RIACS

Analysis of a Parallel Multigrid Algorithm

Tony F. Chan
Ray S. Tuminaro

IN-62
43108
p.24

October 1989

Analysis of a Parallel Multigrid Algorithm

Tony F. Chan†
Ray S. Tuminaro††

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 89.41
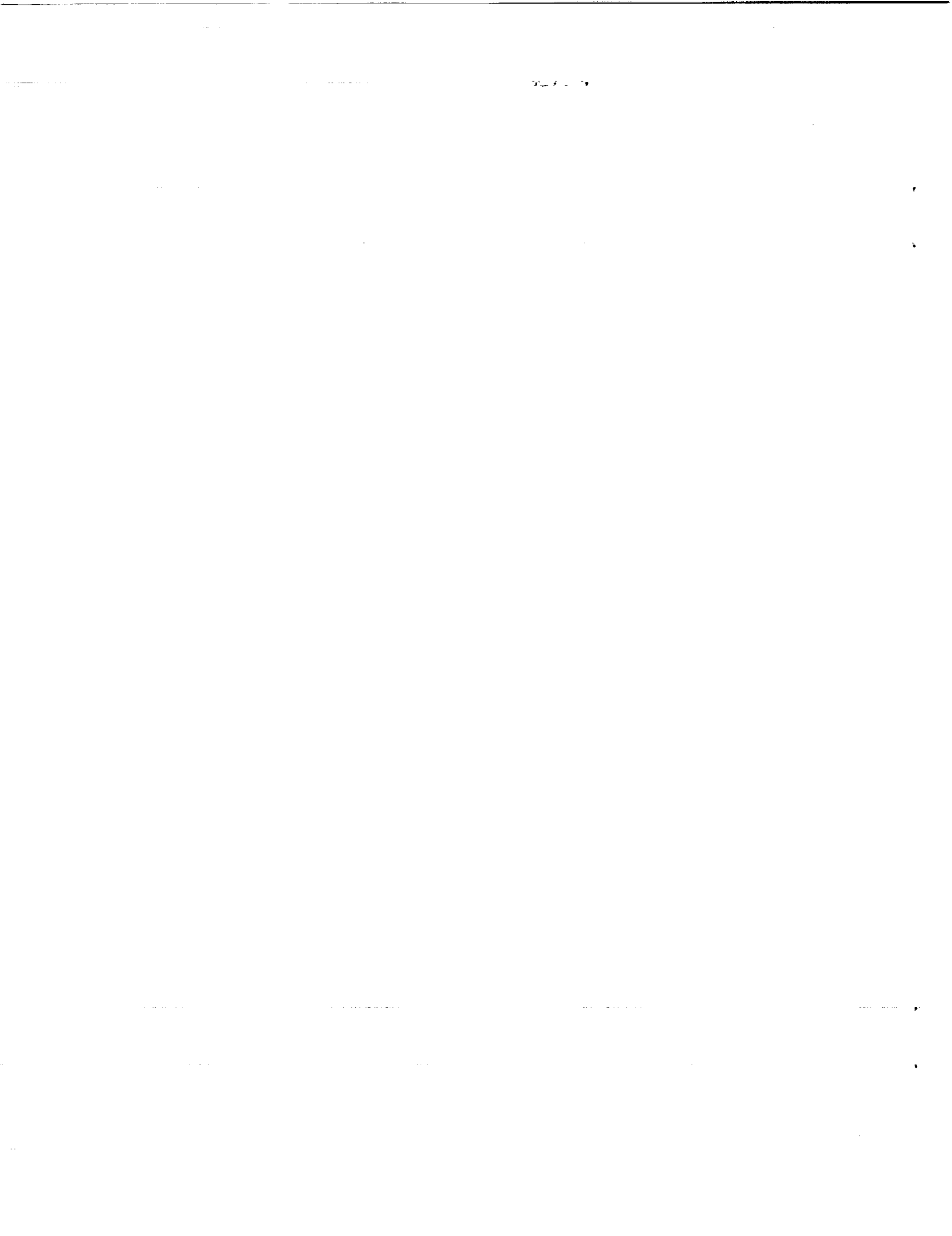October 1989

The introduction of parallel computers has motivated the search for algorithms which run well on them. In this search many new "multigrid- like" algorithms have been proposed. Being inherently more parallel than standard multigrid, these algorithms have the potential of being efficient on massive parallel machines. This paper considers the parallel multigrid algorithm of Frederickson and McBryan [13]. This algorithm uses multiple coarse grid problems (instead of one) in the hope of accelerating convergence. In this paper, we analyze the convergence properties of this new algorithm. This analysis reveals a close relationship with traditional multigrid methods. Specifically, the parallel coarse grid correction operator is identical to a traditional multigrid coarse grid correction operator except that the mixing of high and low frequencies caused by aliasing error is removed. We show how to choose appropriate relaxation operators to take advantage of this property. Comparisions between the standard multigrid and the new method are made.

1. Introduction. The multigrid algorithm is a fast, efficient method for solving a wide class of partial differential equations and is now used in many areas of scientific computing (such as computational fluid dynamics and structural mechanics) [3], [5], [16], [19]. Despite advances in both numerical algorithms and computer hardware, many applications require still greater performance than is offered by traditional computers. Given the success of the serial algorithm, it is natural to consider parallel multigrid algorithms.

Many successful parallel multigrid algorithms based on domain decomposition have been proposed and/or implemented [1], [2], [4], [7], [6], [9], [10], [14], [18], [20], [21], [23], [24], [25], [26], [27], [28], [29], [30], [11]. In these algorithms, parallelism is obtained by subdividing the physical region and assigning the subdivisions to different processors. The parallel algorithm is then identical to the serial algorithm, with each processor updating only grid points in its subdomain. Unfortunately when implemented on a massively parallel computer, this algorithm results in many inactive processors. Consider for example, a problem partitioned so that each processor contains only one point on the finest grid. If an h to $2h$ coarsening is used on a two dimensional problem, 3/4 of the processors are idle when processing the next coarsest grid. Even more processors are idle on still coarser levels. This idle processor problem has spurred research into multigrid-like methods suited for massively parallel systems. See [8], [12], [13], [15], [17], [22].

In this paper we consider a promising new parallel algorithm which avoids the idle processor problem, the Frederickson-McBryan parallel multigrid algorithm [13]. This algorithm (to be described in section 2) uses multiple coarse grid corrections to improve the convergence rate over that of the standard multigrid algorithm. The number of coarse grid corrections is matched so that the same number of grid points are processed when computing on a fine or coarse grid. For the one dimension Poisson equation it is a direct method and is in fact equivalent to odd-even reduction. In this paper we analyze the new algorithm for a model anisotropic problem. This analysis reveals a close relationship with traditional multigrid methods. Specifically, the parallel coarse grid correction operator is identical to a traditional multigrid coarse grid correction operator except that the mixing of high and low frequencies caused by aliasing error is removed. To take advantage of the aliasing removal, we define a new smoothing number (similar to the Brandt smoothing number) and a corresponding method for determining relaxation operators for the method. Comparisons between the new method and the standard algorithm are made.

2

2. Basic Algorithm. For simplicity, we consider only the situation of one grid point per processor. The principle idea in the Frederickson-McBryan algorithm is to use otherwise idle processors to perform additional coarse grid corrections and hopefully improve convergence. For example, on a one-dimensional problem using a two-level method, two coarse grid corrections are performed in parallel. Half the processors produce one correction by projecting the fine grid equations on the odd points and the other processors produce another correction by projection on the even points. The hope is that by combining these coarse grid corrections the convergence rate can be improved.

Below we summarize a 'V' cycle variant of a 2-dimensional algorithm using one post-relaxation sweep to solve the problem

$$(1) \qquad\qquad\qquad A_1 u = b.$$

```
proc multi( A_i, u, b, level )
{
  if ( level = CoarsestLevel ) then u = A_i^+ b
  else
    Relax(b, u, level)
    ComputeResidual(A_i, u, b, level, res)
    ProjResidual(res, r_oo, r_oe, r_eo, r_ee, level)
    multi( A_oo, r_oo, u_oo, level+1)
    multi( A_oe, r_oe, u_oe, level+1)
    multi( A_eo, r_eo, u_eo, level+1)
    multi( A_ee, r_ee, u_ee, level+1)
    Interpolate(u_oo, u_oe, u_eo, u_ee, level, correction)
    u = u+correction
  endif
}
```

A_i is the current fine grid operator, and A_i^+ is the corresponding pseudo-inverse of the fine grid operator.

A_{ee} is the coarse grid operator defined on the even points of the current fine grid,

3

A_{oo} is the operator defined on the odd points of the current fine grid,

A_{oe} is the operator defined on the odd points in the x direction and even points in the y direction, and

A_{eo} is the operator defined on the even points in the x direction and odd points in the y direction. It is further assumed that the same projection and interpolation stencils are used for each correction (distinguishing this algorithm from the highly parallel multigrid algorithms of Hackbusch [17] and Douglas-Miranker [12]).

The propagation of the error, e_k, for a 2 level method can be described by the matrix T:

$$(2) \qquad e_k = T e_{k-1}$$

where

$$(3) \qquad T = S(I - C),$$

$$(4) \qquad S = p(G) = 1 + \sum_{i=1}^{n} p_i G^i,$$

and

$$(5) \qquad C = P A_2^+ R A_1.$$

S represents the smoothing operator and is a matrix polynomial with coefficients p_i. For our discussion the iteration operator, G is simply A_1.

C represents the coarse grid correction operator.

P is the composite prolongation (interpolation) operator (applied at all points).

R is the composite restriction (projection) operator (applied at all points).

A_2 is the composite coarse grid operator applied at all points (i.e. a combination of A_{ee}, A_{oo}, A_{oe}, and A_{eo}).

A_1 is the fine grid operator.

The matrices defined above are square ($n^2 \times n^2$ entries for an $n \times n$ grid) as each operator is applied to every grid point. This is in direct contrast with the standard algorithm where some operators are applied only to a subset of the points resulting in rectangular matrices.

4

The principle difficultly with the above algorithm is to determine smoothing, interpolation and projection operators to produce convergence rates that are significantly better than the standard multigrid approach. The remainder of the paper will address this topic.

3. Fourier Analysis Framework. The model problem analysis of the parallel algorithm is somewhat easier than that of standard multigrid. This is primarily because the Fourier transform diagonalizes the iteration operator. We illustrate the approach for a two dimensional problem defined on the unit square with periodic boundary conditions:

$$(6) \qquad u_{xx} + \beta u_{yy} = f.$$

The basic analysis for the Poisson equation ($\beta = 1$) appears in [13]. Discretization by central differences for an $n \times n$ grid yields the linear system:

$$(7) \qquad A_1 u = h^2 f$$

with $h = 1/n$. From the previous section, the iteration operator is given by:

$$(8) \qquad e_k = p(A_1)(I - PA_2^+ RA_1)e_{k-1}.$$

A simplification occurs if we assume that the matrices P and R are circulant (i.e. the same stencil is applied at each point in the domain). Since A_2^+ and A_1 are also circulant, all the matrices in (8) commute. This implies that the individual choices of P and R are unimportant, only the quantity:

$$(9) \qquad Q = PR.$$

With this simplification the error is given by:

$$(10) \qquad e_k = p(A_1)(I - QA_2^+ A_1)e_{k-1}$$

For the remainder of this paper we assume that the restriction operator is simply point-wise injection (i.e. $R = I$), and the interpolation operator (hence Q) is given by a local symmetric 9 point stencil.

$$(11) \qquad Q = \begin{pmatrix} q_{11} & q_1 & q_{11} \\ q_1 & q_0 & q_1 \\ q_{11} & q_1 & q_{11} \end{pmatrix}$$

5

We do not consider stencils of greater width, or whether larger stencils yield significantly better convergence rates.

We transform the operators into Fourier space by considering

$$(12) \qquad u(x,y) = \sum_{1 \le |k,j| \le p} \bar{u}_{k,j} e^{2\pi i(kx+jy)} \quad \text{where} \quad p = \frac{n}{2}.$$

In Fourier space all the operators are diagonal with elements:

$$
\begin{aligned}
\bar{A}_1 &= 2(1 - x_1) + 2\beta(1 - x_2), \\
(13) \qquad \bar{A}_2 &= (1 - x_1^2) + \beta(1 - x_2^2), \\
\bar{Q} &= q_0 + 2q_1(x_1 + x_2) + 4q_{11}x_1 x_2.
\end{aligned}
$$

where

$$(14) \qquad x_1 = \cos(2\pi k/n) \quad \text{and} \quad x_2 = \cos(2\pi j/n).$$

Notice that $x_1 < 0$ ($x_1 > 0$) corresponds to high (low) frequency in the x direction and that $x_2 < 0$ ($x_2 > 0$) corresponds to high (low) frequency in the y direction. For now we take the coefficients p_i in the polynomial $p(x)$ and q_i in Q as unknowns. Using these we can write down an explicit expression for the error in the parallel algorithm and obtain an 'optimal' method by choosing the free parameters to minimize the error over the entire range. That is

$$(15) \qquad \min_{q_i's, p_i's} \quad \max_{-1 \le x_1, x_2 \le 1} |\bar{T}(x_1, x_2)|$$

where

$$(16) \qquad \bar{T}(x_1, x_2) = p(\bar{A}_1)[I - \bar{Q}\bar{A}_2^+ \bar{A}_1].$$

In [13], Frederickson and McBryan used error expressions derived for their parallel multigrid algorithm combined with an optimization routine to deduce optimal parameters. With these parameters, they illustrate that extremely fast convergence can be obtained. In the next sections the analysis is extended to deduce near optimal parameters analytically. Before proceeding with the smoothing, we explain why the parallel method yields faster convergence rates than the serial one.

6

4. Cancellation of Aliasing with Multiple Coarse Grids. In this section, we show that it is possible to eliminate most of the aliasing error associated with standard multigrid methods by using multiple coarse grid corrections. The result is shown for a one dimensional problem for simplicity. The same arguments extend to higher dimensional problems.

We first state a lemma which is used in the following theorem.

LEMMA 1.

(17)
$$R_o(\alpha v_l + \beta v_h) = \alpha \tilde{v}_l - \beta \tilde{v}_{-l} \quad for \ |l| \le \frac{n}{4},$$

$$R_e(\alpha v_l + \beta v_h) = \alpha \hat{v}_l + \beta \hat{v}_{-l} \quad for \ |l| \le \frac{n}{4},$$

$$R_o^T(\alpha \tilde{v}_l + \beta \tilde{v}_{-l}) = \alpha(v_l - v_{-h}) + \beta(v_{-l} - v_h) \quad for \ |l| \le \frac{n}{4},$$

$$R_e^T(\alpha \hat{v}_l + \beta \hat{v}_{-l}) = \alpha(v_l + v_{-h}) + \beta(v_{-l} + v_h) \quad for \ |l| \le \frac{n}{4},$$

where

(18)
$$[v_k]_j = e^{2\pi i k(j/n)} \qquad |k| = 1, ..., \frac{n}{2} \ and \ j = 1, ..., n,$$

$$[\tilde{v}_k]_j = e^{4\pi i k(j/n)} \qquad |k| = 1, ..., \frac{n}{4} \ and \ j = 1, ..., \frac{n}{2},$$

$$[\hat{v}_k]_j = e^{2\pi i (2k-1)(j/n)} \quad |k| = 1, ..., \frac{n}{4} \ and \ j = 1, ..., \frac{n}{2},$$

R_e and R_o denote injection onto the even and odd points respectively, and $h = \frac{n}{2} - l$ when $l > 0$ and $h = \frac{-n}{2} + l$ when $l < 0$.

PROOF 1. *For $l > 0$:*

(19)
$$R_e(\alpha v_l + \beta v_h) = \alpha e^{4\pi i l(j/n)} + \beta e^{4\pi i h(j/n)}$$
$$= \alpha \tilde{v}_l + \beta e^{4\pi i j(n/2-l)/n}$$
$$= \alpha \tilde{v}_l + \beta [e^{-4\pi i l(j/n)} e^{2\pi i j}]$$
$$= \alpha \tilde{v}_l + \beta \tilde{v}_{-l}.$$

and

(20)
$$R_o(\alpha v_l + \beta v_h) = \alpha e^{2\pi i l(2j-1)/n} + \beta e^{2\pi i h(2j-1)/n}$$
$$= \alpha \hat{v}_l + \beta e^{\pi i (n-2l)(2j-1)/n}$$
$$= \alpha \hat{v}_l + \beta [e^{-2\pi i l(2j-1)/n} e^{(2j-1)\pi i}]$$
$$= \alpha \hat{v}_l - \beta \hat{v}_{-l}$$

7

We summarize R_e and R_o by

$$(21) \qquad R_e = (\hat{V}_l \quad \hat{V}_{-l}) \begin{pmatrix} I & 0 & 0 & I \\ 0 & I & I & 0 \end{pmatrix} \begin{pmatrix} V_l^T \\ V_h^T \\ V_{-l}^T \\ V_{-h}^T \end{pmatrix}$$

$$(22) \qquad R_o = (\bar{V}_l \quad \bar{V}_{-l}) \begin{pmatrix} I & 0 & 0 & -I \\ 0 & -I & I & 0 \end{pmatrix} \begin{pmatrix} V_l^T \\ V_h^T \\ V_{-l}^T \\ V_{-h}^T \end{pmatrix}$$

where

$$(23) \qquad v_l \in V_l, \quad v_h \in V_h, \quad v_{-l} \in V_{-l}, \quad v_{-h} \in V_{-h},$$

$$(24) \qquad \hat{v}_l \in \hat{V}_l, \quad \hat{v}_{-l} \in \hat{V}_{-l}, \quad \bar{v}_l \in \bar{V}_l, \quad and \quad \bar{v}_{-l} \in \bar{V}_{-l}.$$

The properties of R_e^T and R_o^T follow from (21) and (22).

The lemma states how the lth Fourier mode on the fine grid (v_l), transforms when it is projected onto the coarse grid. Specifically, if the lth mode corresponds to a low frequency $(|l| \leq n/4)$, then it appears as the lth Fourier mode on the coarse grid. However, if the mode (v_h) corresponds to high frequency $(|h| > n/4)$, then it is aliased and appears as the $-l^{th}$ mode on the coarse grid. The essential point is that the aliasing on even and odd grid points is of opposite sign.

We illustrate the cancellation of aliasing by defining two coarse grid correction methods: single coarse grid correction (SC) which solves on the even points, and multiple coarse grid correction (MC) which solves on both the even and odd points as described in section 2.

$$(25) \qquad C_{SC} = P_1 A_e^+ R_1$$

and

$$(26) \qquad C_{MC} = .5 * (P_1 A_e^+ R_1 + P_2 A_o^+ R_2)$$

where P_1, R_1, P_2, R_2 correspond to interpolation and restriction on the even and odd points respectively. A_e and A_o denote the coarse grid differential operator on the even and odd points respectively. We assume that the interpolation and restriction operators employ the same stencil for both coarse grids. Notice that the multiple coarse grid correction (26) is simply the average of two standard coarse grid corrections on both the even and odd points. For the theorem that follows, we assume that the operators A_e and A_o are identical except shifted to operate on different points. Additionally, we assume that the operators treat a positive ($l > 0$) and negative ($l < 0$) mode identically. These assumptions corresponds to the following general representations of the pseudo-inverses for arbitrary $\bar{v}_l$ and $\hat{v}_l$

$$(27) \qquad A_e^+ \bar{v}_l = \sum_{i=1}^n k_{i,|l|} \bar{v}_l \quad \text{and} \quad A_o^+ \hat{v}_l = \sum_{i=1}^n k_{i,|l|} \hat{v}_l$$

with the same coefficients ($k_{i,j}$) for both operators. We note that all these assumptions are not unrealistic and hold for periodic constant coefficient operators.

Without loss of generality, we write the interpolation and restriction operators as:

$$(28) \qquad P_1 = Z R_e^T, \quad P_2 = Z R_o^T, \quad R_1 = R_e Y \quad \text{and} \quad R_2 = R_o Y$$

where R_e and R_o are injection operators on the even and odd points, and Z and Y are general $n \times n$ circulant matrices. A property of circulant matrices is that the v_l's defined in the lemma are eigenvectors. We denote the corresponding eigenvalues as z_l and y_l. That is

$$(29) \qquad Z v_l = z_l v_l \quad \text{and} \quad Y v_l = y_l v_l.$$

We now state the theorem.

THEOREM 1. *Consider a splitting of the Fourier modes (V) into high and low frequency:*

$$(30) \qquad V = [V_l, V_h]$$

where

$$(31) \qquad V_l = span_{1 \le |k| \le n/4}\{v_k\}, \quad and \quad V_h = span_{n/4 \le |k| \le n/2}\{v_k\}.$$

Using this Fourier mode splitting, we define a partioning of the operator C_{SC} in Fourier space:

$$(32) \qquad \bar{C}_{SC} = \begin{pmatrix} F_1 & F_3 \\ F_4 & F_2 \end{pmatrix}$$

where

$$(33) \qquad \bar{C}_{SC} = [V_l, V_h] C_{SC} [V_l, V_h]^T.$$

Then, $\tilde{C}_{MC}$ is given by:

$$(34) \qquad \bar{C}_{MC} = \begin{pmatrix} F_1 & 0 \\ 0 & F_2 \end{pmatrix}$$

where

$$(35) \qquad \bar{C}_{MC} = [V_l, V_h] C_{MC} [V_l, V_h]^T.$$

PROOF 2.

We look at a matched pair (l, h) from the low and high frequency space:

$$(36) \qquad v_l \in V_l \quad \text{and} \quad v_h \in V_h.$$

Then

$$
\begin{aligned}
(37) \qquad C_{SC}(\alpha v_l + \beta v_h) &= P_1 A_e^+ R_1 (\alpha v_l + \beta v_h) \\
&= Z R_e^T A_e^+ R_e Y (\alpha v_l + \beta v_h) \\
&= Z R_e^T A_e^+ R_e (\alpha y_l v_l + \beta y_h v_h) \\
&= Z R_e^T A_e^+ (\alpha y_l \tilde{v}_l + \beta y_h \tilde{v}_{-l}) \\
&= Z R_e^T (\alpha y_l \tilde{v}_l + \beta y_h \tilde{v}_{-l}) \sum_{i=1}^{n} k_{i|l|}.
\end{aligned}
$$

$$
\begin{aligned}
(38) \qquad C_{SC}(\alpha v_l + \beta v_h) &= Z(\alpha y_l (v_l + v_{-h}) + \beta y_h (v_{-l} + v_h)) \sum_{i=1}^{n} k_{i|l|} \\
&= \alpha y_l \sum_{i=1}^{n} k_{i|l|} z_l v_l + \beta y_h \sum_{i=1}^{n} k_{i|l|} z_{-l} v_{-l} + \alpha y_l \sum_{i=1}^{n} k_{i|l|} z_{-h} v_{-h} + \\
&\quad \beta y_h \sum_{i=1}^{n} k_{i|l|} z_h v_h.
\end{aligned}
$$

A similar procedure for the odd points yields:

$$P_2 A_o^+ R_2(\alpha v_l + \beta v_h) = \alpha y_l \sum_{i=1}^{n} k_{i|l|} z_l v_l - \beta y_h \sum_{i=1}^{n} k_{i|l|} z_{-l} v_{-l} - \alpha y_l \sum_{i=1}^{n} k_{i|l|} z_{-h} v_{-h} +$$

(39)

$$\beta y_h \sum_{i=1}^{n} k_{i|l|} z_h v_h.$$

Averaging (38) and (39) yields:

(40)
$$C_{MC}(\alpha v_l + \beta v_h) = \alpha y_l \sum_{i=1}^{n} k_{i|l|} z_i v_i + \beta y_h \sum_{i=1}^{n} k_{i|l|} z_h v_h.$$

Notice that the αv_{-h} and the βv_{-l} terms in (38) and (39) cancel when they are averaged. Additionally, the first term in (38) corresponds to F_1, the second term to F_3, the third term to F_4 and the fourth term to F_2. It is clear that the multiple grid method yields the identical expression with the terms from F_3 and F_4 removed.

————-

The theorem states that for every multiple grid method of the type described in section 2, there is a closely related single grid method. In particular, the multiple coarsening correction (MC) is identical to the standard one (SC) except that the mixing of low and high frequencies (F_3 and F_4) is removed. These terms F_3 and F_4 correspond to an artificial mixing of high and low frequencies introduced by the method. Typically, these terms degrade the performance of the multigrid algorithm as the aliased frequencies on the coarse grid bear no physical relationship with the original frequencies on the fine grid. While F_3 and F_4 are error terms, F_1 is responsible for the fast convergence of the multigrid algorithm. That is, the coarse grid correction accurately reflects the low frequency behavior of the fine grid. Since the multiple coarsening method retains the low frequency behavior of the single grid method without the aliasing error, it seems logical that the multiple coarsening method will converge faster than the standard method.

5. Interpolation. In this section, we motivate our use of a scaled bilinear interpolation operator for the new method based on the close connection between the new and standard coarse grid corrections.

The essential idea is to choose an interpolation operator Q that ensures a fast convergence rate for the multiple coarsening algorithm. Notice that the standard multigrid convergence depends on the low to low frequency operator F_1 which in turn depends on the choice of Q. Since the low-to-low frequency operator in the MC method is the same as

that in the standard method, it seems intuitive to choose the same interpolation operator for both methods. Therefore, for the MC algorithm, we are led to the choice of bilinear interpolation which has been shown to be a good interpolation operator in standard multigrid algorithms for the solution of second order equations. This leads to the following for Q:

$$(41) \qquad Q = \frac{1}{4} \begin{pmatrix} .25 & .5 & .25 \\ .5 & 1.0 & .5 \\ .25 & .5 & .25 \end{pmatrix}.$$

Note that the scaling factor $1/4$ is needed because we are averaging 4 interpolants from multiple coarse grids. This is in fact the same operator that is used in [13] for the Poisson equation and is proved to be the 'best' using Fourier Analysis arguments. For the remainder of this paper we assume that the interpolation operator is given by (41). Of course, the possibility exists that it may be better to use interpolation with wider stencils in the new parallel method but we shall not explore that here.

6. Smoothing Criteria/Richardson Relaxation. As we shall see, the typical relaxation criteria for standard multigrid methods is not entirely appropriate for the Frederickson-McBryan method. In this section we develop a new criteria for the parallel method by making use of the close relationship between the standard and new methods. We begin by studying Richardson relaxation as a smoothing operator within multigrid methods.

A general Richardson relaxation method can be described by a matrix polynomial. For example, the Richardson method for the model problem in section 3 can be written as:

$$(42) \qquad S = P(A_1) = 1 + \sum_{i=1}^{n} p_i A_1^i.$$

Given such a scheme, the key question is how to determine the p_i's so that the multigrid method converges rapidly. Probably, the most popular criteria for these parameters is that they minimize the Brandt smoothing number [3]. For our model problem, the Brandt number is defined as follows:

$$(43) \qquad \mu_b = \max |\bar{S}(x_1, x_2)| \quad \text{for} \quad x_1 \leq 0 \quad \text{or} \quad x_2 \leq 0,$$

12

where S in Fourier space is:

$$(44) \qquad \bar{S}(x_1, x_2) = 1 + \sum_{i=1}^{n} p_i [\bar{A}_1(x_1, x_2)]^i$$

and x_1, x_2, and $\bar{A}_1$ are defined in section 3. The Brandt number measures the least that any high frequency error is damped after the relaxation is applied. The intuition is that the coarse grid correction effectively solves in the low frequency range. Thus, the smoothing operator must only damp the error in the high frequency range. Smoothing parameters deduced by minimizing the Brandt smoothing number are usually easy to compute for model problems and work well in multigrid algorithms. In fact, for many problems (including our model problem) the convergence rate for a standard two-level multigrid method (for example using point-wise injection and bilinear interpolation) is approximately equal to the Brandt smoothing number and is close to optimal over all possible p_i's.

EXAMPLE 1. *Consider the model problem*

$$(45) \qquad u_{xx} + \beta u_{yy} = f$$

with periodic boundary conditions defined on the unit square. The Brandt number for this problem can be written as:

$$(46) \qquad \mu_b = \max |\bar{S}(\bar{A}_1)| = |1 + \sum_{i=1}^{n} p_i \bar{A}_1| \quad for \quad 2 \leq \bar{A}_1 \leq 4 + 4\beta.$$

where the Brandt number ranges given in (43) have been transformed into functions of A_1. The minimization of (46) over all p_i's is given by a Tchebycheff polynomial. For $n = 1$, the optimum is:

$$(47) \qquad T(\bar{A}_1) = 1 - \frac{\bar{A}_1}{3 + 2\beta}$$

with

$$(48) \qquad p_1 = \frac{-1}{3 + 2\beta} \quad and \quad \mu_b = \frac{2\beta + 1}{3 + 2\beta}.$$

It is natural to ask how well this choice of p_i performs for a multiple correction method. Specifically, consider a two-level single correction method using one iteration of this smoothing criteria, point-wise injection, and bilinear interpolation. We can compare
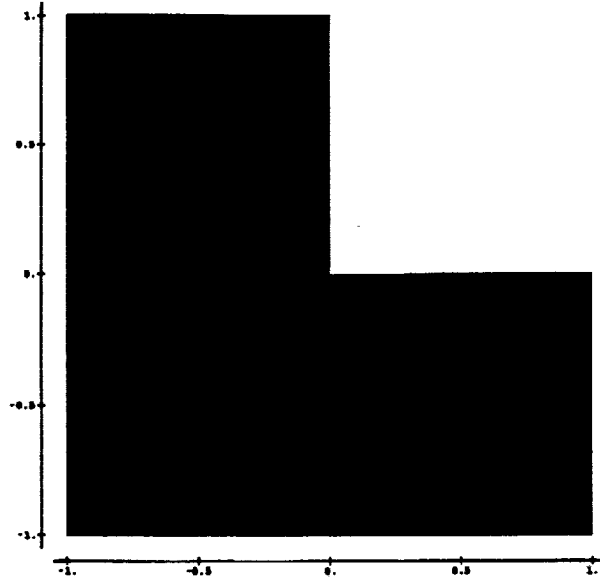
FIG. 1. *Dark region indicates areas where weight function,* $\tilde{N}$, *is equal to 1 in Fourier space. Note: high frequency in both directions corresponds to lower left corner of diagram.*

this with a multiple correction method, whose coarse grid correction is simple the average of four single corrections. We omit the detailed results and simply state that there is almost no improvement using the parallel method over the serial method when the Brandt smoothing number criteria is used (i.e. the convergence rates are identical). This is because the Brandt smoothing number criteria equi-distributes the errors in F_2, F_3, F_4. Now that F_3 and F_4 are zero, we should chose the relaxation operator to minimize the effect of F_2 only.

To deduce a new smoothing criteria, we examine the motivation behind the Brandt number. Specifically, we view the Brandt criteria as a simplification of the general optimization problem. That is, we wish to minimize the spectral radius of the multigrid iteration operator. The optimal parameters, p_i's, are given by

$$(49) \qquad \min_{p_i'\text{s}} \rho(T) = \min_{p_i'\text{s}} \rho(p(A_1)(I - C)).$$

One can view the coarse grid correction, $I - C$, as a somewhat complex weight function applied to $p(A_1)$. We therefore, replace it by a simplier weight function, N. In Fourier space this operator is defined as follows:

$$(50) \qquad \tilde{N}(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 \leq 0 \text{ or } x_2 \leq 0 \\ 0 & \text{if } x_1 > 0 \text{ and } x_2 > 0 \end{cases}$$
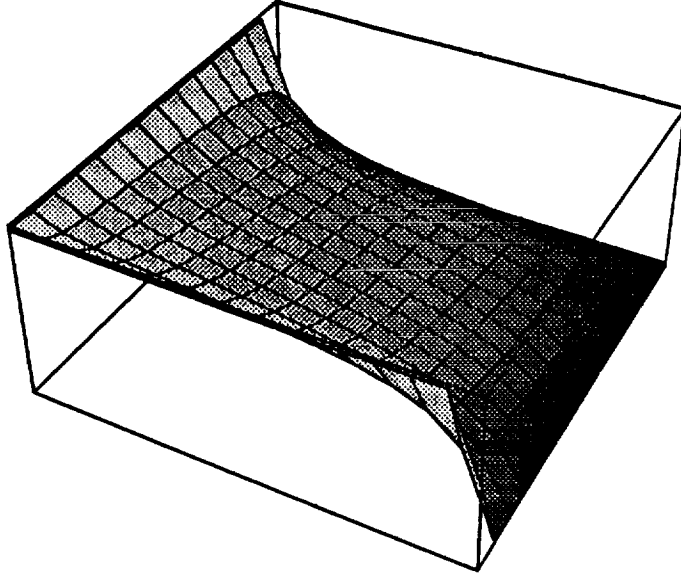
14

FIG. 2. *Damping of Fourier components with the multiple coarse grid correction. High frequency in both directions corresponds to leftmost corner in the diagram where the damping is equal to 1.*

and is depicted in figure 1. The optimization problem with N replacing C is equivalent to minimizing the Brandt smoothing number. The definition of N is motivated by the behavior of smoothing operators as well as the behavior of coarse grid corrections. Specifically, the coarse grid correction provides no improvement over the high frequencies and so this error must be entirely damped by the smoothing operator. Second, almost all damping of low frequencies is effectively handled by the coarse grid correction.

Unfortunately for the multiple grid method, the Brandt smoothing number is not entirely appropriate. However, we can define a new smoothing number by simply choosing a new weight function based on the behavior of the multiple grid correction. Intuitively, the removal of the aliasing error should significantly improve the coarse grid correction over the high and middle frequencies because they contribute more to aliasing error than the low frequencies. Smoothing is still needed to compensate for the fact that the coarse grid operators do not accurately reflect the fine grid operator over the high frequencies. However, the removal of the aliasing error implies that the smoother should focus on damping the highest frequency modes (which are most poorly represented on the coarse grid). Specifically for our model Poisson problem, the damping of the Fourier modes with the coarse grid correction (depicted in figure 2) is given by:

$$(51) \qquad \lambda(I - C(x_1, x_2)) = 1 - \frac{1}{2}(1 + x_1)(1 + x_2)\frac{(1 - x_1) + (1 - x_2)}{(1 - x_1^2) + (1 - x_2^2)}$$
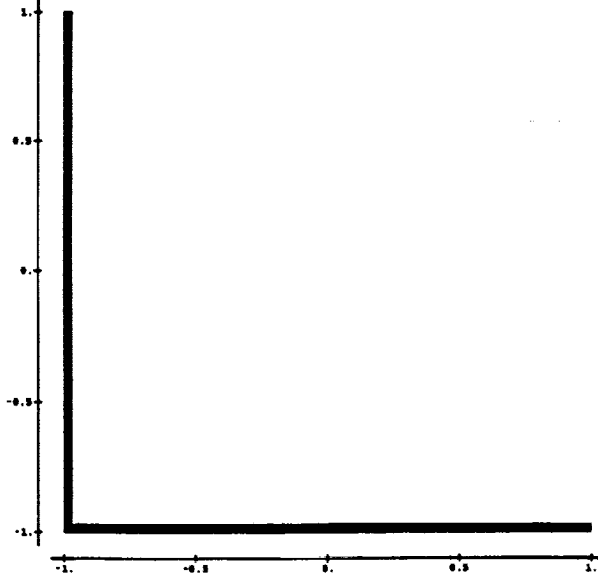
15

FIG. 3. *Dark region indicates areas where weight function, $\bar{N}$, is equal to 1 in Fourier space. High frequency in both directions corresponds to lower left corner of diagram.*

where $\lambda()$ denotes the damping of a Fourier mode due to the coarse grid correction. Further, we observe that

$$(52) \qquad \lambda(I - C(x_1, x_2)) = 1 \quad \text{when} \quad x_1 = -1 \quad \text{or} \quad x_2 = -1$$

and

$$(53) \qquad \lambda(I - C(x_1, x_2)) < 1 \quad \text{for} \quad x_1 > -1 \quad \text{and} \quad x_2 > -1.$$

That is, the coarse grid correction even damps most of the high frequency errors. This is a consequence of the absence of aliasing error (which degrades the performance of the coarse grid correction). However, since the coarse grid correction does not damp the highest frequency modes ($x_1 = -1$ or $x_2 = -1$), all damping of these components must come from the smoothing operator. Based on these observation, we define a new weight function corresponding to figure 3 given by:

$$(54) \qquad \bar{N}(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = -1 \text{ or } x_2 = -1 \\ \\ 0 & \text{if } x_1 \neq 1 \text{ and } x_2 \neq -1 \end{cases}$$

and a corresponding *parallel* smoothing number:

$$(55) \qquad \mu_p = \max S(x_1, x_2) \quad x_1 = -1 \text{ or } x_2 = -1.$$

16

Assuming that the coarse grid correction damps low frequencies sufficiently, we expect that the multigrid convergence rate will be close to the smoothing rate. We emphasize that this new smoothing number is a heuristic based on numerical experimentation.

We demonstrate the use of this new smoothing number with a few examples.

EXAMPLE 2. $n = 1$ *(damped Jacobi smoothing)*.

We determine the extremal values of $\bar{A}_1$ over the intervals given by $x_1 = -1$ and $x_2 = -1$. For our model problem, the parallel smoothing number is then given by:

$$(56) \qquad \mu_p = \max_{4 \leq \bar{A}_1 \leq 4+4\beta} |1 + p_1 \bar{A}_1|$$

which is minimized by the Tchebycheff polynomial

$$(57) \qquad T(\bar{A}_1) = k[\bar{A}_1 - (4 + 2\beta)]$$

with

$$(58) \qquad p_1 = k = \frac{-1}{4 + 2\beta}, \quad \text{and} \quad \mu_p = \frac{\beta}{2 + \beta}.$$

EXAMPLE 3. $n = 2$ *(two step smoothing)*.

The parallel smoothing number is given by

$$(59) \qquad \mu_p = \max_{4 \leq \bar{A}_1 \leq 4+4\beta} |1 + p_1 \bar{A}_1 + p_2 \bar{A}_1^2|.$$

which is minimized by the Tchebycheff polynomial:

$$(60) \qquad k[\bar{A}_1^2 - (8 + 4\beta)\bar{A}_1 + 2\beta^2 + 16\beta + 16].$$

Therefore

$$(61) \qquad p_2 = k = \frac{1}{2\beta^2 + 16\beta + 16}, \quad p_1 = \frac{-8 + -4\beta}{2\beta^2 + 16\beta + 16},$$

and

$$(62) \qquad \mu_p = \frac{\beta^2}{\beta^2 + 8\beta + 8}.$$

Tables 1 and 2 illustrates the results for the Frederickson-McBryan method using the smoothers corresponding to these examples. Specifically, the tables compare the two-level parallel multigrid convergence rate using this smoothing with one obtained using a

β	p_1	$\mu_p(\text{MC})$	$\rho(\text{MC})$	optimum $\rho(\text{MC})$
1.	-.16667	.3333	.3333	.3333
2.	-.125	.5	.5	.498
3.	-.1	.6	.6	.598
4.	-.08333	.6667	.667	.664

TABLE 1

Comparison of convergence rate of parallel method which minimizes parallel smoothing number vs. the optimized parallel method ($n = 1$) on a 64×64 grid.

β	p_1	p_2	$\mu_p(\text{MC})$	$\rho(\text{MC})$	optimum $\rho(\text{MC})$
1.	-.352941	.0294118	.058824	.103	.0938
2.	-.285714	.0178571	.142857	.143	.142
3.	-.243902	.0121951	.219512	.220	.217
4.	-.214286	.0089286	.285714	.286	.283

TABLE 2

Comparison of convergence rate of parallel method which minimizes parallel smoothing number vs. the optimized parallel method ($n = 2$) on a 64×64 grid.

numeric optimization routine which chooses the relaxation parameters, p_i, to minimize the overall convergence rate of the two level multigrid process.

Not only does the new smoothing number produce near optimal relaxation parameters for the model problem, but it also predicts the convergence rate (similar to the Brandt number for the standard multigrid method). With the exception of $\beta = 1$, the convergence rates and the smoothing number are nearly identical. When $\beta = 1$ however, the assumption that the coarse grid correction sufficiently damps the low frequencies is not valid. That is, the smoothing damps the high frequencies better than the coarse grid correction is removing the low frequency components. Nonetheless, the convergence rates obtained are not significantly worse than the optimal. It should be noted that this same phenomena occurs with the Brandt smoothing number. That is, the Brandt number does not accurately reflect the convergence rate when the smoothing number is very small. Typically when this happens, it implies that more smoothing was done than was necessary. In other words, the relaxation operator is smoothing the high frequencies better than the coarse grid correction is removing low frequency errors.

β	$n=1$		$n=2$		$n=3$	
	$\rho(\text{SC})$	$\rho(\text{MC})$	$\rho(\text{SC})$	$\rho(\text{MC})$	$\rho(\text{SC})$	$\rho(\text{MC})$
1.	.6	.333	.220	.103	.0739	.074
2.	.714	.5	.342	.143	.148	.068
3.	.778	.6	.434	.220	.215	.0740
4.	.818	.667	.503	.286	.275	.111
5.	.846	.714	.558	.342	.327	.148
6.	.867	.75	.601	.392	.373	.182
7.	.882	.778	.637	.434	.413	.215
8.	.895	.8	.667	.471	.448	.246
9.	.905	.818	.693	.503	.480	.275
10.	.913	.833	.715	.532	.508	.302

TABLE 3

Comparison of convergence rates for serial (with optimum Brandt smoothing number) and parallel multigrid (with optimum parallel smoothing number) for a 64 × 64 grid.

7. Comparisons with Standard Multigrid.

In this section we compare the new algorithm with the standard multigrid algorithm on our 2-dimensional model problem for a 64 × 64 grid. The standard multigrid method employs a point-wise injection, bilinear interpolation, and a smoothing algorithm which minimizes the Brandt smoothing number. The parallel algorithm uses the same grid transfer operators and minimizes the parallel smoothing number.

Table 3 shows the convergence rates using a 1, 2 and 3 step relaxation algorithm. From the table, we notice that the multiple grid method always yields better convergence rates than the single grid method. Finally in table 4, we compare the serial and the parallel methods using smoothing schemes which result in a convergence rate of one half for the serial method. This is accomplished by choosing both the smoothing operator and the number of smoothing iterations, and varying β to obtain a convergence rate of one half. One iteration of the smoothing operator corresponds to the polynomial which minimizes the Brandt smoothing number for the standard method and minimizing the parallel smoothing number for the parallel method. The degree of the minimizing polynomial is given by the column 'degree' and the number of iterations of the smoothing iteration is

| degree	sweeps	β	$\rho(\text{ SC })$	$\rho(\text{ MC })$
1.	2.	1.914121	.5	.239161
1.	3.	3.34732	.5	.245292
1.	4.	4.78521	.5	.247372
2.	1.	3.94949	.5	.28261
2.	2.	9.63334	.5	.272215
3.	1.	9.71329	.5	.294394

TABLE 4

Comparison of convergence rate of parallel method vs. standard method using different smoothing schemes which correspond to a serial convergence rate of one half.

given by 'sweeps.' From the table, we can conclude that when the standard method yields a convergence rate of one half, the corresponding parallel method yields a convergence rate close to one quarter for this problem. Obviously, additional tests must be performed to fully assess the potential of this parallel method.

8. Conclusions. We have analyzed the Frederickson and McBryan parallel multigrid algorithm and have shown that it can produce convergence rates that are significantly better than the standard multigrid method. We have shown that the reason for this success is that the mixing of high and low frequencies due to aliasing error is removed. In order to take advantage of this, however, the relaxation operator must be appropriately chosen. To this end, we defined a new smoothing number which reflects the behavior of the new multiple grid correction. In general, the relaxation parameters chosen by this smoothing criteria are nearly optimal for our model problem, and result in substantial computational savings over the standard method. More tests with variable coefficient problems as well as more severely anisotropic problems using more sophisticated relaxation operators are planned.

REFERENCES

[1] E. BARSZCZ, T. CHAN, D. JESPERSEN, AND R. TUMINARO, *Performance of a Parallel Euler Equation Code On Hypercube Computers*, Tech. Rep., NASA Ames, Moffett Field, Ca, 1989.

[2] M. BASSETT, *Implementation of multigrid on a hypercube multiprocessor*, in Proceedings of the First Conference on Hypercube Multiprocessors, M. Heath, ed., Siam, Philadelphia, 1985, pp. 211–220.

[3] A. BRANDT, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.

[4] B. BRIGGS, L. HART, S. MCCORMICK, AND D. QUINLAN, *Multigrid methods on a hypercube*, in Proceedings of the Third Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., Marcel Dekker, NY, 1987, pp. 63-84.

[5] W. BRIGGS, *Multigrid Tutorial*, SIAM, Philadelphia, 1987.

[6] T. CHAN AND Y. SAAD, *Multigrid algorithms on the hypercube multiprocessor*, IEEE Trans. Comp., C-35 (1986), pp. 969–977.

[7] T. CHAN AND R. SCHREIBER, *Parallel networks for multi-grid algorithms: architecture and complexity*, SIAM J. Sci. Stat. Compu., 6 (1985), pp. 698–711.

[8] T. CHAN AND R. TUMINARO, *Design and implementation of parallel multigrid algorithms*, in Proceedings of the Third Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., Marcel Dekker, NY, 1987, pp. 101–115.

[9] ——, *Designing Parallel Multigrid Algorithms*, Tech. Rep. 87.21, Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffett Field, Ca., Aug. 1987.

[10] ——, *Multigrid algorithms on hypercube processors*, in Proceedings of the Second Conference on Hypercube Multiprocessors, M. Heath, ed., Siam, Philadelphia, 1987, pp. 730-737

[11] E. V. DE VELDE AND H. KELLER, *The design of a parallel multigrid algorithm*, in Proceedings of the Second International Conference on Supercomputing, L. Kartashev and S. Kartashev, eds., 1987, pp. 76–83.

[12] C. DOUGLAS AND W. MIRANKER, *Constructive interference in parallel algorithms*, SIAM Journal on Numerical Analysis, 25 (1987), pp. 376–398.

[13] P. FREDERICKSON AND O. MCBRYAN, *Parallel superconvergent multigrid*, in Proceedings of the Third Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., Marcel Dekker, NY, 1987, pp. 195–210.

[14] D. GANNON AND J. V. ROSENDALE, *On the structure of parallelism in a highly concurrent pde solver*, Journal of Parallel and Distributed Computing, 3 (1986), pp. 106–135.

21
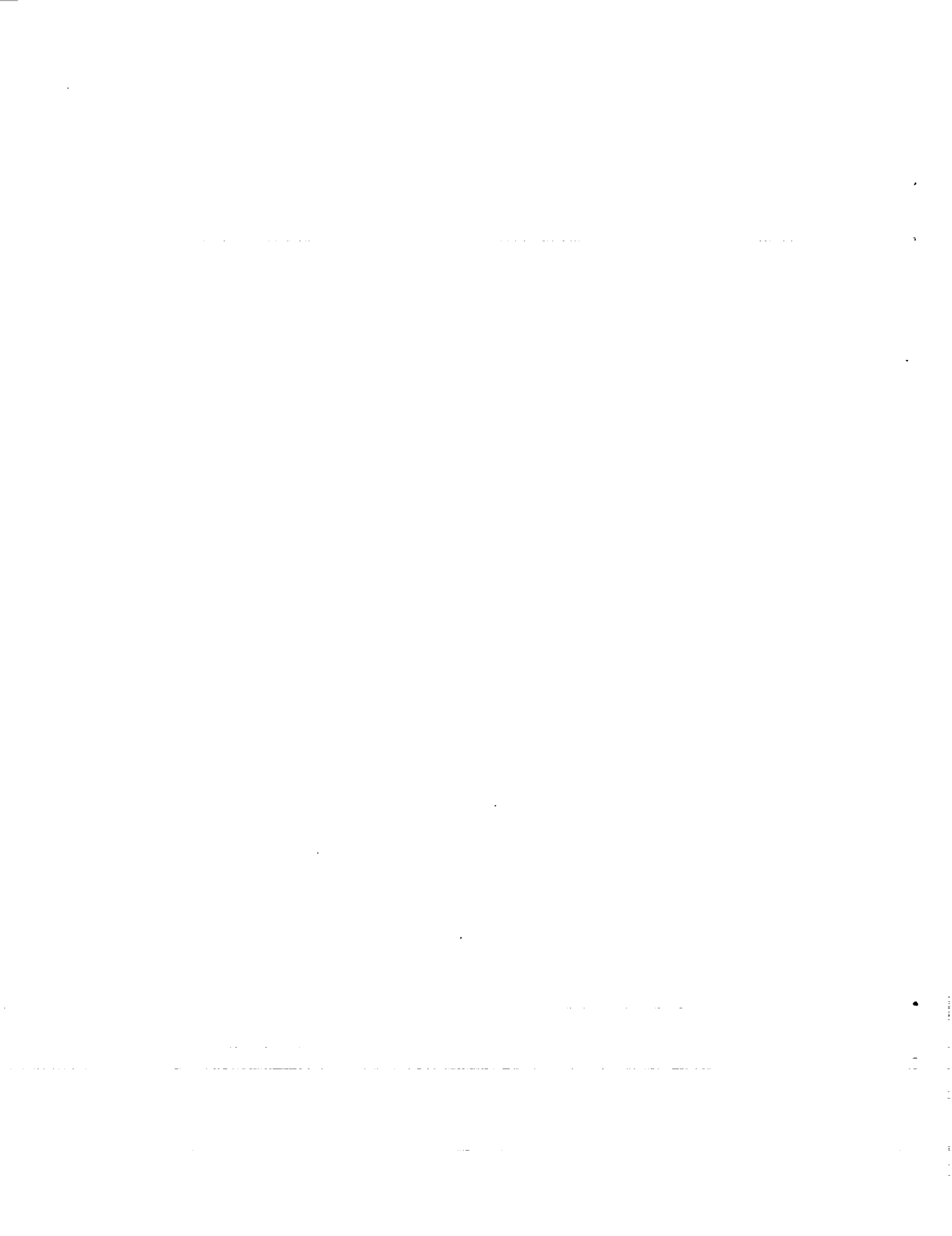
[15] A. GREENBAUM, *A multigrid method for multiprocessors*, in Proceedings of the Second Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., 1986, pp. 75–88.

[16] W. HACKBUSCH, *Multi-grid Methods and Applications*, Springer-Verlag, Berlin, 1985.

[17] ——, *A new approach to robust multi-grid methods*, in First International Conference on Industrial and Applied Mathematics, Paris, 1987.

[18] H. HOPPE AND H. MUHLENBEIN, *Parallel adaptive full-multigrid methods on message-based multiprocessors*, Parallel Computing, 3 (1986), pp. 269–287.

[19] D. JESPERSEN, *Multigrid methods for partial differential equations*, in Studies in Numerical Analysis, G. Golub, ed., 1984, pp. 270–318.

[20] G. JOHNSON AND J. SWISSHELM, *Multigrid for parallel-processing supercomputers*, paper given at the Third Copper Mountain Conference on Multigrid Methods, Copper Mountain, Colorado, April 5-10, 1987.

[21] O. KOLP AND H. MIERENDORFF, *Efficient multigrid algorithms for locally constrained parallel systems*, in Proceedings of the Second Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., 1986, pp. 169–200.

[22] C. KUO, *Parallel Algorithms and Architectures for Solving Elliptic Partial Differential Equations*, M.I.T. Master's thesis, Jan. 1985.

[23] O. MCBRYAN, *The Connection Machine: PDE Solution on 65536 Processors*, Tech. Rep. Research Report LA-UR-86-4218, Los Alamos National Laboratory, 1987.

[24] O. MCBRYAN AND E. V. DE VELDE, *Multigrid method on parallel processors*, in Multigrid Methods II, W. Hackbusch and U. Trottenberg, eds., Springer-Verlag, Berlin, 1986, pp. 232–260.

[25] S. MCCORMICK AND D. QUINLAN, *Multilevel load balancing for multiprocessors - an outline*, paper given at the Third Copper Mountain Conference on Multigrid Methods, Copper Mountain, Colorado, April 5-10, 1987.

[26] H. MIERENDORFF, *Parallelization of multigrid methods with local refinements for a class of non-shared memory systems*, in Proceedings of the Third Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., Marcell Dekker, NY, 1987, pp. 449-466.

[27] V. NIAK AND S. TA'ASAN, *Performance Studies of the Multigrid Algorithms Implemented on Hypercube Multiprocessor Systems*, Tech. Rep. 87-5, ICASE, NASA Langley Research Center, Hampton, VA, Jan. 1987.

[28] J. SWISSHELM, G. JOHNSON, AND S. KUMAR, *Parallel computation of euler and navier-stokes flows*, in Proceedings of the Second Copper Mountain Conference on Multigrid Methods, S. McCormick, ed., 1986, pp. 321–331.

[29] C. THOLE, *A Short Note on Parallel Multigrid Algorithms for 3D-Problems*, Tech. Rep., GMD-Studien Report, St-Augustin, West Germany, 1987.

[30] D. TYLAVSKY, *Assessment of Inherent Parallelism in Explicit CFD Codes*, Tech. Rep., NASA
Ames Research Report, Moffett Field, Ca, 1987.