Lim and Leong *Journal of Inequalities and Applications* (2024) 2024:155
https://doi.org/10.1186/s13660-024-03240-z

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

# Memoryless quasi-Newton-type methods via some weak secant relations for large-scale unconstrained optimization

Keat Hee Lim[1] and Wah June Leong[2*]

*Correspondence:
leongwj@upm.edu.my
[2]Institute for Mathematical Research, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia
Full list of author information is available at the end of the article

**Abstract**

This article presents two variants of memoryless quasi-Newton methods with backtracking line search for large-scale unconstrained minimization. These updating methods are derived by means of a least-change updating strategy subjected to some weaker form of secant relation obtained by projecting the secant equation onto the search direction. In such a setting, the search direction can be computed without the need of calculation and storage of matrices. We establish the convergence properties for these methods, and their performance is tested on a large set of test functions by comparing with standard methods of similar computational cost and storage requirement. Our numerical results indicate that significant improvement has been achieved with respect to iteration counts and number of function evaluations.

**Keywords:** Large-scale optimization; Quasi-Newton-type methods; Weak secant relations; Least-change updating scheme; Armijo line search

## 1 Introduction

We introduce two variants of memoryless quasi-Newton methods for the unconstrained optimization problem

$$\min f(x), \tag{1.1}$$

where $f$ is a continuously twice differentiable function of $n$ variables. Like most quasi-Newton methods, these variants undergo iterative algorithm in the following form:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, 2, \ldots, \tag{1.2}$$

and the search direction $d_k$ used in quasi-Newton-type methods is generally as follows:

$$d_k = -H_k g_k,$$

where $g_k$ is the gradient of $f$ at $x_k$, the $k$th approximation to the solution, $H_k$ is some approximation to the inverse Hessian, $[G(x_k)]^{-1}$ and the scalar $\alpha_k$ is a step length parameter,



Springer

which is obtained through an exact line search [6] or an inexact line search. The inexact line search strategy ensures that the step length $\alpha_k$ gives a sufficient decrease in the objective function $f$ along the descent direction $d_k$, if the following inequality holds (Armijo [4]):

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k d_k^T g_k, \tag{1.3}$$

for some constant $c_1 \in (0, 1)$. The details of this line search algorithm will be given in Sect. 3. Traditionally, the $H_k$ in quasi-Newton methods obeys the secant equation

$$H_{k+1} y_k = s_k, \tag{1.4}$$

where $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$. The general idea of quasi-Newton-type methods is finding the best approximation to the Hessian or its inverse, without explicitly evaluating it at every iteration. Such updating formulae can be generated via variational formulation as follows:

$$\min \|H_{k+1} - H_k\|$$

subject to

$$H_{k+1} y_k = s_k,$$
$$H_{k+1} = H_{k+1}^T,$$

where $\|\cdot\|$ is some matrix norm. The first constraint is included to preserve the symmetry of $H$. A different setup of the minimization problem above will generate a different solution, and hence different quasi-Newton updating formulae (Greenstadt [11], Nocedal and Wright [19]). On the other hand, Dennis and Wolkowicz [9] introduced a weaker form of the secant equation, given by

$$y_k^T H_{k+1} y_k = y_k^T s_k, \tag{1.5}$$

in the derivation of some symmetric rank-one methods. One can see that (1.5) is the scalar-projection of the secant Eq. (1.4) along $y_k$. The motivation is that, by relaxing the secant relation, it encourages a higher degree of freedom in incorporating more desirable properties (such as positive-definiteness of the updating matrix) into the updating formula. Motivated by the approach of Dennis and Wolkowicz [9] (see also Nazareth [15], Sim et al. [19]), the updating formulae proposed in this article are constructed based on the least-change updating strategy via the nonsecant Eq. (1.4) above, as well as a variant of it. The other variant is obtained by performing the same scalar-projection, but along a different vector, namely $s_k$:

$$s_k^T H_{k+1} y_k = s_k^T s_k, \tag{1.6}$$

where $s_k = \alpha_k d_k$ also represents the search direction. We shall give the full derivations of these memoryless formulae based on the weak secant equations in the following section.

The rest of this paper is organized as follows: Sect. 2 presents two types of memoryless updating methods via the least-change updating strategy. Section 3 discusses the convergence properties of these new algorithms, and numerical results are reported in Sect. 4.

## 2 Weak secant updates via the least-change updating strategy

Suppose that a new approximate inverse Hessian is updated from the previous approximation as

$$H_{k+1} = H_k + E_k,$$

where $E_k$ is the correction between two successive approximations of the inverse Hessian $H$. There are choices to achieve the "best" correction in some sense, one of them is looking for the smallest correction $E_k$, preventing the elements of $H$ from getting too large and causing undesirable numerical instability. One could achieve this by minimizing the correction with respect to some norm. By taking the Frobenius norm, we have

$$\|E_k\|_F^2 = \operatorname{tr}(E_k E_k^T), \tag{2.1}$$

where tr represents the trace of a matrix. The minimization of the norm (2.1) can be written as the following variational problem:

$$\min \frac{1}{2} \operatorname{tr}(E_k E_k^T) \tag{2.2}$$

subject to (1.5) with (2.1) and the symmetry condition,

$$E_k = E_k^T. \tag{2.3}$$

In a memoryless setting, the previous approximation $H_k$ is not stored in each iteration. Instead, the update is restarted, i.e., $H_k$ is replaced by a multiple of the identity matrix, $\theta_k I$, where the scalar $\theta_k$ can be further chosen to incorporate $H_{k+1}$ with a desirable property such as hereditary positive-definiteness. This implies that we restrict our update to the following form:

$$H_{k+1} = \theta_k I + E_k. \tag{2.4}$$

In doing so, one can avoid the needs of storing matrices throughout the iterations, and thus renders a method with only $O(n)$ storage requirement. For the rest of this derivation, we shall ignore the subscript $k$. We solve the constrained minimization problem above using Lagrange multipliers, see Greenstadt [11]. The Lagrangian function for (2.2) subject to (1.5) and (2.3), where $H_{k+1}$ is defined by (2.4), can be written as

$$L(E, \omega, \lambda) = \frac{1}{2} \operatorname{tr}\left(EE^T\right) + \omega\left(\theta y^T y + y^T E y - y^T s\right) + \lambda^T (E^T - E)\lambda, \tag{2.5}$$

where $\omega$ and $\lambda$ are the Lagrange multipliers for (1.5) and (2.3), respectively.

It can be observed that the last term of $L(E, \omega, \lambda)$ is expressible as

$$\lambda^T\left(E^T - E\right)\lambda = \lambda^T E^T \lambda - \lambda^T E \lambda = \operatorname{tr}(E^T \lambda\lambda^T) - \operatorname{tr}(E\lambda\lambda^T). \tag{2.6}$$

Recall that

$$\frac{\partial}{\partial E}\left[\operatorname{tr}\left(E\lambda\lambda^T\right)\right] = \lambda\lambda^T \tag{2.7}$$

and

$$\frac{\partial}{\partial E}\left[\operatorname{tr}\left(E^T\lambda\lambda^T\right)\right] = \lambda\lambda^T. \tag{2.8}$$

Hence, differentiating (2.5) we obtain

$$\frac{\partial L}{\partial E} = E + \omega yy^T + \lambda\lambda^T - \lambda\lambda^T. \tag{2.9}$$

Setting the partial derivative to zero yields

$$E = -\omega yy^T. \tag{2.10}$$

Note that (2.10) satisfies the symmetry condition (2.4). Substituting Eq. (2.10) into the weak secant relation (1.5) gives

$$\omega = \frac{\theta y^T y - y^T s}{(y^T y)^2}. \tag{2.11}$$

Replacing $\omega$ in (2.10) with (2.11), we obtain the formula for $E$ as follows:

$$E = \frac{y^T s - \theta y^T y}{(y^T y)^2} yy^T.$$

Thus, the approximate inverse Hessian is updated as follows:

$$H_{k+1} = \theta_k I + \frac{y_k^T s_k - \theta_k y_k^T y_k}{(y_k^T y_k)^2} y_k y_k^T, \tag{2.12}$$

where the choice of $\theta_k$ will be explained later.

Alternatively, if the weak secant Eq. (1.6) is considered, we can solve the constrained minimization problem in a similar fashion and arrive at the following result:

$$E = \frac{s^T s - \theta s^T y}{s^T s y^T y + (s^T y)^2}\left(sy^T + ys^T\right),$$

Hence, the second updating formula for the inverse Hessian approximation is as follows:

$$H_{k+1} = \theta_k I + \frac{s_k^T s_k - \theta s_k^T y_k}{s_k^T s_k y_k^T y_k + \left(s_k^T y_k\right)^2}\left(s_k y_k^T + y_k s_k^T\right). \tag{2.13}$$

It is worth mentioning that, using different norms, such as the Frobenius norm with weighting matrix $W$, namely $\|E\|_W^2 = \operatorname{tr}(WEWE^T)$,

may lead to distinct updates and convergence behaviors, potentially improving the stability or efficiency in certain problems. This approach could open up new possibilities for designing memoryless updates tailored to specific applications or problem structures.

Note that none of the updating formulae (2.12) and (2.13) can preserve positive-definiteness, but the property of hereditary positive-definiteness can be incorporated into the updates with a suitable choice of $\theta_k$. We shall again omit the subscript $k$ for simplicity and denote $H_{k+1} = H_+$ for the following derivation process for $\theta_k$. To ensure positive-definiteness for $H_+$, one possible way is to have $\theta$ satisfying $y^T s - \theta y^T y \geq 0$, i.e.,

$$\theta \leq \frac{y^T s}{y^T y}. \tag{2.14}$$

If $\frac{y^T s}{y^T y} \geq 1$, one can simply take $\theta = 1$. Otherwise, if $\frac{y^T s}{y^T y} < 1$, there are many possible choices for $\theta$ that can satisfy (2.14). To choose the "best" $\theta$ in some sense, one could consider a value that can improve the overall condition number of $H_+$. In general, having a well-conditioned updating matrix will yield a more numerically stable updating scheme, and thus improve convergence speed numerically. It can be observed that the trace of (2.12) can be written as

$$\operatorname{tr}(H_+) = \operatorname{tr}(\theta I) + \operatorname{tr}\left(\gamma y y^T\right) = n\theta + \gamma y^T y,$$

where $\gamma = \frac{y^T s - \theta y^T y}{(y^T y)^2}$ and $\operatorname{tr}\left(y y^T\right) = y^T y$. Since $y \in \boldsymbol{R}^n$, we can assume that there exist $n-1$ vectors $p_i$, $i = 1, 2, \ldots, n-1$, that are orthogonal to vector $y$, and this implies that the set $\left\{y, p_1, \ldots, p_{n-1}\right\}$ forms a basis for $\boldsymbol{R}^n$. By multiplying these vectors with (2.12), we have the following set of equations:

$$H_+ p_i = \theta p_i, \text{ for } i = 1, 2, \ldots, n-1.$$

This suggests that $H_+$has $(n-1)$ repeated eigenvalues, namely $\theta$ and the remaining eigenvalue denoted by $\rho$. Since the trace is the sum of all eigenvalues, it leads to

$$\operatorname{tr}(H_+) = n\theta + \gamma y^T y = (n-1)\theta + \rho$$

and yields

$$\rho = \frac{y^T s}{y^T y}. \tag{2.15}$$

Since we require $y^T s - \theta y^T y \geq 0$, the latter implies that

$$\theta \leq \frac{y^T s}{y^T y} = \rho.$$

Then, $\rho$ would be the largest eigenvalue for matrix $H_+$. Minimizing the condition number with respect to $\theta$ and subject to the positive definiteness condition, we arrive at the following minimization problem:

$$\min_{\theta} \frac{\rho}{\theta} \tag{2.16}$$

subject to

$$y^T s - \theta y^T y \geq 0, \tag{2.17}$$

$$\theta > 0. \tag{2.18}$$

The associated Lagrangian function for (2.16) to (2.18) is then given as follows:

$$L(\theta,\xi) = \frac{\rho}{\theta} + \xi(\theta y^T y - y^T s), \theta > 0, \tag{2.19}$$

where $\xi$ is the Lagrange multiplier associated with (2.17). Differentiating (2.19) with respect to $\theta$ and using the Karush–Kuhn–Tucker (KKT) (see [5, 12, 16]) conditions, we obtain the following:

$$\frac{\partial L(\theta,\xi)}{\partial\theta} = -\frac{\rho}{\theta^2} + \xi y^T y = 0, \tag{2.20}$$

$$\xi\left(\theta y^T y - y^T s\right) = 0, \tag{2.21}$$

$$\theta y^T y - y^T s \leq 0, \tag{2.22}$$

$$\xi \geq 0. \tag{2.23}$$

If we consider $\xi = 0$, then (2.20) becomes

$$\frac{\rho}{\theta^2} = 0,$$

indicating that $\rho = \frac{y^T s}{y^T y} = 0$, which is not admissible as $y^T s \neq 0$. Therefore, $\xi \neq 0$ and (2.21) gives

$$\theta y^T y - y^T s = 0 \Rightarrow \theta = \frac{y^T s}{y^T y}, \tag{2.24}$$

where it also satisfies the inequality (2.14). Combining with the case when $\frac{y^T s}{y^T y} \geq 1$, we can choose

$$\theta_k = \min\left(1, \frac{y_k^T s_k}{y_k^T y_k}\right). \tag{2.25}$$

Similarly, suppose we want to preserve the positive-definiteness of the second updating formula (2.13), that is, $s^T s - \theta s^T y \geq 0$, then we need the following condition:

$$\theta \leq \frac{s^T s}{s^T y}. \tag{2.26}$$

Suppose that $s$ and $y$ are not parallel. Then we can assume that there exist $n-2$ vectors $p_i$, $i = 1, 2, \ldots, n-2$, that are orthogonal to the subspace spanned by $s$ and $y$. By multiplying these vectors with (2.13), we have

$$H_+ p_i = \theta p_i, \text{ for } i = 1, 2, \ldots, n-2.$$

This implies that the $H_+$ given by (2.13) has $(n-2)$ repeated eigenvalues $\theta$, and the remaining two eigenvalues are denoted by $\rho_1$ and $\rho_2$. In a similar manner, we obtain

$$tr(H_+) = n\theta + 2\overline{\gamma} s^T y = (n-2)\theta + \rho_1 + \rho_2,$$

where $\overline{\gamma} = \frac{s^T s - \theta s^T y}{s^T s y^T y + (s^T y)^2}$, and it follows that $\rho_1 + \rho_2 = 2(\theta + \overline{\gamma} s^T y)$. For $\overline{\gamma} s^T y > 0$, it implies that either $\rho_1 > \theta$ or $\rho_2 > \theta$ (or both), and thus one of them would be the largest eigenvalue of $H_+$. Without loss of generality, let us assume $\rho_1$ is the largest eigenvalue of $H_+$ given by (2.13). To obtain $\theta$ such that the condition number of $H_+$ is sufficiently close to 1 while maintaining positive-definiteness, we shall consider the following minimization problem:

$$\min_{\theta} \frac{\rho_1}{\theta} \tag{2.27}$$

subject to

$$s^T s - \theta s^T y \geq 0, \tag{2.28}$$

$$\theta > 0. \tag{2.29}$$

Again the KKT conditions will lead to $\theta = \frac{s^T s}{s^T y}$ and, together with the case $\frac{s^T s}{s^T y} \geq 1$ where we shall take $\theta = 1$, we have

$$\theta_k = \min\left(1, \frac{s_k^T s_k}{s_k^T y_k}\right). \tag{2.30}$$

Finally, our two updates $H_{k+1}$ with their corresponding $\theta_k$ are given respectively by

$$H_{k+1}^{QN1} = \theta_k I + \frac{y_k^T s_k - \theta_k y_k^T y_k}{(y_k^T y_k)^2} y_k y_k^T, \text{ where } \theta_k = \min\left(1, \frac{y_k^T s_k}{y_k^T y_k}\right), \tag{2.31}$$

and

$$H_{k+1}^{QN2} = \theta_k I + \frac{s_k^T s_k - \theta_k s_k^T y_k}{s_k^T s_k y_k^T y_k + (s_k^T y_k)^2} \left(s_k y_k^T + y_k s_k^T\right), \text{ where } \theta_k = \min\left(1, \frac{s_k^T s_k}{s_k^T y_k}\right). \tag{2.32}$$

To obtain the search direction, $d_{k+1} = -H_{k+1} g_{k+1}$, for $k \geq 0$, it is not necessary to compute and store $H_{k+1}$ explicitly. Instead, we can obtain the product $H_{k+1} g_{k+1}$ directly as follows:

$$d_{k+1}^{QN1} = -H_{k+1}^{QN1} g_{k+1} = -\theta_k g_{k+1} - \left(\frac{y_k^T s_k - \theta_k y_k^T y_k}{(y_k^T y_k)^2} y_k^T g_{k+1}\right) y_k \tag{2.33}$$

and

$$d_{k+1}^{QN2} = -H_{k+1}^{QN2} g_{k+1} = -\theta_k g_{k+1} - \frac{s_k^T s_k - \theta_k s_k^T y_k}{s_k^T s_k y_k^T y_k + \left(s_k^T y_k\right)^2} \left(s_k y_k^T g_{k+1} + y_k s_k^T g_{k+1}\right). \tag{2.34}$$

The first search direction (2.33) involves a linear combination of two vectors, namely $g_{k+1}$ and $y_k$, while the second one (2.34) is denoted by a linear combination of three vectors, i.e., $g_{k+1}$, $s_k$, and $y_k$. Both formulae require only $3n$ storage units and a couple of scalar-products to compute. We can now present the full algorithms for the memoryless formulae proposed, incorporating a backtracking line search condition to preserve monotonicity in function values. By a backtracking line search, we mean an algorithm of the following form for computing the step length $\alpha$:

**Algorithm 2.1** (Backtracking line search with Armijo [4] condition)

Step 1. *Set* $\alpha = 1$.

Step 2. *Test the Armijo condition* (1.3), *where* $c_1 \in (0,\ 1)$.

If the condition (1.3) is not satisfied, choose a new $\alpha := \tau\alpha$ where $0 < \tau < 1$ and go to step 2. Else if the condition (1.3) is satisfied, set $\alpha_k = \alpha$.

Using the updating formulae (2.31) and (2.32), along with the backtracking line search above, the full algorithms are as follows:

**QNWS1 algorithm**

Step 1. *Given an initial guess* $x_0$, *set a positive definite matrix* $H_0 = I$. *Set* $k = 0$ *and* $d_0 = -g_0$, *then compute the norm of the gradient*, $\|g_k\|$.

Step 2. *If* $\|g_k\| \leq \varepsilon$ *or* $k \geq 1000$ *then stop. Otherwise, perform Algorithm* 2.1 *to choose a suitable* $\alpha$.

Step 3. *Compute* $x_{k+1} = x_k + \alpha_k d_k$ *and* $g_{k+1}$, *where* $\alpha_k$ *is the latest step length obtained in step* 2.

Step 4. *Compute* $s_k = x_{k+1} - x_k$ *and* $y_k = g_{k+1} - g_k$, *respectively. Then, update* $d_{k+1}$ *by* (2.33).

Step 6. *Set* $k := k + 1$ *and return to step* 2.

**QNWS2 algorithm** *Similar to QNWS*1 *algorithm except that the search direction* $d_k$ *in step* 4 *is computed by* (2.34).

Observe that due to these modifications, these algorithms involve only vectors. As a result, the storage and number of evaluations at each iteration are greatly reduced. We shall study the convergence properties of these algorithms for convex problems in the next section.

## 3 Convergence analysis

To investigate the convergence properties, we first present the following assumption for the objective function $f$. The Hessian matrix of $f$ will be denoted by $G$. The starting point for the algorithm is $x_0$, and we define the level set $D = \{x \in \boldsymbol{R}^n \mid f(x) \leq f(x_0)\}$.

**Assumption 3.1**

1. The objective function $f$ is twice continuously differentiable and the level set $D$ is convex.
2. The gradient of $f$ is Lipschitz continuous, i.e., there exists a positive constant $M > 0$ such that

$$\|g(x) - g(y)\| \leq M \|x - y\|, \forall x, y \in D. \tag{3.1}$$

3. The Hessian matrix $G$ is positive definite and their eigenvalues are bounded away from 0, i.e., there exists a positive constant $m > 0$ such that

$$z^T G(x) z \geq m \|z\|^2 \tag{3.2}$$

for all $x,\ z \in D$.

Note that the assumptions above imply that $f$ has at least one (local) minimizer $x^*$ in $D$. If we define the average Hessian of the function along $s_k$ as

$$\overline{G}_k(x) = \int_0^1 G\left(x_k + \tau s_k\right) d\tau,$$

then, by the mean value theorem, we have

$$y_k = \overline{G}_k s_k. \tag{3.3}$$

Premultiplying (3.3) with vector $s_k$ and using (3.2), one can obtain

$$s_k^T y_k \geq m \left\| s_k \right\|^2.$$

Moreover, using Assumption 3.1.2, we have

$$s_k^T y_k = s_k^T \left(g_{k+1} - g_k\right) \leq \left\| s_k \right\| \left\| g_{k+1} - g_k \right\| \leq M \left\| s_k \right\|^2,$$

and, together with the lower bound, get

$$m \left\| s_k \right\|^2 \leq s_k^T y_k \leq M \left\| s_k \right\|^2. \tag{3.4}$$

On the other hand, condition (3.2) and the inverse of (3.3) imply

$$y_k^T s_k = y_k^T \overline{G}_k^{-1} y_k \leq \frac{1}{m} \left\| y \right\|^2,$$

and from (3.1)– (3.2), we have $y_k^T s_k \geq m \left\| s_k \right\|^2 \geq \frac{m}{M} \left\| y_k \right\|^2$. Hence, we can finally obtain

$$\frac{m}{M} \left\| y_k \right\|^2 \leq y_k^T s_k \leq \frac{1}{m} \left\| y_k \right\|^2. \tag{3.5}$$

To proceed further, we introduce the following lemmas that relate backtracking line search with the convergence to a minimizer $x^*$.

**Lemma 3.1** *Under Assumption 3.1, there exist positive constant $\eta_1$ and $\eta_2$ such that, for any $x_k$ and $d_k$ with $g_k^T d_k < 0$, the steplength $\alpha_k$ produced by backtracking line search with Armijo condition (1.3) will satisfy either*

$$f\left(x_k + \alpha_k d_k\right) - f\left(x_k\right) \leq -\eta_1 \frac{(g_k^T d_k)^2}{\left\| d_k \right\|^2} \tag{3.6}$$

*or*

$$f\left(x_k + \alpha_k d_k\right) - f\left(x_k\right) \leq \eta_2 g_k^T d_k, \tag{3.7}$$

*where $\eta_1$ and $\eta_2$ are positive constants.*

*Proof* See Byrd and Nocedal [5]. □

To establish the convergence of our proposed methods, we present the following result where the boundedness and conditions from Assumption 3.1 and Lemma 3.1, respectively, will guarantee the necessary condition for optimality.

**Lemma 3.2** *Suppose that $f(x)$ satisfies Assumption 3.1. Consider a line search method, namely*

$$x_{k+1} = x_k + \alpha_k d_k,$$

*where $\alpha_k$ is obtained by the backtracking line search with Armijo condition and $d_k = -H_k g_k$, where $H_k$ is chosen such that*

$$c_1 \|z\|^2 \leq z^T H_k z \leq c_2 \|z\|^2, \tag{3.8}$$

*for any nonzero vector $z \in R^n$ and some positive constants $c_1$ and $c_2$. Then,*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{3.9}$$

*Proof* If

$$d_k = -H_k g_k, \tag{3.10}$$

then the first condition (3.6) becomes

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta_1 \frac{\left(g_k^T H_k g_k\right)^2}{g_k^T H_k^2 g_k}. \tag{3.11}$$

Observe that when $z = g_k$, (3.8) becomes

$$c_1 \|g_k\|^2 \leq g_k^T H_k g_k \leq c_2 \|g_k\|^2.$$

On the other hand, from (3.8) we have

$$\begin{aligned} c_1^2 \|z\|^2 &\leq z^T H_k^2 z \leq c_2^2 \|z\|^2, \\ c_1^2 \|g_k\|^2 &\leq g_k^T H_k^2 g_k \leq c_2^2 \|g_k\|^2. \end{aligned} \tag{3.12}$$

Then, (3.11) becomes

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -\eta_1 \frac{c_1^2 \|g_k\|^4}{c_2^2 \|g_k\|^2} = -\frac{\eta_1 c_1^2}{c_2^2} \|g_k\|^2. \tag{3.13}$$

Since the Armijo condition implies that $f(x_{k+1}) < f(x_k)$ and $f$ is bounded below, taking the limit $k \to \infty$ on both sides of (3.13) gives

$$0 = \lim_{k\to\infty} f(x_k + \alpha_k d_k) - f(x_k) \leq -\frac{\eta_1 c_1^2}{c_2^2} \lim_{k\to\infty} \|g_k\|^2.$$

As $\eta_1$, $c_1^2$, and $c_2^2$ are positive constants, the right-hand limit becomes

$$\lim_{k\to\infty} \|g_k\| = 0.$$

Similarly, using (3.10) and (3.8), the second condition (3.7) can be rewritten as

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\eta_2 g_k^T H_k g_k \le -\eta_2 c_1 \|g_k\|^2. \tag{3.14}$$

Taking the limit as $k \to \infty$ on both sides, with $f$ being bounded below and both $\eta_2$, $c_1 >$ 0, one obtains

$$0 = \lim_{k\to\infty} f(x_k + \alpha_k d_k) - f(x_k) \le -\eta_2 c_1 \lim_{k\to\infty} \|g_k\|^2,$$

which also implies that

$$\lim_{k\to\infty} \|g_k\| = 0.$$

□

Using both Assumption 3.1 and Lemma 3.2, we can now show the convergence for the proposed algorithms.

**Theorem 3.1** *Let $x_0$ be the starting point for which $f$ satisfies Assumption 3.1. Suppose that $\{x_{k+1}\}$ is generated by (1.2), with the search direction $d_k = -H_k g_k$, under the line search conditions (3.6) and (3.7) for all $k \ge 0$. Then for $H_0 = I$, the memoryless QNWS1 algorithm generates iterations $x_k$ which converge to the minimizer $x^*$.*

*Proof* Suppose we update the matrix at every iteration according to (2.32), then for a nonzero vector $z \in \boldsymbol{R}^n$,

$$z^T H_0 z = \|z\|^2, \tag{3.15}$$

since $H_0 = I$. We shall show that $H_{k+1}$ along any vector $z$ is bounded both below and above. For any $k \ge 0$, we have two possible updates as in (2.31), which are

*Case I*: If $H_{k+1} = \frac{y_k^T s_k}{y_k^T y_k} I$, when $\theta_k = \frac{y_k^T s_k}{y_k^T y_k}$, then

$$z^T H_{k+1} z = \frac{y_k^T s_k}{y_k^T y_k} \|z\|^2 = \frac{y_k^T s_k}{\|y_k\|^2} \|z\|^2, \tag{3.16}$$

which, using (3.5), becomes

$$\frac{m}{M} \|z\|^2 \le z^T H_{k+1} z \le \frac{1}{m} \|z\|^2, \tag{3.17}$$

clearly showing that the sequence is bounded.

*Case II*: If $H_{k+1} = I + \frac{y_k^T s_k - y_k^T y_k}{(y_k^T y_k)^2} y_k y_k^T$, when $\theta_k = 1$, then

$$z^T H_{k+1} z = \|z\|^2 + \frac{y_k^T s_k - y_k^T y_k}{\left(y_k^T y_k\right)^2} \left(z^T y_k\right)^2, \tag{3.18}$$

which is positive as $y_k^T s_k - y_k^T y_k > 0$. A lower bound of (3.18) can be obtained by simply removing the fraction term $\frac{y_k^T s_k - y_k^T y_k}{\left(y_k^T y_k\right)^2} \left(z^T y_k\right)^2$, that is,

$$z^T H_{k+1} z \geq \|z\|^2. \tag{3.19}$$

For the upper bound, using the Cauchy–Schwarz inequality in (3.18), we have $(z^T y_k)^2 \leq \|z\|^2 \left\|y_k\right\|^2$ which, with (3.5), gives an upper bound as follows:

$$\begin{aligned} z^T H_{k+1} z &= \|z\|^2 + \frac{y_k^T s_k - y_k^T y_k}{\left(y_k^T y_k\right)^2} \left(z^T y_k\right)^2 \\ &\leq \|z\|^2 + \frac{\frac{1}{m} \left\|y_k\right\|^2 - \left\|y_k\right\|^2}{\left\|y_k\right\|^4} \|z\|^2 \left\|y_k\right\|^2 \\ &= \left(1 + \frac{1}{m} - 1\right) \|z\|^2 \\ &= \frac{1}{m} \|z\|^2 \end{aligned} \tag{3.20}$$

since $y_k^T y_k = \left\|y_k\right\|^2$ and $\left(y_k^T y_k\right)^2 = \left\|y_k\right\|^4$. Combining (3.15), (3.17), (3.19), and (3.20), we can generalize the bounds as follows:

$$\beta_1 \|z\|^2 \leq z^T H_{k+1} z \leq \beta_2 \|z\|^2, \tag{3.21}$$

where $\beta_1 = \min\left\{\frac{m}{M}, 1\right\}$ and $\beta_2 = \max\left\{\frac{1}{m}, 1\right\}$. The lower and upper bounds from (3.21), together with Lemma 3.2, will imply that $\lim_{k\to\infty} \left\|g_k\right\| = 0$. This indicates that the iterations from memoryless QNWS1 algorithm will converge to the minimizer, and hence the proof is completed. □

**Theorem 3.2** *Let $x_0$ be the starting point for which $f$ satisfies Assumption 3.1. Suppose that $\{x_{k+1}\}$ is generated by (1.2), with the search direction $d_k = -H_k g_k$, under the line search conditions (3.6) and (3.7) for all $k \geq 0$. Then for $H_0 = I$, the memoryless QNWS2 algorithm generates iterations $x_k$ which converge to the minimizer $x^*$.*

*Proof* The proof is generally the same as that of Theorem 3.1, except that now we use updates from (2.32). At $k = 0$, for a nonzero vector $z \in \boldsymbol{R}^n$, we have

$$z^T H_0 z = \|z\|^2 \tag{3.22}$$

as $H_0 = I$. Once again, for any $k \geq 0$, it can be shown that $z^T H_{k+1} z$ is bounded in both cases below.

*Case I*: $H_{k+1} = \frac{s_k^T s_k}{s_k^T y_k} I$, when $\theta_k = \frac{s_k^T s_k}{s_k^T y_k}$.

For any $k \geq 0$, we have

$$z^T H_{k+1} z = \frac{s_k^T s_k}{s_k^T y_k} \|z\|^2 = \frac{\|s_k\|^2}{s_k^T y_k} \|z\|^2 . \tag{3.23}$$

Using (3.4), the latter becomes

$$\frac{1}{M} \|z\|^2 \leq z^T H_{k+1} z \leq \frac{1}{m} \|z\|^2 , \tag{3.24}$$

which is bounded below and above.

*Case II*: $H_{k+1} = I + \frac{s_k^T s_k - s_k^T y_k}{s_k^T s_k y_k^T y_k + (s_k^T y_k)^2} \left( s_k y_k^T + y_k s_k^T \right)$, when $\theta_k = 1$. Now

$$\begin{aligned} z^T H_{k+1} z &= \|z\|^2 + \frac{s_k^T s_k - s_k^T y_k}{s_k^T s_k y_k^T y_k + (s_k^T y_k)^2} \left( z^T s_k y_k^T z + z^T y_k s_k^T z \right) \\ &= \|z\|^2 + \frac{\|s_k\|^2 - s_k^T y_k}{\|s_k\|^2 \left\|y_k\right\|^2 + (s_k^T y_k)^2} \cdot 2 z^T s_k y_k^T z \end{aligned} \tag{3.25}$$

as $s_k^T s_k = \|s_k\|^2$, $y_k^T y_k = \left\|y_k\right\|^2$ and $z^T s_k y_k^T z = z^T y_k s_k^T z$. Since (3.25) is always positive, removing the fraction term from the equation will give a lower bound as follows:

$$z^T H_{k+1} z \geq \|z\|^2 . \tag{3.26}$$

Consider the following for (3.2):

$$\left\|y_k\right\|^2 = \left\|\overline{G}_k s_k\right\|^2 = \left(\overline{G}_k s_k\right)^T \left(\overline{G}_k s_k\right) = s_k^T \overline{G}_k^2 s_k . \tag{3.27}$$

From (3.4), we have

$$m^2 \|s_k\|^2 \leq \left\|y_k\right\|^2 \leq M^2 \|s_k\|^2$$

or

$$m \|s_k\| \leq \left\|y_k\right\| \leq M \|s_k\| . \tag{3.28}$$

By using the Cauchy–Schwarz inequality in (3.25) and since $z^T s_k y_k^T z \leq \|z\|^2 \|s_k\| \left\|y_k\right\|$, using (3.4) together with inequalities from (3.28) gives an upper bound as follows:

$$\begin{aligned} z^T H_{k+1} z &= \|z\|^2 + \frac{\|s_k\|^2 - s_k^T y_k}{\|s_k\|^2 \left\|y_k\right\|^2 + (s_k^T y_k)^2} \cdot 2 z^T s_k y_k^T z \\ &\leq \|z\|^2 + \frac{\|s_k\|^2 - m \|s_k\|^2}{\|s_k\|^2 \left\|y_k\right\|^2 + m^2 \|s_k\|^4} \cdot 2 \|z\|^2 \|s_k\| \left\|y_k\right\| \\ &\leq \|z\|^2 + \frac{(1-m) \|s_k\|^2}{m^2 \|s_k\|^4 + m^2 \|s_k\|^4} \cdot 2M \|z\|^2 \|s_k\|^2 \\ &= \|z\|^2 + \frac{(1-m)}{m^2} \cdot M \|z\|^2 \end{aligned}$$

$$= \left(1 + \frac{(1-m)M}{m^2}\right) \|z\|^2 . \tag{3.29}$$

Combining the results from (3.22), (3.24), (3.26), and (3.29) gives the bounds for all $k \geq 0$ as follows:

$$\gamma_1 \|z\|^2 \leq z^T H_{k+1} z \leq \gamma_2 \|z\|^2 , \tag{3.30}$$

where $\gamma_1 = \min\left\{\frac{1}{M}, 1\right\}$ and $\gamma_2 = \max\left\{\frac{1}{m}, 1 + \frac{(1-m)M}{m^2}, 1\right\}$. The lower and upper bounds from (3.30), together with Lemma 3.2, will imply that $\lim_{k\to\infty} \|g_k\| = 0$. This suggests that the iterations from memoryless QNWS2 algorithm will converge to the minimizer, and hence the proof is completed. □

## 4 Numerical results

This section reports the numerical performance of the proposed methods on a set of 90 test problems given in Table 1 [2, 3], with dimensions ranging from 100 to 10000. The table below illustrates the problems selected from the respective articles.

The algorithms proposed are compared against some existing conjugate gradient methods, namely Polak–Riberie [17] and Dai–Yuan [8]. For simplicity, we shall denote the algorithms under consideration as QNWS1, QNWS2, CGPR, and CGDY, respectively. These standard CG methods would be good direct competitors to the proposed methods as they require comparable computational cost, i.e., require $O(n)$ storage units at each iteration. Note that all methods incorporate the same line search strategy, Algorithm 2.1. All test problems are tested with their standard starting points, with the following parameters:

- line search parameter, $c = 0.3$ (refer to (1.3));
- $\tau = \frac{1}{2}$ is used in Algorithm 2.1;
- lower bound for the step length $\alpha$ is $10^{-3}$;
- termination criterion: $\|g_k\| \leq \varepsilon$, where $\varepsilon = 10^{-4}$;

A run is declared as a failure when the number of iterations reaches 1000 without satisfying the termination criterion. A total of 1440 runs are conducted. The performances of these methods in terms of iteration counts and function evaluations are assessed using the cumulative distribution functions introduced by Dolan and Moré [10], known as performance profiles. The source code is written in MATLAB R2017a and executed on a laptop with an Intel Core i5 2.71 GHz CPU processor and 8 GB RAM memory. Figures 1 and 2 display algorithm performances in terms of number of iterations and function evaluations, respectively, while Table 2 shows the total number function evaluations per iteration for the tested methods.

According to Figs. 1 and 2, QNWS1 and QNWS2 clearly perform better in terms of the number of iterations and function evaluations. For the iteration count, QNWS1 and QNWS2 scored the best at more than 70% and 50% problems, respectively. Similarly, both memoryless methods showed fewer function evaluations, with QNWS1 scoring the best in over 80% instances. In general, the proposed methods often require fewer iterations and function evaluations, and therefore less CPU time. Regarding efficiency, Table 2 suggests that QNWS1 is significantly better among the tested methods as it requires less than 3 function calls per iteration. From the numerical results shown above, we observe that QNWS1 shows promising potential, often requiring fewer iterations and function evaluations to reach the optimal solution.

**Table 1** Test problems

| Source: Andrei [2] | | Source: Andrei [3] |
|---|---|---|
| Generalized Rosenbrock | Generalized White & Holst | TR-SUMM of Quadratic |
| Extended White & Holst | Extended Tridiagonal 2 | Tridiagonal Double Borded Arrow Up |
| Extended Beale | Perturbed Quadratic Diagonal | Diagonal Double Borded Arrow Up |
| Extended Penalty | Generalized Tridiagonal 1 | Almost Perturbed Quartic |
| Perturbed Quadratic | Extended Tridiagonal 1 | DENSCHNA |
| Raydan 1 | Generalized Tridiagonal 2 | DENSCHNC |
| Raydan 2 | Extended Quadratic Penalty QP1 | GENROSNB |
| Diagonal 1 | Extended Quadratic Penalty QP2 | Tridiagonal TS1 |
| Diagonal 2 | Extended Quadratic Exponential EP1 | Tridiagonal TS2 |
| Diagonal 3 | Partial Perturbed Quadratic | Tridiagonal TS3 |
| Hager | Broyden Tridiagonal | QP3 |
| Extended TET | Almost Perturbed Quadratic | EG1 |
| Diagonal 4 | Perturbed Tridiagonal Quadratic | PROsin |
| Diagonal 5 | Staircase 1 | PROD1 |
| Extended Himmelblau | Staircase 2 | PRODcos |
| Generalized PSC1 | LIARWHD | PROD2 |
| Extended PSC1 | POWER | DIAG-AUP1 |
| Extended Powell | ENGVAL1 | |
| Full Hessian FH1 | EDENSCH | |
| Full Hessian FH2 | INDEF | |
| Extended BD1 | CUBE | |
| Extended Maratos | NONSCOMP | |
| Extended Cliff | VARDIM | |
| Extended Wood | QUARTC | |
| Extended Hiebert | Diagonal 6 | |
| Quadratic QF1 | DIXON3DQ | |
| Quadratic QF2 | COSINE | |
| FLETCBV3 | SINE | |
| FLETCHER | BIGGSB1 | |
| BDQRTIC | Generalized Quartic | |
| TRIDIA | Diagonal 7 | |
| ARGLINB | Diagonal 8 | |
| ARWHEAD | Full Hessian FH3 | |
| NONDIA | Diagonal 9 | |
| NONDQUAR | HIMMELBG | |
| DQDRTIC | HIMMELH | |
| EG2 | | |

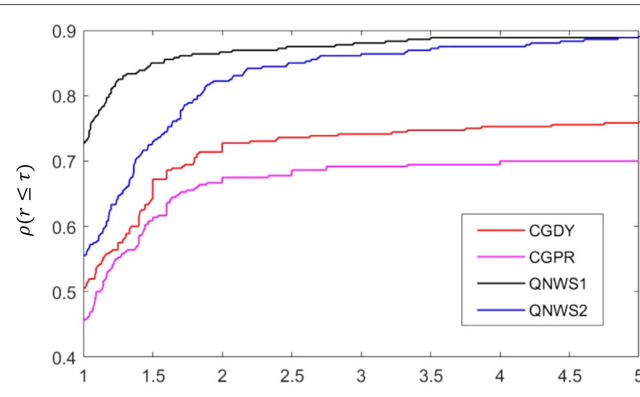


**Figure 1** Performance of the methods in terms of the number of iterations

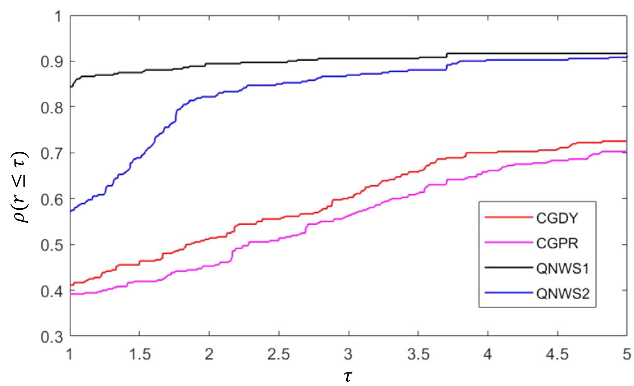


**Figure 2** Performance of the methods in terms of the number of function evaluations

**Table 2** Total number of function evaluations per iteration of the methods

| Methods | QNWS1 | QNWS2 | CGDY | CGPR |
|---|---|---|---|---|
| Average number of function evaluations per iteration | 2.79 | 3.49 | 5.67 | 5.74 |

## 5 Conclusion

This article introduced two variants of quasi-Newton methods that are based on the least-change updating strategy with weak secant equations. Initially, a pair of quasi-Newton methods using the full matrix were derived; one with the weak secant equation suggested by Dennis and Wolkowicz [9] and the other with a nonstandard weak secant Eq. (1.6). Our interest was to derive updating formulas which had smaller computational cost and required less storage, hence both updating formulas were modified by incorporating the concept of a memoryless updating scheme. Numerical results obtained clearly proved that these memoryless versions are encouraging compared to some existing methods. Overall, the proposed memoryless-type methods often require fewer iterations and function evaluations to reach the solutions. Nevertheless, there is still some room for a possible development to further enhance their performances, for instance, one could take a different line search strategy and $\theta$ to fulfil (2.14) and (2.26).

Finally, for a future research, we recommend exploring the application of the proposed memoryless quasi-Newton methods in solving constrained and derivative-free (see, e.g., [20]) optimization problems. Given the efficiency of memoryless quasi-Newton methods in handling large-scale unconstrained optimization, extending these techniques to constrained scenarios (see [1, 13, 14]) may provide valuable insights and potential performance improvements. Additionally, extending our memoryless quasi-Newton methods under other frameworks (see, e.g., [7, 18, 19]) could broaden their applicability in various fields such as engineering design, machine learning, and operations research.

**Author contributions**
K. H. Lim wrote the main text, performed the numerical experiments and prepared the figures. W. J. Leong prepared the theoretical analysis and proofread the convergence results.

**Funding**
The authors would like to thank Ministry of Higher Education, Malaysia, for the financial support given under Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UPM/02/14). The first author would like to express his sincere gratitude to Universiti Putra Malaysia (UPM) for the financial support it has given during his postgraduate study.

**Data Availability**
No datasets were generated or analysed during the current study.

## Declarations

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia. [2]Institute for Mathematical Research, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia.

Received: 29 August 2024 Accepted: 27 November 2024 Published online: 05 December 2024

**References**
1. Abubakar, A.B., Kumam, P., Mohammad, H., Ibrahim, A.H., Seangwattana, T., Hassan, B.A.: A hybrid BFGS-like method for monotone operator equations with applications. J. Comput. Appl. Math. **446**, 115857 (2024)
2. Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim. **10**, 147–161 (2008)
3. Andrei, N.: A collection of 75 unconstrained optimization test functions. Technical Report No. 6/2018, Research Institute for Informatics, Bucharest, Romania (2018)
4. Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. **16**(1), 1–3 (1966)
5. Byrd, R.H., Nocedal, J.: A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM J. Numer. Anal. **26**(3), 727–739 (1989)
6. Cauchy, A.: Méthode générale pour la résolution des systèmes d'équations simultanées. C. R. Hebd. Séances Acad. Sci. **25**, 536–538 (1847)
7. Cheng, Y., Gao, J.: An efficient augmented memoryless quasi-Newton method for solving large-scale unconstrained optimization problems. AIMS Math. **9**(9), 25232–25252 (2024)
8. Dai, Y.H., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**(1), 177–182 (1999)
9. Dennis, J.E., Wolkowicz, H.: Sizing and least change secant methods. SIAM J. Numer. Anal. **30**(5), 1291–1314 (1993)
10. Dolan, E.D., More, J.J.: Benchmarking optimization software with performance profile. Math. Program. **91**(2), 201–213 (2002)
11. Greenstadt, J.: Variation on variable-metric methods. Math. Comput. **24**(109), 1–22 (1970)
12. Luenberger, D.G., Ye, Y.: Linear and Nonlinear Programming, 3rd edn. Springer, New York (2008)
13. Nakayama, S., Narushima, Y., Nishio, H., Yabe, H.: An active-set memoryless quasi-Newton method based on a spectral-scaling Broyden family for bound constrained optimization. Results Control Optim. **3**, 100012 (2021)
14. Nakayama, S., Narushima, Y., Yabe, H.: Memoryless quasi-Newton methods based on spectral-scaling Broyden family for unconstrained optimization. J. Ind. Manag. Optim. **15**(4), 1773–1793 (2019)
15. Nazareth, J.L.: If quasi-Newton then why not quasi-Cauchy? SIAG/OPT Views News **6**, 11–14 (1995)
16. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
17. Polak, E., Ribière, G.: Note sur la convergence de méthodes de directions conjugées. Rev. Franc. Inf. Rech. Opér. **3**, 35–43 (1969)
18. Sim, H.S., Chen, C.Y., Leong, W.J., Li, J.: Nonmonotone spectral gradient method based on memoryless symmetric rank-one update for large-scale unconstrained optimization. J. Ind. Manag. Optim. **18**(6), 3297–3315 (2022)
19. Sim, H.S., Leong, W.J., Chen, C.Y.: Gradient method with multiple damping for large-scale unconstrained optimization. Optim. Lett. **13**, 617–632 (2019)
20. Ullah, N., Sabi'u, J., Shah, A.: A derivative-free scaling memoryless Broyden–Fletcher–Goldfarb–Shanno method for solving a system of monotone nonlinear equations. Numer. Linear Algebra Appl. **28**(5), e2374 (2021)

## Publisher's Note