Comput Optim Appl
DOI 10.1007/s10589-009-9252-7

Convergence analysis of truncated incomplete Hessian Newton minimization method and application in biomolecular potential energy minimization

Dexuan Xie · Mazen G. Zarrouk

Received: 31 March 2008

Abstract This paper gives a general convergence analysis to the truncated incomplete Hessian Newton method (T-IHN). It shows that T-IHN is globally convergent even with an indefinite incomplete Hessian matrix or an indefinite preconditioner, which may happen in practice. It also proves that when the T-IHN iterates are close enough to a minimum point, T-IHN has a Q-linear rate of convergence, and an admissible line search steplength of one. Moreover, a particular T-IHN algorithm is constructed for minimizing a biomolecular potential energy function, and numerically tested for a protein model problem based on a widely used molecular simulation package, CHARMM. Numerical results confirm the theoretical results, and demonstrate that T-IHN can have a better performance (in terms of computer CPU time) than most CHARMM minimizers.

Keywords Truncated-Newton method · Incomplete Hessian Newton method · Convergence analysis · Molecular potential minimization

1 Introduction

A practical and efficient modified Newton minimization method, called the truncated incomplete Hessian Newton method (T-IHN), was recently proposed in [34] to solve the following local minimization problem:

$$f(x^*) \le f(x) \quad \text{for all } x \in \mathcal{D} \subset R^n, \tag{1}$$

This project was partially supported by the National Science Foundation (DMS-0241236, USA) and the National Institutes of Health (PHS R01 EB005825-01, USA).

D. Xie (✉) · M.G. Zarrouk
Department of Mathematical Sciences, University of Wisconsin, Milwaukee, WI 53211, USA
e-mail: dxie@uwm.edu

M.G. Zarrouk
e-mail: mzarrouk@uwm.edu

Published online: 02 May 2009

where $\mathcal{D}$ is an open domain containing only one local minimum point x^*, f is twice continuously differentiable on $\mathcal{D}$, and the Hessian matrix $H(x)$ (i.e., the second derivative of f at $x \in \mathcal{D}$) is dense but can be well approximated by a sparse incomplete Hessian matrix, $M(x)$. In [34], the incomplete Hessian Newton method (IHN) was proposed and analyzed under the assumption that $M(x)$ is symmetric and positive definite (SPD). In practice, however, $M(x)$ may be indefinite when x is far away from x^*. To develop a practical IHN-type scheme, IHN was modified as T-IHN by using the truncation strategies similar to those suggested in [29], resulting in a well defined descent search method even with an indefinite incomplete Hessian matrix or an indefinite preconditioner. T-IHN was well studied numerically in [34] for a chemical database problem described in [23, 31, 35], which is one typical example of problem (1). But, its convergence analysis was not given.

Another typical example of (1), as mentioned in [34], is the minimization problem of biomolecular potential energy function; solving it is a fundamental and challenging task in biomolecular simulations [15, 18, 23]. In this application, $f(x)$ is a sum of bonded and nonbonded potential energy terms, and x is a collective vector of all the atomic coordinates of a biomolecular system; each value of x determines a conformation of the biomolecular system. Mathematically, the energy function $f(x)$ is twice continuously differentiable in some domain but strongly nonlinear with many local minimum points. Evaluating the function and its gradient vector is extremely expensive, and its Hessian matrix is typically dense. Hence, minimization of the energy function is a formidable geometry optimization problem [22]. So far, several numerical optimization algorithms were adapted to some widely-used biomolecular simulation packages such as CHARMM [3], AMBER [5], NAMD [17], and GROMACS [12, 26] for minimizing the biomolecular energy function. For example, CHARMM contains the steepest descent method, the nonlinear conjugate gradient method (CONJ) [9], the modified Newton-Raphson method, the adopted basis Newton-Raphson method (ABNR) [4], and the truncated Newton method (TN) [6, 8, 29].

In this paper, we intend to give T-IHN a general convergence analysis, and develop a particular T-IHN algorithm for minimizing the biomolecular potential energy function. In particular, we prove that T-IHN is globally convergent even with an indefinite incomplete Hessian matrix or an indefinite preconditioner. As the T-IHN iterates are sufficiently close to the minimum point, T-IHN is proved to have a Q-linear rate of convergence, and an admissible line search steplength of one.

In the construction of the particular T-IHN algorithm, we observed that most of the nonzero entries of $H(x)$ are only related to the nonbonded potential energy terms, which can be found in the form $O(1/r_{ij}^3)$, where r_{ij} denotes the distance between two atoms i and j with $i \neq j$. Thus, most entries of $H(x)$ can be simply zeroed out when r_{ij} is large enough, resulting in a sparse incomplete Hessian matrix, $M(x)$, a good approximation to the original $H(x)$. Based on this observation, we constructed an effective and efficient T-IHN algorithm for minimizing the biomolecular potential energy function by using a simple distance cutoff strategy. We then programmed it in FORTRAN based on CHARMM [8, 25, 30, 32], and numerically tested it based on a simple model problem defined by a protein, BPTI (a bovine pancreatic trypsin inhibitor with 586 atoms). We also made numerical tests to compare the performances

of T-IHN with that of three widely-used CHARMM minimizers: CONJ, ABNR, and TN. Numerical results confirm our T-IHN analytical results, and demonstrate the great potential of T-IHN in minimizing the biomolecular potential energy function. About 60% and 95% total CPU times were saved by T-IHN compared to ABNR and CONJ, respectively, for this protein model problem.

The remainder of this paper is organized as follows. We define T-IHN in Sect. 2, and give it a general convergence analysis in Sect. 3. In Sect. 4, the T-IHN for minimizing the biomolecular potential energy function is constructed, and the numerical results for the protein model problem are presented. Finally, conclusions are made in Sect. 5.

2 The T-IHN method

Let $f(x_k)$, $g(x_k)$, $H(x_k)$, $M(x_k)$ and $B(x_k)$ denote the values of function, gradient vector, Hessian matrix, incomplete Hessian matrix, and preconditioner at the kth iterate x_k, respectively. When there is no ambiguity, they are simply denoted by f_k, g_k, H_k, M_k, and B_k. Similarly, H_* and M_* denote their values at the minimum point x^*. Throughout this paper, B_k is assumed to be at least invertible for $k = 0, 1, 2, \ldots$, $\|\cdot\|$ represents the standard Euclidean norm of a vector or a matrix, and the superscript T of a vector or a matrix denotes its transpose. For a given initial guess, x_0, a sequence of the T-IHN iterates, $\{x_k\}$, is defined by

$$x_{k+1} = x_k + \lambda_k p_k, \quad k = 0, 1, 2, \ldots, \tag{2}$$

where p_k is a descent search direction generated by using the truncated Newton strategies described in [30] for solving the incomplete Hessian equation

$$M_k p_k = -g_k, \tag{3}$$

and λ_k is a steplength determined by a line search algorithm (e.g., the one given in [16]), which satisfies the Wolfe's conditions

$$f(x_k + \lambda_k p_k) \le f(x_k) + \alpha \lambda_k g(x_k)^T p_k, \tag{4a}$$

$$g(x_k + \lambda_k p_k)^T p_k \ge \beta g(x_k)^T p_k, \tag{4b}$$

for some $\alpha \in (0, \frac{1}{2})$ and $\beta \in (\frac{1}{2}, 1)$. The details on the construction of the search direction p_k of T-IHN are presented in Algorithm 1 for clarity.

Algorithm 1 (Search direction p_k of T-IHN) Let $p_k^{(j)}$ be the jth iterate of the preconditioned conjugate gradient method (PCG) [11] for solving (3) at the kth T-IHN iteration, B_k denote a preconditioner of PCG, $IT_{\max}$ be the maximum number of PCG iterations (e.g., $IT_{\max} = 80$), and δ the tolerance of the singularity test (e.g., $\delta = 10^{-10}$). Set $\eta_k = \min\{c_r/k, \|g_k\|\}$ (e.g., $c_r = 0.5$). The kth descent search direction p_k of T-IHN is produced by the following steps.

1. [INITIALIZATION]
   Set $j = 1$, $p_k^{(1)} = 0$, $r_k^{(1)} = -g_k$, and $d_k^{(1)} = z_k^{(1)}$, where $z_k^{(1)}$ is the solution of the linear system $B_k z_k^{(1)} = r_k^{(1)}$.
2. [SINGULARITY TEST]
   If either $|(r_k^{(j)})^T z_k^{(j)}| \le \delta \|g_k\| \|d_k^{(1)}\|$ or $|(d_k^{(j)})^T M_k d_k^{(j)}| \le \delta \|d_k^{(j)}\|^2$,
   exit the algorithm with $p_k = p_k^{(j)}$ (for $j = 1$, set $p_k = -g_k$).
3. Compute $p_k^{(j+1)} = p_k^{(j)} + \alpha_k^{(j)} d_k^{(j)}$, where $\alpha_k^{(j)} = \frac{(r_k^{(j)})^T z_k^{(j)}}{(d_k^{(j)})^T M_k d_k^{(j)}}$.
4. [DESCENT DIRECTION TEST]
   If $g_k^T p_k^{(j+1)} \ge g_k^T p_k^{(j)}$,
   exit the algorithm with $p_k = p_k^{(j)}$ (for $j = 1$, set $p_k = -g_k$).
5. Compute $r_k^{(j+1)} = r_k^{(j)} - \alpha_k^{(j)} M_k d_k^{(j)}$.
6. [TRUNCATION TEST]
   If $\|r_k^{(j+1)}\| \le \eta_k \|g_k\|$ or $j + 1 \ge IT_{\max}$
   exit the algorithm with $p_k = p_k^{(j+1)}$.
7. Compute $d_k^{(j+1)} = z_k^{(j+1)} + \beta_k^{(j)} d_k^{(j)}$, where $\beta_k^{(j)} = \frac{(r_k^{(j+1)})^T z_k^{(j+1)}}{(r_k^{(j)})^T z_k^{(j)}}$, and $z_k^{(j+1)}$ is the solution of the linear system $B_k z_k^{(j+1)} = r_k^{(j+1)}$.
8. Increase j to $j + 1$ and go to Step 2.

The preconditioner B_k is usually an SPD matrix that well approximates M_k in some sense. But, in practice, it may be indefinite. Thus, a modified Cholesky factorization method is required to solve the preconditioning equation $B_k z_k^{(j)} = r_k^{(j)}$. In our implementation of T-IHN, we use the unconventional modified Cholesky (UMC) factorization method proposed in [29]. Compared to the traditional modified Cholesky method (see [10], for example), UMC allows the modified B_k to remain to be indefinite to avoid a too large modification when B_k is strongly indefinite. See [29] for the details on UMC.

The case that PCG fails at its first iteration may happen only if B_k or M_k is indefinite. In our experiences, such an extreme case happened rarely. Instead, since $B(x^*)$ and $M(x^*)$ are usually SPD, the continuity of $B(x)$ and $M(x)$ implies that both M_k and B_k are SPD. Thus, the search direction p_k is mostly set as $p_k^{(j)}$ with j determined by the truncation test as x_k is sufficiently close to x^*.

3 T-IHN convergence analysis

We first show that T-IHN is convergent for any initial guess $x_0 \in \mathcal{D}$ even with an indefinite M_k or B_k, which may occur in practice when the T-IHN iterate x_k is far away from the minimum point x^*.

Theorem 1 *Let $\{x_k\}$ be a sequence of T-IHN iterates for solving the local minimization problem (1). For any initial guess $x_0 \in \mathcal{D}$, T-IHN is a descent search method. Furthermore, if $\mathcal{D}$ contains the level set $\{x | f(x) \le f(x_0)\}$, and the Hessian H is*

bounded above, then the T-IHN iterative sequence $\{x_k\}$ *converges in the sense*

$$\lim_{k\to\infty} \|g(x_k)\| = 0. \tag{5}$$

Proof We first prove that T-IHN is a descent method. In T-IHN, the kth search direction p_k is set as $p_k = -g_k$ when PCG fails at its first iteration, which is clearly a descent search direction; otherwise, $p_k = p_k^{(l)}$ with $l \geq 2$. For all $l \geq 2$, the descent direction test of Algorithm 1 guarantees that

$$g_k^T p_k^{(l)} < g_k^T p_k^{(l-1)} < \cdots < g_k^T p_k^{(2)} < g_k^T p_k^{(1)} = 0, \tag{6}$$

where $p_k^{(1)} = 0$. This shows that $p_k^{(l)}$ is descent. Thus, each search direction p_k of T-IHN is descent, and then a positive steplength λ_k that satisfies the Wolfe's conditions (4) exists (see [7], for example). Hence, T-IHN is a well-defined descent search method.

We next prove the convergence (5). Since $\mathcal{D}$ contains the level set, all T-IHN iterates x_k are in $\mathcal{D}$. Thus, with (4a), we can get that

$$f(x^*) \leq f(x_k) \leq f(x_0) \quad \text{for all } k \geq 1, \tag{7}$$

where $f(x^*)$ is the minimum value of f on $\mathcal{D}$. From that H is bounded above it follows that there exists a constant $\mathcal{M} > 0$ such that

$$\begin{aligned} &\|H(x)\| \leq \mathcal{M}, \qquad \|M(x)\| \leq \mathcal{M}, \quad \text{and} \\ &\|g(x) - g(y)\| \leq \mathcal{M}\|x - y\|, \quad \forall x \in \mathcal{D}. \end{aligned} \tag{8}$$

The third inequality of (8) defines the Lipschitz continuity of g. With the Wolfe's conditions (4), the Lipschitz continuity of g, and (7), we can use the same arguments as the ones used in the proof of Theorem 3.2 in [19] to get

$$\sum_{k=0}^{j} \cos^2\theta_k \|g(x_k)\|^2 \leq \frac{\mathcal{M}[f(x_0) - f(x_{j+1})]}{\alpha(1-\beta)} \leq \frac{\mathcal{M}[f(x_0) - f(x^*)]}{\alpha(1-\beta)} < \infty,$$

which immediately implies that

$$\lim_{k\to\infty} \cos^2\theta_k \|g(x_k)\|^2 = 0,$$

where $\cos\theta_k = -g_k^T p_k/(\|g_k\|\|p_k\|)$. We then can obtain $\cos\theta_k \geq c_1/c_2$ for all k and complete the proof by

$$\|g(x_k)\| \leq \frac{c_2}{c_1}\cos\theta_k \|g(x_k)\| \to 0 \quad \text{as } k \to \infty$$

provided that there exist two positive constants c_1 and c_2 (independent of k) such that

$$\text{(a)} \quad -g_k^T p_k \geq c_1\|g_k\|^2, \qquad \text{(b)} \quad \|p_k\| \leq c_2\|g_k\|, \quad \forall k \geq 0. \tag{9}$$

Hence, the remaining part of the proof is to verify the above two inequalities, which is presented in the following.

Proof of (9a) For $j = 1, 2, \ldots, l$ with $l \geq 2$, by some arguments used in the proof of Lemma 1, it is easy to obtain

$$g_k^T p_k^{(j+1)} - g_k^T p_k^{(j)} = -\frac{[(r_k^{(j)})^T B_k^{-1} r_k^{(j)}]^2}{(d_k^{(j)})^T M_k d_k^{(j)}}, \tag{10}$$

from which it immediately follows that

$$g_k^T p_k^{(j+1)} < g_k^T p_k^{(j)} \quad \text{is equivalent to } (d_k^{(j)})^T M_k d_k^{(j)} > 0. \tag{11}$$

Because of the singularity test, we actually obtain that

$$(d_k^{(j)})^T M_k d_k^{(j)} > \delta \|d_k^{(j)}\|^2, \quad \text{or} \quad \frac{\|d_k^{(j)}\|^2}{(d_k^{(j)})^T M_k d_k^{(j)}} < \frac{1}{\delta}, \tag{12}$$

for $j = 1, 2, \ldots, l$ with $l \geq 2$. By (6), (10), and the singularity test, we get that

$$g_k^T p_k \leq g_k^T p_k^{(2)}, \qquad g_k^T p_k^{(2)} = -\frac{[(r_k^{(1)})^T z_k^{(1)}]^2}{(d_k^{(1)})^T M_k d_k^{(1)}}, \quad \text{and}$$

$$|(r_k^{(1)})^T z_k^{(1)}| > \delta \|g_k\| \|d_k^{(1)}\|.$$

From the second inequality of (8) it follows that $(d_k^{(1)})^T M_k d_k^{(1)} \leq \mathcal{M} \|d_k^{(1)}\|^2$. Hence, for $p_k = p_k^{(l)}$ with $l \geq 2$,

$$-g_k^T p_k \geq -g_k^T p_k^{(2)} = \frac{[(r_k^{(1)})^T z_k^{(1)}]^2}{(d_k^{(1)})^T M_k d_k^{(1)}} \geq \frac{\delta^2}{\mathcal{M}} \|g_k\|^2.$$

When PCG fails at its first iteration, $p_k = -g_k$, and thus $-g_k^T p_k = \|g_k\|^2$. Therefore, setting $c_1 = \min\{1, \frac{\delta^2}{\mathcal{M}}\}$ completes the proof of (9a).

Proof of (9b) For $l \geq 2$, the PCG iterate $p_k^{(l)}$ can be written as

$$p_k^{(l)} = \sum_{j=1}^{l-1} \alpha_k^{(j)} d_k^{(j)} = \sum_{j=1}^{l-1} \frac{(r_k^{(1)})^T d_k^{(j)}}{(d_k^{(j)})^T M_k d_k^{(j)}} d_k^{(j)} = \sum_{j=1}^{l-1} \frac{d_k^{(j)} (d_k^{(j)})^T}{(d_k^{(j)})^T M_k d_k^{(j)}} r_k^{(1)},$$

where we have used the identity $(a^T b)b = bb^T a$ for vectors a and b, and the identity $(r_k^{(j)})^T B_k^{-1} r_k^{(j)} = (r_k^{(1)})^T d_k^{(j)}$, which follows from (15) and (16). Hence, for $p_k = p_k^{(l)}$ with $l \geq 2$, by $r_k^{(1)} = -g_k$ and (12),

$$\|p_k\| \leq \left[\sum_{j=1}^{l-1} \frac{\|d_k^{(j)}\|^2}{(d_k^{(j)})^T M_k d_k^{(j)}} \right] \|r_k^{(1)}\| \leq \frac{l-1}{\delta} \|g_k\| \leq \frac{IT_{\max}}{\delta} \|g_k\|.$$

If $p_k = -g_k$, then $\|p_k\| = \|g_k\|$. Thus, setting $c_2 = \max\{1, \frac{IT_{\max}}{\delta}\}$ gives (9a). This completes the proof of this theorem. □

Remark 1 Similar to the proof of Theorem 1, we also can prove that T-IHN is globally convergent in the traditional sense: for any $x_0 \in R^n$, the iterative sequence $\{x_k\}$ satisfies $\lim_{k\to\infty} \|g(x_k)\| = 0$ provided that f is bounded below in R^n, twice continuously differentiable in an open set $\mathcal{N}$ that contains the level set $\{x | f(x) \le f(x_0)\}$, and the Hessian H is bounded above on $\mathcal{N}$.

Remark 2 Obviously, Theorem 1 holds with $\mathcal{D} = R^n$ if f is a convex function on R^n. When f has many local minimizers, we usually consider a local minimization problem of (1) to avoid the difficulty of finding a global minimizer of f. In practice, an open domain of $\mathcal{D}$ can be selected based on our knowledge of f.

One key fact used in the analysis of IHN [34] is that $p_k^T r_k = 0$, which holds naturally since the search direction p_k of IHN is the exact solution of (3), i.e., the residual $r_k = -g_k - M_k p_k = 0$. In T-IHN, however, the residual r_k is usually nonzero since the T-IHN search direction p_k is only an approximate solution of (3). Even so, this key fact still holds in T-IHN according to the following lemma.

Lemma 1 *Let $p_k^{(l)}$ denote the lth PCG iterate for solving the linear system (3). If the initial guess $p_k^{(1)} = 0$, then for $l \ge 2$,*

$$(p_k^{(l)})^T (g_k + M_k p_k^{(l)}) = 0. \tag{13}$$

Proof According to the PCG definition of $p_k^{(l)}$ as given in Algorithm 1, one can easily check that

$$p_k^{(l)} = p_k^{(l-1)} + \alpha_k^{(l-1)} d_k^{(l-1)} = p_k^{(1)} + \sum_{j=1}^{l-1} \alpha_k^{(j)} d_k^{(j)} = \sum_{j=1}^{l-1} \alpha_k^{(j)} d_k^{(j)}, \tag{14}$$

where $p_k^{(1)} = 0$ has been used. Further, $d_k^{(j)}$ can be expressed as

$$d_k^{(j)} = z_k^{(j)} + \beta_k^{(j-1)} d_k^{(j-1)} = \sum_{i=1}^{j} \gamma_{ij} B_k^{-1} r_k^{(i)}, \tag{15}$$

where $\gamma_{jj} = 1$ and $\gamma_{ij} = \prod_{m=i}^{j-1} \beta_k^{(m)}$ for $i = 1, 2, \ldots, j-1$, which can be simplified as $\gamma_{ij} = \frac{(r_k^{(j)})^T z_k^{(j)}}{(r_k^{(i)})^T z_k^{(i)}}$ by using $\beta_k^{(m)} = \frac{(r_k^{(m+1)})^T z_k^{(m+1)}}{(r_k^{(m)})^T z_k^{(m)}}$.

For $j = 1, 2, \ldots, l$, we have that $(r_k^{(j)})^T B_k^{-1} r_k^{(j)} = (r_k^{(j)})^T z_k^{(j)} \ne 0$ and $(d_k^{(j)})^T M_k d_k^{(j)} \ne 0$. Thus, similar to what is done in the PCG theory (see [11], for example), we can obtain that the residuals $r_k^{(j)} = -g_k - M_k p_k^{(j)}$ satisfy

$$(r_k^{(i)})^T B_k^{-1} r_k^{(j)} = 0, \quad \forall i \ne j. \tag{16}$$

Hence, by (14), (15), and (16),

$$
\begin{aligned}
(p_k^{(l)})^T (g_k + M_k p_k^{(l)}) &= -(r_k^{(l)})^T p_k^{(l)} = -\sum_{j=1}^{l-1} \alpha_k^{(j)} (r_k^{(l)})^T d_k^{(j)} \\
&= -\sum_{j=1}^{l-1} \alpha_k^{(j)} (r_k^{(l)})^T \left(\sum_{i=1}^{j} \gamma_{ij} B_k^{-1} r_k^{(i)} \right) \\
&= -\sum_{j=1}^{l-1} \alpha_k^{(j)} \left(\sum_{i=1}^{j} \gamma_{ij} \underbrace{(r_k^{(l)})^T B_k^{-1} r_k^{(i)}}_{\equiv 0} \right) = 0.
\end{aligned}
$$

This completes the proof of Lemma 1. □

We assume the following *basic assumptions* for the remaining analysis.

Assumption 1 (Basic Assumptions)

1. The T-IHN sequence $\{x_k\}$ converges to $x^* \in \mathcal{D}$ for which $g(x^*) = 0$.
2. $H(x)$, $M(x)$ and $B(x)$ are SPD at $x = x^*$.

It is easy to obtain the following lemma (see e.g., [27, p. 52], and [20, p. 46]).

Lemma 2 *Let $A(x)$ be an $n \times n$ real matrix function defined on $\mathcal{D} \subset R^n$. If $A(x)$ is continuous and SPD at some point $x^* \in \mathcal{D}$, then $A(x)^{-1}$ is continuous and SPD at x^* too. Moreover, there exists $\delta > 0$ such that $A(x)$ and $A(x)^{-1}$ are SPD for all $x \in \mathcal{B}(\delta)$, where $\mathcal{B}(\delta) = \{x \in \mathcal{D} | \|x - x^*\| < \delta\}$.*

In practice, the incomplete Hessian $M(x)$ or preconditioner $B(x)$ may be indefinite when x is far away from the minimum point x^*. To deal with these difficult situations, the singularity and descent direction tests are introduced in Algorithm 1, which enhance the robustness and numerical stability of T-IHN. However, when x is sufficiently close to x^*, the following lemma shows that T-IHN can be well defined without them.

Lemma 3 *Let $\{x_k\}$ be a T-IHN sequence. If Assumption 1 holds, then there exists an index k_0 such that for $k \geq k_0$, the search direction p_k of T-IHN is a PCG iterate determined by only involving the truncation test in Algorithm 1.*

Proof By Assumption 1 and Lemma 2, there exists an index k_0 such that B_k and M_k are SPD for all $k \geq k_0$. In particular, for $x \neq 0$, $x^T B_k x$ and $x^T M_k x$ are (strictly) positive. Therefore, the singularity test becomes inactive (theoretically $\delta = 0$ in Algorithm 1). Moreover, (10) further shows that $g_k^T (p_k^{(j+1)} - p_k^{(j)}) < 0$ from which it follows that all the PCG iterates are descent. Hence, Algorithm 1 terminates only in the truncation test, and so for each $k \geq k_0$, the search direction p_k of T-IHN must be set as a PCG iterate, $p_k^{(l)}$ with some $l \geq 2$. This completes the proof of this lemma. □

Next, we give the following two lemmas, which will be used in the proof of our next theorem.

Lemma 4 *If H and M are $n \times n$ SPD real matrices, then the eigenvalues of $M^{-1}H$ are all positive, and its minimum eigenvalue $\lambda_{\min}$ and maximum eigenvalue $\lambda_{\max}$ satisfy*

$$\lambda_{\min} p^T M p \le p^T H p \le \lambda_{\max} p^T M p, \quad \forall p \in R^n. \tag{17}$$

Proof Let $M^{\frac{1}{2}}$ be the unique square root matrix of M, which is SPD, and set $\overline{H} = M^{-\frac{1}{2}} H M^{-\frac{1}{2}}$. Clearly, $\overline{H}$ is SPD. Moreover, Since $\overline{H}$ can be written as $\overline{H} = M^{\frac{1}{2}}(M^{-1}H)M^{-\frac{1}{2}}$, $\overline{H}$ is similar to $M^{-1}H$. Hence, $M^{-1}H$ has the same eigenvalues of $\overline{H}$, which are positive, and

$$\lambda_{\min} y^T y \le y^T \overline{H} y \le \lambda_{\max} y^T y, \quad \forall y \in R^n.$$

Selecting $y = M^{\frac{1}{2}} p$ for $p \in R^n$ gives (17) immediately. □

Lemma 5 *Let $\{x_k\}$ be a T-IHN sequence, and α and β be the two parameters of Wolfe's conditions (4). If Assumption 1 holds, then there exists an index k_0 such that for $k \ge k_0$,*

$$\begin{aligned} & f(x_k + p_k) - f(x_k) - \alpha g_k^T p_k \\ &= \frac{1}{2} p_k^T [(H(u_k) - H_*) - 2(1-\alpha)(M_k - M_*)] p_k \\ &\quad + \frac{1}{2} [p_k^T H_* p_k - 2(1-\alpha) p_k^T M_* p_k], \end{aligned} \tag{18}$$

where $u_k = x_k + t_k p_k$ for some $t_k \in (0,1)$, and

$$\begin{aligned} & g(x_k + p_k)^T p_k - \beta g(x_k)^T p_k \\ &= \int_0^1 p_k^T [H(x_k + t p_k) - H_*] p_k dt - (1-\beta) p_k^T (M_k - M_*) p_k \\ &\quad + p_k^T H_* p_k - (1-\beta) p_k^T M_* p_k. \end{aligned} \tag{19}$$

Proof By Lemmas 1 and 3, we can choose an index k_0 such that

$$p_k^T (g_k + M_k p_k) = 0, \quad \forall k \ge k_0. \tag{20}$$

By Taylor expansion, there exists $u_k = x_k + t_k p_k$, where $t_k \in (0,1)$, such that

$$f(x_k + p_k) = f(x_k) + g_k^T p_k + \frac{1}{2} p_k^T H(u_k) p_k.$$

Hence, by (20) and the obvious algebraic manipulations, we have for $k \ge k_0$

$$f(x_k + p_k) - f(x_k) - \alpha g_k^T p_k$$

$$
\begin{aligned}
&= (1-\alpha) g_k^T p_k + \frac{1}{2} p_k^T H(u_k) p_k \\
&= -(1-\alpha) p_k^T M_k p_k + \frac{1}{2} p_k^T H(u_k) p_k = \frac{1}{2} p_k^T (H(u_k) - 2(1-\alpha) M_k) p_k \\
&= \frac{1}{2} p_k^T [H(u_k) - H_* + H_* - 2(1-\alpha) M_* + 2(1-\alpha) M_* - 2(1-\alpha) M_k] p_k \\
&= \frac{1}{2} p_k^T [(H(u_k) - H_*) - 2(1-\alpha)(M_k - M_*)] p_k + \frac{1}{2} p_k^T [H_* - 2(1-\alpha) M_*] p_k,
\end{aligned}
$$

which gives (18). The proof of (19) is similar. In fact, by Taylor expansion,

$$
g(x_k + p_k) = g(x_k) + \int_0^1 H(x_k + t p_k) p_k dt,
$$

and arguing as previously, we get for all $k \geq k_0$

$$
\begin{aligned}
& g(x_k + p_k)^T p_k - \beta g(x_k)^T p_k \\
&= \int_0^1 p_k^T H(x_k + t p_k) p_k dt + (1-\beta) p_k^T g(x_k) \\
&= \int_0^1 p_k^T H(x_k + t p_k) p_k dt - (1-\beta) p_k^T M_k p_k \\
&= \int_0^1 p_k^T [H(x_k + t p_k) - H_*] p_k dt + p_k^T [H_* - (1-\beta) M_* \\
&\quad + (1-\beta) M_* - (1-\beta) M_k] p_k \\
&= \int_0^1 p_k^T [H(x_k + t p_k) - H_*] p_k dt - (1-\beta) p_k^T (M_k - M_*) p_k \\
&\quad + p_k^T [H_* - (1-\beta) M_*] p_k.
\end{aligned}
$$

This completes the proof of this lemma. □

Now we are in a good position to present and prove the following theorem. Here, $\rho(A)$ denotes the spectral radius of a square matrix A, which is the largest of moduli of the eigenvalues of A.

Theorem 2 (Wolfe's Admissibility) *If Assumption 1 holds, and*

$$
\rho(M_*^{-1} H_* - I) < \min\{1 - 2\alpha, \beta\}, \tag{21}
$$

where α and β are the parameters in Wolfe's conditions (4), then there exists an index k_0 such that $\lambda_k = 1$ satisfies (4) for all $k \geq k_0$.

Proof By Lemma 4, all the eigenvalues of $M_*^{-1} H_*$ are positive real numbers. Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of $M_*^{-1} H_*$, respectively. From (21) it quickly follows that $1 - \beta \leq \lambda_{\min}$ and $\lambda_{\max} \leq 2(1-\alpha)$. Hence,

we can set two small positive numbers ϵ_1 and ϵ_2 by

$$\epsilon_1 = 2(1-\alpha) - \lambda_{\max} \quad \text{and} \quad \epsilon_2 = \lambda_{\min} - (1-\beta).$$

Since M_* is SPD, the minimum eigenvalue $\lambda_{\min}(M_*)$ of M_* is positive. By the continuity of $M(x)$ and $H(x)$ at x^*, for the positive number ϵ defined by

$$\epsilon = \frac{1}{3}\lambda_{\min}(M_*)\min\{\epsilon_1, \epsilon_2\},$$

there exists $\delta > 0$ such that

$$\|M(x) - M(x_*)\| \le \epsilon \quad \text{and} \quad \|H(x) - H(x_*)\| \le \epsilon, \quad \forall x \in \mathcal{B}(\delta), \tag{22}$$

where $\mathcal{B}(\delta) = \{x \in \mathcal{D} | \|x - x^*\| < \delta\}$.

The first term of the right hand side of (19) can be estimated as below:

$$\begin{aligned}
&\left| \int_0^1 p_k^T [H(x_k + tp_k) - H_*] p_k dt \right| \\
&\le \int_0^1 \|H(x_k + tp_k) - H_*\| dt \|p_k\|^2 \\
&\le \sup_{0 \le t \le 1} \|H(x_k + tp_k) - H_*\| \|p_k\|^2 = \|H(x_k + t_k' p_k) - H_*\| \|p_k\|^2,
\end{aligned} \tag{23}$$

where $\|H(x_k + t_k' p_k) - H_*\| = \sup_{0 \le t \le 1} \|H(x_k + tp_k) - H_*\|$ with $t_k' \in [0, 1]$ since $\|H(x_k + tp_k) - H_*\|$ is continuous on $0 \le t \le 1$.

Set $v_k = x_k + t_k' p_k$ and recall that $u_k = x_k + t_k p_k$. Since $x_k \to x^*$, both u_k and $v_k \to x^*$ too. Thus, we can choose an index k_1 such that x_k, u_k, and v_k are in $\mathcal{B}(\delta)$ for $k \ge k_1$. Also, by Lemma 5, we can choose an index k_2 such that (18) and (19) hold. Set $k_0 = \max\{k_1, k_2\}$. Hence, applying (22) to (23) gives

$$\left| \int_0^1 p_k^T [H(x_k + tp_k) - H_*] p_k dt \right| \le \epsilon \|p_k\|^2, \quad \forall k \ge k_0. \tag{24}$$

If $k \ge k_0$, then using (18) and (22), we get

$$\begin{aligned}
&f(x_k + p_k) - f(x_k) - \alpha g_k^T p_k \\
&\le \frac{1}{2}[\|H(u_k) - H_*\| + 2(1-\alpha)\|M_k - M_*\|]\|p_k\|^2 \\
&\quad + \frac{1}{2}[p_k^T H_* p_k - 2(1-\alpha) p_k^T M_* p_k] \\
&\le \frac{1}{2}[\epsilon + 2(1-\alpha)\epsilon]\|p_k\|^2 + \frac{1}{2}[p_k^T H_* p_k - 2(1-\alpha) p_k^T M_* p_k] \\
&= \frac{1}{2}(3 - 2\alpha)\epsilon\|p_k\|^2 + \frac{1}{2}[p_k^T H_* p_k - 2(1-\alpha) p_k^T M_* p_k].
\end{aligned}$$

Now we give estimate of the last term of the above inequality. Since M_* is SPD, $p_k^T M_* p_k \geq \lambda_{\min}(M_*)\|p_k\|^2$. By Lemma 4, $p_k^T H_* p_k \leq \lambda_{\max} p_k^T M_* p_k$. Thus,

$$\begin{aligned}
&\frac{1}{2}[p_k^T H_* p_k - 2(1-\alpha)p_k^T M_* p_k] \\
&\leq \frac{1}{2}[\lambda_{\max} - 2(1-\alpha)]p_k^T M_* p_k \\
&= -\frac{1}{2}\epsilon_1 p_k^T M_* p_k \leq -\frac{1}{2}\epsilon_1 \lambda_{\min}(M_*)\|p_k\|^2 \leq -\frac{3}{2}\epsilon\|p_k\|^2.
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&f(x_k + p_k) - f(x_k) - \alpha g_k^T p_k \\
&\quad \leq \frac{1}{2}(3-2\alpha)\epsilon\|p_k\|^2 - \frac{3}{2}\epsilon\|p_k\|^2 = -\alpha\epsilon\|p_k\|^2 \leq 0.
\end{aligned}$$

This completes the proof that the Wolfe's first condition (4a) holds for $\lambda_k = 1$ for all $k \geq k_0$.

The admissibility of $\lambda_k = 1$ in the Wolfe's second condition (4b) follows from (19), (22), (24), and Lemma 4 as shown below:

$$\begin{aligned}
&g(x_k + p_k)^T p_k - \beta g(x_k)^T p_k \\
&\quad \geq -[\epsilon + (1-\beta)\epsilon]\|p_k\|^2 + [\lambda_{\min} - (1-\beta)]p_k^T M_* p_k \\
&\quad = -(2-\beta)\epsilon\|p_k\|^2 + \epsilon_2 p_k^T M_* p_k \\
&\quad \geq -(2-\beta)\epsilon\|p_k\|^2 + \epsilon_2 \lambda_{\min}(M_*)\|p_k\|^2 \\
&\quad \geq -(2-\beta)\epsilon\|p_k\|^2 + 3\epsilon\|p_k\|^2 = (1+\beta)\epsilon\|p_k\|^2 \geq 0.
\end{aligned}$$

This completes the proof of the theorem. □

Finally, we obtain that T-IHN has a local Q-linear rate of convergence in the following theorem.

Theorem 3 (Q-Linear Rate of Convergence) *Let $\{x_k\}$ be a T-IHN sequence, $IT_{\max}$ a parameter given in Algorithm 1, and $e_k = x_k - x^*$. If Assumption 1 and (21) hold, and $IT_{\max} \geq n$, then there exist an index $k_0 > 0$ and a constant $c \in (0,1)$ such that*

$$\|e_{k+1}\|_* \leq c\|e_k\|_*, \quad \forall k \geq k_0,$$

where $\|\cdot\|_*$ is defined by $\|x\|_* = \|M_*^{\frac{1}{2}} x\|$.

Proof By Theorem 2, there exists $k_1 > 0$ such that $\lambda_k = 1$, and the T-IHN iterate can be written as $x_{k+1} = x_k + p_k$ for $k \geq k_1$. By assumption, $x_k \to x^*$, where $M(x^*)$ is SPD. Hence, by Lemma 2, there exists an index k_2 such that M_k^{-1} exists and SPD.

Therefore, for $k \geq \max\{k_1, k_2\}$, the incomplete Hessian step, $p_k^{IHN} = -M_k^{-1} g_k$, is well defined and $e_{k+1} = e_k + p_k^{IHN} + p_k - p_k^{IHN}$. Thus,

$$\|e_{k+1}\|_* \leq \|e_k + p_k^{IHN}\|_* + \|p_k - p_k^{IHN}\|_*. \tag{25}$$

From Corollary 2 in [34], it can follow that there exists an index $k_3 > \max\{k_1, k_2\}$ and a constant $c_1 \in (0, 1)$ such that

$$\|e_k + p_k^{IHN}\|_* \leq c_1 \|e_k\|_*, \quad \forall k \geq k_3. \tag{26}$$

In terms of residual $r_k = -M_k p_k - g_k$, we have $p_k^{IHN} - p_k = M_k^{-1} r_k$, and

$$\|p_k^{IHN} - p_k\|_* = \|M_k^{-1} r_k\|_* = \|M_*^{\frac{1}{2}} M_k^{-1} r_k\| \leq \|M_*^{\frac{1}{2}}\| \|M_k^{-1}\| \|r_k\|.$$

By the continuity of $M(x)^{-1}$ at x^*, clearly there exists $k_4 > 0$ such that

$$\|M_k^{-1}\| \leq \|M_k^{-1} - M_*^{-1}\| + \|M_*^{-1}\| \leq 1/\|M_*^{\frac{1}{2}}\| + \|M_*^{-1}\|,$$

for all $k \geq k_4$. By Lemma 3 and the assumption that $IT_{\max} \geq n$, there exists an index k_5 such that

$$\|r_k\| \leq \eta_k \|g_k\|, \quad \forall k \geq k_5.$$

Furthermore, by Taylor expansion about x^*,

$$g_k = g(x_k) - g(x^*) = \int_0^1 H(x_k + te_k) e_k dt.$$

Hence,

$$\begin{aligned}
\|g_k\| &\leq \left(\int_0^1 \|H(x_k + te_k) - H(x^*)\| dt \|M_*^{-\frac{1}{2}}\| + \|H_* M_*^{-\frac{1}{2}}\| \right) \|e_k\|_* \\
&\leq \left(\sup_{0 \leq t \leq 1} \|H(x_k + te_k) - H(x^*)\| \|M_*^{-\frac{1}{2}}\| + \|H_* M_*^{-\frac{1}{2}}\| \right) \|e_k\|_* \\
&= (\|H(x_k + t'_k e_k) - H(x^*)\| \|M_*^{-\frac{1}{2}}\| + \|H_* M_*^{-\frac{1}{2}}\|) \|e_k\|_*,
\end{aligned}$$

where $t'_k \in [0, 1]$ has been selected to satisfy that

$$\|H(x_k + t'_k e_k) - H(x^*)\| = \sup_{0 \leq t \leq 1} \|H(x_k + te_k) - H(x^*)\|$$

since $\|H(x_k + te_k) - H(x^*)\|$ is a continuous function of t. Hence, there exists $k_6 > 0$ such that

$$\|H(x_k + t'_k e_k) - H(x^*)\| \leq 1/\|M_*^{-\frac{1}{2}}\|,$$

and

$$\|g_k\| \leq (1 + \|H_* M_*^{-\frac{1}{2}}\|) \|e_k\|_* \quad \text{for } k \geq k_6.$$

Therefore, for $k \geq \max\{k_2, k_4, k_5, k_6\}$,

$$\|p_k^{IHN} - p_k\|_* \leq \eta_k c_2 \|e_k\|_*, \tag{27}$$

where $c_2 = (1 + \|M_*^{\frac{1}{2}}\| \|M_*^{-1}\|)(1 + \|H_* M_*^{-\frac{1}{2}}\|)$.

Set $k_7 = \max_{1 \leq i \leq 6} k_i$. It follows from (25), (26) and (27) that

$$\|e_{k+1}\|_* \leq (c_1 + \eta_k c_2)\|e_k\|_*, \quad \forall k > k_7.$$

Since $\eta_k \to 0$, $c_1 \in (0, 1)$ and c_2 is independent of k, it is clear that there exists an index $k_0 \geq k_7$ and a constant $c \in (0, 1)$ such that $c_1 + \eta_k c_2 \leq c$ for all $k \geq k_0$. This completes the proof. □

4 Application in biomolecular potential energy minimization

In this section, we construct a particular T-IHN algorithm for minimizing a potential energy function, $f(x)$, of a biomolecular system, where $x = (X_1, X_2, \ldots, X_N)$ with $X_i = (x_{1i}, x_{2i}, x_{3i})$ being the position vector of atom i, N is the total number of atoms of the biomolecular system, and $f(x)$ is a sum of bond length potentials, bond angle potentials, torsional potentials, van der Waals potentials, Coulomb potentials, and hydrogen bond interaction potentials. The definitions of these potential energy terms can be found in [3, 13, 14, 18, 23]. For example, the Coulomb potential part of $f(x)$ is defined in the form

$$E_{\text{Coulomb}}(x) = \sum_{i \neq j} c_{ij} \frac{q_i q_j}{r_{ij}},$$

where c_{ij} is a constant, q_i is the charge on atom i, and r_{ij} denotes the distance between atoms i and j, which is defined by

$$r_{ij} = \|X_i - X_j\| = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{3i} - x_{3j})^2}.$$

Clearly, the Hessian matrix of $E_{\text{Coulomb}}(x)$ is fully dense. The van der Waals potential part of $f(x)$ is also defined in terms of all the atomic position vectors X_j for $j = 1, 2, \ldots, N$, which can be described as below:

$$E_{\text{VDW}}(x) = \sum_{i \neq j} \left(\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} \right),$$

where A_{ij} and B_{ij} are the attractive and repulsive coefficients, respectively. Since each van der Waals potential term decays at least in an order of $1/r_{ij}^6$, $E_{\text{VDW}}(x)$ is usually approximately calculated by using a simple cutoff scheme [23]. However, it has been well known that $E_{\text{Coulomb}}(x)$ cannot be simply calculated approximately by a cutoff scheme; or false conformations may be produced from minimizing $f(x)$ [23]. Other bonded potential terms of $f(x)$ involve only a few of neighboring atoms.

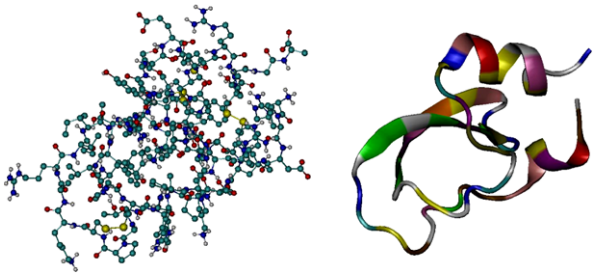

Fig. 1 The X-ray crystal structure of protein BPTI. The *left figure* gives the molecular structure with spherical balls for atoms and lines for bonds. The collection of its atomic coordinates is selected as an initial guess in all the numerical tests. The *right figure* gives the secondary structure, where different residues are shown in different colors, along with two α-helices and two β-sheets

Hence, it is the Coulomb potential part that makes the Hessian matrix of $f(x)$ fully dense.

We observed that each entry of the Hessian matrix of $E_{\text{Coulomb}}(x)$ can be found in a form of $O(1/r_{ij}^3)$, which clearly decays quickly with the increase of r_{ij}. Hence, a good incomplete Hessian matrix, $M(x)$, can be naturally generated by using a distance cutoff strategy, resulting in an efficient T-IHN algorithm for minimizing the biomolecular potential energy function $f(x)$.

In details, we define the sparse pattern P of $M(x)$ by

$$P = \{(i, j) \mid \|X_i - X_j\| \leq \tau \text{ for } i, j = 1, 2, \ldots, N\}, \tag{28}$$

where $\tau > 0$ is a given cutoff radius, and the (i, j)th block entry of $M(x)$ is set as the corresponding block entry of the Hessian matrix $H(x)$ if $(i, j) \in P$; and zero otherwise. In our numerical tests, P is only calculated once by using the initial guess x_0 and used by all the T-IHN iterations.

Based on the TN program routine of CHARMM (version 29b2) [8, 25, 29, 30, 32], we developed a T-IHN program in FORTRAN for minimizing the biomolecular potential energy function $f(x)$. Since $M(x)$ is symmetric and sparse, we only stored and evaluated the nonzero entries of the upper triangular part of $M(x)$ using a compressed sparse matrix format. Moreover, the natural preconditioner of the TN program within CHARMM [24, 28] is also employed by T-IHN for improving its performance. This T-IHN program is complicated since it involves several special numerical schemes and techniques to formulate a compressed sparse matrix of $M(x)$, and many program routines of CHARMM, which is a huge and complex FORTRAN program package. The construction, capability, and usage of this T-IHN program will be discussed in our sequent papers. To confirm our theoretical results and highlight the performance of the T-IHN program, in this section, we present some numerical results made by this T-IHN program for a protein model problem, where the potential energy function $f(x)$ is calculated from CHARMM for the protein BPTI.

Figure 1 displays an X-ray crystal three dimensional structure of BPTI, including its residue sequence and secondary structure. BPTI consists of $N = 586$ atoms (including hydrogens), and so the dimension of the minimization problem is $n = 3N = 1758$. The collection of the atomic coordinates of the X-ray structure of BPTI was

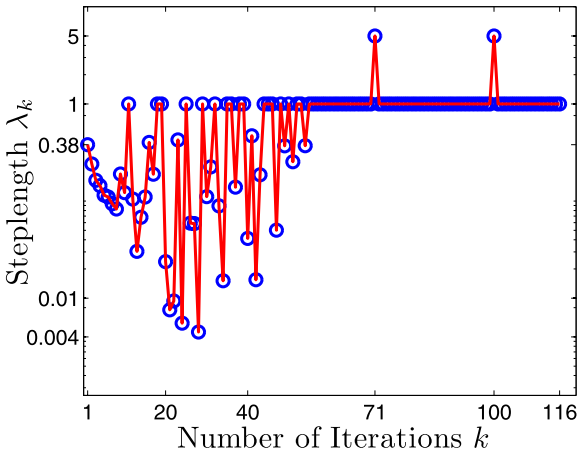

Fig. 2 Evolution of the values of the steplength λ_k of T-IHN with $\tau = 12$ for protein BPTI

used as the initial guess x_0 for all the minimizers considered in this section. Here $f(x_0) = -1782.0048$ and $\|g(x_0)\| = 18.7469$.

Two particular T-IHN algorithms, called T-IHN1 and T-IHN2, were constructed by using cutoff radius $\tau = 12$ Å and 10 Å, respectively. The two corresponding incomplete Hessian matrices have the nonzero entry ratios (i.e., a ratio of the total number of nonzero entries over n^2) 37.13% and 25.31%, respectively. We also carried out numerical tests using the CHARMM minimizers TN, ABNR, and CONJ, as well as D-TN (a version of TN with the numerical option of Hessian-vector product, in which the Hessian-vector product is approximated by the Euler forward finite-difference formula). The tests were made on one processor (600 MHz MIPS R14000 with 1 Gbytes main memory) of the SGI Origin 300 at the University of Wisconsin-Milwaukee. For simplicity, the default values of CHARMM for implementing TN, ABNR and CONJ were used in all the tests, along with the following convergence criterion:

$$\|g(x_k)\| < \epsilon_g(1 + |f(x_k)|), \tag{29}$$

where $\epsilon_g = 10^{-6}$.

Figure 2 displays the evolution of the steplength λ_k of T-IHN1 for BPTI minimization. From the figure we can see that the values of λ_k vary mostly between 0 and 1 but can be fixed as 1 when the T-IHN iterates are close enough to the minimum point x^*. This confirms our theoretical result of Theorem 2.

Figure 3 compares the gradient norms generated by T-IHN, TN, and D-TN. It confirms that T-IHN has a linear rate of convergence as claimed in Theorem 3 while both TN and D-TN have a super-linear rate of convergence. Here, the same preconditioner was used by T-IHN, TN, and D-TN.

Figures 4 and 5 give a comparison of the gradient norms and energy function values generated by T-IHN, ABNR and CONJ for BPTI minimization. From the figures we can see that T-IHN has a much faster rate of convergence than both ABNR and CONJ, and the energy function values are reduced monotonically.

Table 1 compares the performance of T-IHN with that of the CHARMM minimizers TN, D-TN, ABNR, and CONJ. In the second column, the number in parentheses

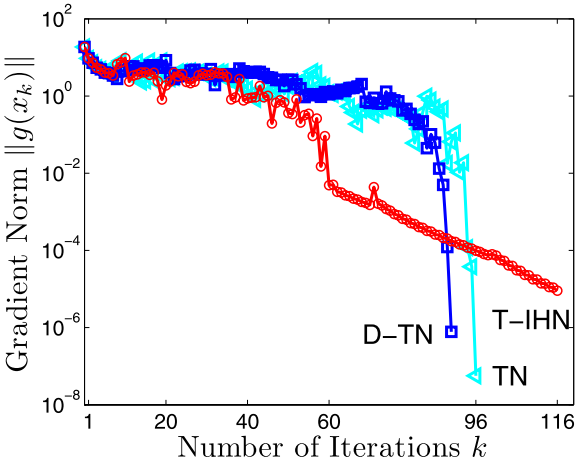

Fig. 3 Comparison of gradient norms generated by T-IHN, D-TN, and TN for BPTI minimization

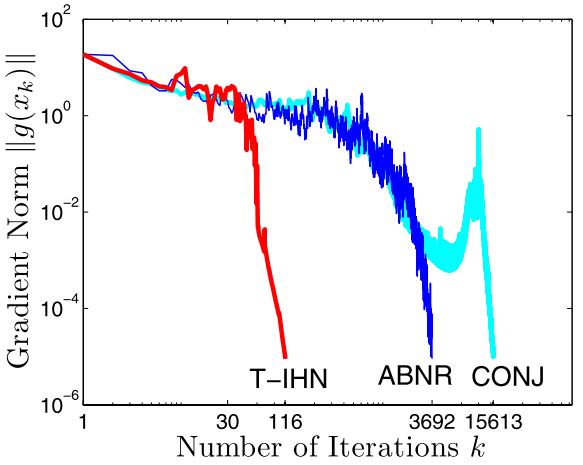

Fig. 4 A comparison of the gradient norms generated by T-IHN and the two CHARMM minimizers ABNR and CONJ for BPTI minimization

is the total number of PCG iterations for producing the descent search directions of TN, D-TN, or T-IHN. As can be seen from the table, T-IHN1 took more iterations but less CPU time than TN and D-TN. This becomes possible since each iteration of T-IHN is much cheaper in computing than TN and D-TN. The table also shows that T-IHN can be much more efficient than ABNR and CONJ in searching for a local minimum point of the biomolecular potential energy function. About 60% and 95% total CPU times were saved by T-IHN1 compared to ABNR and CONJ, respectively.

We noticed that different local minimum points were located by these different minimizers. This multiple-minima problem is beyond the scope of this work. But, we noticed that the corresponding molecular structures of these different minimum points were very similar (as shown in Fig. 6, for example); all of them may be favorable in a biomolecular simulation.

Table 1 A comparison of the performance of T-IHN with that of the CHARMM minimizers TN, D-TN, CONJ, and ABNR for BPTI minimization ($n = 1704$). Here T-IHN1 and T-IHN2 denote the two T-IHN algorithms using the incomplete Hessian using matrices defined from (28) with the cutoff radius $\tau = 12$ and 10, respectively, and D-TN is the TN using the numerical Hessian-vector product option

Minimizer	Iterations	CPU Time (minutes)	Final E	Final $\|g\|$
T-IHN1	137 (9710)	1.71	−2672.73	9.82×10^{-7}
T-IHN2	427 (16308)	3.14	−2724.84	9.95×10^{-7}
TN	96 (6084)	2.88	−2742.63	5.66×10^{-8}
D-TN	69 (3008)	1.78	−2671.59	5.52×10^{-8}
ABNR	4621	4.38	−2756.52	8.03×10^{-7}
CONJ	16025	39.29	−2763.21	9.89×10^{-7}

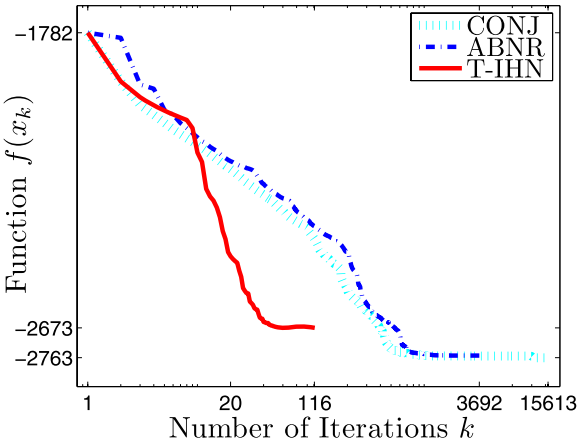

Fig. 5 A comparison of the biomolecular potential energy values generated by T-IHN, ABNR, and CONJ for BPTI minimization

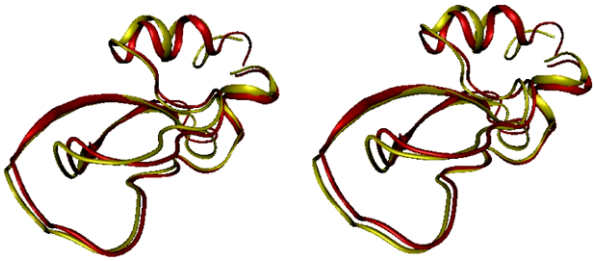

Fig. 6 (Color online) A comparison of the energy minimized structures (in backbone) of BPTI: T-IHN (*dark/red*) vs. CONJ (*light/yellow*) in the *left figure*; and T-IHN (*dark/red*) vs. ABNR (*light/yellow*) in the *right figure*

5 Conclusion

We have theoretically and numerically shown that T-IHN is a robust and practical numerical algorithm for solving a class of minimization problems of (1). We have also obtained a sufficient condition (21) for constructing an efficient T-IHN algorithm, and completed the proofs of the admissibility of a line search steplength of one in T-IHN and the Q-linear rate of convergence of T-IHN. As an important application, an efficient T-IHN algorithm was developed and programmed for minimizing

the biomolecular potential energy function based on the widely-used biomolecular simulation package CHARMM [3, 8, 25, 29, 30, 32]. The numerical experiments reported here confirm our theoretical results and demonstrate that T-IHN can compare favorably (in terms of computer CPU time) to the other available minimizers in CHARMM.

However, further work is needed to make our T-IHN program package a powerful tool in real biomolecular simulations. For example, more numerical experiments on various biomolecular systems are needed to check the capability and reliability of our T-IHN program. Another important issue is to make T-IHN effective for a solvated protein system based on either an explicit solvent approach (e.g., placing the protein at the center of a large box full of water molecules [1, 23]) or an implicit solvent approach (e.g., modeling the solvent as a continuum dielectric media [2, 21, 33]). We plan to study these issues, in the future, to further improve our T-IHN program.

Acknowledgements The authors thank the two anonymous referees for their valuable comments, which led to the improvement of the paper.

References

1. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Oxford University Press, London (1990)
2. Baker, N.A.: Improving implicit solvent simulations: a Poisson-centric view. Curr. Opin. Struct. Biol. **15**, 137–143 (2005)
3. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. **4**, 187–217 (1983)
4. Chu, J.W., Trout, B.L., Brooks, B.R.: A super-linear minimization scheme for the nudged elastic band method. J. Chem. Phys. **119**(24), 12708–12717 (2003)
5. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A.: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. **117**, 5179–5197 (1995)
6. Dembo, R.S., Steihaug, T.: Truncated-Newton algorithms for large-scale unconstrained optimization. Math. Program. **26**, 190–212 (1983)
7. Dennis, J.E. Jr., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Inc., Englewood Cliffs (1983). Reprinted with corrections by SIAM, Philadelphia (1996)
8. Derreumaux, P., Zhang, G., Brooks, B., Schlick, T.: A truncated-Newton method adapted for CHARMM and biomolecular applications. J. Comput. Chem. **15**, 532–552 (1994)
9. Fletcher, R., Reeves, C.M.: Function minimization by conjugate gradients. Comput. J. **7**(2), 149–154 (1964)
10. Gill, P.E., Murray, W.: Newton-type methods for unconstrained and linearly constrained optimization. Math. Program. **28**, 311–350 (1974)
11. Golub, G.H., van Loan, C.F.: Matrix Computations, 2nd edn. John Hopkins University Press, Baltimore (1986)
12. Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E.: Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theor. Comput. **4**, 435–447 (2008)
13. MacKerell, A.D. Jr., Foloppe, N.: All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phased macromolecular target data. J. Comput. Chem. **21**, 86–104 (2000)
14. MacKerell, A.D. Jr., Foloppe, N.: All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. J. Comput. Chem. **21**, 105–120 (2000)
15. Molecular mechanics and modeling. Chem. Rev. **93**(7), (1993). Special issue

16. Moré, J.J., Thuente, D.J.: Line search algorithms with guaranteed sufficient decrease. ACM Trans. Math. Softw. **20**, 286–307 (1994)
17. Nelson, M., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L., Skeel, R.D., Schulten, K.: NAMD—a parallel, object-oriented molecular dynamics program. Int. J. Supercomput. Appl. High Perform. Comput. **10**(4), 251–268 (1996)
18. Neumaier, A.: Molecular modeling of proteins and mathematical prediction of protein structure. SIAM Rev. **39**, 407–460 (1997)
19. Nocedal, J., Wright, S.: Numerical Optimization, 2nd edn. Springer, New York (2006)
20. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, San Diego (1970)
21. Roux, B., Simonson, T.: Implicit solvent models. Biophys. Chem. **78**, 1–20 (1999)
22. Schlick, T.: Geometry optimization. In: von Ragué Schleyer, P.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer, H.F. III (eds.) Encyclopedia of Computational Chemistry, vol. 2, pp. 1136–1157. Wiley, New York (1998)
23. Schlick, T.: Molecular Modeling and Simulation, an Interdisciplinary Guide. Springer, New York (2002)
24. Schlick, T., Overton, M.L.: A powerful truncated Newton method for potential energy functions. J. Comput. Chem. **8**, 1025–1039 (1987)
25. Schlick, T., Fogelson, A.: TNPACK—a truncated Newton minimization package for large-scale problems: I. Algorithm and usage. ACM Trans. Math. Softw. **14**, 46–70 (1992)
26. Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P., van Gunsteren, W.F.: The GROMOS biomolecular simulation program package. J. Phys. Chem. A **103**, 3596–3607 (1999)
27. Sundaram, R.K.: A First Course in Optimization Theory. Cambridge University Press, Cambridge (1996)
28. Xie, D.: An Effective Compressed Sparse Preconditioner for Large Scale Biomolecular Simulations. Lecture Notes in Computer Science, vol. 3314, pp. 64–70. Springer, Berlin (2004)
29. Xie, D., Schlick, T.: Efficient implementation of the truncated Newton method for large-scale chemistry applications. SIAM J. Optim. **10**(1), 132–154 (1999)
30. Xie, D., Schlick, T.: Remark on the updated truncated Newton minimization package, algorithm 702. ACM. Trans. Math. Softw. **25**(1), 108–122 (1999)
31. Xie, D., Schlick, T.: Visualization of chemical databases using the singular value decomposition and truncated-Newton minimization. In: Floudas, C.A., Pardalos, P. (eds.) Optimization in Computational Chemistry and Molecular Biology: Local and Global Approaches, pp. 267–286. Kluwer Academic, Dordrecht (2000)
32. Xie, D., Schlick, T.: A more lenient stopping rule for line search algorithms. Optim. Methods Softw. **17**(4), 683–700 (2002)
33. Xie, D., Zhou, S.: A new minimization protocol for solving nonlinear Poisson-Boltzmann mortar finite element equation. BIT **47**, 853–871 (2007)
34. Xie, D., Ni, Q.: An incomplete Hessian Newton minimization method and its application in a chemical database problem. Comput. Optim. Appl. Published online: 12 January 2008
35. Xie, D., Singh, S.B., Fluder, E.M., Schlick, T.: Principal component analysis combined with truncated-Newton minimization for dimensionality reduction of chemical databases. Math. Program. **95**(1), 161–185 (2003)