Review of Preconditioning Methods for Fluid Dynamics

Abstract

We consider the use of preconditioning methods to accelerate the convergence to a steady state for both the incompressible and compressible fluid dynamic equations. Most of the analysis relies on the inviscid equations though some applications for viscous flow are considered. The preconditioning can consist of either a matrix or a differential operator acting on the time derivatives. Hence, in the steady state the original steady solution is obtained. For finite difference methods the preconditioning can change and improve the steady state solutions. Several preconditioners previously discussed are reviewed and some new approaches are presented.

† ICASE
NASA Langley Research Center
Hampton, Va 23665
and
School of Mathematical Sciences
Sackler Faculty of Exact Sciences
Tel-Aviv University
Tel-Aviv 69978, Israel
Over the past years numerous researchers have tried to solve the steady state incompressible equations for both inviscid and viscous flows. This also lead to attempts to solve the compressible equations over a large range of mach numbers. A standard way of solving the steady state equations is to march the time dependent equations until a steady state is reached. Since the transient is not of any interest one can use acceleration techniques which destroy the time accuracy but enable one to reach the steady state faster. For the incompressible equations the continuity equation does not contain any time derivatives. To overcome this difficulty Chorin [14] added an artificial time derivative of the pressure to the continuity equation together with a multiplicative variable, $\beta$. With this artificial term the resultant scheme is a symmetric hyperbolic system for the inviscid terms. Thus, the system is well posed and and numerical method for hyperbolic systems can be used to advance this system in time. The free parameter $\beta$ is then chosen to reach the steady state quickly. Later Turkel [54] extended this concept by adding the pressure time derivative to the momentum equations and introducing a second free parameter $\alpha$. This system can then be analyzed for optimal $\alpha$, $\beta$. The resulting system after preconditioning is no longer symmetric but can be symmetrized by a change of variables. This will be shown in more detail later.

It is well known that it is difficult to solve the compressible equations for low Mach numbers. For an explicit scheme this is easily seen by looking at the time steps. For stability the time step must be chosen inversely proportional to the largest eigenvalue of the system which is approximately the speed of sound, $c$, for slow flows. However, other waves are convected at the fluid speed, $u$, which is much slower. Hence, these waves don’t change very much over a time step. Thus, thousands of time steps are required to reach a steady state. Should one try a multigrid acceleration one finds that the same disparity in wave speeds slows down the multigrid acceleration. With an implicit method an ADI factorization is usually used so that one can easily invert the implicit factors. The use of ADI introduces factorization errors which again slows down the convergence rate when there are wave speeds of very different magnitudes [49].

For small Mach numbers it can be shown ([28], [31]) that the incompressible equations approximate the compressible equations. Hence, one needs to justify the use of the compressible equations for low Mach flows. We present several reasons one would still use the compressible equations even though the Mach number of the flow is small.

- There are many sophisticated compressible codes available that could be used for such problems especially in complicated geometries

- For low speed aerodynamic problems at high angle of attack most of the flow consists of a low Mach number flow. However, there are localized regions containing shocks.

- In many problems thermal effects are important and the energy equation is coupled to the other equations.

Therefore, one wants to change the transient nature of the system to remove this disparity of the wave speeds. Based on an analogy with conjugate gradient methods such methods were called [54] preconditioned methods since the object is to reduce the condition number of the matrix. Another approach, in one dimension, is to diagonalize
the matrix of the inviscid term. One can then use a different time step for each equation, or wave. Upon returning to the original variables one finds that this is equivalent to multiplying the time derivatives by a matrix. Hence, this same approach is named characteristic time stepping in [55]. In multidimensions one can no longer completely decouple the waves by diagonalizing both the entropy and the shear waves and so the characteristic time stepping is only an approximation.

Thus, for both the incompressible and compressible equations we will consider systems of the form

\[ w_t + f_x + g_y = 0, \]  

This system is written in conservation form though for some applications this is not necessary. Our analysis will be based on the linearized equations so that the conservation form does not appear in the analysis though it does appear in the numerical system. This system is now replaced by

\[ P^{-1}w_t + f_x + g_y = 0, \]  

or in linearized form

\[ P^{-1}w_t + Aw_x + Bw_y = 0, \]  

In order for this system to be equivalent to the original system in the steady state we demand that \( P \) have an inverse. This only need be true in the flow regime under consideration. We shall see later that frequently \( P \) is singular at stagnation points and also along the sonic line. Thus, we will only consider strictly subsonic flow without a stagnation point or else strictly supersonic flow. For transonic flow it is necessary to smooth out the singularity in a neighborhood of the sonic line. We also assume that the Jacobian matrices \( A = \frac{\partial f}{\partial w} \) and \( B = \frac{\partial g}{\partial w} \) are simultaneously symmetrizable. In terms of the ‘symmetrizing’ variables we also demand that \( P \) be positive definite. We shall show later in detail that it does not matter which set of dependent variables are used to develop the preconditioner. One can transform between any two sets of variables. The choice of variables is dictated only by convenience in constructing the preconditioner. Popular choices are two out of density, pressure, enthalpy, entropy or temperature in addition to the velocity components. Thus, when we are finished we will analyze a system which is similar to (3) where the matrices \( A \) and \( B \) are symmetric and \( P \) is both symmetric and positive definite. Such systems are known as symmetric hyperbolic systems. One can then multiply this system by \( w \) and integrate by parts to get estimates for the integral of \( w_t^2 \), i.e. energy estimates. These estimates can then be used to show that the system is well posed. We stress that if \( P \) is not positive then we change the physics of the problem. For example, if \( P = -I \) then we have reversed the time direction and must therefore change all the boundary conditions. Hence, to be sure that the system is well posed with the original type of boundary conditions we shall only consider the symmetric hyperbolic system. For more general systems one must use a more complicated analysis to show well-posedness for the initial-boundary value problem ([30], [63]).

With these assumptions we see that the steady state solutions of the two systems are the same. Assuming the steady state has a unique solution it does not matter which system we march to a steady state. We shall later see that for the finite difference
approximations the steady state solutions are not the necessarily same and usually the preconditioned system leads to a better behaved steady state.

We can also look at (3) from a different viewpoint. We assume that the matrices $A$ and $B$ are symmetric and $P$ is positive definite. It is well known that for the Euler equations that the matrices $A$ and $B$ cannot be simultaneously diagonalized by a similarity transformation. However, the matrix $P$ has changed the equation. Since $P$ is positive definite there exists a matrix $Q$ so that $P = QQ^*$. We then introduce a new variable $w = Qv$. For constant coefficients $A, B$ (3) is replaced by

$$v_t + Q^*AQv_x + Q^*BQv_y = 0,$$

Thus, the diagonalization question changes and we wish to know if $A$ and $B$ can be simultaneously diagonalized by a congruence transformation $(Q^*AQ)$. A sufficient condition for this to be true is that there exist numbers $\omega_1, \omega_2$ so that $\omega_1A + \omega_2B$ is positive definite. It is shown in [53] that this true for supersonic flow. Hence, we have shown that for supersonic flow one can introduce a preconditioning matrix so that the equations (constant coefficients) are diagonalized. However, this is not true for subsonic flow. We shall later show that using differential operators one can diagonalize the system even for subsonic flow.

2 Incompressible equations

We first consider the incompressible inviscid equations in primitive variables.

$$u_x + v_y = 0$$

$$u_t + uu_x + vu_y + p_x = 0$$

$$v_t + uv_x + vv_y + p_y = 0$$

(5)

We consider generalizations of Chorin’s pseudo-compressibility method [14]. Using the preconditioning suggested in [54] we have

$$\frac{1}{\beta^2}p_t + u_x + v_y = 0$$

$$\frac{\alpha u}{\beta^2}p_t + u_t + uu_x + vu_y + p_x = 0$$

$$\frac{\alpha v}{\beta^2}p_t + v_t + uv_x + vv_y + p_y = 0$$

(6)

or in conservation form

$$\frac{1}{\beta^2}p_t + u_x + v_y = 0$$

$$\frac{(\alpha + 1)u}{\beta^2}p_t + u_t + (u^2 + p)_x + (uv)_y = 0$$

$$\frac{(\alpha + 1)v}{\beta^2}p_t + v_t + (uv)_x + (v^2 + p)_y = 0$$

(7)
Hence, (7) reduces to the original pseudo-compressibility method when $\alpha = 0$. The conservative form reduces to the basic method when $\alpha = -1$. We can also write (7) in matrix form using

$$
\begin{aligned}
P_T^{-1} &= \begin{pmatrix}
\frac{1}{\beta^2} & 0 & 0 \\
\alpha u/\beta^2 & 1 & 0 \\
\alpha v/\beta^2 & 0 & 1
\end{pmatrix} &
P_T &= \begin{pmatrix}
\beta^2 & 0 & 0 \\
-\alpha u & 1 & 0 \\
-\alpha v & 0 & 1
\end{pmatrix}
\end{aligned}
$$

i.e.

$$
\begin{aligned}
\begin{pmatrix}
\frac{1}{\beta^2} & 0 & 0 \\
\alpha u/\beta^2 & 1 & 0 \\
\alpha v/\beta^2 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
v
\end{pmatrix}
+ 
\begin{pmatrix}
0 & 1 & 0 \\
1 & u & 0 \\
0 & 0 & u
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
v
\end{pmatrix}
+ 
\begin{pmatrix}
0 & 0 & 1 \\
0 & v & 0 \\
1 & 0 & v
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
v
\end{pmatrix} = 0
\end{aligned}
$$

Multiplying by $P$ we rewrite this as

$$
w_t + PAw_x + PBw_y = 0
$$

We also define

$$
D = \omega_1 A + \omega_2 B \quad -1 \leq \omega_1, \omega_2 \leq 1
$$

where $\omega_1, \omega_2$ are the Fourier transform variables in the x and y directions respectively. The speeds of the waves are now governed by the roots of $det(\lambda I - PA\omega_1 - PB\omega_2) = 0$ or equivalently $det(\lambda P^{-1} - A\omega_1 - B\omega_2) = 0$. Let

$$
q = u\omega_1 + v\omega_2
$$

Then the eigenvalues of $PD$ are

$$
d_0 = q
$$

$$
d_{\pm} = 1/2 \left[ (1 - \alpha)q \pm \sqrt{(1 - \alpha)^2q^2 + 4\beta^2} \right]
$$

Note that in the special case $\alpha = 1$ we have

$$
d_{\pm} = \pm \beta
$$

and so the ‘acoustic’ speed is isotropic.

We see that the spatial derivatives involve symmetric matrices, i.e. $D$ is a symmetric matrix. Thus, while the original system was symmetric hyperbolic the preconditioned system is no longer symmetric. In ([54]) it is shown that as long as

$$
\beta^2 > \alpha(u^2 + v^2)
$$

then the system is symmetrizable. Hence, for any nonnegative $\alpha$ the system is always symmetrizable. Recall that $\alpha = 0$ for the original pseudo-compressibility equations in primitive variables (7) while $\alpha = -1$ for the original pseudo-compressibility method in conservative variables (8) For $\alpha = 1$ we need

$$
\beta^2 > (u^2 + v^2)
$$
On the other hand the eigenvalues are most equalized if \( \beta^2 = (u^2 + v^2) \). Hence, we wish to choose \( \beta^2 \) slightly larger than \( u^2 + v^2 \). However, numerous calculations verify that in general a constant \( \beta \) is the best for the convergence rate. The reasons for this are not clear.

However, we wish to stress that \( \beta \) has the dimensions of a speed. Therefore, \( \beta \) can not be a universal constant. There are papers that claim that \( \beta = 1 \) or \( \beta = 2.5 \) are optimal. Such claims can not be true in general. It is simple to see that if one nondimensionalizes the equation then \( \beta \) gets divided by a reference velocity. Hence, the optimal ‘constant’ \( \beta \) depends on the dimensionalization of the problem and in particular depends on the inflow conditions. In most calculations the inflow mass is fixed at one or else \( p + (u^2 + v^2)/2 = 1 \). Such conditions will give an optimal \( \beta \) close to one. However, if one chose the incoming mass as ten then the optimal \( \beta \) would be closer to ten.

Van Leer, Lee and Roe considered the compressible equations. They wanted a symmetric preconditioner so that there would be no question of well posedness. We now translate their results to the incompressible equations (1). They assume that the flow is aligned with the x direction and so \( v=0 \) and \( |u|^2 \) is the total speed of the fluid. Their preconditioner in this coordinate system is

\[
\hat{P} = \begin{pmatrix}
\frac{\tau}{\beta^2} |u|^2 & -\frac{\tau}{\beta^2} u & 0 \\
-\frac{\tau}{\beta^2} u & 1 + \frac{\tau}{\beta^2} & 0 \\
0 & 0 & \tau
\end{pmatrix}
\]

Choosing \( \tau = 1 \) preserves the speed of the shear wave while choosing \( \beta = 1 \) gives an isotropic ‘acoustic’ wave (20) the magnitude of this acoustic wave is determined. In order to compare this formula with the previous formulas we wish to reformulate this preconditioner for the case where the flow is not aligned in the x direction. We denote the matrices in the streamwise and perpendicular directions as \( A_\parallel \) and \( A_\perp \) respectively. We next define the rotation matrices as

\[
U = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \theta & \sin \theta \\
0 & -\sin \theta & \cos \theta
\end{pmatrix}, \quad U^{-1} = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \theta & -\sin \theta \\
0 & \sin \theta & \cos \theta
\end{pmatrix}
\]

To get the streamwise direction we shall choose

\[
\cos \theta = \frac{u}{\sqrt{u^2 + v^2}}, \quad \sin \theta = \frac{v}{\sqrt{u^2 + v^2}}
\]

One can then verify that given the original matrices \( A, B \),

\[
A_\parallel = U(A \cos \theta + B \sin \theta)U^{-1} \\
A_\perp = U(-A \sin \theta + B \cos \theta)U^{-1}
\]

Given numbers \( \hat{\omega}_1, \hat{\omega}_2 \) for \( A_\parallel, A_\perp \) we define
\[
\omega_1 = \hat{\omega}_1 \cos \theta - \hat{\omega}_2 \sin \theta \\
\omega_2 = \hat{\omega}_1 \sin \theta + \hat{\omega}_2 \cos \theta
\]

note

\[
\hat{\omega}_1^2 + \hat{\omega}_2^2 = \omega_1^2 + \omega_2^2.
\]

Also define

\[
P = U^{-1} \hat{\mathbf{P}} U.
\]

Then it is easy to verify that

\[
\hat{\mathbf{P}} (A \| \hat{\omega}_1 + A \| \hat{\omega}_2) = U \left[ \mathbf{P} (A \omega_1 + B \omega_2) \right] U^{-1}
\]

Therefore, the appropriate preconditioner is \( \mathbf{P} \) given by

\[
\mathbf{P}_V = \begin{pmatrix}
  u^2 + v^2 & -u & -v \\
  -u & 1 + \frac{u^2}{u^2 + v^2} & \frac{uv}{u^2 + v^2} \\
  -v & \frac{uv}{u^2 + v^2} & 1 + \frac{v^2}{u^2 + v^2}
\end{pmatrix}
\]

Note that \( \mathbf{P}, \mathbf{A}, \mathbf{B} \) are symmetric matrices. This does not imply that \( \mathbf{PA} \) or \( \mathbf{PB} \) are symmetric. However, this is still a symmetric hyperbolic system and so the standard energy estimates prove the well posedness of the system. We also see that the eigenvalues do not change if we use the streamwise direction or the full 2D form. Thus, the eigenvalues of the preconditioned system are

\[
d_0 = \sqrt{u^2 + v^2} \omega_1 = u \omega_1 + v \omega_2 = q \\
d_{\pm} = \pm \sqrt{u^2 + v^2}
\]

\( d_{\pm} \) are the same as in (13) if we choose \( \alpha = 1 \) and \( \beta = \sqrt{u^2 + v^2} \).

As noted before, with the preconditioner of Van Leer et. al. one cannot have the usual shear speed together with an isotropic ‘acoustic’ wave speed with an arbitrary magnitude. With therefore, consider a modification of their preconditioner. In streamwise coordinates it is given by

\[
\hat{\mathbf{P}} = \begin{pmatrix}
  \beta^2 & -\hat{\alpha} & 0 \\
  -\hat{\alpha} & \hat{\sigma} & 0 \\
  0 & 0 & 1
\end{pmatrix}
\]

with

\[
\hat{\alpha} = \frac{\beta^2 \pm \beta \sqrt{\beta^2 - u^2}}{u}, \quad \hat{\sigma} = \frac{2\hat{\alpha}}{u}, \quad u \neq 0
\]

Choosing \( \beta^2 = u^2 \) gives the original preconditioner of Van Leer et. al. for incompressible flow. In general nonaligned coordinates this becomes

\[
P_{VM} = \begin{pmatrix}
  \beta^2 & -\alpha u & -\alpha v \\
  -\alpha u & 1 + \frac{(2\alpha - 1)u^2}{u^2 + v^2} & \frac{(2\alpha - 1)uv}{u^2 + v^2} \\
  -\alpha v & \frac{(2\alpha - 1)uv}{u^2 + v^2} & 1 + \frac{(2\alpha - 1)v^2}{u^2 + v^2}
\end{pmatrix}
\]
\[
\alpha = \frac{\beta^2 \pm \beta \sqrt{\beta^2 - (u^2 + v^2)}}{u^2 + v^2}, \quad u^2 + v^2 \neq 0
\]

Now, we have the condition \( \beta^2 \geq u^2 + v^2 \) (cf. 16). The speeds are now given by
\[
d_0 = q \\
d_{\pm} = \pm \beta
\]

This can now be compared with (14) for \( P_T \).

Numerous computer runs have shown that \( P_T \) works best with \( \beta \) constant and not depending on the speed. To date there have been no computer calculations for the incompressible equations with \( P_V \).

These examples show that the preconditioning is not unique. In fact, it is straightforward to see that the transpose of \( P_T \) is also a preconditioner with the same eigenvalues for the preconditioned system. In general, these various systems will have similar eigenvalues but different eigenvectors for the preconditioned system. Numerous calculations show that the system given by \( P_T \) is more robust and converges faster than that with the transpose preconditioner. This shows that it is not sufficient to consider just the eigenvalues but somehow the eigenvectors are also of importance.

3 Compressible equations

The time dependent Euler equations can be written as
\[
\begin{align*}
\frac{1}{\rho c^2} p_t + \frac{1}{\rho c^2} (u p_x + v p_y) + u_x + v_y &= 0 \\
u_t + uu_x + vv_y + \frac{p_x}{\rho} &= 0 \\
v_t + uv_x + vv_y + \frac{p_y}{\rho} &= 0 \\
S_t + uS_x + vS_y &= 0
\end{align*}
\]

(23)

The first general attempt to replace this by other systems of equations with the same steady state was by Viviand ([59],[27]). He considered both incompressible and compressible isoenthalpic flow. We will consider preconditionings that are a generalization of (9)

\[
\begin{pmatrix}
\frac{1}{\rho c^2} & 0 & 0 & 0 \\
\frac{\alpha u}{\rho c^2} & 1 & 0 & 0 \\
\frac{\alpha v}{\rho c^2} & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
v \\
S
\end{pmatrix}_t
+ \begin{pmatrix}
u c^2 & 1 & 0 & 0 \\
\frac{1}{\rho} & u & 0 & 0 \\
0 & 0 & u & 0 \\
0 & 0 & 0 & u
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
u \\
S
\end{pmatrix}_x
+ \begin{pmatrix}
u c^2 & 0 & 1 & 0 \\
0 & v & 0 & 0 \\
\frac{1}{\rho} & 0 & v & 0 \\
0 & 0 & 0 & v
\end{pmatrix}
\begin{pmatrix}
p \\
u \\
v \\
S
\end{pmatrix}_y
= 0
\]

Note that if we use \( \frac{\partial p}{\partial c} \) instead of \( dp \) the matrices become symmetric. We next present the eigenvalues of \( P_D \) (defined in (11)). Let
\(q = u\omega_1 + v\omega_2\)

then

\[d_0 = q\]

\[d_\pm = 1/2 \left( (1 - \alpha + \beta^2/c^2)q \pm \sqrt{((1 - \alpha + \beta^2/c^2)q^2 + 4(1 - q^2/c^2)\beta^2)} \right)\]

If we consider the special case \(\alpha = 1 + \beta^2/c^2\) we find that the ‘acoustic’ eigenvalue is given by

\[d_\pm = \sqrt{(1 - q^2/c^2)\beta^2}\]

Hence, these eigenvalues are isotropic in the limit of \(M\) going to zero. However, this eigenvalue vanishes at the sonic line and so the matrix is singular. In general, if we demand that the acoustic eigenvalues be isotropic then we have a singularity at the sonic line where the eigenvalues cannot be isotropic. The two ways out of this difficulty are either to smooth the formulas near the singular line or else to give up on isotropy. For example in [34] \(\alpha\) is chosen as zero. This results in a ratio of about 2.6 between the fastest and slowest wave speeds at \(M = 0\). However, now the formulas are regular at the sonic line. This difficulty is not a property of the preconditioning just presented but applies equally to all preconditioners e.g. that of Van Leer et. al. which will now be presented.

The Van Leer, Lee, Roe preconditioning [55] for general non-aligned flow in \((\frac{\partial \rho}{\partial t}, du, dv, dS)\) variables is

\[
\mathbf{P}_V = \begin{pmatrix}
\frac{\tau}{\beta^2} M^2 & -\frac{\tau}{\beta^2} u/c & -\frac{\tau}{\beta^2} v/c & 0 \\
-\frac{\tau}{\beta^2} u/c & \left(\frac{\tau}{\beta^2} + 1 \right) \frac{u^2}{u^2 + v^2} + \tau \frac{u^2}{u^2 + v^2} & \left(\frac{\tau}{\beta^2} + 1 \right) \frac{uv}{u^2 + v^2} & 0 \\
-\frac{\tau}{\beta^2} v/c & \left(\frac{\tau}{\beta^2} + 1 \right) \frac{uv}{u^2 + v^2} & \left(\frac{\tau}{\beta^2} + 1 \right) \frac{v^2}{u^2 + v^2} + \tau \frac{v^2}{u^2 + v^2} & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

\[\beta = \begin{cases} \sqrt{1 - M^2}, & M < 1; \\ \sqrt{M^2 - 1}, & M \geq 1; \end{cases}\]

\[\tau = \begin{cases} \sqrt{1 - M^2}, & M < 1; \\ \sqrt{1 - M^{-2}}, & M \geq 1. \end{cases}\]

At the sonic line \(\beta = 0\) and \(\tau = 0\) and the matrix becomes singular. In both these examples the preconditioner was constructed based on using \((p, u, v, S)\) as the dependent variables. The reason for this choice is that the matrices are essentially symmetric which this choice. However, if another choice of variables is more appropriate that introduces no difficulties. Thus, for example [13] recommends the use of \((p, u, v, T)\) variables for the Navier-Stokes equations. Given two sets of dependent variables \(w\) and \(W\) let \(W_w\) be the Jacobian matrix \(\frac{\partial W}{\partial w}\). Then, we have \(dW = W_w dw\). So we can go between any sets of primitive variables or between primitive variables and conservation variables. In particular since the equations are solved in conservation variables we have several ways of going from the primitive variable preconditioner to a conservation variable preconditioner. Thus, the choice of variables used in constructing the preconditioner is dictated by mathematical or physical reasoning and then the preconditioner can be transformed to any other set of variables.
• We can construct the preconditioner matrix for the conservation variables. If \( W \) are the conservative variables and \( w \) are the primitive variables the \( P_{\text{conservative}} = (W_w)^{-1} P_{\text{primitive}}(W_w) \).

Let \( W \) denote the conservative variables \((\rho, m, n, E)^t\), with \( m = \rho u, n = \rho v \), let \( w \) denote the primitive variables \((p, u, v, S)^t\) and let \( \hat{w} \) denote \((p, u, v, T)^t\). Then

\[
\frac{\partial W}{\partial w} = \begin{pmatrix}
\frac{1}{\gamma} & 0 & 0 & -\frac{\rho}{c_p} \\
\frac{\rho}{c_p} & 0 & 0 & -\frac{m}{c_p} \\
\frac{1}{\gamma - 1} + \frac{M^2}{2} & m & n & -\frac{\rho(u^2 + v^2)}{2c_p}
\end{pmatrix}
\]

\[
\frac{\partial w}{\partial W} = \begin{pmatrix}
\frac{(\gamma - 1)(u^2 + v^2)}{2} & -\frac{u(\gamma - 1)}{\rho} & -\frac{v(\gamma - 1)}{\rho} & \gamma - 1 \\
-\frac{\rho}{c_p} & 0 & 0 & 0 \\
-\frac{2}{\rho} & 0 & 0 & 0 \\
\frac{\rho(c_p(\gamma - 1)M^2 - 1)}{\rho} & -\frac{c_p(\gamma - 1)u}{\rho} & -\frac{c_p(\gamma - 1)v}{\rho} & \frac{(\gamma - 1)c_p}{\rho}
\end{pmatrix}
\]

\[
\frac{\partial \hat{w}}{\partial w} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
\frac{(\gamma - 1)T}{\gamma S} & 0 & 0 & \frac{T}{\gamma S}
\end{pmatrix}
\]

\[
\frac{\partial w}{\partial \hat{w}} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-\frac{R}{\rho} & 0 & 0 & \frac{c_p}{T}
\end{pmatrix}
\]

• We calculate the residual \( dW \) in conservative variables. We then transform \( dW \) to \( dw \) as before. Next we multiply by \( P \) and finally transform back to conservative variables \( dW \) and update the solution. This is algebraically equivalent to the first option but requires three matrix multiplies instead of one. However, it offers more flexibility.

• Similar to the previous suggestion we calculate the residual \( dW \) and transform to conservative variables \( dw \) and the multiply by \( P \). At this stage we update the primitive variables \( w \). We then use the nonlinear relations to construct \( W \) from \( w \). This approach has advantages if the boundary conditions are given in terms of the primitive variables (\( p \) or \( T \)) and so they can be specified exactly and not approximately.

These methods are all equivalent for linear systems and the difference between them is mainly one of convenience.

Based on conservative variables Choi and Merkle [35] suggest two other preconditioners. The first is
\begin{align*}
\textbf{P}_{CM1} &= \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
\frac{u^2 + v^2}{2}(M^{-2} - 1) & u(M^{-2} - 1) & v(M^{-2} - 1) & M^{-2}
\end{pmatrix}
\end{align*}

This matrix is closely related to the first preconditioner \( P_T \) with \( \alpha = 0 \) after switching between \((p, u, v, S)\) variables and conservative variables (see [54] for more details). We get a similar looking preconditioner by replacing \( E_t \) in the energy equation by \( \frac{(E + p)u}{\gamma - 1} \) and then

\[
P = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
\frac{u^2 + v^2}{2} & -u & -v & \frac{\gamma}{\gamma - 1}
\end{pmatrix}
\]

For the Navier-Stokes equations they [13] suggest a different preconditioner given by

\begin{align*}
\textbf{P}_{CM2} &= \begin{pmatrix}
\frac{1}{\beta M^2} & 0 & 0 & 0 \\
\frac{\beta M^2}{u} & \rho & 0 & 0 \\
\frac{\nu M^2}{u} & 0 & \rho & 0 \\
\frac{E + p}{\beta M^2} - \delta & \rho u & \rho v & \frac{\gamma \rho R}{\gamma - 1}
\end{pmatrix}
\end{align*}

Choosing \( \delta = 0 \) or 1 made very little difference in their calculations. For inviscid flows \( \beta = c^2 \). As pointed out before, for both these preconditioners the ratio of eigenvalues of the preconditioned system is not one in the limit of \( M = 0 \) but on the other hand the systems are not singular at the sonic line.

We thus again see that the preconditioner is not unique for a given set of variables. Instead many matrices are capable of reducing the spread of the wave speeds at low Mach numbers. The main difference for inviscid flow between all these preconditioners are the eigenvectors that result from the preconditioning. There has been little work comparing the properties and efficiencies of these preconditioners.

### 3.1 Supersonic Flow

We previously mentioned that for supersonic flow one can diagonalize both matrices \( A \) and \( B \) simultaneously with a congruence transform (two dimensions only). We now explicitly give this transformation. We consider the symmetrizing variables \( \left( \frac{\partial \rho}{\partial c}, u, v, S \right) \), then

\[
A = \begin{pmatrix}
u & c & 0 & 0 \\
c & u & 0 & 0 \\
0 & 0 & u & 0 \\
0 & 0 & 0 & u
\end{pmatrix}, \quad B = \begin{pmatrix}
v & 0 & c & 0 \\
0 & v & 0 & 0 \\
c & 0 & v & 0 \\
0 & 0 & 0 & v
\end{pmatrix}
\]

Let \( q^2 = u^2 + v^2 \). We assume \( u \geq 0, v \geq 0 \). Since the flow is supersonic \( q \geq c \). The last row and column decouple and so we consider only a 3x3 submatrix. Define,
and let \( Q = U_1 T U_2 \). Then

\[
Q^* A Q = \begin{pmatrix}
 u + \frac{c v}{\sqrt{q^2 - c^2}} & 0 & 0 \\
0 & u - \frac{c v}{\sqrt{q^2 - c^2}} & 0 \\
0 & 0 & u
\end{pmatrix}
\quad Q^* B Q = \begin{pmatrix}
 v + \frac{a c}{\sqrt{q^2 - c^2}} & 0 & 0 \\
0 & v - \frac{a c}{\sqrt{q^2 - c^2}} & 0 \\
0 & 0 & v
\end{pmatrix}
\]

We then have the following trivial theorem:

**Theorem 1** If we replace the matrices \( A \) and \( B \) by the same congruent transformation then this is equivalent to preconditioning with a non-negative matrix. If the congruent transformation is nonsingular then the preconditioning matrix is positive definite.

The proof follows since \( QQ^* A = Q(Q^* AQ)Q^{-1} \) and similarly for \( B \). Thus, the preconditioner \( P \) is given by \( P = QQ^* \). The converse follows by letting \( Q \) be the square root of \( P \) which exists whenever \( P \) is positive definite.

4 Difference Equations

Until now the entire analysis has been based on the partial differential equation. For long waves it is reasonable to replace the numerical approximation by the original differential equation. Since we are mainly interested in wave speeds these are governed by the low frequencies. It is also possible to extend this analysis to the finite difference approximation. We now make some remarks on important points for any numerical approximation of this system.

- For an upwind difference scheme based on a Riemann solver this Riemann solver should be for the preconditioned system and not the original scheme. In [17] plots are shown to illustrate the greatly improved accuracy for low Mach number flows when the Riemann solver is based on the preconditioning.

- For central difference schemes there is a need to add an artificial viscosity. Accuracy is improved for low Mach number flows if the preconditioner is applied only to the physical convective and viscous terms but not to the artificial viscosity. Volpe [60] shows that the accuracy of the original system deteriorates as the Mach number is
reduced. The author has had a similar experience in three dimensional flows around a fuselage configuration. The use of a matrix artificial dissipation ([51]) should be based on the preconditioned equations as in the upwind difference scheme. On the other hand Merkle (private communication) has indicated that he has no difficulties with accuracy in the very low Mach regime. He can take the solution obtained with a preconditioner and use that as initial data for a nonpreconditioned code which then simply converges in one time step with the same small residual. In this case both the original system and the preconditioned system give the same results even on the difference level. Upwind schemes tend to have more difficulties with accuracy for low Mach flows [17].

Hence, both for upwind and central difference schemes the Riemann solver or artificial viscosity should be based on \( P^{-1}PA \) and not \( |A| \). i.e. in one dimension solve \( w_t + Pf_x = (|PA|w_x)_x \). For a scalar artificial viscosity \( |PA| \) is replaced by the spectral radius of \( PA \) or equivalently the time step associated with the preconditioned matrix. This is equivalent to not multiplying the artificial viscosity by \( P \).

- Similarly, when using characteristics in the boundary conditions these should be based on the characteristics of the modified system and not the physical system.
- When using multigrid it is better to transfer the residuals based on the preconditioned system to the next grid since these residuals are more balanced than the physical residuals.
  Preconditioning is even more important when using multigrid than with an explicit scheme. With the original system the disparity of the eigenvalues greatly affects the smoothing rates of the slow components and so slows down the multigrid method, [56].
- In addition to convergence difficulties there are accuracy difficulties at low Mach numbers [60]. Some of these can be alleviated by preconditioning the dissipation terms as indicated above. For very small Mach numbers there is also a difficulty with roundoff errors as \( \frac{P}{u^2 + v^2} \to \infty \). Several people have suggested subtracting out a constant pressure from the dynamic pressure. A more detailed analysis [20] suggests replacing the pressure \( p \) by \( \tilde{p} \) where \( p = \frac{u_0 + \epsilon \tilde{p}}{2} \) and \( \epsilon \) is a representative Mach number.
- We conclude from the above remarks that the steady state solution of the preconditioned system may be different from that of the physical system. Thus, on the finite difference level the preconditioning can improve the accuracy as well as the convergence rate.

5 Differential Preconditioners

In the previous sections the preconditioner \( P \) was a matrix. For the nonlinear fluid dynamic equations the elements of \( P \) involved the dependent variables. There are several limitations with this approach.

We first consider a scalar equation

\[
(30) \quad w_t + aw_x + bw_y = 0,
\]
We consider a uniform cartesian mesh with constant $\Delta x, \Delta y$. We define the aspect ratio for this problem as

$$ ar = \text{aspect ratio} = \frac{a/\Delta x}{b/\Delta y}. $$

This can be interpreted as the ratio of time for a wave to traverse a mesh in the x direction relative to the time in the y direction. We note that the ratio $\Delta y/\Delta x$ is meaningless since this can be changed by a trivial change of variables.

If this aspect ratio differs greatly from one then the standard schemes will converge slowly since a time step appropriate for one direction is inappropriate for the other direction. For a scalar equation, this is an artificial problem since, in practice, the mesh would be chosen so that the aspect ratio is close to one. However, for a system of equations there are many waves. If the aspect ratio is close to one for one wave it will not be close to one for other waves. In the boundary layer for the acoustic wave $ar = (u+c)/\Delta x \approx \Delta y/\Delta x$. However, for the shear wave $ar \approx u \Delta v/\Delta x$ and away from the wall but in the boundary layer $u$ is much larger than $v$. Hence, any mesh that is appropriate for the acoustic wave is not appropriate for the shear and entropy waves and vice versa. In addition there are viscous effects that we are ignoring, so that in practice the mesh is constructed based on viscous effects and ignores both the acoustic and entropy waves. For the scalar equation we are considering algebraic preconditioning cannot help (Li and Van Leer, private communication). For a system the preconditionings we have considered can partially rectify the difference of speeds between the various waves but does not alleviate the aspect ratio difficulty.

The matrix preconditioners we have considered until now have a second difficulty. For one dimensional flow one can choose the preconditioner as the absolute value of the matrix $A$. Then all the resultant waves have identical speeds with only differences in the direction, positive or negative. However, in two space dimension when the matrices $A$ and $B$ do not commute it is not possible, in general , to equalize all the speeds. Equivalently, we cannot diagonalize the system and reduce it to a sequence of scalar equations even for the frozen coefficient problem.

To alleviate these two problems we shall allow the preconditioner $P$ to contain derivatives. However, as before we still demand that for the symmetric equations that $P$ be invertible and be positive definite.

For the scalar equation (30) we consider a preconditioner based on residual smoothing [26]. This is given by

$$ (1 - \beta_x \partial_{xx})(1 - \beta_y \partial_{yy})Res_{new} = Res_{old} $$

where $Res$ refers to the residual before and after smoothing. This residual smoothing is usually introduced to improve the time step and smoothing properties of an explicit scheme as Runge-Kutta or Lax-Wendroff. Here, we analyze the scheme from a different perspective, that of wave speeds. We assume that the aspect ration for the problem is very large (i.e. $b$ is large compared to $a$ or $\Delta y$ is small compared to $\Delta x$ ). The question we wish to address is whether $\beta_x$ and $\beta_y$ can be chosen so as to reduce this aspect ratio.

We first consider residual smoothing in one space dimension. In this case there is no aspect ratio. Instead we will show how the concept of wave speeds explains the phenomena that one should not use residual smoothing with a very large time step even though it can be stabilized by choosing an appropriately large $\beta$.

$$(1 - \beta \partial_{xx})w_t + aw_x = 0$$
We analyze this for a semi-discrete equation with time continuous, the first \( x \) derivative approximated by a central difference and the second space derivative by a three point central difference. In order to find the phase and group velocities we consider solutions of the form \( w = e^{i(kx - \omega t)} \). Here \( k \) is given and we find \( \omega \) from the dispersion relation. For the one dimensional residual smoothing we have

\[
k = a \sin \theta, \quad \omega = \frac{a \sin \theta / \Delta x}{1 + 2 \beta (1 - \cos \theta)}
\]

\( \theta = k \Delta x \)

To find a stability condition for a Runge-Kutta scheme in time we maximize \( \omega \) and find that the worst case is \( \cos \theta = \frac{2\beta}{1 + 2\beta} \). We then find that the scheme is stable if

\[\beta \geq \frac{1}{4} (r^2 - 1)\]

where \( r = \frac{\Delta t_{\text{new}}}{\Delta t_{\text{original}}} \). Thus, from the viewpoint of stability we can choose any time step we wish by choosing \( \beta \) sufficiently large. Nevertheless, one finds computationally that convergence to a steady state is slowed down by choosing \( \Delta t \), and hence \( \beta \), too large. Optimal values are \( r \sim 2 \). We shall now show from the viewpoint of wave propagation that it is not good to choose a very large time step.

Residual smoothing adds a term \( w_{xx} \) to the original differential equation. Such a term is a dispersive term i.e. the energy is not reduced but now the speed of a plane wave is no longer constant but instead depends on the wave number. The main purpose of this term is to increase the time stability limit. However, as in defining the aspect ratio, increasing the time step is meaningful only if we normalize the solution in some way, otherwise we are merely rescaling the time dimension. Hence, the appropriate quantity is not the time step but rather the time it takes a wave to transverse one cell (assuming \( \Delta x \) is constant). The phase speed of a plane wave is given by

\[v_p = \frac{\omega}{k} = \frac{a}{1 + 4\beta \sin^2 \theta / 2}\]

Let \( \beta = \frac{1}{4} (r^2 - 1) \) and multiply \( v_p \) by \( r \) to get the distance transversed in time \( \Delta t \). Then

\[s_p = \text{relative phase distance} = \frac{2r}{(r^2 + 1) - (r^2 - 1) \cos \theta}\]

For the long wave lengths \( \cos \theta \sim 1 \) and so \( s_p \sim r \), i.e. the long wave lengths move \( r \) times further in one time step. If we look at \( \theta = \pi / 2 \), we have \( s_p = \frac{2r}{r^2 + 1} \leq 1 \). Thus this frequency moves slower than without residual smoothing. For the highest frequency on the mesh we have \( \theta = \pi \) and \( s_p = \frac{1}{r} \). We therefore, see that the high frequencies are actually slowed down by the residual smoothing and so take longer to exit from the domain, furthermore the larger \( \Delta t \) is chosen the slower these waves go. Even more important the larger \( \Delta t \) is chosen the more frequencies that are slowed down even though the lowest frequencies travel faster. The breakeven frequency is given by \( \cos \theta = \frac{r - 1}{r + 1} \).

We can also consider the group velocity. For the optimal \( \beta \) this is given by
\[
v_g = \frac{d\omega}{dk} = 2\frac{(r^2 + 1)\cos\theta - (r^2 - 1)\cos2\theta}{[(r^2 + 1) - (r^2 - 1)\cos\theta]^2}
\]

and

\[
s_g = s_p \frac{(r^2 + 1)\cos\theta - (r^2 - 1)\cos2\theta}{(r^2 + 1) - (r^2 - 1)\cos\theta}
\]

The situation now is even less favorable than before. Again, the lowest frequencies are sped up by a factor \(r\). The frequency \(\theta = \pi/2\) is slowed down by an additional factor of \(\frac{r^2}{r^2 + 1}\) and the highest frequency \(\theta = \pi\) now reverses direction and goes upstream.

In figures (1a-1c) we plot the phase and group relative distances for \(r=2,5,10\). As demonstrated above we gain a factor of \(r\) for the low frequencies but actually lose compared with \(r=1\) for the high frequencies. As \(r\) is increased more frequencies get slowed down. Because we are considering the semi-discrete equation and residual smoothing is purely dispersive there is no damping of the waves. For a Runge-Kutta scheme one finds that as \(r\) is increased that the damping of high frequencies decreases. Thus, for large \(r\) the high frequencies do not propagate very fast and are not damped either. This explains one in practice one chooses an \(r\) of about two for the greatest increase in the convergence rate to a steady state.

We next consider the two dimensional equation. To ease the derivations we shall consider the partial differential equation (31) rather than the finite difference approximation. We rewrite (31) as

\[
(34) \quad u_t - \beta_x u_{xxt} - \beta_y u_{yyt} + \beta_x \beta_y u_{xxyyt} = au_x + bu_y
\]

We are interested in the effect of high aspect ratios. So we consider \(\Delta y << \Delta x\). By rescaling we instead consider a uniform mesh but \(a << b\). In particular we shall choose \(a = \epsilon, \quad b = 1\).

Consider solutions of the form \(u = e^{i(k_x x + k_y y - \omega t)}\) or equivalently \(u = e^{i(\bar{k} \cdot \bar{x} - \omega t)}\) where \(\bar{k} = (k_x, k_y)\) and \(\bar{x} = (x, y)\). Substituting this into (34) we get

\[
\omega(k_x, k_y) = \frac{\epsilon k_x + k_y}{(1 + \beta_x k_x^2)(1 + \beta_y k_y^2)}
\]

Hence, \(\omega(1, 0) = \frac{\epsilon}{1 + \beta_x}\) and \(\omega(0, 1) = \frac{\epsilon}{1 + \beta_y}\). If we want these to be equal then we need \(\beta_x = O(1), \beta_y = O(\frac{1}{\epsilon})\). This is different than what is normally chosen for in residual smoothing ([50]).

We now consider differential preconditioners for the Euler equations. We shall only considered the linearized equations with constant coefficients. This will now be a matrix preconditioner where the elements of the matrix contain partial derivatives. We first rewrite (24) in a more relevant differential form. Thus, the Euler equations can be written as

\[
(35) \quad w_t + L w = 0
\]

with \(w = (p, u, v, S)^t\). We next define

\[
Q = u\partial_x + v\partial_y
\]
Since, all coefficients are assumed constant $Q$ commutes with $\partial_x$ and $\partial_y$ then

$$L = \begin{pmatrix} Q & \rho c^2 \partial_x & \rho c^2 \partial_y & 0 \\ \frac{1}{\rho} \partial_x & Q & 0 & 0 \\ \frac{1}{\rho} \partial_y & 0 & Q & 0 \\ 0 & 0 & 0 & Q \end{pmatrix}$$

(36)

Let

$$D = Q^2 - c^2 (\partial_x^2 + \partial_y^2).$$

(37)

We now replace (35) by the preconditioned system

$$w_t + P_D L w = 0$$

(38)

with

$$P_D = \begin{pmatrix} Q^2 & -\rho c^2 \partial_x Q & -\rho c^2 \partial_y Q & 0 \\ -\frac{1}{\rho} \partial_x & Q^2 - c^2 \partial_x^2 & c^2 \partial_x \partial_y & 0 \\ -\frac{1}{\rho} \partial_y & c^2 \partial_x \partial_y & Q^2 - c^2 \partial_y^2 & 0 \\ 0 & 0 & 0 & D \end{pmatrix}$$

(39)

One can then verify that

$$P_D L = Q D I, \quad P_D^{-1} = D^{-1} Q^{-1} L$$

One can of course replace the D in the lower right corner of $P_D$ by the identity matrix. Then $P_D L$ is not the identity matrix but is still a diagonal matrix. We can use simpler matrices than $P_D$ by considering congruent transformations. We consider the symmetrizing variables $c(\frac{\partial x}{\rho c}, u, v, S)$, then

$$L = \begin{pmatrix} Q & c \partial_x & c \partial_y & 0 \\ c \partial_x & Q & 0 & 0 \\ c \partial_y & 0 & Q & 0 \\ 0 & 0 & 0 & Q \end{pmatrix}$$

Let,

$$P_E = \begin{pmatrix} Q & -c \partial_x & -c \partial_y & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad P_E^t = \begin{pmatrix} Q & 0 & 0 & 0 \\ -c \partial_x & 1 & 0 & 0 \\ -c \partial_y & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(40)

then

$$P_E L P_E^t = \begin{pmatrix} DQ & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & Q & 0 \\ 0 & 0 & 0 & Q \end{pmatrix}$$

so we have diagonalized $L$ by a congruent transformation. But,

$$P_E L P_E^t = P_E^{-1} (P_E^t P_E L) P_E.$$
so the congruent transform is similar to a preconditioning with a positive definite matrix \( P_E^t P_E \). Alternatively, \((P_E^t P_E)L\) is similar to a diagonal matrix.

\[
P_E^t P_E = \begin{pmatrix}
Q & -c\partial_x & -c\partial_y & 0 \\
-c\partial_x & 1 + c^2\partial_x^2 & c^2\partial_x\partial_y & 0 \\
-c\partial_y & c^2\partial_x\partial_y & 1 + c^2\partial_y^2 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

Note that \( P_E^t P_E \) looks similar to \( P_D \) but is not identical. \( P_E^t P_E \) has fewer derivatives along the identical but \( P_E^t P_E L \) is only similar to a diagonal matrix while \( P_D L \) is diagonal and even a scalar differential operator multiplying the identity matrix. These transformations are independent of the flow regime as long as the preconditioner is non-singular.

These preconditioners are connected with the techniques used in distributive Gauss-Seidel smoothers for multigrid methods ([6],[7]).

It remains to show that \( P \) is nonsingular. We have four eigenvalues and corresponding eigenfunctions. As usual the entropy wave decouples. For this wave \( P \) has an eigenvalue \( D \) and an eigenfunction \((0,0,0,1)\). For the shear wave \( P \) has an eigenvalue \( D \) and the eigenvector is \((v_1, v_2, v_3, 0)\) where

\[
Dv_1 = 0 \\
D\left(\frac{\partial v_2}{\partial x} + \frac{\partial v_3}{\partial y}\right) = 0
\]

The other two ‘acoustic’ eigenvalues of \( P \) are \( Q^2 \pm cQ\sqrt{\partial_x^2 + \partial_y^2} \) and the eigenvectors satisfy the pseudo-differential equation

\[
\left[c(\partial_x^2 + \partial_y^2) \mp Q\sqrt{\partial_x^2 + \partial_y^2}\right]\left(\frac{\partial v_3}{\partial x} - \frac{\partial v_2}{\partial y}\right) = 0 \\
\mp\sqrt{\partial_x^2 + \partial_y^2} = \rho c \left(\frac{\partial v_2}{\partial x} + \frac{\partial v_3}{\partial y}\right)
\]

We therefore have to show that the eigenvalues are all nonzero so that \( P \) is nonsingular. The operator \( D \) is just the potential operator i.e. for any variable \( w \)

\[
Dw = (u^2 - c^2)w_{xx} + 2uww_{xy} + (v^2 - c^2)w_{yy}
\]

For subsonic flow this is an elliptic operator and so invertible. For supersonic flow \( D \) is a hyperbolic operator. Similarly, \( Q \) is a hyperbolic operator denoting convection along a streamline. Thus given appropriate boundary conditions this too should be invertible. At a stagnation point \( Q \) is singular and so it is necessary to limit the values of \( u \) and in \( v \) in the definition of \( Q \) so that they do not become too small in a neighborhood of the stagnation point. A similar smoothing is needed near the sonic line. These arguments have been applied to \( P_D \) but similar arguments work for \( P_E^t P_E \).

With residual smoothing and \( P_D \) or \( P_E^t P_E \) we have increased the order of the system and so changed the number of boundary conditions needed for the equation to be well
posed. To avoid this difficulty we do not solve the equation (38). Instead these preconditioners are used as a post processor for the usual Euler or Navier-Stokes equations. Thus, at each time step we calculate a residual based on one’s favorite scheme. This gives a predicted value of the change in time, $\Delta w_{\text{predicted}}$. We also update the boundary conditions for the standard fluid dynamic equations. We then operate on $\Delta w$ with $P$ with the boundary condition that $\Delta w_{\text{corrected}} = 0$, i.e. we don’t change the boundary values calculated by the predictor. When we reach a steady state for the fluids equations we are solving $P\Delta w_{\text{corrected}} = 0$ with zero boundary conditions. Since $P$ is invertible $\Delta w = 0$, i.e. we preserve the steady state. Thus, in essence we are imposing the fluid dynamic boundary conditions between the $P$ operator and the $L$ operator.

6 Alternate Methods, Time Dependent Problems and Viscous Problems

The justification for preconditioned schemes began with low Mach number flows. For such flows other techniques exist beside preconditioning the equations. The method of time inclining has similarities to preconditioning [15].

The basis of one such method is to use an implicit scheme. However, a two dimensional implicit method is too expensive to be efficient. Thus, one classically uses an ADI approach. However, it is known that with ADI one cannot choose a very large time step and converge quickly to the steady state. The splitting errors that occur in the ADI method couples the waves together and one cannot choose an appropriate time step for each wave. Instead one attempts to separate those terms in the equations that contribute to the fast acoustic waves from the slow components. One than can use a semi-implicit method which is implicit for the fast waves and explicit for the slow waves. Thus, the stability limit of the scheme is governed by the convective speed rather than the acoustic speed. The explicit part can be either a leapfrog method ([18], [19]), or a two step method [20]. This can also be extended to the Navier-Stokes equations [21]. Alternatively, once these components are identified, one can split the equations in several pieces and solve each one separately as in the classical splitting methods [2]. In this case one can use an implicit method for the fast waves and an explicit method for the slow waves and in addition one can split off the viscous terms. These methods work for both time dependent and steady state problems.

A different alternative is to add terms to the equations which disappear in the steady state. This has a connection with preconditioned methods when time derivatives are added to the equations. However, in this approach other terms can be added beside time derivatives. One example, is to assume that the total enthalpy is constant in the steady state for the compressible inviscid equations. One can then add terms to the equations that depend on the deviation of the current enthalpy at each point from this constant steady state enthalpy [25]. For the incompressible equations one can add the divergence of the velocity field or time derivatives of the divergence to the momentum equation [41], [43]. One can also consider a more general equation of state that reduces to the physical one at the steady state [44]. In [27] they analyze the general case of such pseudo-unsteady systems.

An extension of this technique is to modify the differential equation to remove the acoustic waves or other ‘bad’ features. One must then justify that the solutions obtained to these modified equations are close to the original equations for some flow regime. Typical examples are the various Low Mach number expansions for the fluid dynamic equations or the geostrophic equations as an approximation to the shallow water
equations in meteorology.

For incompressible flow popular schemes are the SIMPLE [39] and MAC [22] algorithms and their generalizations. These usually require the solution of a Poisson equation for the pressure and then a pressure correction is used to update the momentum equations. These methods can then be generalized to the compressible equations [24]. Merkle, Venkateswaran and Buelow [37] compare such methods to the preconditioned techniques discussed in this paper. We again stress that the difference in these approaches is not whether density or pressure are used as the dependent variable as one can transform between these variables. Thus, for example, one can modify the compressible continuity equation by replacing the time derivative of the density with a time derivative of the pressure. This is just another example of a matrix preconditioning as one can express the pressure derivative as a combination of a density derivative together with momentum and energy derivatives. As described above, it is a programming decision whether one should use this modified equation to update the pressure and then transfer to density or to calculate the the appropriate preconditioning matrix and update the density. For a linear system the two approaches are identical.

For time dependent problems the first approach just discussed is useful. However, the preconditioned methods and the second approach of this section destroy the time accuracy unless the coefficients of the perturbation are chosen as a function of the mesh size and so only affect terms of the order of the accuracy of the scheme. A more popular approach has been to use a two-time scheme. In this approach each new time level is considered as the steady state of some problem. Alternatively, the physical time derivatives are considered a forcing terms. One now uses the preconditioned methods to achieve this ‘steady state’ which in reality is the solution at the next time step. Hence, there is the physical time \( t \) and an artificial time \( \tau \) and \( \tau \) goes to infinity as an inner loop within each time step. [12], [47], [48]). Thus,

\[
P^{-1} \frac{\partial w}{\partial \tau} + \frac{\partial w}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0.
\]

The main difficulty with this approach is its efficiency. It is reasonable to use such a technique only if each ‘steady state’ problem can be solved with little effort. One advantage is that one usually has good initial guess for the solution based on the solution at previous time steps. However, it typically takes 10 subiterations for each time step. Hence, this approach is ten times more expensive than a straight implicit method. One can also use an Newton iteration [38] at each time step. nevertheless, a semi-implicit approach in ([18] - [21]) seems attractive.

All the methods discussed thus far have been based on an inviscid analysis. For the Navier-Stokes equations at high Reynolds number we do not expect any important changes outside the boundary layer. Inside the boundary layer viscous effects modify the eigenvalues of the differential operator. We thus wish to equalize the contribution of three quantities, the acoustic waves, the convective waves and the viscous terms. In particular the viscous eigenvalues are very stiff and so the eigenvalues of the solution operator are no longer well conditioned. All the preconditioners presented above depend on free parameters \((\beta, \alpha, \tau, \delta)\). Optimal values for these parameters were given for inviscid flow. A simple extension of the above methods to viscous flow would keep the same form for the preconditioning matrices but allow these parameters to also depend on the Reynolds or Prandtl number (see for example [10] ,[13] ). Thus, for example one finds that for the original pseudo-compressibility method that \(\beta\) should increase as the Reynolds number
is decreased. In [13] a new preconditioner is introduced Based on a physical analysis of
the Navier-Stokes equations (see 29). The difficulty is that the time steps are governed
by the acoustic and convective speeds and also a viscous contribution.

A basic problem for the preconditioned Navier-Stokes equations is well-posedness. For
the inviscid equations one can show that with the preconditioner \( P^T \) that the equations
can be symmetrized if \( \alpha, \beta \) satisfy the inequality (15), (see [54] ). The preconditioner
\( P_v \) is constructed from the symmetric form. Hence, in both cases we can reduce the
preconditioned equations to a symmetric hyperbolic system and so it is well posed. Once
one adds the viscous terms this analysis is no longer valid. One possibility is to start
with a form that is symmetric for both the inviscid and viscous terms [1]. If one uses
a positive definite preconditioner for these variables then standard energy arguments
shows that the linearized preconditioned system is well-posed.

We now analyze the preconditioner \( P^T \) a little more carefully for the incompressible
Navier-Stokes equations. We also linearizee and so the coefficients \( u, v, \beta \) are considered
as constant. The resultant preconditioned equations are

\[
\frac{1}{\beta^2} p_t + u_x + v_y = 0
\]

\[
\frac{\alpha u_0}{\beta^2} u_t + u_0 u_x + v_0 u_y + p_x = \mu \Delta u
\]

\[
\frac{\alpha v_0}{\beta^2} v_t + u_0 v_x + v_0 v_y + p_y = \mu \Delta v
\]

We next differentiate the second equation with respect to x and the third with respect
to y. We replace the divergence of the velocity from the first equation. Let \( R = u_0(u_{xx} +
v_{xy}) + v_0(u_{xy} + v_{yy}) \). Then the pressure \( p \) satisfies an acoustic-like equation

\[
-\frac{1}{\beta^2} p_t + \Delta p + \frac{\mu}{\beta^2} \Delta p_t + \frac{\alpha}{\beta^2} (u_0 p_x + v_0 p_y)_t = -R
\]

Thus, we replace the Poisson equation used in the MAC type approach by a gener-
alized wave equation for the pressure. We Fourier transform (42) , i.e. \( p = e^{i(k_1 x + k_2 y - \omega t)} \)
and \( |k|^2 = k_1^2 + k_2^2 \). Then,

\[
\omega^2 - \left[ \alpha(u_0 k_1 + v_0 k_2) + i\mu |k|^2 \right] \omega - |k|^2 \beta^2 = 0.
\]

We first consider the case \( \alpha = 0 \) (i.e. the original pseudo-compressibility for the
primitive equations). Then

\[
\omega = \frac{i\mu |k|^2 \pm |k| \sqrt{4\beta^2 - \mu^2 |k|^2}}{2}
\]

We now have two regimes to consider

case 1: \( |k| \) small (i.e. \( |k|^2 < 4\beta^2/\mu^2 \))

Then (44) gives \( \omega \). As expected \( \mu \) introduces a decay in the acoustic wave . The
speed of the wave (real part of \( \omega \)) is now slowed down for the same \( \beta \). We thus should
choose a larger \( \beta \) as \( \mu \) increases to compensate for this (see also [13]).
case 2: \(|k| \text{ large (i.e. } |k|^2 > 4\beta^2/\mu^2|) \) Now, \( \omega = i\mu|k|^2 \left[ 1 \pm \sqrt{1 - 4\beta^2\mu^2|k|^2} \right] \). Hence, \( \omega \) is pure imaginary. Thus, these high frequencies do not propagate and their damping is reduced by \( \beta \) (for the smaller damping mode). Thus, one also wants to increase \( \beta \) so that most of the modes in the domain correspond to small \(|k| \).

We next consider non-zero \( \alpha \). Let \( \gamma = \alpha(u_0k_1 + v_0k_2) \)

\[
\omega = \frac{\gamma + i\mu|k|^2 \pm \sqrt{4\beta^2|k|^2 + \gamma^2 - \mu^2|k|^4 + 2i\gamma\mu|k|^2}}{2}.
\]

Taking real and imaginary parts of the square root we see that only \( \gamma^2 \) enters into the imaginary part of \( \omega \), i.e. the decay rate. So the sign of \( \alpha \) is not important for viscous effects. Thus, it seems that \( \alpha \) has no major impact on viscous flows and its advantage comes from equalizing the flow speeds of the inviscid portion of the flow.

7 Computational Results

Numerous authors have used some of these preconditioners for both incompressible and compressible flows. A selection of papers is presented in the bibliography. Here we summarize a few of these calculations. Most of these computations have used central difference approximations of the spatial derivatives and either a Runge-Kutta explicit scheme or an A.D.I. implicit scheme in time.

For the original pseudo-compressibility equations a number of authors (e.g. [10], [45], [11]) have found that a constant \( \beta \) works best. Rizzi and Eriksson [45] suggest \( \beta^2 = \text{max}(0.3, r(u^2 + v^2)) \) with \( 1 \leq r \leq 5 \), see also [9]. In [38] they also explore similar issues with regard to upwind schemes. As before their constant 0.3 must depend on the normalizations used. Arnone ([3], [4]) has used the original pseudo-compressibility method to solve inviscid and viscous incompressible flow about cascades. A Runge-Kutta method is used which is accelerated by a multigrid technique. This method has been extended by the author to include the preconditioner \( P_T \). In these calculations we find that \( \beta = \text{constant} \) is more robust than choosing \( \beta \) to depend on the speed of the flow. In most cases using a variable \( \beta \) causes the iterations to diverge though when they do converge it is faster than the constant \( \beta \). Paul and Carlson [40] have a similar three dimensional code for external flow over wings. This code has also been extended to include \( P_T \). In both these codes the convergence is also very dependent on the boundary conditions imposed. For some boundary conditions the code converged for a range of \( \alpha \) and then \( \alpha = 1 \) gave the fastest convergence rates as expected. However, for other boundary conditions only the original pseudo-compressibility method \( \alpha = 0 \) would converge. It is suspected that the difficulties are connected with initialization. Thus, \( \alpha = 1 \) though faster may be less robust. It would therefore be necessary to start the calculation with \( \alpha = 0 \) and only once the asymptotic region is reached to change to \( \alpha = 1 \).

Hsu [23] also solves the incompressible equations using \( P_T \). In this case an upwinded approximation is used and the solution is advanced using an A.D.I. method. They examine in more detail the influence of \( \alpha \) and \( \beta \). Due to their implicit solver the code convergences in all the cases they tried, mainly flows about a delta wing. However, they also find that \( \beta = 1 \) is faster than the variable \( \beta \). They principally investigated \( \alpha = -1 \) but indicate that other \( \alpha \)’s behaved similarly. There have been no computations, to date,
for the incompressible equations using the $P_V$ preconditioner due to the newness of this approach.

For the compressible equations at low Mach numbers early calculations were done by Briley, McDonald and Shamroth [8] and a later by D. Choi and Merkle [11], and also Y.H. Choi and Merkle [34]. These methods have mainly used A.D.I. methods though some results with Runge-Kutta schemes have also been achieved. More recently ([17],[55]) results have been achieved with the $P_V$ preconditioner in conjunction with an upwind scheme. Godfrey (private communication) indicates that there is not a great difference between the two preconditioners. The use of the correct Riemann solver was more important than the details of the preconditioner.

Much of the most recent work has gone into extending these results to the Navier-Stokes equations [13] and chemistry ([17], [48], [58]). A number of authors have also investigated extensions to time dependent problems based on a two-time approach ([16], [48], [62]).

Here we present only one set of results. This is for incompressible flow around a VKI cascade with a nonperiodic mesh across the wake. The mesh is shown in figure 2a. A Runge-Kutta multistage scheme is used with a multigrid acceleration. The code is a extension of that of Arnone and Stecco [4]. The flow is turbulent with a Reynolds number of 500,000 and Baldwin-Lomax type turbulence model is used. In table I we present the residual of the pressure after 50 steps on the first mesh, 50 steps on the second mesh and 300 steps on the finest mesh. We thus see that $\alpha = 1$ gave the fastest convergence rates, though the differences were not very large. We were able to run only the modified Van Leer et. al. preconditioner and even that only with a constant $\alpha$ and $\beta$ with $\alpha = \frac{1}{2}$ as opposed to the value of $\alpha$ given in (22). With this value of $\alpha$ the terms with $u^2 + v^2$ do not appear.

<table>
<thead>
<tr>
<th>method</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>precondition AV</td>
<td>1</td>
<td>1</td>
<td>6.63</td>
</tr>
<tr>
<td>precondition AV</td>
<td>1</td>
<td>0</td>
<td>6.07</td>
</tr>
<tr>
<td>precondition AV</td>
<td>1</td>
<td>-1</td>
<td>5.76</td>
</tr>
<tr>
<td>no precondition AV</td>
<td>1</td>
<td>1</td>
<td>6.34</td>
</tr>
<tr>
<td>no precondition AV</td>
<td>1</td>
<td>0</td>
<td>6.23</td>
</tr>
<tr>
<td>no precondition AV</td>
<td>1</td>
<td>-1</td>
<td>5.76</td>
</tr>
<tr>
<td>precon AV eq. (22)</td>
<td>1</td>
<td>0.5</td>
<td>6.34</td>
</tr>
<tr>
<td>no precon AV eq. (22)</td>
<td>1</td>
<td>0.5</td>
<td>6.22</td>
</tr>
</tbody>
</table>

Table 1: Convergence rate

In conclusion these computations show that one can calculate both inviscid and viscous flows and even those with chemical reactions over a large range of Mach numbers going down to $M = 10^{-5}$ in some cases. There is need for further work on the effect of the parameters in the preconditioners on the convergence rates. It is not understood why constant $\beta$ seems to be the best choice. There is also need for further investigation on the effect of boundary conditions on these preconditioners.

acknowledgment
The author would like to thank Bram van Leer for many informative discussions related to the topics discussed here.

References


