Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 226 (2007) 2317–2340
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Globalization strategies for Newton–Krylov methods for stabilized FEM discretization of Navier–Stokes equations

Stefania Bellavia [a], Stefano Berrone [b,*]

[a] *Dipartimento di Energetica, Università di Firenze, Via Lombroso 6/17, 50134 Firenze, Italy*
[b] *Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

Received 13 December 2006; received in revised form 11 July 2007; accepted 13 July 2007
Available online 7 August 2007

Abstract

In this work we study the numerical solution of nonlinear systems arising from stabilized FEM discretizations of Navier–Stokes equations. This is a very challenging problem and often inexact Newton solvers present severe difficulties to converge. Then, they must be wrapped into a globalization strategy. We consider the classical backtracking procedure, a subspace trust-region strategy and an hybrid approach. This latter strategy is proposed with the aim of improve the robustness of backtracking and it is obtained combining the backtracking procedure and the elliptical subspace trust-region strategy. Under standard assumptions, we prove global and fast convergence of the inexact Newton methods embedded in this new strategy as well as in the subspace trust-region procedure. Computational results on classical CFD benchmarks are performed. Comparisons among the classical backtracking strategy, the elliptical subspace trust-region approach and the hybrid procedure are presented. Our numerical experiments show the effectiveness of the proposed hybrid technique.

AMS subject classifications: 65N30; 65N50; 65N55

Keywords: Inexact Newton methods; Globalization strategy; Numerical Navier–Stokes equation; Finite element methods

1. Introduction

We consider the numerical solution of the Navier–Stokes equations modelling incompressible fluid flows. We use a discretization by stabilized Finite Elements [17] that allows circumventing the Babuška–Brezzi stability condition. The discretization of these equations usually gives rise to a large-scale system of nonlinear equations

$$F(x) = 0 \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear differentiable map.

* Corresponding author. Tel.: +39 011 564 75 03; fax: +39 011 564 75 99.
E-mail addresses: stefania.bellavia@unifi.it (S. Bellavia), sberrone@calvino.polito.it (S. Berrone).
URLs: http://ciro.de.unifi.it/~bellavia (S. Bellavia), http://calvino.polito.it/~sberrone (S. Berrone).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2007.07.021

Due to their large dimension these problems are usually solved by Newton–Krylov methods, that are Newton-type methods in which a Krylov method is employed to solve approximately the arising linear systems and compute a so called inexact Newton step. To enhance the robustness these methods are augmented with a suitable globalization strategy.

Well known convergence properties of these methods motivated the works to create robust and locally fast algorithms and to develop reliable and efficient software [1,3,7,11,15,16,26,27]. Recently, there has been much research in investigating the convergence behaviour of globalized Newton–Krylov methods applied to the nonlinear systems arising in the solution of fully coupled large scale CFD problems [14,25,29,30]. In fact, in these papers it has been shown that the numerical solution of these nonlinear systems is very challenging and a globalization strategy is needed in order to successfully solve this class of problems. In the review [25] several representative globalizations are discussed and compared. Moreover in [30] failures of the backtracking strategy along the inexact Newton step have been analyzed. The backtracking technique shortens the inexact Newton step as necessary to get a sufficient reduction of the nonlinear residual Euclidean norm $\|F\|$. In [30] it has been shown that there are some situations, especially if the Jacobian of F is ill-conditioned, where the inexact Newton step becomes increasingly orthogonal to the gradient of the nonlinear residual norm and the backtracking strategy fails. The failures are due to the fact that too many backtracks are necessary to get the desired reduction of $\|F\|$ that a practical backtracking algorithm detects a failure. In this situation the inexact Newton direction must be dropped.

Here, the classical backtracking procedure, an elliptical subspace trust-region strategy and an hybrid approach are considered.

In the subspace trust-region strategy the trial step is computed minimizing the classical quadratic function within the trust-region in the subspace spanned by the gradient of the function $\|F\|^2$ and the inexact Newton step. Once the subspace has been built, the work to compute an approximate solution of the trust-region subproblem is trivial since in the subspace the problem is only two dimensional. We remark that this approach has been considered also in [4] and [5] in the context of affine scaling trust-region methods for bound constrained nonlinear system and bound constrained minimization problems, respectively. On the other hand, to our knowledge, this approach has never been employed in conjunction with inexact Newton–Krylov methods for unconstrained nonlinear systems. We also underline that an alternative approach to trust-region globalization for Inexact Newton methods is treated in the very recent paper [26], where a general inexact Newton dogleg method is developed. In [26] the classical dogleg path, requiring the computation of the Newton step, is replaced by an approximate dogleg curve that is defined using the inexact Newton step together with approximate steepest-descent directions.

The hybrid globalization strategy we propose here is obtained by a suitable combination of the backtracking along the inexact Newton step and the two dimensional subspace trust-region previously mentioned. Specifically, first we use the direction of the inexact Newton step. If it does not work well, i.e. relatively few steps do not suffice to decrease the value of the nonlinear residual norm, we revert to the subspace trust-region strategy. Our motivation for dealing with hybrid approaches is that in our opinion the backtracking method should be the first choice for its simplicity as well its effectiveness [27,29,7], however, in the applications we are involved with, it is crucial to have at disposal an alternative strategy when the inexact Newton step is a poor direction and the backtracking fails. In fact, trust-region methods have the potential advantage of generating directions that may be stronger descent directions than the inexact Newton step. Then, the switch to a trust-region strategy may improve the practical applicability of backtracking Newton–Krylov methods and may prevent failures as described in [30]. On the other hand trust-region techniques require the computation of the gradient of $\|F\|^2$ and this calls for the product of the transpose of the Jacobian of F with a vector, while in backtracking Newton–Krylov methods only the action of the Jacobian times a vector is needed. Then, trust-region Newton–Krylov method cannot be straightforwardly implemented in a matrix-free manner, i.e. avoiding the explicit computation of the Jacobian by approximating the products of the Jacobian and a vector by finite differences. We underline that in [7] a matrix-free trust-region strategy has been proposed. Moreover, matrix-free hybrid globalization strategies have been considered in [1] and [3] and they revealed to be an effective approach that enhances the classical backtracking procedure. However, all these globalization techniques are strictly based on the iterative solver GMRES [28] and cannot be implemented if

a different iterative linear solver is chosen. Anyway, we remark that in our context the Jacobian is available and can be easily evaluated.

The theoretical analysis we performed shows that inexact Newton methods, augmented with the subspace trust-region strategy or the hybrid technique, are globally and fast convergent.

We carried out an extensive numerical experimentation on classical CFD benchmarks. We analyzed some implementational issues that are crucial for the effectiveness of the globalization strategies and we compared the performance of the three globalization procedures in conjunction with the Newton-GMRES method. The obtained numerical results show that the combination of the two strategies outperforms both the classical backtracking approach and the subspace trust-region strategy.

2. Incompressible Navier–Stokes model

In this section, the incompressible steady-state Navier–Stokes model and its finite elements discretization, are presented briefly.

2.1. The continuous problem

We consider the following velocity–pressure formulation of the steady-state, incompressible Navier–Stokes equations:

$$-\frac{1}{Re}\Delta u + (u\cdot\nabla)u + \nabla p = f \quad \text{in } \Omega, \tag{2.1}$$

$$\nabla\cdot u = 0 \quad \text{in } \Omega, \tag{2.2}$$

$$u = 0 \quad \text{on } \Gamma_D, \tag{2.3}$$

$$\frac{1}{Re}\frac{\partial u}{\partial \hat{n}} - p\hat{n} = g_N \quad \text{on } \Gamma_N, \tag{2.4}$$

where Re is the Reynolds number; Ω is a bounded domain in $\mathbb{R}^2$ with a regular boundary $\partial\Omega$ that belongs to the class $C^{0,1}$ ($\partial\Omega$ can be locally described by Lipschitz continuous functions [19]); the boundary $\partial\Omega$ is split into two subsets Γ_D and Γ_N, where Γ_D is closed and the following conditions holds true: $\partial\Omega = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$ and $|\Gamma_D| \neq 0$; $\hat{n}$ is the usual unit outward normal vector to $\partial\Omega$; $f \in [2(\Omega)]^2$; $g_N \in [H^{\frac{1}{2}}(\Gamma_N)]^2$.

Let us first derive a weak formulation of problem (2.1)–(2.4). The functional spaces we deal with are the usual Sobolev space $H^1_{0,D}(\Omega) = \{v \in H^1(\Omega) | v_{|\Gamma_D} = 0\}$ and Lebesgue space $L^2_0(\Omega) = \{q \in L^2(\Omega) | \int_\Omega q \, d\Omega = 0\}$. Moreover we set $V = [H^1_{0,D}(\Omega)]^2$ and $Q = L^2_0(\Omega)$ if $|\Gamma_N| = 0$ or $Q = L^2(\Omega)$ if $|\Gamma_N| > 0$. A weak formulation of the problem can be written as: *Find* $[u,p] \in V \times Q$ *such that* $\forall\ [v,q] \in V \times \mathrm{Q}$ *one has*

$$\frac{1}{Re}(\nabla u, \nabla v) + ((u\cdot\nabla)u, v) - (p, \nabla\cdot v) = (f, v) + (g_N, v)_{\Gamma_N}, \tag{2.5}$$

$$(q, \nabla\cdot u) = 0, \tag{2.6}$$

where $(.,.)$ denotes the usual inner product in $L^2(\Omega)$ or in $[L^2(\Omega)]^2$ and $(.,.)_{\Gamma_N}$ denotes the inner product in $[L^2(\Gamma_N)]^2$. Existence and uniqueness of the solution for all positive Re follows from the usual *coercivity* inequality and *inf–sup* condition (see e.g. [19]).

2.2. The discrete problem

In order to discretize problem (2.1)–(2.4), we assume Ω to be a polygonal domain and we introduce a regular family of partitions $\{\mathcal{T}_h\}_h$ of $\overline{\Omega}$ into triangles which satisfy the usual conformity and minimal-angle conditions [9]. We denote by h_T the diameter of the element $T \in \mathcal{T}_h$. The parameter h of the family $\{\mathcal{T}_h\}_h$ is defined as $h = \max_{T\in\mathcal{T}_h} h_T$. Let $V_h \subset \mathrm{V}$ and $Q_h \subset \mathrm{Q}$ be two conforming finite element spaces based on the partition $\mathcal{T}_h$. If we consider the pure Galerkin approximation of the continuous problem (2.5) and (2.6), we have to satisfy the discrete version of the *inf–sup* condition [6,19].

In what follows, we are going to use continuous finite element for the velocity: $V_h = \{v_h \in V \cap [\mathcal{C}^0(\overline{\Omega})]^2 | v_{h|_T} \in [P_k(T)]^2, \forall T \in \mathcal{T}_h\}$ and the pressure: $Q_h = \{q_h \in Q \cap \mathcal{C}^0(\overline{\Omega}) | q_{h|_T} \in P_l(T), \forall\ T \in \mathcal{T}_h\}$. Here $P_i(T)$ is the space of polynomials of degree $i \geqslant 1$ on the element $T \in \mathcal{T}_h$. With an arbitrary choice of k and l these spaces may not satisfy the discrete *inf–sup* condition [6]. However, this difficulty may be avoided by resorting to a consistent modified approximation of the problem known as the *Streamline Upwind/Petrov Galerkin* (*SUPG*) method [17,21]: *Find* $[u_h, p_h] \in V_h \times Q_h$ *such that* $\forall [v_h, q_h] \in V_h \times Q_h$ *let be*:

$$\frac{1}{Re}(\nabla u_h, \nabla v_h) + ((u_h \cdot \nabla)u_h, v_h) - (p_h, \nabla \cdot v_h) + \sum_{T \in \mathcal{T}_h} \tau_T \left(-\frac{1}{Re}\Delta u_h + (u_h \cdot \nabla)u_h + \nabla p_h, (u_h \cdot \nabla)_h v_h \right)_T + \sum_{T \in \mathcal{T}_h} \delta_T (\nabla \cdot u_h, \nabla \cdot v_h)_T = (f, v_h) + (g_N, v_h)_{\Gamma_N} + \sum_{T \in \mathcal{T}_h} \tau_T (f, (u_h \cdot \nabla)v_h)_T, \tag{2.7}$$

$$(q_h, \nabla \cdot u_h) + \sum_{T \in \mathcal{T}_h} \tau_T \left(-\frac{1}{Re}\Delta u_h + (u_h \cdot \nabla)u_h + \nabla p_h, \nabla q_h \right)_T = \sum_{T \in \mathcal{T}_h} \tau_T (f, \nabla q_h)_T. \tag{2.8}$$

Here τ_T and δ_T depend on the local conditions of the flow in each element expressed by $Re_T = m_k \frac{\|u_h\|_{\infty,T} h_T}{4\frac{1}{Re}}$ and $m_k = \min\left\{\frac{1}{3}, \frac{2}{C_*}\right\}$, C_* being the constant of the inverse inequality [20]: $h_T^2 \|\Delta v_h\|_{0,T}^2 \leqslant C_* \|\nabla v_h\|_{0,T}^2, \forall v_h \in V_h$. For linear elements, obviously, $m_k = \frac{1}{3}$. Practically, following [17] we set $\tau_T = m_k \frac{h_T^2}{8} Re$, $\delta_T = \lambda m_k \frac{\|u_h\|_{\infty,T} h_T^2 Re}{4}$ if $0 \leqslant Re_T < 1$ and $\tau_T = \frac{h_T}{2\|u_h\|_{\infty,T}}$, $\delta_T = \lambda \|u_h\|_{\infty,T} h_T$ if $Re_T \geqslant 1$. Possible values for λ are 1 and 0. By using the *SUPG* method, not only we circumvent the *inf–sup* condition [6,21], but also stabilize the advective operator preventing the oscillations in the velocity field that appear for high Reynolds numbers [17]. The problem (2.7 and 2.8) gives rise to a system of nonlinear equations

$$F(x) = 0, \tag{2.9}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \mathbb{R}^n$ is the vector of unknown velocity components and pressure.

3. The numerical methods

The core of our approach is an inexact Newton method, that is an iterative procedure that constructs a sequence of iterates $\{x_k\}$ such that, at each iteration k, the correction $\bar{s}_k = x_{k+1} - x_k$ satisfies

$$F'(x_k)\bar{s}_k = -F(x_k) + r_k, \quad \|r_k\| \leqslant \bar{\eta}_k \|F(x_k)\|, \tag{3.1}$$

where F' is the system Jacobian, $\bar{\eta}_k$ is a suitable scalar in $[0, 1)$ called forcing term, r_k is commonly referred to as the residual vector and $\|\cdot\|$ always denotes the Euclidean norm. The vector $\bar{s}_k$ is commonly called inexact Newton step. The choice of the forcing term $\bar{\eta}_k$ has a strong relevance on the performance of inexact Newton methods as it is shown in [11]. We recall that, under standard conditions, if $\{\bar{\eta}_k\} \to 0$ then $\{x_k\}$ exhibits local q-superlinear convergence to a solution x^*, while if $\{\bar{\eta}_k\} = O(F(x_k))$ the convergence rate is quadratic. In [16] two choices of $\bar{\eta}_k$'s yielding up to quadratic rate of convergence have been proposed.

Here, we embed the inexact Newton method into three different globalization frameworks: the classical backtracking strategy, an elliptical subspace trust-region procedure and an hybrid approach. This latter technique was developed in order to enhance robustness of the classical backtracking approach. This goal is achieved applying the backtracking strategy whenever it works well and reverting to the subspace trust-region globalization when the inexact Newton step is a poor descent direction.

3.1. Backtracking technique

A widely used inexact Newton method augmented with a backtracking strategy is given by the INB (inexact Newton Backtracking) method [15,27]. The kth iteration of this method can be sketched as follows:

INB Method

Let x_k, $\eta_{\max} \in (0, 1)$, $t \in (0, 1)$, $0 < \theta_m < \theta_{\max} < 1$ be given.

1. Choose $\bar{\eta}_k \in [0, \eta_{\max}]$ and compute $\bar{s}_k$ such that
   $\|F(x_k) + F'(x_k)\bar{s}_k\| \leqslant \bar{\eta}_k \|F(x_k)\|$.
2. Perform the backtracking strategy:
   2.1. Set $s_k = \bar{s}_k$, $\eta_k = \bar{\eta}_k$.
   2.2. While $\|F(x_k + s_k)\| > (1 - t(1 - \eta_k))\|F(x_k)\|$ do:
   Choose $\theta \in [\theta_m, \theta_{\max}]$.
   Update $s_k = \theta s_k$ and $\eta_k = 1 - \theta(1 - \eta_k)$.
3. Set $x_{k+1} = x_k + s_k$.

The backtracking strategy performed at Step 2 provides a sufficient decrease of the nonlinear residual norm $\|F(x)\|$. In particular, moving along the direction of the initial inexact step $\bar{s}_k$, successively shorter steps $s_k = s_k(\eta_k)$ of the form $s_k = (1 - \eta_k)\bar{s}_k/(1 - \bar{\eta}_k)$ are selected. In [15] the following global convergence theorem is proved.

Theorem 3.1. *Assume that F is continuously differentiable. If $\{x_k\}$ generated by the INB method has a limit point x^* such that $F'(x^*)$ is nonsingular, then $F(x^*) = 0$ and the sequence $\{x_k\}$ converges to x^*. Moreover, eventually the inexact Newton step $\bar{s}_k$ is taken.*

Then, the convergence rate is governed by the choice of η_k's and up to quadratic convergence can be obtained with the choices of η_k's described in [16].

3.2. Elliptical subspace trust-region technique

In the classical trust-region approach, at iteration k, the following quadratic model

$$m_k(s) = \frac{1}{2}\|F'(x_k)s + F(x_k)\|^2 = \frac{1}{2}\|F(x_k)\|^2 + F(x_k)^{\mathrm{T}}F'(x_k)s + \frac{1}{2}s^{\mathrm{T}}F'(x_k)^{\mathrm{T}}F'(x_k)s,$$

and a specified trust-region around the current iterate x_k are defined. Letting $f(x)$ be the merit function:

$$f(x) = \frac{1}{2}\|F(x)\|^2, \tag{3.2}$$

the quadratic model $m_k(s)$ is trusted to be an adequate representation of $f(x_k + s)$ within the trust-region. Usually, spherical trust-regions are adopted. On the other hand this choice is not appropriate in case of poorly scaled functions. In order to give more flexibility to our approach we consider elliptical trust-regions. Then, the search direction s_k is the vector solution of the subproblem

$$\min_s\{m_k(s) : \|D_k s\| \leqslant \Delta_k\}, \tag{3.3}$$

where $D_k \in \mathbb{R}^{n\times n}$ is a positive definite diagonal matrix and $\Delta_k > 0$ is the trust-region size. We would like to mention that (3.3) is a specific case of the more general approach where, given a symmetric positive definite matrix M, the trust-region is defined using the M-norm:

$$\|s\|_M = \sqrt{s^{\mathrm{T}}Ms},$$

and the underlying trust-region subproblem is given by:

$$\min_s\{m_k(s) : \|s\|_M \leqslant \Delta_k\}. \tag{3.4}$$

In fact, (3.4) reduces to (3.3), whenever $M = D_k^2$. Note that, as the Newton step $s_k^N = -F'(x_k)^{-1}F(x_k)$ is the global minimum of $m_k(s)$, if $\|s_k^N\|_M \leqslant \Delta_k$, then s_k^N is the solution of (3.4). Moreover, from Theorem 7.4.1 of [10], it follows that (3.4) has a unique solution, whenever $F'(x_k)$ is invertible.

The dogleg method is a very common way to compute an approximate solution of the trust-region subproblem. We refer to [24] for a description of the dogleg procedure. Here, we just recall that it is based on the so called dogleg path, whose definition requires the computation of the Newton step. A straightforward extension of the dogleg method to the inexact context can be obtained by substituting the inexact step $\bar{s}_k$ for s_k^N in the definition of the dogleg path. Unfortunately, as it is nicely shown in [26], this straightforward extension is not reliable.

Therefore, we shared the approach given in [8] and replace the full elliptical trust-region subproblem (3.3) by

$$\min_s\{m_k(s) : \|D_k s\| \leqslant \Delta_k, s \in \mathcal{S}_k\}, \tag{3.5}$$

where $\mathcal{S}_k$ is the two-dimensional subspace given by:

$$\mathcal{S}_k = \mathrm{span}\{D_k^{-2}\nabla f(x_k), \bar{s}_k\}. \tag{3.6}$$

We underline that the inclusion of the scaled gradient vector $D_k^{-2}\nabla f(x_k)$ in $\mathcal{S}_k$ will guarantee convergence to a stationary point of the function $f(x)$. On the other hand, the inclusion of the inexact Newton step will provide fast local convergence of the inexact Newton method embedded into this procedure.

Due to the low dimension of $\mathcal{S}_k$, an approximate solution s_k of problem (3.3) can be easily computed as we will show in the next section. Then, the trial step s_k is accepted if the following standard condition is satisfied:

$$\rho_k^f(s_k) = \frac{f(x_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)} \geqslant \beta_1, \tag{3.7}$$

where β_1 is a given constant such that $\beta_1 \in (0,1)$. If (3.7) does not hold s_k is rejected, the trust-region is reduced and a new trial step is computed. In other words, with condition (3.7) a good agreement between the quadratic model and the merit function f is imposed.

Then, we embed an inexact Newton method into the subspace strategy as follows:

ESTR Method (Elliptical subspace trust-region method)

Given x_k, $\eta_{\max} \in (0,1)$, $0 < \beta_1 < \beta_2 < 1$, $\alpha_1 \in (0,1)$, $\Delta_{\min} > 0$, $\Delta_k > \Delta_{\min}$, D_k diagonal $n \times n$ matrix, with positive diagonal entries.

1. Choose $\bar{\eta}_k \in [0, \eta_{\max}]$ and compute $\bar{s}_k$ such that
   $\|F(x_k) + F'(x_k)\bar{s}_k\| \leqslant \bar{\eta}_k\|F(x_k)\|$.
2. Let $\mathcal{S}_k = \mathrm{span}\{D_k^{-2}\nabla f(x_k), \bar{s}_k\}$.
3. Do
   3.1 Find $s_k = \mathrm{argmin}_{\{\|D_k s\| \leqslant \Delta_k, s \in \mathcal{S}_k\}} m_k(s)$.
   3.2 Set $\Delta_k = \alpha_1 \min\{\Delta_k, \|D_k s_k\|\}$.
   While $\rho_k^f(s_k) < \beta_1$
4. Set $x_{k+1} = x_k + s_k$ and choose D_{k+1}.
5. If $\rho_k^f(s_k) \geqslant \beta_2$ then
   set $\Delta_{k+1} = \max(\Delta_{\min}, 2\|D_k s_k\|)$
   else set $\Delta_{k+1} = \Delta_k$.

Our convergence results for the ESTR method are given in Theorem 3.2 below, whose proof is provided in Appendix.

Theorem 3.2. *Let $r > 0$ and $L = \cup_{k=0}^{\infty}\{x \in \mathbb{R}^n | \|x - x_k\| \leqslant r\}$ be a neighborhood of sequence $\{x_k\}$ generated by the ESTR method. Assume that F' is Lipschitz continuous in L and $\|F'(x)\|$ is bounded above on L. Suppose furthermore that $\nabla f(x_k) \neq 0$ for $k \geqslant 0$, the sequences $\{\|D_k\|\}$ and $\{\|D_k^{-1}\|\}$ are bounded above. Then,*

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{3.8}$$

Further, if there exists a limit point x^* such that $F'(x^*)$ is nonsingular then, $F(x^*) = 0$ and the sequence $\{x_k\}$ converges to x^*. Moreover, eventually a step s_k satisfying

$$\|F(x_k) + F'(x_k)s_k\| \leqslant \bar{\eta}_k\|F(x_k)\|, \tag{3.9}$$

is taken.

Note that all the accumulation points of the sequence are stationary points of the function f. Moreover, eventually an inexact Newton step satisfying (3.1) is taken. Then, the convergence rate is governed by the choice of η_k's.

We close this subsection with some comments on the generalization of the ESTR method to the more general setting given by the trust region problem (3.4). Indeed, in this more general case, let $\bar{\mathcal{S}}_k = \mathrm{span}\{M^{-1}\nabla f(x_k), \bar{s}_k\}$ and $\bar{D}_k$ be such that $M = \bar{D}_k^{\mathrm{T}}\bar{D}_k$, for example $\bar{D}_k$ can be the Cholesky factor of M. Then, substituting $\bar{D}_k$ for D_k and $\bar{\mathcal{S}}_k$ for $\mathcal{S}_k$ in the ESTR method, we obtain a subspace trust-region method for (3.4) and the theoretical results stated in Theorem 3.2 continues to hold. However, note that, in this case a linear system with M must be solved at each iteration.

3.3. The hybrid procedure

Now, we are ready to describe the new hybrid procedure which combines the two globalization strategies. At each iteration the backtracking strategy along the inexact Newton step adopted in algorithm INB is tried first. If within a maximum number $MAXBT$ of backtracks no progress is found in the norm of the nonlinear residual, the direction $\bar{s}_k$ is left and the two dimensional elliptical subspace strategy is applied. Our intention is to leave the direction $\bar{s}_k$ when too many backtracks along it are necessary to satisfy the decrease condition:

$$\|F(x_k + s_k)\| \leqslant (1 - t(1 - \eta_k))\|F(x_k)\|. \tag{3.10}$$

This may happen when $\bar{s}_k$ is nearly orthogonal to the gradient of $\|F\|^2$ at x_k and $F'(x_k)$ is ill-conditioned [30].

The resulting kth iteration can be sketched as follows.

HIN Method (Hybrid inexact Newton Method)

Given x_k, $\eta_{\max} \in (0,1)$, $t \in (0,1)$, $0 < \beta_1 < \beta_2 < 1$, $\alpha_1 \in (0,1)$, $0 < \theta_m < \theta_{\max} < 1$, $MAXBT > 0$, $\Delta_{\min} > 0$, $\Delta_k > \Delta_{\min}$, D_k diagonal $n \times n$ matrix, with positive diagonal entries.

1. Choose $\bar{\eta}_k \in [0, \eta_{\max}]$ and compute $\bar{s}_k$ such that
   $\|F(x_k) + F'(x_k)\bar{s}_k\| \leqslant \bar{\eta}_k\|F(x_k)\|$.
2. Perform the backtracking strategy:
   2.1 Set $s_k = \bar{s}_k$, $\eta_k = \bar{\eta}_k$, $nbt = 0$.
   2.2 While $\|F(x_k + s_k)\| > (1 - t(1 - \eta_k))\ \|F(x_k)\|$ & $nbt < MAXBT$ do:
   Choose $\theta \in [\theta_m, \theta_M]$
   Update $s_k = \theta s_k$, $\eta_k = 1 - \theta(1 - \eta_k)$ and $nbt = nbt + 1$.
3. If $\|F(x_k + s_k)\| > (1 - t(1 - \eta_k))\|F(x_k)\|$
   3.1 Let $\mathcal{S}_k = \mathrm{span}\{D_k^{-2}\nabla f(x_k), \bar{s}_k\}$, $ntr = 0$.
   3.2 Do
   3.2.1 Find $s_k = \mathrm{argmin}_{\{\|D_k s\| \leqslant \Delta_k, s \in \mathcal{S}_k\}} m_k(s)$ and set $ntr = ntr + 1$.
   3.2.2 Set $\Delta_k = \alpha_1 \min\{\Delta_k, \|D_k s_k\|\}$.
   While $\rho_k^f(s_k) < \beta_1$
4. Set $x_{k+1} = x_k + s_k$ and choose D_{k+1}.
5. If $\rho_k^f(s_k) \geqslant \beta_2$ then
   set $\Delta_{k+1} = \max(\Delta_{\min}, 2\|D_k s_k\|)$
6. else set $\Delta_{k+1} = \Delta_k$.

The following theorem formalizes the convergence properties of the HIN method. The proof is provided in Appendix.

Theorem 3.3. *Let $r > 0$ and $L = \cup_{k=0}^{\infty}\{x \in \mathbb{R}^n | \|x - x_k\| \leqslant r\}$ be a neighborhood of the sequence $\{x_k\}$ generated by the HIN method. Assume that F' is Lipschitz continuous in L and $\|F'(x)\|$ is bounded above on L. Suppose furthermore that $\nabla f(x_k) \neq 0$ for $k \geqslant 0$, the sequences $\{\|D_k\|\}$ and $\{\|D_k^{-1}\|\}$ are bounded above. If there exists a limit point x^* of $\{x_k\}$ such that $F'(x^*)$ is invertible, then*

(*a*) $\|F(x^*)\| = 0$.
(*b*) $lim_{k\to\infty} x_k = x^*$;
(*c*) $s_k = \bar{s}_k$, *for sufficiently large k.*

This theorem shows that eventually x_{k+1} has the form $x_{k+1} = x_k + \bar{s}_k$. Then, the method inherits the local convergence properties of the INB method.

4. Implementation details

In this section we discuss some crucial implementation issues. First, we focus on the linear algebra phase.

4.1. Construction of the Jacobian and solution of the linear system

The Jacobian matrix $F'(x)$ of the system (2.9), at x_k, can be easily obtained by differentiating the discrete equations of the system (2.9) with respect to the components of the vector x. In this process we make the following simplification: at each iteration we consider the stabilization parameter τ_T of (2.7, 2.8) constant with respect to the components of the vector x_k, i.e. we compute Re_T and τ_T with the velocity components of the previous iteration.

Common CFD problems require a high number of unknown to be solved. Moreover, usually the Jacobian matrix has a very large condition number that increases whenever the Reynolds number increases and the meshsize decreases. Then, when an iterative linear solver is used to compute the inexact Newton step, preconditioning is unavoidable.

In our implementation of the three methods described in the previous section, the inexact Newton step is computed employing the iterative linear solver GMRES [28] with restart and null initial guess. We use an additive Schwarz domain decomposition preconditioner, with an ILU incomplete LU factorization with a *level of fill-in* equal to 2, Reverse Cuthill–McKee reordering of the factors, and one level of overlapping. We preconditioned on the right as right preconditioning leaves unchanged the residual of the linear system. This is quite useful in the context of inexact Newton methods as condition (3.1) is precisely the small unpreconditioned linear residual termination for iterative linear solvers. If within the maximum number of GMRES iterations, GMRES failed to provide a vector $\bar{s}_k$ satisfying (3.1), we continued using the last GMRES iterate s_k^m. This is done letting $\bar{s}_k = s_k^m$ and $\bar{\eta}_k = \|F(x_k) + F'(x_k)\bar{s}_k\|/\|F(x_k)\|$ after Step 1 of the INB, ESTR and HIN methods. We remark that, choosing the null initial guess, the initial residual is given by $\|F(x_k)\|$. Then, thanks to the minimization properties of GMRES and right preconditioning, $\bar{\eta}_k$ is ensured to be smaller than one.

The choice of the forcing term has a strong relevance on the performance of the algorithms discussed here. In all our computations we consider the following four strategies for choosing η_k:

- *Choice* 1 [16,27]: Select $\bar{\eta}_0 \in [0, 1)$ and set

$$\bar{\eta}_k = \frac{|\|F(x_{k+1})\| - \|F(x_k) + F'(x_k)s_k\||}{\|F(x_k)\|}, \quad k = 1, 2, \ldots$$

with the safeguard: $\bar{\eta}_k = \max\{\bar{\eta}_k, \eta_{k-1}^{(1+\sqrt{5})/2}\}$ if $\eta_{k-1}^{(1+\sqrt{5})/2} > 0.1$;
- *Choice* 2 [16,27]: Select $\bar{\eta}_0 \in [0, 1)$, $\gamma \in [0,1]$, $\alpha \in (1,2]$ and set

$$\bar{\eta}_k = \gamma\left(\frac{\|F(x_{k+1})\|}{\|F(x_k)\|}\right)^{\alpha}, \quad k = 1, 2, \ldots$$

with the safeguard: $\bar{\eta}_k = \max\{\bar{\eta}_k, \gamma\eta_{k-1}^{\alpha}\}$ if $\gamma\eta_{k-1}^{\alpha} > 0.1$;
- *Choice* 3: $\eta_k = 1.0\text{E}{-1}$, $\forall k \geqslant 0$;
- *Choice* 4: $\eta_k = 1.0\text{E}{-4}$, $\forall k \geqslant 0$.

We underline that the adaptive choice *Choice* 1 yields q-superlinear convergence of all the three methods considered. On the other hand, *Choice* 2 (with $\gamma = 0.9$ and $\alpha = 2$) guarantees quadratic convergence.

4.2. Globalization strategies

The backtracking strategy along the inexact Newton step, in Step 2 of the INB and HIN methods, is implemented as in the code Nitsol [27]. That is, θ is chosen minimizing the quadratic function that interpolates $\|F\|$ in the direction of the inexact Newton step, see also [25,22].

In Step 3.1 of the ESTR Method and in Step 3.2.1 of the HIN Method, we compute an approximate dogleg solution of the two-dimensional elliptical subspace trust-region problem as follows. Let $W \in \mathbb{R}^{n\times 2}$ be an orthonormal basis for $D_k\mathcal{S}_k$. Then, a vector $p \in D_k\mathcal{S}_k$ is such that $p = W q$ for some $q \in \mathbb{R}^2$ and we can write (3.5) as the following spherical trust-region problem in $\mathbb{R}^2$:

$$\min_{q\in\mathbb{R}^2}\{\psi_k(q) : \|q\| \leqslant \Delta_k\} \tag{4.1}$$

with ψ_k given by:

$$\psi_k(q) = \frac{1}{2}\|F(x_k) + F'(x_k)D_k^{-1}Wq\|^2 \quad q \in \mathbb{R}^2. \tag{4.2}$$

Therefore, we can compute an approximate solution q_k^{tr} to (4.1) by the dogleg strategy and setting $s_k^{\mathrm{tr}} = D_k^{-1}Wq_k^{\mathrm{tr}}$ we get an approximate solution to (3.5). We recall that q_k^{tr} is the point where the dogleg path intersects the trust-region boundary. The dogleg path is the piecewise linear path connecting the origin, the minimizer q_k^U of $\psi_k(q)$ along $W^{\mathrm{T}}D_k^{-1}\nabla f(x_k)$ and the minimizer of $\psi_k(q)$,i.e.

$$q_k^m = \operatorname*{argmin}_{q\in\mathbb{R}^2}\psi_k(q). \tag{4.3}$$

The above least square problem can be solved without much effort either via the QR decomposition of the $n\times 2$ matrix $F'(x_k)D_k^{-1}W$ or by solving the normal equation, that is the following 2×2 linear system:

$$\begin{bmatrix} \|v_1\|^2 & v_2^{\mathrm{T}}v_1 \\ v_2^{\mathrm{T}}v_1 & \|v_2\|^2 \end{bmatrix}\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} -\|D_k^{-1}\nabla f(x_k)\| \\ 0 \end{bmatrix}$$

where $v_1 = F'(x_k)D_k^{-1}w_1$ and $v_2 = F'(x_k)D_k^{-1}w_2$. We remark that both approaches require two Jacobian-vector products in order to compute v_1 and v_2.

The dogleg strategy can be summarized in the following algorithm where we use the following simplifications. In place of $WW^{\mathrm{T}}D_k^{-1}\nabla f(x_k)$ we compute $D_k^{-1}\nabla f(x_k)$, since these two vectors coincide as $D_k^{-1}\nabla f(x_k)$ belongs to $D_k\mathcal{S}_k$. Further, the first column of W is chosen as $D_k^{-1}\nabla f(x_k)/\|D_k^{-1}\nabla f(x_k)\|$ so that we have $W^{\mathrm{T}}D_k^{-1}\nabla f(x_k) = [\|D_k^{-1}\nabla f(x_k)\|, 0]^{\mathrm{T}}$.

Dogleg two dimensional subspace strategy

Input parameters $x_k \in \mathbb{R}^n$, $\Delta_k > 0$, $\bar{s}_k \in \mathbb{R}^n$.

1. Compute $D_k^{-1}\nabla f(x_k)$ and $w_1 = D_k^{-1}\nabla f(x_k)/\|D_k^{-1}\nabla f(x_k)\|$.
2. Compute $w_2 = D_k\bar{s}_k - (w_1^{\mathrm{T}}D_k\bar{s}_k)w_1$. Set $w_2 = w_2/\|w_2\|$.
3. Set $W = [w_1, w_2]$.
4. Compute $q_k^m = \operatorname{argmin}_{q\in\mathbb{R}^2}\psi_k(q)$.
5. Compute $q_k^U = \operatorname{argmin}_{\tau>0}\psi_k(-\tau W^{\mathrm{T}}D_k^{-1}\nabla f(x_k))$ setting

$$q_k^U = -\frac{\|D_k^{-1}\nabla f(x_k)\|^2}{\|F'(x_k)D_k^{-2}\nabla f(x_k)\|^2}W^{\mathrm{T}}D_k^{-1}\nabla f(x_k).$$

6. Find the dogleg solution q_k^{tr} to (4.1) by setting:

$$q_k^{\mathrm{tr}} = \begin{cases} q_k^m, & \text{if } \|q_k^m\| \leqslant \Delta_k, \\ -\frac{\Delta_k}{\|D_k^{-1}\nabla f(x_k)\|}W^{\mathrm{T}}D_k^{-1}\nabla f(x_k), & \text{if } \|q_k^U\| \geqslant \Delta_k, \\ uq_k^m + (1-u)q_k^U, u \in (0,1), & \text{s.t.}\|q_k^{\mathrm{tr}}\| = \Delta_k, \quad \text{otherwise.} \end{cases} \tag{4.4}$$

7. Compute $s_k^{\mathrm{tr}} = D_k^{-1}Wq_k^{\mathrm{tr}}$.

Summarizing, the step q_k^{tr} is the classical dogleg solution to (4.1), while the approximate solution s_k^{tr} of (3.3), is obtained as the intersection of the trust region with the piecewise-linear path connecting zero, the minimizer of $m_k(s)$ along the scaled steepest descent direction $D_k^{-2}\nabla f(x_k)$ and $D_k^{-1}Wq_k^m$.

Regarding the shape of the trust-region, in our computations we adopted a spherical trust-region ($D_k = I$, for all $k > 0$) as well as an elliptical one. In this latter case, letting $\|(F'_k)_l\|$ be the lth column of $F'(x_k)$, the lth diagonal element of the matrix D_k has been chosen as follows:

$$(D_0)_l = \|(F'_0)_l\|_1,$$

and

$$(D_{k+1})_l = \begin{cases} \max(\|(F'_k)_l\|_1, 0.6(D_k)_l), & \text{if } k = m \cdot \mathit{Diter} \text{ for some integer } m, \\ (D_k)_l, & \text{otherwise.} \end{cases} \tag{4.5}$$

This update technique has been proposed in [12] with $\mathit{Diter} = 1$, i.e. the update is performed at each iteration. Here, due to the quite relevant cost of the computation of $\|(F'_k)_l\|$ at each iteration, we update the diagonal matrix every $\mathit{Diter} = 50$ iterations.

4.3. Scaling and stopping criterion

We enrich all the presented methods with the scaling strategy suggested in [25]. More precisely, at each nonlinear iteration k, each row of the Jacobian matrix $F'(x_k)$ and the corresponding element of the right-hand-side $-F(x_k)$ are rescaled by the sum of the absolute values of the entries in the corresponding row of the Jacobian matrix, i.e. the Jacobian matrix and the right-hand-side are premultiplied by the diagonal matrix whose ith diagonal entry is given by $1/\sum_{j=1}^{n}|(F'(x_k))_{i,j}|$. The same scaling is applied to $F(x_{k+1})$ in the termination test.

For all the methods considered here convergence is declared when the following condition is met:

$$\|F(x_{k+1})\| \leqslant \mathtt{tol}, \tag{4.6}$$

where `tol` is a user supplied tolerance. On the other hand, failure is declared if a maximum number of iterations `maxit` are performed without satisfying condition (4.6). Moreover, a failure of the INB method is detected also if condition (3.10) is not satisfied within 10 backtracks; a failure of the ESTR method and of the HIN method is declared whenever the trust-region size is reduced below $\sqrt{\epsilon_m}$.

5. Numerical results

In this section we present some numerical results. The aim of this experimentation is twofold. First, we intend to investigate the effect of some implementation choices on the behaviour of the globalization strategies. Namely, the scaling of the problems, the trust-region shape in the ESTR method and the maximum number of backtracks along the inexact Newton step before switching to the trust-region strategy in the HIN method. Second, we wish to verify if the proposed hybrid strategy enhances the robustness of the INB method. Namely, we present comparisons between the following six codes:

- INB-S (INB method with scaling of the problem).
- INB (INB method without scaling of the problem).
- SSTR (Spherical Subspace Trust-Region method without scaling of the problem ($D_k = I$ for all k's in the ESTR method)).
- ESTR-S (ESTR method with scaling of the problem (D_k given by (4.5))).
- HIN-S-2 (HIN method, with MAXBT=2, scaling of the problem and elliptical trust-region (D_k given by (4.5))).
- HIN-S-4 (HIN method, with MAXBT=4, scaling of the problem and elliptical trust-region (D_k given by (4.5))).

All the codes are written in C with double precision variables and we run them on a PC with AMD Opteron(tm) Processor 250 and 2 GB of RAM.

For all the methods we set to 600 the maximum number of GMRES iterations, to 200 the restart value, and we adopted the following choices: $\eta_{\max} = 1.0 - 1.0\text{E}{-4}$, $t = 1.0\text{E}{-4}$, $\theta_m = 0.1$, $\theta_{\max} = 0.5$. Further, Algorithms ESTR and HIN were run setting $\Delta_0 = 1.0$, $\Delta_{\min} = 1.0\text{E}{-4}$, $\alpha_1 = 0.5$, $\beta_1 = 0.25$, $\beta_2 = 0.75$. For all the methods, we set the stopping tolerance $\texttt{tol} = 1.0\text{E}{-6}$ and the maximum number of nonlinear iterations $\texttt{maxit} = 200$.

We adopted the four choices of $\bar{\eta}_k$'s described in Section 4.1; specifically *choice* 2 was implemented with $\gamma = 0.9$ and $\alpha = 2$ and $\eta_0 = 0.1$ was used for both the adaptive choices. The linear systems were solved employing codes belonging to the PETSc library [2].

We considered two classical CFD benchmark problems: the Lid Driven Cavity and the Backward Facing Step, with wide-ranging Reynolds numbers and discretization grids.

In all our run we used $\lambda = 0$ in (2.7) and as initial guess the solution of the corresponding Stokes problem.

We remark that we always compare codes with the same choice of the forcing terms.

To visualize the overall performances of the considered globalization strategies, we employ the performance profiles approach [13]. In this approach, when m solvers are compared on a test set, the performance of each solver in the solution of a test is measured by the ratio of its computational effort and the best computational effort by any solver on this test. Here we use two different quantities to measure the computational effort of each code: the total number of GMRES iterations and the number of function evaluations performed during the overall solution of each test. Specifically, for each test t solved by the solver s we denote by $Q_{s,t}$ the total number of GMRES iterations (or the number of function evaluations) required by the solver s to solve the test t. Moreover, we denote by $\underline{Q}_{s,t}$ the lowest number of GMRES iteration (or function evaluations) required by all the solvers to solve test t. Then, the ratio

$$q_{s,t} = \frac{Q_{s,t}}{\underline{Q}_{s,t}}$$

measures the performance on test t by solver s with respect to the best performance among all the solvers on such test. Clearly, $q_{s,t} \geqslant 1$ and $q_{s,t} = 1$ means that the solver s is the most effective in solving the test t over all the solvers. Then, the performance profile of solver s is defined as

$$\pi_s(\tau) = \frac{\text{no. of tests s.t. } q_{s,t} \leqslant \tau}{\text{total no. of tests}}, \quad \tau \geqslant 1.$$

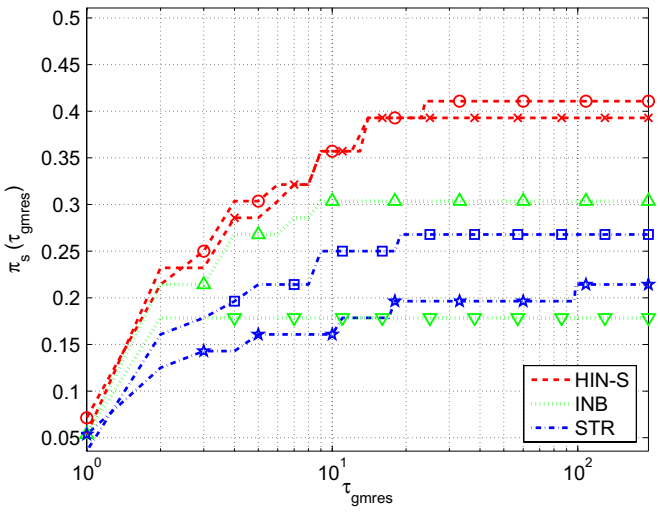

Fig. 1. Cavity problem, η: *choice* 1, GMRES iterations.

Moreover, assume that a parameter $q_M > q_{s,t}$ for all s,t is chosen. Then, if the solver s fails in solving test t, $q_{s,t}$ is set to q_M. Note that $q_{s,t} = q_M$ if and only if solver s does not solve test t. As a result of this convention $\pi_s(q_M) = 1$ and $\lim_{\tau \to q_M^-} \pi_s(\tau)$ is the probability that the solver solves a problem. Moreover, the performance profiles flattens for $\tau \in [\bar{\tau}, q_M]$ for some $\bar{\tau} < q_M$. Then, if $\pi_s(\tau)$ is plotted in $[0, \bar{\tau}]$, the value of $\lim_{\tau \to q_M^-} \pi_s(\tau)$ can be readily seen in the performance profile's plot and the right side of the plot gives the percentage of the test problems that were successfully solved by the solver. On the other hand, the left side of the plot gives the percentage of test problems for which the solver is fastest. We report some performance profiles in Figs. 1–4, 8–11. In all the figures the dashed line with circle markers corresponds to the HIN-S-2 method and the dashed line with the x-marks to the HIN-S-4 method. The dotted line with up-pointing triangular markers corresponds to the INB-S method and the dotted line with down-pointing triangular markers corresponds to INB method. Finally, the dash-dotted line with pentagram markers refers to the ESTR-S method, while the dash-dotted line with square markers refers to the SSTR method.

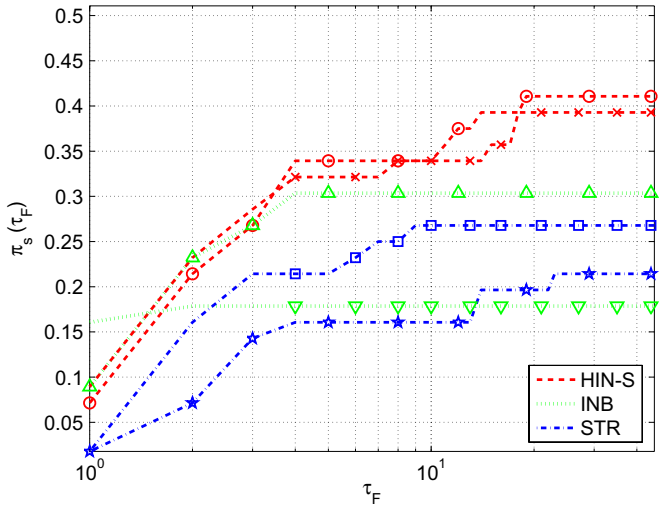

Fig. 2. Cavity problem, η: *choice* 1, function evaluation.

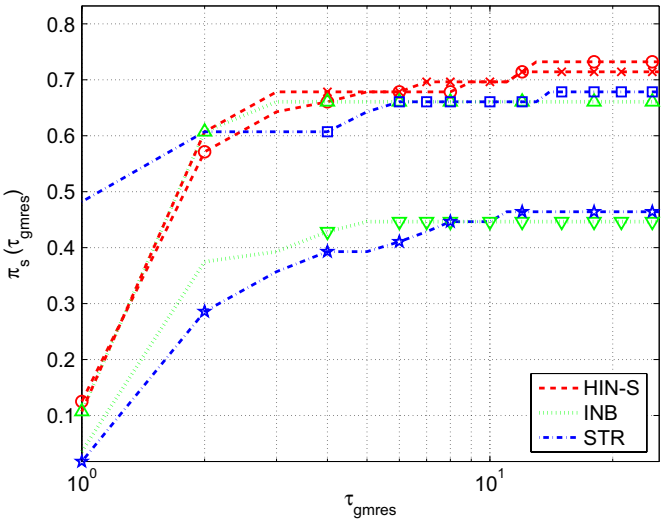

Fig. 3. Cavity problem, η: *choice* 2, GMRES iterations.

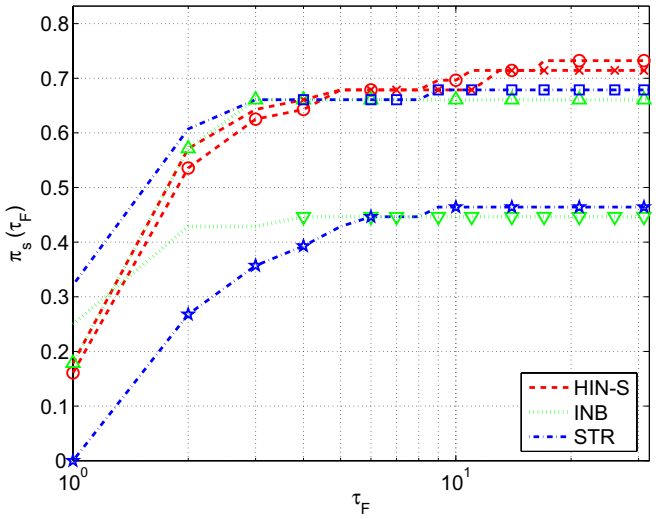

Fig. 4. Cavity problem, η: *choice* 2, function evaluation.

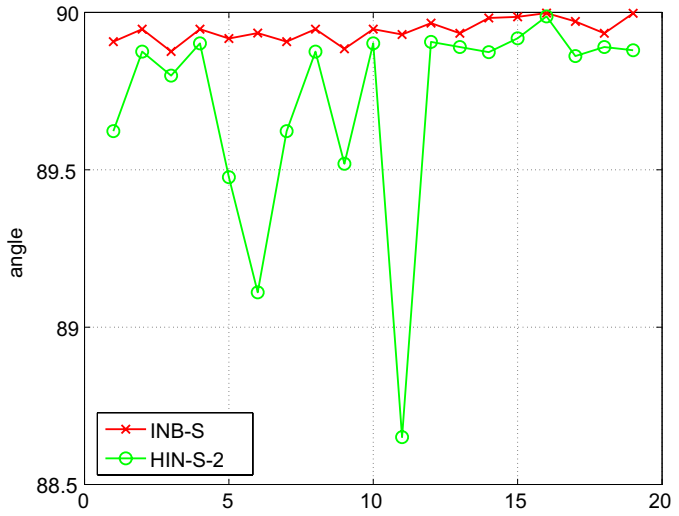

Fig. 5. Cavity problem, η: *choice* 2, comparison of the angles between s_k and $-\nabla f(x_k)$.

5.1. Lid driven cavity

The lid driven cavity problem is a classical CFD benchmark problem [18]. The geometry of this problem is the unit square, we impose no slip boundary conditions ($u_1 = u_2 = 0$) on the two vertical walls and on the bottom of the cavity and $u_1 = 1$ and $u_2 = 0$ on the top of the cavity. We solved it with the following Reynolds numbers: $Re = 1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000$ and using essentially uniform grids with a regular distribution of nodes, with the following number of nodes: 437, 843, 1652, 2684, 4079, 11354, 19803. All these meshes are uniformly subdivided in nine subdomains for the Schwarz preconditioner. Then, from the lid driven cavity problem we obtain 56 tests (8 Reynolds numbers $\times$ 7 meshes).

The considered grids ranges from very coarse grids, where the approximate solution of the problem is widely inaccurate and hard to get, to quite fine grids, where the solution is more accurate and easier to get.

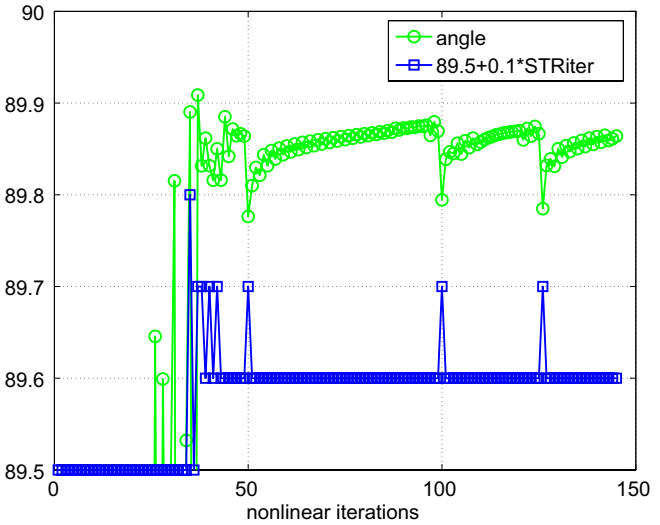

Fig. 6. Cavity problem, η: *choice* 2, HIN-S-2, behaviour of the angle between s_k and $-\nabla f(x_k)$ during the convergence process.

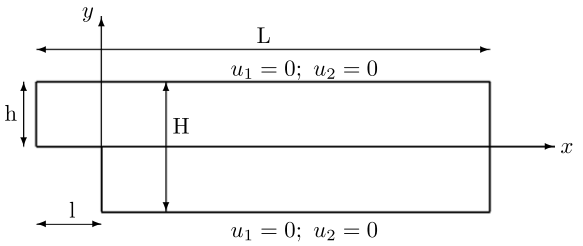

Fig. 7. Geometry of the backward facing step.

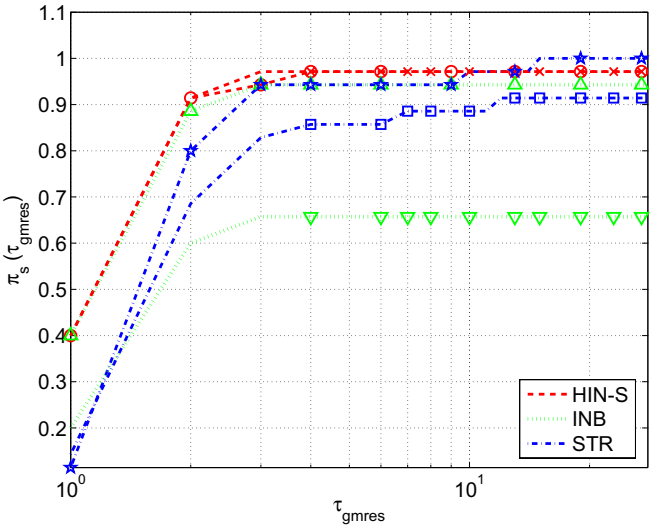

Fig. 8. BFS problem, η: *choice* 1, GMRES iteration.

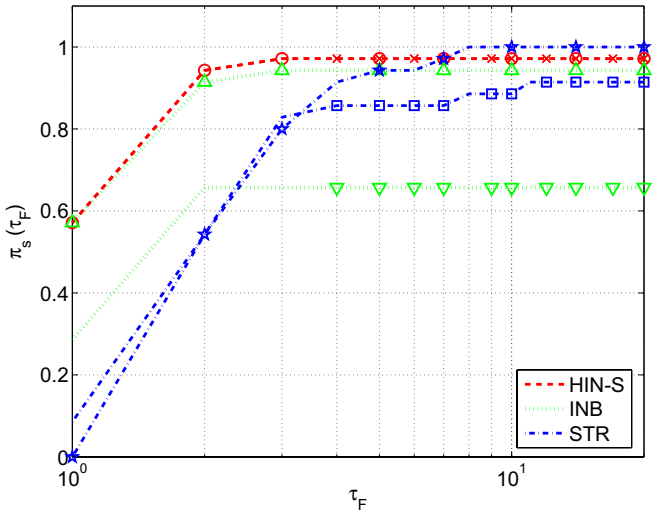

Fig. 9. BFS problem, η: *choice* 1, function evaluation.

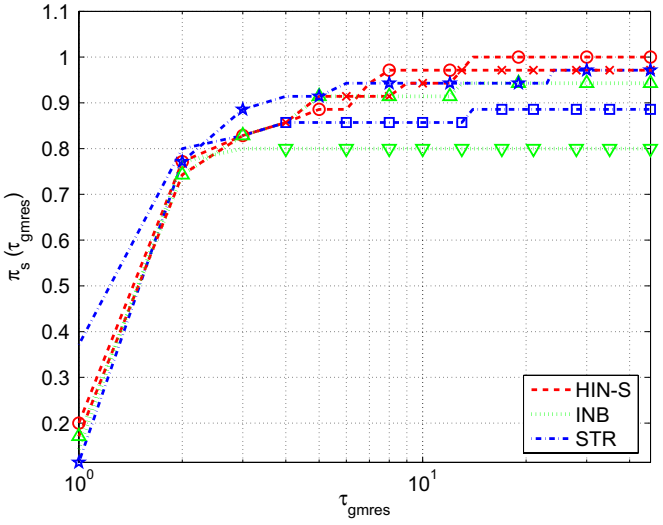

Fig. 10. BFS problem, η: *choice* 2, GMRES iterations.

In Table 1, we report, for each choice of the forcing terms and for each code, the number of successes over the 56 tests. In the last column $\sum_\eta$ is the number of successes of each code with all the considered choices of the forcing terms (224 runs), i.e. is the sum of the previous entries of the same row of the table.

In Figs. 1–4 we compare the performance profiles of the codes, according to the two adopted computational effort measures, with forcing terms given by *choice* 1 and *choice* 2, respectively.

As we can see from Table 1 and Figs. 1–4, these tests are quite difficult to be solved. We underline that the large number of failures is also due to the quite coarse grid we consider. However, in some cases it is important to compute a solution of the nonlinear system on such grids, too. For example, coarse grids are usually the starting point of an adaptive method. Anyway, the hybrid approach is the more robust; the HIN-S-2 method solves about the 42% of tests with *choice* 1 of η_k's, and this percentage increases about to 75% with *choice* 2 of η_k's. Regarding the robustness of the method HIN-S-4 performs slightly worse, as it succeeded on the 39% and 75% of tests with forcing terms given by *choice* 1 and *choice* 2, respectively. On the other hand INB-S solves

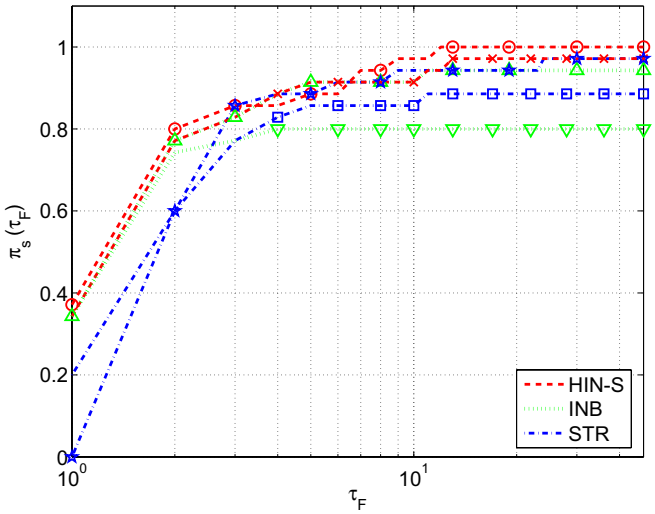

Fig. 11. BFS problem, η: *choice* 2, function evaluation.

Table 1
Cavity: number of successes over 56 tests

	*Choice*1	*Choice*2	$\eta = 1.0\text{E}{-1}$	$\eta = 1.0\text{E}{-4}$	$\sum_\eta$
HIN-S-2	23	41	25	16	105
HIN-S-4	22	40	23	16	101
INB-S	17	37	20	11	85
INB	10	25	11	7	53
ESTR-S	12	26	20	12	70
SSTR	15	38	11	12	76

the 30% of test with *choice* 1 of forcing terms and 65% of tests with η_k's given by *choice* 2. It is also evident that the scaling enhances robustness of the INB approach. Focusing on the trust-region method, it can be observed that the elliptical trust-region with scaling deteriorates the performance of this approach (only 45% of test successfully solved by ESTR-S against the 65% of SSTR, with *choice* 1 of η_k's). This is quite surprising as the elliptical shape of the trust-region plays a key role in the performance of the HIN approach as well as the scaling is crucial to improve robustness of all the other methods.

Regarding the efficiency of the considered codes, we can see that we do not have a clear winner, even if, with η_k's chosen according to *choice* 2, the spherical trust-region performs better. We underline that the performance profile of the HIN-S and the INB-S codes is almost the same for small values of τ. This is nice as it means that the switch to the trust-region strategy does not affect the performance of the HIN method in solving problems where also the backtracking strategy works well.

More insight into the behaviour of the HIN method can be gained monitoring the angle between the gradient of $f(x)$ at x_k and the selected step s_k. We remark that the INB-S and the HIN-S method coincide up to the iteration $\bar{k}$ where the INB-S method performed more than *MAXBT* backtracks and the hybrid method switches to the subspace trust-region strategy. As an example, we focus on HIN-S-2 and INB-S with η_k's given by *choice* 2, and for each test successfully solved by HIN-S-2 method and unsolved by the INB-S method we plot in Fig. 5 two points. That denoted by the x-mark shows the value of the angle between the step $s_{\bar{k}}$ produced by the INB-S method and $-\nabla f(x_{\bar{k}})$; the point marked by a circle displays the value of the angle between the step $s_{\bar{k}}$ employed by the HIN-S-2 method and $-\nabla f(x_{\bar{k}})$, both at the first occurrence of a switch. This figure clearly shows, in a compact form, that when the INB-S method fails, the inexact Newton step is nearly orthogonal to the gradient of the function f. On the other hand,

when the hybrid approach manages to overcome these failures, the elliptical subspace trust-region strategy produces a step $s_{\bar{k}}$ slightly better angled with respect to $-\nabla f(x_k)$ and this yields the convergence of the procedure. We remark that the step $s_{\bar{k}}$ is never nearly parallel to the gradient of f. Thus, the hybrid method does not reduce to the steepest descent algorithm. In Fig. 6 we further restrict our attention to a representative test, namely that obtained with $Re = 6000$ and 843 nodes. This tests corresponds to test thirteen in Fig. 5. The upper graph in Fig. 6 plots the value of the angles between the step s_k chosen by the hybrid method and $-\nabla f(x_k)$ throughout the convergence process; the bottom graph shows the number of tentative steps computed by the trust-region strategy at each nonlinear iteration (denoted by STRiter in the picture's label) of the HIN-S-2 method. From the comparison between the two graphs we can observe that when the trust-region strategy is activate the angle formed by s_k and $-\nabla f(x_k)$ is close to 90^o and reductions of the trust-region radius give rise to reductions of the angle.

We underline that when also the hybrid method fails, failures are due to a quasi stagnation of the procedure: the method achieves sufficient reduction of $\|F\|$ to proceed, but not enough to achieve the satisfaction of the stopping criterion within 200 nonlinear iterations.

We would like to stress that the elliptical shape of the trust-region was fundamental to enhance robustness of the HIN method. In fact, we performed several runs employing spherical trust-regions in the HIN method. However, the switch to the spherical trust-region did not produce any benefit, as the step produced by the trust-region was nearly parallel to the inexact Newton step until the trust-region radius became very small. Then, also the trust-region approach failed in producing a sufficient reduction of the nonlinear residual. This behaviour is to be ascribed to the badly scaling of the problems; we verified that the inexact Newton step became very long while the Cauchy step was very short.

In order to show that these problems become more and more difficult to be solved as the Reynold number increases and the number of nodes decreases, in Tables 2 and 3 we report the number of total failures of the HIN-S-2 method, HIN-S-4 method, SSTR method and INB-S method, all with *choice* 2 of the forcing terms, for each considered Reynolds number and for each considered grid.

Finally, to give an idea on the overall cost of the HIN approach, in Table 4, we give detailed results of our experimentation with the HIN-S-2 method and with *choice* 2 of the forcing terms on the finest grid. The following values are given in the tables:

Table 2
Cavity problem: number of failures over 7 tests for each Reynolds number

Re	INB-S	HIN-S-2	HIN-S-4	SSTR
1000	0	0	0	0
2000	1	1	1	1
3000	1	1	1	1
4000	2	1	1	2
5000	2	1	2	1
6000	3	2	2	3
8000	4	4	4	5
10000	6	5	5	5

Table 3
Cavity problem: number of failures over eight tests for each grid

Node	INB-S	HIN-S-2	HIN-S-4	SSTR
437	7	7	7	6
843	4	2	2	2
1652	4	3	4	1
2684	1	2	1	0
4079	2	1	2	3
11354	0	0	0	2
19803	1	0	0	4

Table 4
Cavity: HIN-S-2, $\eta = Choice$ 2, 19,803 nodes

Re	*iter*	$GMiter_M$	$\lceil\overline{GMiter}\rceil$	nbt_M	ntr_M	*nsw*	$\|f\|_2$
1000	15	141	44.07	1	0	0	2.97E−9
2000	22	134	34.14	1	0	0	6.61E−7
3000	32	129	36.78	1	0	0	2.20E−7
4000	48	128	33.85	1	0	0	6.91E−7
5000	60	126	50.10	1	0	0	9.96E−7
6000	79	130	43.03	1	0	0	9.83E−7
8000	111	109	50.72	2	2	32	9.44E−7
10000	137	108	51.20	2	2	53	8.66E−7

- *iter*: number of nonlinear iterations;
- $\lceil\overline{GMiter}\rceil$: average number of GMRES iterations per inexact Newton step;
- $GMiter_M$: maximum number of GMRES iterations;
- nbt_M: maximum number of backtracks along $\bar{s}_k$;
- ntr_M: maximum number of trust-region radius reductions;
- nsw: number of switches to subspace trust-region strategy.

At this regard, we remark that the method succeeds in solving all tests. The tests corresponding to Reynolds number up to 6000 are solved without switching to the trust-region, that means that they are successfully solved by the INB-S method, too. On the other hand, switching occurs in the last two tests of the table. We note that the number of nonlinear iterations grows with the Reynolds number.

5.2. Backward facing step

The backward facing step is an interesting prototype for an internal, separated flow [23]. The geometry is described in Fig. 7 the parameters used in this simulations are $L = 44$, $l = 6$, $H = 2$, $h = 1$.

A laminar fully developed velocity profile is imposed at the inflow section of dimension h. Homogeneous Neumann boundary conditions $\frac{1}{Re}\frac{\partial u}{\partial \hat{n}} - p\hat{n} = 0$ are imposed at the outflow of dimension H. No-slip boundary conditions are applied on the top and the bottom walls. The characteristic velocity used to define the Reynolds number is the maximal physical velocity in the inflow section, the characteristic length is the difference between the physical height of the channel at outflow and at the inflow sections.

We solved this problem with Reynolds numbers: $Re = 200$, 400, 500, 600, 650, 700, 800 and using quasi uniform grids of 1058, 2412, 6703, 13463, 22181 nodes, all the meshes are subdivided in nine subdomains for the Schwarz preconditioner.Then, we obtain 35 tests.

In Table 5 we report, for each choice of the forcing terms and for each code, the number of successes over the 35 tests. In the last column $\sum_\eta$ is the number of the successes of each code over all the considered choices of the forcing terms (140 runs), i.e. is the sum of the previous entries on the same row of the table.

In Figs. 8 and 9 we plot the performance profiles of the codes with forcing terms chosen according to *choice* 1, while in in Figs. 10 and 11 we plot the performance profiles of the codes with forcing terms chosen according to *choice* 2.

Table 5
Backward facing step: number of successes over 35 tests

	*Choice*1	*Choice* 2	$\eta = 1.0E{-}1$	$\eta = 1.0E{-}4$	$\sum_\eta$
HIN-S-2	34	35	34	30	133
HIN-S-4	34	34	34	31	133
INB-S	33	33	33	24	123
INB	23	28	27	21	99
ESTR-S	35	34	31	25	125
SSTR	32	31	34	26	123

Table 6
BFS: HIN-S-2 method, $\eta = Choice$ 2, 22181 nodes

Re	*iter*	$GMiter_M$	$\lceil \overline{GMiter} \rceil$	nbt_M	ntr_M	*nsw*	$\|f\|_2$
200	7	79	33.29	1	0	0	8.14E−8
400	16	67	27.44	1	0	0	2.17E−8
500	34	60	24.91	1	0	0	6.56E−7
600	84	66	28.37	2	2	9	6.47E−7
650	87	62	27.55	1	0	0	5.84E−7
700	93	59	27.57	1	0	0	9.57E−7
800	107	56	28.43	1	0	0	8.47E−7

From these plots is quite evident that this problem is less difficult to be solved by the considered codes than the cavity problem. In any case, both the hybrid strategies are quite robust and the HIN-S-2 (*choice* 2) method solves all the tests. Moreover, we remark that all the failures of the INB-S method are to be ascribed to the fact that the gradient of f is nearly orthogonal to the inexact Newton step and the switch to the trust-region strategy allowed to overcome these failures.

It should be noted that, in this case the elliptical subspace trust-region strategy performs better, in terms of robustness, than the spherical one as it solves about the 95% of the tests with forcing terms given by *choice* 2 and all the tests with *choice* 1 of η_k's. Regarding the efficiency, again the HIN method and the INB have a very similar behaviour, to confirm that the switching to the trust-region strategy does not add any extra cost in the solution of tests that are easily solved by the backtracking strategy. Comparing all codes with η_k's given by *choice* 1, we can observe that INB-S, HIN-S-2 and HIN-S-4 are the best, in terms of F-evaluations, on about the 60% of tests. Considering the computational effort given by the number of GMRES iterations, they are the most efficient on about the 40% of tests. On the other hand, focusing on *choice* 2 of the forcing terms, there is not a clear winner; the SSTR method requires less GMRES iterations on about the 40% of tests while the INB-S and the hybrid methods are the best according to the number of function evaluations.

To give an idea on the overall cost of the HIN approach, in Table 6, we give detailed results of our experimentation with the HIN-S-2 method and η=*choice* 2 on the finest grid.

5.3. Conclusions on the numerical tests

Some conclusions can be drawn from these numerical results:

- The combination of the two globalization strategies enhances the classical backtracking approach. More precisely, mainly it improves robustness whenever the backtracking strategy fails as the gradient of $f(x)$ and the inexact Newton step are nearly orthogonal.
- At the same time, focusing on the efficiency of the codes, the HIN-S and the INB-S approach perform very similarly; in fact, we underline that the HIN-S method reduces to INB-S in all the situations in which INB-S succeeds performing no more than $MAXBT$ backtracks. More precisely, solving the cavity problem, on a total of 54 tests successfully solved by INB-S, HIN-S-2 and HIN-S-4 switch to the trust-region strategy only in the solution of 5 tests and 1 test, respectively. Regarding BFS problem, HIN-S-4 reduces to INB-S on all the 66 tests successfully solved by INB-S, while HIN-S-2 activates the alternative strategy only in the solution of 1 of such tests.
- Failures of the INB-S method are to be ascribed to failure of the globalization procedure, i.e. the backtracking lacks in producing an acceptable step within 10 steplenghts reductions. This happens in the 100% of failures with the BFS problem and in the 90% of tests arising by the cavity problem.
- Despite the employment of the subspace trust-region strategy allows to enhance robustness of the backtracking procedure, neither the spherical trust-region nor the elliptical one seem to be competitive with the HIN approach for all the tests.

Acknowledgments

The authors would like to thank the referees for their detailed reading and the insightful comments which led to significative improvements in the paper.

The first author was supported by Italian funds "Indam-GNCS" and MIUR-PRIN-2004 "Sviluppo di metodi numerici e algoritmi per applicazioni a problemi di fluidodinamica ambientale". The second author was supported by Italian funds MIUR-PRIN-2004 "Adattività e avanzamento in tempo nei modelli numerici alle derivate parziali" and MIUR-PRIN-2006 "Nuove tecniche di accoppiamento di modelli e di metodi numerici nel trattamento di problemi differenziali".

Appendix A

Here we sketch the proofs of the convergence results (Theorems 3.2 and 3.3) of the ESTR Method and HIN Method. First, we show some key results regarding both methods.

First, we note that, the vector $s_k = \operatorname{argmin}_{\{\|D_k s\| \leqslant \Delta_k, s \in \mathcal{S}_k\}} m_k(s)$ produces a decrease in the model function m_k, greater than the reduction attained by the so called Cauchy point. In other words, the following inequality holds:

$$m_k(0) - m_k(s_k) \geqslant m_k(0) - m_k(s_k^c) \tag{A.1}$$

where s_k^c is the Cauchy point, i.e. the minimizer of m_k along the scaled steepest descent direction subject to satisfying the trust-region bound, that is

$$s_k^c = -\tau_k D_k^{-2} \nabla f(x_k), \tag{A.2}$$

where $\tau_k = \operatorname{argmin}_{\tau>0}\{m_k(-\tau D_k^{-2}\nabla f(x_k)) : \|\tau D_k^{-1}\nabla f(x_k)\| \leqslant \Delta_k\}$ [24]. We underline that relation (A.1) holds because $D_k^{-2}\nabla f(x_k) \in \mathcal{S}_k$ and this obviously yields $m_k(s_k) \leqslant m_k(s_k^c)$. We remark that the approximate trust-region solution s_k^{tr} provided by the dogleg two dimensional subspace strategy satisfies (A.1), too.

Moreover, a classical trust-region result [24] shows that a step s_k satisfying (A.1), satisfies the following inequality:

$$m_k(0) - m_k(s_k) \geqslant \frac{1}{2}\|D_k^{-1}\nabla f(x_k)\| \min\left(\Delta_k, \frac{\|D_k^{-1}\nabla f(x_k)\|}{\|F'(x_k)D_k^{-1}\|^2}\right). \tag{A.3}$$

First, we prove that both the ESTR method and the HIN method are well defined. These methods contain the same repeat loop, that is the repeat loop at Step 3 of the ESTR method and the repeat loop at Step 3.2 of the HIN method. Next theorem shows that this repeat loop terminates.

Lemma 6.1. *Let x_k be the kth iterate generated by either the ESTR method or the HIN method, and $B(x_k,r) = \{x \| \|x - x_k\| \leqslant r\}$. Assume that there exists $r > 0$, such that F' is Lipschitz continuous in $B(x_k,r)$ and $\|F'(x)\|$ is bounded above on $B(x_k,r)$. Then, if $\nabla f(x_k) \neq 0$, there exists $\Delta_k > 0$ such that $s_k = \operatorname{argmin}_{\{\|D_k s\| \leqslant \Delta_k, s \in \mathcal{S}_k\}} m_k(s)$ satisfies (3.7).*

Proof. Let γ_L be the Lipschitz constant of F' in $B(x_k,r)$. From the Lipschitz continuity of F' it follows that $\nabla f(x) = F'(x)^{\mathrm{T}}F(x)$ is Lipschitz continuous in $B(x_k,r)$ with constant $\hat{\gamma} = \gamma_L\tilde{\gamma} + \tilde{\gamma}^2$, where $\tilde{\gamma} = \max\{\sup_{x\in L} f(x), \sup_{x\in L}\|F'(x)\|\}$ and $L = \cup_{k=0}^{\infty}\{x \in \mathbb{R}^n \| \|x - x_k\| \leqslant r\}$ (see [24]). Let $\hat{\beta}$ be a positive constant such that $\|F'(x)\| < \hat{\beta}$, whenever $x \in B(x_k,r)$.

From the Taylor's Theorem it follows

$$f(x_k + s_k) = f(x_k) + \nabla f(x_k)^{\mathrm{T}} s_k + \int_0^1 (\nabla f(x_k + \xi s_k) - \nabla f(x_k))^{\mathrm{T}} s_k \, \mathrm{d}\xi$$

then, as $\|D_k s_k\| \leqslant \Delta_k$, we get

$$\begin{aligned} |m_k(s_k) - f(x_k + s_k)| &= \left|\frac{1}{2} s_k^{\mathrm{T}} F'(x_k)^{\mathrm{T}} F'(x_k) s_k - \int_0^1 (\nabla f(x_k + \xi s_k) - \nabla f(x_k))^{\mathrm{T}} s_k \, \mathrm{d}\xi\right| \leqslant \frac{1}{2}(\hat{\beta}^2\|s_k\|^2 + \hat{\gamma}\|s_k\|^2) \\ &\leqslant \frac{\|D_k^{-1}\|^2 \Delta_k^2}{2}(\hat{\beta}^2 + \hat{\gamma}). \end{aligned}$$

This, along with (3.7) and (A.3), yields

$$|\rho_k^f(s_k) - 1| = \left|\frac{m_k(s_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)}\right| \leqslant \frac{\|D_k^{-1}\|^2 \Delta_k^2(\hat{\beta}^2 + \hat{\gamma})}{\|D_k^{-1}\nabla f(x_k)\| \min\left(\Delta_k, \frac{\|D_k^{-1}\nabla f(x_k)\|}{\hat{\beta}^2\|D_k^{-1}\|^2}\right)}.$$

Then, there exists $\bar{\Delta}_k$ such that $\min\left(\Delta_k, \frac{\|D_k^{-1}\nabla f(x_k)\|}{\hat{\beta}^2\|D_k^{-1}\|^2}\right) = \Delta_k$ and $\frac{\|D_k^{-1}\|^2\Delta_k(\hat{\beta}^2+\hat{\gamma})}{\|D_k^{-1}\nabla f(x_k)\|} < 1 - \beta_1$ whenever $\Delta_k < \bar{\Delta}_k$. This implies $\rho_k^f(s_k) > \beta_1$, whenever $\Delta_k < \bar{\Delta}_k$, i.e. the repeat loop terminates after a finite number of reductions of the trust-region radius. □

The convergence results given in Theorems 3.2 and 3.3 can be obtained exploiting the following theorem of [15].

Theorem 6.1. [15, Th. 3.3–3.4] *Let* $\{x_k\}$ *be a sequence such that, for each* k,

$$\|F(x_k) + F'(x_k)s_k\| \leqslant \eta_k\|F(x_k)\|, \tag{A.4}$$

$$\|F(x_k + s_k)\| \leqslant (1 - t(1 - \eta_k))\|F(x_k)\|, \tag{A.5}$$

where $s_k = x_{k+1} - x_k$, $\eta_k \in [0, 1)$ *and* $t \in (0, 1)$. *Then,*

(*a*) *If* $\sum_{k=0}^{\infty}(1 - \eta_k)$ *is divergent, then* $\|F(x_k)\| \to 0$.
(*b*) *If* $\|F(x_k)\| \to 0$ *and* x^* *is a limit point of* $\{x_k\}$ *such that* $F'(x^*)$ *is invertible, then* $F(x^*) = 0$ *and* $x_k \to x^*$.

It is easy to show that the sequences $\{x_k\}$ and $\{\eta_k\}$ generated by the HIN method satisfy the Assumptions of the above theorem. In fact, if s_k is computed in Step 2, (A.4) is verified (see [15]) and (A.5) obviously holds, too. On the other hand, if s_k is computed in Step 3, then (A.4) is verified with $\eta_k = \|F(x_k) + F'(x_k)(s_k)\|/\|F(x_k)\|$. Note that, η_k is ensured to be smaller than one, since s_k is the minimizer of $m_k(s)$ in $\mathcal{S}_k$, within the trust-region and $m_k(0) = \|F(x_k)\|^2$. Further, (3.7) and (A.4) yield:

$$\begin{aligned}\|F(x_k + s_k)\|^2 &\leqslant \|F(x_k)\|^2 - \beta_1(\|F(x_k)\|^2 - \|F(x_k) + F'(x_k)s_k\|^2) \leqslant (1 - \beta_1)\|F(x_k)\|^2 + \beta_1\eta_k^2\|F(x_k)\|^2 \\ &\leqslant (1 - \beta_1(1 - \eta_k))\|F(x_k)\|^2.\end{aligned}$$

Noting that $\sqrt{1 - \beta_1(1 - \eta_k)} \leqslant 1 - \beta_1(1 - \eta_k)/2$, we can conclude that (A.5) is satisfied with $t = \beta_1/2$. From the above discussion, it is clear that the sequences $\{x_k\}$ and $\{\eta_k\}$ generated by the ESTR method satisfy the assumptions of the above theorem, too.

Then, using the above theorem and classical results of global convergence of trust-region methods, we can prove Theorem 3.2.

Proof of Theorem 3.2. Let γ_L be the Lipschitz constant of $F'(x)$ in L, and χ_D and ξ_D be such that $\|D_k\| < \chi_D$ and $\|D_k^{-1}\| < \xi_D$ for $k \geqslant 0$. First, with a slight modification of Theorems 4.7–4.8 of [24], by (A.3) $\lim_{k\to\infty}\|D_k^{-1}\nabla f(x_k)\| = 0$ follows. The modifications are needed in order to deal with elliptical trust-regions. Then by $\|D_k^{-1}\nabla f(x_k)\| \geqslant \|\nabla f(x_k)\|/\chi_D$, we get (3.8) and since $\nabla f(x^*) = F'(x^*)^{\mathrm{T}}F(x^*)$ and $F'(x^*)$ is invertible by hypothesis, this yields $F(x^*) = 0$. Moreover, as the sequence $\{\|F(x_k)\|\}$ is monotone decreasing and bounded below, it follows $\lim_{k\to\infty}\|F(x_k)\| = 0$.

Second, by using Theorem 6.1 we get $x_k \to x^*$.

Assume that k is sufficiently large that $F'(x_k)$ is invertible and $\|F'(x_k)^{-1}\| \leqslant \xi_F$ for some $\xi_F > 0$. This is ensured by the invertibility of $F'(x^*)$. Further, let $\bar{s}_k$ be the inexact Newton step computed in Step 1 of the ESTR method, $\tilde{\Delta}_k$ be the initial trust-region radius at the first iteration of the repeat-loop in Step 3 and s_k be the solution of the elliptical trust-region subproblem (3.5) with $\Delta_k = \tilde{\Delta}_k$. Now, we will show that for k sufficiently large the step s_k is accepted.

From Theorem 11.1 in [24] it follows that

$$\|F(x_k + s_k)\|^2 = \|F(x_k) + F'(x_k)s_k + w(x_k, s_k)\|^2,$$

where

$$w(x_k, s_k) = \int_0^1 (F'(x_k + \xi s_k) - F'(x_k)) s_k \, \mathrm{d}\xi.$$

From the Lipschitz continuity of F' it follows that $\|w(x_k, s_k)\| \leqslant \frac{\gamma_L}{2}\|s_k\|^2$. Then, as s_k satisfies (3.5) we get

$$\|F(x_k) + F'(x_k)s_k\| \leqslant \|F(x_k)\| \tag{A.6}$$

and

$$\begin{aligned} |\|F(x_k) + F'(x_k)s_k\|^2 - \|F(x_k + s_k)\|^2| &\leqslant 2\|F(x_k) + F'(x_k)s_k\|\|w(x_k, s_k)\| + \|w(x_k, s_k)\|^2 \\ &\leqslant \|F(x_k)\|\gamma_L\|s_k\|^2 + (\gamma_L/2)^2\|s_k\|^4 \end{aligned} \tag{A.7}$$

To prove the thesis we need to show that s_k satisfies condition (3.7). First, note that from (3.1) it follows

$$\|\bar{s}_k\| \leqslant \|(F'(x_k))^{-1}\|(\|F(x_k)\| + \|r_k\|) \leqslant 2\xi_F\|F(x_k)\|. \tag{A.8}$$

Then, as $\tilde{\Delta}_k \geqslant \Delta_{\min}$, $\lim_{k\to\infty}\|F(x_k)\| = 0$ and $\|D_k\| < \chi_D$, it follows

$$\|D_k\bar{s}_k\| \leqslant \tilde{\Delta}_k,$$

for k sufficiently large, i.e. eventually the inexact Newton step belongs to the trust-region with $\Delta_k = \tilde{\Delta}_k$. Consider the denominator of (3.7) and assume $\|D_k\bar{s}_k\| \leqslant \tilde{\Delta}_k$. Then, as s_k is the solution of (3.5) for $\Delta_k = \tilde{\Delta}_k$, we have:

$$\begin{aligned} \|F(x_k)\|^2 - \|F(x_k) + F'(x_k)s_k\|^2 &\geqslant \|F(x_k)\|^2 - \|F(x_k) + F'(x_k)\bar{s}_k\|^2 \geqslant \|F(x_k)\|^2 - \bar{\eta}_k^2\|F(x_k)\|^2 \\ &\geqslant (1 - \eta_{\max}^2)\|F(x_k)\|^2. \end{aligned} \tag{A.9}$$

Moreover, taking into account (A.6) we get

$$\|s_k\| \leqslant 2\|(F'(x_k))^{-1}\|\|F(x_k)\| \leqslant 2\xi_F\|F(x_k)\| \tag{A.10}$$

and this along with (A.9) yields

$$\|F(x_k)\|^2 - \|F(x_k) + F'(x_k)s_k\|^2 \geqslant \frac{(1 - \eta_{\max}^2)}{4\xi_F^2}\|s_k\|^2. \tag{A.11}$$

Then, by (3.7), (A.7), and (A.11) we get

$$|\rho_k^f(s_k) - 1| = \left|\frac{m_k(s_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)}\right| \leqslant \frac{(\|F(x_k)\|\gamma_L + (\gamma_L/2)^2\|s_k\|^2)\|s_k\|^2}{\frac{1}{4\xi_F^2}(1 - \eta_{\max}^2)\|s_k\|^2}$$

Moreover, as from (A.10) it follows that $\lim_{k\to\infty}\|s_k\| = 0$, we get

$$\lim_{k\to\infty}\|F(x_k)\|\gamma_L + (\gamma_L/2)^2\|s_k\|^2 = 0.$$

Then, for k sufficiently large, we have $\rho_k^f(s_k) \geqslant \beta_1$ and the step s_k is accepted. Moreover, as the inexact Newton step eventually is feasible and $\|F'(x_k)s_k + F(x_k)\| \leqslant \|F'(x_k)\bar{s}_k + F(x_k)\|$ it follows that s_k satisfies (3.9) for k sufficiently large.

Now, we turn our attention to the convergence properties of the HIN method. First, we observe that if there are is finite number of steps generated by the ESTR technique, i.e. if eventually the steps are generated by the INB strategy, the convergence behaviour of the HIN method is given by Theorem 3.1. Then, in the following lemma we focus on the case where an infinite number of steps generated by the ESTR technique are taken. We omit the proof of this result as it is a slight modification of Theorems 4.7–4.8 of [24].

Lemma 6.2. *Let $r > 0$ and $L = \cup_{k=0}^{\infty}\{x \in \mathbb{R}^n | \|x - x_k\| \leqslant r\}$ be a neighborhood of the sequence $\{x_k\}$ generated by the HIN method. Assume that F' is Lipschitz continuous in L and $\|F'(x)\|$ is bounded above on L. Suppose furthermore that there exists a limit point x^* of $\{x_k\}$, $\nabla f(x_k) \neq 0$ for $k \geqslant 0$, the sequences $\{\|D_k\|\}$ and $\{\|D_k^{-1}\|\}$ are bounded above. If there exists a subsequence $\{x_{k_j}\} \to x^*$ such that eventually s_{k_j} is computed in Step 3, then*

$$\lim_{k_j\to\infty} \|\nabla f(x_{k_j})\| = 0.$$

Now we are ready to establish the main convergence theorem of the HIN method.

Proof of Theorem 3.3. Let $\{x_{k_j}\}$ be a subsequence such that $\{x_{k_j}\} \to x^*$ and K_{INB} be the set of indices such that s_{k_j} is computed in Step 2 for any $k_j \in K_{\mathrm{INB}}$. The sequence $\{\|F(x_k)\|\}$ is monotone decreasing and bounded from below, then it is convergent. Hence, $\lim_{k\to\infty}\|F(x_k)\| = \|F(x^*)\|$.

First, assume that K_{INB} contains an infinite number of indices. Since in Step 2 at most $MAXBT$ backtracks are performed, it follows that

$$1 - \eta_{k_j} \geqslant \theta^{MAXBT}(1 - \bar{\eta}_{k_j}) \geqslant \theta^{MAXBT}(1 - \eta_{\max}) \quad \text{for any } k_j \in K_{\mathrm{INB}}.$$

Therefore, the series $\sum_{k_j\in K_{\mathrm{INB}}}(1-\eta_{k_j})$ is divergent and this implies the divergence of the series $\sum_{k=0}^{\infty}(1-\eta_k)$. Then from Theorem 6.1, Part a, we get $\lim_{k\to\infty}\|F(x_k)\| = \|F(x^*)\| = 0$.

On the other hand, if K_{INB} is finite, s_{k_j} is computed in Step 3 for k_j sufficiently large. Then, by Lemma 6.2 we obtain

$$\lim_{k\to\infty} \nabla f(x_{k_j}) = \nabla f(x^*) = 0.$$

Since $\nabla f(x^*) = F'(x^*)^{\mathrm{T}}F(x^*)$ and $F'(x^*)$ is invertible by hypothesis, we get $F(x^*) = 0$ and Part a follows.

Part b follows from Theorem 6.1, Part b and Part c derives from the convergence result of Algorithm INB given in Theorem 3.1.

References

[1] H.B. An, Z.Z. Bai, A globally convergent Newton-GMRES method for large sparse systems of nonlinear equations, Appl. Numer. Math. 57 (2007) 235–252.

[2] S. Balay, K Buschelman, V Eijkhout, W.D. Gropp, D. Kaushik, M.G. Knepley, L. Curfman McInnes, B.F. Smith, H. Zhang, PETSc Users Manual, ANL-95/11 – Revision 2.1.5, Argonne National Laboratory, http://www-unix.mcs.anl.gov/petsc, 2004.

[3] S. Bellavia, B. Morini, Newton-GMRES subspace method for systems of nonlinear equations, SIAM J. Sci. Comput. 23 (2003) 940–960.

[4] S. Bellavia, B. Morini, Subspace trust-region methods for large bound constrained nonlinear equations, SIAM J. Numer. Anal. 44 (2006) 1535–1555.

[5] M.A. Branch, T.F. Coleman, Y. Li, A subspace, interior, and conjugate gradient method for large scale bound-constrained minimization problems, SIAM J. Sci. Comput. 21 (1999) 1–23.

[6] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, Springer-Verlag, New York/Berlin, 1991.

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

[8] R.H. Byrd, R.B. Schnabel, M.H. Shultz, Approximate solution of the trust-region problem by minimization over two-dimensional subspaces, Math. Program. 40 (1988) 247–263.

[9] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland Publishing Company, Amsterdam, 1978.

[10] A.R. Conn, N.I.M. Gould, P.L. Toint, Trust Region Methods, Springer Series in Optimization, 2000.

[11] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, SIAM J. Numer. Anal. 19 (1982) 400–408.

[12] J.E. Dennis, D.M. Gay, R.E. Welsh, An adaptive nonlinear least squares algorithm, ACM Trans. Math. Soft. 7 (1981) 348–368.

[13] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, Math. Program. 91 (2 Ser. A) (2002) 201–213.

[14] R.N. Elias, A.L.G.A. Coutinho, M.A.D. Martins, Inexact-Newton-Type methods for non-linear problems arising from the SUPG/PSPG solution of steady incompressible Navier–Stokes Equations, J. Braz. Soc. Mech. Sci. & Eng. 26 (2004) 330–339.

[15] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, SIAM J. Optim. 4 (1994) 393–422.

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

[17] L.P. Franca, S.L. Frey, Stabilized finite element methods: II. The incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Eng. 99 (1992) 209–233.

[18] U. Ghia, K.N. Ghia, C.T. Shin, High-*Re* solutions for incompressible flow using the Navier–Stokes equations and a multigrid method, J. Comput. Phys. 48 (1982) 387–411.

[19] V. Girault, P.A. Raviart, Finite Element Methods for Navier–Stokes Equations, Theory and Algorithms, Springer-Verlag, Berlin/Heidelberg, 1986.

[20] I. Harari, T.J.R. Hughes, What are *C* and *h*?: Inequalities for the analysis and design of finite element methods, Comput. Methods Appl. Mech. Eng. 97 (1992) 157–192.

[21] T.J.R. Hughes, L.P. Franca, M. Balestra, A new finite element formulation for computational fluid dynamics: V. Circumventing the Babuška—Brezzi condition: a stable Petrov–Galerkin formulation of the Stokes problem accommodating equal-order interpolations, Comput. Methods Appl. Mech. Eng. 59 (1986) 85–99.

[22] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.
[23] A GAMM workshop, in: K. Morgan, J. Periaux, F. Thomasset (Eds.), Analysis of Laminar Flow over a Backward Facing Step, Friedr. Vieweg & Sohn, Braunschweig, 1984.
[24] J. Nocedal, S.J. Wright, Numerical Optimization, Springer Series in Operations Research, 1999.
[25] 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.
[26] P. Pawlowski, J.P. Simonis, H.F. Walker, J.N. Shadid, Inexact Newton Dogleg methods, WPI Math. Sciences Dept. Tech. Rep. MS-5-05-36, SIAM J. Numer. Anal., submitted for publication.
[27] M. Pernice, H.F. Walker, NITSOL: a new iterative solver for nonlinear systems, SIAM J. Sci. Comput. 19 (1998) 302–318.
[28] Y. Saad, M.H. Schultz, GMRES: a generalized minimal residual method for solving nonsymmetric linear systems, SIAM J. Sci. Stat. Comput. 6 (1985) 856–869.
[29] J.N. Shadid, R.S. Tuminaro, H.F. Walker, An inexact newton method for fully coupled solution of the Navier–Stokes equations with heat and mass transport, J. Comput. Phys. 137 (1997) 155–185.
[30] R.S. Tuminaro, H.F. Walker, J.N. Shadid, On backtracking failure in Newton–GMRES methods, J. Comput. Phys. 180 (2002) 549–558.