

**Diagonal Dominance and Positive Definiteness of
Upwind Approximations for Advection Diffusion
Problems**

G. Golub

Department of Computer Science, Stanford University, CA

D. Silvester

Department of Mathematics, UMIST, Manchester, M60 1QD

A. Wathen

Oxford University

Numerical Analysis Group

We dedicate this paper to Professor Ron Mitchell who has been an inspiration to many generations of numerical analysts. In addition his generous spirit has infused our discipline with a sense of positive endeavour.

We examine whether three different upwind schemes for the positive definite advection diffusion problem can yield indefinite coefficient matrices. In particular, in line with the increasing use of adaptive meshes, we are concerned with discretisations on arbitrary grids but only in one dimension. We show that indefinite coefficient matrices can arise from certain approaches: this can present difficulties with efficient iterative solution techniques which might be required for corresponding approximations in higher dimensions.

Key words and phrases: advection-diffusion problems, diagonal dominance, positive definiteness, upwinding

Oxford University Computing Laboratory
Numerical Analysis Group
Wolfson Building
Parks Road
Oxford, England OX1 3QD
E-mail: wathen@comlab.oxford.ac.uk

February, 1996

Contents

1	Introduction	3
2	Symmetry and skew-symmetry	4
3	Finite Element and Finite Volume methods	6
4	Conclusions	7

1 Introduction

The numerical approximation of advection-diffusion operators continues to provide a significant challenge for Numerical Analysts. Many schemes have been proposed in the framework of finite differences, finite volumes and finite elements (see for example [6]). A common theme is the use of upwinding: approximations which are biased to the direction of advection. A common goal is to achieve accurate (and possibly also monotone) solutions over a range of Peclet or mesh Peclet numbers. In this paper we examine another aspect of the problem: the structure of the discrete (matrix) equations arising from alternative approximation techniques.

The simplest model problem is in one dimension:

$$-u'' + \sigma u' = 0 \quad , \quad x \in [0, 1], \quad u(0) = 1, \quad u(1) = 0, \quad (1.1)$$

where σ is a positive constant. It is well known that for large σ the solution of this equation develops a boundary layer of thickness $O(1/\sigma)$.

The issue we address is of the qualitative approximation of the advection-diffusion operator by a discrete approximation. In particular we are concerned with whether or not the positive definiteness of the continuous differential operator is mirrored by various discretisation schemes. This is a significant issue when iterative solution methods are employed for the resulting linear systems, since certain methods are convergent only for systems which are positive definite:

$$x^T A x > 0 \quad \text{for nonzero real vectors } x$$

(see for example [1]). The related but different property of diagonal dominance:

$$|a_{i,i}| \geq \sum_j |a_{i,j}| \quad \text{for each } i$$

with strict inequality in at least one row is a sufficient condition for the convergence of most simpler fixed point iterations such as the Jacobi and Gauss-Seidel iterations. The issue of reducibility which rarely arises in differential equation applications is also a general consideration - see [12]. Tridiagonal matrices with non-zero sub- and super-diagonal entries are certainly irreducible and we will only deal with such matrices here without further mention of this issue. The close connection between diagonally dominant and positive definite matrices (through scaling) was established by Tartar [11]. In some situations the related M-matrix property is also useful (for example for preconditioning: see [5]).

In the one-dimensional situation many solution methods are applicable, but in particular in three-dimensional problems the applicability and rapid convergence of iterative methods is the central practical issue. In general a discretisation which is not positive definite in one-dimension will also not be in higher dimensions. We will thus consider only the model problem (1.1) for simplicity with the understanding that corresponding discretisations in higher dimensions will share similar qualitative features.

2 Symmetry and skew-symmetry

Multiplying the differential equation (1.1) by an appropriate test function v which vanishes at the end points of the domain and integrating employing integration by parts on the diffusion (second derivative) term yields

$$\langle u', v' \rangle + \langle \sigma u', v \rangle = 0$$

where $\langle \cdot, \cdot \rangle$ is the L_2 inner product. Now it is apparent that the first term is symmetric and positive definite whereas the second term is skew-symmetric at least if σ' is zero (the corresponding property in higher dimensions would be $\nabla \cdot \sigma = 0$ which for example in incompressible fluid dynamics corresponds to conservation of mass).

Many approximations respect this structure: for example if the Galerkin method is employed using a conforming approximation space $V_h = \text{span}\{\phi_1, \phi_2, \dots, \phi_n\}$ then discrete equations result of the form

$$A\mathbf{u} + C\mathbf{u} = \mathbf{f}$$

where $A = \{a_{i,j}\}$, $a_{i,j} = \langle \phi'_j, \phi'_i \rangle$ is a symmetric and positive definite matrix, $C = \{c_{i,j}\}$, $c_{i,j} = \langle \sigma \phi'_j, \phi'_i \rangle$ is skew-symmetric and \mathbf{f} arises from the inhomogeneous boundary term. Thus, for example, for a Galerkin spectral approximation or a Galerkin finite element approximation on any grid, the association of the symmetric part of the discretised matrix with the self-adjoint part of the continuous problem and correspondingly the skew-symmetric part with the skew-adjoint part of the differential operator holds true. With appropriate scaling, the elementary central finite difference approximation also shares this property - see below.

Much emphasis has however been put on preserving diagonal dominance of discretisations rather than ensuring positive definiteness of the symmetric part. In many situations, diagonal dominance corresponds to the existence of a discrete maximum principle (which may be useful in the suppression of oscillations in the discrete solution - but see [2]). In all cases diagonal dominance implies that all of the eigenvalues lie in the right half plane by simple application of the Gershgorin theorem, but it does not follow that the symmetric part and thus the matrix itself is positive definite. A simple example is illustrative:

Using the simplest finite differences on a grid of variable spacing

$$\dots, h = x_j - x_{j-1}, k = x_{j+1} - x_j, \ell = x_{j+2} - x_{j+1}, \dots$$

the second derivative term $-u''$ might be replaced by

$$-\left(\frac{u_{j+1} - u_j}{k} - \frac{u_j - u_{j-1}}{h}\right) / \left(\frac{1}{2}(h + k)\right)$$

and the first derivative could either be approximated by a central difference

$$\sigma \frac{(u_j + u_{j+1})/2 - (u_j + u_{j-1})/2}{\frac{1}{2}(h + k)} = \sigma \frac{u_{j+1} - u_{j-1}}{h + k}$$

or an upwind difference

$$\sigma \frac{u_j - u_{j-1}}{h}.$$

Thus, after division by the common factor $1/(h+k)$, the coefficient matrix is of the form

$$A = \text{tri} \left(-\sigma - \frac{2}{h}, \frac{2}{k} + \frac{2}{h}, \sigma - \frac{2}{k} \right)$$

in the central difference case. The condition for diagonal dominance (which is the same as the condition that A be an M-matrix) is the well-known mesh Peclet number condition $\sigma h/2 \leq 1$. (One would have to take the maximal h to satisfy dominance in every row of the matrix). However, the symmetric part of the matrix comes only from the second derivative and is diagonally dominant and hence positive definite because of the Gershgorin theorems (see [12]).

Turning to the upwind difference, we obtain the matrix

$$A = \text{tri} (-\sigma(h+k)/h - 2/h, 2/k + 2/h + \sigma(h+k)/h, -2/k)$$

which is diagonally dominant for every σ , h and k (i.e. for any mesh). But considering the symmetric part we have

$$(A + A^T)/2 = \text{tri} (-\sigma(h+k)/2h - 2/h, 2/k + 2/h + \sigma(h+k)/h, -\sigma(k+\ell)/2k - 2/k)$$

which is diagonally dominant only in rows for which $\ell h \leq k^2$. Many practical meshes that might be employed for this simple problem could satisfy this condition: for example a smoothly graded mesh with $k = \alpha h, \ell = \alpha^2 h$ would yield a diagonally dominant matrix (with the strict inequalities holding in the first and last row). However it is clear that there are reasonable meshes (such as one with a regular mesh spacing in one part of the domain and a different regular spacing elsewhere) which do not satisfy this condition in every row. This still does not imply indefiniteness or otherwise. However a specific case shows that indefiniteness is possible: for the simple mesh $0, 0.5, 0.51, 0.52, 1$ with $\sigma = 10$ it is readily checked that

$$H := (A + A^T)/2 = \begin{pmatrix} 214.2 & -210 & 0 \\ -210 & 420 & -445 \\ 0 & -445 & 694.166^r \end{pmatrix}.$$

This matrix has determinant equal to -10579695 and thus has a negative eigenvalue. This is a coarse and not very suitable mesh for the given problem, however for meshes with many more points but similarly clustered around 0.5 we have similarly observed indefiniteness of the symmetric part. For related but more complicated problems an interior layer at 0.5 might be expected: a mesh of the given form might be reasonable in such a situation. Here it is the non-monotonic change in mesh size which is necessary to give indefiniteness: for a monotonically

graded mesh this could not occur (see [6]). In higher dimensions and for example with mesh adaptivity such monotonicity may not be so easy to guarantee.

Of relevance to the matrix theory for this problem, we note that H certainly has positive diagonal entries and non-positive off-diagonal entries, thus by a theorem of Tartar [11] there will be a diagonal scaling matrix D for which DA is positive definite if and only if H^{-1} has positive entries. It is apparent by considering the inner product of the second row of H with the second column of H^{-1} that the second column of H^{-1} must have at least one negative entry, thus the difficulty here is not one which can be overcome by simple scaling.

Further, we comment that without the $h + k$ scaling it is a much simpler matter to demonstrate that indefinite matrices arise even with meshes which grade smoothly into the right hand boundary. The scaling that we have employed (which is unique in preserving symmetry of the approximation when $\sigma = 0$) is therefore apparently the most sensible.

The basic point is that the upwind approximation of the first order derivative contributes to the symmetric part of the coefficient matrix. This is well known on regular meshes where it strengthens the positive definiteness. The example given here shows that weakening is also possible to the extent that the discretisation of the underlying positive definite problem becomes indefinite.

3 Finite Element and Finite Volume methods

The simple example above demonstrates the issue we wish to highlight. In this section we consider two different and popular upwind strategies which are usually described in the frameworks of finite element and finite volume methods respectively.

There are a large number of finite element approaches including the use of upwind test functions in a so called Petrov-Galerkin setting and bubble functions in the context of Galerkin least squares (see [3], [9]). However, we shall consider only one of the simpler and earlier upwind approaches namely that due to Heinrich et al(1977). In this approach standard piecewise linear finite element trial (expansion) functions ϕ_i are used together with test functions of the form $\phi_i + \psi_i$ where

$$\psi_i(x) = \begin{cases} 3\alpha_i(x - x_{i-1})(x_i - x)/(x_i - x_{i-1})^2 & x_{i-1} \leq x \leq x_i \\ -3\alpha_{i+1}(x - x_i)(x_{i+1} - x)/(x_{i+1} - x_i)^2 & x_i \leq x \leq x_{i+1} \end{cases}$$

and α_i is related to the local mesh Peclet number: here we use the popular choice $\alpha_i = \coth(\beta_i) - 1/\beta_i$ where $\beta_i = \sigma(x_i - x_{i-1})/2$ is the local mesh Peclet number (but see [10] for a more recent approach). The form of the (i, j) entry of the coefficient matrix is then

$$\int_0^1 \phi_j'(\phi_i' + \psi_i') + \sigma \phi_j'(\phi_i + \psi_i) dx$$

which is integrated to give $A = K + \sigma C$ where K is the standard ‘stiffness’ matrix

$$K = \text{tri}(-1/(x_i - x_{i-1}), 1/(x_i - x_{i-1}) + 1/(x_{i+1} - x_i), -1/(x_{i+1} - x_i))$$

and

$$C = \text{tri}(-1/2 - \alpha_i/2, \alpha_i/2 + \alpha_{i+1}/2, 1/2 - \alpha_{i+1}/2).$$

Thus here the first derivative (advection) term does give rise to a symmetric as well as a skew symmetric part: the symmetric part being due solely to the augmentation ψ_i of the trial function. However, since $\alpha_j \geq 0$ for any positive β_j , the symmetric part of C is diagonally dominant and it follows that the symmetric part of A has positive diagonal entries and negative sub- and super-diagonal entries and is diagonally dominant. Hence for any mesh, this discretisation yields a positive definite coefficient matrix.

The second method we consider is the four-point cell-vertex finite volume scheme due to Morton, Rudgyard and Shaw [8]. In the notation employed above this has a typical row of the form

$$\begin{aligned} \dots, 0, \frac{-1}{\ell(k+\ell)}, \frac{1}{\ell(k+\ell)} - \frac{\ell}{k^2(k+\ell)} + \frac{h}{k^2(k+h)} + \frac{\sigma}{k}, \\ \frac{1}{h(k+h)} - \frac{h}{k^2(k+h)} + \frac{\ell}{k^2(k+\ell)} - \frac{\sigma}{k}, \frac{-1}{h(k+h)}, 0, \dots \end{aligned}$$

It is apparent that this method is cell-based and so there is little chance of diagonal dominance for the coefficient matrix in this linear system to determine the nodal unknowns. Indeed, it is not immediately apparent which are the diagonal entries. It is possible to consider the matrix as the sum of the two tridiagonal matrices (with different diagonals)

$$\text{tri}\left(\frac{-1}{\ell(k+\ell)}, \frac{1}{\ell(k+\ell)} - \frac{\ell}{k^2(k+\ell)} + \frac{\sigma}{k}, \frac{\ell}{k^2(k+\ell)}\right)$$

and

$$\text{tri}\left(\frac{h}{k^2(k+h)}, \frac{1}{h(k+h)} - \frac{h}{k^2(k+h)} - \frac{\sigma}{k}, \frac{-1}{h(k+h)}\right)$$

at least away from the boundary, however for large enough σ the second of these has a large negative diagonal and must be indefinite or even negative definite.

One approach to construct equations for nodal (rather than cell) residuals is through ‘distribution matrices’ and artificial viscosity: the analysis of such schemes is beyond the scope of this short note.

Though the analysis of coercivity and stability for cell-vertex finite volume methods is achieved without recourse to matrix theory (see [7]), for advection-diffusion equations it remains a challenging problem to find iterative solution techniques for these methods (but see [13]).

4 Conclusions

We have here only considered three of the great many upwind schemes for the advection diffusion equation. Our concern has been to show that some upwind

schemes on certain meshes can give indefinite coefficient matrices even though the partial differential equations problem is positive definite.

One consequence is that certain iterative solution techniques can be expected not to perform well (or quite possibly fail) for discrete systems of equations derived from such schemes. This is a serious practical issue for large three-dimensional problems.

References

- [1] H.C. Elman. Iterative methods for linear systems, in *Advances in Numerical Analysis, Vol III: Large scale matrix problems and the numerical solution of partial differential equations*, J. Gilbert and D. Kershaw, eds., Cambridge University Press, 1994.
- [2] P. Gresho and R.L. Lee. Don't suppress the wiggles – they're telling you something. *Computers and Fluids*, 9:223-253, 1981.
- [3] D.F. Griffiths and J. Lorenz. An analysis of the Petrov-Galerkin finite element method. *Comput. Meths. Appl. Mech. Engrg.*, 14:39-64. 1978.
- [4] J.C. Heinrich, P.S. Huyakorn, A.R. Mitchell and O.C. Zienkiewicz. An upwind finite element scheme for two-dimensional convective transport equations. *Int. J. Numer. Meths. Engrg.*, 11:131-143. 1977.
- [5] J.A. Meijerink and H.A. van der Vorst. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. *Math. Comput.*, 31:148-162, 1977.
- [6] K.W. Morton. Numerical solution of convection-diffusion problems. Chapman Hall, 1996.
- [7] K.W. Morton, Coercivity for one-dimensional cell vertex approximations., in A. R. Mitchell 70th Birthday Volume, World Scientific, 1996.
- [8] K.W. Morton, M.A. Rudgyard and G.J. Shaw. Upwind iteration methods for the cell-vertex scheme in one-dimension. *J. Comput. Phys.*, 114(2):209-226, 1994.
- [9] A. Quarteroni and A. Valli. Numerical approximation of partial differential equations. Springer-Verlag, 1994.
- [10] J. Simo, F. Armero and C. Taylor. Stable and time-dissipative finite element methods for the incompressible Navier-Stokes equations in advection dominated flows. *Int. J. Numer. Meths. Engrg.*, 38:1475-1506. 1995.
- [11] L. Tartar, Une nouvelle caracterisation des M-matrices, *Revue Francaise d'Informatique et de Recherche operationnelle*, R-3:127-128. 1971.

- [12] R.S. Varga. Matrix iterative analysis. Prentice-Hall, 1962.
- [13] N.D. Wash. Upwind iteration techniques for compressible flow computations. D.Phil. thesis, Oxford University. 1995.