Projection techniques for iterative solution of $A\underline{x} = \underline{b}$ with successive right-hand sides

Paul F. Fischer[1]

Abstract

Projection techniques are developed for computing approximate solutions to linear systems of the form $A\underline{x}^n = \underline{b}^n$, for a sequence $n = 1, 2, ...$, e.g., arising from time discretization of a partial differential equation. The approximate solutions are based upon previous solutions, and can be used as initial guesses for iterative solution of the system, resulting in significantly reduced computational expense. Examples of two- and three-dimensional incompressible Navier-Stokes calculations are presented in which $\underline{x}^n$ represents the pressure at time level t^n, and A is a consistent discrete Poisson operator. In flows containing significant dynamic activity, these projection techniques lead to as much as a two-fold reduction in solution time.

1 Introduction

We consider iterative solution of a sequence of linear problems having the form:

$$\mathcal{P}_n : \qquad A\underline{x}^n = \underline{b}^n \quad , \qquad n = \{1, 2, \ldots\} \tag{1}$$

where A is an $m \times m$ matrix, and $\underline{x}^n$ is assumed to be a solution which is evolving with some parameter, e.g., time, in the case where (1) represents an implicit substep in numerical solution of a time dependent partial differential equation. When A is sufficiently sparse and not amenable to eigenfunction decomposition techniques (e.g., fast Poisson solvers), iterative methods are generally preferable to direct factorizations, both from the standpoint of storage and operation count. For the particular class of problems defined by (1), direct methods benefit from amortization of the one-time cost of matrix-factorization. However, they derive no benefit from the fact that successive problems might have very similar solutions, whereas iterative methods can exploit this possibility as a good initial guess can lead to a significant reduction in the number of iterations required to bring the residual to within the specified tolerance.

In this paper we present simple techniques for extracting information from previous problems, $\mathcal{P}_k$, $n - l \leq k \leq n - 1$, to generate initial guesses to the current problem, $\mathcal{P}_n$. The first approach is to simply remove any component of $\underline{b}^n$ for which the solution is already known, by projecting $\underline{b}^n$ onto the set of vectors $\{\underline{b}^{n-l}, \ldots, \underline{b}^{n-1}\}$ having associated solutions $\{\underline{x}^{n-l}, \ldots, \underline{x}^{n-1}\}$, and to solve the problem corresponding to the component of $\underline{b}^n$ orthogonal to $span\{\underline{b}^{n-l}, \ldots, \underline{b}^{n-1}\}$. The second approach is a refinement of the first, which seeks the best approximation to $\underline{x}^n$ in $span\{\underline{x}^{n-l}, \ldots, \underline{x}^{n-1}\}$ with respect to a norm tailored to the convergence properties of the conjugate gradient method for the case when A is symmetric positive definite.

[1]Division of Applied Mathematics, Brown University, Providence, RI 02912, U.S.A.

The idea of using information generated from previous right-hand sides to speed iterative solution processes is not new. It is of course standard to solve only for the *change* in the solution, $\Delta \underline{x}^n \equiv \underline{x}^n - \underline{x}^{n-1}$ (e.g., [9]). It is possible to improve upon this basic approach through higher-order extrapolation in time. However, projection techniques are superior to those based upon extrapolation in that they yield the best possible approximation within a given basis. Moreover, while extrapolation techniques run the risk of generating a poor initial guess, this is not possible with the methods proposed here, as the projection is guaranteed to reduce the error in a relevant norm provided that $\underline{x}^n$ has some component in $span\{\underline{x}^{n-l}, \ldots, \underline{x}^{n-1}\}$. In the case where $\underline{x}^n$ is not well represented in this space, the projection will be void and the residual unchanged.

If Krylov based iterative methods are used to solve (1), it is also possible to generate an initial guess by projecting onto the resultant (orthogonal) Krylov bases, as suggested by many authors, dating back to Lanczos [10]. Widlund and O'Leary [15] proposed this idea in the context of three-dimensional Helmholtz solvers. Simon [21] and Saad [20] considered this approach in the context of Lanczos methods. Saad showed in particular that, in exact arithmetic, it is possible to generate a sequence of A-conjugate bases without resorting to full orthogonalization during each successive solve. In [22] Van der Vorst also presented several techniques for exploiting previously generated Krylov spaces. These techniques have been exploited in several engineering contexts, including time dependent structural problems by Farhat, Crivelli, and Roux [4]; boundary element problems by Prasad, Keyes, and Kane [17]; and fluid dynamics applications by Vuik [23]. A clear analysis of the potential of these methods has been recently presented by Saad [19] and by Chan and Wan [1]. The case of multiple systems for which the right-hand sides are available simultaneously has also been extensively studied; however, we do not consider that case here.

The present technique differs from the Krylov based algorithms in that the approximation space is simply the span of the previous solutions. The proposed method is quite similar to reduced bases methods used in nonlinear finite element problems, e.g. [8, 16]. Although the Krylov-based techniques offer potential for greater reduction in iteration count, the dimension of the required basis sets can become quite large, and there is no clear way to continue or update the set of basis vectors for a continuing sequence of right-hand sides once orthogonality is lost, short of resorting to some type of re-orthogonalization. Moreover, the simplicity of present approach allows it to be implemented as a black box, requiring only calls to the iterative solver, "$solve\_A$" and, for reasons to be discussed, the forward operator application, "$multiply\_by\_A$". It is thus quite versatile in situations where the solver/preconditioner is very complex. We have found the method particularly effective in constrained evolution problems such as unsteady incompressible fluid flows and volume-preserving mesh deformation problems.

The outline of the paper is as follows. In Section 2 we describe two projection techniques for generating initial guesses $\bar{\underline{x}} \simeq \underline{x}^n$ based on l previous solutions $\underline{x}^k$, $n-l \leq k \leq n-1$. In Sections 3 and 4 we describe an application of the technique to the incompressible Navier-Stokes equations and present performance results for several fluid dynamics calculations. In Section 5 we analyze the method for a model problem. Finally, we make some concluding remarks regarding memory usage in Section 6.

2 Projection Methods

2.1 Method 1

We begin by assuming that we have stored a set of vectors $B_l = \{\tilde{\underline{b}}_1, \ldots, \tilde{\underline{b}}_l\}$ and solution vectors $X_l = \{\tilde{\underline{x}}_1, \ldots, \tilde{\underline{x}}_l\}$ satisfying:

$$A\tilde{\underline{x}}_k \quad = \quad \tilde{\underline{b}}_k \qquad k = \{1, \ldots, l\} \quad . \tag{2}$$

To simplify the orthogonalization, B_l is assumed to be orthonormal:

$$< \tilde{\underline{b}}_i, \tilde{\underline{b}}_j > \quad = \quad \delta_{ij} \quad , \tag{3}$$

where δ_{ij} is the Kroenecker delta, and $< . >$ is an appropriately weighted inner-product. B_l and X_l are assumed to be derived from the l most recent problems, $\mathcal{P}_k$, $k = n-l, \ldots, n-1$, i.e., $span\{B_l\} = span\{\tilde{\underline{b}}_{n-l}, \ldots, \tilde{\underline{b}}_{n-1}\}$. However, *any* set of independent vectors satisfying (2-3) would constitute a valid basis for approximation.

The algorithm is based upon the following Gram-Schmidt procedure:

$$\begin{aligned} &At\ time\ level\ n, \quad input\ \ \underline{b}^n\text{:} \\ &\qquad \alpha_k =< \underline{b}^n, \tilde{\underline{b}}_k >, \qquad k = 1, \ldots, l \\ &\qquad \tilde{\underline{b}} \longleftarrow \underline{b}^n - \textstyle\sum \alpha_k \tilde{\underline{b}}_k \\ &\qquad solve\ \ A\tilde{\underline{x}} = \tilde{\underline{b}} \quad \text{to tolerance } \epsilon \\ &\qquad \underline{x}^n \longleftarrow \tilde{\underline{x}} + \textstyle\sum \alpha_k \tilde{\underline{x}}_k \\ &\qquad update\ \{B_l, X_l\} \\ &\qquad return\ \ \underline{x}^n \end{aligned} \tag{4}$$

The computation of $\tilde{\underline{b}}$ is simply standard Gram-Schmidt orthogonalization of $\underline{b}^n$ with respect to B_l. The α_i's are computed in a group prior to modifying $\underline{b}^n$ so that the inner-products may be carried out in a single $O(\log_2 P)$ data exchange of an l-vector when the computation is performed on a P processor distributed memory computer. If necessary, it is possible to employ a more stable modified Gram-Schmidt procedure [2, 22] for the computation of $\tilde{\underline{b}}$, at the expense of l individual vector reductions. However, we have not found this to be necessary in any application to date, probably due to the fact that the orthogonality condition (3) results from a Gram-Schmidt procedure (below) rather than from recursion as in the case of Krylov methods.

To complete the procedure (4), we require a mechanism for updating the basis sets $\{B_l, X_l\}$. Initially, the sets are empty, and can be filled with the first l solutions and data. In fact, since $\tilde{\underline{b}} \perp \tilde{\underline{b}}_k$, $k = \{1, \ldots, l\}$ by construction, the vector pair $\{\tilde{\underline{b}}, \tilde{\underline{x}}\}$ seems like a likely candidate to add to the basis set. However, this will generally not be stable because $A\tilde{\underline{x}} = \tilde{\underline{b}}$ is not satisfied exactly. This situation can be corrected by re-computing the required inhomogeneity, i.e., setting $\hat{\underline{b}} = A\tilde{\underline{x}}$, and enforcing (3) via a second Gram-Schmidt procedure. Additionally, we need a strategy for deciding which vectors to keep when the size of the basis set exceeds available memory capacity. There are several possibilities, e.g., retaining those vectors which repeatedly capture most of the energy in $\underline{b}^n$. Initial trials indicate that a reasonable approach is to save just the solution to the current problem, $\underline{x}^n = \tilde{\underline{x}} + \sum \alpha_k \tilde{\underline{x}}_k$, which is a near optimal linear combination of elements in the current basis set.

We summarize the update procedure as follows. If L is taken to be the maximum number of vector pairs to be stored, i.e., $l \leq L$, then at each time step:

$$
\begin{array}{lll}
\textit{If } (l = L) \textit{ then:} & \tilde{\underline{b}}_1 \longleftarrow A\underline{x}^n / ||A\underline{x}^n|| & (5) \\
& \tilde{\underline{x}}_1 \longleftarrow \underline{x}^n / ||A\underline{x}^n|| & \\
& l = 1 & \\
\textit{else:} & \hat{\underline{b}} \longleftarrow A\tilde{\underline{x}} & \\
& \alpha_k =< \hat{\underline{b}}, \tilde{\underline{b}}_k >, \qquad k = 1, \ldots, l & \\
& \tilde{\underline{b}}_{l+1} \longleftarrow (\hat{\underline{b}} - \sum \alpha_k \tilde{\underline{b}}_k) / ||\hat{\underline{b}} - \sum \alpha_k \tilde{\underline{b}}_k|| & \\
& \tilde{\underline{x}}_{l+1} \longleftarrow (\tilde{\underline{x}} - \sum \alpha_k \tilde{\underline{x}}_k) / ||\hat{\underline{b}} - \sum \alpha_k \tilde{\underline{b}}_k|| & \\
& l = l + 1 & \\
\textit{endif} & &
\end{array}
$$

Here, $||.|| =< . >^{\frac{1}{2}}$. The procedure re-initializes $\{B_l, X_l\}$ with the most recent solution pair when the memory limits are exceeded, and then reconstructs a set which satisfies (2-3).

2.2 Method 2: A-conjugate Projection

The procedure of the preceding section follows the intuitive line of reasoning that if $\underline{b}^n$ is well approximated by $\bar{\underline{b}} = \sum \alpha_j \tilde{\underline{b}}_j$, then $\underline{x}^n$ will be well approximated by $\bar{\underline{x}} = \sum \alpha_j \tilde{\underline{x}}_j$. The degree of approximation can be quantified by noting that $\bar{\underline{b}}$ is the L_2 projection of $\underline{b}^n$ onto B_l, which implies that $\bar{\underline{x}}$ is the best approximation to $\underline{x}^n$ in X_l with respect to the A^2-norm: $||\underline{x}||_{A^2} \equiv < A\underline{x}, A\underline{x} >^{\frac{1}{2}}$. If A is symmetric positive definite and conjugate gradient iteration is employed, it is sensible to begin with a projection which minimizes the distance between $\underline{x}^n$ and X_l in the A-norm, $||\underline{x}||_A \equiv < \underline{x}, A\underline{x} >^{\frac{1}{2}}$, since the conjugate gradient method seeks approximations which successively minimize the error in the A-norm [7].

The derivation of the resultant projection method is based upon a straightforward minimization procedure. Assuming as before that we have a set of previous solution vectors $X_l = \{\tilde{\underline{x}}_i\}$, $i = 1, \ldots, l$, we seek coefficients α_i such that the approximation given by

$$\bar{\underline{x}} = \sum_{i=1}^{l} \alpha_i \tilde{\underline{x}}_i \tag{6}$$

minimizes the error in the A-norm:

$$||\underline{x}^n - \bar{\underline{x}}||_A^2 \quad = \quad (\underline{x}^n)^T A \underline{x}^n \; - \; 2 \sum_{i=1}^{l} \alpha_i (\tilde{\underline{x}}_i^T A \underline{x}^n) \; + \; \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j \, (\tilde{\underline{x}}_i^T A \tilde{\underline{x}}_j) \quad . \tag{7}$$

The minimization procedure is simplified if we insist that the $\tilde{\underline{x}}_i$'s are A-conjugate and normalized to satisfy:

$$\tilde{\underline{x}}_i^T A \tilde{\underline{x}}_j \; = \; \delta_{ij} \quad . \tag{8}$$

Requiring a vanishing first variation of (7) with respect to α_i leads to:

$$\alpha_i \; = \; \tilde{\underline{x}}_i^T A \underline{x}^n \; = \; \tilde{\underline{x}}_i^T \underline{b}^n \quad , \quad i = 1, \ldots, l \quad . \tag{9}$$

Thus, given a set of vectors $X_l = \{\tilde{\underline{x}}_i\}$ satisfying (8), the best approximation to $\underline{x}^n$ is found by simply projecting $\underline{b}^n$ onto X_l. This forms the kernel of Method 2:

$$
\begin{aligned}
&\textit{At time level } n, \quad \textit{input } b^n\textit{:} \\
&\qquad \alpha_i = \underline{\tilde{x}}_i^T \underline{b}^n, \qquad i = 1, \ldots, l \\
&\qquad \underline{\bar{x}} \longleftarrow \textstyle\sum \alpha_i \underline{\tilde{x}}_i \\
&\qquad \underline{\tilde{b}} \longleftarrow \underline{b}^n - A\underline{\bar{x}} \\
&\qquad \textit{solve } A\underline{\tilde{x}} = \underline{\tilde{b}} \quad \text{to tolerance } \epsilon \\
&\qquad \underline{x}^n \longleftarrow \underline{\tilde{x}} + \underline{\bar{x}} \\
&\qquad \textit{update } \{\tilde{X}_l\} \\
&\qquad \textit{return } \underline{x}^n
\end{aligned}
\tag{10}
$$

Notice that the storage for this procedure is roughly half that of Method 1 as it only requires X_l, and not B_l. However, one additional A-multiply is required prior to the *solve_A* stage.

As before, we need a mechanism to update the set X_l. To satisfy (8) it is necessary to project the most recent solution, $\underline{x}^n$, onto $X_l^\perp$ and normalize the result. If we do not insist upon a modified Gram-Schmidt procedure, this can be done with a single multiply by A as follows. If L is taken to be the maximum number of vector pairs to be stored, i.e., $l \leq L$, then at each time level:

$$
\begin{aligned}
&\textit{If } (l = L) \textit{ then:} \quad \underline{\tilde{x}}_1 \longleftarrow \underline{x}^n / ||\underline{x}^n||_A \\
&\qquad\qquad\qquad\quad\; l \longleftarrow 1 \\
&\textit{else:} \qquad\qquad\quad\; \alpha_i = \underline{\tilde{x}}_i^T A\underline{\tilde{x}}, \qquad i = 1, \ldots, l \\
&\qquad\qquad\qquad\quad\; \underline{\tilde{x}}_{l+1} \longleftarrow (\underline{\tilde{x}} - \textstyle\sum \alpha_i \underline{\tilde{x}}_i) / ||(\underline{\tilde{x}} - \sum \alpha_i \underline{\tilde{x}}_i)||_A \\
&\qquad\qquad\qquad\quad\; l \longleftarrow l + 1 \\
&\textit{endif}
\end{aligned}
\tag{11}
$$

Note that in (11) the required normalization satisfies $||(\underline{\tilde{x}} - \sum \alpha_i \underline{\tilde{x}}_i)||_A = (\underline{\tilde{x}}^T A \underline{\tilde{x}} - \sum \alpha_i^2)^{\frac{1}{2}}$ due to the the A-conjugate relationship (8) and can therefore be computed with no additional A multiplies.

3 Navier-Stokes Implementation

We have implemented the above projection techniques in spectral-element solution of the incompressible Navier-Stokes equations:

$$
\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} &= -\nabla p + \tfrac{1}{Re} \nabla^2 \mathbf{u} \qquad \text{in } \Omega, \\
\nabla \cdot \mathbf{u} &= 0 \qquad\qquad\qquad\quad \text{in } \Omega,
\end{aligned}
\tag{12}
$$

where $\mathbf{u}$ is the velocity vector, p the pressure, and $Re = \frac{UL}{\nu}$ the Reynolds number based on a characteristic velocity and length scale, and kinematic viscosity.

Spatial discretization is based upon decomposition of the computational domain into K spectral elements which are locally mapped to $[-1,1]^d$ in $\mathcal{R}^d$. Within each element, the geometry, solution, and data are expanded in terms of high-order tensor-product polynomial bases in each coordinate direction. Variational projection operators are used to discretize the elliptic equations arising from a semi-implicit treatment of (12) and a consistent variational formulation is used for the pressure/divergence treatment. The velocity is represented by Nth-order Lagrange polynomials on the Gauss-Lobatto-Legendre quadrature points, with

C^0 continuity enforced at element interfaces. The pressure is represented by polynomials of degree $N-2$ based upon the Gauss-Legendre quadrature points. Temporal discretization is based upon an operator splitting in which the nonlinear convective terms are treated explicitly via a characteristic/sub-cycling scheme, and the viscous and divergence operators are treated implicitly. The discretization leads to the following linear Stokes problem to be solved at each time step:

$$\begin{aligned} H\,\underline{u}_i - D_i^T\,\underline{p} &= B\,\underline{f}_i, \qquad i = 1, ..., d \quad , \\ D_i\,\underline{u}_i &= 0 \quad . \end{aligned} \tag{13}$$

Here, H is the discrete equivalent of the Helmholtz operator, $\{-\frac{1}{Re}\nabla^2 + \frac{1}{\Delta t}\}$; B is the mass matrix associated with the velocity mesh; $\mathbf{D} = (D_1, ..., D_d)$ is the discrete gradient operator; and underscore refers to basis coefficients. Further details of spectral element discretizations for the Navier-Stokes equations may be found in [11].

The solution of (13) is simplified by a Stokes operator splitting which decouples the viscous and pressure/divergence constraint citempr90. This splitting leads to the solution of a standard Helmholtz equation for each velocity component, while the resulting system for the pressure is similar to (13) save that H is replaced by $\frac{1}{\Delta t}B$. The resulting system can be efficiently treated by formally carrying out block Gaussian elimination (Uzawa decoupling) for $\underline{p}$, leading to:

$$E\underline{p} = \underline{g}, \tag{14}$$

where

$$E = -\sum_{i=1}^{d} D_i\,B^{-1}D_i^T\,, \tag{15}$$

and $\underline{g}$ is the inhomogeneity resulting from the time-split treatment of (12). E corresponds to a consistent Poisson operator for the pressure and, though symmetric-positive definite, is less well conditioned than the Helmholtz problems for the velocity components. Consequently, solution of (14) dominates the Navier-Stokes solution time. The advantage of the Stokes splitting is that no system solves are required when applying E, as B is diagonal.

The consistent Poisson problem (14) is solved via a two-level iteration scheme developed by Rønquist [18] in which a coarse-grid operator is folded into a global conjugate-gradient iteration through deflation [13, 14]. The coarse (subscript c) and fine (subscript f) decomposition is effected through a subdomain-motivated prolongation operator $J \in \mathcal{R}^{m\times K}$, where $m = K(N-1)^d$ is the number of pressure degrees-of-freedom. The column space of the prolongation operator J is intended to approximate the span of the low eigenmodes of the E system; for this particular problem, J maps element-piecewise-constant functions to the m nodes of the underlying spectral element discretization. The pressure is then expressed as $\underline{p} = J\underline{p}_c + \underline{p}_f$, leading to an algebraic reformulation of the original problem as solvable fine and coarse subproblems,

$$E_f\underline{p}_f = \underline{g} - J\,E_c^{-1}J^T\underline{g}\,, \tag{16}$$

$$E_c\underline{p}_c = J^T\underline{g} - J^T E\underline{p}_f\,, \tag{17}$$

respectively. Here $E_f = E - E\,J\,E_c^{-1}J^T E$, and $E_c = J^T E\,J$. The fine system (16) is solved by conjugate-gradient iteration. Once $\underline{p}_f$ is established, the coarse-grid problem is

solved (directly) for $\underline{p}_c$, and the procedure is complete. With appropriate application of a local, element-based preconditioner to E_f, the condition number of the fine system is significantly reduced relative to the originating E matrix.

The projection methods of Section 2 are implemented at the level of equation (14), rather than being applied directly to (16). At each time step, we solve for the change in pressure $\Delta\underline{p}^n \equiv \underline{p}^n - \underline{p}^{n-1}$. Thus, in the notation of Section 2, we take $A = E$, $\underline{x}^n = \Delta\underline{p}^n$, and $\underline{b}^n = \underline{g}^n - E\underline{p}^{n-1}$.

4 Numerical Results

We first consider the problem of two-dimensional start-up flow past a cylinder at $Re = \frac{DU_\infty}{\nu} = 200$. The discretization consists of $K = 116$ spectral elements of degree $N = 9$ ($m = 7424$), with time step $\Delta t = .0168$, non-dimensionalized with respect to U_∞ and D. The tolerance for the L_2 norm of the pressure residual was set to 3×10^{-6}, a value commensurate with the achievable discrete divergence of the resultant velocity field in 32-bit precision.

In Fig. 1 we plot the required number of pressure iterations per step, $\mathcal{N}_E$, for the cases $L = 0$, 2, and 20, using the A-conjugate projection technique of Section 2.2. For clarity, a 50 step windowed average is presented. Over the non-dimensional time simulated, $t = 0$ to 150, the flow passes through three transient regimes: symmetric wake formation, wake destabilization, and periodic (von Karman) vortex shedding. The first and third regimes are characterized by a high level of dynamic activity, while the second is relatively quiescent, as illustrated in the lower half of Fig. 1 by the time trace of u at a point in the near wake region of the cylinder. In flows devoid of dynamics, the pressure at time t^n is well represented by $\underline{p}^{n-1}$. Hence, little improvement results from incorporating information from more than one time step, as seen in the quiescent regime ($t \simeq 10 - 50$). However, for flows having a richer dynamical structure, the enriched basis of the projection method provides potential for significant savings, as seen in the von Karman street regime in which a two-fold reduction in iteration count is attained for $L = 20$. Increasing the number of basis functions to $L = 30$ brings no further significant reduction in this case.

Table 1 compares the average iteration count per step in the von Karman street regime for several values of Re and Δt. In all cases, the Method 2 ($L = 20$) yields a slight improvement over Method 1 ($L = 20$) and roughly a fifty-percent reduction over the standard case ($L = 0$). This is typical of the performance observed in other two- and three-dimensional flows of similar complexity. In large three-dimensional flows, the savings in pressure iterations typically translates into a fifty-percent reduction in CPU time [5].

Table 1: Iteration count for flow past a cylinder.

Re	Δt	Standard	Meth. 1	(ratio)	Meth. 2	(ratio)
100	0.01	65	45	(.68)	39	(.59)
200	0.01	90	48	(.53)	43	(.48)
100	0.04	125	65	(.52)	61	(.49)
200	0.04	159	89	(.56)	85	(.53)

As a second example, we consider the benchmark problem of computing the growth rate of

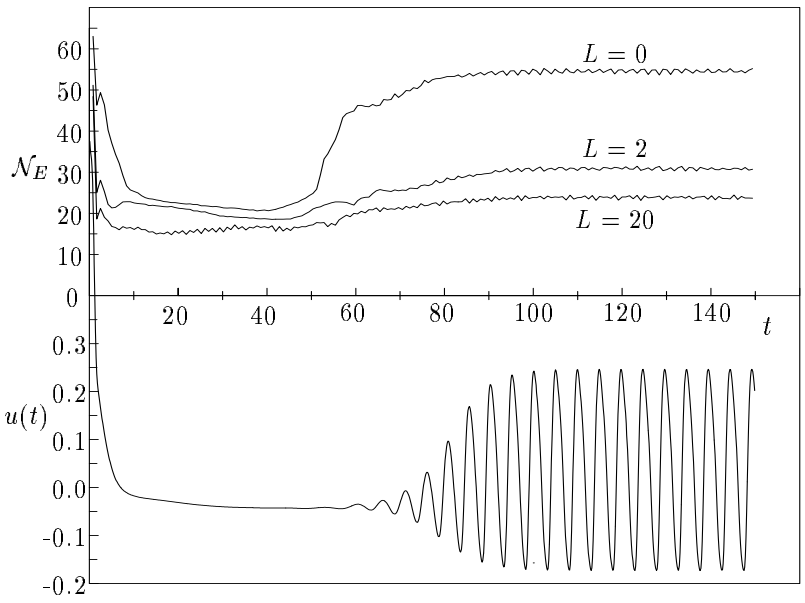

Figure 1: Pressure iteration count and time history of velocity for impulsively started flow past a cylinder at $Re = 200$.

small amplitude three-dimensional Tollmien-Schlichting (TS) waves in plane Poiseuille flow at $Re = 1500$, e.g. [6]. The domain consists of two-flat plates separated by a distance $2h$, periodic boundary conditions in the streamwise and spanwise directions with periodicity lengths $2\pi h$. The initial condition is a parabolic profile with unit centerline velocity, with a superimposed three-dimensional TS wave corresponding to the least damped eigenmode having amplitude 10^{-4} and horizontal wave numbers α and β of unity. For the spectral element calculation with $K = 54$, $N = 7$, and non-dimensional time step $\Delta t = .00625$, the observed growth rate is $Im(\omega_{\rm SE}) = -.028273$, compared to $Im(\omega_{\rm LT}) = -.028230$ predicted by linear theory. The calculations were performed in 64-bit arithmetic and the pressure tolerance was set to 10^{-13} in order to observe high-order spatial and second-order temporal convergence rates.

In Fig. 2a we compare the required number of pressure iterations for the A-conjugate projection method ($L = 80$) to the standard case ($L = 0$). For this problem, the projection method reduces the number of iterations from roughly 60 to as few as *one* per time step, with an average of 3.3. A peak of roughly 20 iterations results when the basis set is restarted, e.g., at step 83. The addition of more basis vectors does not further reduce the iteration count other than by reducing the frequency of restart. The corresponding pre-solver residual and Navier-Stokes solution times shown in Figs. 2b and 2c indicate respective fifty- and four-fold reductions. The computations were carried out on an eight-node Intel iPSC/860. We note that the savings attained is typical for this particular class of problems. However, the performance of the projection techniques for these convergence benchmarks is exceptional and does not reflect the reduction attained in more general engineering flows.

5 Analysis

From the examples of the previous section it is clear that the projection techniques are capable of providing a significant reduction in work. We now analyze a model problem which

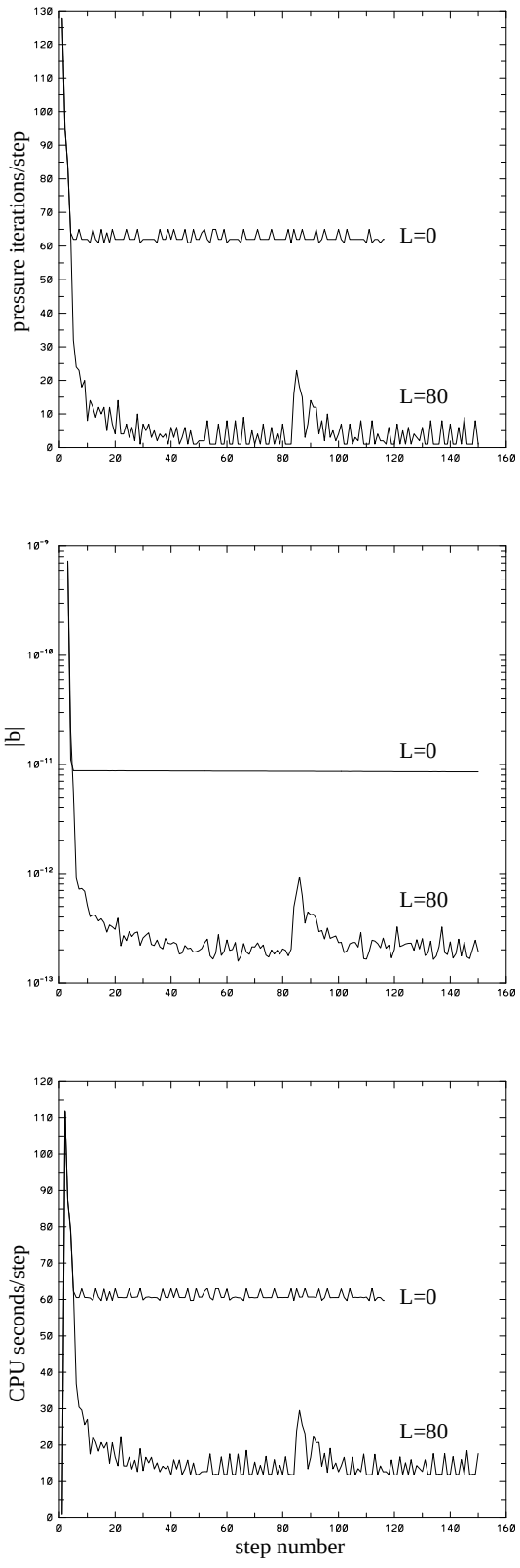

Figure 2: Iteration count (a), initial residual (b), and eight-processor iPSC/860 CPU time (c) for the A-conjugate projection technique applied to the three-dimensional Tollmien-Schlichting wave benchmark problem with $K = 54$ elements of order $N = 7$.

illustrates the leading order terms contributing to the quality of the projected approximation. In particular, we show that for *constrained* evolution problems such as the unsteady Stokes problem, projection techniques based upon previous solutions can have potential advantages over those using Krylov bases in the originating operator.

5.1 Constrained problems

We consider a simplified model of the incompressible Navier-Stokes equations:

$$\begin{bmatrix} H & D^T \\ D & 0 \end{bmatrix} \begin{pmatrix} \underline{u}^{l+1} \\ \underline{p}^{l+1} \end{pmatrix} = \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} \underline{u}^{l-1} \\ \underline{p}^{l-1} \end{pmatrix} \quad . \tag{18}$$

Here, H is assumed to the SPD Helmholtz operator, $H = I + \nu \Delta t A$, where A is the (negative) discrete Laplacian. C is assumed to be a non-symmetric convection operator, $C = I + \Delta t U \mathcal{D}$, where $\mathcal{D}$ is some form of discrete gradient. Note that (18) is representative of *any* evolution problem: $H\underline{u}^{l+1} \approx C\underline{u}^{l}$, subject to the constraint $D\underline{u}^{l+1} = 0$. The crucial observation is that, if (18) represents an *evolution* equation, then the spectrum of $H^{-1}C$ will be contained within a region of diameter $O(\Delta t)$ near $(1,0)$ in the complex plane.

Formally inverting the Stokes operator on the left of (18) yields

$$\begin{pmatrix} \underline{u}^{l+1} \\ \underline{p}^{l+1} \end{pmatrix} = \begin{bmatrix} H^{-1}(I-P)C & 0 \\ E^{-1}DH^{-1}C & 0 \end{bmatrix} \begin{pmatrix} \underline{u}^{l} \\ \underline{p}^{l} \end{pmatrix}, \tag{19}$$

where $E \equiv D\,H^{-1}D^T$ is the Schur complement system governing the pressure, and $I - P \equiv I - D^T E^{-1} D H^{-1}$ is a projection operator which ensures that the solution $\underline{u}^{l+1}$ satisfies the constraint $D\underline{u}^{l+1} = 0$. Let $M \equiv H^{-1}(I-P)C$. The solution at time level $l+1$ is then:

$$\underline{u}^{l+1} = M^{l+1}\,\underline{u}^{0} \tag{20}$$

$$\underline{p}^{l+1} = E^{-1}DH^{-1}\,M^{l}\,\underline{u}^{0} \quad . \tag{21}$$

Aside from a multiplicative factor, $\underline{p}^{l+1}$ is a monomial of degree l in the matrix M times $\underline{u}^0$

Applying the the projection method of Sec. 2.2 to the problem $E\underline{p}^{l+1} = DH^{-1}C\underline{u}^{l}$, yields an approximation $\bar{\underline{p}}$ satisfying: $\bar{\underline{p}} \in span\{\underline{p}^1, \ldots, \underline{p}^l\}$,

$$\| \underline{p}^{l+1} - \bar{\underline{p}} \|_E \leq \| \underline{p}^{l+1} - \underline{q} \|_E \qquad \forall\, \underline{q} \in span\{\underline{p}^1, \ldots, \underline{p}^l\} \quad . \tag{22}$$

Assuming that (21) is exactly satisfied by the elements of the approximation space, then $\bar{\underline{p}}$ satisfies:

$$\| \underline{p}^{l+1} - \bar{\underline{p}} \|_E \leq \| DH^{-1} \left(M^l - P_{l-1}[M] \right) \underline{u}^0 \|_{E^{-1}} \qquad \forall\, P_{l-1}(x) \in \mathbb{P}_{l-1}(x) \tag{23}$$

$$\leq c\, \| \tilde{P}_l[M] \underline{u}^0 \|_2 \qquad \forall\, \tilde{P}_l(x) \in \tilde{\mathbb{P}}_l(x) \quad . \tag{24}$$

Here, c is a constant independent of l and weakly dependent on Δt. $\mathbb{P}_{l-1}(x)$ is the space of all polynomials of degree $l-1$ or less in the argument x and $\tilde{\mathbb{P}}_l$ is the space of all *monic* polynomials of degree l in the argument:

$$\tilde{P}_l[M] = M^l + a_{l-1}M^{l-1} + \ldots + a_0 I \quad .$$

We can readily determine a bound on $\| \tilde{P}_l[M]\underline{u}^0 \|_2$ when C is SPD, corresponding to the case of unsteady Stokes flow. In this case, the eigenvalues of M are non-negative real. M has

several zero eigenvalues as a result of the embedded projection operator $(I - P)$. However, these have no impact on the bound for $\|\tilde{P}_l[M]\underline{u}^i\|_2$ provided $\underline{u}^i \in \mathcal{R}(M)$, which is always true for $i > 0$. (This suggests that the first solution field might not be a good candidate for the approximation space if $\underline{u}^0$ is not divergence free; in practice we always avoid using the first few fields as candidate basis vectors.) The remaining relevant (i.e., nonzero) part of the spectrum of M is bounded on the positive real axis by the extremal eigenvalues of $H^{-1}C$, $\lambda_{\min}$ and $\lambda_{\max}$. A classic mini-max result based on Chebyshev polynomials (e.g. [3]) gives

$$\|\tilde{P}_l[M]\underline{u}^0\|_2 \quad \leq \quad 2\left(\frac{\lambda_{\max} - \lambda_{\min}}{4}\right)^l \|\underline{u}^0\|_2 \quad . \tag{25}$$

Consequently, given the definitions of H and C, we would expect the reduction in $\|\underline{p}^{l+1} - \underline{\bar{p}}\|_E$ to scale as Δt^l. In the case when C is nonsymmetric, it is more difficult to establish the bound explicitly (in particular, with regards to the relevance of the null-space of M). However, one might expect that the results will not change drastically for problems which continue to be evolutionary in nature.

We illustrate the predicted $O(\Delta t^l)$ convergence behavior by reconsidering the cylinder calculation of Fig. 1 in the von Karman street regime. In this case, C is both non-symmetric and non-linear because the full Navier-Stokes equations are being treated. At each step, the iteration tolerance is set to 10^{-13} in order to simulate the case where (21) is satisfied exactly. The spectral element discretization consists of $K = 186$ elements of order $N = 9$, for a total of $m = 11904$ pressure degrees-of-freedom. In Fig. 3 we plot the residual reduction ratio:

$$r_l(\Delta t) \quad \equiv \quad \frac{\|E(\underline{p}^l - \underline{\bar{p}}^l)\|_2}{\|E\underline{p}^l\|_2} \quad , \tag{26}$$

as a function of the number of basis vectors, l. Three cases are shown, Δt=0.05, 0.025, and 0.0125, corresponding to Courant numbers ranging from 4 to 1. Also plotted are the quantities $r_l(0.05)/2^l$ and $r_l(0.05)/4^l$, shown as *model 1* and *model 2*. Initially, the models scale directly as $r_l(0.025)$ and $r_l(0.0125)$, respectively, as should be the case according to (25). However, the residual does not decrease indefinitely because the approximation space only satisfies (21) to a finite tolerance and because the convection operator contributes non-negligible terms to the residual (particularly for larger Δt) which are unaccounted for in the preceding analysis. For reference, the associated pressure histories are also shown in Fig. 3 (right). The flow has been restarted at a time of $t = 150$ (see Fig. 1). For each curve, the projection technique was initiated at the same time, corresponding to (restarted) step numbers 10, 20, and 40 respectively. The impact of the projection scheme on the iteration count is clearly visible at those steps.

Fig. 4 shows the residual reduction ratio when the pressure tolerance is varied. The time step size $\Delta t = .0125$. The *rate* of convergence is observed to be the same as in Fig. 3. However, the magnitude of the residual reduction ratio decreases when less stringent tolerances are imposed; the approximation space fails to satisfy (21) exactly and the favorable polynomial approximation properties are lost. In essence, one is trying to construct a divergence-free flow field, $\underline{\bar{u}}$, with a set of basis functions which are not completely divergence-free and the net improvement over the initial approximation is limited. The associated pressure histories are shown in Fig. 4 (right). Note that even though the 2-norm of the residual is not significantly reduced in the $(\Delta t = .0125, tol = 10^{-5})$ case, the pressure count *is* reduced slightly more than two-fold.

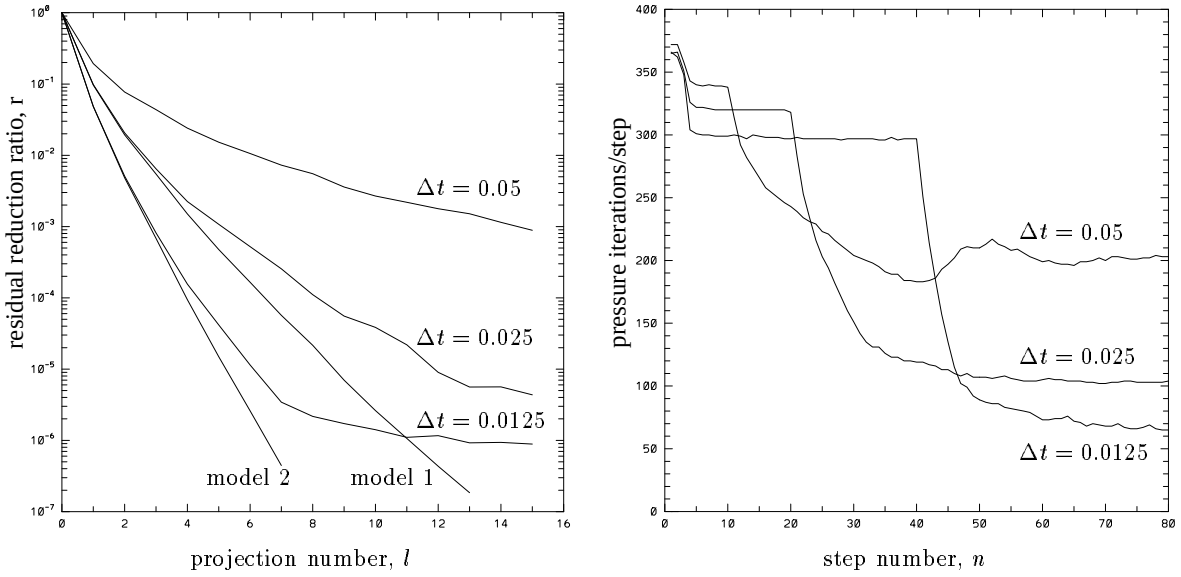

Figure 3: Measured and predicted residual reduction ratios (left) for unsteady two-dimensional flow past a cylinder ($Re_D = 200$) with iteration tolerance set to 10^{-13}, and time step size $\Delta t = .0125$ to 0.05, corresponding to Courant numbers of $U\Delta t/\Delta x = 1$ to 4. Corresponding pressure iteration counts (right).

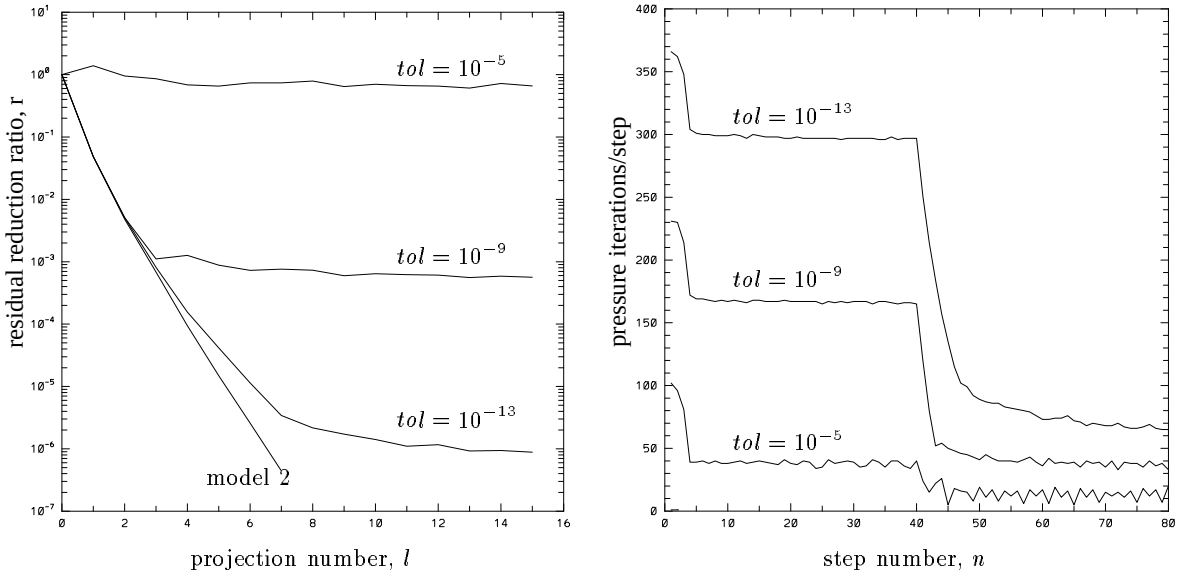

Figure 4: Measured and predicted residual reduction ratio (left) as in Fig. 3, with iteration tolerance varying from to 10^{-5} to 10^{-13} and fixed time step size $\Delta t = .0125$. Corresponding pressure iteration counts (right).

5.2 Other evolution problems

Finally, we comment on the possibility of using an approximation space of the form $span\{\underline{u}^0, \ldots, \underline{u}^{l-1}\}$ to generate approximations to the *unconstrained* evolution equation:

$$H\underline{u}^{l+1} \quad = \quad C\underline{u}^l \quad . \tag{27}$$

Proceeding as before, one would obtain a bound involving the term $\|\tilde{P}_l[H^{-1}C]\underline{u}^0\|_2$, to be minimized over the space of monic polynomials. By contrast, if one were to base the initial guess upon the Krylov subspace $K_k^l = \{\underline{u}^l, (\tilde{H}^{-1}H)\underline{u}^l, \ldots, (\tilde{H}^{-1}H)^{k-1}\underline{u}^l\}$, generated by conjugate gradient solution of the previous time step using preconditioning matrix $\tilde{H}^{-1}$, the bound would be determined in terms of a polynomial in $\tilde{H}^{-1}H$. Since one has complete freedom in choosing $\tilde{H}^{-1}$, there is no *a priori* reason to expect the spectrum of $H^{-1}C$ to be more compact than that of $\tilde{H}^{-1}H$. In fact, the opposite is likely to be true, implying that the prior Krylov subspace will have better approximation properties than an equivalently dimensioned previous-solution space. However, in the *constrained* problem analyzed above, the conditioning of E and the conditioning of $H^{-1}(I-P)C$ are largely decoupled. Depending on the nature of the constraint, it is quite possible that the spectrum of $H^{-1}(I-P)C$ will be more compact than that of the preconditioned matrix, $\tilde{E}^{-1}E$, due to the fact that $H^{-1}(I-P)C$ represents an evolution operator.

6 Concluding Remarks

We remark that the $O(mL)$ memory requirement for the projection methods may at first seem quite high. However, this must be examined in the context of the application. First, the present application is for a general geometry Navier-Stokes solver, rather than just a linear equation solver. Consequently, the *total* memory requirements are already quite high, as it is necessary to store the grid coordinates, metrics, Jacobians, etc., as well as several scalar and vector fields. In addition, efficient iterative solvers generally require significant storage for preconditioners - with memory costs scaling at least as m. Thus, the relative increase in memory demanded by saving a set of basis vectors may not be prohibitive. Secondly, on dedicated distributed memory machines, these algorithms provide a classic example of superlinear speedup; for a problem of fixed size, increasing the number of processors results in increased memory, thus allowing an increased value of L and corresponding decrease in $\mathcal{N}_E$. Our preference is to regard this fact as a flaw in the fixed-problem-size parallel performance metric, rather than to claim that projection techniques are a pathway to super-linear parallel algorithms. It is nonetheless a classic example of a space-time trade-off which can have a very real impact in many circumstances.

Acknowledgements

The author would like to thank Catherine Mavriplis, Anthony Patera, and Einar Rønquist for useful comments in the course of this work. This work was supported by the NSF under Grant # ASC-9405403, and in part by NASA Contract No. NAS1-19480 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA.

References

[1] T.F. Chan and W.L. Wan, "Analysis of projection methods for solving linear systems with multiple right-hand sides," Tech. Rep. CAM-94-26, UCLA, Dept. of Math., Los Angeles, CA 90024 (1994).

[2] G. Dahlquist and Å. Björk, *Numerical Methods*, Prentice Hall, Englewood Cliffs, New Jersey, 1974.

[3] P.J. Davis, *Interpolation and Approximation*, Dover Publications, Inc., N.Y., 1975.

[4] C. Farhat, L. Crivelli, and F.X. Roux, "Extending substructure based iterative solvers to multiple load and repeated analyses," *Comput. Methods Appl. Mech. Engrg.* **117** (1994) 195-209.

[5] P.F. Fischer, "Domain Decomposition Methods for Large Scale Parallel Navier-Stokes Calculations" in *Proceedings of the Sixth International Conference on Domain Decomposition Methods for Partial Differential Equations, Como Italy*, A. Quarteroni, Ed., AMS (1994) p.313-322.

[6] S.E. Krist and T.A. Zang, "Numerical simulation of channel flow transition," *NASA Tech. Paper 2667*, LaRC, Hampton, VA, 1987.

[7] G.H. Golub and C.F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.

[8] M.D. Gunzburger *Finite Element Methods for Viscous Incompressible Flows*, Academic Press, San Diego, CA 1989.

[9] L.A. Hageman and D.M. Young, *Applied Iterative Methods*, Academic Press, Orlando Florida, 1981.

[10] C. Lanczos, "Solution of systems of linear equations by minimized iterations," *J. of Res. of the Natl. Bureau of Standards*, **49** 1 (1952) pp. 33-53.

[11] Y. Maday, and A.T. Patera, "Spectral element methods for the Navier-Stokes equations" in *State of the Art Surveys in Computational Mechanics*, A.K. Noor, ed., ASME, New York. (1989) pp. 71-143.

[12] Y. Maday, A.T. Patera, and E.M. Rønquist, "An operator-integration-factor splitting method for time-dependent problems: application to incompressible fluid flow." *J. Sci. Comput.*, **5**(4), (1990) pp. 310-37.

[13] L. Mansfield, "On the use of deflation to improve the convergence of conjugate gradient iteration," *Comm. in Appl. Numer. Meth.*, **4**, (1988) pp. 151-56.

[14] R.A. Nicolaides, "Deflation of conjugate gradients with applications to boundary value problems," *SIAM J. Numer. Anal.*, **24**(2), (1987) pp. 355-65.

[15] D.P. O'Leary and O. Widlund, "Capacitance matrix methods for the Helmholtz equation on general three-dimensional regions," *Math. Comp.*, **33** 147 (1979) pp. 849-879.

[16] J. Peterson, "The reduced basis method for incompressible viscous flow calculations," *SIAM J. Sci. Statist. Comput.* **10** (1989), pp. 777-786.

[17] K. Prasad, D. Keyes, and J. Kane, "GMRES for sequentially multiple nearby systems," ICASE Rep. (1994), NASA Langley Research Center, Hampton, VA.

[18] E.M. Rønquist, "A Domain Decomposition Method for Elliptic Boundary Value Problems: Application to Unsteady Incompressible Fluid Flow," in *Fifth Conference on Domain Decomposition Methods for Partial Differential Equations* T.F. Chan, D.E. Keyes, G.A. Meurant, J.S. Scroggs, and R.G. Voigt, eds., SIAM, Philadelphia, PA, 1992.

[19] Y. Saad, "Analysis of augmented Krylov subspace methods," *SIAM J. of Matrix Anal. and Appl.*, to appear.

[20] Y. Saad, "On the Lanczos method for solving symmetric linear systems with several right-hand sides," *Math. Comp.* **48** 178 (1987), pp. 651-62.

[21] H.D. Simon, "The Lanczos algorithm with partial reorthogonalization," *Math. Comp.* **42** 165 (1984), pp. 115-142.

[22] H.A. van der Vorst, "An iterative method for solving $f(A)x = b$ using Krylov subspace information obtained for the symmetric positive definite matrix A," *J. Comput. App. Math.*, **18** (1987), pp. 249-63

[23] C. Vuik, "Fast iterative solvers for the discretized incompressible Navier-Stokes equations," *Int. J. for Num. Meth. Fluids* **22** (1996) pp. 195-210.