

# A Curvilinear Method for Large Scale Optimization Problems

M.S. Apostolopoulou\* , D.G. Sotiropoulos and C.A. Botsaris

*Abstract*— We present a new matrix-free method for the computation of the negative curvature direction in large scale unconstrained problems. We describe a curvilinear method which uses a combination of a quasi-Newton direction and a negative curvature direction. We propose an algorithm for the computation of the search directions which uses information of two specific L-BFGS matrices in such a way that avoids both the calculation and the storage of the approximate Hessian. Explicit forms for the eigenpair that corresponds to the most negative eigenvalue of the approximate Hessian are also presented. Numerical results show that the proposed approach is promising.

*Keywords*— large scale unconstrained optimization, curvilinear search, negative curvature direction, eigenvalues, eigenvectors, power inverse method, quasi-Newton method

## I. INTRODUCTION

We consider the following large scale *unconstrained optimization* problem

$$\min\{f(x) \mid x \in \mathbb{R}^n\}, \quad (1)$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is twice continuously differentiable, while the dimension  $n$  is more than one thousand. There exist two principal classes for solving (1), namely, line search and trust region methods [16]. The method described in this work belongs to the class of linesearch procedures; in particular, we consider an iterative scheme of the form

$$x_{k+1} = \begin{cases} x_k + \alpha_k p_k, & \text{when } \nabla^2 f(x) \geq 0; \\ x_k + \alpha_k^2 p_k + \alpha_k d_k, & \text{otherwise,} \end{cases} \quad (2)$$

where  $x_k$  is the current estimate of the minimum,  $p_k$  and  $d_k$  are the pair of directions and  $\alpha_k$  is a step size chosen so that at least  $f(x_{k+1}) < f(x_k)$ . The direction  $p$  is a Newton-type direction, while the direction  $d$  is called a direction of *negative curvature* and is related to the eigenvector that corresponds to the most negative eigenvalue of the Hessian.

One of the first proposals for the use of the negative curvature direction is that of Fiacco and McCormick [8]. In 1977, McCormick [12] showed how a modification of the Armijio's rule [1] could be used to include cases where second derivative information is used, in particular when the Hessian matrix is not positive semi-definite. Moré and Sorensen [13] and Goldfarb [9] have proposed a similar approach along a curve of the form

$$C = \{x(\alpha) : x(\alpha) = x + \phi_1(\alpha)p + \phi_2(\alpha)d, \alpha \geq 0\} \quad (3)$$

\* Corresponding author. E-mail: msa@math.upatras.gr

M.S. Apostolopoulou is with the Department of Mathematics, University of Patras, GR-265 04 Patras, Greece.

D.G. Sotiropoulos is with Ionian University, Department of Informatics, GR-491 00 Corfu, Greece.

C.A. Botsaris is with University of Central Greece, Department of Regional Economic Development, GR-321 00 Levadia, Greece.

with  $\phi_1(0) = \phi_2(0) = 0$ . In [13], the functions  $\phi_1$  and  $\phi_2$  are  $\phi_1(\alpha) = \alpha^2$  and  $\phi_2(\alpha) = \alpha$ , and therefore the curvilinear path (3) leads to iterations of the form

$$x_{k+1} = x_k + \alpha_k^2 p_k + \alpha_k d_k, \quad \alpha \geq 0.$$

However, in [9] the functions  $\phi_1$  and  $\phi_2$  are  $\phi_1(\alpha) = \alpha$  and  $\phi_2(\alpha) = \alpha^2$ , thus the curvilinear path is of the form

$$x_{k+1} = x_k + \alpha_k p_k + \alpha_k^2 d_k, \quad \alpha \geq 0.$$

The main disadvantage of these methods is the calculation and the storage of the Hessian matrix which requires an additional  $n(n+1)/2$  locations of core. Furthermore, both the eigenvalues of the Hessian and the eigenvector corresponding to the most negative eigenvalue have to be calculated, for determining the negative curvature direction. This computation requires the factorization and the storage of a matrix. Therefore, the large dimensions make the computation of both the Hessian matrix and the negative curvature direction very costly.

The aim of this work is to avoid both the computation and the storage of the Hessian matrix and to produce a direction of negative curvature with an inexpensive way, that can be applied in very large problems. To this end, we utilize a specific limited memory BFGS method for the computation of an approximate Hessian matrix. The BFGS update is applied to an initial diagonal matrix  $B^{(0)}$  using information from the  $m = 2$  previous iterations. Each step of this method updates  $B_k$  by means of the BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}. \quad (4)$$

using the vector pair  $\{s_k, y_k\}$ , where  $s_k = x_{k+1} - x_k$  and  $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ . The inverse of  $B_{k+1}$  [15] denoted by  $H_{k+1}$ , is of the form

$$H_{k+1} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left[ 1 + \frac{y_k^T H_k y_k}{s_k^T y_k} \right] \frac{s_k s_k^T}{s_k^T y_k}. \quad (5)$$

The resulting approximation, give us the ability to determine analytically the eigenvalues of  $B$  and the inverse of  $B - \lambda I$ ,  $\lambda \in \mathbb{R}$ , via simple formulas. Therefore, the eigenvector that corresponds to the most negative eigenvalue of  $B$  is computed by applying a single step of the inverse power method [2], [11], [19]. Thus, the computation of the search directions is obtained by performing a sequence of inner products and vector summations.

*Notation.* Throughout the paper  $\|\cdot\|$  denotes the Euclidean norm. The gradient of  $f$  is denoted by  $g$  while a matrix  $A$  is indicated that is positive semi-definite by  $A \geq 0$ . Moreover, the L-BFGS matrices are denoted by  $B^{(1)}$  and  $H^{(1)}$ , for  $m = 1$ , and by  $B^{(2)}$  and  $H^{(2)}$ , for  $m = 2$ .

## II. PRELIMINARIES

For the remainder of this work we will assume that the following regularity properties hold for problem (1) and the initial point  $x_0 \in \mathbb{R}^n$ .

*Assumption 1:* The level set  $\Omega = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$  is compact.

*Assumption 2:* The objective function  $f$  has second order derivatives that are Lipschitz continuous on an open set that contains  $\Omega$ .

Under Assumptions 1 and 2 it is possible to construct a local quadratic model for the objective function  $f$  from the corresponding Taylor series expansion at all iterates  $x_k \in \Omega$  as

$$q_k(p) \simeq f_k + g_k^T p + \frac{1}{2} p^T B_k p$$

where  $B_k$  is an positive definite L-BFGS updated approximate Hessian matrix. Hence, the quasi-Newton direction  $p_k$  can be computed by the formula

$$p_k = \begin{cases} -B_k^{-1} g_k, & \text{if } B_k \geq 0; \\ -g_k, & \text{otherwise,} \end{cases}$$

while the negative curvature direction  $d_k$  is related with the unit eigenvector  $u_k$  that corresponds to the most negative eigenvalue of  $B_k$  and is obtained by the form

$$d_k = \begin{cases} 0, & B_{k+1} \geq 0; \\ -\text{sgn}(u_k^T g_k) u_k, & \text{otherwise.} \end{cases}$$

Both directions must be *descent* directions.

*Definition 1:* A pair of directions  $(p, d)$  is called a *descent pair* at a point  $x$  where  $f(x)$  is twice differentiable, if

$$p^T \nabla f(x) < 0, \quad d^T \nabla f(x) \leq 0 \quad \text{and} \quad d^T \nabla^2 f(x) d = 0,$$

when  $\nabla^2 f(x) \geq 0$ , and

$$p^T \nabla f(x) \leq 0, \quad d^T \nabla f(x) \leq 0 \quad \text{and} \quad d^T \nabla^2 f(x) d < 0,$$

otherwise.

From a more practical point of view, in a given iterate  $x_k$ , the pair  $(p_k, d_k)$  is assumed to be a sufficient descent pair of directions if the directions  $\{p_k\}$  and  $\{d_k\}$  are bounded and satisfy the following conditions:

*Condition 1:*  $g_k^T p_k = 0$  implies  $g_k = 0$  and  $p_k = 0$

*Condition 2:*  $g_k^T p_k \rightarrow 0$  implies  $g_k \rightarrow 0$  and  $p_k \rightarrow 0$

*Condition 3:*  $d_k^T B_k d_k \rightarrow 0$  implies  $\min(\lambda_1, 0) \rightarrow 0$  and  $d_k \rightarrow 0$ , where  $\lambda_1$  is the smallest eigenvalue of  $B_k$ .

Conditions 1 and 2 are the standard ones for Newton-type directions. They ensure that the  $p_k \neq 0$  and  $g_k \neq 0$  are not orthogonal and do not become nearly so too rapidly. Condition 3 ensures that  $d_k$  contains information related to the smallest eigenvalue of the approximate Hessian.

## III. UPDATING THE APPROXIMATE HESSIAN

The computation of the matrix  $B \approx \nabla^2 f(x)$  is accomplished using the L-BFGS philosophy with  $m = 2$ . Denoting by  $B_k^{(0)}$  and  $H_k^{(0)}$  the initial matrices, the matrices  $B_{k+1}$  and  $H_{k+1}$  are updated using  $B_k^{(0)} = (1/\theta_k)I$  and

$H_k^{(0)} = \theta_k I$ ,  $\theta_k \in \mathbb{R}$ , respectively. More analytically, for  $k \geq 2$ , the matrix  $B_{k+1}$  in Eq. (4) is updated using the matrix  $B_k$  defined by

$$B_k = \frac{1}{\theta_k} I - \frac{s_{k-1} s_{k-1}^T}{\theta_k s_{k-1}^T s_{k-1}} + \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}}, \quad (6)$$

while its inverse,  $H_{k+1}$  in Eq. (5), is updated by the matrix

$$H_k = \theta_k I - \theta_k \frac{y_{k-1} s_{k-1}^T + s_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}} + \left[ 1 + \theta_k \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}} \right] \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}. \quad (7)$$

Thus, the matrices  $B_k$  and  $H_k$  are computed by applying the 1-BFGS update on  $B_k^{(0)}$  and  $H_k^{(0)}$  (the set of vector pairs contains *one pair*  $\{(s_{k-1}, y_{k-1})\}$ ). Consequently,  $B_{k+1}$  and  $H_{k+1}$  are computed by applying the 2-BFGS update on  $B_k^{(0)}$  and  $H_k^{(0)}$  (the set of vector pairs contains *two pairs*,  $\{(s_{k-1}, y_{k-1}), (s_k, y_k)\}$ ).

Motivated by the work of Barzilai and Borwein [3] and Birgin and Martínez [4], we use the scalar parameter

$$\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \quad (8)$$

for scaling the initial matrix. This parameter is the inverse of the Rayleigh quotient  $s^T G s / s^T s$ , which lies between the largest and the smallest eigenvalues of the average Hessian  $G$ . Rigorous analysis of methods exclusively based on this approximation may be found in [5], [17], [18].

It is known [16] that the determinant and the trace of BFGS matrices are given by the formulas  $\det(B_{k+1}) = \det(B_k) \frac{s_k^T y_k}{s_k^T B_k s_k}$  and  $\text{tr}(B_{k+1}) = \text{tr}(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{s_k^T y_k}$ .

When  $B_{k+1} \equiv B_{k+1}^{(1)}$ , its determinant is expressed as

$$\det(B_{k+1}^{(1)}) = \frac{1}{\theta_{k+1}^n}, \quad (9)$$

while its trace is given by

$$\text{tr}(B_{k+1}^{(1)}) = \frac{a_k + n - 2}{\theta_{k+1}}, \quad (10)$$

where

$$a_k = 1 + \theta_{k+1} \frac{y_k^T y_k}{s_k^T y_k}. \quad (11)$$

Observing the updating scheme, we can see that  $B_{k+1}^{(0)}$  is updated by the addition of two rank-one matrices. The following lemma, due to Wilkinson [20, pp. 94–98], states that the eigenvalues of a matrix which is updated by a rank-one matrix, interlace with the eigenvalues of the original matrix.

*Lemma 1:* If symmetric matrices  $A$  and  $A^*$  differ by a matrix of rank-one, then their eigenvalues  $\lambda$  and  $\lambda^*$  interpolate each other in a weak sense. In particular, if  $A^* = A + \sigma v v^T$ , where  $\sigma$  is scalar,

$\lambda_n \geq \lambda_{n-1} \geq \dots \geq \lambda_1$  and  $\lambda_n^* \geq \lambda_{n-1}^* \geq \dots \geq \lambda_1^*$ , then

- (1) if  $\sigma \geq 0$ ,  $\lambda_n^* \geq \lambda_n \geq \lambda_{n-1}^* \geq \lambda_{n-1} \geq \dots \geq \lambda_1^* \geq \lambda_1$   
(2) if  $\sigma < 0$ ,  $\lambda_n \geq \lambda_n^* \geq \lambda_{n-1} \geq \lambda_{n-1}^* \geq \dots \geq \lambda_1 \geq \lambda_1^*$

We next give the general form of the characteristic polynomial of  $B_{k+1}^{(1)}$ , and show that the extreme eigenvalues of  $B_{k+1}^{(1)}$  are distinct.

*Lemma 2:* Let the symmetric matrix  $B_{k+1}^{(1)} \in \mathbb{R}^{n \times n}$ . Then, the characteristic polynomial of  $B_{k+1}^{(1)}$  has the general form

$$p_1(\lambda) = \left(\lambda - \frac{1}{\theta_{k+1}}\right)^{n-2} \left(\lambda^2 - \frac{a_k}{\theta_{k+1}}\lambda + \frac{1}{\theta_{k+1}^2}\right). \quad (12)$$

Moreover, if  $a_k > 2$ , then  $\lambda_1 < 1/\theta_{k+1} < \lambda_n$ , where  $\lambda_1$  and  $\lambda_n$  are the smallest and largest eigenvalues of  $B_{k+1}^{(1)}$ , respectively.

*Proof:* First we show that  $B_{k+1}^{(1)}$  has at most two distinct eigenvalues. To this end, we consider the matrix  $\bar{B} = (1/\theta_{k+1})I - s_k s_k^T / \theta_{k+1} s_k^T s_k$  with rank  $(n-1)$ . Using Lemma 1, it is easy to see that  $\bar{B}$  besides the zero eigenvalue, has one more eigenvalue equals to  $1/\theta_{k+1}$  of multiplicity  $(n-1)$ . If  $B_{k+1}^{(0)}$  is positive definite, then the addition of the term  $y_k y_k^T / s_k^T y_k$  on  $\bar{B}$  yields (cf. Lemma 1)

$$\lambda_n^B \geq \frac{1}{\theta_{k+1}} \geq \lambda_{n-1}^B \geq \frac{1}{\theta_{k+1}} \geq \dots \geq \lambda_2^B \geq \frac{1}{\theta_{k+1}} \geq \lambda_1^B \geq 0;$$

otherwise,

$$0 \geq \lambda_n^B \geq \frac{1}{\theta_{k+1}} \geq \lambda_{n-1}^B \geq \frac{1}{\theta_{k+1}} \geq \dots \geq \lambda_2^B \geq \frac{1}{\theta_{k+1}} \geq \lambda_1^B.$$

In both cases, it is obvious that  $\lambda_{n-1}^B = \dots = \lambda_2^B = 1/\theta_{k+1}$ , and consequently,

$$\lambda_n^B \geq 1/\theta_{k+1} \geq \lambda_1^B. \quad (13)$$

Relation (13) implies that  $B_{k+1}^{(1)}$  has at most two distinct eigenvalues and one eigenvalue equals to  $1/\theta_{k+1}$  of multiplicity at least  $(n-2)$ . Denoting by  $\lambda_1$  and  $\lambda_2$  the two unknown distinct eigenvalues, we yield that

$$\det(B_{k+1}^{(1)}) = \frac{\lambda_1 \lambda_2}{\theta_{k+1}^{n-2}}, \quad \text{and} \quad \text{tr}(B_{k+1}^{(1)}) = \lambda_1 + \lambda_2 + \frac{n-2}{\theta_{k+1}}.$$

From Eqs. (9) and (10) we obtain  $\lambda_1 \lambda_2 = 1/\theta_{k+1}^2$ , while  $\lambda_1 + \lambda_2 = a_k + 2/\theta_{k+1}$ . Consequently, the characteristic polynomial of  $B_{k+1}^{(1)}$  has the form

$$p_1(\lambda) = (\lambda - 1/\theta_{k+1})^{n-2} [\lambda^2 - (a_k + 2/\theta_{k+1})\lambda + 1/\theta_{k+1}^2]$$

and relation (12) follows immediately.

Note that the parameter  $a_k$  is bounded from below by 2, since

$$a_k = 1 + \theta_{k+1} \frac{y_k^T y_k}{s_k^T y_k} = 1 + \frac{\|s_k\|^2 \|y_k\|^2}{(s_k^T y_k)^2} = 1 + \frac{1}{\cos^2 \omega} \geq 2, \quad (14)$$

where  $\omega$  is the angle between the vectors  $s_k$  and  $y_k$ . If  $a_k = 2$ , then the characteristic polynomial is reduced to

$p_1(\lambda) = (\lambda - 1/\theta_{k+1})^n$ ; thus  $B_{k+1}^{(1)} = (1/\theta_{k+1})I = B_{k+1}^{(0)}$ . In different case (i.e.,  $a_k > 2$ ), the characteristic polynomial becomes  $p_1(\lambda) = (\lambda - 1/\theta_{k+1})^{n-2} (\lambda - \lambda_1)(\lambda - \lambda_2)$ , where the eigenvalues

$$\lambda_{1,2} = \frac{a_k \pm \sqrt{a_k^2 - 4}}{2\theta_{k+1}}$$

are distinct. From inequalities (13) follows that  $\min(\lambda_1, \lambda_2) < 1/\theta_{k+1} < \max(\lambda_1, \lambda_2)$ . Therefore, the extreme eigenvalues are distinct. ■

*Remark 1:* The scalar parameter  $a_k$  plays an important role in the updating process. In particular, if it happens that  $a_k = 2$ , relation (14) implies that  $s_k = \theta_{k+1} y_k$ , i.e.,  $s_k$  and  $y_k$  are collinear. In this case,  $B_{k+1}^{(1)}$  is reduced to a scaled unit matrix.

When  $B_{k+1} \equiv B_{k+1}^{(2)}$ , the approximate Hessian contains curvature information from the two previous iterations, i.e., the vector set of pairs contains the pairs  $\{(s_{k-1}, y_{k-1}), (s_k, y_k)\}$ . The updated matrix  $B_{k+1}^{(2)}$  is given by

$$B_{k+1}^{(2)} = B_k^{(1)} - \frac{B_k^{(1)} s_k s_k^T B_k^{(1)}}{s_k^T B_k^{(1)} s_k} + \frac{y_k y_k^T}{s_k^T y_k} \quad (15)$$

where  $B_k^{(1)}$  is defined in Eq. (6). The inverse of  $B_{k+1}^{(2)}$  can be expressed by

$$H_{k+1}^{(2)} = H_k^{(1)} - \frac{H_k^{(1)} y_k s_k^T + s_k y_k^T H_k^{(1)}}{s_k^T y_k} + \left(1 + \frac{y_k^T H_k^{(1)} y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k}, \quad (16)$$

where  $H_k^{(1)}$  is calculated by means of Eq. (7). In order to simplify the notations, we set

$$w_k = H_k^{(1)} y_k \quad \text{and} \quad b_k = 1 + \frac{y_k^T w_k}{s_k^T y_k}.$$

Thus, the expression of  $H_{k+1}^{(2)}$  is reduced to the following simpler form

$$H_{k+1}^{(2)} = H_k^{(1)} - \frac{w_k s_k^T + s_k w_k^T}{s_k^T y_k} + b_k \frac{s_k s_k^T}{s_k^T y_k}. \quad (17)$$

Taking into account Eqs. (9) and (10), it is straightforward to compute the determinant and the trace of  $B_{k+1}^{(2)}$ . The computation yields

$$\det(B_{k+1}^{(2)}) = \frac{1}{\theta_k^n} \frac{s_k^T y_k}{s_k^T B_k^{(1)} s_k}, \quad (18)$$

and

$$\text{tr}(B_{k+1}^{(2)}) = \frac{a_{k-1} + n - 2}{\theta_k} - \frac{\|B_k^{(1)} s_k\|^2}{s_k^T B_k^{(1)} s_k} + \frac{\|y_k\|^2}{s_k^T y_k}. \quad (19)$$

*Lemma 3:* The characteristic polynomial of  $B_{k+1}^{(2)}$  has the general form

$$p_2(\lambda) = \left(\lambda - \frac{1}{\theta_k}\right)^{n-4} (\lambda^4 - \beta_3\lambda^3 + \beta_2\beta_0\lambda^2 - \beta_1\beta_0\lambda + \beta_0), \quad (20)$$

where

$$\begin{aligned} \beta_0 &= \frac{1}{\theta_k^4} \frac{s_k^T y_k}{s_k^T B_k^{(1)} s_k}, \\ \beta_1 &= (a_{k-1} + 2)\theta_k - 2 \frac{s_k^T w_k}{s_k^T y_k} + b_k \theta_{k+1}, \\ \beta_2 &= \theta_k(\beta_1 - a_{k-1}\theta_k)(a_{k-1} + 2) - 2\theta_k^2 + \left(\frac{s_k^T w_k}{s_k^T y_k}\right)^2 \\ &\quad - \theta_{k+1} \frac{w_k^T w_k}{s_k^T y_k} + 2 \frac{s_k^T H_k^{(1)} w_k}{s_k^T y_k} - b_k \frac{s_k^T H_k^{(1)} s_k}{s_k^T y_k}, \quad \text{and} \\ \beta_3 &= \frac{a_{k-1} + 2}{\theta_k} - \frac{\|B_k^{(1)} s_k\|^2}{s_k^T B_k^{(1)} s_k} + \frac{\|y_k\|^2}{s_k^T y_k}. \end{aligned} \quad (21)$$

*Proof:* We first show that  $B_{k+1}^{(2)}$  has at most four distinct eigenvalues. Without loss of generality, we assume that  $B_k^{(1)}$  and  $B_{k+1}^{(2)}$  are positive definite matrices (in different case, the proof follows by similar arguments). Let  $\lambda_i^{B^{(1)}}$  and  $\lambda_i^{\bar{B}}$  be the eigenvalues of  $B_k^{(1)}$  and

$$\bar{B} = B_k^{(1)} - \frac{B_k^{(1)} s_k s_k^T B_k^{(1)}}{s_k^T B_k^{(1)} s_k},$$

respectively. Since  $\lambda_2^{B^{(1)}} = \dots = \lambda_{n-1}^{B^{(1)}} = 1/\theta_k$ , by Lemma 1 we have that

$$\begin{aligned} \lambda_n^{B^{(1)}} &\geq \lambda_n^{\bar{B}} \geq \frac{1}{\theta_k} \geq \lambda_{n-1}^{\bar{B}} \geq \frac{1}{\theta_k} \geq \dots \\ &\geq \frac{1}{\theta_k} \geq \lambda_2^{\bar{B}} \geq \lambda_1^{B^{(1)}} \geq \lambda_1^{\bar{B}}. \end{aligned}$$

The addition of the term  $\frac{y_k y_k^T}{s_k^T y_k}$  to the matrix  $\bar{B}$  yields

$$\begin{aligned} \lambda_n^{B^{(2)}} &\geq \lambda_n^{\bar{B}} \geq \lambda_{n-1}^{B^{(2)}} \geq \frac{1}{\theta_k} \geq \dots \\ &\geq \frac{1}{\theta_k} \geq \lambda_2^{B^{(2)}} \geq \lambda_2^{\bar{B}} \geq \lambda_1^{B^{(2)}} \geq \lambda_1^{\bar{B}}, \end{aligned}$$

where  $\lambda_i^{B^{(2)}}$  are the eigenvalues of  $B_{k+1}^{(2)}$ . The above inequalities imply that

$$\lambda_1^{B^{(2)}} \geq \lambda_2^{B^{(2)}} \geq 1/\theta_k \geq \lambda_{n-1}^{B^{(2)}} \geq \lambda_n^{B^{(2)}}, \quad (22)$$

and thus, the matrix  $B_{k+1}^{(2)}$  has at most four distinct eigenvalues. If we denote by  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  the four unknown eigenvalues, the characteristic polynomial of  $B_{k+1}^{(2)}$  takes the form

$$p_2(\lambda) = (\lambda - 1/\theta_k)^{n-4} (\lambda^4 - c_3\lambda^3 + c_2\lambda^2 - c_1\lambda + c_0),$$

where  $c_3 = \sum_{i=1}^4 \lambda_i$ ,  $c_2 = \sum_{i<j} \lambda_i \lambda_j$ ,  $c_1 = \sum_{i<j<\ell} \lambda_i \lambda_j \lambda_\ell$ , and  $c_0 = \prod_{i=1}^4 \lambda_i$ , and moreover,

$\prod_{i=1}^4 \lambda_i = \det(B_{k+1}^{(2)}) \theta_k^{n-4}$  and  $\sum_{i=1}^4 \lambda_i = \text{tr}(B_{k+1}^{(2)}) - \frac{n-4}{\theta_k}$ . Utilizing Eqs. (18) and (19) as well as the fact that the eigenvalues of  $B_{k+1}^{(2)}$  and  $H_{k+1}^{(2)}$  are inverses of each other, after some algebraic manipulations, we yield the expression in Eq. (20) with the parameters in Eqs. (21). ■

In case where  $B_{k+1}^{(2)}$  has four distinct eigenvalues, these can be computed by analytically solving the quartic equation  $\lambda^4 - \beta_3\lambda^3 + \beta_2\beta_0\lambda^2 - \beta_1\beta_0\lambda + \beta_0 = 0$ , using standard methods [6], [7]. However, the exact number of the distinct eigenvalues is strongly depends on the way the vector pairs  $(s_i, y_i)$  in the vector set are related. The following proposition establishes sufficient conditions for the exact number of distinct eigenvalues.

*Proposition 1:* Let the symmetric matrix  $B_{k+1}^{(2)} \in \mathbb{R}^{n \times n}$  as defined in Eq. (15).

1. If  $a_{k-1} = 2$  then  $B_{k+1}^{(2)}$  has at most two distinct eigenvalues. Moreover,

- (a) if  $a_k > 2$ , then two distinct eigenvalues exist,
- (b) if  $a_k = 2$  and  $\theta_k \neq \theta_{k+1}$ , then only one distinct eigenvalue exists, and
- (c) if  $a_k = 2$  and  $\theta_k = \theta_{k+1}$ , then all the eigenvalues are equal to  $1/\theta_k$ .

2. If  $a_{k-1} > 2$ , then  $B_{k+1}^{(2)}$  has at least three distinct eigenvalues. Moreover,

- (a) if  $a_k = 2$ , then only three distinct eigenvalues exist, and
- (b) if  $a_k > 2$ , then only four distinct eigenvalues exist.

*Proof:* Assume that  $a_{k-1} = 2$ . Then, in Lemma 2 we have proved that, in this case,  $B_k^{(1)} = (1/\theta_k)I$ . Consequently,  $B_{k+1}^{(2)}$  has the following form

$$B_{k+1}^{(2)} = \frac{1}{\theta_k} I - \frac{1}{\theta_k} \frac{s_k s_k^T}{s_k^T s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \quad (23)$$

while its characteristic polynomial becomes

$$p_2(\lambda) = \left(\lambda - \frac{1}{\theta_k}\right)^{n-2} \left[\lambda^2 - \left(\frac{a_k-1}{\theta_{k+1}} + \frac{1}{\theta_k}\right)\lambda + \frac{1}{\theta_k \theta_{k+1}}\right]. \quad (24)$$

In view of Eq. (24), it follows immediately that  $B_{k+1}^{(2)}$  has at most two distinct eigenvalues. Consider now the quadratic equation  $\lambda^2 - \left(\frac{a_k-1}{\theta_{k+1}} + \frac{1}{\theta_k}\right)\lambda + \frac{1}{\theta_k \theta_{k+1}} = 0$ , which has the following two solutions

$$\begin{aligned} \lambda_{1,2} &= \frac{\theta_{k+1} + \theta_k (a_k - 1)}{2\theta_k \theta_{k+1}} \\ &\pm \frac{\sqrt{(\theta_k - \theta_{k+1})^2 + \theta_k (a_k - 2)(a_k \theta_k + 2\theta_{k+1})}}{2\theta_k \theta_{k+1}} \end{aligned}$$

Let  $a_k > 2$  and suppose that either  $\lambda_1$  or  $\lambda_2$  equals to  $1/\theta_k$ . Solving the equations  $\lambda_i - 1/\theta_k = 0$  with respect to  $a_k$  we obtain  $a_k = 2$ , which is a contradiction. Thus, when  $a_k > 2$  the matrix has exactly two distinct eigenvalues. When  $a_k = 2$ , the characteristic polynomial becomes  $p_2(\lambda) = (\lambda - 1/\theta_k)^{n-1} (\lambda - 1/\theta_{k+1})$ . Therefore, in case where  $\theta_k \neq \theta_{k+1}$ , the only distinct eigenvalue is  $1/\theta_{k+1}$ ; otherwise  $B_{k+1}^{(2)}$  does not have distinct eigenvalues.

Assume now that  $a_{k-1} > 2$ . In this case,  $B_k^{(1)}$  has two distinct eigenvalues. Thus, the matrix  $\bar{B} = B_k^{(1)} - \frac{B_k^{(1)} s_k s_k^T B_k^{(1)}}{s_k^T B_k^{(1)} s_k}$  besides the zero eigenvalue, has two more distinct eigenvalues. Consequently,  $B_{k+1}^{(2)}$  has at least three distinct eigenvalues. If  $a_k = 2$ , then  $s_k = \theta_{k+1} y_k$  and  $B_{k+1}^{(2)}$  becomes

$$B_{k+1}^{(2)} = B_k^{(1)} - \frac{B_k^{(1)} y_k y_k^T B_k^{(1)}}{y_k^T B_k^{(1)} y_k} + \frac{y_k y_k^T}{\theta_{k+1} y_k^T y_k}. \quad (25)$$

Hence,  $\bar{B} = B_k^{(1)} - \frac{B_k^{(1)} y_k y_k^T B_k^{(1)}}{y_k^T B_k^{(1)} y_k}$ , and therefore, the addition of  $y_k y_k^T / \theta_{k+1} y_k^T y_k$  changes the zero eigenvalue of  $\bar{B}$  to  $1/\theta_{k+1}$  and leaves the others unchanged. Thus,  $B_{k+1}^{(2)}$  has as many distinct eigenvalues as  $\bar{B}$ , i.e., three. Since one of them is  $1/\theta_{k+1}$ , utilizing Lemma 3, the characteristic polynomial of  $B_{k+1}^{(2)}$  becomes

$$p_2(\lambda) = (\lambda - 1/\theta_k)^{n-3} (\lambda - 1/\theta_{k+1})(\lambda^2 - c_1 \lambda + c_2),$$

where  $c_1 = a_{k-1}/\theta_k - \|B_k^{(1)} y_k\|^2 / y_k^T B_k^{(1)} y_k$  and  $c_2 = y_k^T y_k / (\theta_k^2 \theta_{k+1} y_k^T B_k^{(1)} y_k)$ . By solving the quadratic equation  $\lambda^2 - c_1 \lambda + c_2 = 0$ , we obtain the other two distinct eigenvalues. Finally, if  $a_k > 2$ , then  $y_k$  is not an eigenvector of  $\bar{B} = B_k^{(1)} - \frac{B_k^{(1)} s_k s_k^T B_k^{(1)}}{s_k^T B_k^{(1)} s_k}$ , and thus, the addition of the term  $y_k y_k^T / s_k^T y_k$  results one more distinct eigenvalue for  $B_{k+1}^{(2)}$ . ■

*Lemma 4:* If  $B_{k+1}^{(2)}$  has at least two distinct eigenvalues, then

$$\lambda_1 < 1/\theta_k < \lambda_n, \quad (26)$$

where  $\lambda_1$  and  $\lambda_n$  are the smallest and largest eigenvalues of  $B_{k+1}^{(2)}$ , respectively.

*Proof:* When  $B_{k+1}^{(2)}$  has at least three distinct eigenvalues, from relation (22) it is evident that (26) holds. Now, if  $B_{k+1}^{(2)}$  has two distinct eigenvalues, then one pair of the vector set  $\{(s_{k-1}, y_{k-1}), (s_k, y_k)\}$  must be collinear. Due to Proposition (1), these two vectors are  $s_{k-1}$  and  $y_{k-1}$ , which implies that  $B_k^{(1)}$  has no distinct eigenvalues. Under these circumstances, Lemma 1 implies relation (26), which completes the proof. ■

#### IV. COMPUTATION OF THE DESCENT PAIR OF DIRECTIONS

The descent pair of directions used on the curvilinear scheme (2) consists of the quasi-Newton direction  $p$  defined as

$$p_{k+1} = \begin{cases} -H_{k+1}^{(m)} g_{k+1}, & B_{k+1}^{(m)} \geq 0; \\ -g_{k+1}, & \text{otherwise.} \end{cases} \quad (27)$$

and the direction of negative curvature defined as

$$d_k = \begin{cases} 0, & B_{k+1}^{(m)} \geq 0; \\ -\text{sgn}(u_k^T g_k) u_k, & \text{otherwise,} \end{cases} \quad (28)$$

where  $u_k$  is a normalized vector that corresponds to the most negative eigenvalue of  $B^{(m)}$ .

The quasi-Newton direction is obtained by the L-BFGS formula, for  $m = 1, 2$ . The resulting direction is

$$p_{k+1} = \begin{cases} -\theta_{k+1} g_{k+1} - \left( a_k \frac{s_k^T g_{k+1}}{s_k^T y_k} - \theta_{k+1} \frac{y_k^T g_{k+1}}{s_k^T y_k} \right) s_k \\ \quad + \theta_{k+1} \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k, & \text{for } m = 1; \\ -H_k^{(1)} g_{k+1} + \frac{w_k^T g_{k+1} - b_k s_k^T g_{k+1}}{s_k^T y_k} s_k \\ \quad + \frac{s_k^T g_{k+1}}{s_k^T y_k} w_k, & \text{for } m = 2, \end{cases} \quad (29)$$

where the matrix-vector product  $H_k^{(1)} g_{k+1}$  is computed as in the case where  $m = 1$ .

For being able to calculate the second direction, i.e., the eigenvector that corresponds to the most negative eigenvalue, we make use of the inverse power method.

Given a non-zero starting vector  $u^{(0)}$ , inverse iteration generates a sequence of vectors  $u^{(i)}$ , generated recursively by the formula

$$u^{(i)} = \left( B^{(m)} - \hat{\lambda} I \right)^{-1} \frac{u^{(i-1)}}{\|u^{(i-1)}\|},$$

$i \geq 1$ , where  $\hat{\lambda} = \lambda + \epsilon$ ,  $\lambda$  is a distinct eigenvalue of  $B^{(m)}$  and  $\epsilon \rightarrow 0$ . The sequence of iterates  $u^{(i)}$  converges to an eigenvector associated with an eigenvalue closest to  $\hat{\lambda}$ . Usually, the starting vector  $u^{(0)}$  is chosen to be  $(1, 1, \dots, 1)^T$ . Moreover, if this particular eigenvalue  $\lambda$  is known exactly, this method converges in a single iteration [11].

Since the smallest eigenvalue  $\lambda_1$  of  $B^{(m)}$  is known exactly, if it is also a distinct eigenvalue, using the inverse iteration, we can find the analytical form for the corresponding eigenvector.

*Proposition 2:* Let the symmetric matrix  $B_{k+1}^{(m)} \in \mathbb{R}^{n \times n}$ . If the extreme eigenvalues of  $B_{k+1}^{(m)}$  are distinct, then the unit eigenvector corresponding to the most negative eigenvalue  $\lambda_1$  is of the form

$$u_{k+1} = \left[ \left( B_{k+1}^{(m)} \right)^{2m} + \sum_{i=0}^{2m-1} (-1)^i \nu_i^{(m)}(\hat{\lambda}) \left( B_{k+1}^{(m)} \right)^i \right] \frac{u}{\nu^{(m)}(\hat{\lambda}) \|u\|}, \quad (30)$$

where  $\hat{\lambda} = \lambda_1 + \epsilon$ ,  $\epsilon \rightarrow 0$ ,  $u = (1, 1, \dots, 1)^T$  and the functions  $\nu_i^{(m)}(\hat{\lambda})$  are defined as follows: for  $m = 1$ ,

$$\begin{aligned} \nu^{(1)}(\hat{\lambda}) &= (1/\theta_{k+1} - \hat{\lambda})(\hat{\lambda}^2 - a_k \hat{\lambda}/\theta_{k+1} + 1/\theta_{k+1}^2), \\ \nu_1^{(1)}(\hat{\lambda}) &= -\hat{\lambda} + (a_k + 1)/\theta_{k+1} \quad \text{and} \\ \nu_0^{(1)}(\hat{\lambda}) &= -\hat{\lambda}^2 - (a_k + 1)\hat{\lambda}/\theta_{k+1} + (a_k + 1)/\theta_{k+1}^2, \end{aligned}$$

and for  $m = 2$ ,

$$\begin{aligned} \nu_3^{(2)}(\hat{\lambda}) &= -\hat{\lambda} + \beta_3 + 1/\theta_k, \\ \nu_2^{(2)}(\hat{\lambda}) &= -\hat{\lambda} \nu_3^{(2)}(\hat{\lambda}) + \beta_2 \beta_0 + \beta_3/\theta_k, \\ \nu_1^{(2)}(\hat{\lambda}) &= -\hat{\lambda} \nu_2^{(2)}(\hat{\lambda}) + \beta_1 \beta_0 + \beta_2 \beta_0/\theta_k, \\ \nu_0^{(2)}(\hat{\lambda}) &= -\hat{\lambda} \nu_1^{(2)}(\hat{\lambda}) + \beta_0 + \beta_0 \beta_1/\theta_k \quad \text{and} \\ \nu^{(2)}(\hat{\lambda}) &= -\hat{\lambda} \nu_0^{(2)}(\hat{\lambda}) + \beta_0/\theta_k, \end{aligned}$$

while the parameters  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$  are defined as in Lemma 3.

*Proof:* We consider the following cases:

(i)  $m = 1$ : Let  $\lambda_1$  and  $\lambda_2$  be the extreme eigenvalues of  $B_{k+1}^{(1)}$ . Consequently, the matrix  $(B_{k+1} - \lambda I)$  has two distinct eigenvalues,  $\lambda_1 - \lambda$  and  $\lambda_2 - \lambda$ . Using Eq. (12) after some algebraic calculations, the characteristic polynomial of  $(B_{k+1}^{(1)} - \lambda I)$  is

$$q(x) = (x - c)^{n-2} (x^2 - c_1 x + c_0),$$

where the quantities  $c, c_1$  and  $c_0$  are defined as

$$c = \frac{1}{\theta_{k+1}} + \lambda, \quad c_1 = \frac{a_k}{\theta_{k+1}} - 2\lambda, \quad \text{and}$$

$$c_0 = \lambda^2 - \frac{a_k \lambda}{\theta_{k+1}} + \frac{1}{\theta_{k+1}^2}.$$

Therefore, its minimal polynomial is

$$q_m(x) = (x - c) (x^2 - c_1 x + c_0).$$

Using the Caley-Hamilton theorem, we have that  $q_m(B_{k+1}^{(1)} - \lambda I) = 0$ . Multiplying both sides of the equation by  $(B_{k+1}^{(1)} - \lambda I)^{-1}$  and after some algebraic computations, we obtain

$$(B_{k+1}^{(1)} - \lambda I)^{-1} = \frac{1}{\nu^{(1)}(\lambda)} [(B_{k+1}^{(1)})^2 - \nu^{(1)}(\lambda) B_{k+1}^{(1)} + \nu_0^{(1)}(\lambda) I]. \quad (31)$$

Applying the inverse power method, we obtain Eq. (30).

(ii)  $m = 2$ : From Proposition 1 we know that  $B_{k+1}^{(2)}$  has at most four distinct eigenvalues,  $\lambda_i, i = 1, \dots, 4$ . Thus,  $(B_{k+1}^{(2)} - \lambda I)$  has also at most four distinct eigenvalues,  $\lambda_i - \lambda$ . Utilizing Lemma 3, the characteristic polynomial of  $(B_{k+1}^{(2)} - \lambda I)$  takes the form

$$q(x) = (x - c)^{n-4} (x^4 - c_3 x^3 + c_2 x^2 - c_1 x + c_0),$$

where  $c = \theta_k^{-1} - \lambda, c_3 = -4\lambda + \beta_3, c_2 = 6\lambda^2 - 3\beta_3\lambda + \beta_2\beta_0, c_1 = -4\lambda^3 + 3\beta_3\lambda^2 - 2\beta_2\beta_0\lambda + \beta_1\beta_0,$  and  $c_0 = \lambda^4 - \beta_3\lambda^3 + b_2\beta_0\lambda^2 - \beta_1\beta_0\lambda + \beta_0,$  where the constants  $\beta_i$  are defined in Lemma 3. In the general case, its minimal polynomial is

$$q_m(x) = (x - c)(x^4 - c_3 x^3 + c_2 x^2 - c_1 x + c_0).$$

Therefore, using the Caley-Hamilton theorem, we obtain

$$(B_{k+1}^{(2)} - \lambda I)^{-1} = \frac{1}{\nu^{(2)}(\lambda)} \left[ (B_{k+1}^{(2)})^4 - \nu_3^{(2)}(\lambda) (B_{k+1}^{(2)})^3 + \nu_2^{(2)}(\lambda) (B_{k+1}^{(2)})^2 - \nu_1^{(2)}(\lambda) B_{k+1}^{(2)} + \nu_0^{(2)}(\lambda) I \right]. \quad (32)$$

Finally, by applying the inverse power method, we obtain (Eq. 30). ■

When  $m = 1$ , Eq. (30) yields

$$u_{k+1} = -\nu_u^{(1)}(\hat{\lambda}) u + \nu_{us}^{(1)}(\hat{\lambda}) s_k - \nu_{uy}^{(1)}(\hat{\lambda}) y_k, \quad (33)$$

where

$$\nu_u^{(1)}(\hat{\lambda}) = \frac{1 - \nu_1^{(1)}(\hat{\lambda}) \theta_{k+1} + \nu_0^{(1)}(\hat{\lambda}) \theta_{k+1}^2}{\nu^{(1)}(\hat{\lambda}) \theta_{k+1}^2},$$

$$\nu_{us}^{(1)}(\hat{\lambda}) = \frac{[1 - \nu_1^{(1)}(\hat{\lambda}) \theta_{k+1}] s_k^T u + \theta_{k+1} y_k^T u}{\nu^{(1)}(\hat{\lambda}) \theta_{k+1}^2 s_k^T s_k}, \quad \text{and}$$

$$\nu_{uy}^{(1)}(\hat{\lambda}) = \frac{[1 - \nu_1^{(1)}(\hat{\lambda}) \theta_{k+1} + a_k] \theta_{k+1} y_k^T u - s_k^T u}{\nu^{(1)}(\hat{\lambda}) \theta_{k+1}^2 s_k^T y_k}.$$

When  $m = 2$ , the eigenvector can be computed by the following procedure:

*Procedure 1:* (Compute  $u_{k+1}$ )

$u_{k+1}^{(4)} = u;$   
**for**  $j = 4, \dots, 1$  **do**  
 $v \leftarrow B^{(2)} u_{k+1}^{(j)};$   
 $u_{k+1}^{(j-1)} \leftarrow v + (-1)^{j-1} \nu_{j-1}^{(2)}(\hat{\lambda}) u;$

**end for**

**return**  $u_{k+1}^{(0)}/\nu^{(2)}(\lambda);$

The vector  $v \in \mathbb{R}^n$  is computed by the formula

$$v = B_k^{(1)} u_1^{(j)} - \frac{s_k^T B_k^{(1)} u_1^{(j)}}{s_k^T B_k^{(1)} s_k} B_k^{(1)} s_k + \frac{y_k^T u_1^{(j)}}{s_k^T y_k} y_k,$$

where

$$B_k^{(1)} u_1^{(j)} = \frac{1}{\theta_k} u_1^{(j)} - \frac{s_{k-1}^T u_1^{(j)}}{\theta_k s_{k-1}^T s_{k-1}} s_{k-1} + \frac{y_{k-1}^T u_1^{(j)}}{s_{k-1}^T y_{k-1}} y_{k-1},$$

and

$$B_k^{(1)} s_k = \frac{1}{\theta_k} s_k - \frac{s_{k-1}^T s_k}{\theta_k s_{k-1}^T s_{k-1}} s_{k-1} + \frac{y_{k-1}^T s_k}{s_{k-1}^T y_{k-1}} y_{k-1}.$$

When the extreme eigenvalues of  $B^{(m)}$  are not distinct, due to the structure of  $B^{(m)}$  the desirable eigenvector can be obtained by simple algebraic computations. We consider the following cases:

(i.) In case where  $m = 1$ , when the extreme eigenvalues of  $B_{k+1}^{(1)}$  are not distinct, from Lemma 2 we have that  $a_k = 2$  and it is easy to see that Eq. (31) is reduced to

$$(B_{k+1}^{(1)} + \lambda I)^{-1} = \left( \frac{1}{\theta_{k+1}} + \lambda \right)^{-1} I. \quad (34)$$

Using the decomposition of  $B_{k+1}^{(1)}$  we have  $B_{k+1}^{(1)} = U \Lambda U^T$ , where  $U = I$  and  $\Lambda = \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_1)$ . Obviously a corresponding eigenvector to  $\lambda_1$  is

$$u_{k+1} = e_1 = (1, 0, \dots, 0)^T.$$

(ii.) In case where  $m = 2$ , when the extreme eigenvalues of  $B_{k+1}^{(2)}$  are not distinct, from Proposition 1 along with

Lemma 4 we have that  $a_{k-1} = a_k = 2$  and  $B_{k+1}^{(2)}$  has the form

$$B_{k+1}^{(2)} = \frac{1}{\theta_k} I + \left( \frac{1}{\theta_{k+1}} - \frac{1}{\theta_k} \right) \frac{y_k y_k^T}{y_k^T y_k}. \quad (35)$$

(a) if  $\theta_k \neq \theta_{k+1}$  and  $\lambda_1 = 1/\theta_{k+1}$ , from Eq. (35) it is straightforward to see that  $u_{k+1} = y_k / \|y_k\|$ .

(b) if  $\theta_k \neq \theta_{k+1}$  and  $\lambda_1 = 1/\theta_k$  from Eq. (35) easily can be verified that the resulting eigenvector is of the form  $u_{k+1} = \hat{u}_{k+1} / \|\hat{u}_{k+1}\|$ , where

$$\hat{u}_{k+1} = \left( -\frac{y_k^{(n)}}{y_k^{(1)}}, 0, \dots, 0, 1 \right)^T \quad (36)$$

and  $y_k^{(i)}$  denotes the  $i$ th component of  $y_k$ .

(c) finally, if  $\theta_k = \theta_{k+1}$ , using the decomposition of  $B_{k+1}^{(2)}$  we have that  $u_{k+1} = e_1 = (1, 0, \dots, 0)^T$ .

## V. THE ALGORITHM

We consider an iterative scheme of the form

$$x_{k+1} = \begin{cases} x_k + \alpha_k p_k, & \text{when } B_k^{(m)} \geq 0; \\ x_k + \alpha_k^2 p_k + \alpha_k d_k, & \text{otherwise,} \end{cases} \quad (37)$$

where  $p_k$  is the quasi-Newton direction,  $d_k$  is the negative curvature direction and  $\alpha_k$  is a step size. The following algorithm describes a curvilinear search based on an Armijio procedure.

*Algorithm 1:* curvilinear search

*Step 1:* Given  $x_0$ , integer  $m \in [1, 2]$ ,  $0 < \sigma_1 < \sigma_2 \leq 1$ ,  $0 < c_2 < c_1 < 1$ ,  $\epsilon \rightarrow 0$  and  $\epsilon_1 \rightarrow 0$ ; set  $k = 0$  and  $\ell = 0$ ; compute  $g_0$ ;

*Step 2:* If  $\|g_k\| \leq \epsilon$  stop; else compute the eigenvalues  $\lambda_i$  of  $B_k^{(m)}$ ;

*Step 3:* If  $\lambda_1 \geq 0$  then

- a) compute  $p_k$ ; set  $d_k = 0$  and  $\alpha_k = 1$ ;
- b) Find  $\alpha_k > 0$  such that

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k g_k^T p_k;$$

Else

c) set  $p_k = -g_k$  and compute  $u_k$ ; set  $d_k = -\text{sgn}(u_k^T g_k) u_k$  and  $\alpha_k = 1$ ;

d) Choose  $\sigma \in [\sigma_1, \sigma_2]$ ,  $\alpha_k = \sigma \alpha_k$  such that

$$f(x_k + \alpha_k^2 p_k + \alpha_k d_k) \leq f(x_k) + c_2 \alpha_k^2 \left( g_k^T p_k + \frac{1}{2} \lambda_1 \right).$$

*Step 4:* Set

$$x_{k+1} = \begin{cases} x_k + \alpha_k p_k, & \text{if } \lambda_1 \geq 0; \\ x_k + \alpha_k^2 p_k + \alpha_k d_k, & \text{otherwise.} \end{cases}$$

*Step 5:* Compute  $g_{k+1}$ ,  $s_k = x_{k+1} - x_k$  and  $y_k = g_{k+1} - g_k$ ; if  $|s_k^T y_k| > \epsilon_1$ , save the vector pair  $\{s_k, y_k\}$  and set  $\ell = \ell + 1$ ; if  $\ell > m$  discard the vector pair  $\{s_{\ell-m}, y_{\ell-m}\}$  from storage;

*Step 6:* Set  $k = k + 1$  and go to step 2;

When  $m = 1$ , the eigenvalues in Step 2 of Algorithm 1 can be computed by Eq. (12). In Step 3a), the direction  $p_k$  is obtained by Eq. (29) for  $m = 1$ , while in Step 3c) if

$a_k > 2$ , the unit eigenvector  $u_k$  is obtained using Eq. (33), otherwise we set  $u_1 = e_1$ .

When  $m = 2$ , the eigenvalues can be computed by means of Eq. (20). The computation of  $p_k$  in Step 3a), is obtained by Eq. (29) for  $m = 2$ . In Step 3(c), if  $a_{k-1} > 2$  or  $a_k > 2$ , then  $u_k$  is computed using Procedure 1, else if  $a_{k-1} = a_k = 2$ , we have:

1. if  $\lambda_1$  is distinct eigenvalue, then  $u_k = y_k / \|y_k\|$ ,
2. if  $\lambda_1$  has multiplicity  $n - 1$ , then  $u_k = \hat{u}_k / \|\hat{u}_k\|$ , where  $\hat{u}_k$  is defined in Eq (36),
3. if  $\lambda_1$  has multiplicity  $n$ , then  $u_k = e_1$ .

In Step 3b) the line search procedure is based on truncated quadratic interpolation. More analytically, if  $f(x_k + \alpha_k p_k) \geq f(x_k)$ , then a step length  $\alpha_k$  is computed by the formula

$$\alpha_k = \max \{0.1, 0.5 / [1 + (f(x_k) - f(x_k + \alpha_k p_k)) / p_k^T g_k]\},$$

and the new direction is set to be  $p_k := \alpha_k p_k$ . This process is repeated until a lower function value is obtained. In Step 3d) the curvilinear procedure is based on the method of bisection.

## VI. NUMERICAL RESULTS

We have implemented Algorithm 1 using 1-BFGS and 2-BFGS updating schemes. Both algorithms have been coded in MATLAB 7.3 and all numerical experiments were per-

TABLE I  
1- BFGS UPDATE

Dimension	$n = 10^3$	$n = 10^4$	$n = 10^5$
No.	FE-GE	FE-GE	FE-GE
1.	107-69	96-74	93-75
2.	146-51	205-72	150-56
3.	53-28	57-26	65-30
4.	55-54	87-71	127-86
5.	196-79	425-159	561-166
6.	701-301	513-236	730-281
7.	39-17	42-19	1255-293
8.	151-33	174-31	166-27
9.	100-56	56-33	74-35
10.	11-11	12-12	12-12
11.	12-12	12-12	13-13
12.	14-14	14-14	15-15
13.	28-22	30-27	201-55
14.	54-54	70-70	110-88
15.	176-51	112-29	329-63
16.	32-25	73-43	29-23
17.	15-15	16-16	18-18
18.	34-26	31-24	38-30

formed on a Pentium 1.86 (GHz) personal computer with 1GB of RAM running a Linux operating system. Double precision IEEE floating point arithmetic with machine precision approximately  $2.2 \times 10^{-16}$  was employed.

We have selected 18 large-scale unconstrained optimization test problems in extended or generalized form. These problems from 1 to 18 are: Trigonometric, Rosenbrock,

Beale, Penalty, Powell, Wood, Arwead, Nondia, Dqdrtic, Dixmaana, Dixmaanb, Dixmaanc, Edensch, Vardim, Liarwhd, Engvall, Cosine, Denschnb and Denschnf. The first 6 test problems were selected by the 18 problems of Moré, B.S. Garbow, and K.E. Hillstrom [14] while the rest of them were selected by the CUTE collection [10].

For each test function we have considered 3 numerical experiments with number of variables  $n = 10^3, 10^4, 10^5$ . In both algorithms the implementation is terminated if  $\|g_k\| \leq 10^{-5}$ , while  $B_k$  is allowed to be indefinite. Thus, the condition  $s_k^T y_k > 0$  does not always hold and therefore, for being  $B_k$  well defined, the storage of the vector pair  $\{s_k, y_k\}$  is skipped if  $|s_k^T y_k| \leq 10^{-12} \|s_k\| \|y_k\|$ .

TABLE II  
2- BFGS UPDATE

Dimension	$n = 10^3$	$n = 10^4$	$n = 10^5$
No.	FE-GE	FE-GE	FE-GE
1.	88-73	88-74	70-64
2.	87-44	96-46	111-54
3.	24-17	28-22	33-23
4.	46-45	57-57	68-68
5.	620-147	308-90	824-189
6.	465-178	502-184	485-184
7.	13-10	14-10	16-10
8.	25-13	19-11	21-11
9.	57-27	90-44	86-49
10.	11-11	12-12	12-12
11.	12-12	13-13	13-13
12.	14-14	14-14	15-15
13.	30-26	77-35	34-27
14.	54-54	70-70	110-88
15.	47-29	54-27	60-25
16.	18-14	90-50	84-53
17.	12-12	14-14	15-15
18.	28-26	28-26	28-26

In both tables the first row denotes the number of variables of each problem. We list the number of function (FE) and gradient (GE) evaluations, as the number of iterations is the same as the number of gradient evaluations.

Table I summarizes the results of the 1-BFGS algorithm applied to 18 different problems with varying sizes ranging from  $10^3$  to  $10^5$ . To solve all the 18 problems, the CPU time (in seconds) required by the 1-BFGS algorithm is 1.56, 28.69, and 114.00, for each dimension, respectively.

Table II summarizes the results of the performance of the 2-BFGS algorithm applied to the same problems as in Table I. To solve all the 18 problems, the CPU time (in seconds) required by the 2-BFGS algorithm is 5.16, 98.58, and 2357.67, for each dimension, respectively.

As we can see the number of function and gradient evaluations is lower in 2-BFGS algorithm than 1-BFGS algorithm. However, the 1-BFGS algorithm requires fewer CPU time to solve all test problems.

## REFERENCES

- [1] L. Armijio. Minimization of functions having lipschitz continous first partial derivatives. *Pacific Journal of Mathematics*, 16:1-3, 1966.
- [2] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, 1996.
- [3] J. Barzilai and J.M. Borwein. Two point step size gradient method. *IMA J. Numer. Anal.*, 8:141-148, 1988.
- [4] E. G. Birgin and J. M. Martínez. A spectral conjugate gradient method for unconstrained optimization. *Appl. Math. Optim.*, 43:117-128, 2001.
- [5] E.G. Birgin, J.M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10(6):1196-1211, 2000.
- [6] P. Borwein and T. Erdélyi. *Polynomials and Polynomial Inequalities*, volume 161 of *Graduate Texts in Mathematics*. Springer-Verlag, NY, 1995.
- [7] W. M. Faucette. A geometric interpretation of the solution of the general quartic polynomial. *American Mathematical Monthly*, 103(1):51-57, January 1996.
- [8] A. V. Fiacco and G. P. McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*. Wiley, New York, 1968.
- [9] D. Goldfarb. Curvilinear path steplength algorithms for minimization which use directions of negative curvature. *Math. Programming*, 18:31-40, 1980.
- [10] I.Bongartz, A.R.Conn, N.I.M. Gould, and Ph.L. Toint. Cute: constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software*, 21:123-160, 1995.
- [11] I. C. F. Ipsen. Computing an eigenvector with inverse iteration. *SIAM Review*, 39(2):254-291, June 1997.
- [12] G. P. McCormick. A modification of Armijo's step-size rule for negative curvature. *Math. Programming*, 13:111-115, 1977.
- [13] J. Moré and D. Sorensen. On the use of directions of negative curvature in a modified Newton method. *Math. Programming*, 16:1-20, 1979.
- [14] J.J. Moré, B.S. Garbow, and K.E. Hillstrom. Testing unconstrained optimization software. *ACM Transactions on Mathematical Software*, 4:553-572, 1981.
- [15] J. Nocedal. Updating quasi-Newton matrices with limited storage. *Math. Comput.*, 35:773-782, 1980.
- [16] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 1999.
- [17] M. Raydan. On the Barzilai and Borwein choice of steplength for the gradient-method. *IMA J. Numer. Anal.*, 13(3):321-326, 1993.
- [18] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.*, 7(1):26-33, 1997.
- [19] J.H. Wilkinson. The calculation of the eigenvectors of codiagonal matrices. *Computer J.*, 1:90-96, 1958.
- [20] J.H. Wilkinson. *The algebraic eigenvalue problem*. Oxford University Press, 1965.