Comput Optim Appl (2018) 69:225–241
https://doi.org/10.1007/s10589-017-9940-7

On exact linesearch quasi-Newton methods for minimizing a quadratic function

Anders Forsgren[1] · **Tove Odland**[1]

Received: 16 September 2015 / Published online: 31 August 2017

Abstract This paper concerns exact linesearch quasi-Newton methods for minimizing a quadratic function whose Hessian is positive definite. We show that by interpreting the method of conjugate gradients as a particular exact linesearch quasi-Newton method, necessary and sufficient conditions can be given for an exact linesearch quasi-Newton method to generate a search direction which is parallel to that of the method of conjugate gradients. We also analyze update matrices and give a complete description of the rank-one update matrices that give search direction parallel to those of the method of conjugate gradients. In particular, we characterize the family of such symmetric rank-one update matrices that preserve positive definiteness of the quasi-Newton matrix. This is in contrast to the classical symmetric-rank-one update where there is no freedom in choosing the matrix, and positive definiteness cannot be preserved. The analysis is extended to search directions that are parallel to those of the preconditioned method of conjugate gradients in a straightforward manner.

Keywords Method of conjugate gradients · Quasi-Newton method · Unconstrained quadratic program · Exact linesearch method

Research partially supported by the Swedish Research Council (VR).

✉ Anders Forsgren
andersf@kth.se

Tove Odland
odland@kth.se

1 Optimization and Systems Theory, Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

1 Introduction

In this paper we study the behavior of quasi-Newton methods (QN) on an unconstrained quadratic problem of the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} x^T H x + c^T x, \tag{QP}$$

where $H = H^T \succ 0$. Solving (QP) is equivalent to solving a symmetric system of linear equations $Hx + c = 0$. In particular, our concern is to give conditions under which a quasi-Newton method utilizing exact linesearch generates search directions that are parallel to those of the method of conjugate gradients (CG). As exact linesearch is considered, parallel search directions imply identical iterates. At iteration k, the x-iterate and the gradient $Hx + c$ are denoted by x_k and g_k respectively. In a quasi-Newton method, the search direction p_k is computed from $B_k p_k = -g_k$, where B_k is nonsingular.

We give necessary and sufficient conditions on a QN-method for this equivalence with CG on (QP). This is not the first time necessary and sufficient conditions are given. In [12, Theorem 2.2], a necessary and sufficient condition is given, which is based on projections from iterations 0, 1, …, $k-1$, allowing also the preconditioned setting to be considered. In contrast, we interpret the method of conjugate gradients as a particular quasi-Newton method and base the necessary and sufficient conditions on this observation. The result we give is thus directly based on the projection given by the method of conjugate gradients, i.e., based on quantities from iteration $k-1$ and k involving one projection only.

If considering *update matrices* U_k defined by $U_k = B_k - B_{k-1}$, it is well-known that, on (QP), QN using exact linesearch and an update scheme in the one-parameter Broyden family generates identical iterates to those generated by CG, see, e.g., [3, 11, 14]. The unique rank-1 update matrix in the Broyden family is usually referred to as the SR1 update matrix, and it is determined entirely by the so-called secant condition. As a result of our equivalence result, we show that the symmetric rank-1 update matrices that give parallel search directions to CG are given by the family of update matrices on the form

$$U_k = -\frac{1}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} (\gamma_k g_k - g_{k-1})(\gamma_k g_k - g_{k-1})^T, \tag{1}$$

where γ_k is a free parameter. The free parameter can be seen as a relaxation of the secant condition, as the SR1 update matrix is the only matrix in our parameterized rank-1 family which satisfies this condition. We show how to choose the parameter so that positive definiteness of the quasi-Newton matrix is preserved.

To simplify the exposition, we discuss equivalence to CG, which corresponds to the initial Hessian approximation being the identity matrix in the quasi-Newton method in our analysis. We then give the corresponding results in the preconditioned setting, which corresponds to an arbitrary positive definite and symmetric initial Hessian approximation. For the rank-1 case, the family of symmetric update matrices take the form (1) also in the preconditioned setting.

In Sect. 2, we make a brief introduction to CG and QN. In Sect. 3, we present our results which include necessary and sufficient conditions on QN such that CG and QN generate parallel search directions. These results are specialized to update matrices in Sect. 4. In particular, in Sect. 4.1, we give the results on symmetric rank-1 update matrices. Section 5 contains a discussion on how the results would apply if the inverse of the Hessian was updated instead of the Hessian itself. In Sect. 6, the corresponding results in the preconditioned setting are stated. Finally, in Sect. 7 we make some concluding remarks.

2 Background

For solving (QP), we consider linesearch methods on the following form. At iteration k, a search direction p_k is computed. The x-iterate and the gradient are updated as

$$x_{k+1} = x_k + \theta_k p_k, \quad g_{k+1} = g_k + \theta_k H p_k, \text{ for } \theta_k = -\frac{g_k^T p_k}{p_k^T H p_k}.$$

The choice of steplength θ_k corresponds to exact linesearch, i.e., given a search direction p_k the steplength gives the exact minimizer along p_k. This is a natural choice for (QP), as it can be done explicitly. For a given initial point x_0, the iteration process is terminated at an iteration r if $g_r = 0$, in which case x_r is given as the optimal solution to (QP) or equivalently as the unique solution to $Hx + c = 0$. The method is summarized in Algorithm 1.

Algorithm 1 An exact linesearch method for solving $Hx + c = 0$.

$k \leftarrow 0$; $\quad x_k \leftarrow$ initial point; $\quad g_k \leftarrow Hx_k + c$;
while $\|g_k\| \neq 0$ **do**
  $p_k \leftarrow$ search direction;
  $\theta_k \leftarrow -\dfrac{g_k^T p_k}{p_k^T H p_k}$;
  $x_{k+1} \leftarrow x_k + \theta_k p_k$; $\quad g_{k+1} \leftarrow g_k + \theta_k H p_k$;
  $k \leftarrow k + 1$;
end while

The particular linesearch method is defined by the way the search direction p_k is obtained in each iteration k. Our model method is the method of conjugate gradients, CG, by Hestenes and Stiefel [10]. There are different varieties of CG, which are equivalent on (QP). The variety we describe is referred to as the Fletcher-Reeves method of conjugate gradients, as stated in the following definition.

Definition 1 (*The method of conjugate gradients (CG)*) The method of conjugate gradients, CG, is the linesearch method of the form given by Algorithm 1 in which the search direction p_k is given by p_k^{CG}, with

$$p_k^{CG} = \begin{cases} -g_0 & \text{if } k = 0, \\ -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} p_{k-1}^{CG} & \text{if } k \geq 1. \end{cases} \tag{CG}$$

For CG it holds that, for all k, $g_k^T g_i = 0, i = 0, \ldots, k-1$, so the method terminates with $g_r = 0$ for some r, $r \leq n$, and x_r solves (QP). In addition, it holds that $\{p_k^{CG}\}_{k=0}^{r-1}$ are mutually conjugate with respect to H. For an introduction to CG, see, e.g., [2,15,16]. In [5], CG is extended to general unconstrained problems. The reason for CG being our model method is that it requires one matrix-vector product Hp_k per iteration, and it terminates in r iterations, with $r \leq n$.

Next we define what we will refer to as a quasi-Newton method, QN.

Definition 2 (*Quasi-Newton method (QN)*) A quasi-Newton method, QN, is a line-search method of the form given by Algorithm 1 in which the search direction p_k is given by

$$B_k p_k = -g_k, \tag{QN}$$

where the matrix B_k is assumed nonsingular.

Quasi-Newton methods were first suggested by Davidon, see [1], and later modified and formalized by Fletcher and Powell, see [4]. For an introduction to QN-methods, see, e.g., [7, Chapter 4].

Our interest is now to set up conditions on B_k such that p_k and p_k^{CG} are parallel for all k, so that QN also terminates in r iterations. In [6], we derived such conditions based on a sufficient condition to obtain mutually conjugate search directions. Here, we give a direct necessary and sufficient condition based on p_k^{CG} only.

The results of the paper are derived with (CG) as the model method, which corresponds to $B_0 = I$ in (QN) giving $p_0 = p_0^{CG}$. It is also of interest to consider the case when a symmetric positive definite matrix M is given for which a *preconditioned* method of conjugate gradients is defined. This corresponds to $B_0 = M$ in (QN) giving the initial search directions identical. To simplify the exposition, we derive the results for the unpreconditioned case given by (CG) and give the corresponding results for the preconditioned setting in Sect. 6.

3 Necessary and sufficient conditions for QN

In this section we give precise conditions on B_k such that p_k is parallel to p_k^{CG}. The main benefit of the conditions compared to previous work is that our result is based on the single iteration k. The dependence on the previous iterates is contained in the search direction p_{k-1}, and there is no need to check any condition for all the previous iterates.

In the following proposition, we give a necessary and sufficient condition on B_k at a particular iteration k to give a search direction p_k such that $p_k = \delta_k p_k^{CG}$ for a scalar δ_k. We assume that each previous search direction p_i has been parallel to the corresponding search direction of CG, p_i^{CG}, so that QN and CG have generated the same iterate x_k.

Proposition 1 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \le k < r$. Assume that $p_i = \delta_i p_i^{CG}$ with $\delta_i \neq 0$ for $i = 0, \dots, k-1$, where p_i^{CG}, $i = 0, \dots, k-1$, are the search directions of the method of conjugate gradients, as stated in Definition 1. Let A_k be defined as*

$$A_k = I - \frac{1}{g_{k-1}^T p_{k-1}} p_{k-1} g_k^T. \tag{2}$$

Then,

$$A_k^{-1} = I + \frac{1}{g_{k-1}^T p_{k-1}} p_{k-1} g_k^T, \tag{3}$$

and it holds that $A_k p_k^{CG} = -g_k$. In addition, if p_k is given by $B_k p_k = -g_k$ with B_k nonsingular, then, for any nonzero scalar δ_k, it holds that $p_k = \delta_k p_k^{CG}$ if and only if

$$B_k A_k^{-1} g_k = \frac{1}{\delta_k} g_k, \tag{4}$$

or equivalently if and only if

$$B_k = A_k^T W_k A_k, \textit{ with } W_k g_k = \frac{1}{\delta_k} g_k, \textit{for } W_k \textit{ nonsingular.} \tag{5}$$

Finally, it holds that $B_k \succ 0$ if and only if $W_k \succ 0$.

Proof We have

$$p_k^{CG} = -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} p_{k-1}^{CG} = -\left(I + \frac{1}{g_{k-1}^T p_{k-1}^{CG}} p_{k-1}^{CG} g_k^T \right) g_k. \tag{6}$$

Therefore, since $p_{k-1} = \delta_{k-1} p_{k-1}^{CG}$, with $\delta_{k-1} \neq 0$, (6) gives

$$p_k^{CG} = -\left(I + \frac{1}{g_{k-1}^T p_{k-1}} p_{k-1} g_k^T \right) g_k = -A_k^{-1} g_k, \tag{7}$$

with A_k^{-1} given by (3). Since $g_k^T p_{k-1} = 0$, multiplication of A_k of (2) by A_k^{-1} of (3) gives $A_k A_k^{-1} = I$, so that the stated A_k is nonsingular with corresponding inverse A_k^{-1}. Therefore, (7) gives $A_k p_k^{CG} = -g_k$.

Since B_k is assumed nonsingular and $\delta_k \neq 0$, it holds that $p_k = \delta_k p_k^{CG}$ if and only if $B_k(-\delta_k A_k^{-1} g_k) = -g_k$, which is equivalent to (4). Since $g_k^T p_{k-1} = 0$, we obtain $A_k^{-T} g_k = g_k$, so that (4) is equivalent to

$$A_k^{-T} B_k A_k^{-1} g_k = \frac{1}{\delta_k} g_k,$$

which in turn is equivalent to (5). The final result on positive definiteness follows from the nonsingularity of A_k by Sylvester's law of inertia, see, e.g., [8, Theorem 8.1.17]. □

The necessary and sufficient conditions of Proposition 1 give a straightforward way to check if a matrix B_k is such that the corresponding QN-method and CG will generate parallel search directions. The scaling of p_k^{CG} has a special role in our analysis and we relate p_k to p_k^{CG} by a scalar δ_k. The observation that p_k^{CG} may be written as $p_k^{CG} = -A_k^{-1} g_k$ for a nonsingular A_k has been made in [12, Example 2], but the equivalence result of [12] concerns p_k without relating to the scaling of p_k^{CG} explicitly. Therefore, the condition of [12] involves projections on all previous iterations $0, 1, \ldots, k-1$, not one single projection as we obtain. In addition, since there is no relationship to a particular scaling, there is no parameter corresponding to our δ_k. Since such a parameter is vital for deriving later results in our paper, in particular when characterizing symmetric rank-one updates, we cannot apply the equivalence result of [12] directly. A difference in [12] is that they consider matrices N_k that approximate H^{-1} rather than matrices B_k that approximate H. This is not a major difference, we discuss these issues in Sect. 5.

Note that it is not necessary to make $B_k - I$ increase in rank. In particular, $B_k = A_k^T A_k$, corresponding to $W_k = I$ in Proposition 1, is a positive-definite symmetric matrix for which $B_k p_k = -g_k$ gives $p_k = p_k^{CG}$.

We also note that the characterization of B_k does only depend on information from iteration k and $k-1$, since it directly inherits the properties of the method of conjugate gradients. In addition, the characterization of Proposition 1 only depends directly on quantities computed by the quasi-Newton method, the scaling of the method of conjugate gradients is not needed.

4 Results on update matrices

In the previous section we gave results on B_k for a particular iteration k without directly relating to any other B_i, $i \neq k$. It is often the case that B_k is defined in terms of the previous matrix B_{k-1} and an update matrix U_k such that $B_k = B_{k-1} + U_k$, and that conditions are put on U_k. We have in mind a setting where information from the generated gradients is used, so that B_{k-1} may be expressed as $B_{k-1} = I + V_k$, with $\mathcal{R}(V_k) \subseteq span\{g_0, \ldots, g_{k-1}\}$. As we have in mind such a setting where in addition B_k is symmetric, we make the assumption $B_{k-1} g_k = g_k$.

Proposition 1 can then be applied in a straightforward manner to give conditions on U_k such that $p_k = \delta_k p_k^{CG}$. Note that there is a one-to-one correspondence between U_k and B_k given B_{k-1}.

Proposition 2 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \leq k < r$. Assume that $p_i = \delta_i p_i^{CG}$ with $\delta_i \neq 0$ for $i = 0, \ldots, k-1$, where p_i^{CG}, $i = 0, \ldots, k-1$, are the search directions of the method of conjugate gradients,* *as stated in Definition 1. Let B_{k-1} be a nonsingular matrix such that $B_{k-1}p_{k-1} = -g_{k-1}$ and $B_{k-1}g_k = g_k$. Let $U_k = B_k - B_{k-1}$ and assume that B_k and p_k satisfy $B_k p_k = -g_k$, with B_k nonsingular. Then, for any nonzero scalar δ_k, it holds that $p_k = \delta_k p_k^{CG}$ if and only if*

$$U_k\left(g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} p_{k-1}\right) = \left(\frac{1}{\delta_k} - 1\right) g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} g_{k-1}. \tag{8}$$

Proof By assumption, B_k is nonsingular so for $B_k = B_{k-1} + U_k$, Proposition 1 gives $p_k = \delta_k p_k^{CG}$ if and only if

$$\begin{aligned} U_k\left(g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} p_{k-1}\right) &= \frac{1}{\delta_k} g_k - B_{k-1}\left(g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} p_{k-1}\right) \\ &= \frac{1}{\delta_k} g_k - g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} g_{k-1}, \end{aligned}$$

since p_{k-1} is computed from $B_{k-1}p_{k-1} = -g_{k-1}$ and it is assumed that $B_{k-1}g_k = g_k$, so the statement of the proposition follows. □

Note that in the right-hand side of (8) in Proposition 2, the component along g_{k-1} is nonzero and independent of δ_k. The component along g_k is zero for $\delta_k = 1$, i.e., when $p_k = p_k^{CG}$.

4.1 Results on symmetric rank-one update matrices

Next we consider the case when U_k is a symmetric matrix of rank one. It is well known that the secant condition gives a unique update referred to as SR1, see, e.g., [13, Chapter 9]. The secant condition and SR1 will be discussed later in this section. Using Proposition 2 we can give a different result concerning the case when the update matrix U_k is a symmetric matrix of rank one. In particular, we show that the family of rank-1 update matrices can be parameterized by a free parameter and that the matrix is unique for a fixed value of the parameter. This parametrization allows positive definiteness of the quasi-Newton matrix to be preserved.

The situation can be considered in two ways. First, for any given value of the scalar γ_k, except for three distinct values, there is a symmetric rank-1 update matrix U_k of the form

$$U_k = -\frac{1}{(\gamma_k - 1)p_{k-1}^T g_{k-1}}(\gamma_k g_k - g_{k-1})(\gamma_k g_k - g_{k-1})^T, \tag{9}$$

for which $p_k = \delta_k(\gamma_k)p_k^{CG}$, where $\delta_k(\cdot)$ is a real-valued function. Second, if $p_k = \delta_k p_k^{CG}$ is required for any given value of the scalar δ_k, except for three distinct values, and U_k is symmetric and of rank one, U_k must take the form (9), with $\gamma_k = \gamma_k(\delta_k)$, where $\gamma_k(\cdot)$ is the inverse function of $\delta_k(\cdot)$. Consequently, except for three distinct values, there is a one-to-one correspondence between δ_k such that $p_k = \delta_k p_k^{CG}$ and γ_k of the symmetric rank-1 update matrix U_k of (9).

The functions $\delta_k(\cdot)$ and $\gamma_k(\cdot)$ are defined in the following lemma.

Lemma 1 *Consider iteration k of the exact linesearch method of Algorithm 1, where* $1 \le k < r$. *Assume that* $p_i = \delta_i p_i^{CG}$ *with* $\delta_i \neq 0$ *for* $i = 0, \dots, k-1$, *where* p_i^{CG}, $i = 0, \dots, k-1$, *are the search directions of the method of conjugate gradients, as stated in Definition 1. Let* $\hat{\gamma}_k = p_{k-1}^T g_{k-1} / g_k^T g_k$. *For* $\delta_k \neq 0$ *and* $\gamma_k \neq \hat{\gamma}_k$, *let the functions* $\gamma_k(\delta_k)$ *and* $\delta_k(\gamma_k)$ *be defined by*

$$\gamma_k(\delta_k) = -\frac{p_{k-1}^T g_{k-1}}{g_k^T g_k}\left(\frac{1}{\delta_k} - 1\right), \quad \delta_k(\gamma_k) = \frac{1}{1 - \gamma_k \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}}}.$$

Then, the functions $\gamma_k(\cdot)$ *and* $\delta_k(\cdot)$ *are inverses to each other.*

We now characterize the symmetric rank-one update matrices that give search directions which are parallel to those of the method of conjugate gradients. In addition, we give conditions for preserving positive definiteness and a hereditary result.

Proposition 3 *Consider iteration k of the exact linesearch method of Algorithm 1, where* $1 \le k < r$. *Assume that* $p_i = \delta_i p_i^{CG}$ *with* $\delta_i \neq 0$ *for* $i = 0, \dots, k-1$, *where* p_i^{CG}, $i = 0, \dots, k-1$, *are the search directions of the method of conjugate gradients, as stated in Definition 1. Let* B_k *and* p_k *satisfy* $B_k p_k = -g_k$, *and let* B_{k-1} *be a nonsingular matrix such that* $B_{k-1} p_{k-1} = -g_{k-1}$ *and* $B_{k-1} g_k = g_k$. *In addition, let* $\gamma_k(\cdot)$, $\delta_k(\cdot)$ *and* $\hat{\gamma}_k$ *be given by Lemma 1.*

For any scalar γ_k, *except* $\gamma_k = 0$, $\gamma_k = \hat{\gamma}_k$ *and* $\gamma_k = 1$, *let* B_k *be defined by*

$$B_k = B_{k-1} - \frac{1}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} (\gamma_k g_k - g_{k-1})(\gamma_k g_k - g_{k-1})^T. \tag{10}$$

Then, B_k *is nonsingular and* $p_k = \delta_k p_k^{CG}$ *for* $\delta_k = \delta_k(\gamma_k)$.

Conversely, for any scalar δ_k, *except* $\delta_k = 0$, $\delta_k = \delta_k(1)$ *and* $\delta_k = 1$, *assume that* $p_k = \delta_k p_k^{CG}$ *and assume that* $B_k - B_{k-1}$ *is symmetric and of rank one. Then,* B_k *is a nonsingular matrix given by (10) for* $\gamma_k = \gamma_k(\delta_k)$.

If, in addition, $B_{k-1} = B_{k-1}^T \succ 0$, *then* B_k *defined by (10) satisfies* $B_k \succ 0$ *if and only if* $\gamma_k > 1$ *or* $\hat{\gamma}_k < \gamma_k < 0$, *or equivalently if and only if* $\gamma_k = \gamma_k(\delta_k)$ *for* $0 < \delta_k < \delta_k(1)$ *or* $\delta_k > 1$.

Finally, if $B_i p_i = -g_i$, $i = 0, \dots, k$, *with* $B_0 = I$ *and if, for* $i = 1, \dots, k$, B_{i-1} *is updated to* B_i *according to (10) for* γ_i *such that* $\gamma_i \neq 0$, $\gamma_i \neq \hat{\gamma}_i$ *and* $\gamma_i \neq 1$, *then*

$$B_k p_i = \frac{\gamma_{i+1}\theta_i}{\gamma_{i+1} - 1} H p_i, \quad i = 0, \dots, k-1. \tag{11}$$

Proof Let $U_k = B_k - B_{k-1}$. If U_k is symmetric and of rank one, we may write $U_k = \beta_k u_k u_k^T$, where β_k is a scalar and u_k is a vector in $\mathbb{R}^n$, both to be determined. If B_k is nonsingular and $\delta_k \neq 0$, Proposition 2 shows that $p_k = \delta_k p_k^{CG}$ if and only if

$$\beta_k u_k u_k^T A_k^{-1} g_k = \left(\frac{1}{\delta_k} - 1\right) g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} g_{k-1} = -\frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} (\gamma_k g_k - g_{k-1}), \tag{12}$$

with $\gamma_k = \gamma_k(\delta_k)$ given by Lemma 1. Throughout the proof, assume that $\delta_k \notin \{0, \delta_k(1), 1\}$ and $\gamma_k \notin \{0, \hat{\gamma}_k, 1\}$, which is assumed in the statement of the Proposition. Then, Lemma 1 shows that there is a one-to-one correspondence between δ_k and γ_k. Hence, (12) may be considered for either δ_k or γ_k. We choose γ_k for ease of notation.

We first assume that B_k is nonsingular, and verify that this is the case later in the proof. For B_k nonsingular, it follows from (12) that u_k will be equal to the right-hand side vector up to some arbitrary non-zero scaling. Let

$$u_k = \gamma_k g_k - g_{k-1}. \tag{13}$$

The scaling of u_k will be reflected in β_k by insertion into (12) as

$$\beta_k (\gamma_k g_k - g_{k-1})^T \left(g_k + \frac{g_k^T g_k}{p_{k-1}^T g_{k-1}} p_{k-1} \right) = \beta_k (\gamma_k - 1) g_k^T g_k = -\frac{g_k^T g_k}{p_{k-1}^T g_{k-1}},$$

so that

$$\beta_k = -\frac{1}{(\gamma_k - 1) p_{k-1}^T g_{k-1}}. \tag{14}$$

Note that (14) is well defined as $\gamma_k \neq 1$ is assumed. A combination of (12), (13) and (14) gives B_k expressed as in (10).

It remains to show that B_k is nonsingular. It follows from (10) that

$$B_k = B_{k-1} \left(I - \frac{1}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} (\gamma_k B_{k-1}^{-1} g_k - B_{k-1}^{-1} g_{k-1})(\gamma_k g_k - g_{k-1})^T \right),$$

so that

$$\det(B_k) = \det(B_{k-1}) \eta_k, \tag{15}$$

with

$$\begin{aligned} \eta_k &= 1 - \frac{(\gamma_k B_{k-1}^{-1} g_k - B_{k-1}^{-1} g_{k-1})^T (\gamma_k g_k - g_{k-1})}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} \\ &= 1 - \frac{\gamma_k^2 g_k^T g_k - p_{k-1}^T g_{k-1}}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} = \frac{-\gamma_k (\gamma_k g_k^T g_k - p_{k-1}^T g_{k-1})}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} \\ &= \frac{-\gamma_k (\gamma_k - \hat{\gamma}_k)}{(\gamma_k - 1) \hat{\gamma}_k}, \end{aligned} \tag{16}$$

since $B_{k-1} g_k = g_k$, $B_{k-1} p_{k-1} = -g_{k-1}$ and $g_k^T p_{k-1} = 0$, with $\hat{\gamma}_k$ given by Lemma 1. Hence, since B_{k-1} is assumed nonsingular, a combination of (15) and (16) shows that nonsingularity of B_k is equivalent to $\eta_k \neq 0$, i.e., $\gamma_k \neq 0$ and $\gamma_k \neq \hat{\gamma}_k$, which is exactly what is assumed.

To prove the result on positive definiteness, assume that $B_{k-1} = B_{k-1}^T \succ 0$. In this case, since B_{k-1} and B_k differ by a symmetric rank-1 matrix, B_k can have at most one nonpositive eigenvalue, see, e.g., [8, Theorem 8.1.8]. Therefore, (15) shows that positive definiteness of B_k is equivalent to $\eta_k > 0$. Note that $B_{k-1} \succ 0$ implies $p_{k-1}^T g_{k-1} = -p_{k-1}^T B_{k-1}^{-1} p_{k-1} < 0$, which in turn gives $\hat{\gamma}_k < 0$. We may now examine (16) to see what values of γ_k that give $\eta_k > 0$. The numerator of (16) is positive for $\hat{\gamma}_k < \gamma_k < 0$ and negative for $\gamma_k < \hat{\gamma}_k$ and $\gamma_k > 0$. The denominator of (16) is positive for $\gamma_k < 1$ and negative for $\gamma_k > 1$. We conclude that $\eta_k > 0$ if and only if $\hat{\gamma}_k < \gamma_k < 0$ or $\gamma_k > 1$, which by Lemma 1 is equivalent to $0 < \delta_k < \delta_k(1)$ or $\delta_k > 1$.

To prove the final hereditary result, assume that $B_i p_i = -g_i$, $i = 0, \ldots, k$, with $B_0 = I$ and assume that B_{i-1} is updated to B_i according to (10) for γ_i such that $\gamma_i \neq 0$, $\gamma_i \neq \hat{\gamma}_i$ and $\gamma_i \neq 1$. Then, for a given i, $0 < i < k$, k may be replaced by $i+1$ in (10), which gives

$$\begin{aligned} B_{i+1} p_i &= B_i p_i - \frac{1}{(\gamma_{i+1} - 1) p_i^T g_i} (\gamma_{i+1} g_{i+1} - g_i)(\gamma_{i+1} g_{i+1} - g_i)^T p_i \\ &= -g_i + \frac{1}{\gamma_{i+1} - 1} (\gamma_{i+1} g_{i+1} - g_i) = \frac{\gamma_{i+1}}{\gamma_{i+1} - 1} (g_{i+1} - g_i) \\ &= \frac{\gamma_{i+1} \theta_i}{\gamma_{i+1} - 1} H p_i, \end{aligned} \tag{17}$$

where the identities $B_i p_i = -g_i$, $g_{i+1}^T p_i = 0$ and $g_{i+1} - g_i = \theta_i H p_i$ have been used. Finally, (10) gives $B_j p_i = B_{i+1} p_i$ for $j = i+2, \ldots, k$, since $g_j^T p_i = 0$ for $j \geq i+1$. Consequently, $B_k p_i = B_{i+1} p_i$, with $B_{i+1} p_i$ given by (17), proving (11). □

Note that there are two ways in which positive definiteness of a symmetric B_{k-1} may be preserved in a symmetric rank-one update. The first one, $\gamma_k > 1$, or equivalently $0 < \delta_k < \delta_k(1)$, is straightforward, since it corresponds to $U_k \succeq 0$. The second one, $\hat{\gamma}_k < \gamma_k < 0$, or equivalently $\delta_k > 1$, is less straightforward. The corresponding U_k is negative semidefinite, but still the resulting B_k is positive definite.

Proposition 3 gives precise conditions for which rank-one matrices that give a corresponding update matrix that preserves positive definiteness and gives search directions parallel to the method of conjugate gradients. We have the freedom to choose γ_k or δ_k appropriately. This can be compared to SR1, the symmetric rank-one update scheme uniquely defined by the secant condition

$$B_k s_{k-1} = y_k, \text{ for } s_{k-1} = \theta_{k-1} p_{k-1}, \quad y_k = g_k - g_{k-1}. \tag{18}$$

By writing $U_k = B_k - B_{k-1}$, the secant condition gives a requirement on U_k as

$$U_k s_{k-1} = y_k - B_{k-1} s_{k-1}, \tag{19}$$

which for U_k symmetric and of rank one gives the SR1 update matrix U_k^{SR1} on the form

$$U_k^{SR1} = \frac{1}{s_{k-1}^T(y_k - B_k s_{k-1})}(y_k - B_k s_{k-1})(y_k - B_k s_{k-1})^T, \tag{20}$$

see, e.g., [13, Chapter 9]. Since $B_{k-1}p_{k-1} = -g_{k-1}$ holds by the definition of the quasi-Newton method, we may use the definitions of s_{k-1} and y_k of (18) to rewrite U_k^{SR1} of (20) as

$$U_k^{SR1} = \frac{1}{\theta_{k-1}p_{k-1}^T(g_k - (1-\theta_{k-1})g_{k-1}}(g_k - (1-\theta_{k-1})g_{k-1})(g_k - (1-\theta_{k-1})g_{k-1})^T.$$

Since exact linesearch is performed in our case, it holds that $p_{k-1}^T g_k = 0$, so that U_k^{SR1} takes the form

$$\begin{aligned} U_k^{SR1} &= \frac{-1}{\theta_{k-1}(1-\theta_{k-1})p_{k-1}^T g_{k-1}}(g_k - (1-\theta_{k-1})g_{k-1})(g_k - (1-\theta_{k-1})g_{k-1})^T \\ &= \frac{-(1-\theta_{k-1})}{\theta_{k-1}p_{k-1}^T g_{k-1}}\left(\frac{1}{1-\theta_{k-1}}g_k - g_{k-1}\right)\left(\frac{1}{1-\theta_{k-1}}g_k - g_{k-1}\right)^T, \end{aligned} \tag{21}$$

where in the last step, a scaling of the rank-1 vector by a factor $1/(1-\theta_{k-1})$ has been made. A comparison of (10) and (21) shows that the SR1 update is the particular member of the family of symmetric rank-1 updates given by Proposition 3 for which $\gamma_k = 1/(1-\theta_{k-1})$. In particular, for $\theta_{k-1} = 1$, SR1 is not well defined. In addition, as there is no freedom in choosing the rank-one matrix for SR1, there is no way to ensure $B_k \succ 0$ even if $B_{k-1} = B_{k-1}^T \succ 0$. Note that the condition on B_k of (18) giving a condition on U_k of (19) and a unique symmetric rank-1 U_k of (20) is analogous to our condition on B_k of Proposition 1 for a fixed δ_k giving a condition on U_k of Proposition 2 and a unique rank-1 U_k of Proposition 3.

Example 1 illustrates the SR1 update and another rank-1 update of Proposition 3 which preserves positive definiteness. The H and c of the example are parameterized by a positive scalar ϕ. We obtain $\theta_0 = 2/(3\phi)$, so by selecting $\phi = 2/3$, it follows that $\theta_0 = 1$ and the SR1 update becomes undefined. By selecting ϕ slightly smaller than 2/3, for example 0.65, we obtain θ_0 slightly larger than one ($\theta_0 = 40/39$), so that $\gamma_1 = -39$ and the corresponding δ_1 is negative ($\delta_1 = -3/10$). Consequently, B_1^{SR1} is indefinite and the corresponding p_1 is an ascent direction. For comparison, the rank-1 update of Proposition 3 is given for $\delta_1 = 2$, which preserves positive definiteness. As can be seen from (10), the rank-1 update of Proposition 3 is independent of ϕ.

Example 1 For a positive parameter ϕ, consider the example

$$H = \phi\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad c = \phi\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

for which

$$x_1 = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}, \quad x_2 = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}.$$

Then

$$\phi = \frac{2}{3} \quad \Longrightarrow \quad B_1^{SR1} \text{ undefined} \qquad B_1^{\delta_1=2} = \frac{1}{44}\begin{pmatrix} 43 & 5 \\ 5 & 19 \end{pmatrix}$$

$$\phi = \frac{65}{100} \quad \Longrightarrow \quad B_1^{SR1} = \frac{1}{20}\begin{pmatrix} -16 & 42 \\ 42 & -29 \end{pmatrix} \qquad B_1^{\delta_1=2} = \frac{1}{44}\begin{pmatrix} 43 & 5 \\ 5 & 19 \end{pmatrix}$$

Note that the numerical values and the dimension of Example 1 are not important. For a given quadratic problem, there will always exist a particular positive scaling such that the resulting B_1^{SR1} is undefined.

5 On the approximation of the Hessian

The results of the present manuscript have been written based on the search directions of the method of conjugate gradients. The reason for doing so is that it allows a direct treatment of B_k, and there is no need to focus on the update matrix $B_k - B_{k-1}$. This is the choice of the authors, but other choices are of course possible.

The results are stated for a matrix B_k that approximates the Hessian H. We prefer to think of the quasi-Newton method in this way, but there would be little difference if one instead stated the results for a matrix N_k that approximates H^{-1}, which is done for example in [12]. The search direction p_k would then be defined by $p_k = -N_k g_k$ rather than by $B_k p_k = -g_k$ and conditions would be imposed on N_k rather than on B_k. Proposition 1 could be equivalently stated using N_k as the approximation of H^{-1}. Then, the counterparts of (4) and (5) would read

$$N_k g_k = \delta_k A_k^{-1} g_k \text{ and } A_k N_k A_k^T g_k = \delta_k g_k.$$

When considering update matrices, with $V_k = N_k - N_{k-1}$, the counterpart of the update formula (8) of Proposition 2 would read

$$V_k g_k = \delta_k A_k^{-1} g_k - N_{k-1} g_k = (\delta_k - 1) g_k + \delta_k \frac{g_k^T g_k}{g_{k-1}^T p_{k-1}} p_{k-1}. \tag{22}$$

For the rank-one case, the update matrix is unique for a given δ_k, and (22) gives

$$N_k = N_{k-1} - \frac{1}{\gamma_k(\gamma_k g_k^T g_k - p_{k-1}^T g_{k-1})}(\gamma_k g_k + p_{k-1})(\gamma_k g_k + p_{k-1})^T, \tag{23}$$

with $\gamma_k = \gamma_k(\delta_k)$ of Lemma 1. The uniqueness of the update implies that if $N_{k-1} = B_{k-1}^{-1}$, then $N_k = B_k^{-1}$ and (23) follows from (10) by the Sherman–Morrison formula.

As for the rank-one case, in light of the results of Sect. 4.1, one of the referees has pointed out that the parametrization given by δ_k can be replaced by a different parametrization. The secant condition $\theta_{k-1} B_k p_{k-1} = g_k - g_{k-1}$ and its counterpart on the inverse $\theta_{k-1} p_{k-1} = N_k(g_k - g_{k-1})$ may be relaxed by a parameter λ_k so that

$\lambda_k \theta_{k-1} B_k p_{k-1} = g_k - g_{k-1}$ and $\lambda_k \theta_{k-1} p_{k-1} = N_k(g_k - g_{k-1})$ respectively. For the updates, we obtain

$$\lambda_k \theta_{k-1} U_k p_k = g_k - (1 - \lambda_k \theta_{k-1}) g_{k-1} \text{ and} \tag{24a}$$

$$V_k(g_k - g_{k-1}) = -g_k - (1 - \lambda_k \theta_{k-1}) p_{k-1}, \tag{24b}$$

since $B_{k-1} p_{k-1} = -g_{k-1}$, $N_{k-1} g_k = g_k$ and $N_{k-1} g_{k-1} = -p_{k-1}$. If having read the previous sections of this paper, we would see that $p_k = \delta_k p_k^{CG}$, where we can relate γ_k to λ_k by $\gamma_k = 1/(1 - \lambda_k \theta_{k-1})$, by comparing the right-hand side vector of (24a) to the rank-one vector of Proposition 3 or comparing the right-hand side vector of (24b) to the rank-one vector of (23). The corresponding relationship to δ_k is given by Lemma 1. An alternative to reading the previous sections of this paper, however, would be to say that $\lambda_k = 1$ corresponds to the SR1 update, and $\lambda_k = 0$ corresponds to the conjugate projection update [3, Eq. (4.1.10)]. They are considered in the update of the inverse and are both known to give p_k parallel to p_k^{CG} if $N_0 = I$. By replacing γ_k by $1/(1 - \lambda_k \theta_{k-1})$, one could show that a rank-1 matrix of the form (9) would give p_k parallel to p_k^{CG} using induction similar to what is done in [3, Theorem 3.4.1] and give conditions on preserving positive definiteness on λ_k. This would, however, not show that there is no other family of rank-1 updates giving p_k parallel to p_k^{CG}. We prefer to give a direct proof based on our result of Proposition 1, as we from there get both necessary and sufficient conditions.

6 Preconditioning

Our results have been derived in the setting of CG, which corresponds to $B_0 = I$ in QN giving the initial search directions identical. In this section, we give the analogous results in a preconditioned setting. In the preconditioned method of conjugate gradients, there is a positive definite symmetric matrix M, providing an estimate of H. For the quasi-Newton method, this will correspond to $B_0 = M$ giving the initial search directions identical.

The preconditioned method of conjugate gradient takes the following form. If the Cholesky factor of M is denoted by L, so that $M = LL^T$, then the method of conjugate gradients is applied to

$$L^{-1} H L^{-T} \widehat{x} + L^{-1} c = 0, \tag{25}$$

for $\widehat{x} = L^T x$, see, e.g., [15, Chapter 9.2]. Letting "hat" be associated with quantities of (25), we obtain $\widehat{p} = L^T p$ and $\widehat{g} = L^{-1} g$. Since $\widehat{p}$ is associated with a "usual" unpreconditioned system, we write $\widehat{p}^{CG}$, and since p is associated with a preconditioned system, we write p^{PCG}, so that $\widehat{p}^{CG} = L^T p^{PCG}$. It is straightforward to use these relations to derive the result analogous to those given in the the previous sections also for the preconditioned system.

Definition 3 (*The preconditioned method of conjugate gradients (PCG)*) For a positive definite symmetric $n \times n$ matrix M, the preconditioned method of conjugate gradients, PCG, is the linesearch method of the form given by Algorithm 1 in which the search direction p_k is given by p_k^{PCG}, with

$$p_k^{PCG} = \begin{cases} -M^{-1}g_0 & \text{if } k = 0, \\ -M^{-1}g_k + \frac{g_k^T M^{-1} g_k}{g_{k-1}^T M^{-1} g_{k-1}} p_{k-1}^{PCG} & \text{if } k \geq 1. \end{cases} \tag{PCG}$$

For PCG it holds that, for all k, $g_k^T M^{-1} g_i = 0$, $i = 0, \ldots, k-1$, so the method terminates with $g_r = 0$ for some r, $r \leq n$, and x_r solves (QP). In addition, it holds that $\{p_k^{PCG}\}_{k=0}^{r-1}$ are mutually conjugate with respect to H.

Proposition 4 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \leq k < r$. Assume that $p_i = \delta_i p_i^{PCG}$ with $\delta_i \neq 0$ for $i = 0, \ldots, k-1$, where p_i^{PCG}, $i = 0, \ldots, k-1$, are the search directions of the preconditioned method of conjugate gradients, as stated in Definition 3. Let A_k be defined as*

$$A_k = I - \frac{1}{g_{k-1}^T p_{k-1}} p_{k-1} g_k^T.$$

Then,

$$A_k^{-1} = I + \frac{1}{g_{k-1}^T p_{k-1}} p_{k-1} g_k^T,$$

and it holds that $M A_k p_k^{PCG} = -g_k$. In addition, if p_k is given by $B_k p_k = -g_k$ with B_k nonsingular, then, for any nonzero scalar δ_k, it holds that $p_k = \delta_k p_k^{PCG}$ if and only if

$$B_k A_k^{-1} M^{-1} g_k = \frac{1}{\delta_k} g_k,$$

or equivalently if and only if

$$B_k = A_k^T W_k A_k, \text{ with } W_k M^{-1} g_k = \frac{1}{\delta_k} g_k, \text{ for } W_k \text{ nonsingular.}$$

Finally, it holds that $B_k \succ 0$ if and only if $W_k \succ 0$.

In particular, $B_k = A_k^T M A_k$, corresponding to $W_k = M$ in Proposition 4, is a positive-definite symmetric matrix for which $B_k p_k = -g_k$ gives $p_k = p_k^{PCG}$.

Proposition 5 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \leq k < r$. Assume that $p_i = \delta_i p_i^{PCG}$ with $\delta_i \neq 0$ for $i = 0, \ldots, k-1$, where p_i^{PCG}, $i = 0, \ldots, k-1$, are the search directions of the preconditioned method of conjugate gradients using a positive definite symmetric preconditioning matrix M, as stated in Definition 3. Let B_{k-1} be a nonsingular matrix such that $B_{k-1} p_{k-1} = -g_{k-1}$ and $B_{k-1} M^{-1} g_k = g_k$. Let $U_k = B_k - B_{k-1}$ and assume that B_k and p_k satisfy $B_k p_k = -g_k$, with B_k nonsingular. Then, for any nonzero scalar δ_k, it holds that $p_k = \delta_k p_k^{PCG}$ if and only if*

$$U_k \left(M^{-1} g_k + \frac{g_k^T M^{-1} g_k}{p_{k-1}^T g_{k-1}} p_{k-1} \right) = \left(\frac{1}{\delta_k} - 1 \right) g_k + \frac{g_k^T M^{-1} g_k}{p_{k-1}^T g_{k-1}} g_{k-1}.$$

Lemma 2 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \le k < r$. Assume that $p_i = \delta_i p_i^{PCG}$ with $\delta_i \neq 0$ for $i = 0, \ldots, k-1$, where p_i^{PCG}, $i = 0, \ldots, k-1$, are the search directions of the preconditioned method of conjugate gradients using a positive definite symmetric preconditioning matrix M, as stated in Definition 3. Let $\hat{\gamma}_k = p_{k-1}^T g_{k-1} / g_k^T M^{-1} g_k$. For $\delta_k \neq 0$ and $\gamma_k \neq \hat{\gamma}_k$, let the functions $\gamma_k(\delta_k)$ and $\delta_k(\gamma_k)$ be defined by*

$$\gamma_k(\delta_k) = -\frac{p_{k-1}^T g_{k-1}}{g_k^T M^{-1} g_k}\left(\frac{1}{\delta_k} - 1\right), \quad \delta_k(\gamma_k) = \frac{1}{1 - \gamma_k \frac{g_k^T M^{-1} g_k}{p_{k-1}^T g_{k-1}}}.$$

Then, the functions $\gamma_k(\cdot)$ and $\delta_k(\cdot)$ are inverses to each other.

Proposition 6 *Consider iteration k of the exact linesearch method of Algorithm 1, where $1 \le k < r$. Assume that $p_i = \delta_i p_i^{PCG}$ with $\delta_i \neq 0$ for $i = 0, \ldots, k-1$, where p_i^{PCG}, $i = 0, \ldots, k-1$, are the search directions of the preconditioned method of conjugate gradients using a positive definite symmetric preconditioning matrix M, as stated in Definition 3. Let B_k and p_k satisfy $B_k p_k = -g_k$, and let B_{k-1} be a nonsingular matrix such that $B_{k-1} p_{k-1} = -g_{k-1}$ and $B_{k-1} M^{-1} g_k = g_k$. In addition, let $\gamma_k(\cdot)$, $\delta_k(\cdot)$ and $\hat{\gamma}_k$ be given by Lemma 2.*

For any scalar γ_k, except $\gamma_k = 0$, $\gamma_k = \hat{\gamma}_k$ and $\gamma_k = 1$, let B_k be defined by

$$B_k = B_{k-1} - \frac{1}{(\gamma_k - 1) p_{k-1}^T g_{k-1}} (\gamma_k g_k - g_{k-1})(\gamma_k g_k - g_{k-1})^T. \tag{26}$$

Then, B_k is nonsingular and $p_k = \delta_k p_k^{PCG}$ for $\delta_k = \delta_k(\gamma_k)$.

Conversely, for any scalar δ_k, except $\delta_k = 0$, $\delta_k = \delta_k(1)$ and $\delta_k = 1$, assume that $p_k = \delta_k p_k^{PCG}$ and assume that $B_k - B_{k-1}$ is symmetric and of rank one. Then, B_k is a nonsingular matrix given by (26) for $\gamma_k = \gamma_k(\delta_k)$.

If, in addition, $B_{k-1} = B_{k-1}^T \succ 0$, then B_k defined by (26) satisfies $B_k \succ 0$ if and only if $\gamma_k > 1$ or $\hat{\gamma}_k < \gamma_k < 0$, or equivalently if and only if $\gamma_k = \gamma_k(\delta_k)$ for $0 < \delta_k < \delta_k(1)$ or $\delta_k > 1$.

Finally, if $B_i p_i = -g_i$, $i = 0, \ldots, k$, with $B_0 = M$ and if, for $i = 1, \ldots, k$, B_{i-1} is updated to B_i according to (26) for γ_i such that $\gamma_i \neq 0$, $\gamma_i \neq \hat{\gamma}_i$ and $\gamma_i \neq 1$, then

$$B_k p_i = \frac{\gamma_{i+1}\theta_i}{\gamma_{i+1} - 1} H p_i, \quad i = 0, \ldots, k-1.$$

7 Conclusion

In this paper we have derived necessary and sufficient conditions on the matrix B_k in a QN-method such that p_k, obtained by solving $B_k p_k = -g_k$, satisfies $p_k = \delta_k p_k^{PCG}$ for some $\delta_k \neq 0$, where p_k^{PCG} is the search direction of the preconditioned method of conjugate gradients. These conditions are stated in Proposition 4. The results have been derived for the case of CG and then extended to PCG for a symmetric positive definite preconditioning matrix M.

Further, we have characterized the symmetric rank-one update matrices for QN that give parallel search directions to those of PCG. In Proposition 6, we show that the rank-one matrix must be a linear combination of g_k and g_{k-1}, and also that almost any linear combination will do. In addition, we characterize the family of symmetric rank-one updates that preserve symmetry and positive definiteness of B_{k-1}.

Our focus is on the mathematical properties of PCG and QN in exact arithmetic. We want to stress that considering the numerical properties in finite precision is of utmost importance, but such an analysis is beyond the scope of this paper. See, e.g., [9] for an illustration of a case where PCG and QN generate identical iterates in exact arithmetic but the difference between numerically computed iterates for the two methods is large.

The results of the paper are meant to be useful as such, for understanding the behavior of exact linesearch quasi-Newton methods for minimizing a quadratic function. In addition, we hope that they can lead to further research on methods for unconstrained minimization. In particular, understanding the behavior of quasi-Newton methods on near-quadratic functions would be a subject of future research.

Acknowledgements We thank the editor and the anonymous referees for their constructive comments which significantly improved the presentation. Funding was provided by Vetenskapsrådet (Grant No. 621-2014-4772).

References

1. Davidon, W.C.: Variable metric method for minimization. SIAM J. Optim. **1**(1), 1–17 (1991)
2. Demmel, J.W.: Applied Numerical Linear Algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1997)
3. Fletcher, R.: Practical Methods of Optimization, 2nd edn. A Wiley-Interscience Publication. Wiley, Chichester (1987)
4. Fletcher, R., Powell, M.J.D.: A rapidly convergent descent method for minimization. Comput. J. **6**, 163–168 (1963/1964)
5. Fletcher, R., Reeves, C.M.: Function minimization by conjugate gradients. Comput. J. **7**, 149–154 (1964)
6. Forsgren, A., Odland, T.: On the connection between the conjugate gradient method and quasi-Newton methods on quadratic problems. Comput. Optim. Appl. **60**(2), 377–392 (2015)
7. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London (1981)
8. Golub, G.H., Van Loan, C.F.: Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences, 3rd edn., pp. xxx+698. Johns Hopkins University Press, Baltimore, MD (1996)
9. Hager, W.W., Zhang, H.: The limited memory conjugate gradient method. SIAM J. Optim. **23**(4), 2150–2168 (2013)
10. Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Stand. **49**(409–436), 1952 (1953)
11. Huang, H.Y.: Unified approach to quadratically convergent algorithms for function minimization. J. Optim. Theory Appl. **5**, 405–423 (1970)
12. Kolda, T.G., O'Leary, D.P., Nazareth, L.: BFGS with update skipping and varying memory. SIAM J. Optim. **8**(4), 1060–1083 (1998). electronic
13. Luenberger, D.G.: Linear and Nonlinear Programming, 2nd edn. Addison-Wesley Pub Co, Boston (1984)

14. Nazareth, L.: A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms. SIAM J. Numer. Anal. **16**(5), 794–800 (1979)
15. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia (2003)
16. Shewchuk, J.R.: An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie-Mellon University, Pittsburgh, PA, USA (1994)