Jacobian-Free Newton-Krylov Methods: A Survey of Approaches and Applications[*]

Dana A. Knoll[†] David E. Keyes[‡]

Abstract

Jacobian-free Newton-Krylov (JFNK) methods are synergistic combinations of Newton-type methods for superlinearly convergent solution of nonlinear equations and Krylov subspace methods for solving the Newton correction equations. The link between the two methods is the Jacobian-vector product, which may be probed approximately without forming and storing the elements of the true Jacobian, through a variety of means. Various approximations to the Jacobian matrix may still be required for preconditioning the resulting Krylov iteration. As with Krylov methods for linear problems, successful application of the JFNK method to any given problem is dependent on adequate preconditioning. JFNK has potential for application throughout problems governed by nonlinear partial differential equations and integro-differential equations. In this survey article we place JFNK in context with other nonlinear solution algorithms for both boundary value problems (BVPs) and initial value problems (IVPs). We provide an overview of the mechanics of JFNK and attempt to illustrate the wide variety of preconditioning options available. It is emphasized that JFNK can be wrapped (as an accelerator) around another nonlinear fixed point method (interpreted as a preconditioning process, potentially with significant code reuse).

The aim of this article is not to trace fully the evolution of JFNK, nor to provide proofs of accuracy or optimal convergence for all of the constituent methods, but rather to present the reader with a perspective on how JFNK may be applicable to problems of physical interest and to provide sources of further practical information.

[*] A review paper solicited by the Editor-in-Chief of the Journal of Computational Physics and submitted to the Journal for review in September 2002. Available on-line at www.math.odu.edu/~keyes/papers/ in successively revised forms.

[†] Theoretical Division, Fluid Dynamics Group (T-3), Los Alamos National Laboratory, MS B216, Los Alamos, New Mexico 87545, nol@lanl.gov.

[‡] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529-0077, http://www.math.odu.edu/~keyes/, dkeyes@odu.edu.

1 Introduction and Background

The need to solve nonlinear systems of algebraic equations is ubiquitous throughout computational physics. Such systems typically arise from the discretization of partial differential equations (PDEs), whether scalar (such as heat conduction) or a system of coupled equations (such as the Navier-Stokes equations). One may be interested in the steady-state solution of these equations (a boundary value problem, BVP) or in their dynamical evolution (an initial value problem, IVP). For BVPs, implicit nonlinear iterative methods are desirable. The same is true for multiple time-scale IVPs, when discretized implicitly at each time step. A particular unified solution algorithm for these two classes of nonlinear systems, the Jacobian-Free Newton-Krylov method (JFNK), is the focus of this survey article. JFNK methods have been developed and applied in many areas of computational physics, but so far by a relatively small number of researchers. The aim of this article is to review recent advances and help accelerate the development and application of JFNK methods by a broader community of computational physicists.

It is our observation that solution strategies for nonlinearly implicit PDEs have evolved along somewhat different trajectories in the applied mathematics community and the computational physics community. In discussing solution strategies for BVPs [73, 125] the applied mathematics community has emphasized Newton-based methods. Outside of finite element practitioners, the computational fluid dynamics (CFD) community has emphasized Picard-type linearizations and splitting by equation or splitting by coordinate direction [1, 126]. The difference in predominating approach (Newton versus Picard) seems stronger for implicit IVPs. Again, the applied mathematics community has focused on Newton-based methods, and on converging the nonlinear residual within a time step. In the computational physics community, operator splitting [16, 124] has been the "bread and butter" approach, with little attention to monitoring or converging the nonlinear residual within a time step, often allowing a splitting error to remain that is first order (or even worse) in time. In both IVP and BVP contexts, the concept of splitting (a form of divide-and-conquer at the operator level) has been motivated by the desire to numerically integrate complicated problems with limited computer resources. This tension does not vanish with terascale hardware, since the hardware is justified by the need to do ever more refined simulations of more complex physics. On can argue that the stakes for effective methods become higher, not lower, with the availability of advanced hardware.

Recent emphasis on achieving predictive simulations (e.g., in the ASCI [103] and SciDAC [122] programs of the U.S. Department of Energy) has caused computational scientists to take a deeper look at operator splitting methods for IVPs and the resulting errors. As a result, the computational physics community is now increasingly driven towards nonlinear multigrid methods [18, 169] and Jacobian-Free Newton-Krylov methods (JFNK) [21, 39, 73]. These nonlinear iterative methods have grown out of advances in linear iterative methods [5, 58], multigrid methods [19, 62, 157, 169], and preconditioned Krylov methods [143].

The standard nonlinear multigrid method is called the "full approximation scheme" or FAS [18, 169]. Whereas a linear multigrid scheme usually solves for a delta correction for the solution based on linearized equations on coarser grid levels, FAS performs relaxation on the full (nonlinear) problem on each successively coarsened grid. In the FAS approach, the nonlinear correction (either Newton or simpler Picard) is not global, but resides inside the cycle over levels and the sweep over blocks of unknowns at each level. As a result, asymptotic quadratic convergence of the overall nonlinear iteration is not guaranteed. The virtues of FAS include its low storage requirement (if one can use a simple smoother), optimal convergence on some problems, and a tendency for an enlarged domain of convergence, relative to a straight Newton method directly on the finest discretization. Disadvantages include the hurdle of forming hierarchical grids, the expertise required to develop coarse grid representations of a nonlinear operator, and the potential for many expensive nonlinear function evaluations. FAS has been used extensively and successfully in many computational fluid dynamics settings [105, 107, 165].

In JFNK methods [21, 39, 73], the nonlinear iterative method is on the outside, and a linear iterative method on the inside. Typically, the outer Newton iteration is "inexact" [44] and strict quadratic convergence is not achieved. Asymptotic quadratic convergence is achievable, but only with effort on the part of the inner, linear iterative method, which is usually unwarranted when overall time to solution is the metric. An advantage of JFNK is that the code development curve is not steep, given a subroutine that evaluates the discrete residual on the desired (output) grid. Furthermore, inexpensive linearized solvers can be used as preconditioners. The storage required for the preconditioner and Krylov vectors may be a limitation.

There have been limited comparisons between FAS and JFNK methods on identical problems. The authors regard both methods as important and complementary. It is not the purpose of this survey article to provide such comparisons, but rather to hasten the applicability of JFNK to new applications, via "real world" examples. Additionally, we direct the reader's attention to ongoing interactions between these two approaches such as JFNK as a smoother for FAS and multigrid as a preconditioner for JFNK [106, 127].

An important feature of JFNK is that the overall nonlinear convergence of the method is not directly affected by the approximations made in the preconditioning. The overall framework, making use of multiple discrete approximations of the Jacobian operator, has a polymorphic object-oriented flavor that lends itself well to modern trends in software design and software integration. In many cases, including some referenced as case studies as DOE and NASA herein, JFNK has been used to retrofit existing BVP and IVP codes while retaining the most important investments (in the physics routines) of the original code.

The remainder of this article is organized as follows. In section 2, we present the fundamentals of the JFNK approach. Section 3 is devoted to considerations of preconditioning. In section 4, we survey examples of JFNK from a variety of applications. In section 5 we illustrate a number of techniques and "tricks" associated with using JFNK in real problems, including many of the applications discussed in section 4. Section 6 describes applications of the JFNK methodology to nonlinear systems with generally dense Jacobians that arise from nonlinear preconditioning. Section 7 considers a novel application of JFNK to PDE-constrained optimization. We conclude in section 8 with a discussion of future directions for JFNK methodology, as influenced by directions for scientific and engineering applications, computer architecture, mathematical software, and the on-going development of other numerical techniques.

2 Fundamentals of the JFNK method

The Jacobian-free Newton-Krylov (JFNK) method is a nested iteration method consisting of at least two, and usually four levels. The primary levels, which give the method its name, are the loop over the Newton corrections and the loop building up the Krylov subspace out of which each Newton correction is drawn. Interior to the Krylov loop, a preconditioner is usually required, which can itself be direct or iterative. Outside of the Newton loop, a continuation scheme is often required. This can be implicit time stepping, with time steps chosen to preserve a physically accurate transient or otherwise, or this can be some other form of parameter continuation such as mesh sequencing.

2.1 Newton Methods

Newton iteration for $\mathbf{F}(\mathbf{u}) = 0$ derives from a multivariate Taylor expansion about a current point $\mathbf{u}^k$:

$$\mathbf{F}(\mathbf{u}^{k+1}) = \mathbf{F}(\mathbf{u}^k) + \mathbf{F}'(\mathbf{u}^k)(\mathbf{u}^{k+1} - \mathbf{u}^k) + \text{higher order terms.} \tag{1}$$

Setting the right-hand side to zero and neglecting the terms of higher-order curvature yields a strict Newton method, iteration over a sequence of linear systems

$$\mathbf{J}(\mathbf{u}^k)\delta\mathbf{u}^k = -\mathbf{F}(\mathbf{u}^k), \qquad \mathbf{u}^{k+1} = \mathbf{u}^k + \delta\mathbf{u}^k, \qquad k = 0, 1, \ldots, \tag{2}$$

given $\mathbf{u}^{\mathbf{0}}$. Here, $\mathbf{F}(\mathbf{u})$ is the vector-valued function of nonlinear residuals, $\mathbf{J} \equiv \mathbf{F}'$ is its associated Jacobian matrix, $\mathbf{u}$ is the state vector to be found, and k is the nonlinear iteration index. For a scalar problem, discretized into n equations and n unknowns, we have

$$\mathbf{F}(\mathbf{u}) = \{F_1, F_2, ..., F_i, ..., F_n\}, \tag{3}$$

and

$$\mathbf{u} = \{u_1, u_2, ..., u_i, ..., u_n\}, \tag{4}$$

where i is the component index. In vector notation, the $(i,j)^{th}$ element (i^{th} row, j^{th} column) of the Jacobian matrix is

$$J_{ij} = \frac{\partial F_i(\mathbf{u})}{\partial u_j}. \tag{5}$$

In this scalar example there is a one-to-one mapping between grid points and rows in the Jacobian. Forming each element of **J** requires taking analytic or discrete derivatives of the system of equations with respect to **u**. This can be both error-prone and time consuming for many problems in computational physics. Nevertheless, there are numerous examples of forming **J** numerically and solving Eq. (2) with a preconditioned Krylov method [68, 82, 90, 109, 148, 149, 151]. **J** can also formed using automatic differentiation [66].

2.2 Krylov Methods

Krylov subspace methods are approaches for solving large linear systems introduced as direct methods in the 1950's [65], whose popularity took off after Reid reintroduced them as iterative methods in 1971 [134] (see the interesting history in [57]). They are projection (Galerkin) or generalized projection (Petrov-Galerkin) methods [143] for solving $\mathbf{Ax} = \mathbf{b}$ using the Krylov subspace, $\mathbf{K}_j$,

$$\mathbf{K}_j = span(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \mathbf{A}^2\mathbf{r}_0, ..., \mathbf{A}^{j-1}\mathbf{r}_0)$$

where $\mathbf{r_0} = \mathbf{b} - \mathbf{Ax_0}$. These methods require only matrix-vector products to carry out the iteration (not the individual elements of **A**) and this is key to their use with Newton's method, as seen below.

A wide variety of iterative methods fall within the Krylov taxonomy [9, 73, 143]. A principal bifurcation in the family tree is applicability to nonsymmetric systems. Since the vast majority of fully coupled nonlinear applications of primary interest result in Jacobian matrices that are nonsymmetric, we focus the discussion on this side of the tree. A further point of discrimination is whether the method is derived from the Arnoldi orthogonalization procedure or the Lanczos bi-orthogonalization procedure.

The widely used Generalized Minimal RESidual method (GMRES) [144] is an Arnoldi-based method. In GMRES, the Krylov vectors are orthonormalized and the latter form the trial subspace out of which the solution is constructed. One matrix-vector product is required per iteration to create each new trial vector, and the iterations are terminated based on a by-product estimate of the residual that does not require explicit construction of intermediate residual vectors or solutions — a major beneficial feature of the algorithm. GMRES has a residual minimization property in the Euclidean norm but requires the storage of all previous Krylov vectors. Full restarts, seeded restarts, and moving fixed-sized windows of Krylov vectors, are all options for fixed-storage versions. Full restart is simple and historically the most popular, though seeded restarts show promise. The Bi-Conjugate Gradient STABilized (BiCGSTAB) [46] and Transpose-free Quasi Minimal Residual (TFQMR) [53] methods are Lancozos-based alternatives to GMRES for nonsymmetric problems. In neither method are the Krylov vectors normalized and two matrix-vector products are required per iteration. However, these methods enjoy a short recursion relation, so there is no requirement to store many Krylov vectors.

We refer to [9, 11, 73, 143] for more details on Krylov methods, and for preconditioning for linear problems. We also call attention to the delightful article [121], which shows that there is no universal ranking possible for iterative methods for nonsymmetric linear problems. Each of the major candidate methods finishes first, last, and in the middle of the pack over the span of a few insight-provoking examples.

As a result of studies in [95, 110], we tend to use GMRES (and its variants) almost exclusively with JFNK. The resulting pressure on memory has put an increased emphasis on quality preconditioning. We believe that it is only through effective preconditioning that JFNK is feasible on large-scale problems. It is in the preconditioner that one achieves algorithmic scaling and also in the preconditioner that one may stand to lose the natural excellent parallel scaling enjoyed by all other components of the JFNK algorithm as applied to PDEs. For this reason we focus our main attention in this review on innovations in preconditioning.

2.3 Jacobian-Free Newton-Krylov Methods

In the JFNK approach, a Krylov method is used to solve the linear system of equations given by Eq. (2). An initial linear residual, $\mathbf{r}_0$, is defined, given an initial guess, $\delta\mathbf{u_0}$, for the Newton correction:

$$\mathbf{r}_0 = -\mathbf{F}(\mathbf{u}) - \mathbf{J}\delta\mathbf{u_0} \; . \tag{6}$$

Note that the nonlinear iteration index, k, has been dropped. This is because the Krylov iteration is performed at a fixed k. Let j be the Krylov iteration index. Since the Krylov solution is a Newton correction, and since a locally optimal move was just made in the direction of the previous Newton correction, the initial iterate for the Krylov iteration for $\delta\mathbf{u_0}$ is typically zero. This is asymptotically a reasonable guess in the Newton context, as the converged value for $\delta\mathbf{u}$ should approach zero in late Newton iterations. The j^{th} GMRES iteration minimizes $\parallel \mathbf{J}\delta\mathbf{u}_j + \mathbf{F}(\mathbf{u}) \parallel_2$ within a subspace of small dimension, relative to n, in a least squares sense. $\delta\mathbf{u}_j$ is drawn from the subspace spanned by the Krylov vectors, $\{\mathbf{r}_0, \mathbf{J}\mathbf{r}_0, (\mathbf{J})^2\mathbf{r}_0, ..., (\mathbf{J})^{j-1}\mathbf{r}_0\}$, orthonormalized in practice, obtained during the previous $j-1$ GMRES iterations. This linear combination of Krylov vectors can be written as,

$$\delta \mathbf{u}_j = \delta \mathbf{u}_0 + \sum_{i=0}^{j-1} \beta_i (\mathbf{J})^i \mathbf{r}_0, \tag{7}$$

where the scalars β_i minimize the residual.

Upon examining Eq. (7) we see that GMRES requires the action of the Jacobian only in the form of matrix-vector products, which may be approximated by [21, 39]:

$$\mathbf{Jv} \approx [\mathbf{F}(\mathbf{u} + \epsilon \mathbf{v}) - \mathbf{F}(\mathbf{u})] \;/\; \epsilon, \tag{8}$$

where ϵ is a small perturbation.

Equation (8) is simply a first-order Taylor series expansion approximation to the Jacobian, $\mathbf{J}$, times a vector, $\mathbf{v}$. For illustration consider the two coupled nonlinear equations $F_1(u_1, u_2) = 0$, $F_2(u_1, u_2) = 0$. The Jacobian matrix is

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial F_1}{\partial u_1} & \dfrac{\partial F_1}{\partial u_2} \\ \\ \dfrac{\partial F_2}{\partial u_1} & \dfrac{\partial F_2}{\partial u_2} \end{bmatrix}.$$

JFNK does not require the formation of this matrix; we instead form a result vector that approximates this matrix multiplied by a vector. Working backwards from Eq. (8), we have

$$\frac{\mathbf{F}(\mathbf{u} + \epsilon \mathbf{v}) - \mathbf{F}(\mathbf{u})}{\epsilon} = \begin{pmatrix} \frac{F_1(u_1+\epsilon v_1, u_2+\epsilon v_2) - F_1(u_1,u_2)}{\epsilon} \\ \\ \frac{F_2(u_1+\epsilon v_1, u_2+\epsilon v_2) - F_2(u_1,u_2)}{\epsilon} \end{pmatrix}.$$

Approximating $\mathbf{F}(\mathbf{u} + \epsilon \mathbf{v})$ with a first-order Taylor series expansion about $\mathbf{u}$, we have

$$\frac{\mathbf{F}(\mathbf{u} + \epsilon \mathbf{v}) - \mathbf{F}(\mathbf{u})}{\epsilon} \approx \begin{pmatrix} \frac{F_1(u_1,u_2) + \epsilon v_1 \frac{\partial F_1}{\partial u_1} + \epsilon v_2 \frac{\partial F_1}{\partial u_2} - F_1(u_1,u_2)}{\epsilon} \\ \\ \frac{F_2(u_1,u_2) + \epsilon v_1 \frac{\partial F_2}{\partial u_1} + \epsilon v_2 \frac{\partial F_2}{\partial u_2} - F_2(u_1,u_2)}{\epsilon} \end{pmatrix},$$

which simplifies:

$$\begin{pmatrix} v_1 \frac{\partial F_1}{\partial u_1} + v_2 \frac{\partial F_1}{\partial u_2} \\ \\ v_1 \frac{\partial F_2}{\partial u_1} + v_2 \frac{\partial F_2}{\partial u_2} \end{pmatrix} = \mathbf{Jv}.$$

The error in this approximation is proportional to $||\epsilon\mathbf{v}||$. This matrix-free approach has many advantages. The most attractive is Newton-like nonlinear convergence without the costs of *forming* or *storing* the true Jacobian. In practice one forms a matrix (or set of matrices) for preconditioning purposes, so we eschew the common description of this family of methods as fully "matrix-free." However, the matrices employed in preconditioning can be simpler than the true Jacobian of the problem, so the algorithm is properly said to be "Jacobian-free." We briefly discuss options for matrix-free (or nearly matrix-free) preconditioning in Section 3.5. A convergence theory has been developed for JFNK [20].

2.3.1 The Jacobian-Vector Product Approximation

As shown above, the Jacobian-vector product approximation is based on a Taylor series expansion. Here we discuss various options for choosing the perturbation parameter, ϵ in Eq. (8), which is obviously sensitive to scaling, given $\mathbf{u}$ and $\mathbf{v}$. If ϵ is too large, the derivative is poorly approximated and if it is too small the result of the finite difference is contaminated by floating-point roundoff error. The best ϵ to use for a scalar finite-difference of a single argument can be accurately optimized as a balance of these two quantifiable trade-offs. However, the choice of ϵ for a Fréchet finite difference is as much of an art as a science. A simple choice of ϵ is

$$\epsilon = \frac{1}{n||\mathbf{v}||_2} \sum_{i=1}^{n} b|u_i|, \tag{9}$$

where n is the linear system dimension and b is a constant whose magnitude is within a few orders of magnitude of the square root of machine roundoff (typically 10^{-6} for 64-bit double precision). Another more sophisticated approach proposed by Brown and Saad [21] is

$$\epsilon = \frac{b}{||\mathbf{v}||_2} \max[|\mathbf{u}^T\mathbf{v}|, typ\mathbf{u}|\mathbf{v}|] \operatorname{sign}(\mathbf{u}^T\mathbf{v}). \tag{10}$$

Here, *typ***u** is a user-supplied "typical size" of **u**.

Equation (8) is a first-order approximation. It is straightforward to construct a second-order approximation,

$$\mathbf{J}\mathbf{v} \approx [\mathbf{F}(\mathbf{u} + \epsilon\mathbf{v}) - \mathbf{F}(\mathbf{u} - \epsilon\mathbf{v})] \;/\; \epsilon. \tag{11}$$

A disadvantage of second order is the cost of two fresh function evaluations per matrix-vector multiply.

2.3.2 Inexact Newton Methods

Since the use of an iterative technique to solve Eq. (2) does not require the exact solution of the linear system, the resulting algorithm is categorized as an “inexact” Newton’s method [44]. A simple inexact method results in the following convergence criteria on each linear iteration.

$$\| \mathbf{J^k}\delta\mathbf{u^k} + \mathbf{F}(\mathbf{u^k}) \|_\mathbf{2} < \gamma \| \mathbf{F}(\mathbf{u^k}) \|_\mathbf{2}, \tag{12}$$

where γ is a constant smaller than unity. Keeping γ small, which is required for Newton-like nonlinear convergence, makes large demands upon the preconditioner, especially as the dimension of the linear system grows. There is a trade-off between the effort required to solve the linear system to a tight tolerance and the resulting required number of nonlinear iterations. Too large a value for γ results in less work for the Krylov method but more nonlinear iterations, whereas too small a value for γ results in more Krylov iterations per Newton iteration. The forcing function and the issue of “oversolving” a Newton step has gained recent interest [148, 159]. It has been demonstrated that in some situation the Newton connvergence may actually suffer if γ is too small in early Newton iterations. Examples of this trade-off between total nonlinear iterations and execution time are given in [109, 129]. Several strategies for optimizing the computational work with a variable “forcing term” γ are given in [49].

2.4 Globalization

The lack of convergence robustness of Newton’s method is frequently raised. In practice, globalization strategies leading from a convenient initial iterate into the ball of convergence of Newton’s method around the desired root are required. For problems arising from differential equations, there are many choices. The issue of globalization is more vexing for BVPs then IVPs, where accurately following the physical transient often guarantees a good initial guess. Based on the robustness of IVP solvers, BVPs are often approached through a false time-stepping.

2.4.1 Pseudo-transient Continuation

Pseudo-transient continuation solves the steady-state problem $\mathbf{F}(\mathbf{u}) = \mathbf{0}$, for which a solution is presumed to exist, through a series of problems

$$\mathbf{f}_\ell(\mathbf{u}) \equiv \frac{\mathbf{u} - \mathbf{u}^{\ell-\mathbf{1}}}{\tau^\ell} + \mathbf{F}(\mathbf{u}) = \mathbf{0}, \quad \ell = \mathbf{1}, \mathbf{2}, \ldots, \tag{13}$$

which are derived from a method-of-lines model

$$\frac{\partial \mathbf{u}}{\partial t} = -\mathbf{F}(\mathbf{u}),$$

each of which is solved (approximately) for $\mathbf{u}^\ell$. The physical transient is followed when the timestep τ^ℓ is sufficiently small, and the problems at each timestep are well solved, leading the iterations through a physically feasible sequence of states. Furthermore, the Jacobians associated with $\mathbf{f}_\ell(\mathbf{u}) = \mathbf{0}$ are well conditioned when τ^ℓ is small. See [52] for an analysis of this effect based on the spectrum of the preconditioned operator in the case of the constant coefficient heat equation.

τ^ℓ is advanced from $\tau^0 \ll 1$ to $\tau^\ell \to \infty$ as $\ell \to \infty$, so that $\mathbf{u}^\ell$ approaches the root of $\mathbf{F}(\mathbf{u}) = \mathbf{0}$. We emphasize that pseudo-transient continuation does *not* require reduction in $||\mathbf{F}(\mathbf{u}^\ell)||$ at each step, as do typical linesearch or trust region globalization strategies [45]; it can "climb hills."

Strict Newton iteration applied to (13) yields

$$\mathbf{u}^{\ell,\mathbf{k}} = \mathbf{u}^{\ell-\mathbf{1}} - (\mathbf{I} + \tau^\ell \mathbf{F}'(\mathbf{u}^{\ell,\mathbf{k}}))^{-\mathbf{1}}(\mathbf{u}^{\ell,\mathbf{k}} + \tau^\ell \mathbf{F}(\mathbf{u}^{\ell,\mathbf{k}}) - \mathbf{u}^{\ell-\mathbf{1}}), \ \ \mathbf{k} = \mathbf{0}, \mathbf{1}, \ldots . \tag{14}$$

If we take $\mathbf{u}^{\ell,\mathbf{0}} = \mathbf{u}^{\ell-\mathbf{1}}$ (the simplest initial iterate), then the first correction step is

$$\mathbf{u}^{\ell,\mathbf{1}} = \mathbf{u}^{\ell-\mathbf{1}} - (\frac{\mathbf{1}}{\tau^\ell}\mathbf{I} + \mathbf{F}'(\mathbf{u}^{\ell-\mathbf{1}}))^{-\mathbf{1}}\mathbf{F}(\mathbf{u}^{\ell-\mathbf{1}}). \tag{15}$$

In some problems, it may be required to iterate the Newton corrector (14) more than once [69] or until it converges ($\lim_{k\to\infty} \mathbf{u}^{\ell,\mathbf{k}} \equiv \mathbf{u}^\ell$), thus leading in the limit to following the transient implicitly. However, we generally prefer to advance in pseudo-time after just one Newton step (15).

A time-step selection scheme is required to complete the algorithm. One choice is successive evolution-relaxation (SER) [120], which lets the time step grow in inverse proportion to residual norm progress:

$$\tau^\ell = \tau^{\ell-1} \cdot \frac{||\mathbf{F}(\mathbf{u}^{\ell-\mathbf{2}})||}{||\mathbf{F}(\mathbf{u}^{\ell-\mathbf{1}})||}. \tag{16}$$

Alternatively, a temporal truncation error strategy bounds the maximum temporal truncation error in each individual component, based on a local estimate for the leading term of the the error. (The idea is not to control the error, *per se*, but to control the stepsize through its relationship to the error.) Another approach sets target maximum magnitudes for change in each component of the state vector and adjusts the time step so as to bring the change to the target. All such devices are "clipped" into a range about the current time step in practice. Typically, the time step is not allowed to more than double in a favorably converging situation, or to be reduced by more than an order of magnitude in an unfavorable one, unless feasibility is at stake, in which case the time step may be drastically cut [74].

The globalization theory of [74] employs a three-phase approach, whose phases in practice may or may not be cleanly demarcated in residual norm convergence plots. Initially, $||\mathbf{u}^\mathbf{0} - \mathbf{u}^*|| \gg \mathbf{1}$ and $\tau^0 \ll 1$. During an "induction phase" the solution is marched in a method-of-lines sense with relatively small time step until $||\mathbf{u} - \mathbf{u}^*||/||\mathbf{u}^\mathbf{0} - \mathbf{u}^*|| \ll \mathbf{1}$. Success of this phase is governed by stability and accuracy of the integration scheme (we simply use the backward Euler method) and by the choice of initial iterate.

For problems in which a complex feature, such as a shock or a flame front, must arise from a structure-free initial condition, the induction phase is typically by far the longest. In a grid-sequenced problem, in which the initial iterate on a given fine grid is interpolated from a converged solution on a coarser grid, and in which solution features are correctly located (if not fully resolved), the induction phase on the finest grid can be relatively brief [150]. During a second "transition phase" the time step is built up in the neighborhood of the solution. The critical assumption is existence of a β such that $\|(\mathbf{I}+\tau\mathbf{F}'(\mathbf{u}))^{-\mathbf{1}}\| \leq (\mathbf{1}+\beta\tau)^{-\mathbf{1}}$ for all $\tau \geq 0$ if $\|\mathbf{u}-\mathbf{u}^*\| \leq \epsilon$. Finally comes a "polishing phase," during which the the time steps approach infinity (or some user-imposed upper bound) and iterates approach the root with asymptotic Newton-like convergence. This phase is treated by a conventional local analysis, as in [73].

The main result of the theory is that there is either convergence from $\mathbf{u}^{\mathbf{0}}$ to $\mathbf{u}^*$ or an easily detectable (undesirable) contraction of τ^ℓ toward 0, allowing recovery actions before blow-up or floating point faults from infeasible steps. (Robust recovery is particularly important in parallel applications.) The main hypotheses of the theory, including smooth differentiability of $\mathbf{F}(\mathbf{u})$, are difficult to verify in practice. They are also rarely respected in practice, since instantaneous analytical approximations of $\mathbf{F}'(\mathbf{u})$ are too expensive in memory and execution time. The theory for pseudo-transient continuation has recently been extended to index-1 differential-algebraic equations [40], in which not all of the equations possess a time derivative term. This is relevant for systems of PDEs in which temporal evolution takes place on a manifold of constraints, such as incompressibility in Navier-Stokes.

2.4.2 Other Continuation Methods

Besides pseudo-transient continuation, there are two other important types of continuation in the literature of numerical solutions for nonlinear BVPs, namely, continuation in a physical parameter of the problem, and mesh sequencing, which is continuation in a discretization parameter — namely a scale for the mesh spacing.

Physical parameters often provide "knobs" by which the nonlinearity in a problem can be varied. An easily understood example from computational fluid dynamics is the Reynolds number, which directly multiplies the convective terms of Navier-Stokes, but there are many other examples including body forcings and boundary forcings. The solution of $\mathbf{F}(\mathbf{u},\pi^\ell)=\mathbf{0}$, where π^ℓ is such a parameter, can be implicitly defined as $\mathbf{u}(\pi^\ell)$.

We suppose that $\mathbf{F}(\mathbf{u},\pi^{\mathbf{0}})=\mathbf{0}$ is "easy" to solve; for instance, it may be linear in $\mathbf{u}$, as when π is a Reynolds number and the governing equations reduce to the Stokes subset. Given $\mathbf{u}^{\ell-\mathbf{1}}$ corresponding to $\pi^{\ell-1}$, we can posit a good initial guess for $\mathbf{u}^\ell$ at a nearby π^ℓ from the Taylor expansion

$$\mathbf{u}^{\ell,\mathbf{0}} = \mathbf{u}(\pi^{\ell-\mathbf{1}}) + \left(\frac{\partial \mathbf{u}}{\partial \pi}\right)^{\ell-\mathbf{1}} (\pi^\ell - \pi^{\ell-\mathbf{1}}) \ . \tag{17}$$

Implicitly differentiating $\mathbf{F}(\mathbf{u}, \pi) = \mathbf{0}$ with respect to π gives

$$\left(\frac{\partial \mathbf{F}}{\partial \mathbf{u}}\right)\left(\frac{\partial \mathbf{u}}{\partial \pi}\right) + \left(\frac{\partial \mathbf{F}}{\partial \pi}\right) = 0 \ , \tag{18}$$

or

$$\left(\frac{\partial \mathbf{u}}{\partial \pi}\right) = -\left(\frac{\partial \mathbf{F}}{\partial \mathbf{u}}\right)^{-1}\left(\frac{\partial \mathbf{F}}{\partial \pi}\right) \tag{19}$$

whence the right-hand side of (17) can be evaluated. This presumes that one is able to readily solve linear systems with the Jacobian, $\frac{\partial \mathbf{F}}{\partial \mathbf{u}}$; otherwise, poorer approximations are possible, including the simple "bootstrapping" procedure of using just $\mathbf{u}(\pi^{\ell-\mathbf{1}})$, itself, for $\mathbf{u}^{\ell,\mathbf{0}}$.

Mesh sequencing is useful when a nonlinear problem is easier to solve on a coarser grid than the one on which the solution is ultimately desired, either because the nonlinearity, itself, is milder or because the linear conditioning of the sequence of nonlinear correction problems is milder. An initial iterate for the next finer mesh is constructed by interpolation from the solution on the preceding coarser mesh. Asymptotically, under certain assumptions that are natural when the discretization ultimately becomes fine enough to accurate resolve the continuous statement of the BVP, it can be shown that the initial interpolant lies in the domain of convergence of Newton's method [149] on the finer grid. Unfortunately, it is usually not easy to determine when this asymptotic range is reached. Consequently, another continuation method, such as pseudo-transience, may be used to drive the initial interpolant towards the Newton domain on each mesh step. Such nested continuation methods are often required in practice on highly nonlinear problems, such as detailed kinetics combustion. Since a decreasing number of inner continuation steps are required on the finer meshes, the nested approach can be economical.

2.4.3 Other Robustification Techniques

Whatever the combination of continuation strategies that may be invoked to prepare for a full Newton iteration on the ultimate accurately discretized BVP, modified Newton-like systems need to be solved at each stage. Traditional physics-independent, discretization-independent algebraic robustification strategies can be employed on these systems and any code intended for general purpose by non-experts should default to some combination of the strategies of line search, trust region, and back-tracking, [45, 73] or the more crude, but often successful "damping on percentage change" [97, 170].

3 Preconditioning of the JFNK Method

The purpose of preconditioning the JFNK method is to reduce the number of GMRES (Krylov) iterations, as manifested (in the GMRES convergence theory; see [144]) by efficiently clustering eigenvalues of the iteration matrix. Traditionally, for linear problems, one chooses a few iterations of a simple iterative method (applied to the system matrix) as a preconditioner. A goal of the JFNK approach is to avoid forming the system matrix $\mathbf{J}$. As illustrated in the sequel, an effective preconditioner for JFNK can typically be simpler than the strict Jacobian of the system.

A linear preconditioner can be applied on the left (rescaling the matrix rows and the right-hand side) or on the right (rescaling the matrix columns and the solution vector), or on both, if suitably factored. Since left preconditioning changes the residual by which convergence to a linear iterative method is generally measured, right preconditioning is often preferred in comparing the intrinsic merit of different preconditioning strategies. However, in the Newton context, we often use left preconditioning, since the preconditioned residual serves as a useful estimate of the size of the Newton correction, itself, when the preconditioning is of high quality. Either strategy may be employed in a Jacobian-free context.

Using right preconditioning one solves

$$(\mathbf{J}\mathbf{P}^{-\mathbf{1}})(\mathbf{P}\delta\mathbf{u}) = -\mathbf{F}(\mathbf{u}). \tag{20}$$

$\mathbf{P}$ symbolically represents the preconditioning matrix (or process) and $\mathbf{P}^{-1}$ the inverse of preconditioning matrix. Right preconditioning is actually realized through a two step process. First solve

$$(\mathbf{J}\mathbf{P}^{-\mathbf{1}})\mathbf{w} = -\mathbf{F}(\mathbf{u}), \tag{21}$$

for $\mathbf{w}$. Then solve

$$\delta\mathbf{u} = \mathbf{P}^{-\mathbf{1}}\mathbf{w}, \tag{22}$$

for $\delta\mathbf{u}$. Thus, while we may refer to the matrix $\mathbf{P}$, operationally the algorithm only requires the action of $\mathbf{P}^{-1}$ on a vector. Note that if a distributed or segregated approach is used for preconditioning, then $\mathbf{P}^{-1}$ may be formed as a linear combination of approximate inverses of submatrices. An example is the additive Schwarz method of section 3.2. The right-preconditioned version of Eq. (8) is:

$$\mathbf{J}\mathbf{P}^{-1}\mathbf{v} \approx \left[\mathbf{F}(\mathbf{u} + \epsilon\mathbf{P}^{-1}\mathbf{v}) - \mathbf{F}(\mathbf{u})\right] / \epsilon. \tag{23}$$

This operation is done once per GMRES iteration, and is actually done in two steps:

1. Preconditioning: Solve (approximately) for $\mathbf{y}$ in $\mathbf{P}\mathbf{y} = \mathbf{v}$.
2. Perform matrix-free product $\mathbf{J}\mathbf{y} \approx [\mathbf{F}(\mathbf{u} + \epsilon\mathbf{y}) - \mathbf{F}(\mathbf{u})] \; / \; \epsilon$.

Only the matrix elements required for the action of $\mathbf{P}^{-1}$ are formed. There are two primary choices to be made:

1. What linearization should be used to form the matrices required in $\mathbf{P}^{-1}$?

   (A new decision facing the user of a Jacobian-free Newton-Krylov method)
2. What linear iterative method should be used for $\mathbf{y} = \mathbf{P}^{-1}\mathbf{v}$?

   (A standard decision facing the user of a Krylov method)

The following subsections focus on specific issues. In practice, many preconditioning approaches use a combination of these ideas.

3.1 Standard Approaches

In systems where forming the Jacobian is a dominant cost one may employ a "stale" (or frozen) Jacobian from an earlier step in the preconditioner while obtaining the action of the current Jacobian in the JFNK matrix-vector multiply [91, 92, 93]. This is referred to as "MFNK" in [92, 93]. This approach is not truly Jacobian-free since *some* true Jacobians are formed and stored. However this usually expensive task is not done every Newton iteration. This is different from a more traditional modified Newton-Krylov (MNK) method where the actual matrix approximating the local tangent hyperplane in Newton's method (not just its preconditioner) is held constant over several Newton iterations. The MNK approach has much weaker nonlinear convergence properties. The Jacobian-free method "feels" the true Jacobian (to within finite difference truncation error, or other less severe limitation) at each iteration.

In BVPs, incomplete lower-upper (ILU) factorizations have been frequently employed when approximately inverting Jacobian matracies in the preconditioner. For systems of equations characterized by tight intra-equation coupling, a blocked ILU factorization may be more effective than a standard "point" ILU factorization preconditioner [70, 111]. Here the degrees of freedom defined at a common point are interlaced and a full factorization (usually dense for systems typical of CFD or MHD, with a dozen or fewer independent fields) is performed amongst them. The overall factorization is incomplete above the block level, with fill-in limited between degrees of freedom defined at different points.

In systems of conservation laws in which convection dominates, high-order convection schemes are desired for accuracy. Using the Jacobian-free method, one can construct the preconditioner from a low-order upwinded discretization that is more stable with respect to incomplete factorization, saving memory and often resulting in more effective preconditioning [77, 84, 111]. Convergence of the nonlinear system occurs to the higher-order discretization represented in the right-hand side residual. Operationally, this split-discretization Jacobian-free preconditioned product is

$$\mathbf{J}\mathbf{P}^{-1}\mathbf{v} \approx \frac{\mathbf{F}_{high}(\mathbf{u} + \epsilon\mathbf{P}_{low}^{-1}\mathbf{v}) - \mathbf{F}_{high}(\mathbf{u})}{\epsilon}. \tag{24}$$

Here, $\mathbf{F}_{high}(\mathbf{u})$ denotes the nonlinear function evaluated with a high-order discretization, and $\mathbf{P}_{low}^{-1}$ denotes a preconditioning operator formed with a low-order discretization.

3.2 Newton-Krylov-Schwarz

Newton-Krylov-Schwarz (NKS) is a preconditioned Jacobian-free Newton-Krylov method in which the action of the preconditioner is composed from those of preconditioners defined on individual subdomains. Historically, the primary motivation for NKS (first called by this name in [27]) is parallel processing through divide-and-conquer. Scalability studies based on dimensionless ratios of communication and computation parameters for message-passing aspects of Schwarz-type iterative methods appeared in [78, 80]. Recently, a sequential motivation for Schwarz methods has become apparent [168]: their localized working sets can be sized to fit in the Level-2 caches of contemporary microprocessors. Furthermore, multilevel Schwarz, algebraically similar to the AFAC form of multigrid [108], can have bounded condition number in the asymptotic limit of finely resolved meshes, and is therefore an "optimal" method from the perspective of convergence rate.

If we decompose the domain of a PDE problem into a set of possibly overlapping subdomains Ω_i, the standard additive Schwarz preconditioner can be expressed as

$$\mathbf{P}_{ASM}^{-1} = \sum_i {R_i}^T \mathbf{J}_i^{-1} R_i, \tag{25}$$

where the three-phase solution process (reading operators from right to left) consists of first collecting data from the local and neighboring subdomains via global-to-local restriction operators R_i, then performing a local linear solve on each subdomain $\mathbf{J}_i^{-1}$, and finally sending partial solutions to the local and neighboring subdomains via the local-to-global prolongation operators ${R_i}^T$. The solve with the local Jacobian can be replaced with an approximate solve, such as a local incomplete factorization, or a multigrid sweep.

While the three phases are sequential and synchronized by communication requirements, each term in the sum can be computed concurrently, leading to parallelism proportional to the number of subdomains. This is in contrast with a global incomplete factorization, whose concurrency is determined by the discretization stencil and the matrix ordering and cannot be scaled to an arbitrary number of processors.

Parallel experience with NKS methods is growing. We mention the shared-memory implementation of [112] and the distributed-memory implementations of [26, 30, 72]. Domain-based parallelism is recognized as the form of data parallelism that most effectively exploits contemporary microprocessors with multi-level memory hierarchy [41, 168]. Schwarz-type domain decomposition methods have been extensively developed for finite difference/element/volume PDE discretizations over the past decade, as reported in the annual proceedings of the international conferences on domain decomposition methods, of which the most recent volume is [54].

In practice, we advocate the *restricted* additive Schwarz Method (RASM), which eliminates interprocess communication during either the restriction or prolongation phase of the additive Schwarz technique [32]. One version of the RASM preconditioner can be expressed in operator notation as

$$\mathbf{P}_{RASM}^{-1} = \sum_i {R'_i}^T \mathbf{J}_i^{-1} R_i. \tag{26}$$

It performs a complete restriction operation but does not use any communication during the interpolation phase, ${R'_i}^T$. This provides the obvious benefit of a 50% reduction in nearest-neighbor communication overhead. In addition, experimentally, it preconditions better than the original additive Schwarz method over a broad class of problems [32], for reasons that are beginning to be understood in the function space theory that underlies Schwarz methodology [23].

As originally introduced in [48], additive Schwarz preconditioning includes a coarse grid term in the sum (25). Indeed, the coarse grid is essential for optimal conditioning in the scalar elliptic case. Table 1 shows the successive improvements towards optimality of a hierarchy of methods, all of which fit within the additive Schwarz algebraic framework, Eq. (25). The most primitive is point Jacobi, in which each subdomain is one point and there is no overlap. Subdomain Jacobi clusters all the points in one subdomain into a single subdomain solve, which is performed concurrently within each subdomain, with no overlap. One-level additive Schwarz has the same concurrency as Jacobi, except that the subdomains overlap, and nontrivial communication is required to set up the subproblems. To achieve the mesh-independent estimates shown (iterations depending only upon the number of processors or subdomains), some operator-dependent, not very severe in practice, assumptions need to be made about the extent of overlap. Finally, the Schwarz preconditioner supplemented with a low-dimensional but global coarse grid problem (2-level) achieves independence of the number of subdomains, at the price of an increasingly complex problem linking the subdomains.

NKS methods have been developed and studied by Cai and collaborators [24, 25, 26, 30], Knoll and collaborators [93, 95, 111], Pernice and collaborators [130], and Tidriri [152, 154], among many others. The combination of pseudo-transient continuation and NKS has been called ΨNKS, and is discussed in [60].

Table 1: Iteration count scaling of Schwarz-preconditioned Krylov methods, translated from the theory in terms of mesh spacing and subdomain diameter into the corresponding quantities of discrete problem size N and processor number P, assuming quasi-uniform grid, quasi-unit aspect ratio grid and decomposition, and quasi-isotropic operator.

	Iteration Count	
Preconditioning	in 2D	in 3D
Point Jacobi	$\mathcal{O}(N^{1/2})$	$\mathcal{O}(N^{1/3})$
Subdomain Jacobi	$\mathcal{O}((NP)^{1/4})$	$\mathcal{O}((NP)^{1/6})$
1-level Additive Schwarz	$\mathcal{O}(P^{1/2})$	$\mathcal{O}(P^{1/3})$
2-level Additive Schwarz	$\mathcal{O}(1)$	$\mathcal{O}(1)$

3.3 Multigrid Approaches

There has been considerable success in applying the multigrid method as a preconditioner to Krylov methods on linear problems [4, 88, 123, 113]. As a result of this success, JFNK researchers have begun to consider the performance of linear multigrid as a preconditioner to a Jacobian-free Newton-Krylov method [22, 34, 71, 96, 98, 128, 139]. In the multigrid preconditioned Newton-Krylov method (NKMG), the system $\mathbf{y} = \mathbf{P}^{-1}\mathbf{v}$, in Eq. (23), is approximately solved for $\mathbf{y}$ using a linear multigrid algorithm.

Whereas the primary motivation for Schwarz-type preconditioning is concurrency, the primary motivation in NKMG is optimal operation complexity. By this we mean a preconditioner that not only renders the number of preconditioned Krylov iterations per Newton iteration independent of grid resolution, but imposes a cost per iteration that grows only linearly in the number of discrete unknowns. NKMG has also been quite effectively implemented in parallel [22, 128].

The basic building blocks of a multigrid algorithm are the mesh interpolation operators, restriction $\mathcal{R}$ and prolongation $\mathcal{P}$, and a method of constructing the coarse grid operators. In the limit (for nested multilevel methods), the "perfect" coarse grid operator is the Schur complement of the fine grid operator with the degrees of freedom not represented on the coarse grid eliminated. The corresponding perfect restriction is the elimination step, and the corresponding prolongation the backsolve. In practice, one uses much less expensive grid transfer operators, such as multilinear interpolation, or even simple injection restriction and piecewise constant prolongation. In [4, 123] it is shown on some challenging problems that multigrid as a preconditioner may outperform multigrid as a solver, and in general it is also more robust, due to the outer Krylov method. In [88] it is shown that a suboptimal multigrid method, which is not a scalable solver as a result of overly simplified restriction and prolongation operators, can produce a scalable method when used as a preconditioner.

In [98] the restriction and prolongation operators are piecewise constant and piecewise linear. Since the systems considered there contain second-order operators, the choice of $\mathcal{R}$ and $\mathcal{P}$ as piecewise constant violates the level transfer "order rule", $m_{\mathcal{P}} + m_{\mathcal{R}} > 2$ [169]. Here $m_{\mathcal{P}}$ and $m_{\mathcal{R}}$ are the order of interpolation plus one for the prolongation and restriction operators, respectively. Thus, this approach can not be considered an optimal multigrid method. In [98] it is demonstrated that multigrid methods make excellent preconditioners for JFNK, superior to comparably complex SGS and ILU, with a typical approximate inverse being only one V-cycle. It is also demonstrated that the algorithmic simplifications which may result in loss of convergence for multigrid as a solver (such as piecewise constant prolongation in place of piecewise linear prolongation) have a much weaker effect when multigrid is the preconditioner.

In [96] two different approaches for defining the coarse grid approximations to the preconditioner are considered. The first approach is to restrict the dependent variables ($\mathbf{u}$) down through a series of grids, re-discretize the equations $\mathbf{F}(\mathbf{u})$, and then form each of the preconditioner elements independently. This may be troublesome for multi-scale nonlinear physics and/or nonlinear discretizations. The second method is to build the coarse grid operators using an additive correction [67] procedure, which can also be viewed as a Galerkin, or variational, approach [169]. Here, the coarse grid operator, $\mathbf{P}_c$ is constructed from the fine grid operator, $\mathbf{P}_f$ as:

$$\mathbf{P}_c = \mathcal{R} * \mathbf{P}_f * \mathcal{P}. \tag{27}$$

When $\mathcal{R}$ and $\mathcal{P}$ are piecewise constant, this should be viewed as an additive correction multigrid method [67, 146]. This is attractive for complex physics codes and/or unstructured grids since no discretizations on the coarse grids are required. Details of this specific multigrid approach, used as a preconditioner to GMRES, on a scalar problem can be found in [88]. In [96] both approaches of forming the coarse grid operators were found to give good algorithmic performance.

Recently, Mavripilis [106] has considered nonlinear multigrid as a preconditioner to JFNK with encouraging results.

3.4 Physics-based preconditioning

An important new class of preconditioners for the Jacobian-free Newton-Krylov method is referred to as physics-based or PDE-based. The motivation behind this approach is that there exist numerous, legacy algorithms to solve nonlinear systems, both IVPs and BVPs. These algorithms typically were developed with some insight into the time scales or physical behavior of the problem. As a benefit of this insight, a reduced implicit system, or a sequence of segregated explicit or implicit systems may be solved in place of the fully coupled system.

Examples include the semi-implicit method for low-speed flow [64], the SIMPLE algorithm for incompressible flow [126], Gummel's method for the semiconductor drift-diffusion equations [61], and numerous other structure-based operator splitting methods for reaction-diffusion systems [16, 124].

The SIMPLE algorithm [126] is a classical segregated solution method in computational fluid dynamics. Its use as a preconditioner to a Jacobian-free Newton-Krylov method is demonstrated in [128]. One employs multiple iterations of the SIMPLE algorithm in "delta (incremental) form" as the preconditioner. By "delta form" we mean that the linear system $\mathbf{Pu} = \mathbf{b}$ is solved for $\delta\mathbf{u}$, i.e., $\mathbf{P}\delta\mathbf{u} = \mathbf{b} - \mathbf{Pu}_0$ ($\mathbf{y} = \mathbf{P}^{-1}\mathbf{v}$ in Eq. 23) and $\mathbf{u} = \mathbf{u}_0 + \delta\mathbf{u}$. It is necessary to cast the preconditioner in "delta form" since this is the form of the problem upon which the outer Newton-Krylov iteration operates. Split, or segregated, methods are employed as preconditioners for Newton-Krylov on a system of time dependent reaction diffusion equations [116], time-dependent MHD equations [37], and steady state incompressible Navier-Stokes equations [96, 128], and time-dependent incompressible Navier-Stokes equations [101, 128]. Also in [101], a standard approximate linearization method used for phase-change heat conduction problems, is employed as a preconditioner for a JFNK solution of phase-change heat conduction problems. In this subsection we present detail on constructing preconditioners for stiff-wave systems using the semi-implicit method and constructing preconditioners using stucture-based operator splitting.

3.4.1 Stiff wave systems

To demonstrate how to construct a physics-based preconditioner for a stiff wave system, consider the 1D shallow water wave equations with a stiff gravity wave (a hyperbolic system):

$$\frac{\partial h}{\partial t} + \frac{\partial uh}{\partial x} = 0, \tag{28}$$

$$\frac{\partial uh}{\partial t} + \frac{\partial u^2 h}{\partial x} = -gh\frac{\partial h}{\partial x}. \tag{29}$$

Here u is the fluid velocity, h is the hydrostatic pressure, x is the spatial coordinate, t is time, g is gravity, and $\sqrt{gh}$ is the fast wave speed. The fast wave time scale is the time scale we which to "step over" in evolving the mesoscale dynamics of interest. A semi-implicit method is constructed by linearizing and implicitly discretizing only those terms which contribute to the stiff gravity wave. Thus, some physics insight is required to produce the implicit system. With $n+1$ as new time and n as old time, and suppressing spatial discretization, we have

$$\frac{h^{n+1} - h^n}{\Delta t} + \frac{\partial (uh)^{n+1}}{\partial x} = 0 \tag{30}$$

$$\frac{(uh)^{n+1} - (uh)^n}{\Delta t} + \frac{\partial (u^2 h)^n}{\partial x} + g h^n \frac{\partial h^{n+1}}{\partial x} = 0. \tag{31}$$

Note that the nonlinear term in h is linearized by evaluating the square of the wave speed (gh) at old time. We evaluate $\frac{\partial (u^2 h)}{\partial x}$ at old time since it does not contribute to the linearized gravity wave.

We rearrange the momentum equation as;

$$(uh)^{n+1} = -\Delta t g h^n \frac{\partial h^{n+1}}{\partial x} + S^n, \quad (S^n = (uh)^n - \Delta t \frac{\partial (u^2 h)^n}{\partial x}). \tag{32}$$

Equation (32) is then substituted into Eq. (30) to give the following scalar parabolic equation.

$$\frac{h^{n+1} - h^n}{\Delta t} - \frac{\partial}{\partial x}\left(\Delta t g h^n \frac{\partial (h)^{n+1}}{\partial x}\right) = \frac{\partial S^n}{\partial x} \tag{33}$$

Equation (33) is solved for h^{n+1}, and then one can easily solve for $(uh)^{n+1}$ using Eq. (32). For this simple problem the source of the nonlinear inconsistency is the linearized wave speed (a time discretization error) and the fact that advection in Eq. (31) is at a different time level. This is an issue when advection and wave propagation happen on the same time scale. The innovation of physics-based preconditioning is realizing that the nonlinear inconsistent solution of Eqs. (32) and (33) can be used as the preconditioner to a nonlinearly consistent Newton-Krylov solution of Eqs. (28)–(29). This is possible since JFNK does not require the formation of the Jacobian, and thus time-splitting approaches, such as the semi-implicit method, can be used as preconditioners. Since the action of the true operator is maintained in the evaluation of the full nonlinear residual and the forward Jacobian used in the Newton-Krylov iteration, the inverse Jacobian used in the preconditioner can be further weakened without compromise to the solution in the interest of minimizing execution time. For instance, a few cycles of a multigrid method, which is ideal for diffusion problems, can be used to approximate the solution of Eq. (33) in the preconditioner.

To be more specific, the function of a preconditioner is to map $[res_h, res_{uh}]$, or "$\mathbf{y}$", to $[\delta h, \delta uh]$, or "$\mathbf{v}$." Using the semi-implicit method in delta form (suppressing spatial discretization) the linearized equations are:

$$\frac{\delta h}{\Delta t} + \frac{\partial \delta uh}{\partial x} = -res_h, \tag{34}$$

$$\frac{\delta uh}{\Delta t} + g h^n \frac{\partial \delta h}{\partial x} = -res_{uh}. \tag{35}$$

Substituting Eq. (35) into Eq. (34), and eliminating δuh, produces

$$\frac{\delta h}{\Delta t} + \frac{\partial}{\partial x}(\Delta t \, g h^n \frac{\partial \delta h}{\partial x}) = -res_h + \frac{\partial}{\partial x}(\Delta t \; res_{uh}). \tag{36}$$

This parabolic equation can be approximately solved for δh. Then δuh can be evaluated:

$$\delta uh = -\Delta t \, gh^n \frac{\partial \delta h}{\partial x} - res_{uh}. \tag{37}$$

To summarize, we use a classical semi-implicit method, to map (res_h, res_{uh}) to $(\delta h, \delta uh)$ with one approximate parabolic solve. The utility of this preconditioning approach is verified on the 2D shallow water equations including the Coriolis force [115]. In addition, this framework has been used to develop preconditioners for MHD problems [37] and the compressible Euler equations [137].

3.4.2 Structure-based Preconditioning

Traditional techniques based on operator structure, which may be only marginally appealing as solvers or stationary operator splittings, may be effective and efficient preconditioners. Two structure-based techniques of particular interest are direction-based splitting and phenomenon-based splitting. To illustrate, consider a canonical system of unsteady convection-diffusion-reaction problems, symbolized by

$$\frac{\partial \mathbf{u}}{\partial t} + \mathcal{R}(\mathbf{u}) + \mathcal{S}(\mathbf{u}) = 0 \ .$$

$\mathbf{u}$ is a discrete gridfunction of p components per point.

The operator $\mathcal{R}$, representing convection and diffusion, typically has discretization stencils that strongly couple near neighbors of the same gridfunction component but typically only weakly couple different components, except through convective velocities and perhaps through composition-dependent constitutive laws. $\mathcal{S}$, representing reaction, may strongly couple different components, but typically involves only the unknowns defined at a single gridpoint. The remaining, transient term is typically diagonal (or narrow banded in an unlumped finite element discretization).

The combination of the transient term and $\mathcal{S}$ is well preconditioned with a block-diagonal operator, with no spatial coupling. The combination of the transient term with $\mathcal{R}$ is well preconditioned with independent multigrid solves for each component, with no intercomponent coupling. Each of these two preconditionings is substantially less computationally complex than a block ILU preconditioning for the entire discrete operator, and it is natural to consider multiplicative or additive forms of operator splitting in which each is applied independently.

Consider a simple backward difference for the transient term in advancing through timestep δt from $\mathbf{u}$ to $\mathbf{u} + \delta\mathbf{u}$. Discretized implicitly at the advanced level and linearized about $\mathbf{u}$, we have

$$\frac{\delta \mathbf{u}}{\delta t} + \mathbf{S}(\mathbf{u})(\mathbf{u} + \delta\mathbf{u}) + \mathbf{R}(\mathbf{u})(\mathbf{u} + \delta\mathbf{u}) = 0,$$

or, in delta-form,

$$\frac{\delta \mathbf{u}}{\delta t} + \mathbf{S}(\mathbf{u})\delta\mathbf{u} + \mathbf{R}(\mathbf{u})\delta\mathbf{u} = -\mathbf{F}(\mathbf{u}).$$

where $\mathbf{F}(\mathbf{u}) \equiv \mathbf{S}(\mathbf{u})\mathbf{u} + \mathbf{R}(\mathbf{u})\mathbf{u} = 0$.

Let the linear system to be solved by Krylov iteration at a given Newton iteration be written $\mathbf{J}\delta\mathbf{u} = -\mathbf{F}$, where $\mathbf{J} = \alpha\mathbf{I} + \mathbf{S} + \mathbf{R}$, and α is the reciprocal of the timestep. Apply an operator-split preconditioner $\mathbf{P}$ to Krylov vector $\mathbf{v}$ with preconditioned output $\mathbf{v}'$ as follows:

- Phase 1, block-diagonal (reaction) coupling

$$\mathbf{v}^* \leftarrow (\alpha\mathbf{I} + \mathbf{S})^{-1}\mathbf{v}.$$

- Phase 2, segregated scalar (spatial) coupling

$$\mathbf{v}' \leftarrow (\alpha\mathbf{I} + \mathbf{R})^{-1} \cdot \alpha \cdot \mathbf{v}^*.$$

The left-preconditioned Jacobian-vector product, $\mathbf{PJ}$, has the form

$$\mathbf{I} - (\alpha\mathbf{I} + \mathbf{R})^{-1}\mathbf{R} + (\alpha\mathbf{I} + \mathbf{R})^{-1} \cdot \alpha \cdot (\alpha\mathbf{I} + \mathbf{S})^{-1}\mathbf{R},$$

which is equivalent to approximating $\mathbf{J}$ with an operator that has first-order temporal splitting error, namely with $\alpha\mathbf{I} + \mathbf{S} + \mathbf{R} + \alpha^{-1}\mathbf{SR}$. This differs from the unsplit original $\mathbf{J}$ only in the last term. When the time step is small, so is this difference.

Alternatively, one can apply $\mathbf{P}$ to vector $\mathbf{v}$ with preconditioned output $\mathbf{v}'$ as follows:

- Phase 1, block-diagonal (reaction) coupling

$$\mathbf{v}^* \leftarrow (\alpha\mathbf{I} + \mathbf{S})^{-1}\mathbf{v}.$$

- Phase 2, segregated scalar (spatial) coupling

$$\mathbf{v}' \leftarrow \mathbf{v}^* + (\alpha\mathbf{I} + \mathbf{R})^{-1}(\mathbf{v} - J\mathbf{v}^*).$$

An extra Jacobian-vector product is required to update the residual in prior to Phase 2. Surprisingly, this extra Jacobian-vector product buys nothing; this method is mathematically identical to the first method, as is readily verified by algebraic manipulation.

Why one would wish to consider this type of preconditioner is apparent from a back-of-the-envelope complexity analysis. Besides p, the number of components (fields) per storage location, consider n, the number of data storage locations (vertices or cells) in the grid, and m, the number of neighbors in a typical stencil. Under a components-first ordering, the Jacobian consists of n

blockrows of width p, with m nonzero $p \times p$ blocks per row. The discrete dimension of $\mathbf{u}$ is pn. The number of nonzeros in the Jacobian (natively, or after block ILU factorization with no fill) is approximately p^2nm. The number of nonzeros in the spatial operator alone, however, is pnm, and the number in the reaction operator alone is np^2. Hence, the sum of the storage required for the discrete S and R operators is only $pn(m+p)$. The ratio of the storage required by the full Jacobian to the sum of the component operators is therefore $mp/(m+p)$. The same ratio applies to memory traffic per iteration involving the full Jacobian, or its components separately, and to the floating point operation count of the preconditioner application. For unstructured tetrahedral grids, $m \approx 15$. For combustion applications, p can be 10 to 100 or more. The savings in storing and applying the preconditioner can easily range up to an order of magnitude. Moreover, the best multilevel preconditioners for the p systems represented in the S operator may have superior properties — in convergence rate and in cache locality — for the segregated systems than for the combined. In Fig. 1 below, the two matrices on the right would be used in place of the full Jacobian in the upper left.

Structure-based operator-split preconditioning has been employed in radiation transport [116, 22], charge transport in semiconductors [8, 75], and fluid flow [43, 128, 101].

Due to architectural factors in high-end computers, the operator-split preconditioners discussed herein — and perhaps several other varieties — are natural to try, to replace block preconditioners that have heavier storage and memory traffic costs. Where operator-splitting is already used as a solver, it can easily be converted into a preconditioner by putting it into δ-form and wrapping a matrix-free Newton-Krylov acceleration around it. A small number of Newton steps (two or three) cleans up the splitting error. In cases in which there are strong couplings across different components stored at different mesh points, the type of phenomenon-based operator splitting described above is not expected to be successful; hence it is probably best exploited in an adaptive way in a polyalgorithmic preconditioner.

The salient point of this subsection is that a JFNK "wrapper" can provide nonlinear consistency to an inner operator-split solver. This is true even if the operator split preconditioner does not allow one to use the large time step size that is otherwise achievable by the outer JFNK solver.

Structured grids are often still used in practice, being natural for many problems in regular geometries such as rectangular parallelipipeds or spherical annuli. In such contexts, directional splitting, of which the Alternating Direction Implicit method (or "ADI") is paradigmatic, remains a popular solver. The operator algebra is identical to that above, except that $\mathbf{R}$ may represent, for instance, x-directional splitting and $\mathbf{S}$ y-directional. The complexity advantages of ADI are obvious, application of each inverse reduces to set of independent one-dimensional (e.g., block tridiagonal) problems. For all of the same qualitative reasons as for physics-based methods, and with only slight qualitative differences, such direction-based operator splittings may make excellent preconditioners, at least in serial computing environments. An outer Newton-Krylov iteration on each time step should quickly remove the splitting error.

Though the results of this section are developed for first-order implicit time-discretization, the real benefits of JFNK come in its ability to make higher-order implicit time discretizations worthwhile. Operationally, the only changes are in the presence of linear combinations of earlier time values of $\mathbf{u}$ on the right-hand side.

The work of Dawson et. al. [43] on two-phase subsurface flow and that of Kerkhoven [75] on the drift-diffusion semiconductor equations are excellent examples of the use of split-based preconditioners. In [75], the outer Newton-Krylov method is regarded as the accelerator to the inner fixed point method. Understanding and quantifying the limitations of split-based preconditioners is an active area of research.

3.5 Matrix-free Preconditioning Approaches

As shown in the previous subsections there are numerous ways to take advantage of the matrix-free matrix-vector multiply while still forming matrices that are reduced in complexity as compared to the full Jacobian. In the "physics-based" example (section 3.4), the approximate preconditioning matrix is derived from a scalar parabolic problem, while the Jacobian matrix is derived from a three-component hyperbolic system. In the "structure-based" paradigm (section 3.5), several independent systems replace a coupled system of greater overall complexity.

However, preconditioner matrix objects are still formed. Storage and memory bandwidth limitations always provide a motive to investigate preconditioning approaches that do not require the formation of any matrix.

There is a continuum of choices ranging from forming no preconditioner to forming the complete Jacobian. In this subsection, we briefly outline a few ideas that lie closest to true matrix-free preconditioning. The only iterative method that can be implemented in a fashion that is literally matrix-free is a Krylov method. Since Krylov methods may present a different matrix polynomial approximation to the matrix inverse for every initial iterate, they have not always enjoyed a reputation as preconditioners. It is now well known, however, how to employ one Krylov method to precondition an outer Krylov method with $\mathbf{P}^{-1}$ "changing" on each outer Krylov iteration. The price is generally some extra dense vector storage, which must be traded against the cost of a sparse matrix. The Flexible GMRES [142] (FGMRES) and GMRES-R [47] methods were developed to address the issue of using a preconditioner (Krylov or otherwise) that may vary within the GMRES iteration.

These flexible accelerators open up a number of preconditioning options such as using a relaxation method preconditioner with variable termination from outer iteration to outer iteration. On some convection-diffusion problems in

[142], FGMRES with GMRES as a preconditioner (a fully matrix-free option) outperformed GMRES with and ILU preconditioner. In the 2D shallow water problem just discussed the preconditioning matrix is symmetric and positive definite, thus we could use conjugate gradient as the preconditioner iterative method and be truly matrix-free. However, in this case, we should use FGMRES on the outside. In the JFNK context, however, where the Krylov solve is often performed inexactly throughout all but the endgame of the Newton cycle, plain GMRES is surprisingly forgiving of mildly inconsistent preconditioning.

The next step up in matrix formation is to implement a relaxation method in such a manner that only the main diagonal (or main block diagonal) need be formed. This is done in [39, 135]. In [39] a nonlinear SOR method is used as a preconditioner, requiring only the diagonal. Another step up is represented by [132] where an approximate factorization of ADI type is implemented. There storage is retained for only one of the required block inverses. This matrix is repopulated and inverted several times to approximate the preconditioner inverse. In both [132, 135] it is demonstrated that the price paid for this reduced storage method in the preconditioner is an increase in the execution time, as compared to the matrix counterparts.

As is justifiably touted in the multigrid community, the FAS algorithm represents a low storage multigrid algorithm. No global matrices need be formed for simple point smoothers. For block Jacobi smoothers, storage is only required for a local block diagonal. Thus, FAS can be viewed as a matrix-free preconditioner, as in [106].

Finally, we mention an idea that is often exploited in the context of problems that are dominantly elliptic. There exists a technology of "fast Poisson solvers" based on Fast Fourier Transforms (FFTs) or other fast transformations. The FFT and the multidimensional fast Poisson solvers (accommodating some spatially uniform, but directionally varying diffusion coefficients) that can be assembled from it require workspace equal to just a few gridfunction vectors and operation count only a log-factor higher than linear in the size of the gridfunction vector. Such FFT-based preconditioners, defined simply by subroutine calls on vectors, with no explicit matrix formation or storage whatsoever, may be highly effective in linear problems, or in nonlinear problems solved by JFNK in which the nonlinearity is a relatively controllable perturbation of one of the elliptic operators for which a fast inversion is known.

4 Applications

The focus of this section is to survey uses of JFNK in various fields of computational physics. We provide some references to the use of more standard Newton-based methods, as well. The subsequent section is an enumeration of "tricks" and techniques that span individual applications, illustrated in some of the work grouped by application domain here.

4.1 Fluid Dynamics / Aerodynamics

Computational fluid dynamics has been a rich area for algorithmic development, testing, and application, including nonlinear multilevel methods. In this section we can only sample the diverse literature to JFNK and computational fluid dynamics. The majority of this work has been on steady-state BVPs.

Vanka [162] was an early advocate of Newton's method for incompressible fluid flow, as was MacArthur [104]. Early explorations of Newton's method in compressible flow can be traced back Venkatakrishnan [163, 164]. This work is representative of the finite volume / finite difference CFD community. There has also been extensive development of nonlinearly implicit algorithms within the finite element community [56].

The incompressible Navier-Stokes equations have been used as a testbed for much JFNK algorithm development and testing, with emphasis on standard test problems, such as the driven cavity [55] and the natural convection cavity [42]. In fact, the driven cavity problem was considered in one of the original JFNK papers [21]. McHugh and co-workers studied inexact Newton methods and mesh sequencing, the performance of various Krylov methods within JFNK, the use of low-order spatial differencing within the preconditioner, as well as block ILU compared to point ILU preconditioning [70, 109, 110] The work of Shadid et. al. [148] is not JFNK, but NK. It has elucidated important issues related to inexact Newton methods, oversolving, and backtracking in a CFD context. Knoll et al. [96, 98, 101, 118] studied the ideas of multigrid preconditioning in JFNK in the context of the incompressible Navier-Stokes equations. They are also among the first to consider operator-split based preconditioning. Pernice *et al.* [127, 128] studied hybrid combinations of nonlinear multigrid, the SIMPLE algorithm, and JFNK. Most notably, the work in [128] applied JFNK, with a SIMPLE / multigrid preconditioner, to a 512^3-cell thermally driven flow problem on up to 512 processors.

JFNK methods have been developed and applied to the compressible Euler and Navier-Stokes equations primarily by the aerodynamics community. Newton-Krylov-Schwarz (NKS) development occupied ICASE and Boeing in the mid-1990s [30, 26, 77, 82, 153, 155]. The combined impact of parallelization, pseudo-transient continuation, and NKS is documented in [59]. Nearly matrix-free preconditioning techniques have been developed for the compressible Navier-Stokes equations [132]. Mavriplis studied the use of agglomeration multigrid as a preconditioner to JFNK on unstructured grids [106]. Other JFNK applications in compressible flow include [69, 133].

4.2 Plasma Physics

Problems in plasma physics provide a rich variety of time scales and nonlinearities. These result from the free electrons in an ionized gas (plasma) and the ability of the plasma to support and propagate electrostatic and electromagnetic waves. JFNK methods have made a significant impact in computational plasma physics within the past decade, and a number of plasma physics studies have been enabled by JFNK [17, 38, 85, 86, 94, 102, 131, 171, 173]. We briefly discuss the work in three separate areas of plasma physics.

4.2.1 Edge Plasma

The edge plasma (scrape-off, boundary layer) of a tokamak fusion experiment is that region of plasma which lies between the last closed flux surface and the vessel wall. This set of equations describes partially ionized flow with nonlinear and anisotropic heat conduction, thermal nonequilibrium (two temperatures), and finite rate ionization and recombination (i.e., chemistry).

The earliest use of JFNK in computational plasma physics was on the tokamak edge plasma fluid equations [140]. For the original use of Newton's method on the edge plasma equations see [83, 89, 90, 97, 167], with other Newton method applications following soon after [166, 174]. The JFNK method has become the mainstay of edge plasma simulation within the U.S. fusion community, as embodied in the UEDGE code [140].

JFNK applications in this area have utilized the numerical formation of a Jacobian and standard ILU factorization to perform the preconditioning process. In a set of papers [91, 92] on two different edge plasma physics models it is demonstrated that the use of a stale Jacobian in the preconditioning process provides significant CPU savings. It is also demonstrated the a pseudo-transient approach (section 2.4.1) provides a good globalization approach to this challenging boundary value problem

In [84] a higher-order, nonlinear, convection scheme is applied to the edge plasma fluid equations. It is demonstrated the forming the preconditioner from a low-order spatial discretization provided a simultaneous savings in memory requirements and CPU time, as compared to using the higher-order method in both the preconditioner and the residual evaluation. In [95] some initial investigation is done on the application of 1-level Schwarz preconditioning to the edge plasma equations. In [141] Newton-Krylov-Schwarz methods are applied to the edge plasma equation for parallelism. Also in [141], the superiority of JFNK over an operator split approach is demonstrated.

4.2.2 Fokker-Planck

The Fokker-Planck equation is used to model semi-collisional transport of charged particles in phase space. There is a wide variety of applications of the Fokker-Planck approach in plasma physics. JFNK is applied to Fokker-Planck-based models of the tokamak edge plasma [117] and of inertial electrostatic confinement (IEC) devices [33, 34, 35, 38]. The major challenge of the Fokker-Planck model is that it is nonlinear integro-differential. Thus a standard implementation of Newton's method results in a dense Jacobian matrix. The storage requirements of such a problem limit its use, although it is well understood that such an implicit implementation has other significant advantages [51].

In [117] the Landau form of the Fokker-Planck equation is solved where the integral effect arises through coefficients that are integral functions of the distribution function (1D in configuration space and 1D in phase space). The preconditioner is formed by lagging the integral coupling while maintaining this coupling in the residual evaluation. A standard ILU method is applied to the resulting sparse preconditioning matrix. This preconditioner is frequently used by others as a solver. However, as the collision frequency increases, that fastest time scale in the problem is being evaluated at the previous iteration. It is clearly demonstrated in [117] that as collision frequency increased, the JFNK method significantly outperformed the method which lagged the integral coupling.

In [33, 34] the Fokker-Planck equation is solved in Rosenbluth form in 2D phase space. Here, two subsidiary potential (elliptic) equations must be solved. This is done inside the residual evaluations using a multigrid-GMRES method. The preconditioning matrix for JFNK (2D, 9-point) is evaluated with the coefficients lagged and the preconditioning matrix is approximately inverted using simple multigrid ideas [88]. This approach is shown to have superior accuracy and scale efficiently to fine grids.

4.2.3 MHD

The equations of magnetohydrodynamics (MHD) represent a combination of the Navier-Stokes equations and Maxwell's equations without displacement current, from which the speed of light has been removed. The MHD equations are typically used to simulate plasma phenomena on the ion time scale, with the fast electron time scales removed. By retaining the Hall term in the generalized Ohm's law (so-called Hall MHD) some electron physics is retained. This adds an additional fast wave to the system, the Whistler wave. As compared to the Navier-Stokes equations, the MHD equations are more complicated since they support a family of waves that propagate at different speeds in an anisotropic manner. In [37] JFNK is applied to a 2D incompressible MHD equation system (a three-equation system). A semi-implicit, physics-based preconditioner is developed and the resulting scalar elliptic problems are approximately solved with simple multigrid methods (section 3.3). The resulting nonlinearly consistent algorithm is demonstrated to have a second-order accurate time step. It is shown that this algorithm can efficiently step over stiff wave time step constraints while maintaining a second-order accurate time step. Excellent time step and grid refinement scaling is demonstrated, as well as significant CPU gains as compared to an explicit simulation.

In [36] JFNK is applied to a 2.5-D incompressible Hall MHD equation system (a five-equation system). The parabolic operator that arises from the semi-implicit treatment of the Whistler wave is of fourth order in space. The conjugate gradient method is used to approximately invert this system in the preconditioner.

4.3 Reactive flows and flows with phase change

Reactive flows and flows with phase change are examples of nonlinear multiple time scale problems. There are many engineering and industrial application that are simulated by such systems. Operator splitting has been used extensively to simulate reactive flows [124]. Early Newton method research as applied to reactive flows may be found in [76, 149, 151].

In [93, 111] the performance of the JFNK method is studied in low Mach number compressible combustion. The base model is a BVP and represents a laminar diffusion flame. In [93] the SER method is employed for pseudo-transient continuation with standard ILU type preconditioners (similar to [92]). In [111] (within a Schwarz context) standard ILU preconditioners are compared to block ILU preconditioners, where the blocksize follows the number of conservation equations within a control volume. Block ILU precondtioning is shown to be superior in memory and in terms of preconditioner performance. The edge plasma Navier-Stokes neutrals model discussed in section 4.2.1 and in [84, 92] also contains finite rate "chemistry" in table look-up form. This model has mass, momentum, and energy exchange between "phases" as a result of the finite rate "chemistry". A particularly challenging version of this problem results from including both molecules and atoms in the neutral fluid [102]. The addition of molecules brings in fast (stiff) reactions. Here block ILU is essential to effective preconditioning. The standard ILU machinery can be thought of as attacking the intra-equation coupling, while the blocking attacks the inter-equation coupling.

Another recent example of the application of JFNK methods to reactive flows is the work of Mukadi and Hayes [119]. Their application is the transient simulation of an automotive catalytic convertor. In this study the effects of spatial discretization on preconditioner performance are considered.

Shadid [148] uses standard inexact Newton-Krylov methodology (not JFNK) to simulate reactive flows in 3D, on unstructured grids, and on massively parallel architectures. Simulations with several chemical species and reactions have been performed [147].

JFNK methods are applied to phase change heat conduction problems [87, 101]. This is done for pure material (isothermal phase change) alloys, and pure materials with multiple phase transitions. The key to implementing the JFNK methodology on this class of problems is the use of the enthalpy formulation [87] (enthalpy as the dependent energy variable). Here, temperature is a function of enthalphy, the phase diagram, and a local scale model for alloys. In the initial study of this problem the phase-change physics is ignored in the preconditioner and the preconditioning operator is formed only from heat conduction. In a subsequent study, the "effective heat capacity" method [101] (a linearized solver for this problem) is used as a preconditioner. This is a approximate route to bringing the phase change physics into the preconditioner. This is shown to provide a factor of five reduction in GMRES iterations and a factor of four improvement in execution time. VanderHeyden and co-workers [160, 161] have been developing a three-dimensional unstructured grid code for simulating multiphase flow with phase change based on JFNK methods. This work is developing novel hybrid preconditioning stratigies.

4.4 Radiation diffusion and radiation hydrodynamics

Radiation diffusion and radiation hydrodynamics are further examples of nonlinear multiple time scale systems. The equations of radiation hydrodynamics are used to simulate astrophysical phenomena and problems in inertial confinement fusion. These equations are formed by combining the compressible Navier-Stokes equations with one of many models for radiation transport. In many regimes of interest there is strong nonlinear coupling between the flow field (often called the "material") and the radiation (photon) field. Operator splitting has been used extensively to simulate radiation hydrodynamics [16]. Early Newton method research as applied to radiation hydrodynamics is found in [170]. The simplest of radiation transport models is the diffusion model, and this is where one finds most of the initial JFNK effort. As with many other multiple time scale systems, the nonequilbrium radiation diffusion problem has both a dynamical time scale and normal modes. Reaction and diffusion time scales can be very fast compared to the thermal front time scale. One wants to be able to follow the dynamical time scale [138].

Early 1D work [99, 100] focused on presenting the ideas behind JFNK to the radiation transport community, as well as elucidating the ability of JFNK to provide increases in both accuracy and efficiency as compared to more standard linearized and operator split methods. The work in [99] demonstrates that the JFNK approach has superior nonlinear convergence rates compared to a Picard iteration. The results in [100] show that when nonlinearities within a time step are not converged, one may not observe the design accuracy of the chosen time step method.

Two-dimensional examples focus on the development on simple multigrid-based preconditioning strategies [139] and on the use of operator-splitting as a preconditioner [116]. Analysis and results of operator-split based preconditioning on such problems indicate that the approach can break down [22]. Thus, there can exist a window where JFNK can integrate a system accurately but traditional operator splitting does not provide adequate preconditioning. This happens when there is an extremely large spread between the dynamical time scale (where JFNK can integrate a system accurately) and the normal modes of the system. Options to overcome this hurdle include Schur complement approaches considered in [22], coupled multigrid preconditioning, or augmenting the operator splitting with a two-stage approach that includes a defect correction and a fine grid block Jacobi smoother on the Jacobian [114].

Two recent efforts in radiation transport couple JFNK with nonlinear multigrid. In [7] JFNK is used as a smoother in an FAS scheme for a radition transport problem. The problem considered has integral coupling, which is an opportune setting for JFNK. In [106] FAS is used as a preconditioner to JFNK for a nonequilbrium radiation diffusion problem similar to that in [116]. This system is solved on an unstructured grid using agglomeration multigrid in the preconditioner.

Coupled radiation hydrodynamics problems are beginning to come under investigation using JFNK. Typically, in radiation hydrodynamics the radiation transport is done implicitly and coupled via operator splitting to an explicit method for the hydrodynamics. The research in [170] considers the implicit Newton solution to the coupled problem. The work in [10] compares a nonlinearly consistent (JFNK) method applied to the radiation model, and coupled to the hydrodynamics model in a predictor-corrector fashion. While this new approach is more accurate than the standard linearize-and-split approach, it is still able to achieve only first order in time.

4.5 Geophysical flows

Problems in porous media flow and atmospheric flows can possess widespread time scales and/or strong nonlinearities. Both of these problems motivate the consideration of JFNK methods. We briefly mention results coming from two different applications of JFNK methods to subsurface flow and one application of JFNK to atmospheric flow.

In [156] JFNK is applied to Richard's equation, a nonlinear parabolic model for variably saturated flow, and the perfomance of various Krylov methods is considered. In [71] JFNK is applied to Richards' equation. Semicoarsening multigrid [145] and simpler multigrid [4] are applied to approximately invert the preconditioning matrix. The preconditioner is the the symmetric part of the complete Jacobian. The work in [68], while not Jacobian-free, develops effective two-level preconditioners for the Newton-Krylov solution of Richard's equation.

In [43] JFNK methods are applied to multiphase flow in a permeable media. A two-stage preconditioner is developed. The first stage is a decoupling stage (similar to ABF [8]) while the second stage, solves separate (scalar) elliptic systems, as promoted in [101, 116]. The two-stage preconditioner is shown to outperform an additive split-based preconditioner. In the recent work of Hammond and co-workers [63] JFNK and an operator-split preconditioner are applied to a multicomponent reactive subsurface transport model. The JFNK method is shown to provide advantages relative to conventional methods in this field.

There is a growing interest in increasing the predictive nature of atmospheric flow simulations such as those involved in wildfire modeling [136] and hurricane modeling. Both of these problems are by nature highly nonlinear and contain multiple time scales. Currently, simulation efforrts in this community are dominated by operator splitting approaches, and thus they contain unmonitored numerical error. The work in [115, 135, 137] represents an initial effort at bringing JFNK into this community. In [115] JFNK is applied to the shallow water wave equations in 2D with the Coriolis effect. Physics-based preconditioning with simple multigrid is shown to be very effective on this classic stiff wave problem. Here the outer JFNK method is integrating a three-component hyperbolic system, while the preconditioner only requires the approximate implicit solution of a scalar parabolic equation. It is clearly shown that a second order in time JFNK method can integrate this system at the dynamical time scale (stepping over the stiff wave time scale) to obtain excellent time accuracy. The work in [137] extends the work in [115] by considering the compressible Euler equations, and thus a more complicated equation of state. This requires a more sophisticated physics-based preconditioner. Here the outer JFNK method is integrating a four-component hyperbolic system while the preconditioner only requires the approximate implicit solution of a scalar parabolic equation. The JFNK method developed in [137] will be used for simulating wildfires and hurricanes.

5 Illustrations

This section provides computational illustrations of some of the techniques and "tricks" of JFNK methods — making them work, and making them work effectively on real problems. As the heading indicates, this section is illustrative, not exhaustive. We make reference back to the appropriate areas in sections 2 through 4.

5.1 Jacobian lagging

Here we present results from [92], which considers an eight-equation system for the coupled edge plasma/Navier-Stokes neutral model. We consider the effect of Jacobian lagging only in the preconditioner (as discussed in section 3.1), versus Jacobian lagging in the outer Newton iteration. In Table 2 (Table 1 in [92]) a single grid simulation is considered using pseudo-transient continuation. SNK is a standard Newton-Krylov method (not Jacobian-free) and forms the Jacobian every Newton iteration. MFNK performs the Newton-Krylov iteration matrix-free, while the Jacobian used in the preconditioner is formed with a frequency of p Newton iterations. MNK (modified Newton-Krylov) is a standard Newton-Krylov method but the Jacobian is formed with a frequency of p Newton iterations for use in both the matrix vector multiply and the preconditioner. Also note that within the pseudo-transient continuation method, to be consistent, the time step is is advanced only every p Newton iterations. Table 2 clearly shows an execution time advantage for the JFNK approach.

Table 2: Convergence and execution time performance for an 8-equation 2D BVP for a coupled edge plasma/Navier-Stokes neutral model (from [92]).

Solution Method	Newton Iterations	GMRES Iterations	Avg. GMRES Per Newton	Rel. CPU
SNK, $p = 1$	158	485	3.1	3.56
MNK, $p = 5$	681	318	0.5	3.26
JFNK, $p = 5$	234	1570	6.7	1.46
JFNK, $p = 10$	182	2624	14.4	1.0

5.2 Mesh Sequencing

The nonlinear convergence rate enhancement resulting from mesh sequencing has been investigated in several studies for BVPs and is discussed in section 2.4.2. Table 3 (which is Table 4 in [92]) presents single grid and mesh sequencing results from an 8-equation system for the coupled edge plasma/Navier-Stokes neutral model in 2D. The impact of mesh sequencing for this BVP is clear. It is also clear that the impact of mesh sequencing increases with grid refinement.

Table 3: Effect of mesh squencing on total execution time for an 8-equation 2D BVP coupled edge plasma/Navier-Stokes neutral model (from[92]).

Problem Size	Without mesh sequencing	With mesh sequencing	Speedup
32×16	0.4	0.4	1.0
64×32	4.3	0.84	5.1
128×64	14.7	1.5	9.8

5.3 Multigrid Preconditioning

We provide some results on the impact of multigrid as a preconditioner to JFNK (discussed in section 3.3). We consider the steady-state solution of the incompressible Navier-Stokes BVP in the stream function-vorticity formulation. Figure 2 is from [98], and plots the average number of GMRES iterations per Newton iteration as a function of grid dimension for five different solution methods. The solution methods vary in the preconditioner (MG, ILU(0), or BSGS) and the number of GMRES vectors stored before restart. Advection is ignored in all preconditioners. It can be seen that the simple multigrid-based preconditioner significantly outperforms ILU(0) as the grid is refined. Furthermore, this allows storage of fewer GMRES vectors. While restart is employed on this problem, allowing 200 total GMRES iterations, its success is limited. We employ the standard restarting algorithm referred to as "Algorithm 6.11" in [143]. In terms of normalized execution time for a converged solution on the 160 × 160 grid we have: NKMG/GMRES(20) = 1.0, NKMG/GMRES(10) = 2.6, NK-ILU(0)/GMRES(100) = 3.1, NK-ILU(0)/GMRES(50) = 2.7, NK-BSGS(3)/GMRES(50) = 5.3. Only NKMG/GMRES(20) converges in a reasonable time on the 320 × 320 grid, with an execution time of 5.6, relative to the four times smaller 160 × 160 case. On the 160 × 160 grid, NKMG would not converge with less than 10 GMRES vectors stored, while NK-ILU(0) would not converge with less than 50 GMRES vectors stored.

Figure 3 is from [96], and is a plot of CPU time scaling, as a function grid dimension, using both the distributed and the coupled multigrid approaches in the preconditioner, piecewise constant restriction and prolongation, and a Galerkin coarse grid operator. The data are for 80×80, 160×160, and 320×320 problems. We include a reference line for linear scaling, and we see that both approaches scale better than linear.

5.4 Physics-based Preconditioning

Here we present some results from physics-based preconditioning, concentrating on the stiff-wave problem discussed in section 3.4.1. First we present a result from [115] where the JFNK method is applied to the 2D shallow water wave equations with the Coriolis force. This is a three-component hyperbolic system with a stiff gravity wave. As discussed in section 3.4.1, a semi-implicit method is used to construct the preconditioner. Thus, the preconditioner action needs only to approximate inversion of one step of a scalar parabolic equation, and this is accomplished with low complexity multigrid. In [115] a stiff-wave model problem is used to demonstrate that a nonlinearly consistent method (JFNK) can use time steps on the order of the dynamical time scale while maintaining comparable accuracy to a semi-implicit method run at the stiff-wave explicit CFL. An example of the algorithmic scaling of the method is given in Table 4, which is from [115]. As a result of spatial discretization mismatch between the semi-implicit method and the true nonlinear function, the preconditioner actually improves under grid refinement. The number of Newton iterations per time step is fairly constant, while the average number of GMRES iterations per time step decreases. As a result, the actual execution time beats the expected execution time scaling, which would be a factor of eight for each refinement by a factor of two in space and time.

Table 5 is from [37], and demonstrates the algorithmic scaling of physics-based preconditioning (with MG) on a three-equation MHD problem. This MHD problem contains a stiff Alfvén wave whose time scale is typically well separated from the dynamical time scale of interest. The data in Table 5 are from three different stiff wave CFL time step sizes over a range of grids. The scaling

Table 4: CFL study for physics-based preconditioning for a 2D hyperbolic gravity wave problem, showing preconditioner improvement with mesh refinement (from [115]).

$NX \times NY$	$\frac{\text{Newton}}{\text{Timestep}}$	$\frac{\text{GMRES}}{\text{Newton}}$	Advection CFL	Normalized CPU	Normalized Scaling	Error (rel. to finest)
32×32	4.12	26.26	0.1394	1.00	1.0	6.998×10^{-5}
64×64	4.01	15.68	0.2322	5.00	8.0	1.838×10^{-5}
128×128	4.00	8.45	0.2421	24.19	64.0	3.003×10^{-6}
256×256	4.00	5.22	0.2435	397.57	512.0	—

in terms of nonlinear iterations per time step and linear iterations per nonlinear iteration is good over a range of grids. As in the shallow water wave problem, the parabolic problem in each preconditioner application is approximately inverted using simple multigrid methods.

5.5 Newton-Krylov-Schwarz: Parallelism and Scaling

We conclude this section with an illustration of the use of JFNK to parallelize a legacy application code. We consider an aerodynamics application based on the code FUN3D, a tetrahedral, vertex-centered unstructured mesh code originally developed by W. K. Anderson of the NASA Langley Research Center for compressible and incompressible Euler and Navier-Stokes equations [2, 3]. FUN3D employs a control volume discretization with a variable-order Roe scheme for approximating the convective fluxes and a Galerkin discretization for the viscous terms. FUN3D has been used for design optimization of airplanes, automobiles, and submarines, with irregular meshes comprising several million mesh points. The optimization involves many analyses, typically sequential. Thus, reaching the steady-state solution in each analysis cycle in a reasonable amount of time is crucial to conducting the design optimization. A representative achievement to date for million meshpoint simulations on thousands of processors is about 10 μsec per degree of freedom for convergence of the steady-state residuals below the square-root of machine precision.

In work that was recognized with a 1999 Gordon Bell Prize in the "special" category, and subsequently published in [60], FUN3D was ported to the PETSc [6] JFNK framework, using the single program multiple data (SPMD) message-passing programming model, supplemented by multithreading at the physically shared memory level.

Achieving high sustained performance, in terms of solutions per second, requires attention to three factors. The first is a scalable implementation, in the sense that time per iteration is reduced in inverse proportion to the number of

Table 5: Grid convergence study with $\Delta t = 20, 40, 160 \times \Delta t_{CFL}$ for the tearing instability in 3D MHD with $S_L = Re = 10^4$. Results are obtained for a run of $T_f = 30\tau_A$. $\widehat{CPU}$ is the execution time “CPU” normalized to $\frac{\text{GMRES}}{\text{time step}}$(from [37]).

Grid	$\Delta t(\tau_A)$	$\frac{\text{Newton}}{\text{timestep}}$	$\frac{\text{GMRES}}{\text{Newton}}$	$\frac{\text{GMRES}}{\text{timestep}}$	CPU (s)	$\widehat{CPU}$
$\Delta t = 20\Delta t_{CFL}$						
32×32	1.875	3.0	2.6	7.8	12.8	1.6
64×64	0.9375	3.0	2.0	5.9	102.	17.3
128×128	0.46875	2.8	1.4	3.8	793.	209.
256×256	0.234375	3.0	1.0	3.0	6537.	2179.
$\Delta t = 40\Delta t_{CFL}$						
32×32	3.75	3.0	3.8	11.5	8.2	0.71
64×64	1.875	3.0	3.3	10.0	73.6	7.4
128×128	0.9375	3.0	2.0	6	517.	86.
256×256	0.46875	3.0	1.6	5.0	4248.	850.
$\Delta t = 160\Delta t_{CFL}$						
32×32	15	3.0	9.3	28.0	4.2	0.15
64×64	7.5	3.0	6.3	19.0	29.	1.5
128×128	3.75	3.1	4.6	14.2	234.	16.
256×256	1.875	3.6	5.9	21.5	3220.	150.

processors (strong scaling), or that time per iteration is constant as problem size and processor number are scaled proportionally (weak scaling). The second is good per processor performance on contemporary cache-based microprocessors. The third is algorithmic scalability, in the sense that the number of iterations to convergence does not grow with increased numbers of processors. The third factor arises because the requirement of a scalable implementation generally forces parameterized changes in the algorithm as the number of processors grows. If the convergence is allowed to degrade, however, the overall execution is not scalable, and this must be countered algorithmically.

The following is an incomplete list of parameters that need to be tuned in various phases of a pseudo-transient Newton-Krylov-Schwarz algorithm.

- Nonlinear robustness continuation parameters: discretization order, initial time step, pseudo-time step evolution law
- Newton parameters: convergence tolerance on each time step, globalization strategy (line search or trust region parameters), refresh frequency for Jacobian preconditioner
- Krylov parameters: convergence tolerance for each Newton correction, restart dimension of Krylov subspace, overall Krylov iteration limit, orthogonalization mechanism
- Schwarz parameters: subdomain number, amount of subdomain overlap, coarse grid usage
- Subdomain parameters: incomplete factorization fill level, number of sweeps

Many of these parameters have been commented on during the presentation of the JFNK and ΨNKS algorithms in sections 2 and 3. In relation to the FUN3D example, we point out that although convergence is to a second-order convection scheme discretization, much of the early nonlinear iteration uses a first-order convection scheme, until the location of the shock has stabilized. Only after this is the discretization sharpened up in the right-hand side nonlinear residuals (and consequently in the Fréchet derivative Jacobian-vector products). Otherwise, Newton is difficult to use on this problem, even with pseudo-timestepping. First-order discretization for the convective terms is used for the Jacobian preconditioner throughout.

We also call attention to the orthogonalization (or conjugation) mechanism employed in the Krylov method. Conventional GMRES employs a modified Gram-Schmidt procedure to orthogonalize the Krylov subspace; however, this requires a separate inner product for each new vector. On a parallel machine, this degree of synchronization can be counter to high implementation efficiency. It is often better to combine multiple inner products into one synchronization, by following the original Gram-Schmidt process, even though the latter is less numerically stable. Similar efficiency-stability trade-offs emerge throughout the field of parallel algorithms.

Choices for these parameters are extensively studied in [60], where there is no claim that the optimal combination has been found. JFNK algorithms are evidently rich in options. This can be overwhelming to a new user, but it provides a great deal of architectural and application adaptivity to the experienced user.

6 Nonlinear preconditioning

As discussed in Section 2.4, the lack of a global convergence theory for Newton's method is a severe drawback that has been met in practice with a variety of inventions. Some, generally those rooted in the physics known to lie behind particular discrete nonlinear systems, are applied outside of Newton's method and exercise their beneficial effect by changing the system or the initial iterate fed to Newton's method. Others, generally those rooted in mathematical assumptions about the behavior of $\mathbf{F}(\mathbf{u})$ near a root, are applied inside, and have their effect by modifying the strict Newton correction before it is accepted. In this section, we mention a new technique, additive Schwarz preconditioned inexact Newton (or "ASPIN"), that is nested inside multiple applications of Newton's method. ASPIN involves a (generally nonlinear) transformation of the original rootfinding problem for $\mathbf{F}(\mathbf{u})$ to a new rootfinding problem, $\mathcal{F}(\mathbf{u}) = 0$, to which an outer Jacobian-free Newton method is applied. The formation of $\mathcal{F}(\mathbf{u})$ at a given point $\mathbf{u}$, which is required many times in the course of performing the outer Jacobian-free Newton-Krylov iteration, in turn involves the solution of possibly many smaller nonlinear systems by Newton's method.

Without such a transformation, Newton's method may stagnate for many iterations in problems that are "nonlinearly stiff." A classical example is transonic compressible flow with a shock. The size of the global Newton step may be limited in such a problem by high curvature in the neglected terms of the multivariate expansion of $\mathbf{F}(\mathbf{u})$ coming from just a few degrees of freedom defined near the shock. Cai and collaborators [28, 29, 31] devised ASPIN to concentrate nonlinear work at such strong nonlinearities, and produce a more balanced global nonlinear problem, on which Newton behaves better, with less damping.

From an algebraic viewpoint, ASPIN is a generic transformation that requires only the unique solvability of subsystems of the original $\mathbf{F}(\mathbf{u})$ in the neighborhood of the root $\mathbf{u}_*$ to perform. From a physical viewpoint, ASPIN is a family of methods in which the subsystems may be chosen by domain decomposition, segregation of equations arising from different physical phenomena, identification of nonlinear stiffness, or still other criteria. As with all Schwarz methods, many flavors of nonlinear Schwarz preconditioning are possible — additive, multiplicative, or general polynomial combination of sub-operators; single-level or multi-level; overlapping or nonoverlapping. In this section, we discuss the additive, single-level case, with arbitrary overlap. We illustrate domain and equation partitioning.

6.1 A Nonlinear Additive Schwarz Preconditioner

In this section, following [28], we briefly describe a nonlinear preconditioner based on the additive Schwarz method. In a practical application of ASPIN we assume that there is a dominant association of certain components of $\mathbf{F}$ with certain components of $\mathbf{u}$, and that, in fact, the square subblock of the full Jacobian $\mathbf{F}'(\mathbf{u})$ describing this dominant relationship is invertible near the desired root. This is major restriction for general nonlinear algebraic systems, but it is completely natural for systems arising from partial differential equations describing local conservation laws. Just as in additive Schwarz for linear problems, most of the computing in ASPIN is carried out within these local blocks.

Let the unknowns $\mathbf{u} \in R^n$ and residuals $\mathbf{F} \in R^n$ be partitioned into N (possibly overlapping) subsets. Based on the partitioning, we introduce subspaces of R^n and the corresponding restriction and extension matrices. For the i^{th} subset, define $\mathbf{V}_i \subset R^n$ as the subspace of vectors whose components vanish outside of the i^{th} subset.

Recalling the restriction operators of Section 3.2, define the subdomain nonlinear function as

$$\mathbf{F}_i = R_i \mathbf{F}.$$

For any vector $\mathbf{v} \in R^n$, define $\mathbf{T}_i(\mathbf{v}) \in \mathbf{V}_i$ as the solution of the nonlinear system within the i^{th} subspace:

$$\mathbf{F}_i(\mathbf{v} - \mathbf{T}_i(\mathbf{v})) = 0.$$

Because of the restriction of $\mathbf{T}_i$ to $\mathbf{V}_i$, with n_i nontrivial components, this is for each i a system of n_i nonlinear equations in the same number of unknowns. Under the natural association of equations to unknowns in the discretization of a well-posed PDE, we expect the Jacobian of this system to be nonsingular, and the local problem to be solvable, for reasonable $\mathbf{v}$. If, for instance, $\mathbf{v}$ represents the desired discrete solution in every component outside of the i^{th} subspace, solving the i^{th} subproblem amounts to consistently extending the global solution into the i^{th} subspace. The new function

$$\mathcal{F}(\mathbf{u}) = \sum_{i=1}^{N} \mathbf{T}_i(\mathbf{u}), \tag{38}$$

formed by summing together all of these local corrections we refer to as the nonlinearly preconditioned $\mathbf{F}(\mathbf{u})$. A single evaluation of the function $\mathcal{F}(\mathbf{v})$, for a given $\mathbf{v}$, involves the calculation of the $\mathbf{T}_i$, which in turn involves the solution of N nonlinear systems. If the overlap is zero, this is a block nonlinear Jacobi preconditioner. It is shown in [28] that $\mathbf{u}$ is a root of $\mathcal{F}(\mathbf{u})$ if and only if it is a root of $\mathbf{F}(\mathbf{u})$.

In the linear case, this algorithm reduces to the additive Schwarz algorithm. Using the usual notation, if

$$\mathbf{F}(\mathbf{u}) = \mathbf{A}\mathbf{u} - \mathbf{b},$$

then

$$\mathcal{F}(\mathbf{u}) = \left(\sum_{i=1}^{N} R_i^T \mathbf{A}_i^{-1} R_i\right)(\mathbf{A}\mathbf{u} - \mathbf{b}),$$

where $\mathbf{A}_i^{-1}$ is the inverse of $\mathbf{A}_i = R_i \mathbf{A} R_i^T$.

If $\mathcal{F}(\mathbf{u}) = 0$ is to be solved using a Newton type algorithm, then the Jacobian $\mathcal{F}'(\mathbf{u})$ is needed in one form or another. Since this is generally large and dense, Jacobian-free methods are essential. It is shown in [28] that $\mathcal{F}'(\mathbf{u})$ is well approximated at a point $\mathbf{u}$ near a root by $\mathcal{J}(\mathbf{u}) = \sum_{i=1}^{N} R_i^T \mathbf{J}_i^{-1}(\mathbf{u}) R_i \mathbf{J}(\mathbf{u})$, where $\mathbf{J}(\mathbf{u})$ is the Jacobian of the original nonlinear system, $\mathbf{F}'(\mathbf{u})$, and $\mathbf{J}_i$, is the Jacobian of the subdomain nonlinear system, $\mathbf{J}_i(\mathbf{u}) = R_i \mathbf{J}(\mathbf{u}) R_i^T$, for $i = 1, \ldots, N$. If $\mathbf{F}(\mathbf{u})$ is sparse nonlinear function of its arguments, then $\mathbf{J}$ is a sparse matrix, and so are the $\mathbf{J}_i$, so it is economical to apply $\mathcal{J}$ to an arbitrary vector. Therefore, nonlinear systems involving $\mathcal{F}(\mathbf{u})$ are straightforward to solve using JFNK, provided that the outer Newton-Krylov iteration does not need any linear preconditioning. We note that the inner Newton iterations to form the $\mathbf{T}_i(\mathbf{u})$ may also be solved with JFNK, or by any other Newton method, assuming we have the $\mathbf{J}_i$ available.

From a software engineering viewpoint, it is convenient that the action required to apply the Jacobian of the nonlinearly preconditioned system to an arbitrary vector is already present in any Newton-Krylov-Schwarz code, since it is the action of the linearly Schwarz-preconditioned Jacobian of the original nonlinear function $\mathbf{F}(\mathbf{u})$. The action of $\mathbf{J}$ on a vector can be approximated by the usual Fréchet derivative in a matrix-free manner, or with explicit elements. The actions of $\mathbf{J}_i^{-1}$ on subvectors corresponding to the nontrivial components in $\mathbf{V}_i$ can be performed concurrently within (possibly overlapping) partitions. Several techniques are available for computing the $\mathbf{J}_i$, for example, by analytic formula, multi-colored finite differencing, or automatic differentiation. A triangular factorization of $\mathbf{J}_i$ may be performed, since the action is needed multiple times within a single outer Newton step on $\mathcal{F}(\mathbf{u}) = 0$.

The evaluation of $\mathcal{F}(\mathbf{u})$ is not a process with *a priori* deterministic complexity, since it involves summation of local corrections that are the result of inner Newton iterations on $\mathbf{F}_i(\mathbf{u})$. In a parallel implementation, if a different processor is assigned to each partition and the partitions overlap, communication to obtain nontrivial ghost values belonging to and updated in other processors may be necessary. We note, however, that these ghost values do not change during the solution of the subdomain nonlinear system. It is known (Section

2.3.2) that the linear systems of the main Newton iteration for $\mathcal{F}(\mathbf{u})$ do not need to be solved well (in fact, often *should* not be solved well) in early stages, in the sense that it is only necessary to enforce the norm of the Newton correction equation residual

$$||\mathcal{F}(\mathbf{u})\delta\mathbf{u} + \mathcal{F}(\mathbf{u})||_2$$

to be bounded by some "forcing term" [50] that may itself be a function of $\mathcal{F}$ and other by-products of the computation, approaching zero at a rate sufficient to guarantee convergence, superlinear convergence, or even quadratic convergence of the outer iteration. This provides some latitude in the degree of accuracy to which the subdomain nonlinear problems are solved in the early iterations.

It is shown in [28] that Newton's method applied to a nonlinearly preconditioned version of the velocity-vorticity driven cavity problem, based on domain decomposition converges rapidly (e.g., in 5–10 Newton iterations) at Reynolds numbers far beyond those at which Newton's method applied to the original discretization of the problem hopelessly stagnates.

It is shown in [31] that Newton convergence of the nonlinearly transformed version of the problem of shocked flow in a variable cross-section duct is much less sensitive to mesh refinement than the original discretization.

The difficulties of Newton on the driven cavity problem can be ameliorated by continuation in Reynolds number and the shocked flow problem through mesh sequencing in other contexts. Nevertheless, it is interesting to see that a purely algebraic method, ASPIN, is effective at rebalancing nonlinearities so that Newton converges easily. We expect that it will have wide applicability in problems with complex nonlinearities as a means of increasing nonlinear robustness.

Unfortunately, it is difficult to obtain direct approximations to the dense Jacobian of the transformed system, $\mathcal{J}$, so as to improve the *linear conditioning* of the resulting Newton correction problems. Therefore, these problems are subject to linear ill-conditioning as mesh resolution increases. To conquer this linear ill-conditioning, multi-level methods of ASPIN need to be devised. The straightforward FAS multigrid approach on $\mathcal{F}(\mathbf{u})$ may not be practical since the nonlinear correction to be computed at each coarse level requires an evaluation of the fine-grid residual, which is subsequently restricted to compute the coarse-grid defect that drives the correction. Since each fine-grid residual involves a host of fine-grid nonlinear subproblems, this is expensive. An alternative multilevel method is investigated, with promising results, in [29].

A case of special interest in ASPIN is the case of relatively few subspaces, e.g., partitioning by equation type in a multicomponent problem, as opposed to by subdomain in a problem of millions of gridcells. In this case, it is natural to take the overlap to be zero; then the diagonal blocks of $\sum_{i=1}^{N} R_i^T \mathbf{J}_i^{-1} R_i \mathbf{J}$ are all identities, and do not involve any computations when multiplied with vectors. For two subdomains,

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}$$

so

$$\sum_{i=1}^{2} R_i^T \mathbf{J}_i^{-1} R_i \mathbf{J} = \left(\begin{array}{cc} \mathbf{I} & \mathbf{J}_{11}^{-1}\mathbf{J}_{12} \\ \mathbf{J}_{22}^{-1}\mathbf{J}_{21} & \mathbf{I} \end{array} \right).$$

For example, partition 1 could represent the fluid degrees of freedom, and partition 2 the structural degrees of freedom in a fluid-structure interaction. The Jacobians of each need to be inverted only on the portion of each partition that couples directly to the other.

7 PDE-constrained Optimization

It is increasingly recognized that PDE-based analyses are rarely ends in themselves, but more properly part of a scientific process that includes some type of sensitivity analysis or optimization, in which the system of PDEs serves as a constraint. The optimization process may arise, for instance, in design, in control, in parameter identification (inverse problems), or in data assimilation. Some property, such as the integrated dissipation rate or the norm of some discrepency between measured and modeled outcomes, is to be minimized, subject to the constraint that a governing system of PDEs is satisfied. Without the PDE constraint, the optimization algorithm may find more optimal values of the objective that are physically infeasible, and therefore uninteresting.

Jacobian-free Newton-Krylov methods have important roles to play in PDE-constrained optimization. At the very least, the fact that PDEs need to be solved in the inner loop of a conventional constrained optimization algorithm requires time- and memory-efficient PDE solvers. A Newton method makes good use of a "warm" initial guess, and therefore performs well in projecting the result of an optimization step onto the constraint manifold inside an iterative optimization method such as Reduced Sequential Quadratic Programming (RSQP) [172]. However, it has become apparent in recent years that there are potentially much more efficient classes of optimization algorithms that employ a Jacobian-free Newton-Krylov method as the *outer* optimization loop, not as the *inner* projection step. These methods search for saddle points of the Lagrangian formulation of the constrained optimization by looking for the roots of the gradient of the Lagrangian with respect to all of its parameters — the design parameters, the state variables of the PDE, and the Lagrange multipliers. In these contemporary optimization methods, called Lagrange-Newton-Krylov (LNK) methods [12, 13, 81], the PDE constraints are not necessarily satisfied accurately at every step, but are only guaranteed to be satisfied asymptotically, as a design parameter optimum is approached.

Our discussion of the LNK family of methods in this survey is introductory only, since a systematic treatment demands first a survey of constrained optimization methdology, and also since this application of JFNK is relatively young. However, a presentation of the prospects for JFNK would be incomplete without mentioning its use in LNK, and such a presentation can be self-contained with the realization that equality-constrained optimization is treatable as a nonlinear rootfinding problem.

One of the chief practical differences between JFNK applied to PDEs and JFNK applied to Lagrangian constrained optimization is that the Jacobian of the direct analysis involves only first derivatives of the conservation laws with respect to the state variables, whereas the Jacobian of the Lagrangian problem (the so-called Karush-Kuhn-Tucker, KKT, matrix) involves second derivatives of the objective function and the PDE conservation laws. It is relatively routine to obtain approximations to first derivatives of well-scaled objects in standard floating point contexts with finite differences. This is not true for second derivatives since a second difference amplifies the roundoff to levels generally unacceptable for standard double-precision floating point. Therefore, the subject of automatic differentiation makes an important appearance in optimization. Moreover, the Jacobian of the Lagrangian problem involves in a fundamental way the transpose of the Jacobian of the PDE constraints with respect to the state variables. In the JFNK context, it is not known how to form Jacobian-transpose-vector products with finite Frechét derivatives. Automatic differentiation, whose so-called "reverse mode" permits efficient Jacobian-transpose applications, is therefore important for this reason, as well. Indeed, it is no coincidence that LNK is only now becoming a practically important method, with the advent of quality automatic differentiation software [14, 15].

7.1 Newton's Method in Constrained Optimization

Equality constrained optimization leads, as mentioned, through the Lagrangian formulation, to a multivariate nonlinear rootfinding problem for the gradient (the first-order necessary conditions), which is amenable to treatment by Newton's method. To establish notation, consider the following canonical framework, in which we enforce equality constraints on the state variables only. (Design variable constraints require additional notation, and inequality constraints require additional algorithmics, so we leave their generalization to the literature [172].) Choose design variables $u \in R^m$ to minimize the scalar objective function, $\phi(u, x)$, subject to state constraints, $h(u, x) = 0$, where $x \in R^n$ is the vector of state variables. In the Lagrange framework, a stationary point of the scalar Lagrangian function

$$\mathcal{L}(x, u, \lambda) \equiv \phi(x, u) + \lambda^T h(x, u)$$

is sought, where $\lambda \in R^n$. When Newton's method is applied to the first-order optimality conditions, a linear system known as the Karush-Kuhn-Tucker (KKT) system arises at each step. There is a natural "outer" partitioning: the vector of parameters is often of lower dimension than the vectors of states and multipliers. This suggests an approximate Schur complement-like block preconditioning process at the outer level. Within the state-variable subproblem, in turn, Schwarz provides a natural “inner” partitioning for concurrency.

To emphasize differences of computational scale relevant to the algorithmics, we mention three classes of PDE-constrained optimization:

- **Design optimization** (especially shape optimization): u parametrizes the domain geometry of the PDE (e.g., a lifting surface) and ϕ is a cost-to-benefit ratio of forces, energy expenditures, etc. Typically, m is small compared with n and does not scale directly with it as the mesh is refined. However, m may still be hundreds or thousands in industrial applications.

- **Optimal control**: u parametrizes a continuous control function acting in part of the domain or on part of the boundary of the domain, and ϕ is the norm of the difference between desired and actual responses of the system. For boundary control, $m \propto n^{2/3}$.

- **Parameter identification/data assimilation**: u parametrizes an unknown continuous constitutive or forcing function defined throughout the domain, and ϕ is the norm of the difference between measurements and simulation results. Typically, $m \propto n$.

Written out in partial detail, the first-order optimality conditions are:

$$\frac{\partial \mathcal{L}}{\partial x} \equiv \frac{\partial \phi}{\partial x} + \lambda^T \frac{\partial h}{\partial x} = 0 \, , \, \frac{\partial \mathcal{L}}{\partial u} \equiv \frac{\partial \phi}{\partial u} + \lambda^T \frac{\partial h}{\partial u} = 0 \, , \, \frac{\partial \mathcal{L}}{\partial \lambda} \equiv h = 0 \; .$$

Newton’s method iteratively seeks a correction,

$$\begin{pmatrix} \delta x \\ \delta u \\ \delta \lambda \end{pmatrix} \quad \text{to the iterate} \quad \begin{pmatrix} x \\ u \\ \lambda \end{pmatrix}$$

to reduce the gradient of the Lagrangian to zero. With subscript notation for the partial derivatives, the Newton correction (KKT) equations are

$$\begin{bmatrix} (\phi_{,xx} + \lambda^T h_{,xx}) & (\phi_{,xu} + \lambda^T h_{,xu}) & h_{,x}^T \\ (\phi_{,ux} + \lambda^T h_{,ux}) & (\phi_{,uu} + \lambda^T h_{,uu}) & h_{,u}^T \\ h_{,x} & h_{,u} & 0 \end{bmatrix} \begin{pmatrix} \delta x \\ \delta u \\ \delta \lambda \end{pmatrix} = - \begin{pmatrix} \phi_{,x} + \lambda^T h_{,x} \\ \phi_{,u} + \lambda^T h_{,u} \\ h \end{pmatrix}$$

or

$$\begin{bmatrix} W_{xx} & W_{ux}^T & J_x^T \\ W_{ux} & W_{uu} & J_u^T \\ J_x & J_u & 0 \end{bmatrix} \begin{pmatrix} \delta x \\ \delta u \\ \lambda_+ \end{pmatrix} = - \begin{pmatrix} g_x \\ g_u \\ h \end{pmatrix} , \tag{39}$$

where $W_{ab} \equiv \frac{\partial^2 \phi}{\partial a \partial b} + \lambda^T \frac{\partial^2 h}{\partial a \partial b}$, $J_a \equiv \frac{\partial h}{\partial a}$, and $g_a = \frac{\partial \phi}{\partial a}$, for $a, b \in \{x, u\}$, and where $\lambda_+ = \lambda + \delta \lambda$.

7.2 Newton-RSQP and LNK

The RSQP method [172] consists of a three-stage iteration. We follow the language and practice of [12, 13].

- **Design Step** (Schur complement for middle blockrow):

$$H\ \delta u = f\ ,$$

where H and f are the reduced Hessian and gradient, respectively:

$$H \equiv W_{uu} - J_u^T J_x^{-T} W_{ux}^T + \left(J_u^T J_x^{-T} W_{xx} - W_{ux}\right) J_x^{-1} J_u,$$

$$f \equiv -g_u + J_u^T J_x^{-T} g_x - \left(J_u^T J_x^{-T} W_{xx} - W_{ux}\right) J_x^{-1} h.$$

- **State Step** (last blockrow):

$$J_x\ \delta x = -h - J_u\ \delta u.$$

- **Adjoint Step** (first blockrow):

$$J_x^T\ \lambda_+ = -g_x - W_{xx}\ \delta x - W_{ux}^T\ \delta u.$$

In each overall iteration, we must form and solve with the reduced Hessian matrix H, and we must solve separately with J_x and J_x^T. The latter two solves are almost negligible compared with the cost of forming H, which is dominated by the cost of forming the sensitivity matrix $J_x^{-1} J_u$. Because of the quadratic convergence of Newton, the number of overall iterations is few (asymptotically independent of m). However, the cost of forming H at each design iteration is m solutions with J_x. These are potentially concurrent over the m independent columns of J_u, but prohibitive.

In order to avoid computing any Hessian blocks, the design step may be approached in a quasi-Newton (e.g., BFGS) manner [172]. Hessian terms are dropped from the adjoint step right-hand side.

- **Design Step** (severe approximation to middle blockrow):

$$Q\ \delta u = -g_u + J_u^T J_x^{-T} g_x\ ,$$

where Q is a quasi-Newton approximation to the reduced Hessian, H.

- **State Step** (last blockrow):

$$J_x\ \delta x = -h - J_u\ \delta u.$$

- **Adjoint Step** (approximate first blockrow):

$$J_x^T \lambda_+ = -g_x.$$

In each overall iteration of this quasi-Newton RSQP, we must perform a low-rank update on Q or its inverse, and we must solve with J_x and J_x^T. This strategy vastly reduces the cost of an iteration; however, it is no longer a Newton method. The number of overall iterations is many. Since BFGS is equivalent to unpreconditioned CG for quadratic objective functions, $\mathcal{O}(m^p)$ sequential cycles ($p > 0$, $p \approx \frac{1}{2}$) may be anticipated. Hence, quasi-Newton RSQP is not scalable in the number of design variables, and no ready form of parallelism can address this convergence-related defect.

To summarize, conventional RSQP methods apply a (quasi-)Newton method to the optimality conditions: solving an approximate $m \times m$ system to update u, solving an $n \times n$ system to update x and λ consistently, and iterating. The unpalatable expense arises from the exact linearized analyses for updates to x and λ that appear in the inner loop. We therefore consider replacing the exact elimination steps of RSQP with preconditioning steps in an outer loop, arriving at LNK.

Consider applying a Krylov-Schwarz method directly to the $(2n+m) \times (2n+m)$ KKT system, Eq. (39). For this purpose, we require the action of the full matrix on the full-space vector and a good full-system preconditioner, for algorithmic scalability. One Newton SQP iteration is a perfect preconditioner—a block factored solver, based on forming the reduced Hessian of the Lagrangian H—but, of course, far too expensive. Backing off wherever storage or computational expense becomes impractical for large-scale PDEs generates a family of attractive methods.

To precondition the full system, we need approximate inverses to the three left-hand side matrices in the first algorithm of §7.2, namely, H, J, and J^T. If a preconditioner is available for H, and exact solves are available for J, and J^T, then it may be shown [79] that conjugate gradient Krylov iteration on the (assumed symmetrizable) reduced system and conjugate gradient iteration on the full system yield the same sequence of iterates. The iterates are identical in the sense that if one were to use the values of u arising from the iteration on the reduced system in the right-hand side of the block rows for x and λ, one would reconstruct the iterates of the full system, when the same preconditioner used for H in the reduced system is used for the W_{uu} block in the full system. Moreover, the spectrum of the full system is simply the spectrum of the reduced system supplemented with a large multiplicity of unit eigenvalues. If one retreats from exact solves with J and J^T, this equivalence no longer holds; however, if good preconditioners are used for these Jacobian blocks, then the cloud of eigenvalues around unity is still readily shepherded by a Krylov method, and convergence should be nearly as rapid as in the case of exact solves.

This Schur-complement-based preconditioning of the full system was proposed in this equality-constrained optimization context by Biros and Ghattas in 1998 [12]. From a purely algebraic point of view (divorced from optimization), the same Schur-complement-based preconditioning was advocated by Keyes and Gropp in 1987 [79] in the context of domain decomposition. There, the reduced system was a set of unknowns on the interface between subdomains, and the savings from the approximate solves on the subdomain interiors more than paid for the modest degradation in convergence rate relative to interface iteration on the Schur complement. The main advantage of the full system problem is that the Schur complement never needs to be formed. Its exact action is felt on the design variable block through the operations carried out on the full system.

Biros and Ghattas have demonstrated the large-scale parallel effectiveness of the full system algorithm on a 3D Navier-Stokes flow boundary control problem, where the objective is dissipation minimization of flow over a cylinder using suction and blowing over the back portion of the cylinder as the control variables [13]. They performed this optimization with domain-decomposed parallelism on 128 processors of a T3E, using an original optimization toolkit add-on to the PETSc [6] toolkit. To quote one result from [13], for 6×10^5 state constraints and 9×10^3 controls, full-space LNKS with approximate subdomain solves beat quasi-Newton RSQP by an order of magnitude (4.1 hours versus 53.1 hours).

Automatic differentiation has two roles in the new algorithm: formation of the action on a Krylov vector of the full KKT matrix, including the full second-order Hessian blocks, and supply of approximations to the elements of J (and J^T) for the preconditioner. LNK will generally be applied to large problems of n state variables and m parameters. Upon surveying existing AD tools, it is concluded in [81] that the preconditioned matrix-vector product can be formed in time linear in these two parameters.

8 Conclusions and Prospects

We conclude with brief remarks on future directions for JFNK methodology, as influenced by directions for scientific and engineering applications, computer architecture, mathematical software, and the on-going development of other numerical techniques.

Computational simulation of systems governed by PDEs is being relied upon as never before for accurate results on which to base massive economic investments, public and corporate, as well as critical governmental policies. For instance, the ASCI program is intended to provide a computational alternative to nuclear weapons testing and the SciDAC program to help target investments in fusion energy devices and next-generation accelerators. The 40 Teraflop/s, 5120-processor Japanese Earth Simulator is intended to allow unprecedented resolution and forward time integration horizons for climate prediction.

The typical governing system confronted in these Grand Challenge problems is nonlinear, coupled, multiscale, and multirate. The typical computational environment is distributed shared memory. To address combination requires scalable nonlinear implicit solvers, for which we propose preconditioned Jacobian-free Newton-Krylov methods. As shown here, JFNK methods offer asymptotically rapid nonlinear convergence, and with proper preconditioning can also be both linearly scalable and efficiently parallelizable.

Preconditioning is where the battle for scalability is won or lost. Therefore, in this article we have reviewed a set of preconditioning techniques so varied that all that some of them have in common is that their action does not directly rely on the matrix elements of the true Jacobian. The distinction between the implicit forward action of the true Jacobian and the inverse action of the an approximate Jacobian, which may be defined only by a subroutine call that maps a residual into an approximate delta correction, is fundamental to the culture of JFNK. As new ideas and implementations for preconditioners evolve, the JFNK method readily absorbs them.

The generality of preconditioning and multiplicity of Jacobian representations exploited in JFNK dictates an open software infrastructure, such as, e.g., the PETSc [6] or Aztec [158] solver frameworks, and invites the reuse of valuable existing user application solver code, which is reinterpreted as a component of the preconditioner.

Future work planned by the authors include the release, in one or more of these JFNK software frameworks, of tutorial examples of the advanced use of JFNK in a variety of fields. We also invite existing and new users of JFNK to post us with their own successes and challenges and to join in expanding the algorithm and application scope of this compelling methodology.

Acknowledgments

The authors thank all of their co-authors and co-explorers in Newton-Krylov methodology over the past dozen years, especially S. Balay, J. Brackbill, X.-C. Cai, L. Chacon, O. Ghattas, W. D. Gropp, G. Hansen, P. Hovland, D. Kaushik, P. McHugh, L. McInnes, V.A. Mousseau, M. Pernice, W. Rider, B. F. Smith, and D. P. Young, for contributing results from onging research as well as for helpful suggestions in the presentation of this material.

This work was supported under the auspices of the U.S. Department of Energy under DOE contract W-7405-ENG-36 at Los Alamos National Laboratory under DOE subcontract B347882 from Lawrence Livermore National Laboratory to Old Dominion University, and DOE cooperative agreement DE-FC02-01ER25476 with Old Dominion University.

References

[1] D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computa-*

tional Fluid Dynamics and Heat Transfer, Hemisphere Publishing Corp., New York, 1984.

[2] W. K. Anderson and D. L. Bonhaus, *An implicit upwind algorithm for computing turbulent flows on unstructured grids*, Comp. Fluids, 23 (1994), pp. 1–21.

[3] W. K. Anderson, R. D. Rausch, and D. L. Bonhaus, *Implicit/multigrid algorithms for incompressible turbulent flows on unstructured grids*, J. Comput. Phys., 128 (1996), pp. 391–408.

[4] S. F. Ashby and R. D. Falgout, *A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations*, Nucl. Sci. Eng., 124 (1996), pp. 145–159.

[5] O. Axelson, *Iterative Solution Methods*, Cambridge University Press, 1994.

[6] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, *PETSc users manual*, Tech. Rep. ANL-95/11 - Revision 2.1.0, Argonne National Laboratory, Apr. 2001.

[7] D. Balsara, *Fast and accurate discrete ordinates methods for multidimensional radiative transfer. part I, basic methods*, J. Quant. Spectr. Radiat. Transfer, 69 (2001), pp. 671–707.

[8] R. Bank, T. Chan, W. Coughran, and R. Smith, *The alternate block factorization procedure for systems of partial differential equations*, BIT, 29 (1989), pp. 938–954.

[9] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, 1994.

[10] J. Bates, D. A. Knoll, W. Rider, R. Lowrie, and V. Mousseau, *On consistent time-integration methods for radiation hydrodynamics in the equilibrium diffusion limit: low energy density regime*, J. Comput. Phys., 167 (2001), pp. 99–130.

[11] M. Benzi, *Preconditioning techniques for large linear systems: A survey*, J. Comput. Phys., (2002, in press).

[12] G. Biros and O. Ghattas, *Parallel Newton-Krylov methods for PDE-constrained optimization*, in Proceedings of SC99, IEEE Computer Society, 1999.

[13] ——, *A Lagrange-Newton-Krylov-Schur method for PDE-constrained optimization*, SIAG/OPT Views-and-News, 11 (2000), pp. 1–6.

[14] C. Bischof, A. Carle, G. Corliss, A. Griewank, and P. Hovland, *ADIFOR - generating derivative codes from Fortran programs*, Scientific Programming, 1 (1992), pp. 1–29.

[15] C. Bischof, L. Roh, and A. Mauer, *ADIC — An extensible automatic differentiation tool for ANSI-C*, Software–Practice and Experience, 27 (1997), pp. 1427–1456.

[16] R. L. Bowers and J. R. Wilson, *Numerical Modeling in Applied Physics and Astrophysics*, Jones and Bartlett, Boston, 1991.

[17] J. U. Brackbill and D. A. Knoll, *Transient magnetic reconnection and unstable shear layers*, Phys. Rev. Lett., 86 (2001), pp. 2329–2332.

[18] A. Brandt, *Multi-level adaptive solutions to boundary value problems*, Math. Comp., 31 (1977), p. 333.

[19] ——, *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics*, tech. rep., von Karman Institue, 1984.

[20] P. N. Brown, *A local convergence theory for combined inexact-Newton / finite-difference projection methods*, SIAM J. Num. Anal., 24 (1987), pp. 407–434.

[21] P. N. Brown and Y. Saad, *Hybrid Krylov methods for nonlinear systems of equations*, SIAM J. Sci. Stat. Comput., 11 (1990), pp. 450–481.

[22] P. N. Brown and C. Woodward, *Preconditioning strategies for fully implicit radiation diffusion with material-energy coupling*, SIAM J. Sci. Comput., 23 (2001), pp. 499–516.

[23] X.-C. Cai, M. Dryja, and M. Sarkis, *RASHO: A restricted additive Schwarz preconditioner with harmonic overlap*, in Proceedings of the 13th International Conference on Domain Decomposition Methods, CIMNE, 2002.

[24] X.-C. Cai, C. Farhat, and M. Sarkis, *Schwarz methods for the unsteady compressible Navier-Stokes equations on unstructured meshes*, in Proceedings of the Eighth International Conference on Domain Decomposition Methods, Wiley, 1997, pp. 453–460.

[25] ——, *A minimum overlap restricted additive Schwarz preconditioner and applications in 3d flow simulations*, in Proceedings of the Tenth International Conference on Domain Decomposition Methods, AMS, 1998, pp. 238–244.

[26] 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), pp. 246–265.

[27] X.-C. Cai, W. D. Gropp, D. E. Keyes, and M. D. Tidriri, *Newton-Krylov-Schwarz methods in CFD*, in Proceedings of the International Workshop on Numerical Methods for the Navier-Stokes Equations, Vieweg, 1995, pp. 17–30.

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

[29] X.-C. Cai, D. E. Keyes, and L. Marcinkowski, *Nonlinear additive Schwarz preconditioners and applications in computational fluid dynamics*, Int. J. of Num. Meth. Fluids), (2002, in press).

[30] X.-C. Cai, D. E. Keyes, and V. Venkatakrishnan, *Newton-Krylov-Schwarz: An implicit solver for CFD*, in Proceedings of the Eighth International Conference on Domain Decomposition Methods, Wiley, 1997, pp. 387–400.

[31] X.-C. Cai, D. E. Keyes, and D. P. Young, *A nonlinearly additive Schwarz preconditioned inexact Newton method for shocked duct flow*, in Proceedings of the 13th International Conference on Domain Decomposition Methods, Domain Decomposition Press, 2002.

[32] X.-C. Cai and M. Sarkis, *A restricted additive Schwarz preconditioner for general sparse linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 792–797.

[33] L. Chacon, *Fokker-Planck modeling of spherical inertial electrostatic, virtual-cathode fusion systems*, PhD thesis, University of Illinois, 2000.

[34] L. Chacon, D. C. Barnes, D. A. Knoll, and G. H. Miley, *An implicit energy-conservative 2D Fokker-Planck algorithm: II Jacobian-Free Newton-Krylov solver*, J. Comput. Phys., 157 (2000), pp. 654–682.

[35] ——, *A bounce-averaged Fokker-Planck code for Penning fusion devices*, Comput. Phys. Comm., 134 (2001), pp. 182–208.

[36] L. Chacon and D. A. Knoll, *An implicit nonlinear Hall MHD solver*, J. Comput. Phys., (2002, submitted).

[37] L. Chacon, D. A. Knoll, and J. M. Finn, *An implicit nonlinear reduced resistive MHD solver*, J. Comput. Phys., 178 (2002), pp. 15–36.

[38] L. Chacon, G. H. Miley, D. C. Barnes, and D. A. Knoll, *Energy gain calculations in Penning fusion systems using a bounce-averaged Fokker-Planck model*, Physics of Plasmas, 7 (2000), pp. 4547–4560.

[39] T. F. Chan and K. R. Jackson, *Nonlinearly preconditioned Krylov subspace methods for discrete Newton algorithms*, SIAM J. Sci. Stat. Comput., 5 (1984), pp. 533–542.

[40] T. S. Coffey, C. T. Kelley, and D. E. Keyes, *Pseudo-transient continuation and differential-algebraic equations.* submitted to SIAM J. Sci. Comput., 2002.

[41] D. E. Culler, J. P. Singh, and A. Gupta, *Parallel Computer Architecture*, Morgan-Kaufman, 1998.

[42] G. D. V. Davis, *Natural convection of air in a square cavity: a benchmark numerical solution*, Int. J. Num. Meth. Fluids, 3 (1983), p. 249.

[43] C. Dawson, H. Klie, M. Wheeler, and C. Woodward, *A parallel, implicit, cell centered method for two-phase flow with a preconditioned Newton-Krylov solver*, Comp. Geosciences, 1 (1997), pp. 215–249.

[44] R. Dembo et al., *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[45] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, 1983.

[46] H. A. V. der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 631–644.

[47] H. A. V. der Vorst and C. Vuik, *A comparision of some GMRES-like methods*, Linear Algebra and its Applications, 160 (1992), pp. 131–162.

[48] M. Dryja and O. B. Widlund, *An additive variant of the Schwarz alternating method for the case of many subregions*, Tech. Rep. 339, Courant Institute, NYU, 1987.

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

[50] ——, *Choosing the forcing terms in an inexact Newton method*, SIAM J. Sci. Comput., 17 (1996), pp. 16–32.

[51] E. M. Epperlein, *Implicit and conservative difference scheme for the Fokker-Planck equation*, J. Comput. Phys., 112 (1994), p. 291.

[52] A. Ern, V. Giovangigli, D. E. Keyes, and M. D. Smooke, *Towards polyalgorithmic linear system solvers for nonlinear elliptic systems*, SIAM J. Sci. Comput., 15 (1994), pp. 681–703.

[53] R. Freund, *A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems*, SIAM J. Sci. Comput., 14 (1993), pp. 470–482.

[54] M. Garbey, R. Hoppe, D. E. Keyes, Y. Kuznetsov, and J. Périaux, eds., *Proceedings of the 13th International Conference on Domain Decomposition Methods*, CIMNE, 2002. See also www.ddm.org.

[55] U. Ghia, K. Ghia, and C. Shin, *High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method*, J. Comput. Phys., 48 (1982), p. 387.

[56] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer Verlag, 1984.

[57] G. H. Golub and D. O'Leary, *Some history of the conjugate gradient and Lanczos algorithms: 1948–1976*, SIAM Review, (1989), pp. 50–102.

[58] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, 1997.

[59] W. Gropp, D. E. Keyes, L. McInnes, and M. Tidriri, *Globalized Newton-Krylov-Schwarz algorithms and software for parallel implicit CFD*, Int. J. High Performance Computing Application, 14 (2000), pp. 102–136.

[60] W. D. Gropp, D. K. Kaushik, D. E. Keyes, and B. F. Smith, *High performance parallel implicit CFD*, Parallel Computing, 27 (2001), pp. 337–362.

[61] H. Gummel, *A self-consistent iterative scheme for one-dimensional steady state transistor calculations*, IEEE Trans. Electron. Devices, ED-11 (1964), pp. 455–465.

[62] W. Hackbusch, *Iterative Methods for Large Sparse Linear Systems*, Springer, 1993.

[63] G. E. Hammond, A. J. Valocchi, and P. C. Lichtner, *Modeling multicomponent reactive transport on parallel computers using Jacobian-Free Newton-Krylov with operator-splt preconditioning*, Comput. Meths. Water Resources, (2002). submitted.

[64] F. Harlow and A. Amsden, *A numerical fluid dynamical calculation method for all flow speeds*, J. Comput. Phys., 8 (1971), pp. 197–214.

[65] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research of the National Bureau of Standards, Section B, 49 (1952), pp. 409–436.

[66] P. Hovland and L. C. McInnes, *Parallel simulation of compressible flow using automatic differentiation and PETSc*, Parallel Computing, 27 (2001), pp. 503–519.

[67] B. R. Hutchinson and G. D. Raithby, *A multigrid method based on the additive correction strategy*, Numer. Heat Trans., 9 (1986), pp. 511–537.

[68] E. W. Jenkins, C. E. Kees, C. T. Kelley, and C. T. Miller, *An aggregation-based domain decomposition preconditioner for groundwater flow*, SIAM J. Sci. Stat. Comput., 23 (2001), pp. 430–441.

[69] H. Jiang and P. A. Forsyth, *Robust linear and nonlinear strategies for solution of the transonic Euler equations*, Comp. Fluids, 24 (1995), pp. 753–770.

[70] R. W. Johnson, P. R. McHugh, and D. A. Knoll, *High-order scheme implimentation using Newton-Krylov solution methods*, Numer. Heat Trans., Part B, 31 (1997), pp. 295–312.

[71] J. E. Jones and C. S. Woodward, *Newton-Krylov-multigrid solvers for large scale, highly heterogeneous, variably saturated flow problems*, Advances in Water Resources, 24 (2001), pp. 763–774.

[72] D. E. Kaushik, D. E. Keyes, and B. F. Smith, *On the interaction of architecture and algorithm in the domain-based parallelization of an unstructured grid incompressible flow code*, in Proceedings of the Tenth International Conference on Domain Decomposition Methods, AMS, 1998, pp. 311–319.

[73] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[74] C. T. Kelley and D. E. Keyes, *Convergence analysis of pseudo-transient continuation*, SIAM J. Numerical Analysis, 35 (1998), pp. 508–523.

[75] T. Kerkhoven and Y. Saad, *On acceleration methods for coupled nonlinear elliptic systems*, Numer. Math., 60 (1992), pp. 525–548.

[76] D. E. Keyes, *Domain decomposition methods for the parallel computation of reacting flows*, Comp. Phys. Comm., 53 (1989), pp. 181–200.

[77] ——, *Aerodynamic applications of Newton-Krylov-Schwarz solvers*, in Proceedings of the 14th Conf. on Num. Meth. in Fluid Dynamics, Springer, Berlin, 1995, pp. 1–20.

[78] ——, *How scalable is domain decomposition in practice?*, in Proceedings of the 11th International Conference on Domain Decomposition Methods, Domain Decomposition Press, 1999.

[79] D. E. Keyes and W. D. Gropp, *A comparison of domain decomposition techniques for elliptic partial differential equations and their parallel implementation*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. s166–s202.

[80] D. E. Keyes, D. K. Kaushik, and B. F. Smith, *Prospects for CFD on petaflops systems*, in Parallel Solution of Partial Differential Equations, Springer, 1999, pp. 247–278.

[81] D. E. Keyes, D. Keyes, L. McInnes, and W. Samyono, *Using automatic differentiation for second-order matrix-free methods in PDE-constrained optimization*, in Automatic Differentiation Algorithms: From Simulation to Optimization, G. Corliss et al., eds., Springer, 2002, pp. 35–50.

[82] D. E. Keyes and V. Venkatakrishnan, *Newton-Krylov-Schwarz methods: Interfacing sparse linear solvers with nonlinear applications*, Zeitschrift fur angewandte Mathematik und Mechanik, 76 (Suppl. 1) (1996), pp. 147–150.

[83] D. A. Knoll, *Development and Application of a Direct Newton Solver for the Two-Dimensional Tokamak Edge Plasma Fluid Equations*, PhD thesis, University of New Mexico, 1991.

[84] ——, *An improved convection scheme applied to recombining divertor plasma flows*, J. Comput. Phys., 142 (1998), pp. 473–488.

[85] D. A. Knoll and J. U. Brackbill, *The Kelvin-Helmholtz instability, differential rotation, and 3-D, localized, magnetic reconnection*, Physics of Plasmas, (2002, in press).

[86] D. A. Knoll and L. Chacon, *Magnetic reconnection in the two-dimensional Kelvin-Helmholtz instability*, Phys. Rev. Lett., 88 (2002).

[87] D. A. Knoll, D. B. Kothe, and B. Lally, *A new nonlinear solution method for phase-change problems*, Numer. Heat Trans., Part B, 35 (1999), pp. 439–459.

[88] D. A. Knoll, G. Lapenta, and J. Brackbill, *A multilevel Iterative Field Solver for Impilict, Kinetic plasma simulation*, J. Comput. Phys., 149 (1999), pp. 377–388.

[89] D. A. Knoll and P. R. McHugh, *NEWEDGE: A 2-D Fully Implicit Edge Plasma Fluid Code for Advanced Physics and Complex Geometries*, J. Nuc. Mat., 196-198 (1992), pp. 352–356.

[90] ——, *An Inexact Newton Algorithm for Solving the Tokamak Edge Plasma Fluid Equations on a Multiply Connected Domain*, J. Comput. Phys., 116 (1995), pp. 281–291.

[91] ——, *Newton-Krylov methods applied to a system of convection-diffusion-reaction equations*, Comput. Phys. Comm., 88 (1995), pp. 141–160.

[92] ——, *Enhanced nonlinear iterative techniques applied to a nonequilibrium plasma flow*, SIAM J. Sci. Comput., 19 (1998), pp. 291–301.

[93] D. A. Knoll, P. R. McHugh, and D. E. Keyes, *Newton-Krylov methods for low Mach number compressible combustion*, AIAA J., 34 (1996).

[94] D. A. Knoll, P. R. McHugh, S. I. Krasheninnikov, and D. J. Sigmar, *Simulation of dense recombining divertor plasmas with a Navier-Stokes neutral transport model*, Phys. Plasmas, 3 (1996), pp. 293–303.

[95] D. A. Knoll, P. R. McHugh, and V. A. Mousseau, *Newton-Krylov-Schwarz methods applied to the tokamak edge plasma fluid equations*, in Domain-Based Parallelism and Problem Decomposition in Computational Science and Engineering, SIAM, Philadelphia, 1995, pp. 75–96.

[96] D. A. Knoll and V. A. Mousseau, *On Newton-Krylov multigrid methods for the incompressible Navier-Stokes equations*, J. Comput. Phys., 163 (2000), pp. 262–267.

[97] D. A. Knoll, A. K. Prinja, and R. B. Campbell, *A Direct Newton Solver for the Two-Dimensional Tokamak Edge Plasma Fluid Equations*, J. Comput. Phys., 104 (1993), pp. 418–426.

[98] D. A. Knoll and W. Rider, *A multigrid preconditioned Newton-Krylov method*, SIAM J. Sci. Comput., 21 (2000), pp. 691–710.

[99] D. A. Knoll, W. J. Rider, and G. L. Olson, *An efficient nonlinear solution method for non-equilibrium radiation diffusion*, J. Quant. Spectrosc. Radiat. Transfer, 63 (1999), pp. 15–29.

[100] ——, *Nonlinear convergence, accuracy, and time step control in non-equilibrium radiation diffusion*, J. Quant. Spectrosc. Radiat. Transfer, 70 (2001), pp. 25–36.

[101] D. A. Knoll, W. VanderHeyden, V. Mousseau, and D. Kothe, *On preconditioning Newton-Krylov methods in solidifying flow applications*, SIAM J. Sci. Comput., 23 (2002), pp. 381–397.

[102] S. I. Krasheninnikov and et.al., *Plasma recombination in tokamak divertors and divertor simulators*, Physics of Plasmas, 4 (1997), p. 1638.

[103] A. R. Larzelere, *Creating simulation capabilities*, IEEE Computational Science and Engineering, 5 (1988), pp. 27–35.

[104] J. W. MacArthur and S. V. Patankar, *Robust Semidirect Finite Difference Methods for Solving the Navier-Stokes and Energy Equations*, Int. J. Num. Meth. Fluids, 9 (1989), pp. 325–340.

[105] D. Mavriplis, *Multigrid strategies for viscous flow solvers on anisotropic unstructured meshes*, J. Comput. Phys., 145 (1998), p. 141.

[106] ——, *An assessment of linear versus non-linear multigrid methods for unstructured mesh solvers*, J. Comput. Phys., 175 (2002), pp. 302–325.

[107] D. Mavriplis and V. Venkatakrishnan, *Agglomeration multigrid for the two-dimensional viscous flows*, Comp. Fluids, 24 (1995), p. 533.

[108] S. F. McCormick, *Multilevel Adaptive Methods for Partial Differential Equations*, SIAM, 1989.

[109] P. R. McHugh and D. A. Knoll, *Fully implicit finite volume solutions of the incompressible Navier-Stokes and energy equations using inexact Newton's method*, Int. J Num. Meth. Fluids, 18 (1994), pp. 439–455.

[110] ——, *Inexact Newton's method solution to the incompressible Navier-Stokes and energy equations using standard and matrix-free implementations*, AIAA J., 32 (1994).

[111] P. R. McHugh, D. A. Knoll, and D. E. Keyes, *Application of a Newton-Krylov-Schwarz algorithm to low Mach number combustion*, AIAA J., 36 (1998), pp. 290–292.

[112] P. R. McHugh, D. A. Knoll, V. A. Mousseau, and G. A. Hansen, *An investigation of Newton-Krylov solution techniques for low mach number compressible flow*, in Proceedings of the ASME Fluids Engineering Division Summer Meeting, 1995.

[113] J. C. Meza and R. S. Tuminaro, *A multigrid preconditioner for the semiconductor equations*, SIAM J. Sci. Comput., 17 (1996), pp. 118–132.

[114] V. Mousseau and D. A. Knoll, *New physics-based preconditioning of implicit methods for non-equilibrium radiation diffusion*, J. Comput. Phys., (2002, submitted).

[115] V. Mousseau, D. A. Knoll, and J. Reisner, *Nonlinearly consistent method for the shallow water equations*, Mon. Wea. Rev., (2002, in press).

[116] V. Mousseau, D. A. Knoll, and W. Rider, *Physics-based preconditioning and the Newton-Krylov method for non-equilibrium radiation diffusion*, J. Comput. Phys., 160 (2000), pp. 743–765.

[117] V. A. Mousseau and D. A. Knoll, *Fully implicit kinetic solution of collisional plasmas*, J. Comput. Phys., 136 (1997), pp. 308–323.

[118] V. A. Mousseau, D. A. Knoll, and W. J. Rider, *A multigrid Newton-Krylov solver for nonlinear systems*, in Multigrid Methods VI: Lecture Notes in Computational Science and Engineering, Springer, Berlin, 2000, pp. 200–206.

[119] L. S. Mukadi and R. E. Hayes, *Modelling the three-way catalytic converter with mechanistic kinetics using the Newton-Krylov method on a parallel computer*, Comput. Chem. Eng., 26 (2002), pp. 439–455.

[120] W. Mulder and B. V. Leer, *Experiments with implicit upwind methods for the Euler equations*, J. Comput. Phys., 59 (1985), pp. 232–246.

[121] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Analysis and Applications, 13 (1992), pp. 778–795.

[122] U. S. D. of Energy, *Scientific discovery through advanced computing*, tech. rep., U. S. Department of Energy, March 2000.

[123] C. W. Oosterlee and T. Washio, *An evaluation of parallel multigrid as a solver and a preconditioner for singularly perturbed problems*, SIAM J. Sci. Stat. Comput., 19 (1998), pp. 87–110.

[124] E. S. Oran and J. P. Boris, *Numerical Simulation of Reactive Flow*, Elsevier, New York, 1987.

[125] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, Boston, 1970.

[126] S. V. Patankar, *Numerical Heat Transfer and Fluid Flow*, Hemisphere Publishing Corp., New York, 1980.

[127] M. Pernice, *A hybrid multigrid method for the steady-state incompressible Navier-Stokes equations*, Elec. Trans. Num. Anal., 10 (2000), pp. 74–91.

[128] M. Pernice and M. Tocci, *A multigrid-preconditioned Newton-Krylov method for the incompressible Navier-Stokes equations*, SIAM J. Sci. Comput., 23 (2001), pp. 398–418.

[129] M. Pernice and H. F. Walker, *NITSOL: A Newton iterative solver for nonlinear systems*, SIAM J. Sci. Comput., 19 (1998), pp. 302–318.

[130] M. Pernice, L. Zhou, and H. F. Walker, *Parallel solution of nonlinear partial differential equations using a globalized inexact Newton-Krylov-Schwarz method*, Tech. Rep. 48, University of Utah Center for High Performance Computing, 1997.

[131] G. D. Porter and et. al, *Detailed comparison of simulated and measured plasma profiles in the scrape-off layer and edge plasma of DIII-D*, Physics of Plasmas, 7 (2000), pp. 3663–3680.

[132] N. Qin, D. K. Ludlow, and S. T. Shaw, *A matrix-free preconditioned Newton/GMRES method for unsteady Navier-Stokes solutions*, Int. J Num. Meth. Fluids, 33 (2000), pp. 223–248.

[133] P. Rasetarinera and M. Hussaini, *An efficient implicit discontinuous spectral Galerkin method*, J. Comput. Phys., 172 (2001), pp. 718–738.

[134] J. K. Reid, *On the method of conjugate gradients for the solution of large sparse systems of linear equations*, in Large Sparse Sets of Linear Equations, J. K. Reid, ed., New York, 1971, Academic Press, pp. 231–254.

[135] J. Reisner, V. Mousseau, and D. A. Knoll, *Application of the Newton-Krylov method to geophysical flows*, Mon. Wea. Rev., 129 (2001), pp. 2404–2415.

[136] J. Reisner, S. Wynne, L. Margolin, and R. Linn, *Coupled atmospheric-fire modeling employing the method of averages*, Mon. Wea. Rev., 128 (2000), pp. 3683–3691.

[137] J. Reisner, A. Wyszogrodzki, V. Mousseau, and D. A. Knoll, *Physics-based preconditioner for the fully implicit solution of the euler equations: An alternative dynamical core for next generation weather prediction codes*, J. Comput. Phys., (2002, submitted).

[138] W. J. Rider and D. Knoll, *Time step size selection for radiation diffusion calculations*, J. Comput. Phys., 152 (1999), pp. 790–795.

[139] W. J. Rider, D. A. Knoll, and G. Olson, *A multigrid preconditioned Newton-Krylov method for multimaterial equilibrium radiation diffusion*, J. Comput. Phys., 152 (1999), pp. 164–191.

[140] T. D. Rognlien, J. L. Milovich, M. E. Rensink, and G. D. Porter, *A fully implicit, time-dependent 2-d fluid code for modeling tokamak edge plasmas*, J. Nuc. Mat., 196-198 (1992), pp. 347–351.

[141] T. D. Rognlien, X. Q. Xu, and A. C. Hindmarsh, *Application of parallel implicit methods to edge-plasma numerical simulations*, J. Comput. Phys, 175 (2002), pp. 249–268.

[142] Y. SAAD, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Stat. Comput., 14 (1993), pp. 461–469.

[143] ——, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.

[144] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving non-symetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), p. 856.

[145] S. SCHAFFER, *A semi-coarsening multigrid method for elliptic partial differential equations with highly discontinuous and anisotropic coefficients*, SIAM J. Sci. Stat. Comput., 20 (1999), pp. 228–242.

[146] A. SETTARI AND K. AZIZ, *A generalization of the additive correction methods for the iterative solution of matrix equations*, SIAM J. Numer. Anal., 10 (1973), pp. 506–521.

[147] J. N. SHADID AND ET. AL, *Efficient parallel computation of unstructured finite element reacting flow solutions*, Parallel Computing, 23 (1997), pp. 1307–1325.

[148] 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), pp. 155–185.

[149] M. D. SMOOKE, *Solution of Burner-Stabilized Premixed Laminar Flames by Boundary Value Methods*, J. Comput. Phys., 48 (1982), pp. 72–105.

[150] M. D. SMOOKE AND R. M. MATTHEIJ, *On the solution of nonlinear two-point boundary value problems on successively refined grids*, Applied Numerical Mathematics, 1 (1985), pp. 463–487.

[151] M. D. SMOOKE, R. E. MITCHELL, AND D. E. KEYES, *Numerical Solution of Two-Dimensional Axisymmetric Laminar Diffusion Flames*, Combust. Sci. and Tech., 67 (1989), pp. 85–122.

[152] M. D. TIDRIRI, *Schwarz-based algorithms for compressible flows*, Tech. Rep. 96-4, ICASE, Jan. 1996.

[153] ——, *Preconditioning techniques for the Newton-Krylov solution of compressible flows*, J. Comput. Phys., 132 (1997), pp. 51–61.

[154] ——, *Hybrid Newton-Krylov/domain decomposition methods for compressible flows*, in Proceedings of the Ninth International Conference on Domain Decomposition Methods in Sciences and Engineering, 1998, pp. 532–539.

[155] ——, *Development and study of Newton-Krylov-Schwarz algorithms*, Int. J. Comput. Fluid Dyn., 15 (2001), pp. 115–126.

[156] M. D. Tocci, C. T. Kelley, C. T. Miller, and C. E. Kees, *Inexact Newton methods and the method of lines for solving Richards' equation in two space dimensions*, Comp. Geosciences, 2 (1998), pp. 291–309.

[157] U. Trottenberg, A. Schuller, and C. Oosterlee, *Multigrid*, Academic Press, 2000.

[158] R. S. Tuminaro, M. Heroux, S. A. Hutchinson, and J. N. Shadid, *Official Aztec user's guide*, Tech. Rep. SAND99-8801J, Sandia National Laboratory, Dec. 1999.

[159] 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., (2002 (accepted)).

[160] W. B. VanderHeyden, E. D. Dendy, D. Livescu, and N. T. Padial-Collins, *CartaBlanca - an object-oriented Jacobian-Free Newton-Krylov solver environment for multiphase flow with phase change*, SIAM J. Sci. Comput., (2002 (submitted)).

[161] W. B. VanderHeyden, E. D. Dendy, and N. T. Padial-Collins, *Cartablanca-a java-component-based systems simulation tool for coupled non-linear physics on unstructured grids*, in Proceedings of the ACM 2001 Java Grande/ISCOPE Conference, June 2-4, 2001.

[162] S. P. Vanka and G. K. Leaf, *Fully-Coupled Solution of Pressure-Linked Fluid-Flow Equations*, Tech. Rep. ANL-83-73, Argonne National Laboratory, 1983.

[163] V. Venkatakrishnan, *Newton Solution of Invisid and Viscous Problems*, AIAA J., 27 (1989).

[164] ——, *Viscous Computations Using a Direct Solver*, Comp. Fluids, 18 (1990).

[165] V. Venkatakrishnan and D. Mavriplis, *Agglomeration multigrid for the three-dimensional Euler equations*, AIAA J., 33 (1995), p. 633.

[166] R. A. Vessy and D. Steiner, J. Comput. Phys., 116 (1995), pp. 300–313.

[167] H. X. Vu, *Plasma Collection by an Obstacle*, PhD thesis, California Institute of Technology, 1990.

[168] G. Wang and D. K. Tafti, *Performance enhancement on microprocessors with hierarchical memory systems for solving large sparse linear systems*, International J. for Supercomputer Applications and High Performance Computing, 12 (1998 (to appear)).

[169] P. Wesseling, *An Introduction to Multigrid Methods*, John Wiley & Sons, Chichester, 1992.

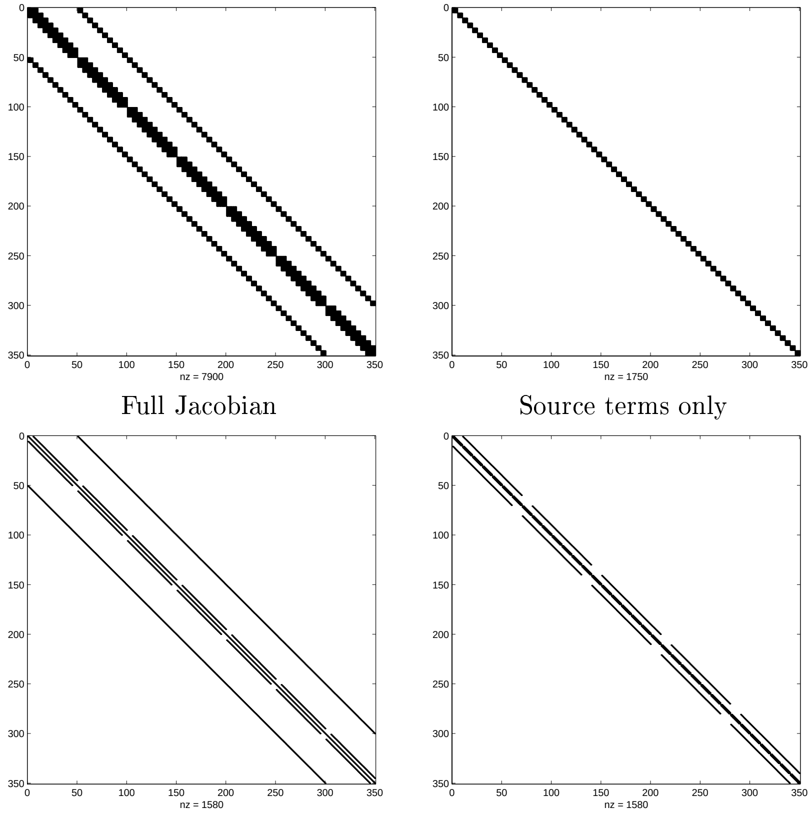
[170] K. A. Winkler, M. L. Norman, and D. Mihalas, *Implicit Adaptive-Grid Radiation Hydrodynamics*, in Multiple Time Scales, J. Brackbill and B. Cohen, eds., Academic Press Inc., 1985.

[171] F. Wising, D. A. Knoll, S. I. Krasheninnikov, T. D. Rognlien, and D. J. Sigmar, *Simulation of the Alcator C-Mod divertor with an improved neutral fluid model*, Contrib. to Plasma Phys., 36 (1996), p. 136.

[172] S. J. Wright and J. Nocedal, *Numerical Optimization*, Springer, 1999.

[173] X. Q. Xu, R. H. Cohen, T. D. Rognlien, and J. Myra, *Low-to-high confinement transition simulations in divertor geometry*, Physics of Plasmas., 7 (2000), p. 1951.

[174] R. Zanino, *Advanced finite element modeling of the tokamak plasma edge*, J. Comput. Phys., 138 (1997), pp. 881–906.



Full Jacobian

Source terms only

Transport terms only
(fields interlaced)

Transport terms permuted
(fields segregated)

Figure 1: Illustration of structure-based operator splitting for a small multi-component Jacobian (five components, five-point stencil, 10 × 7 2D Cartesian grid).

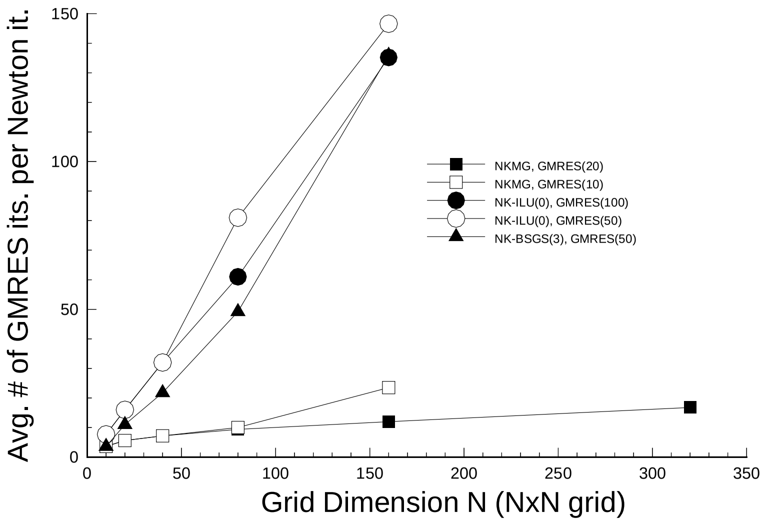


Figure 2: Comparison of NKMG and NK-ILU(0) on 2D Navier-Stokes, varying number of GMRES vectors stored (from [98]).

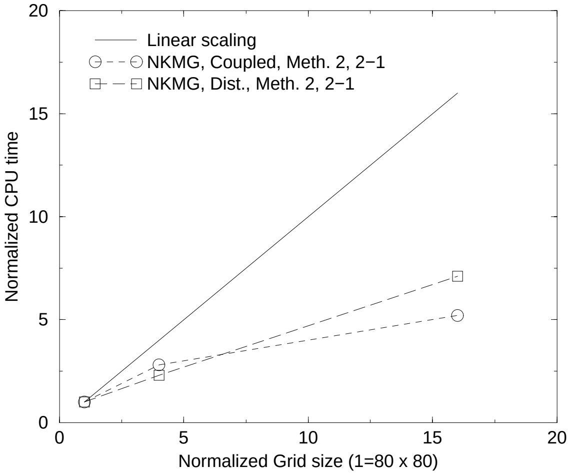


Figure 3: Scaling (CPU time vs problem size) of Newton-Krylov-Multigrid (NKMG) for coupled and distributed preconditioners, for $Gr = 1.0 \times 10^5$ (from [96]).

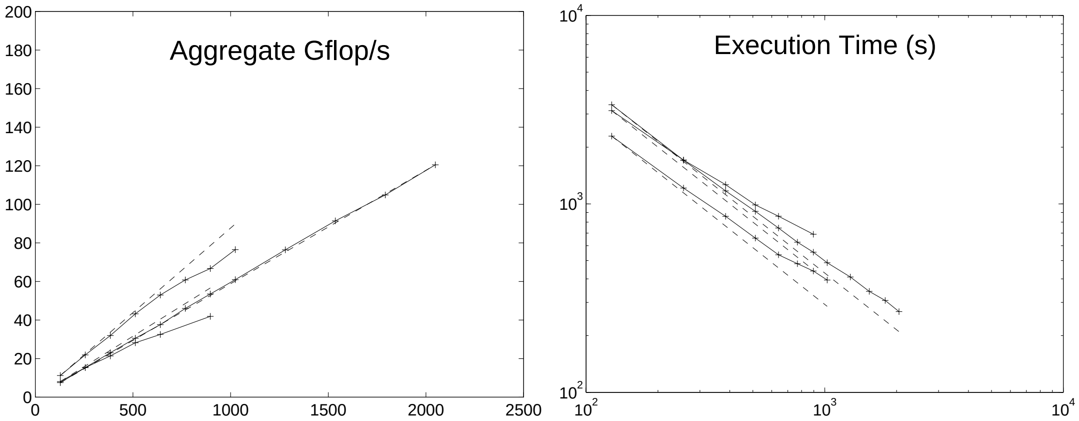


Figure 4: Gigaflop/s ratings and execution times on ASCI Red (up to 3072 dual-processor nodes), ASCI Pacific Blue (up to 768 processors), and a Cray T3E (up to 1024 processors) for a 2.8M-vertex case, along with dashed lines indicating "perfect" scalings, from a baseline of 128 processors (from [60]).