

Nonlinear Krylov-Secant Solvers*

Héctor Klíe, and Mary F. Wheeler

The Center for Subsurface Modeling (CSM)
The Institute for Computational Engineering and Sciences (ICES)
The University of Texas at Austin, Austin, TX, 78712

Contents

1	Introduction	2
2	Newton-Krylov Framework	3
3	The Family of Secant Solvers	5
3.1	Broyden's Method	7
3.2	The Nonlinear Eirola-Nevanlinna (NEN) Method	7
4	Krylov-Secant Updates	9
4.1	Secant Updates constrained to the Krylov subspace	9
4.2	Krylov Secant vs. Standard Secant Update	12
4.3	The Nonlinear Krylov-Eirola-Nevanlinna Algorithm	14
4.4	A high-order Newton-Krylov algorithm	15
5	Implementation issues	15
5.1	Preconditioning	15
5.2	Globalization Strategy and Forcing Terms	16
6	Numerical Experiments	17
7	Conclusions and Further Work	19

Abstract

This report describes a new family of Newton-Krylov methods for solving nonlinear systems of equations arising from the solution of Richards' equation and in fully implicit formulations in air-water systems. The basic approach is to perform secant (Broyden) updates restricted to the Krylov subspace generated by the GMRES iterative solver. This approach is introduced as Krylov-secant methods. One of the most attractive features of these methods is their performance of sequence of rank-one updates without explicitly recalling the computation or action of the Jacobian matrix. Implications of these updates in line-search globalization strategies, computation of dynamic

*Keywords and phrases: Krylov-secant, Broyden's method, Newton-Krylov, GMRES, Arnoldi factorization, rank-one updates;
Author's e-mail: {klie,mfw}@ices.utexas.edu; UT, Austin.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2006		2. REPORT TYPE		3. DATES COVERED 00-00-2006 to 00-00-2006	
4. TITLE AND SUBTITLE Nonlinear Krylov-Secant Solvers				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Center for Subsurface Modeling (CSM),The Institute for Computational Engineering and Sciences (ICES),The University of Texas at Austin,Austin,TX,78712				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 24	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

tolerances (forcing terms) and the use of preconditioning strategies are presented. Numerical results show improvements over traditional implementations.

1 Introduction

It is well known by environmental engineers that solving Richards's equation can be a challenging problem due to the high nonlinearities induced by pressure and saturation dependent coefficients in the temporal and spatial terms [21]. Moreover, when sufficient computer power is available for solving more complex air-water groundwater applications, fully implicit formulations are preferred over IMPES formulations since they are unconditionally stable and allow the performance of larger simulation timesteps [3, 11, 14]. These timesteps are numerically constrained to Newton's radius of local convergence, and the cost of performing each one of them may be computational high.

On the other hand, recent advances in Newton-Krylov methods [17, 19] have brought up new opportunities for environmental agencies to speedup the computation in fully implicit implementations [4, 16, 17]. Nowadays, the use of Krylov iterative methods [27, 31] such as Orthomin, QMR, BiCGSTAB and GMRES provides powerful means to implement inexact versions of Newton's method for solving large scale problems in an efficient way.

Despite these rapid advances, linear preconditioning is still a major issue for fully implicit implementations in multiphase flow due to the coupling of different physical variables [4, 15, 16]. Nevertheless, it is clear that improving the nonlinear convergence may result in significant savings in CPU time since each nonlinear iteration typically involves dozens to hundreds of linear iterations.

The present work proposes an efficient way to exploit the underlying Krylov subspace information from the linear solver in order to accelerate the nonlinear convergence of Newton-Krylov methods. The basic approach is to perform secant (Broyden) updates restricted to the Krylov subspace generated by the GMRES iterative solver, namely, updates to the Hessenberg matrix resulting from the Arnoldi factorization. This approach is introduced as a new family of Krylov-secant methods. One of their most attractive features, beside achieving higher rates of nonlinear convergence, is the performance of sequences of rank-one updates without explicitly recalling the computation or action of the Jacobian matrix. The procedure may be described as the composition of several chord steps for every single Newton

step in a predictor-corrector fashion.

Therefore, in contrast to other fully implicit implementations for ground-water flow implementations this work focuses on reducing the number of nonlinear iterations at a low computational cost. In fact, Krylov-secant methods may be implemented as a practical extension to the Newton-Krylov solver framework. We discuss implications of these updates in the formulation of line-search globalization strategies, computation of dynamic tolerances (forcing terms) and the use of preconditioning strategies.

Numerical results show improvements over traditional implementations, especially in those cases where nonlinearities increase due to stringent pressure and saturation changes throughout the simulation.

Further potentials of the proposed method are devised with the current advances in augmenting, reorthogonalizing and shifting the Krylov basis information [5, 23, 26].

2 Newton-Krylov Framework

We are interested in solving the following nonlinear system of equations

$$F(u) = 0, \quad (1)$$

where $F : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$. A variety of different discretization techniques applied to nonlinear partial differential equations give rise to an algebraic system such as (1), where u is the unknown vector representing the nodal values of the state variable.

In our particular case, this nonlinear system of equations arises from discretizing either Richards' equation or the coupled partial differential equations describing the air-water system. The nonlinearities in both equations are present in the temporal term and the diffusive term, and, when fully implicit schemes are employed, the transient and steady-state cases lead to a nonlinear algebraic system such as (1). Newton's method implies the solution of the Jacobian system

$$J(u) \delta u = -F(u), \quad (2)$$

where $F(u)$ and $J(u)$ denote the evaluation of the function at $u \in \mathbb{R}^n$ and its derivative at any Newton step, respectively.

Due to the high cost or absence of derivatives for the explicit construction of the Jacobian J , the action of this operator onto a vector (in the form of matrix-vector products) may be performed by finite differences. This is, in fact, one of the most appealing features of Newton-Krylov methods. These

methods are also amenable to globalization by line-search or trust region strategies and for dynamic control of linear tolerances (i.e., *forcing terms*). The reader is referred to [19] to see a state-of-the-art discussion of the subject. Specific theory supporting the description of the line-search backtracking method and the selection of the forcing terms to control the convergence of the linear solver at each nonlinear step may be found in [10, 17, 25]. A globalized Newton-Krylov method with a line-search backtracking procedure may be stated as follows:

Algorithm 2.1 (Newton-Krylov algorithm with forcing terms and line-search backtracking)

1. Let $u^{(0)}$ be an initial guess.
2. For $k = 0, 1, 2, \dots$ until convergence do
 - 2.1 Choose $\eta^{(k)} \in [0, 1)$, $t \in (0, 1)$.
 - 2.2 Using some Krylov subspace iterative method, compute a vector $s^{(k)}$ satisfying

$$J^{(k)} s^{(k)} = -F^{(k)} + r^{(k)}, \quad (3)$$
 with $\frac{\|r^{(k)}\|}{\|F(u^{(k)})\|} \leq \eta^{(k)}$.
 - 2.3 While $\|F(u^{(k)} + s^{(k)})\|_2 > [1 - t(1 - \eta^{(k)})] \|F^{(k)}\|$ do
 - 2.3.1 Choose λ to minimize a quadratic polynomial over $[\underline{\lambda}, \bar{\lambda}] \subset (0, 1)$ that interpolates

$$F(\lambda) = \|F(u^{(k)} + \lambda s^{(k)})\|^2.$$
 - 2.3.2 Update $s^{(k)} = \lambda s^{(k)}$.
 - 2.4 Update $\eta^{(k)} = 1 - \lambda(1 - \eta^{(k)})$.
3. Update solution $u^{(k+1)} = u^{(k)} + s^{(k)}$.

We can make the following observations:

- The loop at step 2.3 represents a global backtracking based on the Goldstein-Armijo condition [7]. In practice, the parameter t is selected on the order of 10^{-4} . The interval $[\underline{\lambda}, \bar{\lambda}] = [.1, .5]$ is typically

chosen, and the quadratic interpolant p is constructed in such a way that $p(0) = F(0) = \|F^{(k)}\|^2$, $p(1) = F(1) = \|F^{(k)} + s^{(k)}\|^2$ and $p'(0) = F'(0) = 2\langle F^{(k)}, J^{(k)}s^{(k)} \rangle$.

- The forcing term $\eta^{(k)}$ is selected according to the following criterion:

$$\eta^{(k)} = \min \left\{ \eta_{\max}, \max \left\{ \tilde{\eta}^{(k)}, \left(\eta^{(k-1)} \right)^2 \right\} \right\}, \quad (4)$$

where

$$\tilde{\eta}^{(k)} = \frac{\left| \|F^{(k)}\| - \|F^{(k-1)} + J^{(k-1)}s^{(k-1)}\| \right|}{\|F^{(k-1)}\|}. \quad (5)$$

The choice of $\tilde{\eta}^{(k)}$ reflects the agreement between F and its linear model at the previous iteration, and it depends on how λ is chosen. Thus, the linear solver tolerance is larger when the Newton step is less likely to be productive and smaller when the step is more likely to lead to a good approximation. This operation is performed once the backtracking procedure has returned the adequate nonlinear step size to update the solution. Fig. 1 illustrates the savings in computation using the forcing terms for a fully-implicit formulation for a 3-D two-phase problem. The staircase shape displayed by an inexact Newton method with a fixed forcing term suggests the amount of over-computation in generating decreasing Newton directions. In contrast, the criterion given by (4) and (5) avoids flat portions of the curve resulting then in a significant overall saving of GMRES iterations (about 400 iterations).

3 The Family of Secant Solvers

Rank-one updates for solving nonlinear equation are sometimes referred to as secant or *quasi-Newton methods* since no “true” Newton equation is ever formed throughout all iterations. These updates obey a secant condition that must be satisfied by the new Jacobian approximation. The best of these methods still seems to be the first one originally introduced by Broyden (see e.g., [7]).

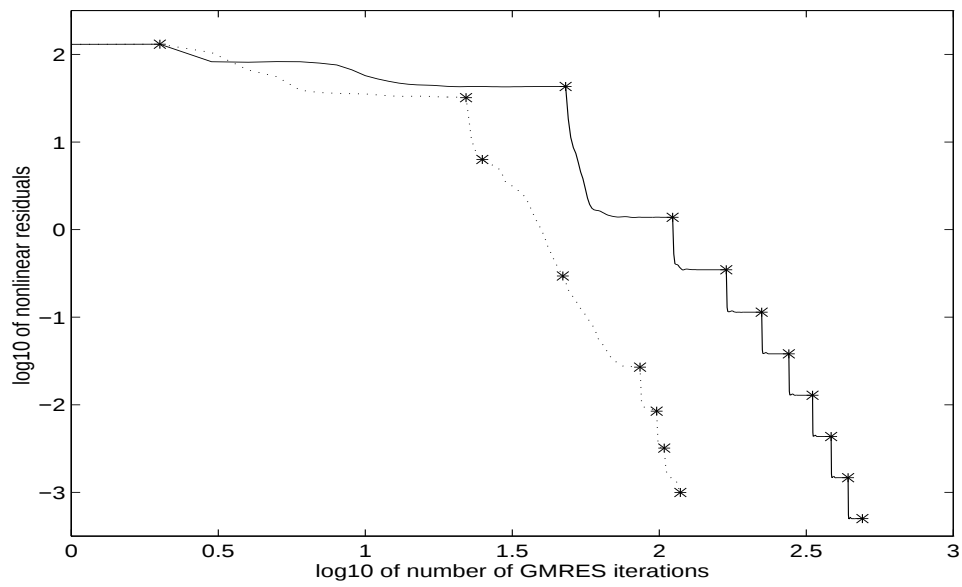


Figure 1: The use of the forcing term criterion for dynamic control of linear tolerances. The solid line represents a standard inexact Newton implementation with fixed linear tolerances 0.1. The dotted line is the inexact Newton implementation with the forcing term criterion. Each symbol * indicates the start of a new Newton iteration.

3.1 Broyden's Method

Given $u \approx u^*$ and $A \approx J(u)$, we can find an approximate new Newton step, u^+ , by

$$u^+ = u - A^{-1}F(u). \quad (6)$$

Broyden's method computes a new A^+ by means of the following rank-one update

$$A^+ = A + \frac{[F^+(u) - F(u) - As]g^t}{g^t s}, \quad (7)$$

that, whenever the approximated Jacobian equation is solved exactly, i.e. $As = -F(u)$, reduces to

$$A^+ = A + \frac{F^+(u)g^t}{g^t s},$$

for $g^t s \neq 0$. The vector g may be chosen in several ways. For instance, when $g \equiv s$, we obtain the “good Broyden's update” and when $g \equiv A^t[F^+(u) - F(u)]$, we have the “bad Broyden's update”.

One of the key features of Broyden's method is to provide a sparse approximation to the true Jacobian that may be exploited in structure. In fact, there are efficient procedures to factorize such an operator under low-rank modifications [7].

In the setting of groundwater flow problems, Broyden's method may be a convenient option to avoid the computation of several hundred thousands to millions of entries in the Jacobian matrix. However, caution must be taken when the updated operator A^+ becomes numerically singular. In that case, the Jacobian must be reconstructed at the new point to resume the nonlinear iterations.

This algorithm converges q -superlinearly to $F(u^*)$ if the rate at which the Jacobian is approximated is faster than the reduction in the nonlinear step (i.e., the Dennis-Moré characterization [7]).

3.2 The Nonlinear Eirola-Nevanlinna (NEN) Method

The nonlinear Eirola-Nevanlinna (NEN) was proposed by [34] as the nonlinear counterpart to the linear EN algorithm [9]. This method is a generalization of Broyden's method for solving linear/nonlinear equations and has received particular attention in recent years [13, 32]. Interestingly enough, the algorithm has connections with the GMRES algorithm [31]. The NEN algorithm may be regarded as the composition of two Broyden's iterations as the following presentation suggests:

Algorithm 3.1 (Nonlinear EN)

1. Give an initial guess $u^{(0)}$ and Jacobian approximation $A^{(0)}$.
2. For $k = 0, 1, \dots$ until convergence do
 - 2.1 Solve $A^{(k)} s^{(k)} = -F^{(k)}$.
 - 2.2 $q^{(k)} = F^{(k+1)} - F^{(k)}$.
 - 2.3 $A^{(k+1)} = A^{(k)} + \frac{(q^{(k)} - A^{(k)} s^{(k)})(s^{(k)})^t}{(s^{(k)})^t s^{(k)}}$.
 - 2.4 Solve $A^{(k+1)} \tilde{s}^{(k)} = -F^{(k)}$.
 - 2.5 Update solution $u^{(k+1)} = u^{(k)} + \tilde{s}^{(k)}$.

Note that each of the iterations involves the solution of two linear systems. However, we still maintain one rank-one update and one function evaluation per step as in Broyden's method. It can be shown that the update at step 2.5 can be expressed as

$$\begin{aligned} u^{(k+1)} &= u^{(k)} + \tilde{s}^{(k)} \\ &= u^{(k)} + s^{(k)} + \theta^{(k)} \hat{s}^{(k)}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} s^{(k)} &= -\left(A^{(k)}\right)^{-1} F^{(k)}, \\ \hat{s}^{(k)} &= -\left(A^{(k)}\right)^{-1} F\left(u^{(k)} + s^{(k)}\right), \end{aligned}$$

and

$$\theta^{(k)} = \frac{1}{1 - \frac{(s^{(k)})^t \hat{s}^{(k)}}{(s^{(k)})^t s^{(k)}}},$$

provided that $(s^{(k)})^t s^{(k)} \neq 0$. Furthermore,

$$u^{(k+1)} = u^{(k)} - \left(A^{(k)}\right)^{-1} \left[F^{(k)} + \theta^{(k)} F\left(u^{(k)} - \left(A^{(k)}\right)^{-1} F^{(k)}\right) \right], \tag{9}$$

The scalar parameter $\theta^{(k)}$ represents the acute angle between the directions generated by step 2.1 and a chord step. Note that if the directions $s^{(k)}$ and $\hat{s}^{(k)}$ are mutually orthogonal then the chord step is taken in full. The composition of steps of this kind has been devised as a mechanism to generate

higher-order nonlinear algorithms [17, 24]. In fact, if $A^{(k)} = J^{(k)}$ and $\theta^{(k)} = 1$ for $k = 0, 1, \dots$ then (9) becomes

$$u^{(k+1)} = u^{(k)} - \left(J^{(k)}\right)^{-1} \left[F^{(k)} + F \left(u^{(k)} - \left(J^{(k)}\right)^{-1} F^{(k)} \right) \right], \quad (10)$$

for $k = 0, 1, \dots$. This recurrence represents a higher-order modification of Newton's method. Iterates generated by (10) converge q -superlinearly with q -order 3 [24]. These methods were studied by Shamanskii [28] and Traub [29]. They pointed out that even higher-order methods can be built out of a longer sequence of chord steps alternated with regular Newton steps. In a more recent treatment, Kelley names those methods after Shamanskii and compares the particular case (10) numerically against Newton's method [17].

Along the lines of Gay's local convergence analysis for Broyden's method, Yang shows that the NEN algorithm converges n -step q -quadratically for n -dimensional problems [34]. Therefore, as in the linear case, the NEN method converges twice as fast as Broyden's method. In addition, efficient limited memory BFGS implementations lately proposed for Broyden's method apply here as well [22, 30].

4 Krylov-Secant Updates

Let $A^{(k)}$ be an approximation of the current Jacobian $J^{(k)}$. We are interested in looking at a minimum change of $A^{(k)}$ consistent with the secant equation $A^{(k)} s^{(k)} = F^{(k+1)} - F^{(k)}$ and restricted to the underlying Krylov subspace. A basis for this subspace arises as a result of using a linear iterative solver such as GMRES for solving the approximated Jacobian system associated with $A^{(k)}$.

4.1 Secant Updates constrained to the Krylov subspace

The Arnoldi factorization provides valuable information that should not be discarded at every GMRES restart. This observation has been a cornerstone of several efforts to initiate the reuse or recycling of Krylov basis information to generate improved restarted versions of GMRES [5, 23, 26]. We show how to reflect secant updates on the Jacobian matrix without altering the current Krylov basis. For the sake of simplicity, let us omit the sources of inexactness induced by the use of GMRES whose relative residuals are supposed to converge at a predefined tolerance (i.e., to a prescribed forcing term value).

For a given nonlinear iteration k , recall that the Arnoldi factorization for $A^{(k)}$ is given by [27]

$$A^{(k)}V_l^{(k)} = V_l^{(k)}H_l^{(k)} + f_{l+1}^{(k)}e_l^t, \quad (11)$$

where $V_l^{(k)} \in \mathbb{R}$ holds the Krylov orthogonal basis along its columns, $H_l^{(k)} \in \mathbb{R}$ is an upper Hessenberg matrix and $f_{l+1}^{(k)}$ is a residual vector orthogonal to $V_l^{(k)}$. Equation (11) may also be expressed in the more compact form:

$$A^{(k)}V_l^{(k)} = V_{l+1}^{(k)}\overline{H}_l^{(k)}, \quad (12)$$

with

$$\overline{H}_l^{(k)} = \begin{pmatrix} H_l^{(k)} \\ \|f_{l+1}^{(k)}\|e_l^t \end{pmatrix},$$

and

$$V_{l+1}^{(k)} = \left(V_l^{(k)} \mid \frac{f_{l+1}^{(k)}}{\|f_{l+1}^{(k)}\|} \right).$$

Consider the solution of the following approximated Jacobian equation at the k -th nonlinear iteration

$$A^{(k)}s^{(k)} = -F^{(k)}, \quad (13)$$

with the GMRES algorithm. This linear solution may be realized as part of an inexact Newton or Broyden's method. Let $s_l^{(k)} = s_0^{(k)} + V_l^{(k)}y^{(k)}$ be the solution obtained after l steps within the final GMRES restart cycle. The associated Krylov subspace for this problem is given by $\mathcal{K}_l^{(k)}(A^{(k)}, r_0^{(k)})$. Now, we seek to use the information gathered during the solution of (13) to provide an approximation to the system

$$A^{(k+1)}s^{(k+1)} = -F^{(k+1)}. \quad (14)$$

Equation (14) may be seen as a perturbed version of linear system (13) and hence, in general, we can not guarantee that $\mathcal{K}_l^{(k+1)}(A^{(k+1)}, r_0^{(k+1)}) = \mathcal{K}_l^{(k)}(A^{(k)}, r_0^{(k)})$. However, rank-one updates to the Arnoldi factorization of (13) may be done without destroying the Krylov basis. That is,

$$\left(A^{(k)} + V_l^{(k)}zw^t \left(V_l^{(k)} \right)^t \right) V_l^{(k)} = V_l^{(k)} \left(H_l^{(k)} + zw^t \right) + h_{l+1,m}^{(k)} v_{l+1}^{(k)} e_l^t, \quad (15)$$

or equivalently,

$$\left(V_l^{(k)} \right)^t \left[A^{(k)} + V_l^{(k)}zw^t \left(V_l^{(k)} \right)^t \right] V_l^{(k)} = H_l^{(k)} + zw^t, \quad (16)$$

for any vectors $z, w \in \mathbb{R}^l$. Expression (16) suggests a clearer way to update $H_l^{(k)}$ rather than $A^{(k)}$. Note that the current Jacobian approximation appears to be updated by a rank-one matrix whose range lies on $\mathcal{K}_l^{(k)}(A^{(k)}, r_0^{(k)})$.

In order to proceed, it would be convenient to express the secant equation

$$A^{(k+1)}s^{(k)} = F^{(k+1)} - F^{(k)}, \quad (17)$$

in terms of a solution strictly lying on the Krylov subspace. In addition, if the initial guess to GMRES is other than zero, we may rewrite (17) in terms of a system whose solution lies on the Krylov subspace. That is, equation (13) would satisfy

$$A^{(k)}s^{(k)} = -F^{(k)} - A^{(k)}s_0^{(k)} = r_0^{(k)}, \quad (18)$$

with $s_l^{(k)} = V_l^{(k)}y^{(k)}$, and thus, the corresponding secant equation (17) would become

$$A^{(k+1)}s^{(k)} = F^{(k+1)} - r_0^{(k)}. \quad (19)$$

Multiplying both sides by $(V_l^{(k)})^t$, it follows that $H_l^{(k+1)}$ should satisfy the secant equation

$$H_l^{(k+1)}y^{(k)} = (V_l^{(k)})^t F^{(k+1)} + \beta e_1, \quad (20)$$

where $\beta = \|r_0^{(k)}\|$. Therefore, the Krylov subspace projected version of the secant equation (17) can be written as

$$H_l^{(k+1)} = H_l^{(k)} + \frac{\left((V_l^{(k)})^t F^{(k+1)} + \beta e_1 - H_l^{(k)}y^{(k)} \right) (y^{(k)})^t}{(y^{(k)})^t y^{(k)}}. \quad (21)$$

This update in terms of $A^{(k)}$ is as follows

$$A^{(k+1)} = A^{(k)} + P \frac{\left(q^{(k)} - A^{(k)}s^{(k)} \right) (s^{(k)})^t}{(s^{(k)})^t s^{(k)}}, \quad (22)$$

with $P = V_l^{(k)} (V_l^{(k)})^t$. We refer to the update (21) as the *Krylov-secant update*. Note that the operator $P^{(k)}$ is an orthogonal projector onto the Krylov subspace $\mathcal{K}_m(A^{(k)}, r_0^{(k)})$. That is,

- $(P^{(k)})^2 = P^{(k)}$ (Idempotency).

- $\left(P^{(k)}\right)^t = P^{(k)}$ (Symmetry).
- $\text{Range}\left(P^{(k)}\right) = \mathcal{K}_m\left(A^{(k)}, r_0^{(k)}\right)$.

4.2 Krylov Secant vs. Standard Secant Update

It can be shown that expression (22) corresponds to the solution of the following minimization problem

$$\min_{B \in \mathbb{R}^{n \times n}} \|B - A^{(k)}\|_F \quad \text{subject to} \quad \left(P^{(k)} B P^{(k)}\right) s^{(k)} = P^{(k)} F^{(k+1)} + r_0^{(k)},$$

where $\|\cdot\|_F$ stands for the Frobenius norm. In other words, the Krylov-secant update produces the closest matrix in Frobenius norm satisfying the secant equation that lies on the Krylov subspace $\mathcal{K}_l^{(k)}\left(A^{(k)}, r_0^{(k)}\right)$. This problem has the equivalent formulation in terms of an Hessenberg matrix, namely,

$$\min_{G \in \mathbb{R}^{m \times m}} \|G - H_m^{(k)}\|_F \quad \text{subject to} \quad Gy = \left(V^{(k)}\right)^t F^{(k)} + \beta e_1.$$

A geometric interpretation may be drawn out from this fact. Consider $\mathcal{Q}$, the set of matrix quotients of $q = F^{(k+1)} - F^{(k)}$ by $s = s^{(k+1)} - s^{(k)}$ defined by

$$\mathcal{Q} = \{B \in \mathbb{R}^{n \times n} \mid Bs = y\}, \quad (23)$$

and $\mathcal{X}$, the set of matrices generating the same Krylov subspace $\mathcal{K}_m \equiv \mathcal{K}_m\left(A^{(k)}, r_0^{(k)}\right)$. That is,

$$\mathcal{X} = \{B \in \mathbb{R}^{n \times n} \mid \mathcal{K}_m(B, r_0^{(k)}) = \mathcal{K}_m\}. \quad (24)$$

The resulting matrix $A^{(k+1)}$ in (22) may be viewed as the nearest to $A^{(k)}$ in $\mathcal{X}$ among the set of nearest matrices in $\mathcal{X}$ to $\mathcal{Q}$. Furthermore, if the intersection of these two sets is not empty, then $A^{(k+1)} \in \mathcal{X} \cap \mathcal{Q}$. This observation is key to the construction of least-change secant updates consistent with operators satisfying the standard secant condition and other properties prescribed by a given affine subspace (e.g., sparsity pattern, positive definiteness) in $\mathbb{R}^{n \times n}$ (see [7, 8]).

Figure 2 provides a geometric interpretation of Krylov-secant updates and standard secant updates. Here, A_K^+ represents the new approximation obtained by (22), A_B^+ represents the new approximation obtained by (7), and J is the Jacobian operator at the new point. Depending on the behavior of the function F , either of these two updates will yield an operator closer to the Jacobian.

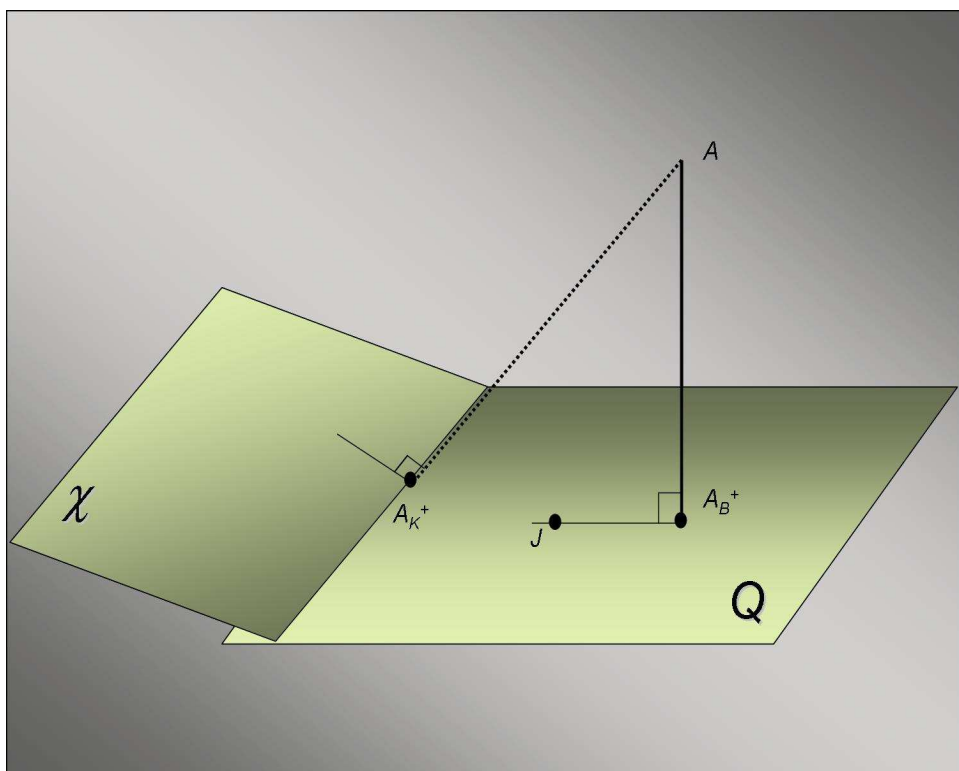


Figure 2: Geometric interpretation of Krylov-secant updates as compared to standard secant updates.

4.3 The Nonlinear Krylov-Eirola-Nevanlinna Algorithm

We are now in a position to describe an inexact NEN algorithm that exploits the information left behind by the GMRES method. We introduce the nonlinear Krylov-Eirola-Nevanlinna (nonlinear KEN) algorithm as follows:

Algorithm 4.1 (Nonlinear KEN)

1. Give an initial guess $u^{(0)}$ and a Jacobian approximation $A^{(0)}$.
2. For $k = 0, 1, \dots$ until convergence do
 - 2.1 $[s^{(k)}, y^{(k)}, H_l^{(k)}, V_l^{(k)}, h_{m+1,m}^{(k)}, \beta \equiv \|r_0^{(k)}\|] = \text{GMRES}(A^{(k)}, -F^{(k)}, s^{(k)})$.
 - 2.2 $q^{(k)} = (V_l^{(k)})^t F^{(k+1)} + \beta e_1$.
 - 2.3 $H_l^{(k+1)} = H_l^{(k)} + \frac{(q^{(k)} - H_l^{(k)} y^{(k)}) (y^{(k)})^t}{(y^{(k)})^t y^{(k)}}$.
 - 2.4 Solve

$$\min_{y \in \mathcal{K}_l(A^{(k)}, r_0^{(k)})} \|\beta e_1 + \overline{H}_l^{(k+1)} y\|,$$
 with $\overline{H}_l^{(k+1)} = \begin{pmatrix} H_l^{(k+1)} \\ h_{m+1,m}^{(k)} e_l^t \end{pmatrix}$.
 - 2.5 $\tilde{s}^{(k)} = V_l^{(k)} \tilde{y}^{(k)}$.
 - 2.6 $u^{(k+1)} = u^{(k)} + \tilde{s}^{(k)}$.
 - 2.7 Update $A^{(k)}$ or construct $A^{(k+1)}$.

Note that the basic difference between the NEN algorithm and the nonlinear KEN algorithm is the way in which the composed nonlinear step is computed. In the nonlinear KEN algorithm there are two updates: one in n -th dimensional space and another restricted to the Krylov subspace of dimension l . The computational implications are evident. Whereas the NEN algorithm solves two, generally large, linear systems of the same size, the nonlinear KEN solves one large and one significantly smaller system (since, in general, $l \ll n$). For the sake of updating the Jacobian approximation, step 2.7 in the nonlinear KEN algorithm may be implemented as step 2.3 of

the NEN algorithm. Another option is to form $A^{(k+1)}$ from scratch at the new nonlinear step so that a true Newton step is composed with this Krylov secant update.

As one may expect from the discussion of the NEN algorithm leading to the update (9), it may be shown that the resulting composite step is the Broyden or Newton method with a chord step constrained to the Krylov basis.

4.4 A high-order Newton–Krylov algorithm

An even faster version of the nonlinear KEN algorithm may be attained by performing exhaustive rank-one updates of the Hessenberg matrix before making the next GMRES call. The extent of these updates is determined by the capability of the residual minimization problem to produce descent directions. Hence, we stop performing Hessenberg updates when we are unable to generate a sufficient decrease of $\|F\|$. Details on this procedure were described in [33].

5 Implementation issues

Since Krylov-secant updates imply several operations on a restricted basis (or smaller space), computations may be carried out in an efficient fashion. We proceed to mention the implications of Krylov-secant updates for preconditioning, globalization and forcing term strategies.

5.1 Preconditioning

Expression (16) shows that if a preconditioner is used, the Krylov-secant update (21) will reflect the implicit joint update of the Jacobian times its preconditioner. More precisely, if $M^{(k)}$ is an operator whose inverse approximates the inverse of the Jacobian (or an approximate Jacobian operator), then the Krylov-secant update for right preconditioning implies

$$\left(AM^{-1}\right)^{(k+1)} = A^{(k)} \left(M^{(k)}\right)^{-1} + P \frac{\left(q^{(k)} - A^{(k)}s^{(k)}\right) \left(\bar{s}^{(k)}\right)^t}{\left(\bar{s}^{(k)}\right)^t \bar{s}^{(k)}}, \quad (25)$$

with $\bar{s}^{(k)} = M^{(k)} s^{(k)}$. On the other hand, left preconditioning reads as follows

$$\left(M^{-1}A\right)^{(k+1)} = \left(M^{(k)}\right)^{-1} A^{(k)} + P \frac{\left(q^{(k)} - A^{(k)} s^{(k)}\right) \left(s^{(k)}\right)^t}{\left(s^{(k)}\right)^t s^{(k)}}. \quad (26)$$

Note that, in the latter case, there are no changes in the search direction with respect to the unpreconditioned problem. These two implicit joint updates must be treated with care in order to maintain consistency in the computations. The key point is that, in general, one-rank updates do not distribute with respect to the product of matrices and therefore we are unable to guarantee that either

$$\left(AM^{-1}\right)^+ = A^+ \left(M^+\right)^{-1} \quad \text{or} \quad \left(M^{-1}A\right)^+ = \left(M^+\right)^{-1} A^+.$$

In addition, particularly for multiphase groundwater flow purposes, computation of an updated version of the preconditioner may be a time consuming task or even impossible to carry out when there is not an explicit form of such preconditioner as it happens in two-stage preconditioning [6, 20] multilevel [4] or in general inner-outer iterations [2]. Assuming that the preconditioner is fixed after each Krylov-secant update seems to be the simplest and most efficient strategy, even though this may introduce some extra inaccuracies in the computations [18].

5.2 Globalization Strategy and Forcing Terms

The utilization of GMRES offers the opportunity to avoid explicit representation of the Jacobian operator after each Krylov-secant update. To implement the globalization strategy we must construct the interpolating polynomial that satisfies

$$\begin{aligned} \frac{1}{2}F'(0) &= \frac{1}{2} \langle F, Js_m \rangle = \langle F, J(s_0 + V_m y_m) \rangle \\ &= \langle F, -F - r_0 \rangle + \left\langle F, V_{m+1} \overline{H}_m y_m \right\rangle \\ &= -\|F\|^2 - \left\langle F, V_{m+1} (\beta e_1 - \overline{H}_m y_m) \right\rangle. \end{aligned} \quad (27)$$

Hence, no explicit reference to the Jacobian is required. Note that (27) simplifies to $\langle F, Js_m \rangle = -\|F\|^2 + \|r_m\|^2$ when $s_0 = 0$. Furthermore, to

compute residuals associated with the damped nonlinear step we need only to compute

$$\begin{aligned} \left\| F^{(k)} + J^{(k)} \left(\lambda^{(k)} s^{(k)} \right) \right\|^2 &= \left\| F^{(k)} \right\|^2 + 2\lambda^{(k)} \left\langle F^{(k)}, J^{(k)} s^{(k)} \right\rangle \\ &\quad + \left(\lambda^{(k)} \left\| J^{(k)} s^{(k)} \right\| \right)^2, \end{aligned} \quad (28)$$

where

$$\left\| J^{(k)} s^{(k)} \right\|^2 = \left\| r_m^{(k)} \right\|^2 - 2 \left\langle F^{(k)}, J^{(k)} s^{(k)} \right\rangle - \left\| F^{(k)} \right\|^2. \quad (29)$$

Finally, note that the Jacobian operator may also be bypassed in the forcing term computation, since

$$\tilde{\eta}^{(k)} = \frac{\left\| \left\| F^{(k)} \right\| - \left\| \beta e_1 - \overline{H}_m y \right\| \right\|}{\left\| F^{(k-1)} \right\|} = \frac{\left\| \left\| F^{(k)} \right\| - \beta \|q_{m+1}^t e_1\| \right\|}{\left\| F^{(k-1)} \right\|}. \quad (30)$$

6 Numerical Experiments

Now we looked at a two-phase problem consisting of wetting and a non-wetting phases. Primary variables are non-wetting phase pressures and water concentrations. The model consists of a bottom hole pressure specified water injection well at coordinate (1,1) of the plane and a bottom hole pressure specified production well in the opposite corner. All terms are solved implicitly.

Absolute permeabilities are set according the following diagonal tensor $\bar{K}$ (in milidarcies)

$$\bar{K} = \begin{pmatrix} 100 & 0 & 0 \\ 0 & 500 & 0 \\ 0 & 0 & 50 \end{pmatrix}.$$

Grid sizes are defined as $DX = DY = 20$ ft and $DZ = 2$ ft. The timestep increase factor is set to 1.25 in order to reach rapidly the maximum DT size of 45 days. This allows the production of appreciable changes in pressures and saturations and cause the activation of the several backtracking steps. Table 1 shows the rest of the input parameters as specified in this simulation study. Figure 3 show the relative permeability curves and capillary pressure associated with the case model.

We use the Gauss-Seidel two-stage preconditioner [18] for solving the linear coupled system. In brief, this preconditioner consists of a decoupling stage defined by the inverse diagonals blocks of the coefficients representing

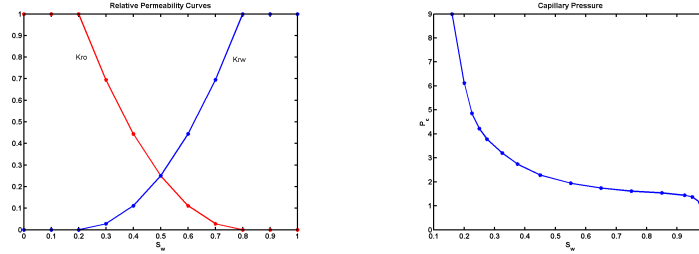


Figure 3: Relative permeability and capillary pressure curves.

the contribution from oil pressures, water saturation and their interaction within the same grid block. The second stage consists of computing an iterative solution of pressures and saturations from the decoupled system. The solutions are combined in a Gauss-Seidel fashion and hence, the whole procedure entails an inner-outer iteration for the solution of the original coupled system. Numerical experiments were carried out on a Linux-based cluster consisting of 45 dual processors and interconnected through Myrinet cards.

Table 1: Simulation input data.

Simulation Time (days)	500
Max. linear solver tolerance	1×10^{-5}
Nonlinear solver tolerance	5.0×10^{-3}
Number of gridblocks	$48 \times 48 \times 8 = 18,432$
DT (days)	45
Water compressibility (psi^{-1})	3.3×10^{-6}
Non-wetting compressibility (psi^{-1})	4.2×10^{-5}
Water density (lb/ft^3)	62.00
Non-wetting density (lb/ft^3)	48.00
Water viscosity (cp)	0.23
Non-wetting viscosity (cp)	1.6
Initial water saturation	.2
Initial pressure (psi)	1800
Injector BHP (psi)	2000
Producer BHP (psi)	1600

Table 2 shows how in this relatively simple problem, the Krylov-secant is able to produce important reductions in the computation. This is reflected

Table 2: Performance comparison between Newton method and the nonlinear KEN algorithm.

Method	Linear Iter.	Nonlinear Iter.	K-S updates	Back's	CPU time (s)
Newton	160	1076	-	2	779
N. KEN	107	780	92	6	482

in the nonlinear KEN algorithm as a smaller number of nonlinear iterations, linear iterations and CPU time in contrast to the Newton method. It is worth mentioning that the Krylov-secant steps produce marginally more backtracking steps than Newton for the successful iterations. However, Newton produces several backtracking steps at the beginning of the simulation that causes a halving of the timestep. There were indications that the nonlinear KEN may have a larger convergence region than Newton under certain situations. This fact has motivated other researchers to combine Broyden or Picard iterations with the Newton method in order to develop efficient continuation methods for timestepping.

Figure 4 compares the performance of both methods throughout the whole simulation. Clearly, the nonlinear KEN outperforms Newton's method. We can observe how the difference in performance builds up as the simulation progresses. As one might expect, the CPU time response is a direct consequence of the number of linear and nonlinear iterations.

7 Conclusions and Further Work

We have proposed a new family of Krylov-secant updates with the potential for a high-order of convergence in environmental problems. These updates do not explicitly rely on the computation or use of the Jacobian matrix that, in general, involves millions of floating point operations. We described the nonlinear KEN algorithm and a higher order version of it that showed to be superior to Newton's method for the numerical cases explored in this paper. The nonlinear KEN seems to be worthy of consideration in fully implicit formulations where the solution of a nonlinear algebraic system must be efficiently overcome.

Nevertheless, the potential for improving the present method is enormous. More robust versions may be developed by enriching or refreshing the Krylov basis in each of the nonlinear iterations. Advances along these lines have been used to iteratively solve linear systems with different right hand sides or a sequence of slowly changing linear systems in the matrix and the

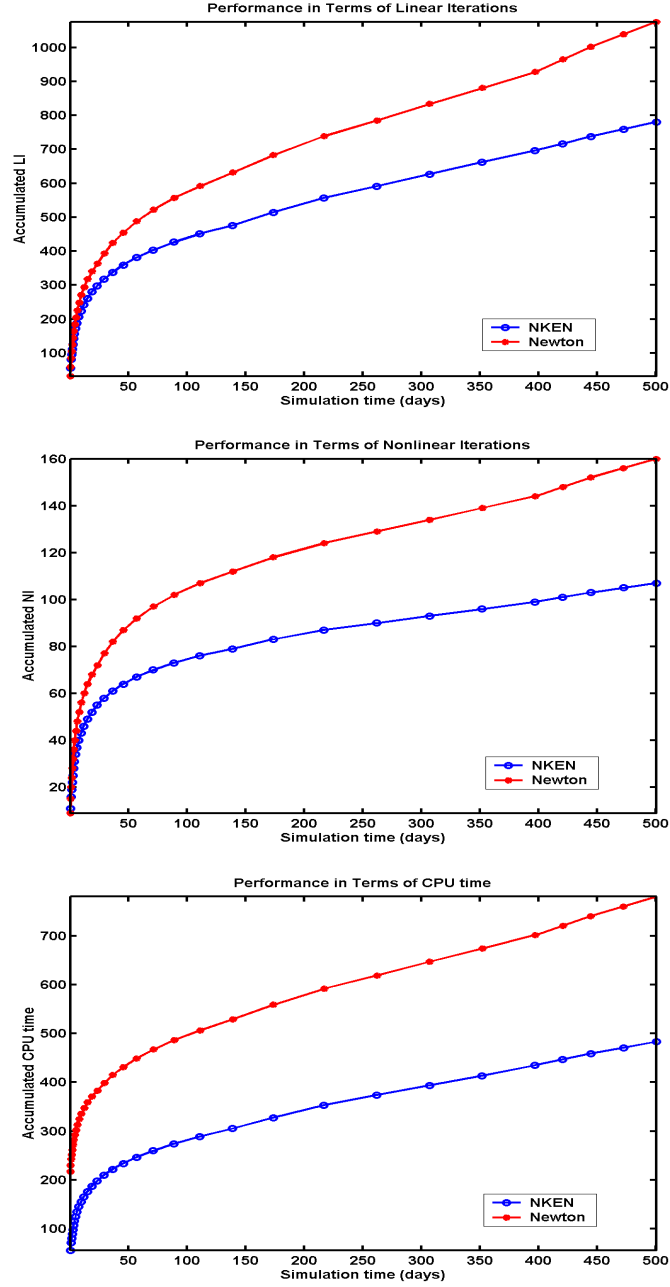


Figure 4: Comparison of linear (top), nonlinear iterations (middle) and CPU time (bottom) between the nonlinear KEN and Newton method.

right hand side [5, 23, 26]. We also find promising the use of deflation strategies oriented to capture the physics of the problem through eigen-information [1, 12].

We remark that our results are promising but must be evaluated under more stringent physical situations and at a larger scale. Among target applications we consider useful the insights coming from two-phase air-water, chemical processes and their coupling to geomechanics. The different strengths in nonlinearities between the typical coupled equations of the subsurface model (i.e., one for the pressure potential or flow driving force and another or others for the saturations/ concentrations) and the coupling itself between reservoir variables are sufficiently good reasons to investigate more of the preconditioning ideas discussed in this project.

Acknowledgements

The authors acknowledge DOD-PET Agency for its financial support through the project EQM K4-001, grant N62306-01-D-7110.

References

- [1] B. Aksoylu and H. Klie. Physics based isotropic deflation. In preparation, 2005.
- [2] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, 1994.
- [3] K. Aziz and A. Sethari. *Petroleum Reservoir Simulation*. Applied Science Publisher, 1983.
- [4] P. Bastian and R. Helmig. Efficient fully-coupled solution techniques for two-phase flow in porous media: Parallel multigrid solution and large scale computations. *Advances in Water Resources*, 23:199–216, 1999.
- [5] K. Burrage, J. Erhel, B. Pohl, and A. Williams. Deflation Technique for Linear System of Equations. *SIAM J. Sci. Statist. Comput.*, 19(4):1245–1260, 1998.
- [6] C. Dawson, H.M. Klie, M.F. Wheeler, and C. Woodward. A Parallel, implicit, cell-centered method for two-phase flow with a preconditioned Newton-Krylov solver. *Comp. Geosciences*, 1:215–249, 1997.

- [7] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
- [8] J.E. Dennis and H.F. Walker. Convergence theorems for least-change secant updates methods. *SIAM J. Numer. Anal.*, 18:949–987, 1981.
- [9] T. Eirola and O. Nevanlinna. Accelerating with rank-one updates. *Linear Alg. and its Appl.*, 121:511–520, 1989.
- [10] S.C. Eisenstat and H.F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM J. Sci. Comput.*, 17:16–32, 1996.
- [11] P.A. Forsyth. Comparison of single phase and two phase numerical model formulation for saturated-unsaturated groundwater flow. *Comp. Meth. Appl. Mech. Eng.*, 69:243–259, 1988.
- [12] J. Frank and C. Vuik. On the construction of deflation-based preconditioners. *SIAM J. Sci. Comput.*, 23(2):442–462, 2001.
- [13] G.-D. Gu and H-B. Wu. A Block EN Algorithm for Nonsymmetric Linear Systems with Multiple Right Hand Sides. *Lin. Alg. And its Appl.*, 299:1–20, 1999.
- [14] R. Helmig. *Multiphase flow and transport processes in the Subsurface-A contribution to the Modeling Systems*. Springer-Verlag, 1997.
- [15] 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. Comput.*, 23(2):430–441, 2001.
- [16] J.E. Jones and C. S. Woodward. Newton-Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24:733–774, 2001.
- [17] C.T. Kelley. Iterative methods for linear and nonlinear equations. In *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1995.
- [18] H. Klie. *Krylov-Secant methods for solving Large Scale Systems of Coupled Nonlinear Parabolic Equations*. PhD thesis, Dept. of Computational and Applied Mathematics, Rice University, Houston, TX, 1996.
- [19] D. A. Knoll and D. E. Keyes. Jacobian free Newton-Krylov methods: A survey of approaches and Applications. *Journal of Computational Physics*, 20(2):357–397, 2004.

- [20] S. Lacroix, Yu. Vassileski, J. Wheeler, and M.F. Wheeler. Iterative solution methods for modeling multiphase flow in porous media fully implicitly. *SIAM J. Sci. Comput.*, 25(3):905–926, 2003.
- [21] C.T. Miller, C. Abhishek, A.B. Sallerson, J.F. Prins, and M.W. Farthing. A comparison of computational and algorithm advances for solving Richard’s equation. In *Computational Methods in Water Resources 2004 International Conference*, Chapell Hill, North Carolina, USA., 2004.
- [22] J.L. Morales and J. Nocedal. Automatic preconditioning by limited memory quasi-newton updating. *SIAM J. Optim.*, 10(4):1079–1096, 2000.
- [23] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Statist. Comput.*, 24(1):20–27, 2002.
- [24] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
- [25] M. Pernice and H.F. Walker. NITSOL: A Newton iterative solver for nonlinear systems. *SIAM J. Sci. Comput.*, 19:302–318, 1998.
- [26] F. Risler and C. Rey. Iterative accelerating algorithms with Krylov subspaces for the solution to large-scale nonlinear problems. *Numerical Algorithms*, 23:1–30, 2000.
- [27] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003. 2nd. edition.
- [28] V.E. Shamanskii. A modification of Newton’s method. *Ukrain. Mat. Zh.*, 19:133–138, 1967. In Russian.
- [29] J.F. Traub. *Iterative methods for the solution of equations*. Prentice Hall, Englewood Cliffs, 1964.
- [30] B. van de Rotten and S.M. Verduyn Lunel. A limited memory Broyden method to solve high-dimensional systems of nonlinear equations. Technical Report Tech. Report NI 2003-6, Mathematical Institute, University of Leiden, Leiden, The Netherlands, 2003.
- [31] H. van der Vorst. *Iterative Krylov Methods for Large Linear Systems*. Cambridge University Press, 2003.

- [32] J. Xue W. Gao and Y. Qu. A New Implementation of EN Method. *Appl. Math. And Comp.*, 98:199–208, 1999.
- [33] Mary F. Wheeler, H. Klie, C. Dawson, O. Eslinger, and X. Gai. Linear and nonlinear solvers for environmental applications. Technical Report N62306-01-D-7110, No. 0062 PET Program, Center for Subsurface Modeling (CSM), The University of Texas at Austin, May, 2004.
- [34] U.M. Yang. *A family of preconditioned iterative solvers for sparse linear systems*. PhD thesis, Dept. of Computer Science, University of Illinois, Urbana-Champaign, 1995.