# Improving Robustness and Parallel Scalability of Newton Method Through Nonlinear Preconditioning*

Feng-Nan Hwang[1] and Xiao-Chuan Cai[2]

[1] Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA (hwangf@colorado.edu)
[2] Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA (cai@cs.colorado.edu,http://www.cs.colorado.edu/~cai/)

**Summary.** Inexact Newton method with backtracking is one of the most popular techniques for solving large sparse nonlinear systems of equations. The method is easy to implement, and converges well for many practical problems. However, the method is not robust. More precisely speaking, the convergence may stagnate for no obvious reason. In this paper, we extend the recent work of Tuminaro, Walker and Shadid [2002] on detecting the stagnation of Newton method using the angle between the Newton direction and the steepest descent direction. We also study a nonlinear additive Schwarz preconditioned inexact Newton method, and show that it is numerically more robust. Our discussion will be based on parallel numerical experiments on solving some high Reynolds numbers steady-state incompressible Navier-Stokes equations in the velocity-pressure formulation.

## 1 Introduction

Many computational science and engineering problems require the numerical solution of large, sparse nonlinear systems of equations. Several classes of approaches are available, including Newton type methods, multigrid type methods, and continuation type methods. However, for some difficult problems, such as incompressible flows with high Reynolds number ($Re$), none of the methods works well, except the continuation methods, e.g. parameter continuation [8] and pseudo time stepping [10], which are often too slow to be considered practical. In general, nonlinear iterative methods are fragile. They may converge rapidly for a well-selected set of parameters (for example, certain initial guesses, certain range of $Re$), but diverge if we slightly change some of the parameters. They may converge well at the beginning of

* The work was partially supported by the Department of Energy, DE-FC02-01ER25479, and by the National Science Foundation, CCR-0219190, ACI-0072089 and ACI-0305666.

the iterations, then suddenly stall for no apparent reason. In this paper we develop some techniques for detecting the bad behavior of Newton method, and focus on a class of nonlinear preconditioning methods that make Newton more robust; i.e., not too sensitive to some of the unfriendly parameters such as large $Re$. The preconditioner is constructed using the nonlinear additive Schwarz method, which not only increases the robustness of Newton, but also maintains the parallel scalability of the algorithm.

## 2 A brief review of inexact Newton method

Solving a nonlinear system of equations,

$$F(x) = 0, \tag{1}$$

using inexact Newton with backtracking (INB) [6] can be described briefly as

$$x^{(k+1)} = x^{(k)} - \lambda^{(k)} s^{(k)},$$

where $\lambda^{(k)}$ is the step length computed using a linesearch technique, and $s^{(k)}$ is a good search direction if a non-zero $\lambda^{(k)}$ can be found. $s^{(k)}$ is computed, often from a *linearly preconditioned* Jacobian equation

$$M_k^{-1} J s^{(k)} = M_k^{-1} F(x^{(k)}),$$

where $J$ is a Jacobian of $F$ and $M_k^{-1}$ is a linear preconditioner. It has been known for a long time that, even with global strategies, INB often stagnates for many problems. A recent study [15] shows that this is likely because the angle between the Newton direction and the steepest descent direction is too close to $\pi/2$. In this case, the Newton direction becomes only a weak descent direction. As a result, only extremely small steps can be accepted by linesearch. More precisely, let $\theta$ be the angle between $s^{(k)}$ and the negative gradient direction of $\|F\|$ at $x^{(k)}$. Then, according to [15], in the worst case,

$$\frac{1}{\kappa(J)} \leq \cos(\theta) \leq \frac{2}{\kappa(J)}, \tag{2}$$

where $\kappa(J)$ is the condition number of $J$. This means that the Newton direction can be nearly orthogonal to the gradient of $\|F\|$ when $\kappa(J)$ is large. In the incompressible Navier-Stokes equations, $\kappa(J)$ becomes very large when $Re$ is high or when the mesh size is fine. Estimate (2) also suggests that sometimes solving the Jacobian system too accurately is not a good idea, even without considering the issue of computational cost. It might be better to stop the Jacobian iteration earlier. The following stopping conditions were suggested in [6],

$$||F(x^{(k)}) - J s^{(k)}||_2 \leq \eta_k ||F(x^{(k)})||_2$$

- Choice 0: $\eta_k$ is a constant (not too small)
- Choice 1:
$$\eta_k = \frac{\left| ||F(x^{(k)})||_2 - ||F(x^{(k-1)}) - Js^{(k-1)}||_2 \right|}{||F(x^{(k-1)})||_2}$$
- Choice 2:
$$\eta_k = \gamma \left( \frac{||F(x^{(k)})||_2}{||F(x^{(k-1)})||_2} \right)^{\alpha}, \gamma \in [0,1], \alpha \in (1,2].$$

INB with these forcing terms is more robust, but is still not enough to solve the Navier-Stokes equations for a large range of $Re$ because the parameters in the "choices" are too problem-dependent [14]. A closer look at (2) and its proof in [15] shows that the linear preconditioner $M_k^{-1}$ does not appear in the estimate (2), which means that even though the linear preconditioning may speed up the solution algorithm for the Jacobian system, it does not help improve the quality of the search direction. Therefore, to enhance the robustness of Newton method by finding a better search direction we believe that the preconditioner has to be *nonlinear*. An alternative approach to improve the quality of the search direction is based on the affine invariant Newton methods [5] using the natural monotonicity test for highly nonlinear systems.

## 3 Nonlinear additive Schwarz preconditioning

This section describes a nonlinearly preconditioned inexact Newton algorithm (ASPIN) [3, 9]. Suppose that $F(x) = 0$ is a nonlinear system of equations arising from a finite element discretization. The finite element mesh on $\Omega$ is partitioned into non-overlapping subdomains $\Omega_i$, $i = 1, \ldots, N$, then, each subdomain is extended into a larger overlapping subdomain $\Omega_i'$. Let $R_i$ be a restriction operator on $\Omega_i'$, we define the subdomain nonlinear function

$$F_i = R_i F.$$

For any given $x \in R^n$, $T_i(x)$ is defined as the solution of the subspace nonlinear systems,

$$F_i\left(x - R_i^T T_i(x)\right) = 0, \text{ for } = 1, ..., N. \tag{3}$$

Using the subdomain functions, we introduce a new global nonlinear system

$$\mathcal{F}(x) = \sum_{i=1}^{N} R_i^T T_i(x) = 0, \tag{4}$$

which we refer to as the nonlinear additive Schwarz preconditioned system. Then, ASPIN algorithm is defined as: find a solution of (1) by solving

$$\mathcal{F}(x) = 0,$$

with INB, starting with an initial guess $x^{(0)}$. As shown in [3, 9], an approximation of the Jacobian of $\mathcal{F}$ takes the form $\sum_{i=1}^{N} J_i^{-1} J$. Through nonlinear preconditioning, we have:

- an improved angle estimate

$$\frac{1}{\kappa(\sum_{i=1}^{N} J_i^{-1} J)} \leq \cos(\theta) \leq \frac{2}{\kappa(\sum_{i=1}^{N} J_i^{-1} J)}; \text{ and}$$

- an improved conditioning of the Jacobian system

$$\left( \sum_{i=1}^{N} J_i^{-1} J \right) s^{(k)} = \mathcal{F}(x^{(k)}); \text{ and}$$

- an improved merit function $\|\mathcal{F}\|^2/2$ for the linesearch.

## 4 Stabilized finite element method for incompressible Navier-Stokes equations in the primitive variable

Consider two-dimensional steady-state incompressible Navier-Stokes equations in the primitive variable form [8, 13]:

$$\begin{cases} \boldsymbol{u} \cdot \nabla \boldsymbol{u} - 2\nu \nabla \cdot \epsilon(\boldsymbol{u}) + \nabla p = 0 & \text{in} \quad \Omega, \\ \nabla \cdot \boldsymbol{u} = 0 & \text{in} \quad \Omega, \\ \boldsymbol{u} = \boldsymbol{g} & \text{on} \quad \Gamma, \end{cases} \tag{5}$$

where $\boldsymbol{u}$ is the velocity, $p$ is the pressure, $\nu = 1/Re$ is the dynamic viscosity, and $\epsilon(\boldsymbol{u}) = 1/2(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T)$ is the symmetric part of the velocity gradient. The pressure $p$ is determined up to a constant. To make $p$ unique, we impose an additional condition $\int_\Omega p \, dx = 0$.

To discretize (5), we use a stabilized $Q_1 - Q_1$ finite element method ([7]). For simplicity, we consider only rectangular bilinear mesh $\mathcal{T}_h = \{K\}$. Let $V^h$ and $P^h$ be a pair of finite element spaces for the velocity and pressure, given by

$$V^h = \{\boldsymbol{v} \in (C^0(\Omega) \cap H^1(\Omega))^2 : \ \boldsymbol{v}|_K \in Q_1(K)^2, \ K \in \mathcal{T}_h \ \}$$
$$P^h = \{p \in C^0(\Omega) \cap L^2(\Omega) : \ p|_K \in Q_1(K), \ K \in \mathcal{T}_h\}.$$

The weighting and trial velocity function spaces $V_0^h$ and $V_g^h$ are

$$V_0^h = \{\boldsymbol{v} \in V^h : \boldsymbol{v} = 0 \text{ on } \Gamma\} \text{ and } V_g^h = \{\boldsymbol{v} \in V^h : \ \boldsymbol{v} = g \text{ on } \Gamma\}.$$

Similarly, let the finite element space $P_0^h$ be both the weighting and trial pressure function spaces:

$$P_0^h = \left\{ p \in P^h : \int_\Omega p \, dx = 0 \right\}.$$

Following [7], the stabilized finite element method for steady-state incompressible Navier-Stokes equations reads: Find $\boldsymbol{u}^h \in V_g^h$ and $p^h \in P_0^h$, such that

$$B\left(\boldsymbol{u}^h, p^h; \boldsymbol{v}, q\right) = 0 \qquad \forall (\boldsymbol{v}, q) \in V_0^h \times P_0^h \tag{6}$$

with

$$B(\boldsymbol{u}, p; \boldsymbol{v}, q) = ((\nabla \boldsymbol{u}) \cdot \boldsymbol{u}, \boldsymbol{v}) + (2\nu\epsilon(\boldsymbol{u}), \epsilon(\boldsymbol{v})) - (\nabla \cdot \boldsymbol{v}, p) - (\nabla \cdot \boldsymbol{u}, q) + \sum_{K \in \mathcal{T}_h} ((\nabla \boldsymbol{u}) \cdot \boldsymbol{u} + \nabla p, \tau((\nabla \boldsymbol{v}) \cdot \boldsymbol{v} - \nabla q))_K + (\nabla \cdot \boldsymbol{u}, \delta \nabla \cdot \boldsymbol{v})$$

We use the stability parameters $\delta$ and $\tau$ suggested in [7]. The stabilized finite element formulation (6) can be written as a nonlinear algebraic system

$$F(x) = 0, \tag{7}$$

which is often large, sparse, and highly nonlinear when the value of Reynolds number is large. A vector $x$ corresponds to the nodal values of $\boldsymbol{u}^h = (u_1^h, u_2^h)$ and $p^h$ in (6). Now, we define the subdomain velocity space as

$$V_i^h = \left\{ v^h \in V^h \cap \left(H^1(\Omega_i')\right)^2 : v^h = 0 \text{ on } \partial\Omega_i' \right\}$$

and the subdomain pressure space as

$$P_i^h = \left\{ p^h \in P^h \cap L^2(\Omega_i') : p^h = 0 \text{ on } \partial\Omega_i' \backslash \Gamma \right\}.$$

Using these subspaces we can define subspace nonlinear problems as in (3). Note that, implicitly defined in the subspaces $V_i^h$ and $P_i^h$, we impose Dirichlet conditions according to the original equations (5) on the physical boundaries, and on artificial boundaries, we assume both $\boldsymbol{u} = 0$ and $p = 0$. This is similar to the conditions used in [11].

## 5 Experimental results

To show the convergence properties of ASPIN and its robustness with respect to high Reynolds numbers, in this section we consider a lid-driven cavity flow problem described by (5) on the unit square. We also compare the results with those obtained using a standard Newton-Krylov-Schwarz algorithm [2], which is here referred to as INB. GMRES is used for solving Jacobian systems. A zero initial guess is used for all test cases, and a constant nonlinear tolerance $10^{-6}$ is used for ASPIN and INB. Other parameters to be studied are described briefly as follows. Two meshes of size $64 \times 64$ and $128 \times 128$ are considered.

Reynolds numbers range from $10^3$ to $10^4$. The subdomains are obtained by partitioning the mesh uniformly into either a 2×2 or a 4×4 partition. The number of processors is the same as the number of subdomains. Our parallel software is developed using PETSc of Argonne National Laboratory [1]. More implementation details and numerical results are available in [9].

Figure 1 compares the nonlinear residual history of ASPIN with those of INB with three different choices of forcing terms as described in Section 2. Ten tests are run for Reynolds numbers ranging from $10^3$ to $10^4$, with an increment of $10^3$. All results are obtained by on a 128×128 mesh using 16(=4×4) processors. We see that nonlinear residuals of INB with all choices of forcing terms behave similarly. Except for a few cases with low Reynolds numbers, INB nonlinear residuals stagnate around $10^{-3}$ without any progress after about the first 15 iterations. Different choices of forcing terms do not help much. On the other hand, ASPIN converges for the whole range of Reynolds numbers. Furthermore, ASPIN preserves the local quadratic convergence of Newton when the intermediate solution is near the desired solution.

To understand the robustness of ASPIN and INB, we next compare the minimum values of $\cos(\theta)$ for ASPIN and INB with different forcing terms in Table 1. The values marked with asterisks in the table indicate that INB fails to converge either after 150 nonlinear iterations, or the backtracking step fails. For INB, the minimum value of $\cos(\theta)$ is tiny when INB fails. This agrees well with estimate (2), since $\kappa(J)$ is expected to be very large for this high $Re$. On the other hand, the minimum value of $\cos(\theta)$ for ASPIN is always away from zero and is not sensitive to the change of $Re$ as well as the refinement of the mesh.

**Table 1.** Comparison of the minimum values of $\cos(\theta)$ for ASPIN and INB.

| | $Re = 10^3$ | $Re = 5 \times 10^3$ | $Re = 10^4$ |
|---|---|---|---|
| **Mesh size:** $64 \times 64$ | | | |
| Choice 0 | 1.68e-03 | 8.50e-12* | 6.70e-11* |
| Choice 1 | 4.21e-03 | 6.22e-08* | 1.09e-04* |
| Choice 2 | 4.80e-03 | 4.91e-05* | 1.54e-04 |
| ASPIN | 7.37e-03 | 1.74e-03 | 1.82e-03 |
| **Mesh size:** $128 \times 128$ | | | |
| Choice 0 | 8.65e-04 | 1.97e-07* | 3.31e-07* |
| Choice 1 | 3.78e-03 | 3.30e-05* | 1.82e-08* |
| Choice 2 | 3.33e-03 | 1.20e-04* | 9.27e-05* |
| ASPIN | 2.98e-03 | 2.94e-03 | 3.90e-03 |

Scalability is an important issue in parallel computing and the issue becomes significant when we solve large scale problems with many processors. Table 2 shows that the number of ASPIN iterations does not change much, while the average number of GMRES iterations increases when the number

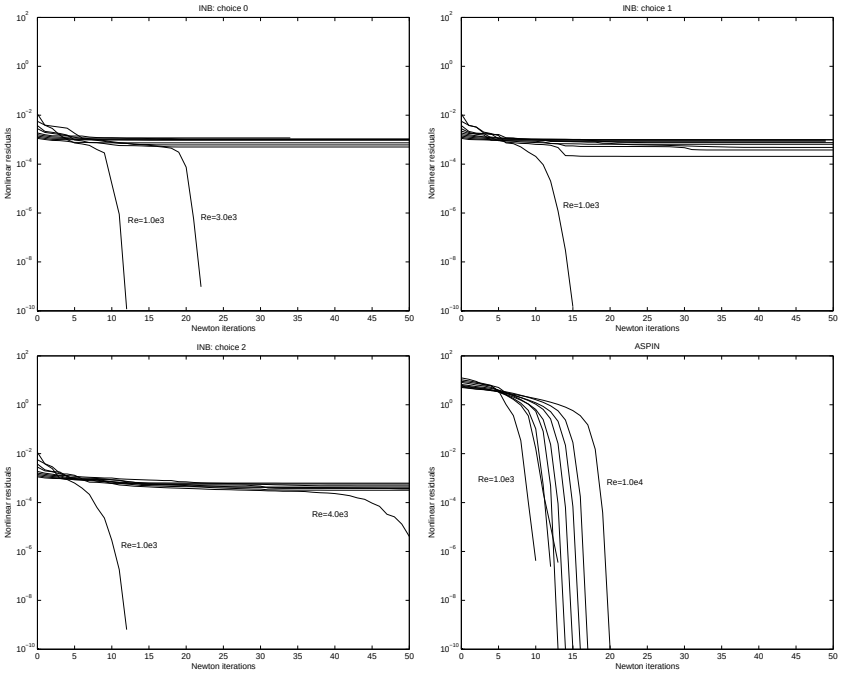


**Fig. 1.** Nonlinear residual curves of ASPIN and INB with three different forcing terms. $Re$ ranges from $10^3$ to $10^4$.

of processors increases from 4 to 16 on a fixed 128×128 mesh. The increase of GMRES iteration numbers is not unexpected since we do not have a coarse space in the preconditioner. The number of GMRES iterations can be kept near a constant if a multilevel ASPIN is used [4, 12].

**Table 2.** Varying the number of processors and the Reynolds number on a 128×128 mesh.

| $np$ | $Re = 10^3$ | $Re = 5 \times 10^3$ | $Re = 10^4$ |
|---|---|---|---|
| | **ASPIN iterations** | | |
| $2 \times 2 = 4$ | 11 | 13 | 19 |
| $4 \times 4 = 16$ | 14 | 13 | 20 |
| | **Average GMRES iterations** | | |
| $2 \times 2 = 4$ | 67 | 71 | 74 |
| $4 \times 4 = 16$ | 128 | 132 | 140 |

## References


[1] S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang, *PETSc Users Manual*, ANL-95/11 - Revision 2.1.5, Argonne National Laboratory, 2002.

[2] X.-C. Cai, W. D. Gropp, D. E. Keyes, R. G. Melvin, and D. P. Young, *Parallel Newton-Krylov-Schwarz algorithms for the transonic full potential equation*, SIAM J. Sci. Comput. 19 (1998) 246-265.

[3] X.-C. Cai and D. E. Keyes, *Nonlinearly preconditioned inexact Newton algorithms*, SIAM J. Sci. Comput. 24 (2002) 183-200.

[4] X.-C. Cai, D. E. Keyes, and L. Marcinkowski, *Nonlinear additive Schwarz preconditioners and applications in computational fluid dynamics*, Int. J. Numer. Meth. Fluids 40 (2002) 1463-1470.

[5] P. Deuflhard, *Global inexact Newton methods for very large scale nonlinear problems*, IMPACT Comp. Sci. Eng. 3 (1991) 366-393.

[6] S. C. Eisenstat and H. F. Walker, *Choosing the forcing terms in an inexact Newton method*, SIAM J. Sci. Comput. 17 (1996) 16-32.

[7] L. P. Franca and S. L. Frey, *Stabilized finite element method: II. The incompressible Navier-Stokes equation*, Comput. Methods Appl. Mech. Engrg. 99 (1992) 209-233.

[8] M. D. Gunzburger, *Finite Element Methods for Viscous Incompressible Flows*, Academics Press, New York, 1989.

[9] F.-N. Hwang and X.-C. Cai, *A parallel nonlinear additive Schwarz preconditioned inexact Newton algorithm for incompressible Navier-Stokes equations*, 2003. (preprint)

[10] C. T. Kelley and D. E. Keyes, *Convergence analysis of pseudo-transient continuation*, SIAM J. Sci. Comput. 35 (1998) 508-523.

[11] A. Klawonn and L. F. Pavarino, *Overlapping Schwarz method for mixed linear elasticity and Stokes problems*, Comput. Methods Appl. Mech. Engrg. 165 (1998) 233-245.

[12] L. Marcinkowski and X.-C. Cai, *Parallel performance of some two-level ASPIN algorithms*, This Proceedings.

[13] J. N. Reddy and D. K. Gartling, *The Finite Element Method in Heat Transfer and Fluid Dynamics*, CRC Press, Florida, 2000.

[14] J. N. Shadid, R. S. Tuminaro, and H. F. Walker, *An inexact Newton method for fully coupled solution of the Navier-Stokes equations with heat and mass transport*, J. Comput. Phys. 137 (1997) 155-185.

[15] R. S. Tuminaro, H. F. Walker, and J. N. Shadid, *On backtracking failure in Newton-GMRES methods with a demonstration for the Navier-Stokes equations*, J. Comput. Phys. 180 (2002) 549-558.