ON THE USE OF RELAXATION PARAMETERS IN HYBRID SMOOTHERS *

ULRIKE MEIER YANG †

Abstract. The use of relaxation parameters in hybrid smoothers within AMG is analyzed both theoretically and practically. Optimal relaxation parameters are determined. The implementation of a procedure to automatically determine an optimal outer relaxation parameter for symmetric positive definite smoothers is described. Numerical results are presented, which show significant improvements over AMG with undamped hybrid smoothers.

1. Introduction. With the advent of large high performance computers with large number of processors, it has become necessary to design parallel algorithms of all sorts. Particular emphasis has been placed on the development of scalable algorithms, such as multigrid methods. With this in mind, the parallelization of algebraic multigrid, a method that can be applied to a linear system, $Ax = b$, without additional knowledge, such as the underlying finite elements or a grid, has become very important. AMG proceeds by determining a subset of the original degrees of freedom through a coarsening algorithm, a restriction operator that transfers vectors from the fine space to the coarse space, and an interpolation operator that transfers vectors from the coarser space to the finer space. One important component of AMG is the smoother. A good smoother will reduce the oscillatory error components, whereas the "smooth" error is transferred to the coarser grids. Although the classical approach of AMG focused mainly on the Gauss-Seidel method [7], the use of other iterative solvers as smoothers has been considered [3, 2].

Gauss-Seidel has proven to be an effective smoother for many problems, however its main disadvantage is its sequential nature. On the other hand, highly parallel smoothers such as Jacobi or block-Jacobi often fail, unless an appropriate smoothing parameter is used and even then their convergence is often slow. Additionally, the user is faced with the challenge on how to choose an appropriate smoothing parameter. Many efforts to parallelize Gauss-Seidel have been made. Possible variants include the use of multi-coloring techniques [1] or hybrid schemes [6]. Multi-coloring techniques are a nuisance to implement and can be inefficient, if too many colors are involved (which is most likely to happen on the coarser levels of AMG). Hybrid schemes use an iterative method, e.g. Gauss-Seidel on each processor, but update in a Jacobi-like approach across boundaries. They are equivalent to block Jacobi methods that use one or more iterations of a smoother within each block instead of a direct solve. Clearly, this approach is very suitable for parallel processing, however, just like the block Jacobi method, it often requires a suitable smoothing parameter for convergence.

In this paper, we investigate the use of relaxation parameters in hybrid smoothers. There are two types of relaxation parameters: the smoothing parameter, mentioned above, which will be denoted the *outer* relaxation parameter, ω_J; and the so-called *inner* relaxation parameters, ω_i, which occur, if we smooth locally, on processor i, using SOR or its symmetric variant, SSOR. For both cases, the question is how to determine good parameters. Additionally, since one deals with a new system on each level of AMG, the development of an automatic procedure to determine such param-

*This work was performed under the auspices of the U.S. Department of Energy University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

† Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Box 808, L-560, Livermore, CA 94551

eters is important. Since the outer relaxation parameter affects the matrix across all processors, it appears that this parameter would be more crucial in improving convergence (or leading to convergence in cases of divergence). Therefore our main focus will be on the determination of an optimal ω_J. However, this paper also contains some results on the use of inner relaxation parameters.

In Section 2, we give some basic definitions. In Sections 3 through 5 we present conditions, under which the smoothing properties are fulfilled. Section 4 focuses specifically on the outer relaxation parameters and presents the determination of optimal parameters. Section 5 analyzes the use of inner relaxation parameters. Section 6 describes a procedure to determine outer relaxation parameters automatically, and in Section 7, we present numerical results that show that this approach can lead to significant improvements or even convergence in cases, for which AMG with an undamped hybrid smoother does not converge.

2. Definitions. Since our goal is to solve the linear system $Ax = b$ on a parallel computer with p processors, we partition the linear system as follows:

$$A = \begin{pmatrix} A_{11} & . & . & A_{1p} \\ . & . & . & . \\ . & . & . & . \\ A_{p1} & . & . & A_{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ . \\ . \\ x_p \end{pmatrix} = \begin{pmatrix} b_1 \\ . \\ . \\ b_p \end{pmatrix}. \tag{2.1}$$

A general definition of a smoother based on a splitting $Q - (Q - A)$ of A is

$$Qu_{n+1} = b + (Q - A)u_n \tag{2.2}$$

where Q can be any matrix. For example, with Gauss-Seidel $Q = D - L$, where D is the diagonal matrix with the diagonal of A and $-L$ the strict lower triangular part of A, while for Jacobi $Q = D$. Since we are interested in parallel smoothers, we will only consider splittings of the form

$$Q = \begin{pmatrix} Q_1 & & & \\ & Q_2 & & \\ & & . & \\ & & & Q_p \end{pmatrix}. \tag{2.3}$$

This is equivalent to performing any iterative solver such as Gauss-Seidel or block Gauss-Seidel, etc, locally on each processor, but updating the unknowns that are located on the neighbor processors only after each iteration step. Specific examples for hybrid smoothers are hybrid Gauss-Seidel with $Q_k = D_k - L_k$, or hybrid symmetric Gauss-Seidel with $Q_k = (D_k - L_k)D_k^{-1}(D_k - L_k^T)$.

As mentioned in the previous section, there are two types of relaxation parameters, the outer relaxation parameter ω_J, and the inner relaxation parameters $\omega_i, i = 1, ..., p$. Therefore consider the following smoothing matrix with outer relaxation parameter ω_J

$$Q = \frac{1}{\omega_J}\tilde{Q} = \frac{1}{\omega_J} \begin{pmatrix} \tilde{Q}_1 & & & \\ & . & & \\ & & . & \\ & & & \tilde{Q}_p \end{pmatrix}. \tag{2.4}$$

Inner relaxation parameters, $\omega_k, k = 1, ..., p$ occur in hybrid SOR with Q defined as above and

$$\tilde{Q}_k = \frac{1}{\omega_k} D_k - L_k \tag{2.5}$$

as well as the corresponding symmetric variant, hybrid SSOR, with

$$\tilde{Q}_k = \frac{\omega_k}{2 - \omega_k} (\frac{1}{\omega_k} D_k - L_k) D_k^{-1} (\frac{1}{\omega_k} D_k - L_k^T). \tag{2.6}$$

Further notations needed include the spectral radius of a matrix A, $\rho(A)$, and the smallest eigenvalue of a matrix A, $\lambda_{min}(A)$.

3. Smoothing Properties. Denote by A^m the operator on the m-th level of AMG, P^m the interpolation operator that interpolates from the $(m+1)$th level to the mth level, and R^m the restriction operator, that transfers from the mth to the $(m+1)$th level, which in general, if A^m is symmetric, is defined as $R^m = (P^m)^T$.

Then the coarse grid correction operator is given by

$$T^m = I - R^m (A^{m+1})^{-1} P^m A^m.$$

We define the smoothing operator on the mth level of AMG as

$$G^m = I - (Q^m)^{-1} A^m,$$

where Q^m is the matrix defined by the relaxation process $Q^m u_{n+1} = b^m + (Q - A) u_n$.

Important conditions for convergence of algebraic multigrid methods are the smoothing properties

$$\|G^m e^m\|_1^2 \leq \|e^m\|_1^2 - \sigma_1 \|e^m\|_2^2 \quad \sigma_1 > 0, \tag{3.1}$$

$$\|G^m e^m\|_1^2 \leq \|e^m\|_1^2 - \sigma_2 \|G^m e^m\|_2^2 \quad \sigma_2 > 0, \tag{3.2}$$

with the following norms

$$\|x\|_1 = (x^T A^m x)^{\frac{1}{2}}, \qquad \|x\|_2 = (x^T (A^m)^T (D^m)^{-1} A^m x)^{\frac{1}{2}}, \tag{3.3}$$

where D^m is the diagonal part of A^m. Here (3.1) refers to postsmoothing, (3.2) to presmoothing. In conjunction with the approximation property

$$\|T^m e^m\|_1^2 \leq \beta \|e^m\|_2^2, \tag{3.4}$$

either (3.1) or (3.2) imply two-level convergence with the convergence factor bounded above by $\sqrt{1 - \delta_1}$ for (3.1) and by $1/\sqrt{1 + \delta_2}$ for (3.2), with $\delta_1 = \sigma_1/\beta$ and $\delta_2 = \sigma_2/\beta$. This shows that larger σ_1 and σ_2 will lead to better smoothing and ultimately better convergence. For a more detailed discussion on convergence see [7, 8].

Obviously, the approximation property is determined by the choice of interpolation and restriction, and therefore of no concern for us in this context, in which we only focus on smoothing.

Now according to [7], (3.1) and (3.2) are equivalent to the following inequalities, which are somewhat easier to deal with (for simplicity, the indices m will be omitted in the remainder of the paper)

$$\sigma_1 e^T Q^T D^{-1} Q e \leq e^T (Q^T + Q - A) e, \tag{3.5}$$

$$\sigma_2 e^T (Q - A)^T D^{-1} (Q - A) e \leq e^T (Q^T + Q - A) e. \tag{3.6}$$

If A is a symmetric positive matrix, then the matrix $Q + Q^T - A$ is symmetric positive definite if and only if $A = Q-(Q-A)$ is a splitting that leads to a convergent iterative method, i.e. $\rho(I - Q^{-1}A) < 1$, see [9]. Using this fact it is easy show that both smoothing properties can always be fulfilled.

THEOREM 3.1. *Assume A and D are symmetric positive and Q is a matrix, for which $\rho(I - Q^{-1}A) < 1$.*

Then both smoothing properties (3.5) and (3.6) are fulfilled with

$$\sigma_1 \leq \frac{\lambda_{min}(Q + Q^T - A)}{\rho(Q^T D^{-1} Q)} \tag{3.7}$$

and

$$\sigma_2 \leq \frac{\lambda_{min}(Q + Q^T - A)}{\rho((Q - A)^T D^{-1}(Q - A))}. \tag{3.8}$$

Proof:

Let us assume for simplicity that $e^T e = 1$. Since $Q+Q^T-A$ is symmetric positive definite, all its eigenvalues are real and positive and therefore

$$e^T(Q + Q^T - A)e \geq \lambda_{min}(Q + Q^T - A).$$

The matrices $Q^T D^{-1} Q$ and $(Q-A)^T D^{-1}(Q-A)$ are symmetric positive semidefinite and thus have real nonnegative eigenvalues. The following inequalities hold

$$\begin{aligned} e^T Q^T D^{-1} Q e &\leq \rho(Q^T D^{-1} Q), \\ e^T (Q - A)^T D^{-1}(Q - A)e &\leq \rho((Q - A)^T D^{-1}(Q - A)), \end{aligned}$$

and one obtains inequalities (3.7) and (3.8) for σ_1 and σ_2.

q.e.d.

Thus, if the underlying iterative scheme is convergent, the smoothing properties are fulfilled. Further, it is easy to come up with a sample vector e, for which the properties are not fulfilled, if the method does not converge. The theorem also explains, why hybrid smoothers fail: for many problems, as can be seen in Section 7, $\rho(I - Q^{-1}A) > 1$, or, equivalently, $\rho(Q^{-1}A) > 2$.

However, the estimated bounds for σ_1 and σ_2 are not very useful if one wants to determine good relaxation parameters. Therefore, we investigate symmetric positive definite Q and derive more meaningful bounds for σ_1 and σ_2.

THEOREM 3.2. *Assume that A, Q and D are symmetric positive definite and*

$$\rho(Q^{-1}A) < 2.$$

Then the smoothing property

$$\sigma_1 e^T Q^T D^{-1} Q e \leq e^T (Q + Q^T - A)e$$

is fulfilled with

$$\sigma_1 \leq \frac{2 - \rho(Q^{-1}A)}{\rho(D^{-1}Q)}. \tag{3.9}$$

Proof:

$$\begin{aligned}\sigma_1 e^T Q^T D^{-1} Q e + e^T A e &\leq \sigma_1 \rho(D^{-1}Q) e^T Q^T e + \rho(Q^{-1}A) e^T Q e \\ &= (\sigma_1 \rho(D^{-1}Q) + \rho(Q^{-1}A)) e^T Q e \\ &\leq 2 e^T Q e \\ &= e^T (Q^T + Q) e.\end{aligned}$$

q.e.d.

A similar results is obtained for the second smoothing property (3.6).

THEOREM 3.3. *Assume that A, Q and D are symmetric positive definite and*

$$\rho(Q^{-1}A) < 2.$$

Then the smoothing property

$$\sigma_2 e^T (A-Q)^T D^{-1} (A-Q) e \leq e^T (Q + Q^T - A) e$$

is fulfilled with

$$\sigma_2 \leq \frac{2 - \rho(Q^{-1}A)}{\rho(D^{-1}Q)[\rho(I - Q^{-1}A)]^2}. \tag{3.10}$$

Proof: First let us consider $\rho((I - Q^{-1}A)D^{-1}(Q - A))$. Also, let $\|.\|_S$ denote the spectral norm or 2-norm.

$$\begin{aligned}\rho((I - Q^{-1}A)D^{-1}(Q-A)) &= \rho((I - Q^{-1}A)D^{-1}Q(I - Q^{-1}A)) \\ &= \rho(Q^{\frac{1}{2}}(I - Q^{-1}A)Q^{-\frac{1}{2}} Q^{\frac{1}{2}} D^{-1} Q (I - Q^{-1}A) Q^{-\frac{1}{2}}) \\ &= \rho((I - Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}}) Q^{\frac{1}{2}} D^{-1} Q^{\frac{1}{2}} (I - Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}})) \\ &= \|D^{-\frac{1}{2}} Q^{\frac{1}{2}} (I - Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}})\|_S^2 \\ &\leq \|D^{-\frac{1}{2}} Q^{\frac{1}{2}}\|_S^2 \|I - Q^{-\frac{1}{2}} A Q^{-\frac{1}{2}}\|_S^2 \\ &= \rho(D^{-1}Q)[\rho(I - Q^{-1}A)]^2\end{aligned}$$

Using this result, we get

$$\begin{aligned}\sigma_2 e^T (A-Q)^T D^{-1} (A-Q) e + e^T A e &\leq \sigma_2 \rho(Q^{-T}(A-Q)^T D^{-1}(A-Q)) e^T Q^T e + \rho(Q^{-1}A) e^T Q e \\ &\leq \sigma_2 \rho(D^{-1}Q)[\rho(I - Q^{-1}A)]^2 e^T Q e + \rho(Q^{-1}A)) e^T Q e \\ &\leq 2 e^T Q e \\ &= e^T (Q^T + Q) e.\end{aligned}$$

q.e.d.

4. Determination of an Optimal Outer Relaxation Parameter. In this section, an optimal outer relaxation parameter is determined, using the results of the previous section.

Using (2.4), (3.9) can be expressed as

$$\sigma_1 \leq \frac{\omega_J (2 - \omega_J \rho(\tilde{Q}^{-1}A))}{\max\limits_{1 \leq k \leq p} \rho(D_k^{-1} \tilde{Q}_k)},$$

and (3.10) as

$$\sigma_2 \leq \frac{\omega_J(2-\omega_J\rho(\tilde{Q}^{-1}A))}{\max\limits_{1\leq k\leq p}\rho(D_k^{-1}\tilde{Q}_k)[\rho(I-\omega_J\tilde{Q}^{-1}A)]^2}.$$

Since σ_1 and σ_2 are positive, one obtains the following condition for ω_J:

$$0<\omega_J<\frac{2}{\rho(\tilde{Q}^{-1}A)}.$$

Note that in the case of inner relaxation parameters ω_k, for $1\leq k\leq p$, the value of ω_J depends on the choice of these parameters.

In order to maximize σ_1 in (3.5), one must choose

$$\omega_J=\frac{1}{\rho(\tilde{Q}^{-1}A)}, \tag{4.1}$$

which leads to

$$\sigma_1\leq\frac{1}{\rho(D^{-1}\tilde{Q})\rho(\tilde{Q}^{-1}A)}.$$

Consider now the second smoothing inequality (3.6). Note that since $\omega_J>0$ and both Q and A are symmetric positive definite, so are $\tilde{Q}$ and $\tilde{Q}^{-\frac{1}{2}}A\tilde{Q}^{-\frac{1}{2}}$. Therefore

$$\begin{aligned}\rho(I-\omega_J\tilde{Q}^{-1}A)&=\rho(I-\omega_J\tilde{Q}^{-\frac{1}{2}}A\tilde{Q}^{-\frac{1}{2}})\\&=\begin{cases}1-\omega_J\lambda_{min}(\tilde{Q}^{-1}A) & \text{for}\quad 0<\omega<\dfrac{2}{\rho(\tilde{Q}^{-1}A)+\lambda_{min}(\tilde{Q}^{-1}A)}\\ |1-\omega_J\rho(\tilde{Q}^{-1}A)| & \text{for}\quad \omega\geq\dfrac{2}{\rho(\tilde{Q}^{-1}A)+\lambda_{min}(\tilde{Q}^{-1}A)}.\end{cases}\end{aligned}$$

The optimal ω_J can now be determined by maximizing

$$\psi_1(\omega)=\frac{\omega(2-\omega\rho(\tilde{Q}^{-1}A))}{(1-\omega\lambda_{min}(\tilde{Q}^{-1}A))^2}\qquad\text{or}\qquad\psi_2(\omega)=\frac{\omega(2-\omega\rho(\tilde{\mathrm{Q}}^{-1}\mathrm{A}))}{(1-\omega\rho(\tilde{\mathrm{Q}}^{-1}\mathrm{A}))^2}$$

with respect to ω.

If $\rho(\tilde{Q}^{-1}A)>2\lambda_{min}(\tilde{Q}^{-1}A)$, which is usually the case for finite element problems, the solution to maximizing $\psi_1(\omega)$ is

$$\omega_J=\frac{1}{\rho(\tilde{Q}^{-1}A)-\lambda_{min}(\tilde{Q}^{-1}A)},$$

which leads to

$$\sigma_2\leq\frac{1}{\rho(D^{-1}\tilde{Q})(\rho(\tilde{Q}^{-1}A)-2\lambda_{min}(\tilde{Q}^{-1}A))}.$$

Since in general for finite element problems $\lambda_{min}(\tilde{Q}^{-1}A)$ is small compared to the largest eigenvalue, the best choice for the first smoothing property (4.1) is also a good choice for the second smoothing property with

$$\sigma_2\leq\frac{\rho(\tilde{Q}^{-1}A)}{\rho(D^{-1}\tilde{Q})(\rho(\tilde{Q}^{-1}A)-\lambda_{min}(\tilde{Q}^{-1}A))^2}.$$

$\psi_2(\omega)$ is decreasing in the considered range, consequently it is maximal for

$$\omega = \frac{2}{\lambda_{min}(\tilde{Q}^{-1}A) + \rho(\tilde{Q}^{-1}A)}. \tag{4.2}$$

with

$$\sigma_2 \leq \frac{4\lambda_{min}(\tilde{Q}^{-1}A)}{\rho(D^{-1}\tilde{Q})(\rho(\tilde{Q}^{-1}A) - \lambda_{min}(\tilde{Q}^{-1}A))^2}.$$

Clearly if

$$4\lambda_{min}(\tilde{Q}^{-1}A) < \rho(\tilde{Q}^{-1}A)$$

(4.1) is a better choice for ω_J than (4.2). Note that (4.2) minimizes $\rho(I - \omega\tilde{Q}^{-1}A)$ and leads therefore to the fastest convergent splitting, but does not lead to the best smoother!

5. Analysis of the Inner Relaxation Parameters. Inner relaxation parameters occur when one uses SOR or SSOR locally on each processor. Determining a best choice for the inner relaxation parameter in SSOR is a very difficult task. Even in the case of only one processor, one needs to analyze the following complicated functions

$$\phi_1(\omega) = \frac{\omega(2-\omega)(2-\omega(2-\omega)\rho(\tilde{Q}(\omega)^{-1}A))}{\rho(D^{-1}\tilde{Q}(\omega))},$$
$$\phi_2(\omega) = \frac{\omega(2-\omega)(2-\omega(2-\omega)\rho(\tilde{Q}(\omega)^{-1}A))}{\rho(D^{-1}\tilde{Q}(\omega))(\rho(I-\omega(2-\omega)\tilde{Q}(\omega)^{-1}A))^2}.$$

Although there are some interesting results on the choice of relaxation parameters for SSOR as an iterative solver in [9], these results do not transfer to smoothers, and applying the same techniques to analyze the parameter gives nonconclusive results. However, one can derive some interesting results for SOR and hybrid SOR.

LEMMA 5.1. *Assume A is symmetric positive definite, $Q = \frac{1}{\omega}D - L$, where D is the diagonal and $-L$ the lower triangular part of A, and $0 < \omega < 2$.*

Then

$$\sigma_1 e^T Q^T D^{-1} Q e \leq e^T (Q^T + Q - A) e.$$

with

$$\sigma_1 \leq \frac{(2-\omega)\omega}{(1+\omega\gamma^-)(1+\omega\gamma^+)},$$

where

$$\gamma^- = \|D^{-1}L\|,$$
$$\gamma^+ = \|D^{-1}L^T\|,$$

and $\|.\|$ denotes any vector induced matrix norm.

In the above estimate, the upper bound is maximal, if

$$\omega = \frac{\sqrt{(2\gamma^+ + 1)(2\gamma^- + 1)} - 1}{\gamma^+ + \gamma^- + 2\gamma^-\gamma^+} \leq 1.$$

Proof:
Since $Q = \frac{1}{\omega}D - L$,

$$Q + Q^T - A = \frac{2}{\omega}D - L - L^T - D + L + L^T = (\frac{2}{\omega} - 1)D$$

Now

$$e^T Q^T D^{-1} Q e \leq \rho(D^{-1}Q^T D^{-1}Q) e^T D e$$

and

$$\begin{aligned}\rho(D^{-1}Q^T D^{-1}Q) &\leq \|D^{-1}Q^T\| \, \|D^{-1}Q\| \\ &= \|\frac{1}{\omega}I - D^{-1}L^T\| \|\frac{1}{\omega}I - D^{-1}L\| \\ &\leq \frac{1}{\omega^2}(1 + \omega\gamma^+)(1 + \omega\gamma^-).\end{aligned}$$

This leads to the following inequality for σ_1

$$\sigma_1 \leq \phi_1(\omega) = \frac{(2 - \omega)\omega}{(1 + \omega\gamma^-)(1 + \omega\gamma^+)}.$$

Now,

$$\frac{d\phi_1(\omega)}{d\omega} = \frac{2 - 2\omega - (\gamma^- + \gamma^+ + 2\gamma^-\gamma^+)\omega^2}{(1 + \omega\gamma^-)^2(1 + \omega\gamma^+)^2},$$

which vanishes for

$$\omega_1 = \frac{-\sqrt{(2\gamma^+ + 1)(2\gamma^- + 1)} - 1}{\gamma^+ + \gamma^- + 2\gamma^-\gamma^+} < 0$$

and

$$\omega_2 = \frac{\sqrt{(2\gamma^+ + 1)(2\gamma^- + 1)} - 1}{\gamma^+ + \gamma^- + 2\gamma^-\gamma^+}.$$

Since

$$\frac{d^2\phi_1(\omega_2)}{d\omega^2} = -\frac{\sqrt{(2\gamma^+ + 1)(2\gamma^- + 1)}}{(1 + \omega_2\gamma^-)^2(1 + \omega_2\gamma^+)^2} < 0,$$

and the relative minimum ω_1 is outside of (0,1), ω_2 is the maximum of $\phi_1(\omega)$ in (0,2).
Since

$$\gamma^+ + \gamma^- + 2\gamma^+\gamma^- = \frac{1}{2}((2\gamma^+ + 1)(2\gamma^- + 1) - 1)$$

it is easy to show that $\omega_2 \leq 1$.

q.e.d.

This result is interesting, since it shows that the best ω in this context does not lead to overrelaxation, as is the case when SOR is used as an iterative solver, but to underrelaxation. In the special case $\gamma^+ = \gamma^- = \gamma$, one obtains $\omega = \frac{1}{1+\gamma} < 1$.

Note that Ruge and Stüben [7] suggest the norm

$$\|A\|_v = \max_{1 \le i \le n} \left\{ \frac{1}{v_i} \sum_{j=1}^{n} v_j |a_{ij}| \right\}$$

where v is a vector with positive elements v_i. This choice leads to

$$\gamma^- = \max_k \left\{ \frac{1}{v_k a_{kk}} \sum_{j<k} v_k |a_{kj}| \right\},$$
$$\gamma^+ = \max_k \left\{ \frac{1}{v_k a_{kk}} \sum_{j>k} v_k |a_{kj}| \right\}.$$

Using a similar argument, one can show that the second smoothing property (3.6) is also fulfilled for SOR.

LEMMA 5.2. *Assume A is symmetric positive definite, $Q = \frac{1}{\omega} D - L$, where D is the diagonal and $-L$ the lower triangular part of A, and $0 < \omega < 2$.*

Then

$$\sigma_2 e^T (Q - A)^T D^{-1} (Q - A) e \le e^T (Q + Q^T - A) e$$

with

$$\sigma_2 \le \frac{(2-\omega)\omega}{(|1-\omega| + \omega\gamma^-)(|1-\omega| + \omega\gamma^+)}.$$

The upper bound is maximal in most cases if $\omega = 1$. For

$$\gamma^- > \frac{1}{2}, \qquad \gamma^+ > \frac{\gamma^-}{2\gamma^- - 1} \tag{5.1}$$

the optimal ω is given by

$$\omega = \frac{\sqrt{(2\gamma^+ - 1)(2\gamma^- - 1)} - 1}{2\gamma^+\gamma^- - \gamma^- - \gamma^+} \le 1. \tag{5.2}$$

The proof is similar to the proof of Lemma 5.1. The inequality for σ_2 is determined in a similar way as was done for σ_1 in Lemma 5.1. In order to determine the best ω, one needs to examine

$$\phi_2(\omega) = \begin{cases} \dfrac{(2-\omega)\omega}{(1+\omega(\gamma^- - 1))(1+\omega(\gamma^+ - 1))} & \text{for} \quad 0 < \omega \le 1 \\ \dfrac{(2-\omega)\omega}{(\omega(\gamma^- + 1) - 1)(\omega(\gamma^+ + 1) - 1)} & \text{for} \quad 1 \le \omega < 2. \end{cases}$$

In most cases it turns out that this function is increasing in (0,1) and decreasing in (1,2), which shows that the optimal ω is 1. Only in the special case $2\gamma^+\gamma^- >$

$\gamma^+ + \gamma^-$, is the absolute maximum in (0,1) and is as given by (5.2) by straightforward differentiation of $\phi_2(\omega)$.

In this case it appears that the use of a relaxation parameter is, in general, not beneficial. Only if conditions (5.1) are fulfilled, which imply a matrix that is not diagonally dominant, should underrelaxation lead to better convergence.

In the following theorem, the SOR hybrid method is investigated.

THEOREM 5.3.

Assume that A is symmetric positive definite, D_B the block diagonal matrix with diagonal blocks A_{kk}, $Q = \frac{1}{\omega_J}\tilde{Q}$ as defined in (2.4), $\tilde{Q}_k = \frac{1}{\omega_k}D_k - L_k$ with $0 < \omega_k < 2$ for $k = 1, ..., p$ and $0 < \omega_J \leq \frac{1}{\rho(D_B^{-1}A)}$.

Then the smoothing property

$$\sigma_1 e^T Q^T D^{-1} Q e \leq e^T (Q + Q^T - A) e$$

is fulfilled for

$$\sigma_1 \leq \omega_J \min_{1 \leq k \leq p} \frac{(2 - \omega_k)\omega_k}{(1 + \omega_k \gamma_k^+)(1 + \omega_k \gamma_k^-)}$$

with

$$\gamma_k^+ = \|D_k^{-1} L_k^T\|,$$
$$\gamma_k^- = \|D_k^{-1} L_k\|,$$

where $\|.\|$ denotes any matrix norm, induced by a vector norm.

Proof:

Lemma 5.1 leads to the following inequality for σ_i

$$\sigma_i \leq \frac{(2 - \omega_i)\omega_i}{(1 + \omega_i \gamma_i^-)(1 + \omega_i \gamma_i^+)}.$$

Now

$$e^T A e \leq \rho(D_B^{-1}A) e^T D_B e \leq \frac{1}{\omega_J} e^T D_B e$$

using the assumption $\omega_J \leq \frac{1}{\rho(D_B^{-1}A)}$. We thus obtain

$$\begin{aligned}
e^T(Q^T + Q - A)e &= \frac{1}{\omega_J}\sum_{i=1}^{p} e_i^T (\frac{2}{\omega_i} D_i - L_i - L_i^T) e_i - e^T A e \\
&= \frac{1}{\omega_J}\sum_{i=1}^{p} (\frac{2}{\omega_i} - 1) e_i^T D_i e_i + \frac{1}{\omega_J} e^T D_B e - e^T A e \\
&\geq \frac{1}{\omega_J}\sum_{i=1}^{p} e_i^T (\tilde{Q}_i^T + \tilde{Q}_i - A_{ii}) e_i + (\frac{1}{\omega_J} - \rho(D_B^{-1}A)) e^T D_B e \\
&\geq \frac{1}{\omega_J}\sum_{i=1}^{p} \frac{(2 - \omega_i)\omega_i}{(1 + \omega_i \gamma_i^-)(1 + \omega_i \gamma_i^+)} e_i^T \tilde{Q}_i^T D_i^{-1} \tilde{Q}_i e_i \\
&\geq \omega_J \min_{1 \leq i \leq p} \frac{(2 - \omega_i)\omega_i}{(1 + \omega_i \gamma_i^-)(1 + \omega_i \gamma_i^+)} e^T Q^T D^{-1} Q e \\
&\geq \sigma_1 e^T Q^T D^{-1} Q e. \qquad \textit{q.e.d.}
\end{aligned}$$

This shows that a good choice of inner relaxation parameters should improve the hybrid SOR method, but even more crucial is a good choice of ω_J for (3.5) to be fulfilled. The best choice according to this theorem, is ω_J, $1/\rho(D_B^{-1}A)$. It is possible to obtain a good estimate of $\rho(D_B^{-1}A)$ using the procedure described in the next section. However, this approach would be very expensive and is therefore not practical, since it requires solving a linear system on each processor in each CG iteration step.

6. Practical Determination of the Outer Relaxation Parameter. The result on the optimal outer relaxation parameter obtained in Section 4 is only useful if it can be applied in practice. It is very important to get good estimates for $\rho(\tilde{Q}^{-1}A)$. This can be achieved by applying k steps of preconditioned conjugate gradient to $Ax = b$ with the preconditioner $\tilde{Q}$. Note that $\tilde{Q}$ needs to be symmetric positive definite. The preconditioning step is here just the application of one sweep of the smoother, which is fairly inexpensive. Due to the relationship of the conjugate gradient method and the Lanczos method [5], one can derive the tridiagonal Lanczos matrix T_k from the parameters obtained within CG as follows [5].

In the mth iteration of CG one computes the parameters

$$\alpha_m = \frac{r_{k-1}^T \tilde{Q}^{-1} r_{k-1}}{p_k^T A p_k}, \quad \beta_m = \frac{r_{k-1}^T \tilde{Q}^{-1} r_{k-1}}{r_{k-2}^T \tilde{Q}^{-1} r_{k-2}}. \tag{6.1}$$

with the residual r_m and the direction vector p_m, which in turn give

$$T_k = \begin{pmatrix} d_1 & c_1 & & & \\ c_1 & d_2 & . & & \\ & . & . & . & \\ & & . & . & c_{k-1} \\ & & & c_{k-1} & d_k \end{pmatrix} \tag{6.2}$$

where

$$d_m = \frac{\beta_m}{\alpha_{m-1}} + \frac{1}{\alpha_m}. \quad c_m = -\frac{\sqrt{\beta_{m+1}}}{\alpha_m} \tag{6.3}$$

Since the eigenvalues of T_k approach the eigenvalues of $\tilde{Q}^{-1}A$ with increasing k, one can estimate the eigenvalues of T_k. This can be done using the Gershgorin estimate $\max_{1\le m\le k}\{|d_m|+|c_m|+|c_{m-1}|\}$, or, since T_k is very small, using an eigenvalue solver for tridiagonal systems (such as the QR algorithm or bisection [5]). It is possible to get good estimates with this procedure with a fairly small number of CG iterations, e.g. $k = 10$ or $k = 15$. The use of this procedure increases the setup time of AMG. However, for problems that require a good smoothing parameter, the resulting decrease in number of iterations and solve time far outweighs this increase in setup time, as can be seen in the next section.

7. Numerical Results. The methods described in the previous sections are applied to various very large 3-dimensional elasticity problems composed of 3 concentric spherical shells. An octant of this domain is shown in Figure 7.1. The outer shells are composed of steel, the inner shell is composed of lucite. We consider problems without and with slide surface boundary conditions. In the case of slide surface boundary conditions the steel and lucite spheres are allowed to slide tangentially relative to each other. Adding the slide surface boundary conditions leads to an indefinite problem. It is, however, possible to reduce the system to a positive definite system through

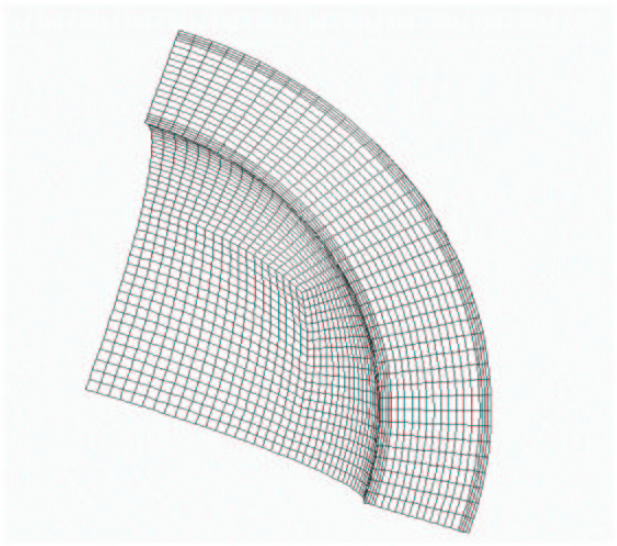

FIG. 7.1. *Finite element discretization of a sphere using quadrilateral elements.*

the elimination of a subset of equations [4]. In our experiments we use the reduced system. The problems were run on the ASCI White Computer at LLNL. Two different problem sizes are considered: the smaller problem is 497,664 elements and is run on 32 processors; the larger problem consists of almost 4 million elements and is run using 256 processors. Since the considered problem has multiple degrees of freedom per grid point, we use AMG for systems problems, employing the function, or "unknown", approach [7]. This approach coarsens each physical variable separately and interpolates only within variables of the same type. The smoothers are symmetric Gauss-Seidel with $\omega_J = 1$ and the optimal ω_J in (4.1), which is obtained using at most 10 CG-iterations. We are able to use a nodal hybrid Gauss-Seidel, i.e. a block Gauss-Seidel method with 3x3 blocks, due to the structure of the problem. Unfortunately, the nodal structure is destroyed after the first level, so use of a nodal smoother beyond the finest level does not make sense. The optimal ω_J was here estimated with 15 CG-iterations, since 10 CG-iterations turned out to be not good enough. We also consider hybrid SSOR. Since we have no procedure to determine the best inner relaxation parameters, we present results for uniform $\omega = 0.75$, 0.5 and 0.25 for the moderate size problem and use $\omega = 0.5$ for the large problem. In all these experiments, AMG is used as a preconditioner for CG. Therefore, in order to not destroy the symmetry of the problem, only symmetric smoothers are used.

Table 7.1 gives the estimates of the optimal outer relaxation weights that have been used for the larger elasticity problems. The fact that the relaxation parameters on the finer levels are smaller than 0.5, and thus $\rho(\tilde{Q}^{-1}A) > 2$, indicates that hybrid (block) Gauss-Seidel is not a convergent iterative scheme for these problems. Table 7.2 contains the notations used for Tables 7.3 through 7.6. $\tilde{Q}$ denotes the symmetric Gauss-Seidel matrix, and $\tilde{Q}_B$ denotes the nodal or 3x3 block Gauss-Seidel matrix.

For moderate size problems, Table 7.3 shows a fairly small improvement of scaled smoothers over unscaled smoothers. However, for large problems, Table 7.4 shows

without slide surfaces				with slide surfaces			
level	ω_J	level	ω_J	level	ω_J	level	ω_J
1	0.425	12	0.588	1	0.465	12	0.588
2	0.458	13	0.585	2	0.462	13	0.584
3	0.480	14	0.591	3	0.453	14	0.605
4	0.460	15	0.614	4	0.453	15	0.612
5	0.498	16	0.604	5	0.502	16	0.627
6	0.555	17	0.603	6	0.556	17	0.639
7	0.585	18	0.601	7	0.580	18	0.613
8	0.567	19	0.587	8	0.588	19	0.620
9	0.585			9	0.583		
10	0.570	nodal:		10	0.585	nodal:	
11	0.572	1	0.397	11	0.575	1	0.419

TABLE 7.1
Relaxation parameters for an elasticity problem, 3,981,312 elements, 256 procs

a significant improvement for scaled smoothers. Overall, the best time achieved is about 10 times as fast as the original test (which uses hybrid Gauss-Seidel without any relaxation parameter). Interestingly enough, it turns out that for this problem, the use of the nodal smoother does not improve convergence; apparently point smoothers are sufficient to smooth the error.

Table 7.5 shows this changes when we include slide surfaces. On the moderate size problem, using hybrid SGS without any smoothing parameter converges in 390 iterations, whereas scaled SGS converges about 3 times as fast. It is interesting that just applying the nodal smoother without any scaling parameter entails a similar number of iterations, showing that for this problem a nodal smoother is more suitable. Scaling the nodal smoother leads to a further improvement of another factor of about 2.3. Scaling the point smoother improves this result only slightly. Reducing the number of sweeps increases the number of iterations, but decreases the time per iteration. The best result is about 7 times faster than the original solver.

The larger problem with slide surfaces diverges without any relaxation parameters, even when a nodal smoother is used on the finest level (Table 7.6). Scaled SGS converges within 192 iterations, but convergence is twice as fast when a scaled nodal smoother is used on the finest level. The overall fastest (with regard to time) combination solves this very large problem, which diverges when unscaled smoothers are employed, in about 5 minutes.

The results in Tables 7.3 through 7.6 show that a good inner relaxation parameter for hybrid SSOR is 0.5. Overall, underrelaxation, i.e. choosing an inner relaxation parameter smaller than 1, beats hybrid SGS, which is equivalent to hybrid SSOR $omega = 1$. Overrelaxation ($\omega > 1$), which is not presented here, leads to a further decrease in performance.

8. Conclusions. The use of relaxation parameters for hybrid smoothers is analyzed. Both outer as well as inner parameters are considered. Analysis of the inner SOR relaxation parameter show that, in most cases, underrelaxation (i.e. $\omega < 1$) is preferred to overrelaxation. The optimal outer relaxation parameter for symmetric positive definite matrices and splittings is determined, and an automatic procedure to determine it is implemented. Numerical experiments show that for certain elasticity problems significant improvements can be achieved using relaxation parameters.

Smoother	Description
G	hybrid symmetric Gauss-Seidel (SGS)
S, ω	hybrid SSOR, $\omega_k = \omega$ for $k = 1, ..., p$
SG	scaled hybrid SGS, $\omega_J = \frac{1}{\tilde{\rho}(\bar{Q}^{-1}A)}$
N	hybrid nodal (3x3 blocks) SGS
SN	scaled hybrid nodal SGS, $\omega_J = \frac{1}{\tilde{\rho}(\bar{Q}_B^{-1}A)}$
<S1>/<S2>	<S1> used only on finest level, <S2> used on coarser levels

TABLE 7.2
Smoother notations

Smoother	no. of sweeps	no. of its	setup time	solve time	total time
G	2	74	15	266	281
S, 0.75	2	63	15	233	248
S, 0.5	2	45	15	164	179
S, 0.25	2	54	15	197	212
SG	2	42	25	157	182
N/G	2	59	16	223	240
SN/G	2	39	22	148	170
SN/G	1	47	22	108	130
SN/SG	2	41	28	161	189
SN/SG	1	50	28	119	148

TABLE 7.3
Elasticity problem without slide surfaces, 497,664 elements, 32 procs, n = 1,545,483

Acknowledgments. The author would like to thank Panayot Vassilevski, Robert Falgout and Van Emden Henson for their helpful comments.

REFERENCES

[1] M. Adams, *A Distributed Memory Unstructured Gauss-Seidel Algorithm for Multigrid Smoothers*, in ACM/IEEE Proceedings of SC01: High Performance Networking and Computing, 2001.

[2] M. Adams, M. Brezina, J. Hu, R. Tuminaro, *Parallel Multigrid Smoothing: Polynomial versus Gauss-Seidel* To appear in JCP.

[3] A. Brandt, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput. 19 (1986), 23–56.

[4] E. Chow, T. Manteuffel, C. Tong, B. Wallin, *Algebraic Elimination of Slide Surface Constraints in Implicit Structural Analysis*, submitted to Intern. J. Num.. Meth. Engrg. Also available as Lawrence Livermore Mational Laboratory technical report UCRL-JC-144528, July 2001.

[5] G.H. Golub, C.F. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, 1989).

[6] V.E. Henson, U.M. Yang, *BoomerAMG: a parallel algebraic multigrid solver and preconditioner.* Applied Numerical Mathematics 41 (2002), 155–177. Also available as LLNL technical report UCRL-JC-133948 (2000)

[7] J. W. Ruge and K. Stüben, *Algebraic multigrid (AMG)*, in : S. F. McCormick, ed., Multigrid Methods, vol. 3 of Frontiers in Applied Mathematics (SIAM, Philadelphia, 1987) 73–130.

[8] K. Stüben, *Algebraic multigrid (AMG): an introduction with applications*, in : U. Trottenberg, C. Osterlee and A. Schüller, eds., Multigrid (Academic Press, 2000).

[9] D. M. Young, *Iterative Solution of Large Linear Systems* (Academic Press, 1971).

Smoother	no. of sweeps	no. of its	setup time	solve time	total time
G	2	484	25	2097	2122
S, 0.5	2	67	25	301	326
SG	2	61	38	270	308
N/G	2	562	25	2477	2502
SN/G	2	57	32	258	290
SN/G	1	67	32	184	216
SN/SG	2	60	41	276	317
SN/SG	1	71	41	196	237

TABLE 7.4
Elasticity problem without slide surfaces, 3,981,312 elements, 256 procs, n = 12,152,595

Smoother	no. of sweeps	no. of its	setup time	solve time	total time
G	2	390	15	1427	1443
S, 0.75	2	201	15	750	765
S, 0.5	2	149	15	555	570
S, 0.25	2	185	15	696	711
SG	2	135	26	517	543
N/G	2	142	17	547	564
SN/G	2	63	23	245	268
SN/SG	2	60	31	237	267
SN/SG	1	74	31	181	212

TABLE 7.5
Elasticity problem with slide surfaces, 497,664 elements, 32 procs, n = 1,587,825

Smoother	no. of sweeps	no. of its	setup time	solve time	total time
G	2	fail	23	-	-
S, 0.5	2	342	23	1375	1399
SG	2	192	37	767	804
N/G	2	fail	27	-	-
SN/G	2	146	29	600	630
SN/SG	2	86	41	365	406
SN/SG	1	106	41	272	313

TABLE 7.6
Elasticity problem with slide surfaces, 3,981,312 elements, 256 procs, n = 12,320,217