A non-Secant quasi-Newton Method for Unconstrained Nonlinear Optimization

Issam A.R. Moghrabi*

Abstract: The Secant equation has long been the foundation of quasi-Newton methods, as updated Hessian approximations satisfy the equation with each iteration. Several publications have lately focused on modified versions of the Secant relation, with promising results. This study builds on that idea by deriving a Secant-like modification that uses non-linear quantities to construct Hessian (or its inverse) approximation updates. The method uses data from the two most recent iterations to provide an alternative to the Secant equation with the goal of producing improved Hessian approximations that induce faster convergence to the objective function optimal solution. The reported results provide strong evidence that the proposed method is promising and warrants further investigation.

Subjects: nonlinear programming; optimization; decision science; engineering

Keywords: quasi-Newton methods; Secant-like methods; BFGS; unconstrained optimization; multi-step methods

1. Introduction
The techniques examined in this paper are those employed in the solution of problems of the form

\[ \min f(x), \ \text{where} \ \mathbf{R}^n \rightarrow \mathbf{R}. \]

Such issues are common in all types of engineering and manufacturing projects. The above problem is solved iteratively using quasi-Newton methods. For the current iteration \( i \), the gradient of \( f \) at \( x_i \) is indicated as \( g_i \) and the matrix \( B_i \) is approximates \( G(x_i) \), the actual Hessian of \( f \). The next iteration Hessian matrix approximation must traditionally satisfy the so-called Secant equation.
The new solution approximation is defined as

\[ x_{i+1} = x_i + \alpha p_i, \]

where \( p_i \) the direction vector, obtained through solving

\[ B_i p_i = -g_i \text{ or } p_i = -H_i g_i. \]

The step length \( \alpha_i \) in (2) is computed as the solution to

\[ \alpha_i = \min_{\alpha \in [0,1]} f(x_i + \alpha p_i), \]

by means of some line search methodology to satisfy conditions the following Wolfe-Powell-conditions are satisfied (Fletcher (1987))

\[ f(x_{i+1}) \leq f(x_i) + 10^{-4} s_i^T g_i \]

and

\[ s_i^T g_{i+1} \geq 0.9 p_i^T g_i. \]

Satisfying conditions of the type in (4) and (5) is necessary to guarantee convergence.

The derivation and coding of the actual Hessian matrix is susceptible to human error and demands considerable storage for large problems. The identity matrix (or some weighted version of it) is usually chosen as the initial approximation to the Hessian, and it is updated step-by-step using the most recent available computed data. It has been demonstrated that such methods exhibit superlinear convergence under reasonable assumptions about the objective functions (Fletcher, 1970, 1994; Dennis & Schnabel, 1979). When the line searches are exact, the methods converge in at most \( n \) iterations on quadratic functions (Broyden, 1970).

The most numerically successful update formula used to approximate the Hessian on a step-wise basis is the BFGS update and is given as

\[ B_{i+1} = B_i - \frac{y_i y_i^T}{s_i^T y_i} + \frac{B_i s_i s_i^T B_i}{s_i^T B_i s_i}. \]

The global convergence of the BFGS method for convex objective functions has been established by some authors (see, Dai et al., 2002; Dennis & Schnabel, 1979; Fletcher, 1987; Shanno, 1978; Shanno & Phua, 1978; Xiao et al., 2006; Yuan et al., 2018, 2017). Dai et al. (2002) demonstrate that when the line searches are inexact, the standard BFGS method may fail to converge on non-convex functions. Wei et al. (2004, 2006) show that even with accurate-line searches, the classical BFGS
method can fail to converge. Li and Fukushima (2001) modified the classical BFGS formula to create a variant of the method that exhibits global convergence even when the objective function \( f \) lacks the convexity condition.

### 2. Variants of the Secant equation

Much research has been conducted in order to develop methods that outperform the classical BFGS update numerically and are globally convergent under reasonable assumptions. In this section, we provide a brief overview of some of the successful methods that have outperformed the classical BFGS. The algorithms’ development is primarily based on the introduction of modifications to the Secant equation (1). The concept of using more of the data available at each iteration in the update of the Hessian approximation that would otherwise be disregarded motivates one particularly successful class of approaches documented in the literature. These data include, though not limited to, the readily computed step vectors as well as the gradient difference vectors in (1) from the \( m \) latest iterations \( (m > 1) \). The standard quasi–Newton methods utilize just from the most recent iteration. We next present a particularly successful class of methods known as the multi-step methods (Ford & Moghrabi, 1993, 1994, 1996).

Applying the Chain Rule is applied to \( g(x(r)) \), for a differentiable path in \( R^n \), \( X = \{x(r)\} \) (for \( r \in R \)), and differentiate with respect to \( r \), the result is

\[
G(x_{i-1})X'(\tau_m) = g'(x(\tau_m)).
\]

In particular, if \( m \) is chosen to pass through the most recent point \( x_{i-1} \) (so that \( x(\tau_m) = x_{i-1} \), say), then equation (6), known as the “Newton Equation” (see, Al-Baali, 1985; Broyden, 1970; Dennis & Schnabel, 1979), presents generally an alternative that the Hessian matrix \( G(x_{i-1}) \) satisfies. Equation (6) may be regarded as a basis for the Secant equation when \( m \) is chosen to be \( i + 1 \) (Broyden, 1970). In Ford and Moghrabi (1994), the path \( X \) is constructed as the polynomial interpolating the \( m + 1 \) most recent points \( \{x_{i-m \cdot k+1}^m\}_{k=0}^m \). The polynomial \( g(r) \), say, is used to approximate the vector \( g'(x(\tau_m)) \) that by differentiating the corresponding polynomial which interpolates the corresponding available gradient points \( \{g(x_{i-m \cdot k+1})\}_{k=0}^m \).

The scalar values \( \{\tau_k\}_{k=0}^m \) correspond to the points \( \{x_{i-m \cdot k+1}^m\}_{k=0}^m \) on the curve \( X = \{x(r)\} \) as follows

\[
x(\tau_{k}) = x_{i-m \cdot k+1}, \text{for } k = 0, 1, \ldots, m.
\]

If \( B_{i+1} \) is a given approximation to \( G(x_{i+1}) \) in (6) and we make the following definitions

\[
\lim_{i \to \infty} f_i = f^*.
\]

\[
l_1 p^2 \leq d^T \nabla^2 f(x) d \leq l_2 p^2
\]

(where the vectors \( s_i \) and \( y_i \) are as defined in (1) and \( w_i \) represents an approximation to \( g'(x(\tau_m)) \), it is not unreasonable (by (6)) to force the matrix \( B_{i+1} \) to satisfy a relation similar to the one in (6). We, thus, have

\[
B_{i+1} f_i = w_i.
\]

There remains the issue of choosing values for the \( \tau \) parameters. One possible choice is (Ford & Moghrabi, 1993)
\( r_0 = -((||s_{i2}|| + ||s_{i-12}||)). \quad r_2 = 0, \text{and} \quad r_1 = -||s_{i2}||. \) (10)

This choice has the merit of reflecting the distances among the iterates in the space of the variables and updated for each new point.

The multi-step BFGS Hessian update is given by

\[ B_{i+1}^{MS} = B_i + \frac{w_i \omega_i^T}{r_i^2} - \frac{B_i r_i r_i^T B_i}{r_i^T B_i r_i}. \] (11)

The main advantage of (10) is in exploiting several recent steps as well as available gradient vectors in contrast to the standard BFGS method that utilizes only latest one-step and the corresponding gradient difference vectors.

Other non-secant methods are motivated by the fact that the classical BFGS formula is confined to using single gradient and step vectors in the construction of the Hessian approximation while neglecting other potentially useful readily computed data. For example, function values might prove to be valuable in the updating process. So, focusing on building better ‘quality’ updates, variants of the classical Secant relation (1) have been explored (see, Ford & Moghrabi, 1996; Wei et al., 2004; Ortiz et al., 2004; G. Yuan & Wei, 2010; Yuan et al., 2010, 2017, 2018; Zhong et al., 1999). Wei et al. (2006), using Taylor’s series, derived a modification to (1) as follows:

\[ B_{i+1} s_i = z_i, \]

where, \( z_i = y_i + \frac{\theta_i}{||r_i||} \) and \( \theta_i = 2(f_i - f_{i-1}) + (g_{i-1} + g_i)^T s_i. \)

G. Yuan and Wei (2010) used another alternative, given as:

\[ z_i = y_i + \frac{\max(0, \theta_i)}{||s_i||}. \]

Some of the recent and well-known Secant-like methods that are motivated by the need to derive accelerated convergence variants of the classical BFGS may be summarized in Table 1.

Although the methods listed in Table 1 represent some of the well-known published Secant-variant formulae that in addition to incorporating gradient and step vectors function values obtained from the latest iterate are exploited in the update of the Hessian approximation. To differ, the methods derived in Moghrabi (2017) and Ford and Moghrabi (1993), utilize nonlinear interpolating polynomials, and introduce a completely different Secant-like equation utilizing available data from the two/three most recent steps. The methods of Moghrabi (2017) are based on the idea of performing multiple updates on each iteration so that multiple Secant and Secant-variant conditions are observed. The updates are performed in a manner that ensures the first update obeys (1) while subsequent updates satisfy

\[ B_{i+1}^{(t+1)} u^t = u^t, \quad t = 1, 2, m - 1 \quad (m > 1) \]

where \( u^t = r_{i-m+1} \) and \( u^t = w_{i-m+1} \), for \( r \) and \( w \) as in (7) and (8).

A sample of the work done on the derivation of modified Secant equations is summarized in Table 1. For interested readers, there are many other similar methods available in the literature.
(Wei et al., 2006; Xiao et al., 2006; Yuan et al., 2018; G. Yuan & Wei, 2010; Yuan et al., 2017, 2010; Zhang et al., 1999). Several of the approaches mentioned here have had their convergence qualities investigated. Yuan et al. (2018), for example, demonstrate global convergence using a less stringent variant of the Powell-Wolfe line search criteria (4) and (5).

Such methods may prove useful in various domains in which the classical BFGS has been a serious contender. One possible domain is in sustainable transportation planning for dealing with the issue of minimizing air pollution, congestion, and noise (Sayyadi & Awasthi, 2018). Another possible application could be in the bi-objective inventory model created by Gharaei et al. (2020). The goal of to optimize the lot-sizing of the replenishment while meeting stochastic limitations and calculating the optimal number of lots and volume of each lot. Gharaei et al. (2015) aim to optimize incentives in single machine scheduling in reward-driven systems such that total reward is maximized while limitations such as total reward limitation for earliness and learning, independence from earliness and learning, and others are met. The goal is to maximize overall rewards in the mentioned system by using quadratic rewards for both learning and earliness, where non-secant methods may prove useful. Gharaei et al. (2019) present a mathematical model that combines the buyers’ and vendors’ total costs (TC) in a supply chain (TC) under penalty, green, and quality assurance (QC) rules, as well as a VMI-CS agreement. The goal is to find the optimal batch-sizing policy in the integrated SC with the lowest TC that finds both the number of the vendor’s batches for each of the shipped product lines and the size of the batches shipped to the buyers in order to minimize the integrated SC’s TC while satisfying the stochastic barriers. Due to the complexity of the optimization model, the possibility of employing the methods proposed in this paper for solving the integrated supply chain problem. Another attractive application might be in areas where quality control and green production policies are considered when designing a bi-objective multi-product limited and integrated economic production quantity model. There are stochastic constraints in the model mentioned above. Such research aims to find the best total inventory cost and total profit while keeping stochastic restrictions in mind (see, Gharaei et al. (2021)). Awasthi & Omrani, (2019) propose a goal-oriented strategy for prioritizing sustainable mobility projects based on fuzzy axiomatic design. The affinity diagram technique is used to establish sustainability evaluation attributes. The criteria and alternatives are rated using the fuzzy Delphi technique. Fuzzy axiomatic design provides final rankings for urban transportation initiatives, and the best project(s) are picked for implementation. Sensitivity analysis is used to examine the

| Author(s)                          | The derived Secant variant                                                                 | Ref.       |
|-----------------------------------|-------------------------------------------------------------------------------------------|------------|
| Fletcher                          | \( b_{i+1} = y_i + (1 - \lambda) b_{i} \)                                                | Fletcher (1987) |
| Zhang et al.                      | \( b_{i+1} = y_i + \frac{k(x_{i+1} - x_i) + \sigma}{\sqrt{k}} b_i \)                    | Zhang et al. (1999) |
| Wei et al.                        | \( b_{i+1} = y_i + \frac{2(x_{i+1} - x_i) + \sigma}{\sqrt{k}} b_i \)                    | Wei et al. (2006) |
| Li and Fukushima                  | \( b_{i+1} = y_i + \sigma, \sigma < 10^{-6} \)                                          | Li and Fukushima (2001) |
| Yuan et al.                       | \( b_{i+1} = y_i + \frac{\max(0, x_{i+1} - x_i) + \sigma}{\sqrt{k}} b_i \)           | Yuan et al. (2017) |
| Hassan                            | \( b_{i+1} = \frac{1}{2} y_i + \frac{2(x_{i+1} - x_i) + \sigma}{\sqrt{k}} y_i \)      | Hassan (2020) |
| Hassan and Mohammed               | \( b_{i+1} = y_i - \frac{\sigma}{\sqrt{k}} y_i \)                                      | Hassan and Mohammed (2020) |
| Hassan and Ghada                  | \( b_{i+1} = y_i + \frac{2(x_{i+1} - x_i) - \sigma}{\sqrt{k}} y_i \)                   | Hassan and Ghada (2020) |
impact of changes in input parameters on the model’s stability, and this is a possible venue for the method proposed in this paper to prove its usefulness.

Next a new method that is motivated by the same concept of incorporating data available from several of the latest iterations for deriving a new variant to the Secant equation (1) is developed.

3. A New Non-Secant equation

We will use data from three most recent iterations to derive the new Secant-like relationship that the updated Hessian (or its inverse) must fulfil, just as we did with the nonlinear multi-step quasi-Newton methods (equations (5–10)) discussed earlier, thus choosing m = 2 in (6). In particular, an interpolation of the points $x_{i-1}$, $x_i$ and $x_{i+1}$ is constructed using a differentiable curve in $R^n$, $x(\tau)$, such that $x(\tau_0) = x_{i-1}$, $x(\tau_1) = x_i$ and $x(\tau_2) = x_{i+1}$.

The corresponding objective function is modelled as $\omega(\tau) = f(x(\tau))$, using Taylor’s expansion relation around the point $\tau_2 = 0$ (corresponds to $x_{i+1}$) as follows

$$\omega(\tau) \approx \omega(0) + \tau \omega'(0) + \frac{1}{2} \tau^2 \omega''(0), \quad (12)$$

for the $\tau$-values chosen in (10).

Using (12), we have

$$\omega'(0) = x'(0)^T g_{i+1}$$
$$\omega''(0) = x'(0)^T B_{i+1} x'(0) + x'(0)^T g_{i+1}. \quad (13)$$

where $x'(\tau) = \frac{dx}{d\tau}$ and $x''(\tau) = \frac{d^2x}{d\tau^2}$.

Choosing the Lagrange representation for $x(\tau)$, we have (obtained by setting $m = 2$ in (7) and (8)), we obtain

$$x'(\tau) = L_2'(\tau)(s_i - \rho s_{i-1}) \quad (14)$$

and

$$x''(\tau) = L_2''(\tau)(s_i - \gamma s_{i-1}). \quad (15)$$

for

$$\rho = \frac{\partial^2}{\partial \tau^2}, \gamma = \frac{\partial^3 - \partial^2}{\partial \tau^3}.$$

and, for all $\tau$, the following quantities hold

$$L_2'(\tau) = \frac{2\tau - \tau_0 - \tau_1}{(\tau_2 - \tau_1)(\tau_2 - \tau_0)}, \quad L_2''(\tau) = \frac{2\tau - \tau_1 - \tau_2}{(\tau_0 - \tau_1)(\tau_0 - \tau_2)}. \quad (16)$$
\[ \mathcal{L}'(r) = \frac{2}{(r_2 - r_1)(r_2 - r_0)} \quad \text{and} \quad \mathcal{L}''(r) = \frac{2}{(r_0 - r_1)(r_0 - 2)}. \]  

(17)

For the \( \tau \)-values selected in (10), we have \( \gamma = \frac{\|S\|}{\|S_{i-1}\|}. \) It is not unreasonable then to introduce a generalization of \( \gamma \) obtained by plugging in a scaling constant, \( \mu \geq 0 \) (see, Moghrabi, 2017). Such scaling gives an easier more convenient mechanism to switch to the standard one-step Secant update method by setting \( \gamma = 0. \) Therefore,

\[ \delta = \mu \frac{\|S\|}{\|S_{i-1}\|}. \]

Should \( r_0 \) be chosen to be used in (12) if the true Hessian at \( x_{i+1} \) be substituted by its approximation \( B_{i+1} \), as is the case in the implementation of the standard quasi-Newton methods, it is reasonable to require that

\[ r_0^2 \omega'(0) \approx 2|f_{i-1} - f_{i+1} - \tau_0 \omega'(0)|. \]  

(18)

or, alternatively,

\[ f_{i-1} = f_{i+1} + \tau_0 x'(0)^T g_{i+1} + \frac{1}{2} r_0^2 x'(0)^T B_{i+1} x'(0) + x'(0)^T \left( \frac{2}{r_1 r_0} S_i - \frac{2}{r_0 - r_1} S_{i-1} \right) g_{i+1}. \]

(19)

For the quantities \( x'(0) \) and \( x''(0) \) as in (14) and (15), respectively. It follows that

\[ f_{i-1} = f_{i+1} + \tau_0 \left[ \frac{-r_0 - r_1}{r_1 r_0} S_i^T g_{i+1} + \frac{r_1}{(r_0 - r_1) r_0} \right] + \frac{1}{2} r_0^2 x'(0)^T B_{i+1} x'(0) + x'(0)^T \left( \frac{2}{r_1 r_0} S_i - \frac{2}{r_0 - r_1} S_{i-1} \right) g_{i+1}. \]

Now, from (13) and (19), we obtain

\[ \tau_0^2 \left[ \frac{-r_0 - r_1}{r_1 r_0} \right]^2 \left( S_i - \vartheta S_{i-1} \right)^T B_{i+1} \left( S_i - \vartheta S_{i-1} \right) + \frac{2}{r_1 r_0} \left( S_i - \delta S_{i-1} \right)^T g_{i+1} \]

\[ = 2|f_{i-1} - f_{i+1} - \frac{-r_0 - r_1}{r_1} \left( S_i - \vartheta S_{i-1} \right) g_{i+1}. \]

(20)

for \( \vartheta \) and \( \delta \) are as defined in (14) and (15), respectively.

Upon defining the quantities

\[ r_j \triangleq S_i - \rho S_{i-1} \quad u_j \triangleq S_i - \gamma S_{i-1} \quad \text{and} \quad \vartheta \triangleq \frac{u_j}{r_j}, \]

then from (19) we have
\[
\partial^2 r_i^T B_{i+1} r_i = 2 \left[ f_{i-1} - f_{i+1} + \partial r_i^T g_{i+1} - \frac{\tau_0}{\tau_1} g_{i+1}^T r_i \right].
\]  

(21)

Equation (21) can be rewritten as

\[
z_i^T B_{i+1} z_i = z_i^T w_i + \sigma_i
\]

(22)

for

\[
z_i = \partial r_i, \quad w_i = y_i - \rho y_{i-1}
\]

and

\[
\sigma_i = 2 \left( f_{i-1} - f_{i+1} + z_i^T g_{i+1} - \frac{\tau_0}{\tau_1} g_{i+1}^T u_i \right) - z_i^T w_i.
\]

The equation obtained in (22) suggests a new Secant-like equation of the form

\[
B_{i+1} z_i = v_i
\]

(23)

for \( z_i \) as in (21) and \( v_i = w_i + \frac{1}{2} z_i \).

The computed search direction is downhill if \( B_{i+1} \) is positive definite. By analogy with the standard Secant equation (1), \( B_{i+1} \) is positive definite if and only if \( B_i \) is already so and \( z_i^T v_i > 0 \). As this cannot be guaranteed in this new formulation, (23) is replaced with

\[
B_{i+1} z_i = \tilde{v}_i
\]

(24)

where

\[
\tilde{v}_i = v_i + \epsilon_i z_i, \quad \epsilon_i = \gamma ||g_i||^2 + \max \left( \frac{z_i^T z_i}{||z_i||^2}, 0 \right)
\]

for some positive constant \( \gamma \). Relation (24) is an appropriate replacement to (23) since

\[
\tilde{v}_i^T z_i \geq \gamma ||g_i||^2 ||z_i||^2 > 0.
\]

(25)

ascertains the positive definiteness of \( B_{i+1} \).

The outline of the new method is as follows

**Algorithm NMBFGS**

Input: \( x_0 \in \mathbb{R}^n \), and set \( H_0 = I \). Let iteration count \( i = 0 \).

Output: (near) optimal solution

1. Stop if the size of the gradient satisfies \( ||g_i|| \leq \epsilon \) (convergence threshold).
(2) compute $d_i = -H_i g_i$.

(3) Minimize, $f(x_i + \alpha d_i)$ to determine $\alpha_i$ using $\alpha \in R$

Cubic interpolation, such that conditions (4) and (5) are satisfied (an $O(n)$ operation).

(1) Compute the new iterate $x_{i+1} = x_i + \alpha_i d_i$. If (24) is satisfied, update $H_i$ using (10) with $w_i$ replaced by $\tilde{v}_i$ and $r_i$ replaced by $z_i$ in (23), else set $z_i = s_i$ and $\tilde{v}_i = y_i$ in (23) and update to satisfy condition (1) (an $O(n^2)$ operation).

(2) Set $i = i + 1$ and go to 1.

The overall cost of the above method, asymptotically, is the same as the standard quasi-Newton algorithm though practically the recorded savings in both iteration, function/gradient evaluations count differentiate the new method as opposed to others in the same category.

4. Convergence properties

The convergence analysis presented next rests on the following assumptions:

A1. The level set $D = \{x | f(x) < f(x_0)\}$ is bounded, for a starting point $x_0$.

A2. The objective function $f$ is twice continuously differentiable on $D$ and in an open set $M$ containing $D$, there exists a constant $\gamma > 0$ such that

$$||g(x) - g(y)|| \leq \gamma ||x-y||,$$

for all $x, y \in M$.

Given that the sequence $\{f_i\}$ is a decreasing sequence, then the sequence $\{x_i\}$ generated by the new method belongs to $D$, and there exists a constant $f^*$ such that:

$$\lim_{i \to \infty} f_i = f^*.$$  \hspace{1cm} (26)

A3. The objective function $f$ is uniformly convex, in that there are positive constants $l_1$ and $l_2$ such that

$$l_1 ||p||^2 \leq \nabla^2 f(x) d \leq l_2 ||p||^2$$

holds for all $x \in D$ and $p \in \mathbb{R}^n$.

**Theorem 1.** Let $f$ satisfy assumptions 1 and 2 and $\{x_i\}$ be generated by algorithm NBFGS and there exist constants $m_1$ and $m_2$ such that

$$||B_i s_i|| \leq m_1 ||s_i|| \quad \text{and} \quad s_i^T B_i s_i \geq m_2 ||s_i||^2, \forall i.$$  \hspace{1cm} (27)

then the following holds

$$\lim_{i \to \infty} \inf ||g_i|| = 0.$$  \hspace{1cm} (28)

**Proof:** Since $s_i = \alpha_i d_i$, $B_i s_i = -\alpha_i g_i$, $||g_i||^2 = \alpha_i^{-2} ||B_i s_i||$, it is straightforward to show that (27) holds with $\rho_i$ replacing $s_i$. Using (27) and $g_i = -B_i \rho_i$, we get
\( p_i^T B_i p_i \geq m_2 p_i^2 \)

and

\( m_2 p_i \leq g_i \leq m_1 p_i \).

Let us define the set \( \Delta \) of indices \( i \), for which (27) holds. By the Powell-Wolfe condition (5) and Assumption 2, we obtain

\( \xi \alpha_i p^2_i \geq p_i^T y_i \geq -0.1 p_i^T g_i \).

It follows that, for any

\[ i \in \Delta, \]

\[ \alpha_i \geq -0.1 p_i^T g_i \geq 0.1 \xi^{-1} m_1. \]

Using (26), one gets

\[ \sum_{j=1}^{\infty} (f_j - f_{j+1}) = \lim_{\epsilon \to \infty} \sum_{j=1}^{\infty} (f_j - f_{j+1}) = \lim_{\epsilon \to \infty} f_1 - f_{\epsilon} = f_1 - f_{\text{min}}, \]

and, thus,

\[ \sum_{j=1}^{\infty} (f_j - f_{j+1}) < \infty. \]

By the Powell-Wolfe condition (4), it follows that

\[ \sum_{j=1}^{\infty} s_j^T g_j < \infty, \]

and, consequently,

\[ \lim_{\epsilon \to \infty} \alpha_i \xi^2 g_i = 0. \]

Hence,

\[ \lim_{i \to \infty} p_i^T B_i p_i = \lim_{i \to \infty} p_i^T g_i = 0, \]

and (28) follows, given (27).

We now proceed to prove that algorithm NBFGS has global convergence.

**Theorem 2.** Let \( f \) satisfy assumptions 1 and 2 above and \( \{x_i\} \) generated by algorithm NBFGS. Then, the following holds

\[ \lim \inf_{i \to \infty} g_i = 0. \]  \( (29) \)

**Proof.** As per Theorem 2, it suffices to prove that (29) holds for all \( i \). By contradiction, assume this is not the case. Thus, there exists a positive constant \( \varnothing \) such that, for all \( i \),
It is then easy to show that
\[ \hat{v}_i \leq \theta s_i, \text{ for some positive constant } \theta. \] (30)

From (25) and the fact that
\[ u_i \geq \sigma (s_i - e s_{i-1}) \geq \frac{1}{2} \sigma s_i, \]
it follows that
\[ u_i^T \hat{v}_i \geq \nu^2 u_i^T u_i \geq \frac{\nu^2}{2} \sigma^2 s_i^T s_i. \] (31)

Thus, from (30) and (31), we obtain
\[ \frac{u_i^T \hat{v}_i}{u_i^T u_i} \leq \rho, \]
for \( \rho = \frac{\rho^2}{\nu^2 \sigma^2}. \)

The sequence \( \{B_i\} \), by theorem 2, for there exist constants \( m_1 \) and \( m_2 \) such that (29) applies for all \( i \). The proof completes on the basis of the following (see Theorem 2.1 in Dai et al. (2002)).

If there are two positive constants \( l_1 \) and \( l_2 \) such that for all \( i \)
\[ \frac{u_i^T \hat{v}_i}{u_i^T u_i} \leq l_1 \text{ and } \frac{\hat{v}_i^T v_i}{\hat{v}_i^T v_i} \geq l_2, \]
such that for any positive integer \( j \) (28) holds for no less than \( j/2 \) iterations of \( i \in \{1, \ldots, j\} \).

4.1. Non-asymptotic Superlinear Convergence of the New Method
The analysis presented here is quite similar to that of Jin and Mokhtari (2021). The convergence results presented by the authors apply primarily to the standard BFGS method. We follow a similar approach to prove the superlinear convergence of NMBFGS. To do so, the following assumptions are needed:

A3. The objective function \( f \) is twice differentially continuous. It is also convex with convexity parameter \( \sigma > 0 \) and a gradient \( g(x) \) which is Lipschitz continuous with parameter \( \rho > 0 \). Thus, we have
\[ \sigma x - v \leq g(x) - g(v) \leq \rho x - v, \forall x, y \in \mathbb{R}^n. \]

\[ \sigma I \leq G(x) \leq \rho I, \forall x \in \mathbb{R}^n. \]

A4. The Hessian matrix \( G(x) \) of \( f \) satisfies the condition
\[ G(x) - G(x^*) \leq x - x^*, \forall x \in \mathbb{R}^n, \]

for some constant \( \delta > 0 \).

**Theorem 3.** Consider the method described in Algorithm NMBFGS and suppose the objective function \( f \) satisfies the conditions in Assumptions A3 and A4. Choose two real parameters, for any arbitrary \( \alpha \) in (0, 1), choose \( \varepsilon > 0 \) and \( \theta > 0 \) such that

\[
\varepsilon \left( \frac{2\alpha \theta + \alpha_2}{1 - \alpha} \right) \leq \theta, \quad \frac{1}{\delta} \left( 2\rho \theta + (2\theta + 1) \right) \leq \alpha,
\]

where \( \alpha_1 = (15/4) \frac{\theta}{\delta^2} \) and \( \alpha_2 = 2 \left( 1 + 2n^2 \right) \frac{\theta}{\delta^2} G(x^*)^\frac{1}{2} \). Should the start point \( x_0 \) and the initial inverse Hessian approximation \( H_0 \) satisfy the relations

\[
||x_0 - x^*|| < \min \left\{ \varepsilon, \frac{\sigma}{3n\sqrt{\rho}} \right\}, ||H_0 - G(x^*)^{-1}||_{G(x^*)^{-1/2}} \leq \theta,
\]

then the sequence \( (x_i)_{i=0}^\infty \) generated by the new method converges to the optimal point \( x^* \) superlinearly at a rate of

\[
x_0 - x^* \leq \left( \frac{4\theta^2(1+\alpha)}{\sqrt{2n^2(1-\alpha)} \sqrt{\delta + \frac{\rho\theta(1+\alpha)}{\sigma^2(1-\alpha)^2}}} \right)^\delta x_0 - x^*. \quad (32)
\]

**Proof.** Similar to the proof done in Jin and Mokhtari (2021).

Theorem 4 suggests that the sequence of points generated by NMBFGS advance to the optimal solution at a superlinear rate of \( \left( \frac{\alpha \sqrt{\theta + \alpha}}{\delta} \right)^\delta \) in the local neighbourhood of the optimal solution. The constants \( \varepsilon_1 \) and \( \varepsilon_2 \) depend on the parameter of the problem being solved. For sufficiently large \( \delta \), the convergence rate is \( O \left( \left( \frac{1}{\delta} \right)^\delta \right) \).

There is a compromise between the speed the method converges superlinearly and the size of the local neighbourhood in that increasing the size of \( \varepsilon \) and \( \theta \) results in an increase in the size of the vicinity the method converges at a superlinear rate. However, this will lower the speed of convergence since (32) increases therefore, by selecting different values for the parameters \( \alpha, \varepsilon \) and \( \delta \) both the speed and region of superlinear convergence change.

**5. Numerical test results and discussions**

Summaries of the numerical results are tabulated in this section. Table 2 summarizes the list of test problems. The chosen test set is for issues of varying dimensions, allowing for the examination of various test problem sizes. The functions that were tested, and hence the results that were reported, relate to problems that were categorized by size/dimension. Those problem sizes fall into four categories, low \( (2 \leq n \leq 20) \), medium \( (21 \leq n \leq 40) \), moderately high \( (41 \leq n \leq 1000) \) and high \( (n > 1000) \). The findings result from experiments on 17 different test problems with dimensions ranging from 2 to 100,000. Each of the problems in the list has been tested from four different perspectives. A total of 950 test items were obtained. Moré et al. (1981), Fletcher (1987), Tajadod et al. (2016), and Xiao et al. (2006) provided the test problems.
Table 2. Test problems and dimensions

| ID | Problem                                      | Dimension range               |
|----|----------------------------------------------|-------------------------------|
| 1  | Watson function                              | $3 \leq n \leq 31$            |
| 2  | Extended Rosenbrock                          | $2 \leq n \leq 10,000$, $n$ even |
| 3  | Extended Powell                              | $2 \leq n \leq 100,000$, $n \mod 4 = 0$ |
| 4  | Penalty function 1                           | $2 \leq n \leq 1000$         |
| 5  | Variably dimensioned function                | $2 \leq n \leq 10,000$       |
| 6  | Trigonometric function                       | $2 \leq n \leq 10,000$       |
| 7  | Modified Trigonometric function              | $2 \leq n \leq 100,000$      |
| 8  | Brayden Tridiagonal function                 | $2 \leq n \leq 1000$         |
| 9  | Discrete Boundary value function             | $2 \leq n \leq 1000$         |
| 10 | Oren & Specicato Power function              | $2 \leq n \leq 10,000$       |
| 11 | Full Set of Distinct Eigen-Values Problem    | $2 \leq n \leq 10,000$       |
| 12 | Tridiagonal function                         | $2 \leq n \leq 10,000$       |
| 13 | Wolfe function                               | $2 \leq n \leq 1000$         |
| 14 | Diagonal Rosenbrock’s function               | $2 \leq n \leq 1000$, $n$ even |
| 15 | Generalized Shallow function                 | $2 \leq n \leq 1000$, $n$ even |
| 16 | Powell Singular                              | $n = 10,000$                 |
| 17 | Helical Valley                               | $n = 1000$                   |
| 18 | Extended Freudenstein & Roth function        | $n = 100,000$                |
| 19 | Extended White & Holst function              | $n = 10,000$, $n$ even       |
| 20 | Perturbed Quadratic function                 | $n = 10,000$, $n$ even       |

In (8) (see, Ford & Moghrabi, 1993), (Ford & Moghrabi, 1994, 1996), the new algorithm NMBFGS is compared to Yuan et al. (2017) (see Table 1) and the multi-step BFGS (MSBFGS; see, Ford & Moghrabi, 1993). The tests are run using Yuan’s technique as a benchmark. Table 3 presents a summary of the overall findings. Table 4, 5, 6 and 7 show the results of the small, medium, moderately large, and huge problems, respectively. Because there isn’t enough room to provide the figures for each dimension separately, the findings are summarized by dimension category. Iteration, function/gradient assessments, and total execution times are all included in the scores. The average optimality error is also reported for each dimension category. Figures 1, 2 and 3 provide a diagrammatic summary for each of function/gradient evaluations, iteration count and timing, respectively. The implementation code is C++ carried out on a 64-bit machine with i7-3770, 3.4 GHZ CPU.

For all the methods tested here, the next point $x_{i+1}$ is obtained from $x_i$ by employing a line search algorithm that applies safeguarded cubic interpolation with step-doubling, such that the conditions in (4) are satisfied.

For the standard BFGS method, it is well known that a necessary and sufficient condition for ensuring that the generated updated matrices $\{H_i\}$ are positive definite and hence the computed search direction is downhill, is that $s^T_i y_i > 0$ (Broyden, 1970). By analogy, for the new method MSBFGS, condition (24) is imposed in order to guarantee that $\hat{v}^T_i z_i$ is “sufficiently” positive to circumvent any
### Table 3. Overall iteration, function evaluations count and timing

| Method      | Evaluations | Iterations | Time (sec.) | Scores | Average Optim. error |
|-------------|-------------|------------|-------------|--------|----------------------|
| Yuan et al. | 51,791      | 41,525     | 29,132.91   | 200    | 1.0*E-05             |
|             | 100%        | 100%       | 100%        |        |                      |
| MSBFGS      | 45378       | 34,588     | 23,506.54   | 325    | 1.0*E-05             |
|             | 89.35%      | 84.69%     | 96.61%      |        |                      |
| NMBFGS      | 41607       | 30,715     | 21,741.62   | 425    | 1.0*E-05             |
|             | 81.88%      | 75.12%     | 89.21%      |        |                      |

### Table 4. Iteration, function evaluation count and timing-large problems

| Method      | Evaluations | Iterations | Time (sec.) | Scores | Avg. Optim. error |
|-------------|-------------|------------|-------------|--------|-------------------|
| Yuan et al. | 21,871      | 18,910     | 468.8       | 34     | 1.0*E-04          |
|             | 100%        | 100%       | 100%        |        |                    |
| MSBFGS      | 18042       | 14,199     | 344.703     | 55     | 1.0*E-05          |
|             | 82.41%      | 74.96%     | 66.53%      |        |                    |
| NMBFGS      | 16309       | 13,335     | 327.412     | 84     | 1.0*E-05          |
|             | 74.45%      | 70.36%     | 61.87%      |        |                    |

### Table 5. Iteration, function evaluation count and timing-moderately large problems

| Method      | Evaluations | Iterations | Time (sec.) | Scores | Avg. Optim. error |
|-------------|-------------|------------|-------------|--------|-------------------|
| Yuan et al. | 11,264      | 8509       | 3009.78     | 95     | 1.0*E-05          |
|             | 100%        | 100%       | 100%        |        |                    |
| MSBFGS      | 9855        | 7105       | 2529.59     | 134    | 1.0*E-04          |
|             | 87.38%      | 83.31%     | 83.51%      |        |                    |
| NMBFGS      | 9311        | 6989       | 2375        | 173    | 1.0*E-06          |
|             | 82.51%      | 81.93%     | 78.20%      |        |                    |

### Table 6. Iteration, function evaluations count and timing-medium problems

| Method      | Evaluations | Iterations | Time (sec.) | Scores | Avg. Optim. error |
|-------------|-------------|------------|-------------|--------|-------------------|
| Yuan et al. | 10,311      | 9716       | 16,311.66   | 31     | 1.0*E-04          |
|             | 100%        | 100%       | 100%        |        |                    |
| MSBFGS      | 9270        | 8977       | 11,499.19   | 99     | 1.0*E-05          |
|             | 89.81%      | 92.32%     | 70.32%      |        |                    |
| NMBFGS      | 7909        | 6101       | 10,112.13   | 134    | 1.0*E-05          |
|             | 76.48%      | 62.39%     | 61.68%      |        |                    |
potential numerical instability in the construction of $H_{i+1}$. Should condition (24) fail to hold, the algorithm reverts to using an MSBFGS iteration on that particular instance. An initial scaling is applied to the start inverse Hessian approximation using the methods introduced in Shanno and Phua (1978).

It is generally known that $s_j^T y > 0$ is a necessary and sufficient condition for assuring that the generated updated matrices $\{H_i\}$ are positive definite and thus the computed search direction is

Table 7. Iteration, function evaluation count and timing-small problems

| Method      | Evaluations | Iterations | Time (sec.) | Scores | Avg. Optim. error |
|-------------|-------------|------------|-------------|--------|------------------|
| Yuan et al. | 8345        | 4390       | 9342.67     | 40     | 1.0*E-05         |
| MSBFGS      | 8211        | 4307       | 9133.06     | 37     | 1.0*E-05         |
| NMBFGS      | 8078        | 4290       | 8927.08     | 34     | 1.0*E-05         |

Figure 1. Overall evaluations.

Figure 2. Overall iterations.
downhill for the typical BFGS method (Broyden, 1970). By analogy, condition (24) is applied for the new technique MSBFGS in order to ensure that \( \bar{v}_i^Tz_i \) is “sufficiently” positive to avoid any potential numerical instability in the building of \( H_{i+1} \). If condition (24) fails, the algorithm falls back to utilizing an MSBFGS iteration on that specific instance. Using the approaches proposed in Shanno and Phua (1978), an initial scaling is applied to the start inverse Hessian approximation.

Figure 4 shows the graphical results of the test problems in terms of the most expensive component of the minimization process, the overall number of function/gradient evaluations. That is, we depict the proportion \( \rho(\tau) \) of problems for which each technique is within a factor of \( \tau \) the best number of evaluations for each approach. The algorithm that performs better in terms of evaluations that is within a factor of the best number of function evaluations is shown on the top curve. Figure 4 shows that NMBFGS outperforms the other two methods in terms of number of iterations and evaluations, thus making it the fastest and best performer.

6. Conclusions
In this study, Secant-like approaches based on modifications to the classical Secant relation (1) are investigated, and a new non-Secant method is devised. To increase the quality of the computed
Hessian approximations and, as a result, expedite convergence to a solution for a particular issue, the approach employs nonlinear quantities that integrate data readily available from the previous three iterations. The outperforming algorithm has been justified both theoretically and practically. The primary motivation to exploring alternatives to the standard Secant equation has been to obtain possibly better Hessian approximations that expedite the convergence of quasi-Newton methods. The proposed algorithm’s convergence behaviour has been investigated. The numerical data achieved above have shown promising results, drawing attention to the methods’ potential practical significance. The findings demonstrate that past iteration data, such as function values, gradient points, and step vectors, yield performance gains practically, and hence provide incentives to explore them further.

7. Future work
Future research ought to concentrate on modified Secant methods for various τ-parameter choices in order to investigate the numerical behaviour of these methods’ sensitivity to such choices. It would also be interesting to study these methods’ ability to offer equivalent improvements when used to solve systems of non-linear equations, as well as their convergence characteristics.

Similar approaches are also being investigated, in which the computed search direction \(d_i\) is adjusted before the line search is performed. This strategy looks to be effective. A variation of the technique is also being considered, in which the update of the Hessian approximation is skipped every other iteration (or more frequently). Also, the sensitivity of the algorithmic behaviour to the accuracy of the line search deserves investigation. The numerical performance of such methods when replacing the line search with Trust Region methods is also worth exploring.

**Funding**
The author received no direct funding for this research.

**Author details**
Issam A.R. Moghrabi
E-mail: Moghrabi.I@gust.edu.kw
ORCID ID: http://orcid.org/0000-0002-4517-7630
Department of Accounting and M.I.S., College of Business Administration, Gulf University for Science and Technology, Kuwait.

**Disclosure statement**
No potential conflict of interest was reported by the author(s).

**Citation information**
Cite this article as: A non-Secant quasi-Newton Method for Unconstrained Nonlinear Optimization, Issam A.R. Moghrabi, Cogent Engineering (2022), 9: 2018929.

**References**
Al-Baali, M. (1985). New property and global convergence of the Fletcher–Reeves method with inexact line searches. *IMA Journal of Numerical Analysis*, 5(1), 122–124. https://doi.org/10.1093/imanum/5.1.121
Awasthi, A., & Omrani, H. (2019). A goal-oriented approach based on fuzzy axiomatic design for sustainable mobility project selection. *International Journal of Systems Science: Operations & Logistics*, 6(1), 86–98. https://doi.org/10.1080/233052674.2018.1435834
Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms - Part 2: The new algorithm. *IMA Journal of Applied Mathematics*, 6(3), 222–231. https://doi.org/10.1093/imamat/6.3.222
Dai, Y., Yuan, J., & Yuan, Y. (2002). Modified two-point step size Gradient methods for unconstrained optimization. *Computational Optimization and Applications*, 22(1), 103–109. https://doi.org/10.1023/A:1016838619611
Dennis, J. E., & Schnabel, R. B. (1979). Least change Secant updates for quasi-Newton methods. *SIAM Review*, 21(4), 443–459. https://doi.org/10.1137/1021091
Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3), 317–322. https://doi.org/10.1093/comjnl/13.3.317
Fletcher, R. (1987). *Practical Methods of Optimization* (second ed.). Wiley.
Ford, J. A., & Moghrabi, I. A. R. (1993). Modified parameter choices for multi-step quasi-Newton methods. *Optimization Methods & Software*, 2(3–4), 357–370. https://doi.org/10.1080/10556789308805550
Ford, J. A., & Moghrabi, I. A. R. (1994). Multi-step quasi-Newton methods for optimization. *Journal of Computational and Applied Mathematics*, 50(1–3), 305–324. https://doi.org/10.1016/0771-050X(94)90309-3
Ford, J. A., & Moghrabi, I. A. R. (1996). Using function-values in multi-step quasi-Newton methods. *Journal of Computational and Applied Mathematics*, 66(1–2), 201–211. https://doi.org/10.1016/0377-0427(95)00178-6
Gharaei, A., Hoseini, S. S., & Karimi, M. (2020). Modelling And optimal lot-sizing of the replenishments in constrained, multi-product and bi-objective EPQ models with defective products: Generalised Cross
Decomposition. *International Journal of Systems Science: Operations & Logistics*, 7(3), 262–274. https://doi.org/10.1080/23302674.2019.1574364

Gharaei, A., Naderi, B., & Mohammadi, M. (2015). Optimization of rewards in single machine scheduling in the rewards-driven systems. *Management Science Letters*, 5(6), 629–638.

Gharaei, A., Naderi, B., & Mohammadi, M. (2019). An integrated multi-product, multi-buyer supply chain under penalty, green, and quality control policies and a vendor managed inventory with consignment stock agreement: The outer approximation with equality relaxation and augmented penalty algorithm. *Applied Mathematical Modelling*, 69, 223–254. https://doi.org/10.1016/j.apm.2018.11.035

Gharaei, A., Shekarabi, S. A., Karimi, M., Pourjavad, E., & Amjadian, A. (2021). An integrated stochastic EPQ model under quality and green policies: Generalised cross decomposition under the separability approach. *International Journal of Systems Science: Operations & Logistics*, 8(2), 119–131. https://doi.org/10.1080/23302674.2019.1656296

Hassan, B., & Ghada, M. (2020). A new quasi-newton equation on the gradient methods for optimization minimization problem. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 737–744. https://doi.org/10.11591/ijeecs.v19i2.pp737-744

Hassan, B., & Mohammed, W. T. (2020). A new variants of quasi-newton equation based on the quadratic function for unconstrained optimization. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2), 701–708. https://doi.org/10.11591/ijeecs.v19i2.pp701-708

Hassan, B. (2019). A new quasi-Newton Updating formula based on the new quasi-Newton equation. *Numerical Algebra Control and Optimization*, 4(1), 63–72.

Hassan, B. (2020). A new type of quasi-Newton updating formulas based on the new quasi-Newton equation. *SIAM Journal Numerical Algebra, Control, and Optimization*, 10(2), 227–235. https://doi.org/10.3934/naco.2019069

Jin, Q., & Mokhtari, A. (2021). Non-asymptotic super linear convergence of standard quasi-Newton methods. *Journal of Strategic Decision Sciences*, 121, https://archiv.pdf/2003.13607v1.pdf

Li, D., & Fukushima, M. (2001). A modified BFGS method and its global convergence in nonconvex minimization. *Journal of Computational and Applied Mathematics*, 129(1–2), 15–35. https://doi.org/10.1016/S0377-0427(00)00540-9

Moghribi, J. A. R. (2017). Implicit extra-update multi-step quasi-Newton methods. *International Journal of Operational Research*, 28(2), 69–81. https://doi.org/10.1504/IJOR.2017.081472

Moré, J. J., Garbow, B. S., & Hillstrom, K. E. (1981). Testing unconstrained optimization software. *ACM Transactions on Mathematical Software*, 7(1), 17–41. https://doi.org/10.1145/355934.355936

Ortiz, F., Simpson, J. R., Pignottiello, J. J., & Heredia-Langner, A. (2004). A genetic algorithm approach to multiple-response optimization. *Journal of Quality Technology*, 36(4), 432–450. https://doi.org/10.1080/00224065.2004.11980289

Soyyodi, R., & Awasthi, A. (2018). A simulation-based optimisation approach for identifying key determinants for sustainable transportation planning. *International Journal of Systems Science: Operations & Logistics*, 5(2), 161–174. https://doi.org/10.1080/23302674.2016.1244301

Shanno, D. F., & Phua, K. H. (1978). Matrix conditioning and nonlinear optimization. *Mathematical Programming*, 14(1), 149–160. https://doi.org/10.1007/BF01588962

Shanno, D. F. (1978). On the convergence of a new conjugate gradient algorithm. *SIAM Journal on Numerical Analysis*, 15(6), 1247–1257. https://doi.org/10.1137/0715085

Tejador, M., Abedini, M., Rateegari, A., & Mobin, M. (2016). A comparison of multi-criteria decision making approaches for maintenance strategy selection (a case study). *International Journal of Strategic Decision Sciences*, 7(3), 51–56. https://doi.org/10.1016/j.ijstds.2016.07.003

Wei, Z., Li, G., & Qi, L. (2000). The superlinear convergence of a modified BFGS- type method for unconstrained optimization. *Computational Optimization and Applications*, 29(3), 315–332. https://doi.org/10.1023/B:COAP.000004184.25410.39

Wei, Z., Li, G., & Qi, L. (2000). New quasi-Newton methods for unconstrained optimization problems. *Mathematics Program Applied Mathematics and Computation*, 175(2), 1156–1188. https://doi.org/10.1016/j.amc.2005.08.027

Xiao, Y. H., Wei, Z. X., & Zhang, L. (2006). A modified BFGS method without line searches for nonconvex unconstrained optimization. *Advances in Theoretical and Applied Mathematics*, 1(2), 149–162. https://www.rpublication.com/Volume/atamv1n2.htm

Yuan, G., Sheng, Z., Wang, B., Hu, W., & Li, C. (2018). The global convergence of a modified BFGS method for nonconvex functions. *Journal of Computational and Applied Mathematics*, 327(1), 274–294. https://doi.org/10.1016/j.cam.2017.05.030

Yuan, G., Wei, Z., & Lu, X. (2017). Global convergence of BFGS and PRP methods under a modified weak Wolfe–Powell line search. *Applied Mathematical Modelling*, 47(2), 811–825. https://doi.org/10.1016/j.apm.2017.02.008

Yuan, G., Wei, Z., & Wu, Y. (2010). Modified limited memory BFGS method with non-monotone line search for unconstrained optimization. *Journal of the Korean Mathematical Society*, 47(4), 767–788. https://doi.org/10.4134/JKMS.2010.47.4.767

Yuan, G., & Wei, Z. (2010). Convex minimizations. *Computational Optimization and Applications*, 47(1), 237–255. https://doi.org/10.1007/s10589-008-9219-0

Zhang, J. Z., Deng, N. Y., & Chen, L. H. (1999). Quasi-Newton equation and related methods for unconstrained optimization. *Journal of Optimization Theory and Applications*, 102(1), 147–167. https://doi.org/10.1023/A:1021898630001
