ 1 & a & D \end{vmatrix} \\
A + Br^4 + Cr^{-4} + D's^4 + Es^{-4} &= N_4 = \begin{vmatrix} D & a & 1 & 0 \\ a & D & a & 1 \\ 1 & a & D & a \\ 0 & 1 & a & D \end{vmatrix}
\end{aligned}
$$

On inverting this for the coefficients, we can find lengthier expressions for them which, on simplification, leads to

$$
\begin{aligned}
N_k = {} & \frac{2rs}{(r-1)^2(s-1)^2} \\
& + \frac{1}{(r-1)^2(r+1)(r-s)(rs-1)} \Big\{ sr^2(1+r)\big(r^{k+2} + r^{-k-2}\big) - (1+s+s^2)r\big(r^{k-2} + r^{-k+2}\big) \cdot \\
& \qquad \big[r^4 + 1 + \frac{r}{s}(1+s)^2(1+r^2) + \frac{r^2}{s^2}\big(1 + 2s + 4s^2 + 2s^3 + s^4\big)\big] \Big\} \\
& - \frac{1}{(s-1)^2(s+1)(r-s)(rs-1)} \Big\{ rs^2(1+s)\big(s^{k+2} + s^{-k-2}\big) - (1+r+r^2)s\big(s^{k-2} + s^{-k+2}\big) \cdot \\
& \qquad \big[s^4 + 1 + \frac{s}{r}(1+r)^2(1+s^2) + \frac{s^2}{r^2}\big(1 + 2r + 4r^2 + 2r^3 + r^4\big)\big] \Big\}
\end{aligned}
$$

An important question surrounding any finite difference scheme is whether or not it is stable. The same holds for these compact finite difference schemes. We will attempt to sketch a possible proof of the stability of a CFD scheme though I suspect it will be highly dependent on the choice of boundary conditions.

To review a simple stability argument, consider the advection equation

$$u_{,t} = v u_{,x}$$

where v is the velocity of the wave. We will (semi-)discretize in space leaving the problem continuous in t: $u(t,x) \rightarrow u(t,x_i) \equiv u_i(t)$. A simple FD approximation becomes

$$u_{i,t} = \frac{v}{h}\big(u_i - u_{i-1}\big)$$

which, on Fourier transforming with $\tilde{u}(t,k) = \frac{1}{\sqrt{2\pi}}\int e^{-ikx}u(t,x)dx$ becomes

$$\tilde{u}_{,t}(t,k) = \frac{v}{h}(1 - e^{-ikh})\,\tilde{u}(t,k)$$

which can be integrated in t to give

$$\tilde{u}(t,k) = \tilde{u}(0,k)\,e^{ikvt\sin(kh)/kh}\;e^{vt(1-\cos(kh))/h}$$

for which all $k > 0$ modes grows exponentially in time for $v > 0$. Hence the scheme is unstable. To note a stable alternative, consider the centered FDA given by

$$u_{i,t} = \frac{v}{2h}\big(u_{i+1} - u_{i-1}\big)$$

(with $v > 0$) which, on Fourier transforming, and integrating in t yields

$$\tilde{u}(t,k) = \tilde{u}(0,k)\,e^{ikvt\sin(kh)/kh}$$

so that, for mode number k, the phase velocity becomes $v' = v\sin(kh)/kh$, and is different from the group velocity, $\partial(kv')/\partial k$. Thus, we have dispersion, but not unbounded growth. Hence this scheme is stable.[†]

If we apply this to a compact finite difference scheme for the above advection equation with periodic boundary conditions, we can write the semi-discrete equation as

$$u_{i,t} = v\,C_{ij}u_j(t)$$

where the matrix $C = A^{-1}B$ comes from our CFD scheme which approximates the derivative via

$$A_{ij}{u_j}' = B_{ik}u_k$$

and where the prime denotes the derivative with respect to x. If we take a three point stencil for both u and its derivative, in the interior of the grid, we would have something like

$$\alpha u'_{i-1} + u'_i + \alpha u'_{i+1} = \frac{a}{2h}(u_{i+1} - u_{i-1})$$

where α and a are chosen for to get as high an order of accuracy for the smallest stencil. In this case, if we would like an a $\mathcal{O}(h^4)$ accurate scheme, we would set $\alpha = 1/4$ and $a = 3/2$. As a result, the matrix A will

[†] A note worth making here is that the "left shifted" (Euler) FD approximation is unstable if $v > 0$ (so that we have a left moving wave) while the "right shifted" analog would be stable. Conversely, if $v < 0$ (right moving wave) the "left shifted" FD approximation would be stable and the "right shifted" FD approximation would be unstable. Of course, the centered FDA is stable for right or left moving waves, i.e. any value of v.

look like our earlier matrix M ($\tilde{M}$, really as we have periodic boundary conditions), i.e. $A = \alpha\tilde{M}$ where $D = 1/\alpha$. The (rescaled) matrix B_{ij} will then best be represented by ones and negative ones as

$$B_{ij} = \delta_{i,j-1} - \delta_{i,j+1} + \delta_{i,N-1}\delta_{j,0} - \delta_{i,0}\delta_{j,N-1}$$

where the indices i and j here take values from 0 to $N-1$ (we assume that we have a total of N points), the first term is for the first superdiagonal of B, the second term gives the first subdiagonal of B, the third term gives the upper rightmost element of B and the final terms is the lower leftmost element of B.

So, our full CFD scheme would look like

$$\alpha\tilde{M}_{ij}u'_j = \frac{a}{2h}B_{ij}u_j$$

On taking the inverse, we have

$$u'_i = \frac{a}{2h\alpha}(\tilde{M}^{-1})_{ij}B_{jk}u_k$$

such that our earlier matrix, C, becomes $C_{ik} = a/(2h\alpha)(\tilde{M}^{-1})_{ij}B_{jk}$. Using previous expressions, we can now write this as

$$C_{ij}u_j = \frac{a(4h\alpha\sinh\lambda)^{-1}}{(-1)^{N+1}+\cosh(N\lambda)}\sum_{j=0}^{N-1}(-1)^{i+j}\Big\{\sinh\big[(N-|j-i|)\lambda\big]+(-1)^N\sinh\big[|j-i|\lambda\big]\Big\}\big(u_{j+1}-u_{j-1}\big)$$

where $\lambda = \cosh^{-1}(1/2\alpha)$ and recall $\alpha = 1/4$ for the CFD scheme we are considering.

At this point, we can Fourier transform our semi-discrete differential equation and we see that we again get a $2i\sin(kh)\,\tilde{u}_j$. So, provided the stuff under the summation is real and positive definite, the integration of our ODE in time should now yield modes that propagate in the same direction, possibly with dispersion, but without exponential growth; hence the CFD scheme should be stable. To show this, note that the quantity under the sum can be written as

$$2\tilde{u}_j\,(-1)^{|j-i|}\begin{cases}\sinh\big(\frac{N\lambda}{2}\big)\cosh\Big[\big(\frac{N}{2}-|j-i|\big)\lambda\Big] & N \text{ even}\\ \cosh\big(\frac{N\lambda}{2}\big)\sinh\Big[\big(\frac{N}{2}-|j-i|\big)\lambda\Big] & N \text{ odd}\end{cases}$$

A compact finite difference (CFD) approx for first order derivatives à la Brady and Livescu

We are ultimately interested in following the Brady-Livescu procedure to derive possible stable compact finite derivative stencils for second derivatives. Their procedure, as applied to first derivatives, is similar to the standard method where interior stencils are found and shifted stencils are used for near-boundary points, the number of which points can be determined, in part, by the order of accuracy desired of the scheme.

We begin with the simple idea as found in the derivation of the interior stencil. In particular, we assume a linear coupling between the derivative of a function, $\phi(x)$, at a point x_i with neighboring values. We then equate this to a linear combination of the function values at similar points. An example of this would be

$$\beta\phi'_{i-2} + \alpha\phi'_{i-1} + \phi'_i + \alpha\phi'_{i+1} + \beta\phi'_{i+2} = \frac{a}{2h}\big(\phi_{i+1} - \phi_{i-1}\big) + \frac{b}{4h}\big(\phi_{i+2} - \phi_{i-2}\big)$$

where the parameters α, β, a and b are to be found based on conditions that we want the CFD approximation to satisfy; often the order of accuracy. Note that in this example we have used centered approximations so that the symmetry about the point x_i permits some nice simplifications. We now simply Taylor expand all the terms on both sides of this equation about the point x_i, collect according to powers of the spacing, h, and write down a set of equations that can provide some of our conditions on the above mentioned parameters. On doing this, we get the equations

$$\begin{aligned}
1 + 2\alpha + 2\beta &= a + b & &\mathcal{O}(h^2)\\
\frac{2}{2!}\big(\alpha + 2^2\beta\big) &= \frac{1}{3!}\big(a + 2^2 b\big) & &\mathcal{O}(h^4)\\
\frac{2}{4!}\big(\alpha + 2^4\beta\big) &= \frac{1}{5!}\big(a + 2^4 b\big) & &\mathcal{O}(h^6)\\
&\vdots\\
\frac{2}{(2n)!}\big(\alpha + 2^{2n}\beta\big) &= \frac{1}{(2n+1)!}\big(a + 2^{2n} b\big) & &\mathcal{O}(h^{2n+2})
\end{aligned}$$

where, to the right of each equation, we have placed the order of the accuracy of the scheme if that equation and all the others above it are used to solve for the parameters. Of course, in the present example, we have four parameters that are to be solved for. We need (at most) four conditions or equations. If we were to choose the first four equations in the set above, we would close the system, be able to solve for α, β, a and b and have a CFD scheme (in the interior) that was eighth order accurate. Alternatively, we could solve three equations for three of the four parameters and leave one unknown parameter which we could fix using some other condition.

If we wanted to go to yet higher order, we need only increase the size of the stencils on the left and/or right hand sides. For example, with the pattern hopefully clear, if we were to go to two additional points on both sides, we would have

$$\gamma\phi'_{i-3} + \beta\phi'_{i-2} + \alpha\phi'_{i-1} + \phi'_i + \alpha\phi'_{i+1} + \beta\phi'_{i+2} + \gamma\phi'_{i+3} = \frac{a}{2h}\big(\phi_{i+1} - \phi_{i-1}\big) + \frac{b}{4h}\big(\phi_{i+2} - \phi_{i-2}\big) + \frac{c}{6h}\big(\phi_{i+3} - \phi_{i-3}\big)$$

which, on Taylor expanding as before, results in the conditions

$$\begin{aligned}
1 + 2\alpha + 2\beta + 2\gamma &= a + b + c & &\mathcal{O}(h^2)\\
\frac{2}{2!}\big(\alpha + 2^2\beta + 3^2\gamma\big) &= \frac{1}{3!}\big(a + 2^2 b + 3^2 c\big) & &\mathcal{O}(h^4)\\
\frac{2}{4!}\big(\alpha + 2^4\beta + 3^4\gamma\big) &= \frac{1}{5!}\big(a + 2^4 b + 3^4 c\big) & &\mathcal{O}(h^6)\\
&\vdots\\
\frac{2}{(2n)!}\big(\alpha + 2^{2n}\beta + 3^{2n}\gamma\big) &= \frac{1}{(2n+1)!}\big(a + 2^{2n} b + 3^{2n} c\big) & &\mathcal{O}(h^{2n+2})
\end{aligned}$$

Thus we see that if we have n_L parameters on the left hand side, n_R parameters on the right hand side, and n_f undetermined parameters (by these conditions), we will have an $\mathcal{O}(h^{2(n_L+n_R-n_f+1)})$ accurate scheme in the interior.

A fly in this nice ointment is that we often have grid boundaries for which the interior stencils will be incorrect or unusable. So there will need to be specialized, usually shifted, FD approximations used on and near boundary points. This will necessitate new parameters and hence new equations for those parameters. An example of a CFD first derivative valid at a (left) boundary point, x_i, would be a totally shifted expression such as

$$\alpha_0 \phi'_i + \alpha_1 \phi'_{i+1} + \alpha_2 \phi'_{i+2} + \cdots = \frac{1}{h}\big(a_0\phi_i + a_1\phi_{i+1} + a_2\phi_{i+2} + \cdots\big)$$

Note that the symmetry of the previous centered approximations is lost so a larger stencil (i.e. more parameters) will be required to obtain a similar order of accuracy. Of course, additional considerations may also come into play. On the left hand side, to preserve the tri- (or penta-)diagonality of the overall matrix operator, we would need to set $\alpha_i = 0$ for $i > 1$ ($i > 2$). In addition, if we require that the scaling of the overall system be such that the first diagonal element of the matrix operator agree with the diagonal elements of the interior stencil, we might include the condition $\alpha_0 = 1$. Doing this for a tridiagonal system, we can reproduce what Brady and Livescu do for their construction of first derivative CFD operators. Doing the Taylor expansions of the resulting expression about x_i and collecting terms, we find

$$\begin{aligned}
0 = a_0 + a_1 + a_2 + \cdots &= \sum_{k=0}^{n_R-1} a_k & \mathcal{O}(h^0)\\
1 + \alpha_1 = a_1 + 2a_2 + 3a_3 + \cdots &= \sum_{k=0}^{n_R-1} k\, a_k & \mathcal{O}(h^1)\\
\alpha_1 = \frac{1}{2!}\big(a_1 + 2^2 a_2 + 3^2 a_3 + \cdots\big) &= \frac{1}{2!}\sum_{k=0}^{n_R-1} k^2\, a_k & \mathcal{O}(h^2)\\
&\vdots\\
\frac{\alpha_1}{(n-1)!} = \frac{1}{n!}\big(a_1 + 2^n a_2 + 3^n a_3 + \cdots\big) &= \frac{1}{n!}\sum_{k=0}^{n_R-1} k^n\, a_k & \mathcal{O}(h^n)
\end{aligned}$$

where n_R is, again, the number of parameters on the right hand side. We have a total of $n_L + n_R$ parameters where $n_L = 1$ such that α_1 is the sole parameter on the left hand side to ensure the left hand matrix is tridiagonal. If we choose these conditions in order to close the system for all our parameters, we will have an n^{th} order accurate scheme with $n = n_R$. Alternatively, if we leave n_f of the parameters free, we will have a scheme that is $n_R - n_f$ order accurate.

If, instead of a totally shifted stencil, we shift only partially, we might have something like

$$\beta_{-1}\phi'_{i-1} + \beta_0\phi'_i + \beta_1\phi'_{i+1} + \cdots = \frac{1}{h}\big(b_{-1}\phi_{i-1} + b_0\phi_i + b_1\phi_{i+1} + \cdots\big)$$

Expanding around x_i and collecting, we find the following equations for a strictly tridiagonal system on the left hand side

$$0 = b_{-1} + b_0 + b_1 + b_2 + \cdots = \sum_{k=-1}^{n_R-2} b_k \qquad \mathcal{O}(h^0)$$

$$\beta_{-1} + \beta_0 + \beta_1 = -b_{-1} + b_1 + 2b_2 + 3b_3 + \cdots = \sum_{k=-1}^{n_R-2} k\, b_k \qquad \mathcal{O}(h^1)$$

$$-\beta_{-1} + \beta_1 = \frac{1}{2!}\big(b_{-1} + b_1 + 2^2 b_2 + 3^2 b_3 + \cdots\big) = \frac{1}{2!}\sum_{k=-1}^{n_R-2} k^2\, b_k \qquad \mathcal{O}(h^2)$$

$$\vdots$$

$$\frac{1}{(n-1)!}\big((-1)^{n+1}\beta_{-1} + \beta_1\big) = \frac{1}{n!}\big((-1)^n b_{-1} + b_1 + 2^n b_2 + 3^n b_3 + \cdots\big) = \frac{1}{n!}\sum_{k=-1}^{n_R-2} k^n\, b_k \qquad \mathcal{O}(h^n)$$

For this system, we have $n_L + n_R$ total parameters with $n_L = 3$ (or 2 if we choose an overall scaling such that $\beta_0 = 1$). If we have n_f free parameters, we will have a scheme that is $n_L + n_R - n_f - 1$ order accurate.

Playing a similar game for another partially shifted stencil, we can write

$$\gamma_{-1}\phi'_{i-1} + \gamma_0\phi'_i + \gamma_1\phi'_{i+1} + \cdots = \frac{1}{h}\big(c_{-2}\phi_{i-2} + c_{-1}\phi_{i-1} + c_0\phi_i + c_1\phi_{i+1} + \cdots\big)$$

Again, expanding around x_i, rescaling so that $\gamma_0 = 1$, and collecting we get

$$0 = c_{-2} + c_{-1} + c_0 + c_1 + \cdots = \sum_{k=-2}^{n_R-3} c_k \qquad \mathcal{O}(h^0)$$

$$\gamma_{-1} + 1 + \gamma_1 = -2c_{-2} - c_{-1} + c_1 + 2c_2 + \cdots = \sum_{k=-2}^{n_R-3} k\, c_k \qquad \mathcal{O}(h^1)$$

$$-\gamma_{-1} + \gamma_1 = \frac{1}{2!}\big(2^2 c_{-2} + c_{-1} + c_1 + 2^2 c_2 + \cdots\big) = \frac{1}{2!}\sum_{k=-2}^{n_R-3} k^2\, c_k \qquad \mathcal{O}(h^2)$$

$$\vdots$$

$$\frac{1}{(n-1)!}\big((-1)^{n+1}\gamma_{-1} + \gamma_1\big) = \frac{1}{n!}\big((-1)^n c_{-2} + (-1)^n c_{-1} + c_1 + 2^n c_2 + \cdots\big) = \frac{1}{n!}\sum_{k=-2}^{n_R-3} k^n\, c_k \qquad \mathcal{O}(h^n)$$

and this would give us the third row of our boundary-appropriate matrix operators.

Brady and Livescu include one more shifted stencil of the form

$$\delta_{-1}\phi'_{i-1} + \delta_0\phi'_i + \delta_1\phi'_{i+1} + \cdots = \frac{1}{h}\big(d_{-3}\phi_{i-3} + d_{-2}\phi_{i-2} + d_{-1}\phi_{i-1} + d_0\phi_i + \cdots\big)$$

which yields (rescaling yet again so that $\delta_0 = 1$) the equations

$$0 = \sum_{k=-3}^{n_R-4} d_k \qquad \mathcal{O}(h^0)$$

$$\delta_{-1} + 1 + \delta_1 = \sum_{k=-3}^{n_R-4} k\, c_k \qquad \mathcal{O}(h^1)$$

$$-\delta_{-1} + \delta_1 = \frac{1}{2!}\sum_{k=-3}^{n_R-4} k^2\, c_k \qquad \mathcal{O}(h^2)$$

$$\vdots$$

$$\frac{1}{(n-1)!}\big((-1)^{n+1}\delta_{-1} + \delta_1\big) = \frac{1}{n!}\sum_{k=-3}^{n_R-4} k^n\, c_k \qquad \mathcal{O}(h^n)$$

So far, all of these shifted CFD approximations have been for shifts to the right and are appropriate for left boundaries. We should note that for right boundaries, we will need to use stencils shifted in the other direction. These will result in analogous equations for the paremeters, but with some extra sign changes. For example, for the completely left shifted stencil appropriate to the rightmost boundary point, we might have

$$\cdots + \alpha_{-2}\phi'_{i-2} + \alpha_{-1}\phi'_{i-1} + \alpha_0\phi'_i = \frac{1}{h}\big(\cdots + a_{-2}\phi_{i-2} + a_{-1}\phi_{i-1} + a_0\phi_i\big)$$

and which would result in the following conditions (assuming we maintain, again, tridiagonality and diagonal element $\alpha_0 = 1$)

$$\begin{aligned}
0 &= \sum_{k=0}^{-(n_R-1)} a_k && \mathcal{O}(h^0)\\
\alpha_{-1} + 1 &= \sum_{k=0}^{-(n_R-1)} k\, a_k && \mathcal{O}(h^1)\\
-\alpha_{-1} &= \frac{1}{2!}\sum_{k=0}^{-(n_R-1)} k^2\, a_k && \mathcal{O}(h^2)\\
&\vdots\\
(-1)^{n-1}\frac{\alpha_1}{(n-1)!} &= \frac{1}{n!}\sum_{k=0}^{-(n_R-1)} k^n\, a_k && \mathcal{O}(h^n)
\end{aligned}$$

These equations are clearly close to those for the completely right-shifted stencil. Indeed, if one solves these equations, one finds that the solutions are the same as in the right-shifted case, but that the parameters on the right hand side flip signs. The parameters on the left hand side remain the same. The same holds true for other shifted stencils.

To conclude, we construct some example matrices that might get used for these first order derivatives. We consider a vector of function values, ϕ_i and their derivatives, ϕ'_i with i ranging between 0 and N. The CFD is written as

$$A_{ij}\phi'_j = B_{ij}\phi_j$$

where A and B are banded matrices. Let us pick an order of accuracy such that in the interior we have 6$^{\text{th}}$ order and the boundary stencils are 5$^{\text{th}}$ order. We will also choose A to be strictly tridiagonal. If we use our previous notation, these matrices become

$$A = \begin{pmatrix}
1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\beta_{-1} & 1 & \beta_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & \gamma_{-1} & 1 & \gamma_1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \delta_{-1} & 1 & \delta_1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \alpha & 1 & \alpha & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & \alpha & 1 & \alpha & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
 & & & \vdots & & & & \vdots & & & & \vdots & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \alpha & 1 & \alpha & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \alpha & 1 & \alpha & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \delta_{-1} & 1 & \delta_1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \gamma_{-1} & 1 & \gamma_1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \beta_{-1} & 1 & \beta_1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \alpha_1 & 1
\end{pmatrix}$$

$$
B = \begin{pmatrix}
a_0 & a_1 & a_2 & a_3 & a_4 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{-1} & b_0 & b_1 & b_2 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_{-2} & c_{-1} & c_0 & c_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
d_{-3} & d_{-2} & d_{-1} & d_0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a & 0 & a & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a & 0 & a & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & \vdots & & & & & \vdots & & & & \vdots & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & a & 0 & a & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & a & 0 & a & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & -d_0 & -d_{-1} & -d_{-2} & -d_{-3} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & -c_1 & -c_0 & -c_{-1} & -c_{-2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & -b_2 & -b_1 & -b_0 & -b_{-1} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & -a_4 & -a_3 & -a_2 & -a_1 & -a_0
\end{pmatrix}
$$

Note that we choose to take all four (first and last) rows of B for near-boundary stencils. This follows Brady and Livescu who generalize this to having free parameters.

A compact finite (second order) derivative à la Brady and Livescu

We can now construct second derivatives in an analogous way. We begin with a interior problem with a (symmetric) formula

$$\beta\phi''_{i-2}+\alpha\phi''_{i-1}+\phi''_{i}+\alpha\phi''_{i+1}+\beta\phi''_{i+2} = \frac{a}{h^2}\big(\phi_{i+1}-2\phi_i+\phi_{i-1}\big)+\frac{b}{(2h)^2}\big(\phi_{i+2}-2\phi_i+\phi_{i-2}\big)+\frac{c}{(3h)^2}\big(\phi_{i+3}-2\phi_i+\phi_{i-3}\big)$$

Expanding in the usual way, we find the conditions on our parameters

$$\begin{aligned}
1+2(\alpha+\beta) &= a+b && \mathcal{O}(h^2)\\
\frac{2}{2!}\big(\alpha+2^2\beta\big) &= \frac{2}{4!}\big(a+2^2b+3^2c\big) && \mathcal{O}(h^4)\\
\frac{2}{4!}\big(\alpha+2^4\beta\big) &= \frac{2}{6!}\big(a+2^4b+3^4c\big) && \mathcal{O}(h^6)\\
&\vdots\\
\frac{2}{(2n!}\big(\alpha+2^{2n}\beta\big) &= \frac{2}{(2n+2)!}\big(a+2^{2n}b+3^{2n}c\big) && \mathcal{O}(h^{2n+2})
\end{aligned}$$

This could clearly be generalized to n_L parameters on the left hand side and n_R parameters on the right hand side.[‡] On closing this system with n_L+n_R conditions, we can have a scheme of $\mathcal{O}(h^{2(n_L+n_R)})$ accuracy. If we choose to leave one or more free parameters, the scheme's order of accuracy would then go down by two for each free parameter.

Consider now shifted stencils for such a second derivative operator, in particular, begin with the totally right shifted formula

$$\alpha_0\phi''_i+\alpha_1\phi''_{i+1}+\alpha_2\phi''_{i+2}+\cdots = \frac{1}{h^2}\big(a_0\phi_i+a_1\phi_{i+1}+a_2\phi_{i+2}+\cdots\big)$$

The equations, on expansion, become

$$\begin{aligned}
0 &= \sum_{k=0}^{n_R-1} a_k && \mathcal{O}(\frac{1}{h})\\
0 &= \sum_{k=0}^{n_R-1} k\,a_k && \mathcal{O}(h^0)\\
\sum_{k=0}^{n_L-1}\alpha_k &= \frac{1}{2!}\sum_{k=0}^{n_R-1} k^2\,a_k && \mathcal{O}(h)\\
\frac{1}{1!}\sum_{k=0}^{n_L-1} k\,\alpha_k &= \frac{1}{3!}\sum_{k=0}^{n_R-1} k^3\,a_k && \mathcal{O}(h^2)\\
&\vdots\\
\frac{1}{(n-1)!}\sum_{k=0}^{n_L-1} k^{n-1}\,\alpha_k &= \frac{1}{(n+1)!}\sum_{k=0}^{n_R-1} k^{n+1}\,a_k && \mathcal{O}(h^n)
\end{aligned}$$

where to get n^{th} order accuracy we must have $n = n_L+n_R-n_f-2$. If we choose a tridiagonal system so that $\alpha_k = 0$ for $k>1$ and the diagonal element is scaled to one: $\alpha_0 = 1$, this requires that the number of parameters on the right hand side be $n_R = n+n_f+1$, i.e. one more than the desired order of accuracy

[‡] Note that these equations do not agree with those for second order derivative CFDs in Johnathan Tyler's thesis, Eq. 2.13–2.18. I believe the equations above are correct.

on the boundary *plus* the number of free parameters we might want (and which, of course, are used in the Brady-Livescu approach).

Now consider the partially right shifted tridiagonal formula

$$\beta_{-1}\phi''_{i-1} + \beta_0\phi''_i + \beta_1\phi''_{i+1} + \cdots = \frac{1}{h^2}\big(b_{-1}\phi_{i-1} + b_0\phi_i + b_1\phi_{i+1} + b_2\phi_{i+2}\cdots\big)$$

If we again choose the diagonal element on the left to be rescaled to one, $\beta_0 = 1$, we find

$$\begin{aligned}
0 &= \sum_{k=-1}^{n_R-2} b_k && \mathcal{O}(\frac{1}{h}) \\
0 &= \sum_{k=-1}^{n_R-2} k\, b_k && \mathcal{O}(h^0) \\
\beta_{-1} + 1 + \beta_1 &= \frac{1}{2!}\sum_{k=-1}^{n_R-2} k^2\, b_k && \mathcal{O}(h) \\
\frac{1}{1!}\big(-\beta_{-1} + \beta_1\big) &= \frac{1}{3!}\sum_{k=-1}^{n_R-2} k^3\, b_k && \mathcal{O}(h^2) \\
&\vdots \\
\frac{1}{(n-1)!}\big((-1)^{n-1}\beta_{-1} + \beta_1\big) &= \frac{1}{(n+1)!}\sum_{k=-1}^{n_R-2} k^{n+1}\, b_k && \mathcal{O}(h^n)
\end{aligned}$$

For this tridiagonal system, being n^{th} order accurate again requires that $n_R = n + n_f$.

Maintaining tridiagonality, we have two more right shifted formulas to consider. The first is

$$\gamma_{-1}\phi''_{i-1} + \gamma_0\phi''_i + \gamma_1\phi''_{i+1} + \cdots = \frac{1}{h^2}\big(c_{-2}\phi_{i-2} + c_{-1}\phi_{i-1} + c_0\phi_i + c_1\phi_{i+1} + \cdots\big)$$

Again, taking $\gamma_0 = 1$, we find

$$\begin{aligned}
0 &= \sum_{k=-2}^{n_R-3} c_k && \mathcal{O}(\frac{1}{h}) \\
0 &= \sum_{k=-2}^{n_R-3} k\, c_k && \mathcal{O}(h^0) \\
\gamma_{-1} + 1 + \gamma_1 &= \frac{1}{2!}\sum_{k=-2}^{n_R-3} k^2\, c_k && \mathcal{O}(h) \\
\frac{1}{1!}\big(-\gamma_{-1} + \gamma_1\big) &= \frac{1}{3!}\sum_{k=-2}^{n_R-3} k^3\, c_k && \mathcal{O}(h^2) \\
&\vdots \\
\frac{1}{(n-1)!}\big((-1)^{n-1}\gamma_{-1} + \gamma_1\big) &= \frac{1}{(n+1)!}\sum_{k=-2}^{n_R-3} k^{n+1}\, c_k && \mathcal{O}(h^n)
\end{aligned}$$

The last shifted formula is

$$\delta_{-1}\phi''_{i-1} + \delta_0\phi''_i + \delta_1\phi''_{i+1} + \cdots = \frac{1}{h^2}\big(d_{-3}\phi_{i-3} + d_{-2}\phi_{i-2} + d_{-1}\phi_{i-1} + d_0\phi_i + \cdots\big)$$

With $\delta_0 = 1$, we find

$$
\begin{aligned}
0 &= \sum_{k=-3}^{n_R-4} d_k && \mathcal{O}(\frac{1}{h}) \\
0 &= \sum_{k=-3}^{n_R-4} k\, d_k && \mathcal{O}(h^0) \\
\delta_{-1} + 1 + \delta_1 &= \frac{1}{2!} \sum_{k=-3}^{n_R-4} k^2\, d_k && \mathcal{O}(h) \\
\frac{1}{1!}\big(-\delta_{-1} + \delta_1\big) &= \frac{1}{3!} \sum_{k=-3}^{n_R-4} k^3\, d_k && \mathcal{O}(h^2) \\
&\vdots \\
\frac{1}{(n-1)!}\big((-1)^{n-1}\delta_{-1} + \delta_1\big) &= \frac{1}{(n+1)!} \sum_{k=-3}^{n_R-4} k^{n+1}\, d_k && \mathcal{O}(h^n)
\end{aligned}
$$

In this and the previous, we again have $n_R = n + n_f$ for an n^{th} order accurate second derivative.