Asia Pac. J. Math. 2026 **13**:178

HIGH-ORDER FROZEN-JACOBIAN CORRECTIONS IN A GLOBALIZED INEXACT NEWTON–KRYLOV FRAMEWORK

AHMED K. AL-JABERI[1,*], MIMOON ISMAEL[2], SAFA H. OBAID[3], MOHANNAD M. EISA[1]

[1]Department of Mathematics, College of Education for Pure Sciences, University of Basrah, Basra 61001, Iraq

[2]Department of Mathematics, College of Education for Pure Sciences, University of Anbar, Ramadi 31001, Iraq

[3]Directorate General of Salahaldeen Education, Tikrit, Iraq

[*]Corresponding author: ahmed.shanan@uobasrah.edu.iq

Received Aug. 24, 2026

Abstract. Reusing a single Jacobian across high-order steps can reduce nonlinear error without rebuilding linear-algebra setup, but sparse applications require controlled inexact solves and globalization. We present Gain-FJ, a globalized inexact Newton–Krylov framework in which frozen-Jacobian high-order corrections are optional and a separately computed inexact Newton step is retained as a descent fallback. We prove a three-power gain for each exact correction block, derive distinct sufficient forcing scales for predictor and correction solves, establish asymptotic inactivity of continuation, rejection, and refresh safeguards near a simple root, and prove residual convergence under explicit level-set assumptions. Arbitrary-precision MATLAB experiments recover orders five, eight, eleven, and fourteen; adaptive forcing reduces Krylov work by 30.0–45.4% in controlled work-to-tolerance tests and by 67.7% on a difficult cyclic start. Sparse reaction–diffusion, convection–diffusion–reaction, and Bratu tests through 102400 unknowns show fewer nonlinear updates and Jacobian/preconditioner constructions, though not uniformly fewer residual evaluations or Krylov iterations. A break-even analysis therefore characterizes Gain-FJ as a setup-reuse strategy whose benefit depends on the relative cost of setup and repeated linear work.

2020 Mathematics Subject Classification. 65H10; 65F10; 65N22; 65Y20.

Key words and phrases. nonlinear systems; frozen Jacobian; high-order iteration; inexact Newton–Krylov methods; adaptive forcing; preconditioning.

1. Introduction

Consider a nonlinear system

$$F(X) = 0, \qquad F : D \subset \mathbb{R}^n \to \mathbb{R}^n, \tag{1.1}$$

where D is open. Unless stated otherwise, $\|\cdot\|$ denotes the Euclidean vector norm and its induced matrix norm. For bilinear and trilinear maps, the same notation denotes the corresponding induced

DOI: 10.28924/APJM/13-178

operator norm. Large instances of (1.1) arise after spatial or temporal discretization of nonlinear partial differential equations, in inverse problems, and in multiphysics models. For such systems, the practical cost of Newton's method depends on much more than the number of outer iterations: Jacobian formation, preconditioner setup, Krylov work, memory, and globalization can dominate the computation [1,5,8,9].

High-order methods for nonlinear systems pursue a different objective. They use additional function evaluations or correction stages to increase the local convergence order. Frozen-Jacobian multistep schemes are particularly attractive because multiple right-hand sides can be processed with one Jacobian or one factorization [18,20–23]. Recent work continues to develop convergence-order accelerators and damped high-order constructions for nonlinear systems [24–26]. The price of high local order is that the underlying linearization can become inaccurate away from the root, and exact or over-accurate linear solves can make the additional stages expensive.

Inexact Newton methods address linear oversolving by imposing a residual tolerance on the Newton equation rather than solving it exactly [2–4]. Newton–Krylov implementations combine this idea with Krylov subspace solvers and preconditioning, thereby avoiding dense direct factorizations and enabling matrix-free Jacobian-vector products when appropriate [5–7,9]. The quality and update frequency of the preconditioner can be decisive; preconditioner construction itself may be a substantial part of the nonlinear solve [13]. Nonlinear preconditioning and elimination methods further show that robustness often depends on how nonlinear structure is treated rather than on the Newton step alone [11,14–17].

The question considered here is more specific: can an inexact Newton–Krylov method use a high-order frozen-Jacobian block without changing its local order, requiring unnecessarily accurate linear solves, or losing a globally convergent Newton fallback? In Gain-FJ, the correction block serves as an optional accelerator rather than as the main update. Each outer iteration retains an inexact Newton predictor as a fallback. During a frozen-Jacobian epoch, additional three-solve correction blocks are attempted only while they continue to reduce the residual sufficiently. Whenever the linearization-defect test or the Krylov-work test shows that reuse is no longer effective, the method rebuilds the Jacobian and preconditioner. In contrast, the growth test rejects any unsafe internal trial, preventing it from being the beginning of a new epoch. We then apply Armijo backtracking to the selected direction: if the high-order candidate fulfils the safeguards we accept it, otherwise we accept the retained Newton direction.

1.1. **Related work, closest methods, and the open question.** Three established areas of research are directly relevant to this study. First, high-order vectorial methods increase local convergence order by composing Newton-like corrections or reusing derivative information [18,20–24]. Second, inexact Newton theory characterizes how the residual of the linearized equation affects nonlinear convergence

and how forcing terms should be selected [2–4]. Third, large-scale Newton–Krylov methods emphasize globalization, preconditioning, and nonlinear structure [8, 10–12, 15–17].

The closest multistep literature is particularly important for positioning the present work. Ahmad, Tohidi, and Carrasco [19] developed a parameterized multistep Newton method that requires one Jacobian inversion per outer iteration and can be implemented with Krylov subspace methods. Al-Obaidi and Darvishi [21] constructed frozen-Jacobian multistep schemes with an explicit emphasis on formal order and efficiency. Erfanifar and Hajarian [22] proposed a multistep family with one Fréchet-derivative evaluation and inversion per iteration and studied order, efficiency, and attraction behavior; their later high-order construction [23] further increases the formal order and applies the method to differential-equation and image-processing models. Most closely related to the present framework, Cordero, Rojas-Hiciano, and Torregrosa [26] recently combined an eighth-order vectorial method with Armijo-type globalization, frozen sparse-factorization reuse, large PDE experiments, and a Newton–Krylov comparison. It therefore provides the closest comparison in terms of globalization and frozen linear-algebra reuse. The question addressed here is narrower and complementary: how accurately may the repeated inverse actions be computed when they are performed inexactly, and how can the frozen high-order block be embedded in a Newton–Krylov iteration that retains an independent certified descent fallback?

TABLE 1. Position of the present framework relative to selected closely related multistep methods. The table summarizes the primary emphasis of each cited work rather than attempting an exhaustive feature checklist.

Work	Reuse/linear-algebra feature	Primary emphasis relevant to this paper
Ahmad et al. [19]	One Jacobian inversion per iteration; Krylov implementation is explicitly discussed	Parameterized multistep Newton iteration, convergence behavior, radius of convergence, and PDE-derived nonlinear systems
Al-Obaidi–Darvishi [21]	Frozen-Jacobian multistep construction	High local order, efficiency index, and nonlinear-system experiments
Erfanifar–Hajarian [22]	One Fréchet-derivative evaluation and inversion per iteration	Multistep order, efficiency, accuracy, attraction basins, and applications
Erfanifar–Hajarian [23]	High-order multistep reuse strategy	Higher formal order with differential-equation and image-processing applications
Cordero et al. [26]	Frozen sparse-factorization reuse with an eighth-order vectorial scheme	Armijo-type globalization, large convection/diffusion-reaction PDEs, stability, and Newton–Krylov comparison
Present GAIN-FJ framework	Frozen Jacobian/preconditioner reused across several right-hand sides; residual-controlled Krylov solves	Inexactness scales that preserve the high-order block, internal safeguards, retained Newton descent fallback, Armijo globalization, and residual global convergence

The novelty is therefore not an isolated claim of a higher formal order, nor is it the use of Armijo globalization or frozen linear algebra by itself. Relative to the closest globalized high-order construction [26], the main analytical point here is a compatibility result between a specific frozen-Jacobian correction map and residual-controlled *inexact* inverse actions: first-order predictor forcing and third-order correction forcing are sufficient to retain the local gain. The main algorithmic point is that the accelerator is disposable: the outer iteration keeps a separately computed inexact Newton direction, so an unsuccessful correction, refresh, or candidate evaluation cannot remove the certified descent fallback. Conditional Jacobian/preconditioner refresh is then incorporated without entering the asymptotic local regime. The resulting analysis connects high-order local behavior, inexact Krylov accuracy, safeguard inactivity, and residual global convergence in one framework.

1.2. **Contributions.** The paper makes four main contributions.

(i) We give a direct local analysis of the exact frozen-Jacobian correction map and prove that each correction block contributes three additional powers of the outer error once its input is of Newton-predictor size. The stage-count convention is made explicit, and the resulting local-order lower bounds for two through five stage points are five, eight, eleven, and fourteen.

(ii) We derive distinct sufficient inexactness scales for the predictor and correction systems and prove preservation of the local high-order lower bound under residual-controlled inexact solves. This establishes a direct link between high-order frozen-Jacobian analysis and inexact Newton forcing.

(iii) We embed the block in a safeguarded architecture that retains the outer inexact Newton step, rejects unsafe internal growth, refreshes the Jacobian/preconditioner conditionally, and uses Armijo backtracking. We prove uniform local inactivity of the internal safeguards, eventual acceptance of the full high-order step, and residual global convergence under explicit level-set assumptions.

(iv) We implement the method with right-preconditioned GMRES so that the monitored Krylov residual is the original-system residual used in the theory. High-precision and sparse PDE experiments quantify both benefits and costs. In addition to raw work counts, a break-even cost criterion identifies when reductions in Jacobian/preconditioner setups can offset the extra residual and Krylov work.

The remainder of the paper is organized as follows. Section 2 analyzes the exact high-order block. Section 3 defines Gain-FJ. Section 4 proves the local and global convergence properties. Section 5 gives the work model and implementation principles. Section 6 reports the numerical study. Sections 7–9 discuss the results, limitations, and conclusions.

2. Exact Frozen-Jacobian Correction Block

Let $\xi \in D$ be a simple root of (1.1),

$$F(\xi) = 0, \qquad J_* := F'(\xi) \text{ nonsingular}, \tag{2.1}$$

and define

$$e_k = X_k - \xi, \qquad J_k = F'(X_k). \tag{2.2}$$

We assume throughout the local analysis that F is three times continuously differentiable in a neighborhood of ξ.

Remark 2.1 (Order convention). In the local arguments, we say that the iteration has "Q-order at least p" when there is a constant C, independent of k, such that for all sufficiently large k (or, in a local one-step statement, for every sufficiently small nonzero error),

$$\frac{\|X_{k+1} - \xi\|}{\|X_k - \xi\|^p} \le C. \tag{2.3}$$

The estimate therefore gives only a lower bound for the order. For some problems, additional cancellations may lead to an actual order greater than p.

2.1. Newton predictor and frozen-Jacobian correction formulas.

The exact Newton equation defines the predictor point as

$$J_k(Z_1 - X_k) = -F(X_k). \tag{2.4}$$

At any correction stage Z_j, keep J_k fixed and introduce

$$S_{j,1} = J_k^{-1} F(Z_j), \qquad U_j = Z_j - \frac{1}{3} S_{j,1}, \tag{2.5}$$

$$S_{j,2} = J_k^{-1} F(U_j), \qquad V_j = Z_j - S_{j,2}, \tag{2.6}$$

$$S_{j,3} = J_k^{-1} F(V_j), \qquad Z_{j+1} = Z_j - 3S_{j,3}. \tag{2.7}$$

An m-stage cycle contains the predictor Z_1 and $m-1$ correction blocks, and sets

$$X_{k+1} = Z_m, \qquad m \ge 2. \tag{2.8}$$

This convention is important: m counts stage points, not correction blocks.

2.2. Uniform local gain of one correction block.

The following result supplies the local mechanism used throughout the paper.

Theorem 2.2 (Three-power gain of the exact correction block). *Fix $C_Z > 0$. There exist $\delta > 0$ and $C_B > 0$ such that, whenever $\|e_k\| \le \delta$ and*

$$\|Z_j - \xi\| \le C_Z \|e_k\|^2, \tag{2.9}$$

one exact block (2.5)–(2.7)*, with* $J_k = F'(X_k)$ *frozen, satisfies*

$$\|Z_{j+1} - \xi\| \le C_B \|e_k\|^3 \|Z_j - \xi\| . \tag{2.10}$$

Consequently, if $Z_j - \xi = \mathcal{O}(\|e_k\|^q)$ *for* $q \ge 2$*, then*

$$Z_{j+1} - \xi = \mathcal{O}(\|e_k\|^{q+3}). \tag{2.11}$$

Proof. Write $z = Z_j - \xi$. For a nonsingular matrix A near J_* define

$$H_A(z) = A^{-1}F(\xi + z), \qquad B = A^{-1}J_*. \tag{2.12}$$

The block is represented by

$$u_A(z) = z - \frac{1}{3}H_A(z), \tag{2.13}$$

$$v_A(z) = z - H_A(u_A(z)), \tag{2.14}$$

$$\Psi_A(z) = z - 3H_A(v_A(z)). \tag{2.15}$$

Thus $Z_{j+1} - \xi = \Psi_A(z)$ for $A = J_k$.

Because $H_A(0) = 0$ and $DH_A(0) = B$, direct differentiation gives

$$Du_A(0) = I - \frac{1}{3}B, \tag{2.16}$$

$$Dv_A(0) = I - B + \frac{1}{3}B^2, \tag{2.17}$$

$$D\Psi_A(0) = I - 3B + 3B^2 - B^3 = (I - B)^3. \tag{2.18}$$

For $A = J_k$, local Lipschitz continuity of F' and uniform boundedness of J_k^{-1} near ξ imply

$$\left\|I - J_k^{-1}J_*\right\| = \left\|J_k^{-1}(J_k - J_*)\right\| \le C \|e_k\| , \tag{2.19}$$

so

$$\|D\Psi_{J_k}(0)\| \le C \|e_k\|^3 . \tag{2.20}$$

We next control the quadratic term uniformly in A. At $A = J_*$ let

$$\mathcal{C} = J_*^{-1}F''(\xi), \tag{2.21}$$

viewed as a bounded symmetric bilinear map. Taylor expansion gives

$$H_{J_*}(z) = z + \frac{1}{2}\mathcal{C}[z, z] + \mathcal{O}(\|z\|^3). \tag{2.22}$$

Substitution into (2.13)–(2.15) yields

$$u_{J_*}(z) = \frac{2}{3}z - \frac{1}{6}\mathcal{C}[z, z] + \mathcal{O}(\|z\|^3), \tag{2.23}$$

$$H_{J_*}(u_{J_*}(z)) = \frac{2}{3}z + \frac{1}{18}\mathcal{C}[z, z] + \mathcal{O}(\|z\|^3), \tag{2.24}$$

$$v_{J_*}(z) = \frac{1}{3}z - \frac{1}{18}\mathcal{C}[z,z] + \mathcal{O}(\|z\|^3), \tag{2.25}$$

$$H_{J_*}(v_{J_*}(z)) = \frac{1}{3}z + \mathcal{O}(\|z\|^3). \tag{2.26}$$

Therefore

$$\Psi_{J_*}(z) = \mathcal{O}(\|z\|^3), \tag{2.27}$$

which implies $D^2\Psi_{J_*}(0) = 0$.

To make the preceding cancellation uniform, one must also ensure that the composed arguments appearing in $H_A(u_A(z))$ and $H_A(v_A(z))$ remain inside the domain of F. Choose $r_0 > 0$ so that $\xi + \overline{B}_{r_0}(0) \subset D$, and choose a compact neighborhood $\mathcal{A}$ of J_* consisting only of nonsingular matrices, with A^{-1} uniformly bounded on $\mathcal{A}$. By the joint continuity of $(A, z) \mapsto u_A(z)$ at $(J_*, 0)$, after shrinking $\mathcal{A}$ and choosing $0 < r_1 < r_0$ we have

$$u_A(\overline{B}_{r_1}(0)) \subset B_{r_0}(0), \qquad A \in \mathcal{A}.$$

On this set v_A is well defined and jointly continuous; shrinking once more, choose $0 < r \le r_1$ such that

$$v_A(\overline{B}_r(0)) \subset B_{r_0}(0), \qquad A \in \mathcal{A}.$$

Thus H_A, u_A, v_A, and Ψ_A are well defined and C^3 jointly in (A, z) on an open neighborhood of the compact set $\mathcal{A} \times \overline{B}_r(0)$. By continuity of F' and the input bound (2.9), after shrinking δ we may also assume $J_k \in \mathcal{A}$ and $z \in B_r(0)$. Hence $D^2\Psi_A(0)$ is locally Lipschitz in A, and there is a constant C such that

$$\left\|D^2\Psi_A(0)\right\| \le C \left\|A - J_*\right\|. \tag{2.28}$$

The third derivative $D^3\Psi_A(z)$ is uniformly bounded on $\mathcal{A} \times \overline{B}_r(0)$. Taylor's theorem with integral remainder, together with (2.20) and (2.28), gives for $A = J_k$

$$\|\Psi_{J_k}(z)\| \le C\left(\|e_k\|^3 \|z\| + \|e_k\| \|z\|^2 + \|z\|^3\right). \tag{2.29}$$

Under (2.9),

$$\|e_k\| \|z\|^2 + \|z\|^3 \le C \|e_k\|^3 \|z\| \tag{2.30}$$

for $\|e_k\|$ sufficiently small. This proves (2.10), and (2.11) follows immediately. □

Corollary 2.3 (Fixed-stage order)**.** *The exact predictor satisfies $Z_1 - \xi = \mathcal{O}(\|e_k\|^2)$. Therefore the exact m-stage cycle (2.8) satisfies*

$$\|X_{k+1} - \xi\| = \mathcal{O}(\|X_k - \xi\|^{p_m}), \qquad p_m = 2 + 3(m-1) = 3m - 1. \tag{2.31}$$

Hence its local order is at least p_m. In particular, $m = 2, 3, 4, 5$ give the lower bounds $5, 8, 11, 14$.

Proof. Standard Newton analysis gives $Z_1 - \xi = \mathcal{O}(\|e_k\|^2)$, so (2.9) holds for Z_1 once k is sufficiently large. If $Z_j - \xi = \mathcal{O}(\|e_k\|^{q_j})$ with $q_j \geq 2$, then Theorem 2.2 gives $Z_{j+1} - \xi = \mathcal{O}(\|e_k\|^{q_j+3})$. Induction over the $m-1$ correction blocks yields $q_m = 2 + 3(m-1) = 3m-1$. □

Remark 2.4. The indexing in (2.31) is the one used throughout this paper. It also agrees with the observed order sequence in the motivating frozen-Jacobian construction [23]. Using a different definition of "step" shifts the formula, so comparisons should be made from the explicit stage count rather than from notation alone.

3. The Globalized Adaptive Inexact Method

The outer iteration advances from X_k to X_{k+1}. Candidate construction is allowed to perform several frozen-Jacobian corrections, but the first inexact Newton predictor is retained throughout the outer iteration as a fallback direction. This separation is what permits local acceleration without sacrificing a standard descent mechanism.

3.1. **Scaled residual and inexact linear systems.** Define

$$R(X) = \frac{\|F(X)\|}{1 + \|F(X_0)\|}. \tag{3.1}$$

Successful termination requires

$$R(X_k) \leq \tau_F. \tag{3.2}$$

A relative step size may be recorded as a stagnation diagnostic, but a small step is not by itself declared convergence when the residual test fails.

Every formal inverse action $A^{-1}b$ is replaced by an approximate solution s satisfying

$$As = b - r, \qquad \|r\| \leq \eta \|b\|, \qquad 0 \leq \eta < 1. \tag{3.3}$$

The residual in (3.3) is the residual of the original linear system, independent of the preconditioning convention. In the MATLAB implementation we enforce this convention by explicit *right preconditioning*. If M denotes the current Jacobi or ILUTP preconditioner, GMRES is applied to

$$AM^{-1}y = b, \qquad s = M^{-1}y,$$

without passing M through MATLAB's `gmres` preconditioner arguments. The GMRES residual is then exactly $b - As$, so the stopping test coincides with (3.3). This implementation detail is important because a stopping test based only on a preconditioned residual need not certify the original-system tolerance used in the convergence analysis. FGMRES can be substituted when the preconditioning operation changes during a solve [7].

3.2. **Stage-dependent forcing.** Let Y denote the point at which the current Jacobian is frozen and set

$$\theta(Y) = \min\left\{1, \frac{\|F(Y)\|}{1 + \|F(X_0)\|}\right\}. \tag{3.4}$$

We use

$$\eta_{\text{pred}}(Y) = \min\{\bar{\eta}, c_p\theta(Y)\}, \tag{3.5}$$

$$\eta_{\text{corr}}(Y) = \min\{\bar{\eta}, c_c\theta(Y)^3\}, \tag{3.6}$$

where $0 < \bar{\eta} < 1$ and $c_p, c_c > 0$. For a practical finite-precision implementation, if a lower forcing floor $0 \le \eta_{\min} < \bar{\eta}$ is used, we define explicitly

$$\eta^{\text{fp}}_{\text{pred}}(Y) = \min\{\bar{\eta}, \max\{\eta_{\min}, c_p\theta(Y)\}\}, \tag{3.7}$$

$$\eta^{\text{fp}}_{\text{corr}}(Y) = \min\{\bar{\eta}, \max\{\eta_{\min}, c_c\theta(Y)^3\}\}. \tag{3.8}$$

The floor must be disabled in the asymptotic order analysis because a fixed positive $\eta_{\min}$ is not $\mathcal{O}(\|e_k\|)$ or $\mathcal{O}(\|e_k\|^3)$ as $e_k \to 0$. For concise algorithmic notation define

$$(\eta^{\text{use}}_{\text{pred}}, \eta^{\text{use}}_{\text{corr}}) = \begin{cases} (\eta_{\text{pred}}, \eta_{\text{corr}}), & \text{floor-free theoretical iteration,} \\ (\eta^{\text{fp}}_{\text{pred}}, \eta^{\text{fp}}_{\text{corr}}), & \text{finite-precision implementation.} \end{cases} \tag{3.9}$$

Section 4 shows why the predictor and correction exponents differ.

3.3. **Residual-based stage continuation and growth rejection.** Suppose an accepted block maps Z_j to Z_{j+1}. Define

$$\rho_j = \frac{\|F(Z_{j+1})\|}{\|F(Z_j)\|}. \tag{3.10}$$

Additional blocks are attempted only while

$$\rho_j \le \rho_{\text{cont}} \tag{3.11}$$

and the prescribed maximum number $m_{\max}$ of stage points has not been reached. This simple rule replaces the more heavily parameterized work-normalized stage score used in an earlier draft.

Before a trial block is committed, its residual ratio

$$\rho_{\text{trial}} = \frac{\|F(Z_{\text{trial}})\|}{\|F(Z_j)\|} \tag{3.12}$$

is tested against a fixed growth threshold $\rho_{\text{rej}} > 1$. If

$$\rho_{\text{trial}} > \rho_{\text{rej}}, \tag{3.13}$$

the block is discarded, the most recent committed point is retained, and the high-order candidate is marked invalid. Crucially, a rejected point is never used as a new Jacobian anchor. The outer line search then works with the retained inexact Newton direction.

3.4. **Conditional Jacobian and preconditioner refresh.** Let Y be the active frozen anchor, $A = F'(Y)$, and Z the most recent accepted internal point. We use the normalized linearization defect

$$\chi(Z;Y,A) = \begin{cases} \dfrac{\|F(Z) - F(Y) - A(Z-Y)\|}{\|F(Z) - F(Y)\|}, & F(Z) \neq F(Y), \\ 0, & Z = Y, \\ +\infty, & F(Z) = F(Y) \text{ and } Z \neq Y. \end{cases} \tag{3.14}$$

The piecewise definition removes the zero-denominator ambiguity. Near a simple root the local inverse function theorem makes F injective, so the last case cannot occur once Y and Z are sufficiently close to ξ. In floating-point code a small safety floor may still be used in the denominator. Let $K_{j,\max}$ be the maximum number of Krylov iterations among the three solves in the current correction block. A refresh is triggered when

$$\chi_j > \chi_{\text{ref}} \qquad \text{or} \qquad K_{j,\max} > K_{\text{ref}}^{\max}. \tag{3.15}$$

If the refresh budget has not been exhausted, the accepted point becomes the new anchor, the Jacobian and preconditioner are rebuilt, and a new inexact Newton predictor starts a fresh frozen-Jacobian epoch. Compared with the earlier version of the algorithm, displacement and residual-stagnation refresh thresholds have been removed because they were partly redundant with the stage-ratio and defect tests.

3.5. **Safeguarded direction selection and Armijo globalization.** Use the standard residual merit function and Armijo backtracking [27]

$$\Phi(X) = \frac{1}{2}\|F(X)\|^2, \qquad \Phi'(X;d) = F(X)^{\mathsf{T}} F'(X) d. \tag{3.16}$$

The retained inexact Newton direction d_k^N satisfies

$$J_k d_k^N = -F_k + r_k^N, \qquad \left\|r_k^N\right\| \le \eta_{k,0} \|F_k\|, \tag{3.17}$$

where $F_k = F(X_k)$. Hence

$$F_k^{\mathsf{T}} J_k d_k^N \le -(1 - \eta_{k,0}) \|F_k\|^2, \tag{3.18}$$

so the predictor is a strict descent direction whenever $\eta_{k,0} < 1$.

If candidate construction returns a valid high-order point $\widehat{X}_{k+1}$, define

$$d_k^H = \widehat{X}_{k+1} - X_k. \tag{3.19}$$

The high-order direction is used only if

$$F_k^{\mathsf{T}} J_k d_k^H \le -\sigma_d \|F_k\|^2, \tag{3.20}$$

$$\left\|d_k^H\right\| \le \omega_d \left\|d_k^N\right\|, \tag{3.21}$$

with $0 < \sigma_d < 1$ and $\omega_d > 1$. Otherwise $d_k = d_k^N$.

Starting with $\lambda = 1$, backtracking replaces λ by $\beta\lambda$ until

$$\Phi(X_k + \lambda d_k) \le \Phi(X_k) + c_A \lambda F_k^{\mathsf{T}} J_k d_k, \qquad 0 < c_A < \frac{1}{2}, \tag{3.22}$$

where $0 < \beta < 1$.

Algorithm 1. Adaptive inexact frozen-Jacobian candidate construction

3.6. **Complete algorithm.**

Require: X_k, cached F_k, outer Jacobian J_k, solver parameters

Ensure: retained Newton data (d_k^N, r_k^N), optional candidate $\widehat{X}_{k+1}$ and residual caches, validity flag, or a converged point

Immediate stopping rule: after every successful evaluation of $F(W)$, if $R(W) \le \tau_F$, return W immediately with status converged.

$Y \leftarrow X_k$, $F_Y \leftarrow F_k$, $A \leftarrow J_k$; build/update preconditioner P
$n_{\mathrm{ref}} \leftarrow 0$, $m_{\mathrm{epoch}} \leftarrow 1$, validH $\leftarrow$ true
Solve $As_0 = F_Y - r_0$ with relative residual tolerance $\eta_{\mathrm{pred}}^{\mathrm{use}}(Y)$
if the outer predictor solve fails **then**
    return predictor failure
end if
Retain $d_k^N \leftarrow -s_0$ and $r_k^N \leftarrow r_0$
$Z \leftarrow X_k + d_k^N$
Attempt to evaluate $F_Z = F(Z)$
if $Z \notin D$ or F_Z is unavailable/nonfinite **then**
    validH $\leftarrow$ false
    return d_k^N, r_k^N, validH with no full-step residual cache
end if
Cache $F_N \leftarrow F_Z$
while $R(Z) > \tau_F$ and $m_{\mathrm{epoch}} < m_{\max}$ **do**
    $Z_{\mathrm{old}} \leftarrow Z$, $F_{\mathrm{old}} \leftarrow F_Z$, $\eta_c \leftarrow \eta_{\mathrm{corr}}^{\mathrm{use}}(Y)$
    Solve $As_1 = F_{\mathrm{old}} - r_1$ with relative residual tolerance η_c
    if the solve fails **then**
        validH $\leftarrow$ false; **break**
    end if
    $U \leftarrow Z_{\mathrm{old}} - s_1/3$; attempt to evaluate $F_U = F(U)$
    if $U \notin D$ or F_U is unavailable/nonfinite **then**
        validH $\leftarrow$ false; **break**
    end if
    Solve $As_2 = F_U - r_2$ with relative residual tolerance η_c
    if the solve fails **then**
        validH $\leftarrow$ false; **break**
    end if
    $V \leftarrow Z_{\mathrm{old}} - s_2$; attempt to evaluate $F_V = F(V)$
    if $V \notin D$ or F_V is unavailable/nonfinite **then**
        validH $\leftarrow$ false; **break**

end if
Solve $As_3 = F_V - r_3$ with relative residual tolerance η_c
if the solve fails **then**
    validH $\leftarrow$ false; **break**
end if
$Z_{\text{trial}} \leftarrow Z_{\text{old}} - 3s_3$; attempt to evaluate $F_{\text{trial}} = F(Z_{\text{trial}})$
if $Z_{\text{trial}} \notin D$ or F_{trial} is unavailable/nonfinite **then**
    validH $\leftarrow$ false; **break**
end if
$\rho_{\text{trial}} \leftarrow \|F_{\text{trial}}\| \,/\, \|F_{\text{old}}\|$ ▷ $F_{\text{old}} \neq 0$ here
if $\rho_{\text{trial}} > \rho_{\text{rej}}$ **then**
    validH $\leftarrow$ false; **break** ▷ reject before commitment
end if
Commit $Z \leftarrow Z_{\text{trial}}$, $F_Z \leftarrow F_{\text{trial}}$, $\rho \leftarrow \rho_{\text{trial}}$
$m_{\text{epoch}} \leftarrow m_{\text{epoch}} + 1$; compute $\chi(Z; Y, A)$ and $K_{j,\max}$
if $R(Z) \leq \tau_F$ **then**
    break
end if
if $\chi(Z; Y, A) > \chi_{\text{ref}}$ or $K_{j,\max} > K_{\text{ref}}^{\max}$ **then**
    if $n_{\text{ref}} < N_{\text{ref}}^{\max}$ **then**
        $Y \leftarrow Z$, $F_Y \leftarrow F_Z$, $A \leftarrow F'(Y)$; rebuild/update P
        $n_{\text{ref}} \leftarrow n_{\text{ref}} + 1$, $m_{\text{epoch}} \leftarrow 1$
        Solve $As_0^{\text{new}} = F_Y - r_0^{\text{new}}$ with relative residual tolerance $\eta_{\text{pred}}^{\text{use}}(Y)$
        if the refresh predictor solve fails **then**
            validH $\leftarrow$ false; **break**
        end if
        $Z_{\text{trial}} \leftarrow Y - s_0^{\text{new}}$; attempt to evaluate $F_{\text{trial}} = F(Z_{\text{trial}})$
        if $Z_{\text{trial}} \notin D$ or F_{trial} is unavailable/nonfinite **then**
            validH $\leftarrow$ false; **break**
        end if
        $\rho_{\text{refresh}} \leftarrow \|F_{\text{trial}}\| \,/\, \|F_Y\|$ ▷ $F_Y \neq 0$ here
        if $\rho_{\text{refresh}} > \rho_{\text{rej}}$ **then**
            validH $\leftarrow$ false; **break**
        end if
        Commit $Z \leftarrow Z_{\text{trial}}$, $F_Z \leftarrow F_{\text{trial}}$; **continue**
    else
        break
    end if
end if
if $\rho > \rho_{\text{cont}}$ **then**
    break

end if

end while

if validH **then**

$\widehat{X}_{k+1} \leftarrow Z$, $F_H \leftarrow F_Z$

end if

return d_k^N, r_k^N, F_N (if cached), $\widehat{X}_{k+1}, F_H$ (if valid), validH, and work counters

Algorithm 2 Globalized GAIN-FJ outer iteration

Require: $X_0 \in D$ with finite $F(X_0)$, nonlinear tolerance, stage/refresh/Krylov/line-search parameters

Evaluate and cache $F_0 = F(X_0)$

for $k = 0, 1, 2, \ldots, k_{\max} - 1$ **do**

if $R(X_k) \le \tau_F$ **then**

return X_k

end if

Build/define $J_k = F'(X_k)$

Call Algorithm 1

if Algorithm 1 returned a converged point X_* **then**

return X_*

end if

if the outer predictor solve failed **then**

return solver failure

end if

if validH **then**

$d_k^H \leftarrow \widehat{X}_{k+1} - X_k$ and $\delta_H \leftarrow F_k^{\mathsf{T}} J_k d_k^H$

end if

if validH and (3.20)–(3.21) hold **then**

$d_k \leftarrow d_k^H$, $\delta_k \leftarrow \delta_H$; use cached F_H for the trial $\lambda = 1$

else

$d_k \leftarrow d_k^N$, $\delta_k \leftarrow -\|F_k\|^2 + F_k^{\mathsf{T}} r_k^N$

Use cached F_N for $\lambda = 1$ if it is available; otherwise evaluate the trial

end if

$\lambda \leftarrow 1$

Evaluate/reuse $F_{\text{trial}} = F(X_k + \lambda d_k)$ whenever this trial lies in D and is finite

while the trial is outside D, nonfinite, or violates the Armijo condition (3.22) **do**

$\lambda \leftarrow \beta\lambda$

Evaluate $F_{\text{trial}} = F(X_k + \lambda d_k)$ when defined

end while

$X_{k+1} \leftarrow X_k + \lambda d_k$; cache $F_{k+1} \leftarrow F_{\text{trial}}$

end for

Return nonconvergence if the budget is exhausted

Remark 3.1 (Optional Jacobian-free product). When explicit Jacobian storage is undesirable, a baseline finite-difference product is

$$F'(Y)v \approx \frac{F(Y + h_v v) - F(Y)}{h_v}, \qquad h_v = \sqrt{\varepsilon_{\text{mach}}}\frac{1 + \|Y\|}{\|v\|}, \tag{3.23}$$

for $v \neq 0$. This is an implementation option, not part of the convergence proof. Scaling or noisy residual evaluations may require a problem-specific perturbation rule [9].

4. Convergence Analysis

We separate local order preservation from global residual convergence. The local results apply near a simple root and explain the forcing exponents. The global theorem is a line-search result and therefore uses stronger assumptions on the relevant level set.

4.1. **Assumptions and residual–error equivalence.** Let

$$\mathcal{L}_0 = \{X \in D : \Phi(X) \leq \Phi(X_0)\}. \tag{4.1}$$

Assumption 4.1 (Local smoothness)**.** F is three times continuously differentiable in a neighborhood of the simple root ξ.

Assumption 4.2 (Global level-set regularity)**.** The level set $\mathcal{L}_0$ is compact and contained in D, and there is $M_J > 0$ such that

$$\left\|F'(X)^{-1}\right\| \leq M_J, \qquad X \in \mathcal{L}_0. \tag{4.2}$$

There exists $\delta_D > 0$ such that the closed δ_D-neighborhood

$$\mathcal{N}_{\delta_D}(\mathcal{L}_0) = \{X \in \mathbb{R}^n : \operatorname{dist}(X, \mathcal{L}_0) \leq \delta_D\} \subset D. \tag{4.3}$$

The mapping F is continuously differentiable on an open neighborhood of $\mathcal{N}_{\delta_D}(\mathcal{L}_0)$, and $\nabla\Phi$ is Lipschitz continuous there with constant L_Φ.

Assumption 4.3 (Outer predictor solves)**.** At every outer iteration with $F(X_k) \neq 0$, the retained inexact Newton predictor solve is successful and satisfies

$$J_k d_k^N = -F_k + r_k^N, \qquad \left\|r_k^N\right\| \leq \eta_{k,0}\left\|F_k\right\|, \qquad 0 \leq \eta_{k,0} \leq \bar{\eta} < 1. \tag{4.4}$$

Failures of optional correction or refresh solves are allowed; they only invalidate the high-order candidate and trigger use of d_k^N.

Assumption 4.4 (Local inner-solver regularity)**.** There exists $\delta_{\text{lin}} > 0$ such that, whenever $0 < \|X_k - \xi\| \leq \delta_{\text{lin}}$, the outer predictor solve and every correction solve attempted before the first internal refresh are successful at their requested original-system residual tolerances. The chosen preconditioner also keeps the corresponding correction-block Krylov counts at or below $K_{\text{ref}}^{\max}$. In all local-order

statements the fixed positive forcing floor is disabled. This is an assumption on the inner linear algebra, not an assumption that a refresh is absent.

Lemma 4.5 (Residual–error equivalence)**.** *Under Assumption 4.1, there are constants $c_F, C_F > 0$ and a neighborhood of ξ such that*

$$c_F \|X - \xi\| \le \|F(X)\| \le C_F \|X - \xi\| . \tag{4.5}$$

Proof. Taylor expansion gives

$$F(X) = J_*(X - \xi) + \mathcal{O}(\|X - \xi\|^2). \tag{4.6}$$

The upper bound follows from continuity of F'. Since J_* is nonsingular, its smallest singular value is positive; after shrinking the neighborhood, the quadratic remainder cannot cancel more than a fixed fraction of the leading linear term. This yields the lower bound. □

4.2. **Effect of one inexact inverse action.**

Lemma 4.6 (Linear-solve perturbation)**.** *Let A be nonsingular with $\left\|A^{-1}\right\| \le M$. If $s = A^{-1}b$ and $\widetilde{s}$ satisfies*

$$A\widetilde{s} = b - r, \qquad \|r\| \le \eta \|b\| , \tag{4.7}$$

then

$$\|\widetilde{s} - s\| \le M\eta \|b\| . \tag{4.8}$$

Proof. Subtracting $As = b$ from $A\widetilde{s} = b - r$ gives $A(\widetilde{s} - s) = -r$. Multiplication by A^{-1} proves (4.8). □

4.3. **Inexact predictor.**

Lemma 4.7 (Quadratic predictor under first-order forcing)**.** *Let X_k be sufficiently close to ξ and let $\widetilde{Z}_1 = X_k - \widetilde{s}_0$ where*

$$J_k\widetilde{s}_0 = F(X_k) - r_{k,0}, \qquad \|r_{k,0}\| \le \eta_{k,0} \|F(X_k)\| . \tag{4.9}$$

If

$$\eta_{k,0} = \mathcal{O}(\|e_k\|), \tag{4.10}$$

then

$$\widetilde{Z}_1 - \xi = \mathcal{O}(\|e_k\|^2). \tag{4.11}$$

If $\eta_{k,0} = o(\|e_k\|)$, the perturbation due to the linear solve is $o(\|e_k\|^2)$.

Proof. Let Z_1 be the exact predictor. Standard Newton analysis gives $Z_1 - \xi = \mathcal{O}(\|e_k\|^2)$. Lemmas 4.6 and 4.5 give

$$\left\|\widetilde{Z}_1 - Z_1\right\| \le C\eta_{k,0} \|F(X_k)\| = \mathcal{O}(\|e_k\|^2). \tag{4.12}$$

Adding the exact and inexact errors proves (4.11). The little-o statement follows from the same estimate. □

4.4. **Inexact correction block.**

Lemma 4.8 (Three-power gain under third-order forcing)**.** *Assume the hypotheses of Theorem 2.2. Let Z_j be the* actual *current stage point and suppose (2.9) holds. Keep J_k frozen and let the three approximate solves satisfy*

$$J_k\widetilde{S}_{j,1} = F(Z_j) - r_{j,1}, \qquad \|r_{j,1}\| \le \eta_{k,j,1}\,\|F(Z_j)\|\,, \tag{4.13}$$

$$J_k\widetilde{S}_{j,2} = F(\widetilde{U}_j) - r_{j,2}, \qquad \|r_{j,2}\| \le \eta_{k,j,2}\left\|F(\widetilde{U}_j)\right\|, \tag{4.14}$$

$$J_k\widetilde{S}_{j,3} = F(\widetilde{V}_j) - r_{j,3}, \qquad \|r_{j,3}\| \le \eta_{k,j,3}\left\|F(\widetilde{V}_j)\right\|, \tag{4.15}$$

where

$$\widetilde{U}_j = Z_j - \frac{1}{3}\widetilde{S}_{j,1}, \qquad \widetilde{V}_j = Z_j - \widetilde{S}_{j,2}, \qquad \widetilde{Z}_{j+1} = Z_j - 3\widetilde{S}_{j,3}. \tag{4.16}$$

If

$$\eta_{k,j,\ell} \le C_\eta\,\|e_k\|^3\,, \qquad \ell = 1,2,3, \tag{4.17}$$

for a constant C_η independent of k and the stage, then there is $\widetilde{C}_B > 0$ such that

$$\left\|\widetilde{Z}_{j+1} - \xi\right\| \le \widetilde{C}_B\,\|e_k\|^3\,\|Z_j - \xi\| \tag{4.18}$$

for all sufficiently large k. Consequently, if $Z_j - \xi = \mathcal{O}(\|e_k\|^q)$ with $q \ge 2$, then

$$\widetilde{Z}_{j+1} - \xi = \mathcal{O}(\|e_k\|^{q+3}). \tag{4.19}$$

Proof. Apply the exact block to the same input Z_j and denote its exact intermediate quantities by $S_{j,\ell}, U_j, V_j, Z_{j+1}$. Theorem 2.2 gives

$$\|Z_{j+1} - \xi\| \le C_B\,\|e_k\|^3\,\|Z_j - \xi\|\,. \tag{4.20}$$

Write $z_j = Z_j - \xi$. Since $J_k \to J_*$, there is $M > 0$ such that $\left\|J_k^{-1}\right\| \le M$ for all sufficiently large k. The exact intermediate points satisfy

$$\|U_j - \xi\| + \|V_j - \xi\| \le C\,\|z_j\|\,. \tag{4.21}$$

Indeed, $J_k^{-1}F(W) = \mathcal{O}(\|W - \xi\|)$ uniformly for W near ξ.

For the first solve, Lemmas 4.6 and 4.5 give

$$\left\|\widetilde{S}_{j,1} - S_{j,1}\right\| \le M\eta_{k,j,1}\,\|F(Z_j)\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.22}$$

Therefore

$$\left\|\widetilde{U}_j - U_j\right\| \le C\,\|e_k\|^3\,\|z_j\|\,, \qquad \left\|\widetilde{U}_j - \xi\right\| \le C\,\|z_j\|\,. \tag{4.23}$$

After shrinking the local neighborhood if necessary, U_j and $\widetilde{U}_j$ lie in a fixed compact subset of D. Consequently the local Lipschitz constant for F can be chosen uniformly, and

$$\left\|F(\widetilde{U}_j) - F(U_j)\right\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.24}$$

For the second solve,

$$\widetilde{S}_{j,2} - S_{j,2} = J_k^{-1}\big(F(\widetilde{U}_j) - F(U_j) - r_{j,2}\big), \tag{4.25}$$

so, using (4.24), (4.17), and $\left\|F(\widetilde{U}_j)\right\| \le C\,\|z_j\|$,

$$\left\|\widetilde{S}_{j,2} - S_{j,2}\right\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.26}$$

Consequently,

$$\left\|\widetilde{V}_j - V_j\right\| \le C\,\|e_k\|^3\,\|z_j\|\,, \qquad \left\|\widetilde{V}_j - \xi\right\| \le C\,\|z_j\|\,, \tag{4.27}$$

After shrinking the neighborhood once more if necessary, V_j and $\widetilde{V}_j$ also lie in the same fixed compact subset of D. Hence

$$\left\|F(\widetilde{V}_j) - F(V_j)\right\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.28}$$

The third solve is identical:

$$\left\|\widetilde{S}_{j,3} - S_{j,3}\right\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.29}$$

Since the final update multiplies the third solve by the fixed factor 3,

$$\left\|\widetilde{Z}_{j+1} - Z_{j+1}\right\| \le C\,\|e_k\|^3\,\|z_j\|\,. \tag{4.30}$$

Combining (4.20) and (4.30) proves (4.18). □

Theorem 4.9 (Local order preservation)**.** *Fix $m \ge 2$ and consider an asymptotic frozen-Jacobian cycle with anchor $Y = X_k$ and no internal refresh. Suppose the predictor and all three solves in every correction block are successful and satisfy*

$$\eta_{k,0} \le C_0\,\|e_k\|\,, \qquad \eta_{k,j,\ell} \le C_c\,\|e_k\|^3\,, \qquad \ell = 1,2,3, \tag{4.31}$$

with constants independent of k and j. Then

$$\|X_{k+1} - \xi\| = \mathcal{O}\Big(\|X_k - \xi\|^{3m-1}\Big)\,, \tag{4.32}$$

so the inexact fixed-stage scheme has local order at least $3m-1$.

Moreover, because $Y = X_k$ in an unrefreshed cycle, Lemma 4.5 implies that the residual laws (3.5)–(3.6) satisfy (4.31) for all sufficiently large k (with the finite-precision forcing floor disabled).

Proof. Lemma 4.7 gives an actual predictor stage Z_1 with

$$Z_1 - \xi = \mathcal{O}(\|e_k\|^2). \tag{4.33}$$

Thus (2.9) holds. Lemma 4.8, applied to the actual current stage at each block, increases the error exponent by three. Induction over the $m-1$ blocks gives

$$\|Z_m - \xi\| = \mathcal{O}\Big(\|e_k\|^{2+3(m-1)}\Big) = \mathcal{O}\Big(\|e_k\|^{3m-1}\Big)\,. \tag{4.34}$$

For the forcing laws, near ξ the minimum in (3.4) is attained by its residual argument, and Lemma 4.5 gives

$$\theta(X_k) = \mathcal{O}(\|e_k\|). \tag{4.35}$$

Hence (3.5) is $\mathcal{O}(\|e_k\|)$ and (3.6) is $\mathcal{O}(\|e_k\|^3)$. □

4.5. **Asymptotic inactivity of the safeguards.**

Proposition 4.10 (Uniform local inactivity of the internal safeguards)**.** *Assume Assumptions 4.1 and 4.4. Fix a finite integer $m_{\max} \geq 2$ and parameters*

$$\rho_{\text{cont}} > 0, \qquad \rho_{\text{rej}} > 1, \qquad \chi_{\text{ref}} > 0.$$

Use the forcing laws (3.5)–(3.6) with no positive forcing floor. Then there exist $\delta_{\text{loc}} > 0$, $C_\rho > 0$, and $C_H > 0$ such that, whenever

$$0 < \|e_k\| = \|X_k - \xi\| \leq \delta_{\text{loc}},$$

Algorithm 1 with $\tau_F = 0$ has the following properties; with a positive practical tolerance, the same conclusions hold up to the first stopping event:

(*a*) *every attempted correction block is accepted by the growth-rejection test;*

(*b*) *no internal refresh is triggered;*

(*c*) *every trial and committed block ratio satisfies*

$$\rho_{\text{trial}} = \rho_j \leq C_\rho \|e_k\|^3 ; \tag{4.36}$$

(*d*) *unless an exact root is reached at an intermediate stage, the construction reaches $m_{\max}$ stage points and its final candidate satisfies, with $p = 3m_{\max} - 1$,*

$$\left\| \widehat{X}_{k+1} - \xi \right\| \leq C_H \|e_k\|^p . \tag{4.37}$$

Proof. Near ξ, Lemma 4.5 and (3.4) give

$$\theta(X_k) \leq C_\theta \|e_k\| . \tag{4.38}$$

Therefore, while no refresh has yet occurred and hence $Y = X_k$,

$$\eta_{k,0} \leq C_0 \|e_k\| , \qquad \eta_{k,j,\ell} \leq C_c \|e_k\|^3 , \qquad \ell = 1, 2, 3, \tag{4.39}$$

for constants independent of k and of the finitely many stages.

Lemma 4.7 then gives

$$\|Z_1 - \xi\| \leq C_1 \|e_k\|^2 . \tag{4.40}$$

Choose a local neighborhood small enough that Theorem 2.2, Lemma 4.8, and Lemma 4.5 all hold there. Suppose inductively that the current committed stage Z_j is in this neighborhood and satisfies the input-size condition (2.9). Lemma 4.8 gives the *multiplicative* estimate

$$\|Z_{\text{trial}} - \xi\| \le \widetilde{C}_B \|e_k\|^3 \|Z_j - \xi\| . \tag{4.41}$$

If $Z_j \neq \xi$, residual–error equivalence therefore yields

$$\rho_{\text{trial}} = \frac{\|F(Z_{\text{trial}})\|}{\|F(Z_j)\|} \le \frac{C_F}{c_F} \widetilde{C}_B \|e_k\|^3 . \tag{4.42}$$

This direct estimate is essential: one must not obtain the ratio by dividing two unrelated big-$\mathcal{O}$ bounds. Set

$$C_\rho = \frac{C_F}{c_F} \widetilde{C}_B .$$

After reducing δ_{loc} if necessary, $C_\rho \|e_k\|^3$ is smaller than both ρ_{rej} and ρ_{cont}. Thus the trial is committed and the continuation test is satisfied. Equation (4.41) also shows that the new stage is again of size at most $C \|e_k\|^2$, so the argument can be repeated for the finitely many blocks.

It remains to exclude the defect refresh. By the inverse function theorem, there are neighborhoods U_0 of ξ and V_0 of 0 such that $F : U_0 \to V_0$ is a C^1 diffeomorphism. Choose a closed ball $\overline{B}_r(0) \subset V_0$ and put $U = F^{-1}(B_r(0))$. Since $D(F^{-1})$ is continuous, it is bounded on $\overline{B}_r(0)$; because the ball is convex, the mean-value integral for F^{-1} gives a constant $M_{\text{inv}} > 0$ such that

$$\|Z - Y\| \le M_{\text{inv}} \|F(Z) - F(Y)\| , \qquad Y, Z \in U.$$

Thus, with $c_{\text{bi}} = M_{\text{inv}}^{-1} > 0$,

$$\|F(Z) - F(Y)\| \ge c_{\text{bi}} \|Z - Y\| \tag{4.43}$$

for all $Y, Z \in U$. Reduce δ_{loc} once more so that the anchor and every committed local stage considered below lie in U. Local Lipschitz continuity of F' gives

$$\left\|F(Z) - F(Y) - F'(Y)(Z - Y)\right\| \le C_{\text{lin}} \|Z - Y\|^2 . \tag{4.44}$$

Hence, for $Z \neq Y$,

$$\chi(Z; Y, F'(Y)) \le \frac{C_{\text{lin}}}{c_{\text{bi}}} \|Z - Y\| . \tag{4.45}$$

Before a refresh, $Y = X_k$, and every committed stage satisfies

$$\|Z - Y\| \le \|Z - \xi\| + \|e_k\| \le C_Z \|e_k\|^2 + \|e_k\| \le C \|e_k\| . \tag{4.46}$$

Thus (4.45) is below χ_{ref} for all sufficiently small $\|e_k\|$. The Krylov branch is inactive by Assumption 4.4. Therefore no refresh occurs.

Starting with (4.40) and applying (4.41) $m_{\max} - 1$ times gives

$$\left\|\widehat{X}_{k+1} - \xi\right\| \le C_H \|e_k\|^{2+3(m_{\max}-1)} = C_H \|e_k\|^{3m_{\max}-1} , \tag{4.47}$$

unless an intermediate stage equals ξ, in which case the nonlinear system has already been solved exactly. □

4.6. **Eventual acceptance of the high-order full step.**

Theorem 4.11 (Local convergence and eventual full-step acceptance of GAIN-FJ)**.** *Assume the hypotheses of Proposition 4.10, and consider the theoretical iteration with the positive stopping tolerance disabled. Let*

$$p = 3m_{\max} - 1. \tag{4.48}$$

There exists $\delta_* > 0$ such that, for every starting point satisfying

$$0 < \|X_0 - \xi\| \le \delta_*, \tag{4.49}$$

either the method reaches ξ in finitely many steps, or every sufficiently local outer step selects the high-order direction, accepts $\lambda_k = 1$, remains in the same neighborhood, and satisfies

$$\|X_{k+1} - \xi\| \le C_H \|X_k - \xi\|^p . \tag{4.50}$$

Consequently $X_k \to \xi$ with Q-order at least $p = 3m_{\max} - 1$.

Proof. For a nonzero error e_k sufficiently small, Proposition 4.10 gives a valid candidate satisfying

$$\widehat{X}_{k+1} - \xi = \mathcal{O}(\|e_k\|^p). \tag{4.51}$$

Hence

$$d_k^H = \widehat{X}_{k+1} - X_k = -e_k + \mathcal{O}(\|e_k\|^p). \tag{4.52}$$

Using

$$F_k = J_* e_k + \mathcal{O}(\|e_k\|^2), \qquad J_k = J_* + \mathcal{O}(\|e_k\|),$$

we obtain

$$J_k d_k^H = -F_k + \mathcal{O}(\|e_k\|^2) \tag{4.53}$$

and therefore

$$F_k^{\mathsf{T}} J_k d_k^H = -\|F_k\|^2 + \mathcal{O}(\|e_k\|^3). \tag{4.54}$$

Since $\|F_k\|^2 = \Theta(\|e_k\|^2)$, the quotient of the left-hand side of (4.54) by $\|F_k\|^2$ tends to -1 as $e_k \to 0$. Thus, after shrinking the neighborhood, the descent safeguard holds for any fixed $0 < \sigma_d < 1$.

Lemma 4.7 gives

$$d_k^N = -e_k + \mathcal{O}(\|e_k\|^2). \tag{4.55}$$

Therefore

$$\frac{\|d_k^H\|}{\|d_k^N\|} \to 1 \qquad (e_k \to 0), \tag{4.56}$$

so the size safeguard holds for any fixed $\omega_d > 1$ after a further reduction of the neighborhood.

For the full step,

$$\Phi(\widehat{X}_{k+1}) = \mathcal{O}(\|e_k\|^{2p}), \tag{4.57}$$

whereas (4.54) gives

$$\Phi(X_k) + c_A F_k^{\mathsf{T}} J_k d_k^H = \left(\frac{1}{2} - c_A\right) \|F_k\|^2 + \mathcal{O}(\|e_k\|^3). \tag{4.58}$$

Because $0 < c_A < 1/2$ and $p \geq 5$, the right-hand side is positive of order $\|e_k\|^2$, while the left-hand side is of higher order. Hence $\lambda_k = 1$ is accepted whenever e_k is sufficiently small.

Proposition 4.10 provides the uniform constant C_H in (4.50). Choose $\delta_* > 0$ small enough that all preceding local conclusions hold and

$$C_H \delta_*^{p-1} \leq \frac{1}{2}. \tag{4.59}$$

Then (4.50) implies

$$\|e_{k+1}\| \leq \frac{1}{2} \|e_k\| \leq \delta_* \tag{4.60}$$

whenever $0 < \|e_k\| \leq \delta_*$. Induction proves that the iterates remain in the local neighborhood and converge to ξ; the bound (4.50) gives Q-order at least p. If an iterate or any residual-evaluated internal point equals ξ exactly, the immediate stopping rule in Algorithm 1 gives finite termination instead. □

4.7. **Global residual convergence.**

Theorem 4.12 (Residual global convergence)**.** *Consider the theoretical outer iteration with no finite outer-iteration cutoff. Assume Assumptions 4.2 and 4.3. Suppose Algorithm 2 uses a high-order direction only when (3.20)–(3.21) hold, and treats an undefined or nonfinite line-search trial as unacceptable and continues backtracking. Then Armijo backtracking terminates after finitely many reductions at every iteration with $F(X_k) \neq 0$, and all accepted iterates remain in $\mathcal{L}_0$. Moreover, either the method reaches an exact root in finitely many outer iterations, or it generates an infinite sequence satisfying*

$$\|F(X_k)\| \to 0. \tag{4.61}$$

Optional correction-block, refresh, or candidate-evaluation failures do not affect this conclusion because they only force the method to use the retained inexact Newton direction.

Proof. For the retained inexact Newton direction, Assumption 4.3 gives

$$F_k^{\mathsf{T}} J_k d_k^N = -\|F_k\|^2 + F_k^{\mathsf{T}} r_k^N \leq -(1-\bar{\eta}) \|F_k\|^2 . \tag{4.62}$$

If a high-order direction is selected, (3.20) gives

$$F_k^{\mathsf{T}} J_k d_k^H \leq -\sigma_d \|F_k\|^2 . \tag{4.63}$$

Thus every selected direction d_k satisfies

$$F_k^{\mathsf{T}} J_k d_k \leq -c_d \|F_k\|^2 , \qquad c_d := \min\{1-\bar{\eta}, \sigma_d\} > 0. \tag{4.64}$$

Because J_k is nonsingular on $\mathcal{L}_0$, (4.4) implies

$$\left\|d_k^N\right\| \le \left\|J_k^{-1}\right\| \left(\|F_k\| + \left\|r_k^N\right\|\right) \le M_J(1+\bar{\eta})\,\|F_k\|\,. \tag{4.65}$$

The size safeguard gives the same bound for an accepted high-order direction up to the factor ω_d. Hence

$$\|d_k\| \le C_d\,\|F_k\|\,, \qquad C_d := \omega_d M_J(1+\bar{\eta}). \tag{4.66}$$

Every accepted Armijo step decreases Φ, so, inductively, $X_k \in \mathcal{L}_0$. Since $\mathcal{L}_0$ is compact and F is continuous,

$$M_F := \max_{X\in\mathcal{L}_0} \|F(X)\| < \infty. \tag{4.67}$$

Thus (4.66) also gives the uniform direction bound

$$\|d_k\| \le C_d M_F =: M_d. \tag{4.68}$$

If $M_d = 0$, then $F(X_k) = 0$ and there is nothing to prove. Otherwise, for

$$0 < \lambda \le \lambda_D := \min\left\{1, \frac{\delta_D}{M_d}\right\}, \tag{4.69}$$

we have

$$\operatorname{dist}(X_k + \lambda d_k, \mathcal{L}_0) \le \lambda\,\|d_k\| \le \delta_D. \tag{4.70}$$

Hence every sufficiently short trial lies in the neighborhood (4.3), so it belongs to D and has a finite residual.

On that neighborhood, the Lipschitz continuity of $\nabla\Phi$ yields the descent lemma:

$$\Phi(X_k + \lambda d_k) \le \Phi(X_k) + \lambda F_k^{\mathsf{T}} J_k d_k + \frac{L_\Phi}{2}\lambda^2\,\|d_k\|^2\,. \tag{4.71}$$

Using (4.64) and (4.66), the Armijo inequality is guaranteed whenever

$$0 < \lambda \le \lambda_A := \frac{2(1-c_A)c_d}{L_\Phi C_d^2}, \tag{4.72}$$

with the convention $\lambda_A = +\infty$ if $L_\Phi = 0$. Indeed,

$$\Phi(X_k + \lambda d_k) - \Phi(X_k) - c_A\lambda F_k^{\mathsf{T}} J_k d_k \le (1-c_A)\lambda F_k^{\mathsf{T}} J_k d_k + \frac{L_\Phi}{2}\lambda^2\,\|d_k\|^2 \tag{4.73}$$

$$\le -\lambda(1-c_A)c_d\,\|F_k\|^2 + \frac{L_\Phi}{2}\lambda^2 C_d^2\,\|F_k\|^2 \le 0. \tag{4.74}$$

Therefore geometric backtracking must terminate after finitely many reductions. Moreover, because trials are generated as $1, \beta, \beta^2, \ldots$, the accepted step lengths satisfy the uniform bound

$$\lambda_k \ge \underline{\lambda} := \beta\min\{1, \lambda_D, \lambda_A\} > 0. \tag{4.75}$$

The accepted Armijo decrease and (4.64) now give

$$\Phi(X_{k+1}) \le \Phi(X_k) - c_A\lambda_k c_d\,\|F_k\|^2 \le \Phi(X_k) - c_A\underline{\lambda} c_d\,\|F_k\|^2\,. \tag{4.76}$$

Summing from $k = 0$ and using $\Phi \geq 0$ yields

$$\sum_{k=0}^{\infty} \|F(X_k)\|^2 < \infty. \tag{4.77}$$

Consequently $\|F(X_k)\| \to 0$. Since the accepted iterates lie in the compact set $\mathcal{L}_0$, every accumulation point of an infinite sequence is therefore a root of F. □

Remark 4.13. The uniform nonsingularity condition (4.2) is intentionally explicit. It is stronger than the assumptions used in the local theorem. Near singular roots or strongly ill-conditioned branches, a trust-region, pseudo-transient, or Levenberg–Marquardt fallback would be a distinct extension and is not covered by Theorem 4.12.

5. Work Model and Implementation Principles

5.1. **Fixed-stage counts.** With the stage convention of Section 2, a fixed m-stage frozen-Jacobian cycle uses

$$N_{\text{solve}} = 1 + 3(m - 1) = 3m - 2 \tag{5.1}$$

linear systems with one coefficient matrix. If $F(X_k)$ is already cached, candidate construction requires one new residual evaluation at the predictor and three per correction block, so

$$N_F^{\text{new}} = 3m - 2. \tag{5.2}$$

These counts explain why a direct dense efficiency index is not sufficient for a large sparse implementation: the cost of each inverse action depends on the Krylov and preconditioner behavior.

5.2. **Sparse cost model.** A useful outer-iteration model is

$$T_k \approx N_{F,k}T_F + N_{J,k}T_J + N_{P,k}T_P + \sum_{\ell=1}^{N_{\text{lin},k}} K_{k,\ell}(T_{Jv} + T_{P^{-1}}) + T_{\text{orth},k} + T_{\text{misc},k}, \tag{5.3}$$

where $N_{F,k}$ counts *all new* residual evaluations in outer iteration k, including line-search trials and excluding cached reuse; $N_{J,k}$ and $N_{P,k}$ count Jacobian and preconditioner constructions; and $N_{\text{lin},k}$ is the number of linear systems attempted. The quantities T_F, T_J, T_P, T_{Jv}, and $T_{P^{-1}}$ denote the corresponding unit costs. The term $T_{\text{orth},k}$ collects GMRES orthogonalization/restart costs, and $T_{\text{misc},k}$ collects lower-order vector updates and scalar tests. This accounting avoids counting a residual evaluation twice when the same value is reused by the line search. It also makes clear that the Krylov iteration count represents only one part of the total computational effort. The numerical study therefore reports residual evaluations, Jacobian constructions, preconditioner constructions, Krylov iterations, and accepted outer steps separately. Since a matched campaign of repeated timings is not available, we use these counters as indicators of computational work rather than as runtime measurements. This is more informative than comparing nonlinear iteration counts alone.

The expected favorable regime for GAIN-FJ is one in which $T_J + T_P$ is significant and the repeated systems with a frozen matrix remain sufficiently well preconditioned. The method need not be advantageous if a standard Newton–Krylov step is already very cheap or if the frozen systems require many Krylov iterations.

5.3. **A setup-reuse break-even criterion.** The work counters can be converted into an explicit cost condition without inventing wall-clock data. Consider a matched GAIN-FJ–NK comparison on the same nonlinear problem, and use superscripts G and N for GAIN-FJ and NK, respectively. Let

$$\Delta N_F = N_F^G - N_F^N, \qquad S_J = N_J^N - N_J^G, \qquad S_P = N_P^N - N_P^G, \qquad \Delta K = K^G - K^N,$$

For a transparent diagnostic, collapse the average matrix-vector, preconditioner-application, and orthogonalization cost associated with one reported Krylov iteration into a representative quantity T_K. Ignoring only unmatched lower-order overhead, the modeled cost difference is

$$T^G - T^N \approx \Delta N_F T_F + \Delta K T_K - S_J T_J - S_P T_P. \tag{5.4}$$

Hence the setup-reuse strategy is predicted to be cheaper whenever

$$S_J T_J + S_P T_P > \Delta N_F T_F + \Delta K T_K. \tag{5.5}$$

If $S_J = S_P = s$, this reduces to $s(T_J + T_P) > \Delta N_F T_F + \Delta K T_K$. Equation (5.5) is not a timing result; it states exactly how expensive Jacobian/preconditioner setup must be, relative to residual and Krylov work, for the measured counter savings to translate into lower modeled cost. Section 6.9 evaluates this condition for the largest PDE runs.

5.4. **Preconditioner reuse.** All three systems in a correction block share the same active matrix. A preconditioner can therefore be reused across those right-hand sides. This is a central design feature rather than a bookkeeping detail: updating a good preconditioner can itself be expensive, and earlier work has shown that preconditioner-update strategies can materially affect Newton–Krylov efficiency [13]. The experiments below use incomplete LU so that setup reuse has a measurable cost; a production PDE solver would normally exploit multilevel or physics-based preconditioning.

6. NUMERICAL EXPERIMENTS

The numerical section is generated by the accompanying MATLAB reproduction code. The recorded environment is MATLAB `24.2.0.2712019 (R2024b)`. The arbitrary-precision calculations use variable-precision symbolic arithmetic, while the sparse tests use IEEE double precision. The code writes machine-readable CSV files, MATLAB data, LaTeX fragments, logs, and figures under the `code/results/` tree.

A run-level *linear-failure* is reserved for failure of the retained outer predictor solve to satisfy the requested original-system residual tolerance within the allowed Krylov budget. Failure of an optional correction or refresh solve invalidates the high-order candidate and invokes the retained Newton fallback, consistently with Algorithms 1–2. In the sparse experiments, explicit right preconditioning is used so that the GMRES residual is exactly the original-system residual $b - As$.

6.1. **MATLAB implementation and common settings.** The sparse experiments use MATLAB's built-in restarted `gmres` routine and `ilu` with `type='ilutp'`. For a preconditioner M, the code applies GMRES to the right-preconditioned operator AM^{-1} and reconstructs $s = M^{-1}y$. Thus the Krylov stopping test directly certifies

$$\frac{\|b - As\|}{\|b\|} \le \eta.$$

The common nonlinear tolerance is $\tau_F = 10^{-10}$. Restarted GMRES uses restart length 50 and at most 20 restart cycles for the PDE tests. The default PDE preconditioner uses ILUTP drop tolerance 10^{-4}. Table 2 lists the principal parameters.

The matched NK baseline uses the same scaled residual, predictor forcing, right-preconditioned GMRES implementation, ILUTP construction, and Armijo line search. Consequently, differences in the tables are attributable to the nonlinear strategy and its Jacobian/preconditioner reuse rather than to different inner stopping conventions.

6.2. **Arbitrary-precision local-order verification.** For local-order verification we use the cyclic exponential system

$$f_r(x) = e^{x_r} + x_{r+1}^2 - 1, \qquad r = 1, \dots, n-1, \tag{6.1}$$

$$f_n(x) = e^{x_n} + x_1^2 - 1. \tag{6.2}$$

The zero vector is a simple root. We set $n = 5$ and

$$X_0 = (0.672, 0.736, 0.800, 0.864, 0.928)^{\mathsf{T}}. \tag{6.3}$$

Four nonlinear cycles are executed so that the asymptotic order can be observed. The computational order is

$$\delta_k = \frac{\log(\|e_{k+1}\| \,/\, \|e_k\|)}{\log(\|e_k\| \,/\, \|e_{k-1}\|)}, \qquad e_k = X_k. \tag{6.4}$$

Working precisions of 300, 800, 2200, and 5200 decimal digits are used for $m = 2, 3, 4, 5$, respectively. The full reference drives each five-dimensional Krylov solve through the complete basis, whereas the adaptive calculation uses (3.5)–(3.6) without a finite-precision forcing floor.

TABLE 2. Reference parameters used in the MATLAB double-precision GAIN-FJ experiments.

Parameter	Value	Purpose
τ_F	10^{-10}	scaled nonlinear residual tolerance
$m_{\max}$	5	maximum stage points per frozen epoch
$\bar{\eta}$	0.5	maximum relative linear residual
c_p, c_c	0.1, 0.1	predictor/correction forcing scales
$\eta_{\min}$	10^{-12}	finite-precision PDE forcing floor
η_{fixed}	10^{-6}	fixed forcing in A0–A3 ablations
ρ_{cont}	0.5	continue a successful correction block
ρ_{rej}	2.0	reject unsafe residual growth
χ_{ref}	0.5	linearization-defect refresh threshold
$K_{\text{ref}}^{\max}$	150	Krylov deterioration threshold
$N_{\text{ref}}^{\max}$	2	maximum internal refreshes per outer step
β	0.5	backtracking factor
c_A	10^{-4}	Armijo parameter
σ_d	10^{-3}	high-order descent safeguard
ω_d	10	high-order direction-size safeguard
GMRES restart	50	Krylov restart length
GMRES max cycles	20	maximum PDE restart cycles per solve
ILUTP drop tolerance	10^{-4}	default PDE setting

TABLE 3. MATLAB arbitrary-precision verification of the local order on (6.1)–(6.2).

m	Linear-solve strategy	p_m lower bound	δ_1	δ_2	δ_3
2	full GMRES reference	5	3.004061	4.815117	4.997814
2	adaptive forcing	5	3.004938	4.814807	4.998031
3	full GMRES reference	8	5.443562	7.949142	7.999669
3	adaptive forcing	8	5.403755	7.952145	7.999869
4	full GMRES reference	11	8.104562	10.993809	11.000845
4	adaptive forcing	11	8.043460	10.989613	11.000431
5	full GMRES reference	14	10.915709	13.996775	14.000033
5	adaptive forcing	14	10.842990	13.995791	13.999990

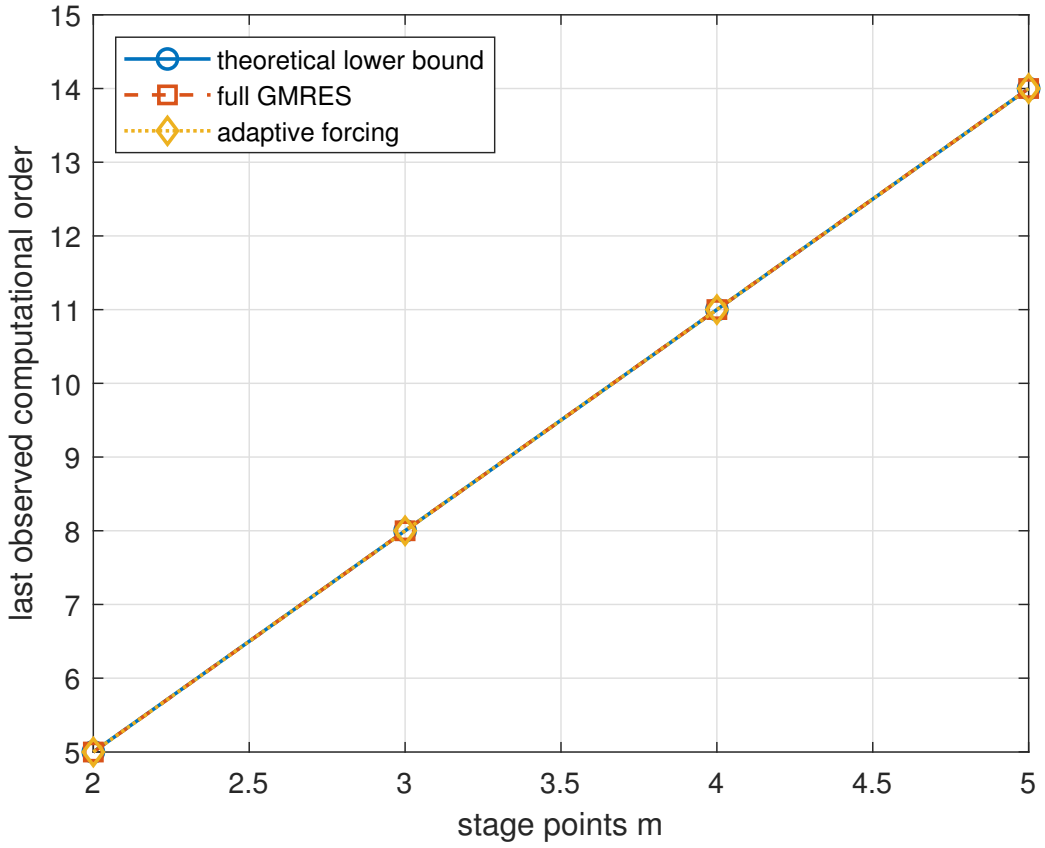


FIGURE 1. Observed computational orders from the arbitrary-precision experiment.

The last computational orders approach 5, 8, 11, and 14 for both solve strategies. Thus Table 3 and Figure 1 are consistent with the proved lower bound $p_m = 3m - 1$; for these tests the observed asymptotic order attains the bound to the displayed accuracy.

6.3. **Linear work to the nonlinear tolerance.** The same problem is repeated at 500 decimal digits and terminated immediately when $R(X_k) \le 10^{-10}$. Table 4 gives the measured work.

Adaptive forcing reduces the measured Krylov work by 30.0–45.4% at the same stage count while reaching the same nonlinear tolerance. The savings increase with the stage count in this experiment.

6.4. **Globalization diagnostic on a difficult algebraic start.** We use (6.1)–(6.2) with $n = 20000$ and

$$x_{0,i} = c\left[1 + 0.10\sin\left(\frac{2\pi i}{n}\right) + 0.05\cos\left(\frac{6\pi i}{n}\right)\right], \qquad c = -0.25. \tag{6.5}$$

A Jacobi preconditioner is used. A0–A3 use the fixed forcing tolerance $\eta_{\text{fixed}} = 10^{-6}$; full GAIN-FJ uses the stage-dependent forcing law. Table 5 reports the resulting work counts.

The two nonglobalized Krylov variants A0–A1 do not converge from this start, whereas A2 converges after Armijo globalization is enabled. The adaptive-refresh test is inactive in A2/A3, as their work counts coincide. Replacing the fixed forcing by stage-dependent forcing reduces the Krylov count from

TABLE 4. MATLAB linear work required to reach $R(X_k) \le 10^{-10}$ in the high-precision cyclic test. Savings are relative to the full-GMRES reference at the same m.

m	Strategy	cycles	Krylov its.	final R	saving
2	full GMRES reference	3	60	3.133×10^{-18}	–
2	adaptive forcing	3	42	3.267×10^{-18}	30.0%
3	full GMRES reference	3	105	2.288×10^{-68}	–
3	adaptive forcing	3	73	1.415×10^{-67}	30.5%
4	full GMRES reference	2	100	3.902×10^{-17}	–
4	adaptive forcing	2	59	5.571×10^{-17}	41.0%
5	full GMRES reference	2	130	1.006×10^{-26}	–
5	adaptive forcing	2	71	1.548×10^{-26}	45.4%

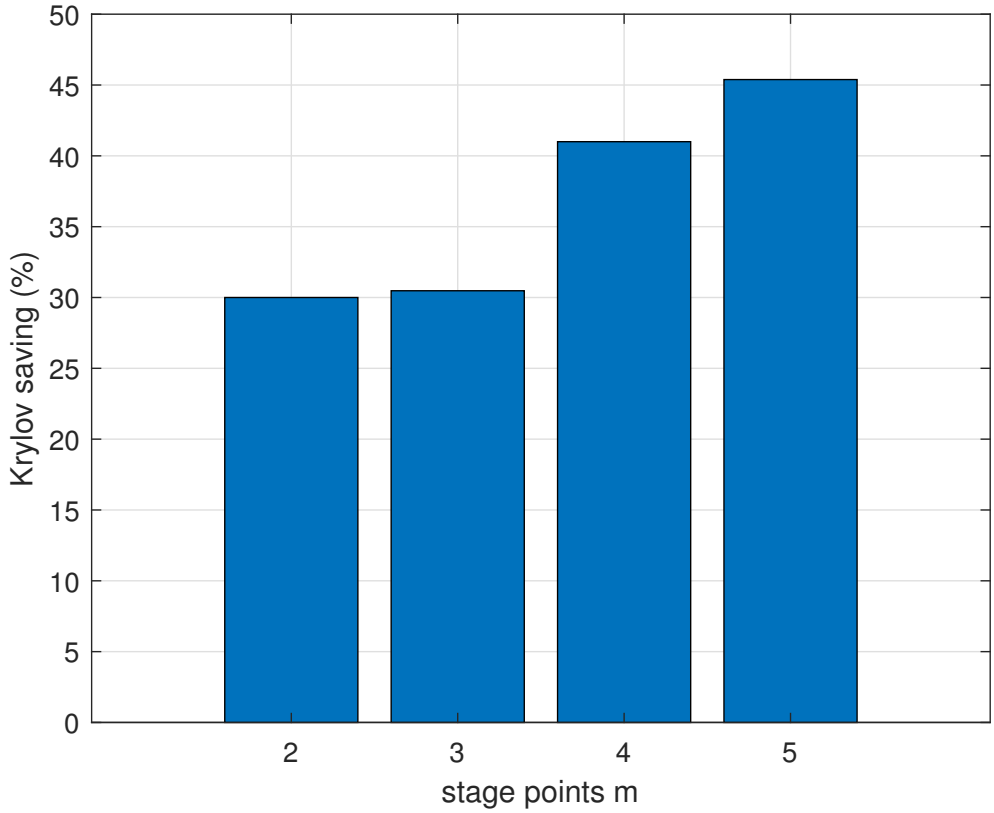


FIGURE 2. Krylov-iteration savings from adaptive forcing in the high-precision work-to-tolerance experiment.

99 to 32, a reduction of

$$\frac{99 - 32}{99} \times 100\% \approx 67.7\%,$$

TABLE 5. MATLAB globalization diagnostic for the cyclic exponential system at $n = 20000$ and $c = -0.25$.

Method	status	accepted	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
fixed direct high-order	nonconvergence	0	4	1	0	0	0	9.572×10^{-1}
NK baseline	converged	5	7	5	5	9	1	6.460×10^{-14}
fixed-stage Krylov A0	nonconvergence	0	4	1	1	41	0	9.572×10^{-1}
adaptive stages A1	nonconvergence	0	4	1	1	41	0	9.572×10^{-1}
A2 + line search	converged	3	19	3	3	99	1	1.309×10^{-14}
A3 + refresh	converged	3	19	3	3	99	1	1.309×10^{-14}
full GAIN-FJ	converged	3	20	3	3	32	1	4.013×10^{-16}

with the same three accepted outer updates. The matched NK baseline remains much cheaper in Krylov work on this structured cyclic problem.

6.5. **Refresh-policy comparison on eight cyclic starts.** The refresh policy is compared on eight starts with $n = 20000$ and

$$c \in \{-0.25,\ 0.2,\ 0.5,\ 0.8,\ 1.2,\ 2.0,\ 2.5,\ 3.0\}.$$

All three policies use globalization and adaptive forcing. The aggregate results are reported in Table 6.

TABLE 6. MATLAB refresh-policy comparison on eight cyclic starts at $n = 20000$.

Policy	success	$\#F$	$\#J$	$\#P$	Krylov	backtracks
no internal refresh	8/8	176	20	20	388	1
refresh every accepted block	8/8	107	26	26	253	1
adaptive refresh	8/8	156	24	24	386	1

All policies converge on all eight starts. Refreshing after every accepted block uses the fewest residual evaluations and Krylov iterations (107 and 253), at the cost of 26 Jacobian/preconditioner setups. The adaptive rule uses two fewer setups, but 386 Krylov iterations; its Krylov work is essentially the same as no internal refresh (388). Thus the current defect/Krylov refresh heuristic is not uniformly work-optimal on this inexpensive cyclic problem.

6.6. **A larger cyclic check with 100000 unknowns.** The $n = 100000$ check uses the same nonconstant profile as (6.5) with the explicitly specified amplitude $c = 0.8$. Table 7 gives the corresponding counts.

The full method reduces A3 from 65 to 44 Krylov iterations while preserving the same setup count. By comparison, the matched NK run needs only 10 Krylov iterations and fewer residual evaluations.

TABLE 7. MATLAB cyclic exponential check at $n = 100000$ and $c = 0.8$.

Method	status	accepted	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
fixed direct high-order	converged	2	21	2	0	0	0	1.004×10^{-14}
NK baseline	converged	6	7	6	6	10	0	4.481×10^{-16}
A0 fixed-stage Krylov	converged	2	21	2	2	60	0	1.004×10^{-14}
A1 adaptive stages	converged	2	21	2	2	60	0	1.004×10^{-14}
A2 + line search	converged	2	21	2	2	60	0	1.004×10^{-14}
A3 + refresh	converged	2	19	3	3	65	0	6.551×10^{-17}
full GAIN-FJ	converged	2	19	3	3	44	0	6.585×10^{-17}

Hence, for this structured problem, the higher formal order does not reduce the amount of linear-solver work.

6.7. **Two-dimensional sparse PDE models.** The PDE experiments are based on three sparse nonlinear systems produced by finite-difference discretizations on the unit square. If q denotes the number of interior grid points along each coordinate axis, then $h = (q+1)^{-1}$ and $n = q^2$. We write L_h for the standard five-point approximation of $-\Delta$ and D_x, D_y for centered first-derivative matrices incorporating homogeneous Dirichlet boundary conditions. All problems start from the zero vector unless stated otherwise.

P1: reaction–diffusion.

$$F_h(u) = L_h u + \alpha u^{\circ 3} - f_h = 0, \qquad \alpha = 50, \tag{6.6}$$

where $u^{\circ 3}$ is componentwise. The manufactured discrete solution is

$$u^*_{ij} = \sin(\pi x_i)\sin(\pi y_j), \qquad f_h = L_h u^* + \alpha (u^*)^{\circ 3},$$

and

$$J_h(u) = L_h + 3\alpha \,\mathrm{diag}(u^{\circ 2}). \tag{6.7}$$

P2: nonsymmetric convection–diffusion–reaction. Let

$$A_h = \nu L_h + b_x D_x + b_y D_y, \qquad \nu = 0.02, \quad b_x = 2, \quad b_y = 1.$$

We solve

$$F_h(u) = A_h u + \alpha \sinh(u) - f_h = 0, \qquad \alpha = 8, \tag{6.8}$$

with componentwise $\sinh$ and manufactured solution

$$u^*_{ij} = 0.8 \sin(\pi x_i)\sin(\pi y_j), \qquad f_h = A_h u^* + \alpha \sinh(u^*).$$

The Jacobian is

$$J_h(u) = A_h + \alpha \operatorname{diag}(\cosh u). \tag{6.9}$$

P3: two-dimensional Bratu problem.

$$F_h(u) = L_h u - \lambda \exp(u) = 0, \qquad \lambda = 6, \tag{6.10}$$

where $\exp(u)$ is componentwise, with

$$J_h(u) = L_h - \lambda \operatorname{diag}(\exp u). \tag{6.11}$$

No exact discrete root is used; convergence is assessed through the scaled nonlinear residual.

6.8. **PDE work counts with ILUTP.** Table 8 contains the deterministic runs for $q = 64, 128, 192$ with ILUTP drop tolerance 10^{-4}. Every reported run converges.

TABLE 8. PDE work counts with ILUTP drop tolerance 10^{-4}. "Out" is the number of accepted outer updates.

Problem	n	Method	status	Out	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
P1 RD	4096	NK	converged	5	7	5	5	12	1	1.030×10^{-11}
		GAIN-FJ	converged	3	19	3	3	26	1	3.412×10^{-14}
P1 RD	16384	NK	converged	5	7	5	5	15	1	1.816×10^{-11}
		GAIN-FJ	converged	3	19	3	3	43	1	1.221×10^{-13}
P1 RD	36864	NK	converged	5	7	5	5	21	1	5.542×10^{-12}
		GAIN-FJ	converged	3	19	3	3	48	1	2.977×10^{-13}
P2 CDR	4096	NK	converged	4	5	4	4	9	0	3.402×10^{-15}
		GAIN-FJ	converged	1	11	1	1	10	0	6.912×10^{-11}
P2 CDR	16384	NK	converged	4	5	4	4	11	0	1.241×10^{-14}
		GAIN-FJ	converged	1	11	1	1	10	0	2.690×10^{-11}
P2 CDR	36864	NK	converged	4	5	4	4	12	0	2.815×10^{-14}
		GAIN-FJ	converged	1	14	1	1	17	0	1.265×10^{-13}
P3 Bratu	4096	NK	converged	5	6	5	5	24	0	1.236×10^{-13}
		GAIN-FJ	converged	2	15	2	2	33	0	4.318×10^{-13}
P3 Bratu	16384	NK	converged	5	6	5	5	35	0	4.747×10^{-13}
		GAIN-FJ	converged	2	15	2	2	45	0	6.747×10^{-13}
P3 Bratu	36864	NK	converged	5	6	5	5	45	0	1.014×10^{-12}
		GAIN-FJ	converged	2	15	2	2	59	0	1.173×10^{-12}

Across these nine problem/size pairs, GAIN-FJ uses fewer accepted outer updates and fewer Jacobian/preconditioner constructions in every comparison: P1 uses 3 updates instead of 5, P2 uses 1 instead of 4, and P3 uses 2 instead of 5. This setup reduction is not accompanied by a uniform Krylov reduction. For P1 and P3, GAIN-FJ uses more Krylov iterations than NK at all three sizes. For P2 the counts are comparable: 10 versus 9 at $q = 64$, 10 versus 11 at $q = 128$, and 17 versus 12 at $q = 192$. Therefore the PDE evidence supports Jacobian/preconditioner reuse, but not a general claim of lower linear work.

6.9. **A 102400-unknown stress test.** The same experiment is repeated at $q = 320$ $(n = 102400)$, with results shown in Table 9.

TABLE 9. MATLAB stress test at $q = 320$ $(n = 102400)$, ILUTP drop tolerance 10^{-4}.

Problem	Method	status	Out	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
P1 RD	NK	converged	5	7	5	5	30	1	1.137×10^{-11}
	GAIN-FJ	converged	3	19	3	3	68	1	8.132×10^{-13}
P2 CDR	NK	converged	4	5	4	4	18	0	7.386×10^{-14}
	GAIN-FJ	converged	1	11	1	1	20	0	2.186×10^{-11}
P3 Bratu	NK	converged	5	6	5	5	71	0	3.015×10^{-12}
	GAIN-FJ	converged	2	15	2	2	92	0	3.238×10^{-12}

All six runs converge. GAIN-FJ again reduces the number of Jacobian/preconditioner setups substantially, but its Krylov counts are larger: 68 versus 30 on P1, 20 versus 18 on P2, and 92 versus 71 on P3. The $n = 102400$ results therefore reinforce the setup-reuse interpretation rather than a linear-work superiority claim.

Table 10 applies the criterion (5.5) to these largest runs. Here T_K denotes the representative cost per reported Krylov iteration introduced in Section 5.3. For P2 the additional Krylov burden is small, so

TABLE 10. Break-even conditions implied by the $q = 320$ work counters. A condition being satisfied would predict lower modeled cost for GAIN-FJ; the table is not a substitute for measured timing.

Problem	setups saved	extra $\#F$	extra Krylov	condition for lower modeled GAIN-FJ cost
P1 RD	2 Jacobians, 2 preconditioners	12	38	$2(T_J + T_P) > 12T_F + 38T_K$
P2 CDR	3 Jacobians, 3 preconditioners	6	2	$3(T_J + T_P) > 6T_F + 2T_K$
P3 Bratu	3 Jacobians, 3 preconditioners	9	21	$3(T_J + T_P) > 9T_F + 21T_K$

setup reuse can be favorable even when setup costs are only moderately larger than residual and Krylov costs. P1 and P3 require substantially more expensive setup operations before the saved constructions can compensate for the extra repeated work. These inequalities make the practical claim deliberately conditional and identify the measurements needed in a future timing study.

6.10. **ILUTP drop-tolerance sensitivity.** Table 11 compares ILUTP drop tolerances 10^{-3} and 10^{-4} at $q = 128$. Every run converges.

TABLE 11. Sensitivity to the ILUTP drop tolerance at $q = 128$ ($n = 16384$).

drop	Problem	Method	status	Out	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
10^{-3}	P1 RD	NK	converged	5	7	5	5	28	1	1.098×10^{-11}
		GAIN-FJ	converged	3	19	3	3	68	1	1.214×10^{-13}
10^{-3}	P2 CDR	NK	converged	4	5	4	4	19	0	1.254×10^{-14}
		GAIN-FJ	converged	1	11	1	1	20	0	1.435×10^{-11}
10^{-3}	P3 Bratu	NK	converged	5	6	5	5	70	0	4.678×10^{-13}
		GAIN-FJ	converged	2	15	2	2	86	0	2.472×10^{-12}
10^{-4}	P1 RD	NK	converged	5	7	5	5	15	1	1.816×10^{-11}
		GAIN-FJ	converged	3	19	3	3	43	1	1.221×10^{-13}
10^{-4}	P2 CDR	NK	converged	4	5	4	4	11	0	1.241×10^{-14}
		GAIN-FJ	converged	1	11	1	1	10	0	2.690×10^{-11}
10^{-4}	P3 Bratu	NK	converged	5	6	5	5	35	0	4.747×10^{-13}
		GAIN-FJ	converged	2	15	2	2	45	0	6.747×10^{-13}

For these particular runs the smaller drop tolerance 10^{-4} reduces Krylov work for both methods on all three PDEs. For example, P3 decreases from 70 to 35 iterations for NK and from 86 to 45 for GAIN-FJ. This is a result for the present discretizations and ILUTP implementation, not a general monotonicity theorem for drop tolerance.

6.11. **Bratu mechanism stress test.** The Bratu ablation uses P3 with $q = 64$, $u_0 \equiv 1.5$, and ILUTP drop tolerance 5×10^{-4}. A2 uses adaptive stages and line search with fixed forcing $\eta_{\text{fixed}} = 10^{-6}$ and no internal refresh; A3 adds adaptive refresh; full GAIN-FJ additionally uses stage-dependent forcing. Table 12 reports the resulting mechanism comparison.

All three variants converge. A2 and A3 have identical residual-call, setup, and Krylov totals, even though A3 requires only two accepted outer updates; the refresh mechanism changes where the setups occur rather than the aggregate counts in this run. Full GAIN-FJ reaches a much smaller final residual

TABLE 12. Bratu mechanism stress test at $q = 64$, $u_0 \equiv 1.5$.

Variant	status	Out	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
A2: stages + line search	converged	3	22	3	3	232	0	5.100×10^{-12}
A3: A2 + refresh	converged	2	22	3	3	232	0	5.100×10^{-12}
full GAIN-FJ	converged	4	26	5	5	283	0	5.337×10^{-15}

but uses more outer updates, setups, and Krylov iterations. Thus this stress test does not show a work advantage for the complete adaptive mechanism.

6.12. **Sensitivity of the difficult Bratu test to parameters.** For this one-factor-at-a-time experiment, we take P3 at $q = 128$ with $u_0 \equiv 1.5$ and use the default ILUTP drop tolerance 10^{-4}. Convergence is obtained for every parameter setting listed below.

TABLE 13. One-at-a-time sensitivity on the Bratu problem at $q = 128$, $u_0 \equiv 1.5$.

Setting	status	Out	$\#F$	$\#J$	$\#P$	Krylov	backtracks	final R
$m_{\max} = 2$	converged	3	16	4	4	180	0	1.913×10^{-11}
$m_{\max} = 3$	converged	3	16	4	4	180	0	1.913×10^{-11}
$m_{\max} = 4$	converged	3	16	4	4	180	0	1.913×10^{-11}
$m_{\max} = 5$	converged	3	16	4	4	180	0	1.913×10^{-11}
$m_{\max} = 6$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\rho_{\mathrm{rej}} = 1.5$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\rho_{\mathrm{rej}} = 2$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\rho_{\mathrm{rej}} = 5$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\chi_{\mathrm{ref}} = 0.25$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\chi_{\mathrm{ref}} = 0.5$	converged	3	16	4	4	180	0	1.913×10^{-11}
$\chi_{\mathrm{ref}} = 1$	converged	4	22	4	4	258	0	1.765×10^{-13}
$c_p = c_c = 0.05$	converged	3	16	3	3	91	0	1.544×10^{-14}
$c_p = c_c = 0.10$	converged	3	16	4	4	180	0	1.913×10^{-11}
$c_p = c_c = 0.20$	converged	3	22	3	3	253	0	5.707×10^{-15}

For this initial point, changing $m_{\max}$ between 2 and 6 or ρ_{rej} between 1.5 and 5 does not alter the recorded trajectory. In contrast, setting $\chi_{\mathrm{ref}} = 1$ increases both the number of accepted outer steps and

the Krylov work. The largest change is associated with the forcing scale: $c_p = c_c = 0.05$ requires 91 Krylov iterations, whereas the counts are 180 for 0.10 and 253 for 0.20. Thus the work does not vary monotonically with the nominal looseness of the inner tolerances, because the forcing choice also alters the nonlinear trajectory, the refresh decisions, and how many linear systems are solved.

6.13. **Checks supporting reproducibility.** The implementation maintains counters for each residual call, Jacobian construction, preconditioner construction, Krylov iteration, accepted outer update, and line-search backtrack. A candidate residual that has already been computed is taken from cache instead of being evaluated again. Each frozen epoch constructs its preconditioner once. Because preconditioning is applied explicitly on the right, the quantity monitored by GMRES is the original-system residual required by (3.3). The reproduction code writes the CSV files underlying Tables 3–13 together with the MATLAB figures for Figures 1–2. The present data package includes no matched set of repeated wall-clock timings; consequently, we report no CPU-speed comparison.

7. Discussion

The results lead to a more specific interpretation of Gain-FJ. This aim differs from the efficiency-index focus of the closest multistep schemes summarized in Table 1. The goal is not simply to obtain a higher formal order, but to determine whether a reusable high-order block can still work effectively with inexact Krylov solves within a globally safeguarded iteration. Under the stated local and global assumptions, the analysis shows that these features are compatible.

First, the local high-order mechanism is strongly supported. The arbitrary-precision calculations recover the predicted lower-bound sequence $5, 8, 11, 14$, and stage-dependent forcing preserves those observed orders while reducing work by 30.0–45.4% in the high-precision work-to-tolerance experiment.

Second, globalization is essential for the difficult cyclic start. The nonglobalized high-order variants do not converge, whereas A2 converges once the Armijo line search is enabled. Stage-dependent forcing then reduces the fixed-forcing Krylov count from 99 to 32 (67.7%). This does not imply superiority to Newton–Krylov: the matched NK baseline requires only 9 Krylov iterations on the same structured problem.

Third, in the PDE campaign with explicit right preconditioning and an original-system residual stopping test, both methods converge on every P1, P2, and P3 case reported for $q = 64, 128, 192, 320$. Gain-FJ consistently reduces accepted outer steps and Jacobian/preconditioner constructions, which is exactly the regime targeted by frozen-Jacobian reuse. However, it usually increases Krylov work and residual evaluations. The main computational tradeoff is therefore *setup reuse versus repeated linear/right-hand-side work*, not robustness versus outright failure.

The break-even conditions in Table 10 make this tradeoff quantitative. At the largest P2 case, three Jacobian/preconditioner constructions are saved at the cost of six additional residual evaluations and

only two additional Krylov iterations. By contrast, the largest P1 case saves two setups of each type but incurs twelve additional residual evaluations and thirty-eight additional Krylov iterations. The data therefore identify where a timing advantage is plausible, but they do not establish one without measured T_F, T_J, T_P, and T_K. Runtime values from the large-scale globalized high-order experiments in [26] are not directly comparable with ours because the two studies differ in algorithm, test problems, linear-algebra setting, and hardware.

The refresh-policy results show the same tradeoff. On the eight inexpensive cyclic starts, rebuilding after every accepted block produces the smallest Krylov count. The adaptive strategy uses fewer setups, but its total Krylov work is almost identical to that of the no-refresh case. A cost-aware refresh rule should therefore consider setup time, factor quality, and accumulated Krylov work together instead of relying only on the defect threshold.

Finally, the Bratu studies show convergence for all reported ablation and parameter-sensitivity cases. They also show that more adaptive machinery need not reduce work: the full method uses more Krylov iterations than A2/A3 in the $q = 64$ stress test, and the forcing scale changes work substantially at $q = 128$.

8. Limitations

Several limitations remain.

(1) The local theory assumes a simple root and C^3 smoothness. Multiple roots, singular Jacobians, and nonsmooth residuals require different analysis.

(2) The global theorem assumes uniform nonsingularity of $F'(X)$ on the relevant merit-function level set, which is strong and excludes nearly singular branches.

(3) The sparse study uses only MATLAB ILUTP and restarted GMRES. It does not assess algebraic multigrid, domain decomposition, block, or physics-based preconditioners. Its largest PDE system has 102400 unknowns; recent globalized high-order work has demonstrated substantially larger discretizations under a different sparse-factorization regime [26], so the present study should be interpreted as an inexact-Krylov/setup-reuse study rather than a scale record.

(4) The deterministic PDE tables establish convergence and work counts, but no matched repeated wall-clock timing campaign is available. Equations (5.4)–(5.5) therefore provide only a conditional cost interpretation, not a measured runtime-speedup claim.

(5) Gain-FJ is not uniformly superior in Krylov work. In most P1 and P3 rows it uses more Krylov iterations than the matched NK baseline, despite fewer Jacobian/preconditioner setups.

(6) The current adaptive refresh heuristic is not work-optimal on the eight-start cyclic test; refresh-after-every-accepted-block uses substantially fewer Krylov iterations.

(7) The Bratu parameter study is one-at-a-time and problem-specific. The apparent insensitivity to $m_{\max}$ and ρ_{rej} should not be generalized beyond the reported start and discretization.

(8) A fixed positive forcing floor such as $\eta_{\min} = 10^{-12}$ is a finite-precision device and is incompatible with the formal asymptotic scales $\eta_{k,0} = \mathcal{O}(\|e_k\|)$ and $\eta_{k,j,\ell} = \mathcal{O}(\|e_k\|^3)$ as $e_k \to 0$. The local-order theorem therefore concerns the floor-free theoretical iteration.

(9) The right-preconditioned implementation assumes that applying the preconditioner is a fixed linear operation during one GMRES solve. Variable preconditioning would require a flexible Krylov implementation.

9. Conclusions

We developed a globalized inexact Newton–Krylov framework that uses high-order frozen-Jacobian corrections as conditional accelerators. The local analysis proves a three-power gain per correction block and therefore the fixed-stage Q-order lower bound

$$p_m = 3m - 1,$$

giving the lower bounds $5, 8, 11, 14$ for two through five stage points. The inexact analysis identifies sufficient first-order predictor forcing and third-order correction forcing, while the retained inexact Newton direction supplies a descent fallback for global convergence.

The arbitrary-precision MATLAB experiments attain the predicted orders to the displayed accuracy. Adaptive forcing reduces the high-precision Krylov work by 30.0–45.4%. On the difficult $n = 20000$ cyclic start, Armijo globalization restores convergence of the fixed-forcing high-order variant, and stage-dependent forcing reduces its Krylov count from 99 to 32.

The sparse results show convergence of both NK and Gain-FJ on every reported reaction–diffusion, convection–diffusion–reaction, and Bratu case through 102400 unknowns. Gain-FJ consistently reduces accepted outer steps and Jacobian/preconditioner constructions, but it is not uniformly cheaper in Krylov iterations or residual evaluations. The break-even analysis therefore states the practical conclusion in the appropriate conditional form: frozen-Jacobian high-order corrections can be advantageous when saved Jacobian/preconditioner setup cost exceeds the additional residual and Krylov work. Whether that condition holds is problem-, preconditioner-, and hardware-dependent.

Future work should therefore focus on cost-aware refresh rules, recycled or flexible Krylov methods across repeated right-hand sides, and multilevel or physics-based preconditioning. A complete performance study should also include repeated wall-clock timings, memory use, and preconditioner setup/application costs on a documented hardware/software platform.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] C.T. Kelley, Solving Nonlinear Equations with Newton's Method, Society for Industrial and Applied Mathematics, 2003. https://doi.org/10.1137/1.9780898718898.

[2] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton Methods, SIAM J. Numer. Anal. 19 (1982), 400–408. https://doi.org/10.1137/0719025.

[3] S.C. Eisenstat, H.F. Walker, Globally Convergent Inexact Newton Methods, SIAM J. Optim. 4 (1994), 393–422. https://doi.org/10.1137/0804022.

[4] S.C. Eisenstat, H.F. Walker, Choosing the Forcing Terms in an Inexact Newton Method, SIAM J. Sci. Comput. 17 (1996), 16–32. https://doi.org/10.1137/0917003.

[5] P.N. Brown, Y. Saad, Hybrid Krylov Methods for Nonlinear Systems of Equations, SIAM J. Sci. Stat. Comput. 11 (1990), 450–481. https://doi.org/10.1137/0911026.

[6] Y. Saad, M.H. Schultz, GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems, SIAM J. Sci. Stat. Comput. 7 (1986), 856–869. https://doi.org/10.1137/0907058.

[7] Y. Saad, A Flexible Inner-Outer Preconditioned GMRES Algorithm, SIAM J. Sci. Comput. 14 (1993), 461–469. https://doi.org/10.1137/0914028.

[8] M. Pernice, H.F. Walker, NITSOL: a Newton Iterative Solver for Nonlinear Systems, SIAM J. Sci. Comput. 19 (1998), 302–318. https://doi.org/10.1137/S1064827596303843.

[9] D.A. Knoll, D.E. Keyes, Jacobian-Free Newton–Krylov Methods: a Survey of Approaches and Applications, J. Comput. Phys. 193 (2004), 357–397. https://doi.org/10.1016/j.jcp.2003.08.010.

[10] R.P. Pawlowski, J.N. Shadid, J.P. Simonis, H.F. Walker, Globalization Techniques for Newton–Krylov Methods and Applications to the Fully Coupled Solution of the Navier–Stokes Equations, SIAM Rev. 48 (2006), 700–721. https://doi.org/10.1137/S0036144504443511.

[11] X.-C. Cai, D.E. Keyes, Nonlinearly Preconditioned Inexact Newton Algorithms, SIAM J. Sci. Comput. 24 (2002), 183–200. https://doi.org/10.1137/S106482750037620X.

[12] F.-N. Hwang, X.-C. Cai, A Parallel Nonlinear Additive Schwarz Preconditioned Inexact Newton Algorithm for Incompressible Navier–Stokes Equations, J. Comput. Phys. 204 (2005), 666–691. https://doi.org/10.1016/j.jcp.2004.10.025.

[13] S. Bellavia, D. Bertaccini, B. Morini, Nonsymmetric Preconditioner Updates in Newton–Krylov Methods for Nonlinear Systems, SIAM J. Sci. Comput. 33 (2011), 2595–2619. https://doi.org/10.1137/100789786.

[14] V. Dolean, M.J. Gander, W. Kheriji, F. Kwok, R. Masson, Nonlinear Preconditioning: How to Use a Nonlinear Schwarz Method to Precondition Newton's Method, SIAM J. Sci. Comput. 38 (2016), A3357–A3380. https://doi.org/10.1137/15M102887X.

[15] L. Liu, F.-N. Hwang, L. Luo, X.-C. Cai, D.E. Keyes, A Nonlinear Elimination Preconditioned Inexact Newton Algorithm, SIAM J. Sci. Comput. 44 (2022), A1579–A1605. https://doi.org/10.1137/21M1416138.

[16] L. Luo, X.-C. Cai, PIN$^{\mathcal{L}}$: Preconditioned Inexact Newton with Learning Capability for Nonlinear System of Equations, SIAM J. Sci. Comput. 45 (2023), A849–A871. https://doi.org/10.1137/22M1507942.

[17] Y. Gong, L. Luo, X.-C. Cai, Stagnation-Shortening Inexact Newton Method Based on Unsupervised Learning for Highly Nonlinear Algebraic Problems and Applications, SIAM J. Sci. Comput. (2026), S513–S534. https://doi.org/10.1137/24M1672869.

[18] M.Z. Ullah, S. Serra-Capizzano, F. Ahmad, E.S. Al-Aidarous, Higher Order Multi-Step Iterative Method for Computing the Numerical Solution of Systems of Nonlinear Equations: Application to Nonlinear PDEs and ODEs, Appl. Math. Comput. 269 (2015), 972–987. https://doi.org/10.1016/j.amc.2015.07.096.

[19] F. Ahmad, E. Tohidi, J.A. Carrasco, A Parameterized Multi-Step Newton Method for Solving Systems of Nonlinear Equations, Numer. Algorithms 71 (2016), 631–653. https://doi.org/10.1007/s11075-015-0013-7.

[20] A. Cordero, C. Jordán, E. Sanabria-Codesal, J.R. Torregrosa, Highly Efficient Iterative Algorithms for Solving Nonlinear Systems with Arbitrary Order of Convergence $p+3$, $p \geq 5$, J. Comput. Appl. Math. 330 (2018), 748–758. https://doi.org/10.1016/j.cam.2017.02.032.

[21] R.H. Al-Obaidi, M.T. Darvishi, Constructing a Class of Frozen Jacobian Multi-Step Iterative Solvers for Systems of Nonlinear Equations, Mathematics 10 (2022), 2952. https://doi.org/10.3390/math10162952.

[22] R. Erfanifar, M. Hajarian, A New Multi-Step Method for Solving Nonlinear Systems with High Efficiency Indices, Numer. Algorithms 97 (2024), 959–984. https://doi.org/10.1007/s11075-023-01735-2.

[23] R. Erfanifar, M. Hajarian, An Efficient High-Order Iterative Method to Solve Systems of Nonlinear Equations with Applications to Differential Equations and Image Processing, J. Franklin Inst. 363 (2026), 108332. https://doi.org/10.1016/j.jfranklin.2025.108332.

[24] A. Cordero, R.V. Rojas-Hiciano, J.R. Torregrosa, M.P. Vassileva, High-Level Convergence Order Accelerators of Iterative Methods for Nonlinear Problems, Appl. Numer. Math. 217 (2025), 390–411. https://doi.org/10.1016/j.apnum.2025.07.003.

[25] A. Cordero, R.V. Rojas-Hiciano, J.R. Torregrosa, M.P. Vassileva, Maximally Efficient Damped Composed Newton-Type Methods to Solve Nonlinear Systems of Equations, Appl. Math. Comput. 492 (2025), 129231. https://doi.org/10.1016/j.amc.2024.129231.

[26] A. Cordero, R.V. Rojas-Hiciano, J.R. Torregrosa, Efficient Solution of Convection-Dominated and Diffusion-Reaction Problems via a Novel Eighth-Order Iterative Method, Appl. Numer. Math. 229 (2026), 209–236. https://doi.org/10.1016/j.apnum.2026.06.004.

[27] L. Armijo, Minimization of Functions Having Lipschitz Continuous First Partial Derivatives, Pac. J. Math. 16 (1966), 1–3. https://doi.org/10.2140/pjm.1966.16.1.