Math. Program., Ser. A (2013) 138:501–530
DOI 10.1007/s10107-012-0522-2

FULL LENGTH PAPER

A perfect example for the BFGS method

Yu-Hong Dai

Received: 12 May 2010 / Accepted: 28 February 2012 / Published online: 22 March 2012

Abstract Consider the BFGS quasi-Newton method applied to a general non-convex function that has continuous second derivatives. This paper aims to construct a four-dimensional example such that the BFGS method need not converge. The example is perfect in the following sense: (a) All the stepsizes are exactly equal to one; the unit stepsize can also be accepted by various line searches including the Wolfe line search and the Arjimo line search; (b) The objective function is strongly convex along each search direction although it is not in itself. The unit stepsize is the unique minimizer of each line search function. Hence the example also applies to the global line search and the line search that always picks the first local minimizer; (c) The objective function is polynomial and hence is infinitely continuously differentiable. If relaxing the convexity requirement of the line search function; namely, (b) we are able to construct a relatively simple polynomial example.

Keywords Unconstrained optimization · Quasi-Newton method · Non-convex function · Global convergence

Mathematics Subject Classification 49M37 · 90C30

1 Introduction

Consider the unconstrained optimization problem

$$\min f(\mathbf{x}), \quad \mathbf{x} \in \mathcal{R}^n, \tag{1.1}$$

Y.-H. Dai (✉)
State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
P.O. Box 2719, Beijing 100190, People's Republic of China
e-mail: dyh@lsec.cc.ac.cn

where f is a general non-convex function that has continuous second derivatives. The quasi-Newton method is a class of well-known and efficient methods for solving (1.1). It is of the iterative scheme

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k H_k \mathbf{g}_k, \tag{1.2}$$

where $\mathbf{x}_1$ is a starting point, H_k is some approximation to the inverse Hessian $[\nabla^2 f(\mathbf{x}_k)]^{-1}$, $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$, and α_k is a stepsize obtained in some way. Defining the vectors

$$\boldsymbol{\delta}_k = \mathbf{x}_{k+1} - \mathbf{x}_k, \quad \boldsymbol{\gamma}_k = \mathbf{g}_{k+1} - \mathbf{g}_k,$$

the quasi-Newton method asks the next approximation matrix H_{k+1} to satisfy the secant equation

$$H_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k. \tag{1.3}$$

This similarity to the identical equation $[\nabla^2 f(\mathbf{x}_{k+1})]^{-1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k$ in the quadratic case enables the quasi-Newton method to be superlinearly convergent (see Dennis and Moré [5], for example) and makes it very attractive in practical optimization.

The first quasi-Newton method was dated back to Davidon [4] and Fletcher and Powell [8]. The DFP method updates the approximation matrix H_k to H_{k+1} by the formula

$$H_{k+1} = H_k - \frac{H_k\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T H_k}{\boldsymbol{\gamma}_k^T H_k \boldsymbol{\gamma}_k} + \frac{\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k}. \tag{1.4}$$

Nowadays, the most efficient quasi-Newton method is perhaps the BFGS method, which was proposed by Broyden [2], Fletcher [6], Goldfarb [9], and Shanno [18], independently. The matrix H_{k+1} in the BFGS method can be updated by the way

$$H_{k+1} = H_k - \frac{\boldsymbol{\delta}_k\boldsymbol{\gamma}_k^T H_k + H_k\boldsymbol{\gamma}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k} + \left(1 + \frac{\boldsymbol{\gamma}_k^T H_k\boldsymbol{\gamma}_k}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k}\right)\frac{\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k}. \tag{1.5}$$

To pass the positive definiteness of the matrix H_k to H_{k+1}, practical quasi-Newton algorithms make use of the Wolfe line search to calculate the stepsize α_k, which ensures a positive curvature to be found at each iteration; namely, $\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k > 0$. More exactly, defining the one-dimensional line search function

$$\Psi_k(\alpha) = f(x_k - \alpha H_k \mathbf{g}_k), \quad \alpha \geq 0, \tag{1.6}$$

the Wolfe line search consists in finding a stepsize α_k such that

$$\Psi_k(\alpha_k) \leq \Psi_k(0) + \mu\, \Psi_k'(0)\, \alpha_k \tag{1.7}$$

and

$$\Psi_k'(\alpha_k) \geq \eta \, \Psi_k'(0), \tag{1.8}$$

where μ and η are constants with $0 < \mu \leq \eta < 1$.

There have been a large quantity of research works devoting to the global convergence of the quasi-Newton method (see Yuan [21] and Sun and Yuan [20], for example). Specifically, Powell [14] showed that the DFP method with exact line searches is globally convergent for uniformly convex functions. In another paper [15], Powell established the global convergence of the BFGS method with the Wolfe line search for uniformly convex functions. This result was extended by Byrd, Nocedal and Yuan [1] to the whole Broyden's convex family of methods except the DFP method. Therefore it is natural to ask the following question: *Does the DFP method with the Wolfe line search converge globally for uniformly convex functions?* On the other hand, since the BFGS method with the Wolfe line search works well for both convex and non-convex functions, we might ask another question: *Does the BFGS method with the Wolfe line search converge globally for general functions?* The difficulty and importance of the two convergence problems has been addressed in many situations, including Nocedal [13], Fletcher [7], and Yuan [21].

Recent studies provide a negative answer to the convergence problem of the BFGS method for nonconvex functions. As a matter of fact, in an early paper [16], which analyzes the convergence properties of the conjugate gradient method, Powell mentioned that the BFGS method need not converge if the line search can pick any local minimizer of the line search function $\Psi_k(\alpha)$. After further studies on the two-dimensional example in [16], Dai [3] presented an example with six cycling points and showed by the example that the BFGS method with the Wolfe line search may fail for nonconvex functions. Later, Mascarenhas [12] constructed a three-dimensional counter-example such that the BFGS method does not converge if the line search picks the global minimizer of the function $\Psi_k(\alpha)$. It should be noted that the stepsize in the counter-example of [12] also satisfies the Wolfe line search conditions. However, neither examples are such that the stepsize is the first local minimizer of the line search function $\Psi_k(\alpha)$.

Surprisingly enough, if there are only two variables, and if the stepsize is chosen to be the first local minimizer of $\Psi_k(\alpha)$; namely,

$$\alpha_k = \arg\min \{\alpha > 0 : \ \alpha \text{ is a local minimizer of } \Psi_k(\alpha)\}, \tag{1.9}$$

Powell [17] established the global convergence of the BFGS method for general twice continuously differentiable function. Powell's proof makes use of the principle of contradiction and is quite sophisticated. Assuming the nonconvergence of the method and the relation $\liminf_{k\to\infty} \|\mathbf{g}_k\| \neq 0$, Powell showed that the limit points of the BFGS path

$$\mathcal{P} = \{\mathbf{x} : \ \mathbf{x} \text{ lies on the line segment connecting } \mathbf{x}_k \text{ and } \mathbf{x}_{k+1} \text{ for some } k \geq 1\}$$

(that is exactly the same as the Polak-Ribière conjugate gradient path if $n = 2$ and $\mathbf{g}_{k+1}^T \boldsymbol{\delta}_k = 0$ for all k) forms some line segment $\mathcal{L}$. The use of the specific line search (1.9) is such that the objective function is monotonically decreasing on the line segment $\mathcal{L}_k$ that connects $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$. It follows that $f(\mathbf{x}) \equiv \lim_{k\to\infty} f(\mathbf{x}_k)$ for all $\mathbf{x} \in \mathcal{L}$. Consequently, the line segment $\mathcal{L}_k$ cannot cross $\mathcal{L}$ and all the points $\mathbf{x}_k$ with large indices lies in the same side of $\mathcal{L}$. Furthermore, Powell showed that starting around from one end of the line segment $\mathcal{L}$, the BFGS path can not get any close to the other end of $\mathcal{L}$ and finally obtained a contradiction. Intuitively, it is difficult to extend Powell's result to the case when $n \geq 3$. This is because, even if it has been shown that the BFGS path tends to some limit line segment $\mathcal{L}$, each line segment $\mathcal{L}_k$ could turn around $\mathcal{L}$ providing that $n \geq 3$, making the similar analysis of the BFGS path complicated and nearly impossible.

The purpose of this paper is to construct a four-dimensional counter-example for the BFGS method with the following features:

(a) *All the stepsizes in the example are exactly equal to one; namely,*

$$\alpha_k \equiv 1; \tag{1.10}$$

For the search direction defined by the BFGS update in the example, a unit stepsize satisfies various line search conditions, including the Wolfe conditions and the Armijo conditions;

(b) *The objective function is strongly convex along each search direction; namely, the line search function $\Psi_k(\alpha)$ is strongly convex for all k, although the objective function is not in itself. The unit stepsize is the unique minimizer of $\Psi_k(\alpha)$. Hence the example also applies to the global line search and the specific line search* (1.9)*;*

(c) *The objective function is polynomial and hence is infinitely continuously differentiable.*

On the other hand, the objective function in our example is linear in the third and fourth variables and thus has no local minimizer. However, the iterations generated by the BFGS method tend to a non-stationary point.

As seen from Sect. 3.4, the construction of the example is quite complicated. To be such that the line search function $\Psi_k(\alpha)$ is strongly convex, a polynomial of degree 38 is introduced. Nevertheless, if we relax the convexity requirement of $\Psi_k(\alpha)$; namely, (b) in the above, it is possible to construct a relatively simple polynomial example of low degree (see Sect. 3.3).

The construction of our examples can be divided into four procedures: (1) Prefix the special forms of the steps $\{\boldsymbol{\delta}_k\}$ and gradients $\{\mathbf{g}_k\}$, leaving several parameters to be determined later. In this case, once $\mathbf{x}_1$ is given, the whole sequence $\{\mathbf{x}_k\}$ is fixed. As seen in Sect. 2.1, our steps $\{\boldsymbol{\delta}_k\}$ and gradients $\{\mathbf{g}_k\}$ are asked to possess some symmetrical and cyclic properties and to push the iterations $\{\mathbf{x}_k\}$ tend to the eight vertices of a regular octagon. This is very helpful in simplifying the construction of the examples and we can focus our attention on the choice of the parameter t, that answers for the decay of the last two components of $\boldsymbol{\delta}_k$ and the first two components of $\mathbf{g}_k$. (2) To enable the BFGS method to generate those prefixed steps, investigate the consistency conditions on the steps $\{\boldsymbol{\delta}_k\}$ and gradients $\{\mathbf{g}_k\}$. With the prefixed forms of $\{\boldsymbol{\delta}_k\}$ and $\{\mathbf{g}_k\}$, we show in Sect. 2.2 that the unit stepsize is indispensable, whereas this stepsize is usually used as the first trial stepsize in the implementation of quasi-Newton methods. Four more consistency conditions on $\{\boldsymbol{\delta}_k\}$ and $\{\mathbf{g}_k\}$ are obtained in Sect. 2.3 by expressing the vectors $H_{k+1}\boldsymbol{\gamma}_k$, $H_{k+1}\mathbf{g}_{k+1}$, $H_{k+1}\mathbf{g}_{k+2}$ and $H_{k+1}\mathbf{g}_{k+3}$ by some $\boldsymbol{\delta}_k$'s and $\mathbf{g}_k$'s instead of considering the quasi-Newton matrix H_{k+1} itself. (3) Choose the parameters in some way to satisfy the consistency conditions and other necessary conditions on the objective function. As seen in Sect. 2.4, the first three consistency conditions are actually corresponding to some under-determined linear system. Substituting its general solution by the Cramer's rule into the fourth consistency condition, we are led to a nonlinear equation from which the exact value can be obtained for the decay parameter t. By suitably choosing the other parameters, we can express all the quasi-Newton updating matrices including the initial choice for H_1 in Sect. 2.5. (4) Construct a suitable objective function f whose gradients are the preassigned values; namely, $\nabla f(\mathbf{x}_k) = \mathbf{g}_k$ for all $k \geq 1$ and the line search has the desired properties. This will be done in the whole third section. To make full use of symmetrical and cyclic properties of the steps $\{\boldsymbol{\delta}_k\}$ and $\{\mathbf{g}_k\}$, we carefully choose a special form of the objective function. In addition, the introduction of element function ϕ in Sect. 3.4 helps us greatly to convexify the line search function and finally complete the perfect example that meets all the requirements (a), (b) and (c). Some concluding remarks are given in the last section.

2 Looking for consistent steps and gradients

2.1 The forms of steps and gradients

Consider the case of four dimension. Inspired by [3,16] and [12], we assume that the steps $\{\boldsymbol{\delta}_k\}$ have the following form:

$$\boldsymbol{\delta}_1 = (\eta_1,\ \xi_1,\ \gamma_1,\ \tau_1)^T; \quad \boldsymbol{\delta}_{k+1} = M\,\boldsymbol{\delta}_k\ (k \geq 1), \tag{2.1}$$

where M is the 4×4 matrix defined by

$$M = \begin{bmatrix} \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} & 0 \\ 0 & t\begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \end{bmatrix}. \tag{2.2}$$

In the above, t is a parameter satisfying $0 < |t| < 1$. This parameter answers for the decay of the last two components of the steps $\{\boldsymbol{\delta}_k\}$. The angles θ_1 and θ_2 are chosen so that $\theta_1 = \frac{2\pi}{m_1}$ and $\theta_2 = \frac{2\pi}{m_2}$ for some positive integers m_1 and m_2. Specifically, we choose in this paper

$$\theta_1 = \frac{1}{4}\pi, \qquad \theta_2 = \frac{3}{4}\pi. \tag{2.3}$$

Consequently, the first two components of the steps $\{\boldsymbol{\delta}_k\}$ turn to the same after every eight iterations, and the last two components will shrink at a factor of t^8 after every eight iterations. Further, we can see that the iterations $\{\mathbf{x}_k\}$ will tend to turn around the eight vertices of some regular octagon (see Sect. 3.1 for more details).

Accordingly, the gradients $\{\mathbf{g}_k\}$ are assumed to be of the form

$$\mathbf{g}_1 = (l_1,\ h_1,\ c_1,\ d_1)^T; \quad \mathbf{g}_{k+1} = P\,\mathbf{g}_k\ (k \geq 1), \tag{2.4}$$

where

$$P = \begin{bmatrix} t\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \end{bmatrix}. \tag{2.5}$$

Since $0 < |t| < 1$, we see that the first two components of $\{\mathbf{g}_k\}$ vanish, whereas the last two components of the gradient $\{\mathbf{g}_k\}$, turn to the same after eight iterations. It follows that none of the cluster points generated by the BFGS method are stationary points.

2.2 Unit stepsizes

It is well known that the unit stepsize is usually used as the first trial stepsize in practical implementations of the BFGS method. Under suitable assumptions on f, the unit stepsize will be accepted by the line search as the iterates tend to the solution and will enable a superlinear convergence step (see [5], for example). In the following, we are going to show that if the line search satisfies

$$\mathbf{g}_{k+1}^T\boldsymbol{\delta}_k = 0, \quad \text{for all } k \geq 1, \tag{2.6}$$

and if the steps generated by the BFGS method have the form (2.1)–(2.2) and the gradients have the form (2.4)–(2.5), then the use of unit stepsizes is also indispensable.

To begin with, we see that the line search condition (2.6), the secant equation (1.3) and the definition of the search direction

$$H_{k+1}\mathbf{g}_{k+1} = -\alpha_{k+1}^{-1}\boldsymbol{\delta}_{k+1} \tag{2.7}$$

indicate that

$$\boldsymbol{\delta}_{k+1}^T\boldsymbol{\gamma}_k = -\alpha_{k+1}\mathbf{g}_{k+1}^T H_{k+1}\boldsymbol{\gamma}_k = -\alpha_{k+1}\mathbf{g}_{k+1}^T\boldsymbol{\delta}_k = 0. \tag{2.8}$$

The above relation (2.8) is sometimes called as the *conjugacy condition* in the context of nonlinear conjugate gradient methods.

By multiplying the BFGS updating formula (1.5) with $\mathbf{g}_{k+1}$ and using (2.6),

$$H_{k+1}\mathbf{g}_{k+1} = H_k\mathbf{g}_{k+1} - \frac{\boldsymbol{\gamma}_k^T H_k\mathbf{g}_{k+1}}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k}\boldsymbol{\delta}_k,$$

which with (2.7) gives

$$H_k \mathbf{g}_{k+1} = -\alpha_{k+1}^{-1} \boldsymbol{\delta}_{k+1} + \frac{\boldsymbol{\gamma}_k^T H_k \mathbf{g}_{k+1}}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \boldsymbol{\delta}_k. \tag{2.9}$$

Multiplying (2.9) by $\mathbf{g}_k^T$ and noticing $\mathbf{g}_k^T H_k \mathbf{g}_{k+1} = 0$, we get that

$$0 = -\alpha_{k+1}^{-1} \mathbf{g}_k^T \boldsymbol{\delta}_{k+1} + \frac{\boldsymbol{\gamma}_k^T H_k \mathbf{g}_{k+1}}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \mathbf{g}_k^T \boldsymbol{\delta}_k = -\alpha_{k+1}^{-1} \mathbf{g}_k^T \boldsymbol{\delta}_{k+1} - \boldsymbol{\gamma}_k^T H_k \mathbf{g}_{k+1}.$$

Thus $\boldsymbol{\gamma}_k^T H_k \mathbf{g}_{k+1} = -\alpha_{k+1}^{-1} \mathbf{g}_k^T \boldsymbol{\delta}_{k+1} = -\alpha_{k+1}^{-1} \mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}$. Hence, by (2.9),

$$H_k \mathbf{g}_{k+1} = -\alpha_{k+1}^{-1} \boldsymbol{\delta}_{k+1} - \alpha_{k+1}^{-1} \frac{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \boldsymbol{\delta}_k. \tag{2.10}$$

It follows from (2.10) and (2.7) with k replaced by $k-1$ that

$$H_k \boldsymbol{\gamma}_k = -\alpha_{k+1}^{-1} \boldsymbol{\delta}_{k+1} + \left[\alpha_k^{-1} - \alpha_{k+1}^{-1} \frac{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \right] \boldsymbol{\delta}_k. \tag{2.11}$$

Further, substituting this into the BFGS updating formula (1.5) yields

$$H_{k+1} = H_k + \alpha_{k+1}^{-1} \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_{k+1}^T + \boldsymbol{\delta}_{k+1} \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} + \left[1 - \alpha_k^{-1} + \alpha_{k+1}^{-1} \frac{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \right] \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k}. \tag{2.12}$$

The above new updating formula requires the quantity $\boldsymbol{\delta}_{k+1}$, that depends on H_{k+1} itself, and hence has theoretical meanings only.

Lemma 2.1 *Assume that (2.6) holds. Then for all $k \geq 1$ and $i \geq 0$, the vector $H_k \mathbf{g}_{k+i} + \alpha_{k+i}^{-1} \boldsymbol{\delta}_{k+i}$ belongs to the subspace spanned by $\boldsymbol{\delta}_k, \boldsymbol{\delta}_{k+1}, \ldots, \boldsymbol{\delta}_{k+i-1}$; namely,*

$$H_k \mathbf{g}_{k+i} + \alpha_{k+i}^{-1} \boldsymbol{\delta}_{k+i} \in Span\{\boldsymbol{\delta}_k, \boldsymbol{\delta}_{k+1}, \ldots, \boldsymbol{\delta}_{k+i-1}\}. \tag{2.13}$$

Proof For convenience, we write (2.12) as

$$H_{k+1} = H_k + V(\mathbf{s}_k, \mathbf{s}_{k+1}), \tag{2.14}$$

where $V(\mathbf{s}_k, \mathbf{s}_{k+1})$ means the rank-two matrix in the right hand of (2.12). Therefore we have for all $i \geq 1$,

$$H_{k+i} = H_k + \sum_{j=1}^{i} V(\mathbf{s}_{k+j-1}, \mathbf{s}_{k+j}). \tag{2.15}$$

The statement follows by multiplying the above relation with $\mathbf{g}_{k+i}$ and using (2.7) with k replaced by $k+i-1$ and $\mathbf{g}_{k+i}^T \boldsymbol{\delta}_{k+i-1} = 0$. □

Lemma 2.2 *To construct a desired example with Det* $(S_1) \neq 0$, *we must have that*

$$\alpha_k = 1, \qquad \textit{for all } k \geq 4.$$

Proof Define the following matrices

$$G_k = \left[\boldsymbol{\gamma}_{k-1}\ \mathbf{g}_k\ \mathbf{g}_{k+1}\ \mathbf{g}_{k+2}\right],$$
$$S_k = \left[\boldsymbol{\delta}_{k-1}\ \ \boldsymbol{\delta}_k\ \ \boldsymbol{\delta}_{k+1}\ \ \boldsymbol{\delta}_{k+2}\right].$$

By $H_k \boldsymbol{\gamma}_{k-1} = \boldsymbol{\delta}_{k-1}$ and Lemma 2.1, we can get that

$$\begin{aligned}\operatorname{Det}(H_k)\operatorname{Det}(G_k) &= \operatorname{Det}\left[H_k\boldsymbol{\gamma}_{k-1}\ H_k\mathbf{g}_k\ H_k\mathbf{g}_{k+1}\ H_k\mathbf{g}_{k+2}\right] \\ &= \operatorname{Det}\left[\boldsymbol{\delta}_{k-1}\ \ -\alpha_k^{-1}\boldsymbol{\delta}_k\ \ -\alpha_{k+1}^{-1}\boldsymbol{\delta}_{k+1}\ \ -\alpha_{k+2}^{-2}\boldsymbol{\delta}_{k+2}\right] \\ &= -\alpha_k^{-1}\alpha_{k+1}^{-1}\alpha_{k+2}^{-1}\operatorname{Det}(S_k).\end{aligned} \tag{2.16}$$

Replacing k with $k+1$ in the above yields

$$\operatorname{Det}(H_{k+1})\operatorname{Det}(G_{k+1}) = -\alpha_{k+1}^{-1}\alpha_{k+2}^{-1}\alpha_{k+3}^{-1}\operatorname{Det}(S_{k+1}). \tag{2.17}$$

On the other hand, due to the special forms of $\{\mathbf{g}_k\}$ and $\{\boldsymbol{\delta}_k\}$, we know that $G_{k+1} = PG_k$ and $S_{k+1} = MS_k$. Hence

$$\begin{aligned}\operatorname{Det}(G_{k+1}) &= \operatorname{Det}(P)\operatorname{Det}(G_k) = t^2\operatorname{Det}(G_k), \\ \operatorname{Det}(S_{k+1}) &= \operatorname{Det}(M)\operatorname{Det}(S_k) = t^2\operatorname{Det}(S_k).\end{aligned} \tag{2.18}$$

Due to the basic determinant relation of the BFGS update, (2.6) and (2.7), it is not difficult to see that

$$\operatorname{Det}(H_{k+1}) = \frac{\boldsymbol{\delta}_k^T H_k^{-1}\boldsymbol{\delta}_k}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k}\operatorname{Det}(H_k) = \alpha_k\operatorname{Det}(H_k). \tag{2.19}$$

If $\operatorname{Det}(S_1) \neq 0$, (2.16) with $k=1$ implies that $\operatorname{Det}(G_1) \neq 0$. Then by (2.18),

$$\operatorname{Det}(G_k) \neq 0, \quad \operatorname{Det}(S_k) \neq 0, \quad \text{for all } k \geq 1. \tag{2.20}$$

Dividing (2.17) by (2.16) and using the above relations, we then obtain

$$\alpha_{k+3} = 1.$$

So the statement holds due to the arbitrariness of $k \geq 1$. □

The deletion of the first finite iterations does not influence the whole example. Thus we will ask our counter-example to satisfy

$$\mathrm{Det}(S_1) \neq 0 \tag{2.21}$$

and

$$\alpha_k \equiv 1. \tag{2.22}$$

In this case, the updating formula (2.12) of H_k can be simplified as

$$H_{k+1} = H_k + \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_{k+1}^T + \boldsymbol{\delta}_{k+1} \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} - \frac{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}}{\mathbf{g}_k^T \boldsymbol{\delta}_k} \frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k}. \tag{2.23}$$

2.3 Consistency conditions

Now we ask what else conditions, besides the requirement of unit stepsizes, have to be satisfied by the parameters in the definitions of $\{\mathbf{g}_k\}$ and $\{\boldsymbol{\delta}_k\}$ so that the steps $\{\mathbf{s}_k;\ k \geq 1\}$ can be generated by the BFGS update. Our early idea is based on the observation on the updating formula (1.5) that H_{k+1} is linear with H_k and the linear system

$$H_{8j+9} = \mathrm{diag}(t^{-4} E_2, t^4 E_2)\, H_{8j+1}\, \mathrm{diag}(t^{-4} E_2, t^4 E_2), \tag{2.24}$$

where E_2 is the two-dimensional identity matrix. However, this way seems to be quite complicated.

Notice that the dimension is $n = 4$ and by Lemma 2.2, the assumption (2.21) implies (2.20). Then the matrix H_{k+1} can be uniquely defined by the equations given by $H_{k+1}\boldsymbol{\gamma}_k$, $H_{k+1}\mathbf{g}_{k+1}$, $H_{k+1}\mathbf{g}_{k+2}$ and $H_{k+1}\mathbf{g}_{k+3}$. As a matter of fact, we have that

$$H_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k \quad \text{(the secant equation)}, \tag{2.25}$$

$$H_{k+1}\mathbf{g}_{k+1} = -\boldsymbol{\delta}_{k+1} \quad \text{(by (2.7) and } \alpha_k \equiv 1), \tag{2.26}$$

$$H_{k+1}\mathbf{g}_{k+2} = -\boldsymbol{\delta}_{k+2} + \frac{\mathbf{g}_{k+2}^T \boldsymbol{\delta}_{k+2}}{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}} \boldsymbol{\delta}_{k+1} \quad \text{(by (2.10) and } \alpha_k \equiv 1), \tag{2.27}$$

$$\begin{aligned} H_{k+1}\mathbf{g}_{k+3} = {} & -\boldsymbol{\delta}_{k+3} + \left(\frac{\mathbf{g}_{k+3}^T \boldsymbol{\delta}_{k+3}}{\mathbf{g}_{k+2}^T \boldsymbol{\delta}_{k+2}} + \frac{\mathbf{g}_{k+3}^T \boldsymbol{\delta}_{k+1}}{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}} \right) \boldsymbol{\delta}_{k+2} \\ & - \left(\frac{\mathbf{g}_{k+2}^T \boldsymbol{\delta}_{k+2}}{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}} \right) \left(\frac{\mathbf{g}_{k+3}^T \boldsymbol{\delta}_{k+1}}{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}} \right) \boldsymbol{\delta}_{k+1}. \end{aligned} \tag{2.28}$$

The last equality is obtained by multiplying (2.23) by $\mathbf{g}_{k+2}$, using (2.27) and finally replacing k with $k+1$.

Relations (2.25)–(2.28) provide a system of 16 equations, while the symmetric matrix H_{k+1} only has 10 independent entries. How to ensure that this linear system has a symmetric solution H_{k+1}? We have the following general lemma.

Lemma 2.3 *Assume that* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ *and* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *are two sets of* n*-dimensional linearly independent vectors. Then there exists a symmetric matrix* $H \in \mathcal{R}^{n\times n}$ *satisfying*

$$H\mathbf{u}_i = \mathbf{v}_i, \quad i = 1, 2, \ldots, n \tag{2.29}$$

if and only if

$$\mathbf{u}_i^T \mathbf{v}_j = \mathbf{u}_j^T \mathbf{v}_i, \quad \forall i, j = 1, 2, \ldots, n. \tag{2.30}$$

Proof The "only if" part. If $H = H^T$ satisfies (2.29), we have for all $i, j = 1, 2, \ldots, n$, $\mathbf{u}_i^T \mathbf{v}_j = \mathbf{u}_i^T H \mathbf{u}_j = \mathbf{u}_i^T H^T \mathbf{u}_j = (H\mathbf{u}_i)^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{u}_j$.

The "if" part. Assume that (2.30) holds. Defining the matrices

$$U = (\mathbf{u}_1,\ \mathbf{u}_2\ \ldots,\ \mathbf{u}_n), \qquad V = (\mathbf{v}_1,\ \mathbf{v}_2\ \ldots,\ \mathbf{v}_n), \tag{2.31}$$

direct calculations show that

$$U^T H U = U^T V = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v}_1 & \mathbf{u}_1^T\mathbf{v}_2 & \cdots & \mathbf{u}_1^T\mathbf{v}_n \\ \mathbf{u}_2^T\mathbf{v}_1 & \mathbf{u}_2^T\mathbf{v}_2 & \cdots & \mathbf{u}_2^T\mathbf{v}_n \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{u}_n^T\mathbf{v}_1 & \mathbf{u}_n^T\mathbf{v}_2 & \cdots & \mathbf{u}_n^T\mathbf{v}_n \end{pmatrix} := A. \tag{2.32}$$

By (2.30), A is symmetric. So $H = U^{-T} A U^{-1}$ satisfies $H = H^T$ and (2.29). This completes our proof. □

By the above lemma, the following six conditions are sufficient for the linear system (2.25)–(2.28) to allow a symmetric solution matrix H_{k+1}.

$$\begin{aligned}
\mathbf{g}_{k+1}^T(H_{k+1}\boldsymbol{\gamma}_k) &= (H_{k+1}\mathbf{g}_{k+1})^T\boldsymbol{\gamma}_k,\\
\mathbf{g}_{k+2}^T(H_{k+1}\boldsymbol{\gamma}_k) &= (H_{k+1}\mathbf{g}_{k+2})^T\boldsymbol{\gamma}_k,\\
\mathbf{g}_{k+3}^T(H_{k+1}\boldsymbol{\gamma}_k) &= (H_{k+1}\mathbf{g}_{k+3})^T\boldsymbol{\gamma}_k,\\
\mathbf{g}_{k+2}^T(H_{k+1}\mathbf{g}_{k+1}) &= (H_{k+1}\mathbf{g}_{k+2})^T\mathbf{g}_{k+1},\\
\mathbf{g}_{k+3}^T(H_{k+1}\mathbf{g}_{k+1}) &= (H_{k+1}\mathbf{g}_{k+3})^T\mathbf{g}_{k+1},\\
\mathbf{g}_{k+3}^T(H_{k+1}\mathbf{g}_{k+2}) &= (H_{k+1}\mathbf{g}_{k+3})^T\mathbf{g}_{k+2}.
\end{aligned}$$

Further, considering the whole sequence of $\{H_{k+1}; k \geq 1\}$ and combining the line search condition $\mathbf{g}_{k+1}^T\boldsymbol{\delta}_k = 0$, we can deduce the following four consistency conditions, which should be satisfied for all $k \geq 1$,

$$\mathbf{g}_{k+1}^T \boldsymbol{\delta}_k = 0, \tag{2.33}$$
$$\boldsymbol{\delta}_{k+1}^T \boldsymbol{\gamma}_k = 0, \tag{2.34}$$
$$\mathbf{g}_{k+2}^T \boldsymbol{\delta}_k = -\boldsymbol{\delta}_{k+2}^T \boldsymbol{\gamma}_k, \tag{2.35}$$
$$\mathbf{g}_{k+3}^T \boldsymbol{\delta}_k = -\boldsymbol{\delta}_{k+3}^T \boldsymbol{\gamma}_k + \left(\frac{\mathbf{g}_{k+3}^T \boldsymbol{\delta}_{k+3}}{\mathbf{g}_{k+2}^T \boldsymbol{\delta}_{k+2}} + \frac{\mathbf{g}_{k+3}^T \boldsymbol{\delta}_{k+1}}{\mathbf{g}_{k+1}^T \boldsymbol{\delta}_{k+1}} \right) \boldsymbol{\delta}_{k+2}^T \boldsymbol{\gamma}_k. \tag{2.36}$$

The positive definiteness of the matrix H_{k+1} will be further considered in Sect. 2.5.

2.4 Choosing the parameters

In this subsection we focus on how to choose the decay parameter t and suitable vectors $\boldsymbol{\delta}_1$ and $\mathbf{g}_1$ such that the consistency conditions (2.33), (2.34), (2.35) and (2.36) hold with $k = 1$. Due to the special structure of this example, we then know that the consistency conditions hold for all $k \geq 2$.

Define $\mathbf{v} = (\Delta_1 \;\; \Delta_2 \;\; \Delta_3 \;\; \Delta_4)^T$, where

$$\Delta_1 = l_1\eta_1 + h_1\xi_1, \quad \Delta_2 = h_1\eta_1 - l_1\xi_1, \quad \Delta_3 = c_1\gamma_1 + d_1\tau_1, \quad \Delta_4 = d_1\gamma_1 - c_1\tau_1. \tag{2.37}$$

As will be seen, the first three consistent conditions provide an under-determined linear system with $\mathbf{v}$, from which we can get a general solution by the Cramer's rule. The substitution of $\mathbf{v}$ into the fourth condition (2.36), which is nonlinear, yields a desired value for the decay parameter t.

At first, the condition $\boldsymbol{\delta}_1^T \mathbf{g}_2 = \boldsymbol{\delta}_1^T P \mathbf{g}_1 = 0$, that is (2.33) with $k = 1$, asks

$$[t\cos\theta_1 \quad -t\sin\theta_1 \quad \cos\theta_2 \quad -\sin\theta_2]\,\mathbf{v} = 0. \tag{2.38}$$

The condition $\boldsymbol{\delta}_2^T \boldsymbol{\gamma}_1 = \boldsymbol{\delta}_1^T M^T (P - I)\mathbf{g}_1 = \boldsymbol{\delta}_1^T (tI - M^T)\mathbf{g}_1 = 0$, that is (2.34) with $k = 1$, requires the vector $\mathbf{v}$ to satisfy

$$[t - \cos\theta_1 \quad -\sin\theta_1 \quad t(1 - \cos\theta_2) \quad -t\sin\theta_2]\,\mathbf{v} = 0. \tag{2.39}$$

The requirement (2.35) with $k = 1$, that is,

$$\begin{aligned} 0 \;=\; \boldsymbol{\delta}_1^T \mathbf{g}_3 + \boldsymbol{\delta}_3^T \boldsymbol{\gamma}_1 &= \boldsymbol{\delta}_1^T P^2 \mathbf{g}_1 + \boldsymbol{\delta}_1^T (M^2)^T (P - I)\mathbf{g}_1 \\ &= \boldsymbol{\delta}_1^T [P^2 + tM^T - (M^2)^T]\mathbf{g}_1, \end{aligned}$$

yields the equation

$$\Big[t\cos\theta_1 + (t^2 - 1)\cos 2\theta_1 \quad t\sin\theta_1 - (1 + t^2)\sin 2\theta_1 \quad t^2(\cos\theta_2 - \cos 2\theta_2) + \cos 2\theta_2 \quad t^2(\sin\theta_2 - \sin 2\theta_2) - \sin 2\theta_2 \Big]\,\mathbf{v} = 0. \tag{2.40}$$

By (2.38), (2.39), (2.40) and the choices of θ_1 and θ_2, we see that $\{\Delta_i\}$ must satisfy

$$\begin{bmatrix} \frac{\sqrt{2}}{2}t & -\frac{\sqrt{2}}{2}t & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ t-\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & (1+\frac{\sqrt{2}}{2})t & -\frac{\sqrt{2}}{2}t \\ \frac{\sqrt{2}}{2}t & \frac{\sqrt{2}}{2}t-(1+t^2) & -\frac{\sqrt{2}}{2}t^2 & (\frac{\sqrt{2}}{2}+1)t^2+1 \end{bmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \end{pmatrix} = 0. \tag{2.41}$$

By the Cramer's rule, to meet (2.41), we may choose $\{\Delta_i\}$ as follows:

$$\begin{cases} \Delta_1 = \pm\left[(2+\sqrt{2})t^3 + (1+\sqrt{2})t + 1\right], \\ \Delta_2 = \pm\left[(2+\sqrt{2})t^3 + (1+\sqrt{2})t - 1\right], \\ \Delta_3 = \pm\left[-\sqrt{2}\,t^3 + 2t^2 + (1-\sqrt{2})t + 1\right], \\ \Delta_4 = \pm\left[\sqrt{2}\,t^3 - 2t^2 + (1+\sqrt{2})t - 1\right]. \end{cases} \tag{2.42}$$

Since $\mathbf{g}_1^T\boldsymbol{\delta}_1 = \Delta_1 + \Delta_3 = \pm 2(t+1)(t^2+1)$ and $0 < |t| < 1$, we choose all the signs in (2.42) as $-$ so that $\mathbf{g}_1^T\boldsymbol{\delta}_1 < 0$.

We now want to substitute the general solution to the fourth consistent condition (2.36). Noting that $\boldsymbol{\delta}_3^T\boldsymbol{\gamma}_1 = -\mathbf{g}_3^T\boldsymbol{\delta}_1$ and $\mathbf{g}_4^T\boldsymbol{\delta}_4 = t\,\mathbf{g}_3^T\boldsymbol{\delta}_3$, we know that (2.36) with $k = 1$ is equivalent to

$$\mathbf{g}_4^T\boldsymbol{\delta}_1 + \mathbf{g}_2^T\boldsymbol{\delta}_4 - \mathbf{g}_1^T\boldsymbol{\delta}_4 + \mathbf{g}_3^T\boldsymbol{\delta}_1\left[t + \frac{\mathbf{g}_4^T\boldsymbol{\delta}_2}{\mathbf{g}_2^T\boldsymbol{\delta}_2}\right] = 0. \tag{2.43}$$

Further, using $\mathbf{g}_2^T\boldsymbol{\delta}_2 = t\,\mathbf{g}_1^T\boldsymbol{\delta}_1$, $\mathbf{g}_4^T\boldsymbol{\delta}_2 = t\,\mathbf{g}_3^T\boldsymbol{\delta}_1$ and $\mathbf{g}_2^T\boldsymbol{\delta}_4 = t\,\mathbf{g}_1^T\boldsymbol{\delta}_3$, (2.43) can be simplified as

$$\mathbf{g}_1^T\boldsymbol{\delta}_1\left[\mathbf{g}_4^T\boldsymbol{\delta}_1 - \mathbf{g}_1^T\boldsymbol{\delta}_4 + t(\mathbf{g}_3^T\boldsymbol{\delta}_1 + \mathbf{g}_1^T\boldsymbol{\delta}_3)\right] + (\mathbf{g}_3^T\boldsymbol{\delta}_1)^2 = 0. \tag{2.44}$$

Consequently, we obtain the following equation for the parameter t:

$$(t+1)^2(t^2+1)^2\left[p^2(t) + 2tp(t) - 2q(t)\right] = 0, \tag{2.45}$$

where

$$p(t) = -(2+\sqrt{2})t^2 + (2+\sqrt{2})t - 1,$$
$$q(t) = (2+3\sqrt{2})t^3 - (4+3\sqrt{2})t^2 + (3+2\sqrt{2})t - 2\sqrt{2}.$$

Further calculations provide

$$\left(6+4\sqrt{2}\right)^{-1}\left[p^2(t) + 2tp(t) - 2q(t)\right]$$
$$= \left[t^2 + (1-\sqrt{2})t + \left(1 - \frac{\sqrt{2}}{2}\right)\right]\left[t^2 + (1-3\sqrt{2})t + \left(-3 + \frac{7}{2}\sqrt{2}\right)\right].$$

Considering the requirement that $t \in (-1,\ 1)$, one can deduce the following quadratic equation from (2.45),

$$t^2 + \left(1 - 3\sqrt{2}\right) t + \left(-3 + \frac{7}{2}\sqrt{2}\right) = 0. \tag{2.46}$$

The above equation has a unique root of t in the interval $(-1,\ 1)$,

$$t = \frac{3\sqrt{2} - 1 - \sqrt{31 - 20\sqrt{2}}}{2}. \tag{2.47}$$

The numerical value of t is equal to 0.7973 approximately.

Therefore if we choose the above t and the vectors $\boldsymbol{\delta}_1$ and $\mathbf{g}_1$ such that (2.42) holds with minus sign, then the prefixed steps and gradients satisfy the consistency conditions (2.33), (2.34), (2.35) and (2.36) for all $k \geq 1$.

There are many ways to choose $\boldsymbol{\delta}_1$ and $\mathbf{g}_1$ satisfying (2.42). Specifically, we choose

$$\begin{pmatrix} \eta_1 \\ \xi_1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} c_1 \\ d_1 \end{pmatrix} = \begin{pmatrix} 0 \\ -\sqrt{2} \end{pmatrix}. \tag{2.48}$$

In this case, by (2.42) with minus signs and the definitions of Δ_i's, we can get that

$$\begin{aligned} \begin{pmatrix} \gamma_1 \\ \tau_1 \end{pmatrix} &= \begin{pmatrix} (17 - 8\sqrt{2})t + (-17 + 9\sqrt{2}) \\ (-17 + 9\sqrt{2})t + (17 - 9\sqrt{2}) \end{pmatrix}, \\ \begin{pmatrix} l_1 \\ h_1 \end{pmatrix} &= \begin{pmatrix} (-4 - 13\sqrt{2})t + (-1 + 11\sqrt{2}) \\ (-4 - 13\sqrt{2})t + (-1 + 12\sqrt{2}) \end{pmatrix}. \end{aligned} \tag{2.49}$$

The initial step $\boldsymbol{\delta}_1$ and the initial gradient $\mathbf{g}_1$ are then determined.

2.5 Quasi-Newton updating matrices

In this subsection we discuss the form of the BFGS quasi-Newton updating matrices implied by (2.25)–(2.28) under the consistent conditions (2.33)–(2.36). A byproduct is that, we will know how to choose the initial matrix H_1 for the example.

For this aim, we provide a follow-up lemma of Lemma 2.3.

Lemma 2.4 *Assume that* (2.30) *holds and the matrix H satisfies* (2.29). *If further, the matrix A in* (2.32) *is positive definite, there must exist nonsingular triangular matrices T_1 and T_2 such that*

$$A = V^T U = T_1^{-T} T_2. \tag{2.50}$$

Further, denoting $\hat{V} = VT_1 = (\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_n)$ and the diagonal matrix $T_1^{-1}T_2^{-1} = diag(t_1, t_2, \ldots, t_n)$, we have that

$$H = \sum_{i=1}^{n} t_i\, \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^T. \tag{2.51}$$

Proof The truth of (2.50) is obvious by the Cholesky factorization of positive definite matrices. Further, by (2.50) and the definition of $\hat{V}$, we get that

$$\hat{V}^T U = T_2.$$

Since $HU = V$, we obtain

$$H = VU^{-1} = \left(\hat{V} T_1^{-1}\right) \left(T_2^{-1} \hat{V}^T\right) = \hat{V} \left(T_1^{-1} T_2^{-1}\right) \hat{V}^T,$$

which with the definitions of $\hat{\mathbf{v}}_i$'s and t_i's implies the truth of (2.51). □

By the above lemma, we can express the form of H_{k+1} determined by (2.25)–(2.26) under the consistent conditions (2.33)–(2.36). At first, we make use of the steps $\{\boldsymbol{\delta}_{k+i};\ i = 0, 1, 2, 3\}$ to introduce the following two vectors that are orthogonal to $\boldsymbol{\gamma}_k$ and $\mathbf{g}_{k+1}$:

$$\begin{aligned} \mathbf{z}_k &= -\frac{\boldsymbol{\delta}_{k+2}^T \boldsymbol{\gamma}_k}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \boldsymbol{\delta}_k - \frac{\boldsymbol{\delta}_{k+2}^T \mathbf{g}_{k+1}}{\boldsymbol{\delta}_{k+1}^T \mathbf{g}_{k+1}} \boldsymbol{\delta}_{k+1} + \boldsymbol{\delta}_{k+2}, \\ \mathbf{w}_k &= -\frac{\boldsymbol{\delta}_{k+3}^T \boldsymbol{\gamma}_k}{\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k} \boldsymbol{\delta}_k - \frac{\boldsymbol{\delta}_{k+3}^T \mathbf{g}_{k+1}}{\boldsymbol{\delta}_{k+1}^T \mathbf{g}_{k+1}} \boldsymbol{\delta}_{k+1} + \boldsymbol{\delta}_{k+3}. \end{aligned} \tag{2.52}$$

The vectors $\mathbf{z}_k$ and $\mathbf{w}_k$ are well defined because $\boldsymbol{\delta}_k^T \boldsymbol{\gamma}_k = -\boldsymbol{\delta}_k^T \mathbf{g}_k > 0$ is positive due to the descent property of $\boldsymbol{\delta}_k$. Direct calculations show that

$$\mathbf{z}_k^T \boldsymbol{\gamma}_k = \mathbf{z}_k^T \mathbf{g}_{k+1} = 0, \quad \mathbf{w}_k^T \boldsymbol{\gamma}_k = \mathbf{w}_k^T \mathbf{g}_{k+1} = 0. \tag{2.53}$$

Further, we use z_k and w_k to define the vector that is orthogonal to $\mathbf{g}_{k+2}$:

$$\mathbf{v}_k = -\frac{\mathbf{w}_k^T \mathbf{g}_{k+2}}{\mathbf{z}_k^T \mathbf{g}_{k+2}} \mathbf{z}_k + \mathbf{w}_k. \tag{2.54}$$

By the choice of $\mathbf{v}_k$, it is easy to see that

$$\mathbf{v}_k^T \boldsymbol{\gamma}_k = \mathbf{v}_k^T \mathbf{g}_{k+1} = \mathbf{v}_k^T \mathbf{g}_{k+2} = 0. \tag{2.55}$$

Now we could express the matrix H_{k+1} by

$$H_{k+1} = -\frac{\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T \mathbf{g}_k} - \frac{\boldsymbol{\delta}_{k+1} \boldsymbol{\delta}_{k+1}^T}{\boldsymbol{\delta}_{k+1}^T \mathbf{g}_{k+1}} - \frac{\mathbf{z}_k \mathbf{z}_k^T}{\mathbf{z}_k^T \mathbf{g}_{k+2}} - \frac{\mathbf{v}_k \mathbf{v}_k^T}{\mathbf{v}_k^T \mathbf{g}_{k+3}}. \tag{2.56}$$

Since $\boldsymbol{\delta}_k^T \mathbf{g}_k < 0$ for all $k \geq 1$, we know that the matrix H_{k+1} is positive definite if and only if $\mathbf{z}_k^T \mathbf{g}_{k+2} < 0$ and $\mathbf{v}_k^T \mathbf{g}_{k+3} < 0$. Direct calculations show that

$$\begin{aligned} \mathbf{z}_1^T \mathbf{g}_3 &= \left(5808 - 3348\sqrt{2}\right) t - \left(6912 - 4280\sqrt{2}\right) < 0, \\ \mathbf{v}_1^T \mathbf{g}_4 &= \left(-88803 + 63514\sqrt{2}\right) t + \left(109307 - 77868\sqrt{2}\right) < 0. \end{aligned}$$

Therefore we know that the matrix H_2 is positive definite. In addition, it is difficult to build the relations $\boldsymbol{\delta}_{k+1}^T \mathbf{g}_{k+1} = t\, \boldsymbol{\delta}_k^T \mathbf{g}_k$, $\mathbf{z}_{k+1}^T \mathbf{g}_{k+3} = t\, \mathbf{z}_k^T \mathbf{g}_{k+2}$, $\mathbf{v}_{k+1}^T \mathbf{g}_{k+4} = t\, \mathbf{v}_k^T \mathbf{g}_{k+3}$, $\mathbf{z}_{k+1} = M\mathbf{z}_k$ and $\mathbf{v}_{k+1} = M\, \mathbf{v}_k$ for all $k \geq 1$. Thus we have by (2.56) and $\boldsymbol{\delta}_{k+1} = M\, \boldsymbol{\delta}_k$ that

$$H_{k+1} = \frac{1}{t}\, M H_k M^T, \tag{2.57}$$

holds for all $k \geq 2$ and hence $\{H_k : k \geq 2\}$ are all positive definite.

With the above procedure, we can calculate the values of $\mathbf{z}_1$, $\mathbf{v}_1$ and H_2. Further, by (2.23), we can obtain the initial matrix H_1 required by the example of this paper.

$$H_1 = H_2 - \frac{\boldsymbol{\delta}_1 \boldsymbol{\delta}_2^T + \boldsymbol{\delta}_2 \boldsymbol{\delta}_1^T}{\boldsymbol{\delta}_1^T \boldsymbol{\gamma}_1} + \frac{\mathbf{g}_2^T \boldsymbol{\delta}_2}{\mathbf{g}_1^T \boldsymbol{\delta}_1} \frac{\boldsymbol{\delta}_1 \boldsymbol{\delta}_1^T}{\boldsymbol{\delta}_1^T \boldsymbol{\gamma}_1} := \frac{\bar{H}_1}{87278}, \tag{2.58}$$

where $\bar{H}_1$ is a symmetric matrix with entries

$$\begin{aligned}
\bar{H}_1(1,1) &= \left(-3690 - 13280\sqrt{2}\right) t + \left(79982 - 13694\sqrt{2}\right),\\
\bar{H}_1(1,2) &= \left(4474 + 1590\sqrt{2}\right) t - \left(1990 + 11308\sqrt{2}\right),\\
\bar{H}_1(1,3) &= \left(1428 + 18496\sqrt{2}\right) t - \left(33966 - 11118\sqrt{2}\right),\\
\bar{H}_1(1,4) &= \left(-23256 + 952\sqrt{2}\right) t + \left(25092 - 2108\sqrt{2}\right)\\
\bar{H}_1(2,2) &= \left(-1954 - 10928\sqrt{2}\right) t + \left(65266 - 15580\sqrt{2}\right),\\
\bar{H}_1(2,3) &= -10268\sqrt{2}\, t - \left(5134 - 10268\sqrt{2}\right),\\
\bar{H}_1(2,4) &= \left(13600 + 5508\sqrt{2}\right) t - \left(12512 + 9996\sqrt{2}\right),\\
\bar{H}_1(3,3) &= \left(78234 - 415769\sqrt{2}\right) t + \left(235654 + 163183\sqrt{2}\right),\\
\bar{H}_1(3,4) &= \left(-875432 + 576963\sqrt{2}\right) t + \left(943194 - 626093\sqrt{2}\right),\\
\bar{H}_1(4,4) &= \left(83606 - 164543\sqrt{2}\right) t + \left(104210 + 49521\sqrt{2}\right).
\end{aligned} \tag{2.59}$$

Direct calculations show that (2.57) also holds with $k = 1$ and hence H_1 is a positive definite matrix.

3 Constructing a suitable objective function

3.1 Recovering the iterations and the function values

Denote the rotation matrices

$$R_1 = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}, \qquad R_2 = \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix}. \tag{3.1}$$

If we write the steps $\{\boldsymbol{\delta}_k\}$ and gradients $\{\mathbf{g}_k\}$ into the forms

$$\boldsymbol{\delta}_{8j+i} = \begin{pmatrix} \eta_i \\ \xi_i \\ t^{8j}\gamma_i \\ t^{8j}\tau_i \end{pmatrix}, \quad \mathbf{g}_{8j+i} = \begin{pmatrix} t^{8j} l_i \\ t^{8j} h_i \\ c_i \\ d_i \end{pmatrix}; \quad i = 1, \ldots, 8, \tag{3.2}$$

we have from (2.1), (2.2), (2.4) and (2.5) that

$$\begin{aligned} \begin{pmatrix} \eta_{i+1} \\ \xi_{i+1} \end{pmatrix} &= R_1 \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, \quad \begin{pmatrix} \gamma_{i+1} \\ \tau_{i+1} \end{pmatrix} = t\, R_2 \begin{pmatrix} \gamma_i \\ \tau_i \end{pmatrix}, \\ \begin{pmatrix} l_{i+1} \\ h_{i+1} \end{pmatrix} &= t\, R_1 \begin{pmatrix} l_i \\ h_i \end{pmatrix}, \begin{pmatrix} c_{i+1} \\ d_{i+1} \end{pmatrix} = R_2 \begin{pmatrix} c_i \\ d_i \end{pmatrix}. \end{aligned} \tag{3.3}$$

For simplicity, we want the iterations $\{\mathbf{x}_k\}$ asymptotically to turn around the eight vertices of some regular octagon Ω of the subspace spanned by the first and second coordinates. More exactly, such a regular octagon Ω has the origin as its center and its eight vertices V_i are given by $V_i = (a_i,\ b_i,\ 0,\ 0)^T$ with

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ -1 - \frac{\sqrt{2}}{2} \end{pmatrix}, \qquad \begin{pmatrix} a_{i+1} \\ b_{i+1} \end{pmatrix} = R_1 \begin{pmatrix} a_i \\ b_i \end{pmatrix}. \tag{3.4}$$

Here we should note that if there is no confusion, we also regard that Ω and V_i's are defined in the subspace spanned by the first and second coordinates. To this aim, we ask the iterations $\{\mathbf{x}_k\}$ to be of the form

$$\mathbf{x}_{8j+i} = \begin{pmatrix} a_i \\ b_i \\ t^{8j} p_i \\ t^{8j} q_i \end{pmatrix}, \tag{3.5}$$

where

$$\begin{pmatrix} p_{i+1} \\ q_{i+1} \end{pmatrix} = t\, R_2 \begin{pmatrix} p_i \\ q_i \end{pmatrix}. \tag{3.6}$$

To decide the values of p_1 and q_1, noting that

$$\mathbf{x}_1 - V_1 = \mathbf{x}_1 - \lim_{j\to\infty} \mathbf{x}_{8j+1} = -\sum_{i=0}^{\infty} \boldsymbol{\delta}_i,$$

we have that

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = -\sum_{i=0}^{\infty} \begin{pmatrix} \gamma_i \\ \tau_i \end{pmatrix} = -(E_2 - tR_2)^{-1} \begin{pmatrix} \gamma_1 \\ \tau_1 \end{pmatrix}, \tag{3.7}$$

where again E_2 is the two-dimensional identity matrix. Direct calculations show that

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \frac{1}{4633} \begin{pmatrix} (-15720 + 4019\sqrt{2})t + (32931 - 17534\sqrt{2}) \\ (4376 - 1977\sqrt{2})t + (9563 - 9433\sqrt{2}) \end{pmatrix}. \tag{3.8}$$

Now, let us assume that the limit of $f(\mathbf{x}_k)$ is f^*. Since for any j, the first two coordinates of $\{\mathbf{x}_{8j+i}\}$ always keep the same as those of its limit V_i, we can get that

$$\begin{aligned} f(\mathbf{x}_{8j+i}) - f^* &= t^{8j} \begin{pmatrix} c_i \\ d_i \end{pmatrix}^T \begin{pmatrix} p_i \\ q_i \end{pmatrix} \\ &= t^{8j} \left[R_2^{i-1} \begin{pmatrix} c_1 \\ d_1 \end{pmatrix} \right]^T \left[(tR_2)^{i-1} \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \right] \\ &= t^{8j+i-1} \begin{pmatrix} c_1 \\ d_1 \end{pmatrix}^T \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \\ &= t^{8j+i-1}(f(\mathbf{x}_1) - f^*), \end{aligned}$$

where

$$f(\mathbf{x}_1) - f^* = c_1 p_1 + d_1 q_1 = \frac{(3954 - 4376\sqrt{2})\,t + (18866 - 9563\sqrt{2})}{4633}.$$

Now we see the value of $\mathbf{g}_{8j+i}^T \boldsymbol{\delta}_{8j+i}$. Direct calculations show that

$$\begin{aligned} \mathbf{g}_{8j+i}^T \boldsymbol{\delta}_{8j+i} &= \begin{pmatrix} t^{8j} l_i \\ t^{8j} h_i \\ c_i \\ d_i \end{pmatrix}^T \begin{pmatrix} \eta_i \\ \xi_i \\ t^{8j}\gamma_i \\ t^{8j}\tau_i \end{pmatrix} = t^{8j} \begin{pmatrix} l_i \\ h_i \\ c_i \\ d_i \end{pmatrix}^T \begin{pmatrix} \eta_i \\ \xi_i \\ \gamma_i \\ \tau_i \end{pmatrix} \\ &= t^{8j+i-1}(l_1\eta_1 + h_1\xi_1 + c_1\gamma_1 + d_1\tau_1) = t^{8j+i-1}\mathbf{g}_1^T\boldsymbol{\delta}_1, \end{aligned}$$

where

$$\mathbf{g}_1^T\boldsymbol{\delta}_1 = l_1\eta_1 + h_1\xi_1 + c_1\gamma_1 + d_1\tau_1 = (-44 + 13\sqrt{2})t + (40 - 18\sqrt{2}).$$

Therefore

$$\begin{aligned} \frac{f(\mathbf{x}_{8j+i+1}) - f(\mathbf{x}_{8j+i})}{\alpha_{8j+i}\mathbf{g}_{8j+i}^T\boldsymbol{\delta}_{8j+i}} &= \frac{(f(\mathbf{x}_{8j+i+1}) - f^*) - (f(\mathbf{x}_{8j+i}) - f^*)}{\mathbf{g}_{8j+i}^T\boldsymbol{\delta}_{8j+i}} \\ &= \frac{(t^{8j+i} - t^{8j+i-1})(f(\mathbf{x}_1) - f^*)}{t^{8j+i-1}\mathbf{g}_1^T\boldsymbol{\delta}_1} \\ &= \frac{(t-1)(f(\mathbf{x}_1) - f^*)}{\mathbf{g}_1^T\boldsymbol{\delta}_1} \\ &\approx 2.6483E-02. \end{aligned} \tag{3.9}$$

The above relation, together with $\mathbf{g}_{8j+i+1}^T \boldsymbol{\delta}_{8j+i} = 0$, implies that the stepsize $\alpha_{8j+i} = 1$ can be accepted by the Wolfe line search, the Armijo line search and the Goldstein line search with suitable line search parameters.

3.2 Seeking a suitable form for the objective function

We now consider how to construct a smooth function f such that its gradient at any point $\mathbf{x}_{8j+i}$ given in (3.5) are the one given in (3.2); namely,

$$\nabla f(\mathbf{x}_{8j+i}) = \mathbf{g}_{8j+i}, \quad \text{for all } j \geq 0 \text{ and } i = 1, \ldots, 8. \tag{3.10}$$

To this aim, we assume that f is of the form

$$f(x_1, x_2, x_3, x_4) = \lambda(x_1, x_2)\, x_3 + \mu(x_1, x_2)\, x_4, \tag{3.11}$$

where λ and μ are two-dimensional functions to be determined. Since the function (3.11) is linear with x_3 and x_4, we know that f has no lower bound in $\mathcal{R}^n$. Consequently, for the sequence $\{\mathbf{x}_k\}$ generated by some optimization method for the minimization of this function, it is expected that $f(\mathbf{x}_k)$ tends to $-\infty$, but this will not happen in our examples.

With the prefixed form (3.11), we have that

$$\nabla f(x_1, x_2, x_3, x_4) = \begin{pmatrix} \frac{\partial \lambda}{\partial x_1}\, x_3 + \frac{\partial \mu}{\partial x_1}\, x_4 \\ \frac{\partial \lambda}{\partial x_2}\, x_3 + \frac{\partial \mu}{\partial x_2}\, x_4 \\ \lambda(x_1, x_2) \\ \mu(x_1, x_2) \end{pmatrix}. \tag{3.12}$$

Comparing the last two components of the right hand vector in (3.12) with the gradients in (3.2), we must have that

$$\lambda(V_i) = c_i, \quad \mu(V_i) = d_i. \tag{3.13}$$

Consequently, by (2.48) and (3.3), we can obtain the concrete values of λ and μ at the eight vertices of Ω, which are listed in the second and third rows in Table 1.

Further, for each i, let J_i be the Jacobian of (λ, μ) at vertex V_i and denote

$$J_i = \left. \begin{pmatrix} \frac{\partial \lambda}{\partial x_1} & \frac{\partial \mu}{\partial x_1} \\ \frac{\partial \lambda}{\partial x_2} & \frac{\partial \mu}{\partial x_2} \end{pmatrix} \right|_{V_i}, \quad J_1 = \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_3 & \omega_4 \end{pmatrix}. \tag{3.14}$$

The comparison of the first two components of the right hand vector in (3.12) with the gradients in (3.2) leads to the relation

$$J_i \begin{pmatrix} p_i \\ q_i \end{pmatrix} = \begin{pmatrix} l_i \\ h_i \end{pmatrix}. \tag{3.15}$$

Table 1 Function values and gradients of λ and μ at V_i's

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
λ	0	1	$-\sqrt{2}$	1	0	-1	$-\sqrt{2}$	-1
μ	$-\sqrt{2}$	1	0	-1	$\sqrt{2}$	-1	0	1
$\frac{\partial\lambda}{\partial x_1}$	ω_1	$-\omega_1$	ω_1	-ω_1	ω_1	$-\omega_1$	ω_1	$-\omega_1$
$\frac{\partial\lambda}{\partial x_2}$	ω_3	$-\omega_3$	ω_3	$-\omega_3$	ω_3	$-\omega_3$	ω_3	$-\omega_3$
$\frac{\partial\mu}{\partial x_1}$	ω_3	$-\omega_3$	ω_3	$-\omega_3$	ω_3	$-\omega_3$	ω_3	$-\omega_3$
$\frac{\partial\mu}{\partial x_2}$	$-\omega_1$	ω_1	-ω_1	ω_1	$-\omega_1$	ω_1	$-\omega_1$	ω_1

To be such that (3.15) always holds, we ask J_{i+1} and J_i to meet the condition

$$J_{i+1} = R_1 J_i R_2^T \tag{3.16}$$

for all $i \geq 1$. In this case, if (3.15) holds for some i, we have by this, (3.6) and (3.3) that

$$J_{i+1}\begin{pmatrix} p_{i+1} \\ q_{i+1} \end{pmatrix} = (R_1 J_i R_2^T)(tR_2)\begin{pmatrix} p_i \\ q_i \end{pmatrix} = tR_1 J_i \begin{pmatrix} p_i \\ q_i \end{pmatrix} = tR_1 \begin{pmatrix} l_i \\ h_i \end{pmatrix} = \begin{pmatrix} l_{i+1} \\ h_{i+1} \end{pmatrix},$$

which means that (3.15) holds with $i+1$. Therefore by the induction principle, to be such that (3.15) holds for all $i \geq 1$, it remains to choose J_1 such that (3.15) holds with $i = 1$. Noticing that there are still two degrees of freedom, we ask the special relations

$$\omega_2 = \omega_3, \qquad \omega_4 = -\omega_1. \tag{3.17}$$

Then the values of ω_1 and ω_3 can be solved from (3.15) with $i = 1$ and (3.17),

$$\omega_1 = \frac{(163 + 106\sqrt{2})\,t - (195 + 129\sqrt{2})}{34}, \; \omega_3 = \frac{(57 + 33\sqrt{2})\,t + (53 + 45\sqrt{2})}{34}. \tag{3.18}$$

Therefore if we choose J_1 to be the one in (3.14) with the values in (3.17) and (3.18) and ask the relation (3.16) for all $i \geq 1$, then we will have (3.15) for all $i \geq 1$.

Now, using (3.13), (3.16) and (3.17), we can list the function values and gradients of λ and μ at the vertices V_i's of the octagon Ω into Table 1.

By using the special choice of Ω and observing the values in Table 1, we can ask the functions λ and μ to have the properties

$$\mu(x_1, x_2) = \lambda(-x_2, x_1) \tag{3.19}$$

and

$$\lambda(x_1, x_2) = -\lambda(-x_1, -x_2) \tag{3.20}$$

for all $(x_1, x_2) \in \mathcal{R}^2$. Further, by (3.20), we could think of constructing the function λ by a polynomial function with only odd orders.

In the next subsection, we will construct a simple function that only meets the requirements (a) and (c) that are described in the first section. In Sect. 3.4, we construct a complicated function that can meet all the requirements (a), (b) and (c).

3.3 A simple function that meets (a) and (c)

If we ignore the convexity requirement (b) of the line search function, we can construct a relatively simple function $\lambda(x_1, x_2)$ to meet the interpolation conditions listed in Table 1 and the function $\mu(x_1, x_2)$ is then given by (3.19).

In fact, we can check that the following function

$$\lambda_f(x_1,x_2)=\left(15-\frac{21}{2}\sqrt{2}\right)x_1^5+\left(6-\frac{9}{2}\sqrt{2}\right)\left(x_1^4x_2+2x_1^2x_2^3\right)+\left(-15+10\sqrt{2}\right)x_1^3 \\ +(-1+\sqrt{2})(6x_1^2x_2+x_2^3)+\left(\frac{15}{4}-\frac{15}{8}\sqrt{2}\right)x_1-\frac{9}{8}\sqrt{2}x_2.$$

has values of $0, 1, -\sqrt{2}, 1$ at vertices V_1, V_2, V_3, V_4, respectively, and zero derivatives at all the vertices V_i's. Further, we can check the following function

$$\lambda_g(x_1, x_2) = -\frac{3-2\sqrt{2}}{4}(2x_1^2-1)[2x_1^2-(3+2\sqrt{2})]x_2$$

has zero function values at all the vertices V_i's, but could be used to interpolate the derivatives of $\lambda(x_1, x_2)$ at V_i's. Consequently, we know that

$$\bar{\lambda}(x_1, x_2) = \lambda_f(x_1, x_2) + \omega_1\lambda_g(x_1, x_2) + \omega_3\lambda_g(-x_2, x_1) \tag{3.21}$$

meets all the interpolation conditions required by $\lambda(x_1, x_2)$ in Table 1. Consequently, we could say that for the function defined by (3.11), (3.21) and (3.19), the BFGS method with the Wolfe line search using the unit initial step does not converge.

Observing that

$$\Psi_i(\alpha) = t\Psi_i(\alpha), \quad \Psi_i'(\alpha) = t\Psi_i'(\alpha) \quad \text{at } \alpha = 0,\ 1,$$

we ideally wish there is the following relation between $\Psi_i(\alpha)$ and $\Psi_{i+1}(\alpha)$:

$$\Psi_{i+1}(\alpha) = t\,\Psi_i(\alpha), \quad \text{for all } \alpha \ge 0 \tag{3.22}$$

so that the neighboring line search functions only differ up to a constant multiplier of t. However, the choice (3.21) of $\bar{\lambda}(x_1, x_2)$ does not lead to (3.22). Nevertheless, we can propose the following compensation function

$$\lambda_c(x_1, x_2) = x_1\left[x_1^2 + x_2^2 - (2+\sqrt{2})\right]^2$$

and define

$$\hat{\lambda}(x_1, x_2) = \bar{\lambda}(x_1, x_2) + \bar{c}_1 \lambda_c(x_1, x_2) + \bar{c}_2 \lambda_c(-x_2, x_1), \tag{3.23}$$

where

$$\bar{c}_1 = \frac{(-39 + 15\sqrt{2})t - (2529 - 1756\sqrt{2})}{272},$$
$$\bar{c}_2 = \frac{(-65 + 8\sqrt{2})t + (681 - 462\sqrt{2})}{272}.$$

In this case, the relation (3.22) will always hold and the corresponding line search function at the first iteration is

$$\hat{\Psi}_1(\alpha) = \sum_{i=0}^{6} \rho_i \, \alpha^i, \tag{3.24}$$

where

$$\rho_6 = \frac{(-3186 + 2233\sqrt{2})t + (3414 - 2398\sqrt{2})}{2},$$

$$\rho_5 = \frac{(40347402 - 28052171\sqrt{2})t - (44367385 - 31014036)}{9266},$$

$$\rho_4 = \frac{(-18674096 + 12804379\sqrt{2})t + (20961677 - 14552090\sqrt{2})}{4633},$$

$$\rho_3 = \frac{(12350618 - 8366705\sqrt{2})t - (13736673 - 9508816\sqrt{2})}{9266},$$

$$\rho_2 = \frac{(-211982 + 366273\sqrt{2})t + (40258 - 176758\sqrt{2})}{9266},$$

$$\rho_1 = \mathbf{g}_1^T \boldsymbol{\delta}_1 = (-44 + 13\sqrt{2})t + (40 - 18\sqrt{2}),$$

$$\rho_0 = f(\mathbf{x}_1) = \frac{(3954 - 4376\sqrt{2})t + (18866 - 9563\sqrt{2})}{4633}.$$

To sum up at this stage, for the function (3.11) with $\lambda(x_1, x_2)$ given in (3.21) or (3.23) and $\mu(x_1, x_2) = \lambda(-x_2, x_1)$, if the initial point is $\mathbf{x}_1 = (a_1, b_1, p_1, q_1)^T$ (see (3.4), (3.8), (2.47) for their values) and if the initial matrix is H_1 given by (2.58) and (2.59), then the BFGS method (1.2) and (1.5) with $\alpha_k \equiv 1$ will generate the iterations in (3.5) whose gradients are given by (2.4)–(2.5). Therefore the method will asymptotically cycle around the eight vertices of a regular octagon without approaching a stationary point or pushing $f(\mathbf{x}_k) \to -\infty$.

3.4 A complicated function that meets (a), (b) and (c)

To meet the requirement (*b*), this subsection gives a further compensation to the function $\hat{\lambda}(x_1, x_2)$ in (3.23) such that each line search function $\Psi_k(\alpha)$ is strongly convex.

To this aim, we consider the straight line connecting $\mathbf{x}_{8j+i}$ and $\mathbf{x}_{8j+i+1}$:

$$\mathbf{l}_i(\alpha) = \mathbf{x}_{8j+i} + \alpha\, \boldsymbol{\delta}_{8j+i} = \begin{pmatrix} a_i + \alpha\, \eta_i \\ b_i + \alpha\, \xi_i \\ t^{8j}(p_i + \alpha\, \gamma_i) \\ t^{8j}(q_i + \alpha\, \tau_i) \end{pmatrix}. \tag{3.25}$$

By (3.12), the line search function at the $(8j+i)$th iteration is

$$\begin{aligned}\Psi_{8j+i}(\alpha) = t^{8j}\big[&\lambda(a_i + \alpha\, \eta_i, b_i + \alpha\, \xi_i)\big(p_i + \alpha\, \gamma_i\big) \\ &+\mu(a_i + \alpha\, \eta_i, b_i + \alpha\, \xi_i)\big(q_i + \alpha\, \tau_i\big)\big].\end{aligned} \tag{3.26}$$

To be such that each line search function is a strongly convex function that has the unique minimizer $\alpha = 1$, we firstly construct $\Psi_1(\alpha)$ as follows

$$\Psi_1(\alpha) = \zeta_1(\alpha - 1)^{38} + \zeta_2(\alpha - 1)^2 + \zeta_3, \tag{3.27}$$

where

$$\begin{aligned}\zeta_1 &= \frac{(173256 - 38127\sqrt{2})t + (-138064 + 48586\sqrt{2})}{166788} \approx 1.5468E - 1, \\ \zeta_2 &= \frac{(377472 - 359709\sqrt{2})t + (-712544 + 577958\sqrt{2})}{166788} \approx 1.0405E - 3, \\ \zeta_3 &= \frac{(-11344 + 6675\sqrt{2})t + (42494 - 26967\sqrt{2})}{4633} \approx 6.1270E - 1.\end{aligned}$$

Due to our special construction, we know from (3.3) and $|t| < 1$ that the third and fourth components of $\{\mathbf{x}_k\}$ tend to zero. This with the general function form (3.11) implies that the limit of $f(\mathbf{x}_k)$ is $f^* = 0$. So the above Ψ_1 is constructed such that

$$\Psi_1(0) = f(\mathbf{x}_1),\ \Psi_1(1) = f(\mathbf{x}_2),\ \Psi_1'(0) = \mathbf{g}_1^T \boldsymbol{\delta}_1,\ \Psi_1'(1) = \mathbf{g}_2^T \boldsymbol{\delta}_1 = 0, \tag{3.28}$$

where the values of $f(\mathbf{x}_1)$, $f(\mathbf{x}_2)$ and $\mathbf{g}_1^T \boldsymbol{\delta}_1$ are given in Sect. 3.1 and by $f^* = 0$. In addition, the positiveness of ζ_i's implies that Ψ_1 is strongly convex. Thus we see that Ψ_1 is a desired strongly convex function that takes $\alpha = 1$ as the unique minimizer. A reason why we choose a polynomial (3.27) of degree 38 will be explained in the last section.

To utilize the function $\hat{\lambda}(x_1, x_2)$ in (3.23), we firstly develop the following relation between (3.27) and (3.24),

$$\Psi_1(\alpha) = \hat{\Psi}_1(\alpha) + \sum_{i=0}^{4} \big[\alpha\,(\alpha - 1)\big]^{4i+2} \sum_{j=0}^{7} \sigma_{8i+j}\, \alpha^j, \tag{3.29}$$

where $\sigma_i = 0\,(i = 39, \ldots, 35)$ and

$$
\begin{array}{lll}
\sigma_{34} = \zeta_1, & \sigma_{33} = -20\zeta_1, & \sigma_{32} = 190\zeta_1,\\
\sigma_{31} = -1140\zeta_1, & \sigma_{30} = 9405\zeta_1, & \sigma_{29} = -41724\zeta_1,\\
\sigma_{28} = 134406\zeta_1, & \sigma_{27} = -346104\zeta_1, & \sigma_{26} = 735471\zeta_1,\\
\sigma_{25} = -1307504\zeta_1, & \sigma_{24} = 1961256\zeta_1, & \sigma_{23} = -2496144\zeta_1,\\
\sigma_{22} = 12688732\zeta_1, & \sigma_{21} = -28289632\zeta_1, & \sigma_{20} = 38155344\zeta_1,\\
\sigma_{19} = -37442160\zeta_1, & \sigma_{18} = 30421755\zeta_1, & \sigma_{17} = -21474180\zeta_1,\\
\sigma_{16} = 13123110\zeta_1; & \sigma_{15} = -6906900\zeta_1, & \sigma_{14} = 30735705\zeta_1,\\
\sigma_{13} = -55057860\zeta_1, & \sigma_{12} = 51389130\zeta_1, & \sigma_{11} = -28048800\zeta_1,\\
\sigma_{10} = 10518300\zeta_1, & \sigma_{9} = -3365856\zeta_1, & \sigma_{8} = 906192\zeta_1,\\
\sigma_{7} = -201376\zeta_1, & \sigma_{6} = 841464\zeta_1, & \sigma_{5} = -1357056\zeta_1,\\
\sigma_{4} = 1041600\zeta_1, & \sigma_{3} = -376992\zeta_1, & \sigma_{2} = 58905\zeta_1 - \rho_6,\\
\multicolumn{3}{l}{\sigma_1 = -7140\zeta_1 - \rho_5 + 4\rho_6,\ \sigma_0 = 630\zeta_1 - \rho_4 + 3\rho_5 - 6\rho_6.}
\end{array}
$$

Secondly, noting that the following relations hold for all $i \geq 1$,

$$
\begin{pmatrix} a_{i+1} + \alpha\,\eta_{i+1} \\ b_{i+1} + \alpha\,\xi_{i+1} \end{pmatrix} = R_1 \begin{pmatrix} a_i + \alpha\,\eta_i \\ b_i + \alpha\,\xi_i \end{pmatrix}, \quad \begin{pmatrix} p_{i+1} + \alpha\,\gamma_{i+1} \\ q_{i+1} + \alpha\,\tau_{i+1} \end{pmatrix} = t\,R_2 \begin{pmatrix} p_i + \alpha\,\gamma_i \\ q_i + \alpha\,\tau_i \end{pmatrix},
$$

we look for some element functions $\phi(x_1, x_2)$ that satisfy for all $(x_1, x_2) \in \mathcal{R}^2$,

$$
\begin{pmatrix} \phi(\bar{x}_1, \bar{x}_2) \\ \phi(-\bar{x}_2, \bar{x}_1) \end{pmatrix} = R_2 \begin{pmatrix} \phi(x_1, x_2) \\ \phi(-x_2, x_1) \end{pmatrix}, \quad \text{where } \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = R_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.30}
$$

There are a lot of possibilities for the choice of $\phi(x_1, x_2)$. The following are some of them required in this paper.

$$
\begin{aligned}
&\phi_1(x_1, x_2) = x_1^3 - 3x_1x_2^2,\ \phi_3(x_1, x_2) = x_1^3x_2^2 - x_1x_2^4,\\
&\phi_5(x_1, x_2) = x_1^5 - 5x_1^3x_2^2,\ \phi_7(x_1, x_2) = x_1^5x_2^2 - x_1x_2^6.
\end{aligned} \tag{3.31}
$$

We take ϕ_1 as an illustrative example. In fact, for any $(x_1, x_2) \in R^2$, $(\bar{x}_1, \bar{x}_2)^T = R_1(x_1, x_2)^T$ means that

$$
\bar{x}_1 = \frac{\sqrt{2}}{2}(x_1 - x_2), \quad \bar{x}_2 = \frac{\sqrt{2}}{2}(x_1 + x_2).
$$

Therefore we have that

$$
\begin{aligned}
\begin{pmatrix} \phi(\bar{x}_1, \bar{x}_2) \\ \phi(-\bar{x}_2, \bar{x}_1) \end{pmatrix} &= \begin{pmatrix} \bar{x}_1(\bar{x}_1^2 - 3\bar{x}_2^2) \\ \bar{x}_2(3\bar{x}_1^2 - \bar{x}_2^2) \end{pmatrix}\\
&= \begin{pmatrix} \frac{\sqrt{2}}{4}(x_1 - x_2)((x_1 - x_2)^2 - 3(x_1 + x_2)^2) \\ \frac{\sqrt{2}}{4}(x_1 + x_2)(3(x_1 - x_2)^2 - (x_1 + x_2)^2) \end{pmatrix}\\
&= \begin{pmatrix} -\frac{\sqrt{2}}{2}(x_1 - x_2)(x_1^2 + 4x_1x_2 + x_2^2) \\ \frac{\sqrt{2}}{2}(x_1 + x_2)(x_1^2 - 4x_1x_2 + x_2^2) \end{pmatrix}
\end{aligned}
$$

$$= \begin{pmatrix} -\frac{\sqrt{2}}{2}[(x_1^3 - 3x_1x_2^2) + (3x_1^2x_2 - x_2^3)] \\ \frac{\sqrt{2}}{2}[(x_1^3 - 3x_1x_2^2) - (3x_1^2x_2 - x_2^3)] \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{\sqrt{2}}{2}[\phi(x_1, x_2) + \phi(-x_2, x_1)] \\ \frac{\sqrt{2}}{2}[\phi(x_1, x_2) - \phi(-x_2, x_1)] \end{pmatrix}$$

$$= R_2 \begin{pmatrix} \phi(x_1, x_2) \\ \phi(-x_2, x_1) \end{pmatrix}.$$

So $\phi_1(x_1, x_2)$ is a desired element function. Similarly, we can check that the other ϕ_i's in (3.31) have the same property. In addition, it is easy to see that if $\phi(x_1, x_2)$ has the property (3.30), then $\phi(-x_2, x_1)$ keeps the same property. So we also consider the following element function

$$\phi_{2i}(x_1, x_2) = \phi_{2i-1}(-x_2, x_1), \quad \text{for } i = 1,\ 2,\ 3.$$

Thirdly, we consider the linear combination of $\phi_i (i = 1, \ldots, 8)$ and define

$$\lambda_\phi(x_1, x_2) = \sum_{i=1}^{8} \nu_i\, \phi_i(x_1, x_2). \tag{3.32}$$

Meanwhile, we define $\mu_\phi(x_1, x_2) = \lambda_\phi(-x_2, x_1)$. For any vector $\boldsymbol{\kappa} = (\kappa_0, \ldots \kappa_7)^T$ in $\mathcal{R}^8$, we are going to claim that there exists a unique solution of $\boldsymbol{\nu} := (\nu_1, \ldots, \nu_8)^T$ such that the related functions λ_ϕ and μ_ϕ satisfy

$$\begin{aligned} &\lambda_\phi(a_1 + \alpha\, \eta_1, b_1 + \alpha\, \xi_1)\big(p_1 + \alpha\, \gamma_1\big) \\ &\quad + \mu_\phi(a_1 + \alpha\, \eta_1, b_1 + \alpha\, \xi_1)\big(q_1 + \alpha\, \tau_1\big) = \sum_{i=0}^{7} \kappa_i\, \alpha^i. \end{aligned} \tag{3.33}$$

In fact, it is easy to see that the above relation leads to a linear system of $\boldsymbol{\nu}$:

$$W\, \boldsymbol{\nu} = \boldsymbol{\kappa}. \tag{3.34}$$

Direct calculations show that

$$W = \big(W_1 + W_2 \sqrt{2}\big)\, t + (W_3 + W_4 \sqrt{2}), \tag{3.35}$$

where W_i $(i = 1, \ldots, 4)$ are given in the Appendix.

By further calculations, we know that

$$\mathrm{Det}(W) = \big(\bar{c}_3 + \bar{c}_4 \sqrt{2}\big)\, t + \big(\bar{c}_5 + \bar{c}_6 \sqrt{2}\big) \neq 0,$$

where

$$\bar{c}_3 = -272578451122583219133441287912493910710378004\,48,$$
$$\bar{c}_4 = -193544036254601538702134048281429136977578475\,52,$$

$$\bar{c}_5 = 21709509966956378375726991567920438736281239552,$$
$$\bar{c}_6 = 15449462608361267330121261246186521297040162816.$$

Hence W is nonsingular and (3.34) is a nonsingular linear system. Therefore for any vector $\boldsymbol{\kappa} \in \mathcal{R}^8$, there exists a unique $\boldsymbol{\nu}$ or λ_ϕ such that the relation (3.33) holds. For convenience, we denote such λ_ϕ by $\lambda_{\phi,\boldsymbol{\kappa}}$.

Denoting the following vectors related to the coefficients $\{\sigma_i;\ i = 0, \ldots, 39\}$ in (3.29),

$$\boldsymbol{\kappa}_i = (\sigma_{8i},\ \ldots, \sigma_{8i+7})^T, \quad i = 0, \ldots, 4,$$

and noticing that for any point $\mathbf{x} = (x_1,\ x_2,\ x_3,\ x_4)^T$ in the line $\mathbf{l}_i(\alpha)$ (i any),

$$x_1^2 + x_2^2 - (2 + \sqrt{2}) = 2\,\alpha\,(\alpha - 1), \tag{3.36}$$

we can finally present the desired function of λ:

$$\lambda(x_1, x_2) = \hat{\lambda}(x_1, x_2) + \sum_{i=0}^{4} \left[\frac{x_1^2 + x_2^2 - (2 + \sqrt{2})}{2} \right]^{4i+2} \lambda_{\phi,\boldsymbol{\kappa}_i}(x_1, x_2). \tag{3.37}$$

Thus by the choice of $\hat{\lambda}(x_1, x_2)$ in (3.23), the relation (3.36), the definitions of $\boldsymbol{\kappa}_i$ and $\lambda_{\phi,\boldsymbol{\kappa}_i}$ and the relation (3.29), we know that the function (3.11) with $\lambda(x_1,\ x_2)$ given in (3.37) and $\mu(x_1,\ x_2) = \lambda(-x_2,\ x_1)$ not only satisfies those necessary interpolation conditions but gives the line search function

$$\Psi_{i+1}(\alpha) = t^i\ \Psi_1(\alpha), \quad \text{for any } i \geq 0, \tag{3.38}$$

where $\Psi_1(\alpha)$ is given by (3.27).

Since the polynomial function $\lambda(x_1, x_2)$ in (3.37) is of order 43, we see that the final objective function in (3.11) is a polynomial of order 44 and hence is infinitely times differentiable. The line search functions, which are given by (3.38) and (3.27), is strongly convex and has the unit stepsize as its unique minimizer. The requirement (b) is then satisfied.

In summary, for the function (3.11) with $\lambda(x_1,\ x_2)$ given in (3.37) and $\mu(x_1,\ x_2) = \lambda(-x_2,\ x_1)$, if the initial point is $\mathbf{x}_1 = (a_1,\ b_1,\ p_1,\ q_1)^T$ (see (3.4), (3.8), (2.47) for their values) and if the initial matrix is H_1 given by (2.58) and (2.59), then the BFGS method (1.2) and (1.5) with $\alpha_k \equiv 1$ will generate the iterations in (3.5) whose gradients are given by (2.4)–(2.5). Therefore the method will asymptotically cycle around the eight vertices of a regular octagon without approaching a stationary point or pushing $f(\mathbf{x}_k) \to -\infty$. This counter-example is perfect in the sense that all the requirements (a), (b) and (c) are satisfied.

4 Concluding remarks

No matter whether the simple example(s) in Sect. 3.3 or the complicated example in Sect. 3.4, we can see that the BFGS method with unit stepsizes produces the same iterations $\{\mathbf{x}_k\}$ and simultaneously, their objective functions provide the same gradients $\{\mathbf{g}_k\}$. As analyzed in Sect. 3.1, the unit stepsize is acceptable by the Wolfe line search, the Armijo line search or the Goldstein line search. If we do not impose the convexity on the line search function $\Psi_k(\alpha)$ for all k, it is possible for us to construct a relatively simple polynomial example shown in Sect. 3.3. The examples have the dimension 4, but can be used to show that the BFGS method may also fail for general functions when $n \geq 5$ since four-dimensional functions can be regarded as special cases of higher-dimensional functions. In the case that $n \geq 5$, we can just take the same objective function and ask the starting point $\mathbf{x}_1$ to have zero components except the first four. Then for all k, the first four components of $\mathbf{x}_k$ remain the same as in the example and its components from the fifth will be always zero.

The counter-example in Sect. 3.4 is perfect in stepsize choices and function properties. However, it is still not perfect in the sense that the objective function is very complicated with large numbers and some coefficients are expressed by nonsingular linear systems. Here we provide the reason why a polynomial (3.27) of degree 38 is used to guarantee the strong convexity of $\Psi_1(\alpha)$. Equivalently, the problem of finding a strictly convex polynomial that satisfies (3.28) (in this case $\Psi_1''(\alpha)$ is nonnegative for all $\alpha \in \mathcal{R}$) can be transferred to the feasibility problem of some semi-definite program (SDP). This is because, by Shor [19], if $n = 1$, a polynomial is nonnegative if and only if it can be written into a sum of squares (s.o.s.); further, by Lasserre [10], for an any-dimensional polynomial, it has the form of s.o.s. if and only if the coefficient matrix, which is formed when lifting the outer product of the vector of monomials and its transpose into the matrix variable, is semi-definite. The other interpolation conditions can be treated as linear constraints. However, even when we chose some relatively high numbers for the order of the desired polynomial, our numerical calculations showed that the corresponding SDP feasibility problems have no solution. This is not expected since there are many freedoms in these polynomials, but there are only four interpolation conditions in (3.28). Instead, we considered the following simple form for $\Psi_1(\alpha)$ with variable but even p.

$$\Psi_1(\alpha) = \bar{c}_1(\alpha - 1)^p + \bar{c}_2(\alpha - 1)^2 + \zeta_3, \tag{4.1}$$

where $\bar{c}_1$ and $\bar{c}_2$ are parameters and ζ_3 is the same constant in (3.27). With this form, it can be deduced from (3.28), $f^* = 0$ and the calculations in Sect. 3.1 that

$$\frac{1}{p} \leq \frac{\Psi_1(1) - \Psi_1(0)}{\Psi_1'(0)} = \frac{f(\mathbf{x}_2) - f(\mathbf{x}_1)}{\mathbf{g}_1^T \boldsymbol{\delta}_1} \approx 2.6483E-02. \tag{4.2}$$

Since the reciprocal of $2.6483E-02$ is about 37.76, we choose p to be the least even integer, that is 38, to meet the condition (4.2).

In spite of the existence of large numbers, we have observed the cycle exactly predicted by the simple example in Sect. 3.3 [with the function $\hat{\lambda}(x_1, x_2)$ given by (3.23)], thanks to the powerful symbolic computation software MAPLE. With such software, it is also possible to observe how numerical errors affect the example. When the machine error is set to 10^{-64} (this is possible in MAPLE), we found that the cycle can go on for ten rounds and during the rounds, the iterations really tend to the eight vertices of the octagon according to the predicted way. Influenced by the numerical errors, however, the iterations jump out from the cycle during the eleventh round.

For quasi-Newton methods, there is global convergence if the norms of the matrix H_k and its inverse are uniformly bounded for all k, in which case the angle between the quasi-Newton direction and the negative gradient must be uniformly less than some angle strictly less than $\pi/2$. For our examples, with the help of (2.2) and (2.57), we can see that the matrices $\{H_{8j+1};\ j=1,2,\ldots\}$ satisfy the relation (2.24). Consequently, we have for all $j \geq 1$,

$$H_{8j+1} = \mathrm{diag}(t^{-4j}E_2, t^{4j}E_2)\, H_1\, \mathrm{diag}(t^{-4j}E_2, t^{4j}E_2). \tag{4.3}$$

By comparing H_{8j+1} and the right matrices in (4.3) but with the middle matrix H_1 replaced by $\lambda_{\min}(H_1)E_4$ and $\lambda_{\max}(H_1)E_4$, respectively, we can show that

$$\lambda_{\max}(H_{8j+1}) \geq t^{-8j}\lambda_{\min}(H_1), \quad \lambda_{\min}(H_{8j+1}) \leq t^{8j}\lambda_{\max}(H_1), \tag{4.4}$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdots)$ mean the largest and smallest eigenvalues of the matrix. Since t is the decay parameter that lies in the interval $(0,\ 1)$, the relation (4.4) implies that both $||H_{8j+1}||_2$ and $||H_{8j+1}^{-1}||_2$ tend to infinity at an exponential rate. This analysis is also valid for $\{H_{8j+i};\ j=1,2,\ldots\}$ with other i's.

The number of the cyclic points in the above example(s) is eight due to the special choices of θ_1 and θ_2 in (2.3). This number of eight could be decreased to seven since if

$$\theta_1 = \frac{2}{7}\pi, \qquad \theta_2 = \frac{4}{7}\pi,$$

the consistent conditions (2.33)–(2.36) also allow a nonzero solution of t, whose numerical value is approximately 0.8642. However, it is difficult to obtain its analytic expression. In addition, we found that the system (2.33)–(2.36) has no solution of t in $(-1,\ 1)$ for

$$\theta_1 = \frac{i}{6}\pi, \quad \theta_2 = \frac{j}{6}\pi, \quad \text{where} i,\ j \text{is any integer in [1,6]},$$

which implies that the number of cyclic points cannot be decreased to six.

As mentioned in Sect. 1, if the stepsize is chosen to be the first local minimizer along the line; namely, by (1.9), Powell [17] established the global convergence of the BFGS method for general differentiable functions when $n = 2$. We argued there that it is difficult to extend Powell's result to the case that $n = 3$. Considerable attentions have also been drawn to the construction of a three-dimensional example (see also Section 5 of Powell [17]). This construction is almost successful, but remains one condition always not satisfied. It is not known yet whether the BFGS method with the specific line search (1.9) converges for three-dimensional general differentiable functions.

Another direction of research is how to present a suitable modification of the BFGS algorithm, with which global convergence can be established for general nonconvex functions. A typical work of this kind is Li and Fukushima [11].

Acknowledgments This work is initiated while the author was visiting Cambridge from July to October in 1997. He is very grateful to Professor Michael Powell and Professor Ya-xiang Yuan for their financial supports of his visit and consistent encouragements in many years. He also thanks the two anonymous referees very much for their useful suggestions and comments on this paper. This work was partially supported by the Chinese NSF grants (no. 10571171 and 10831106), the CAS grant (no. kjcx-yw-s7-03) and the China National Funds for Distinguished Young Scientists (no. 11125107).

Appendix

The matricies W_i $(i = 1, \ldots, 4)$ in the relation (3.35) are given in the following sequentially.

$$
\begin{bmatrix}
-62168 & -15364 & -50467 & -19383 & -50467 & -19383 & -170784 & -69850 \\
325732 & 7844 & 164094 & 202386 & 412766 & -356414 & 617846 & 774894 \\
-993360 & 196008 & -4160 & -201438 & -2561400 & 1153290 & -149012 & -1366496 \\
921688 & 0 & 3698860 & -613452 & -114596 & 463300 & -869492 & -653360 \\
-296512 & -333576 & -407704 & 890484 & -2569920 & -2316500 & -168300 & 3018560 \\
0 & 0 & 296512 & -333576 & 1546864 & 1667880 & 1142176 & -3712088 \\
0 & 0 & 0 & 0 & -593024 & -667152 & -1408432 & 2448120 \\
0 & 0 & 0 & 0 & 0 & 0 & 593024 & -667152
\end{bmatrix}
$$

$$
\begin{bmatrix}
-38766 & -23402 & -34925 & -15542 & -34925 & -15542 & -120317 & -50467 \\
64624 & 102164 & 94362 & 165678 & 249426 & -238058 & 422668 & 564826 \\
94320 & 101688 & 144215 & -260041 & -1444145 & 928375 & 25696 & -1083960 \\
-563244 & -500364 & -488404 & -117052 & 1201020 & -648620 & -1191110 & 3882 \\
315044 & 315044 & 213118 & 206782 & 1009350 & 880270 & 1283770 & 1018070 \\
0 & 0 & -18532 & -18532 & -1700980 & -1575220 & -1143596 & -1018308 \\
0 & 0 & 0 & 0 & 630088 & 630088 & 463300 & 450628 \\
0 & 0 & 0 & 0 & 0 & 0 & -37064 & -37064
\end{bmatrix}
$$

$$\begin{bmatrix} 66580 & -22806 & 48866.5 & -2955.5 & 48866.5 & -2955.5 & 162176 & -419 \\ -305222 & 82838 & -149974 & -138071 & -193910 & 377473 & -543381 & -565187 \\ 1072146 & -156966 & 2300535 & 8467 & 135353 & -905635 & 75885 & 1149501 \\ -1085404 & -74128 & 43774 & 654056 & -3792690 & -741280 & 700702 & 782668 \\ 333576 & 333576 & 444768 & -845924 & 3017860 & 2409160 & 429750 & -2916010 \\ 0 & 0 & -296512 & 296512 & -1948424 & -1704944 & -1357948 & 3555920 \\ 0 & 0 & 0 & 0 & 667152 & 667152 & 1482560 & -2284872 \\ 0 & 0 & 0 & 0 & 0 & 0 & -593024 & 593024 \end{bmatrix}$$

$$\begin{bmatrix} 31153 & 16895 & 32221.5 & 2746 & 32221.5 & 2746 & 113309.5 & 2536.5 \\ 49886 & -124014 & -72850 & -147478 & -49206 & 243454 & -360117 & -438519 \\ -280980 & 12822 & -169365 & 304282 & 980815 & -762350 & -120120 & 1044365 \\ 632088 & 444768 & -554550 & 465018 & 10702 & 509630 & 1166840 & -228060 \\ -315044 & -315044 & -194586 & -93378 & -1538890 & -880270 & -1213170 & -445520 \\ 0 & 0 & 18532 & -18532 & 1838668 & 1575220 & 1059760 & 467608 \\ 0 & 0 & 0 & 0 & -630088 & -630088 & -426236 & -149692 \\ 0 & 0 & 0 & 0 & 0 & 0 & 37064 & -37064 \end{bmatrix}$$

References

1. Byrd, R.H., Nocedal, J., Yuan, Y.X.: Global convergence of a class of quasi-Newton methods on convex problems. SIAM J. Numer. Anal. **24**, 1171–1190 (1987)
2. Broyden, C.G.: The convergence of a class of double rank minimization algorithms: 2. The new algorithm. J. Inst. Math. Appl. **6**, 222–231 (1970)
3. Dai, Y.H.: Convergence properties of the BFGS algorithm. SIAM J. Optim. **13**(3), 693–701 (2002)
4. Davidon, W.C.: Variable metric methods for minimization. SIAM J. Optim. **1**, 1–17 (1991)
5. Dennis, J.E., Moré, J.J.: Quasi-Newton method, motivation and theory. SIAM Rev. **19**, 46–89 (1977)
6. Fletcher, R.: A new approach to variable metric algorithms. Comput. J. **13**, 317–322 (1970)
7. Fletcher, R.: An Overview of Unconstrained Optimization. In: Spedicato, E. (ed.) Algorithms for Continuous Optimization: The State of Art, pp. 109–143. Kluwer, Dordrecht (1994)
8. Fletcher, R., Powell, M.J.D.: A rapidly convergent descent method for minimization. Comput. J. **6**, 163–168 (1963)
9. Goldfarb, D.: A family of variable metric methods derived by variational means. Math. Comp. **24**, 23–26 (1970)
10. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
11. Li, D.H., Fukushima, M.: On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. SIAM J. Optim. **11**, 1054–1064 (2001)

12. Mascarenhas, W.F.: The BFGS algorithm with exact line searches fails for nonconvex functions. Math. Program. **99**(1), 49–61 (2004)
13. Nocedal, J.: Theory of algorithms for unconstrained optimization. Acta Numer. **1**, 199–242 (1992)
14. Powell, M.J.D.: On the convergence of the variable metric algorithm. J. Inst. Math. Appl. **7**, 21–36 (1971)
15. Powell, M.J.D.: Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In: Cottle, R.W., Lemke, C.E. (eds.) Nonlinear Programming, SIAM-AMS Proceedings Vol. IX, pp. 53–72. SIAM, Philadelphia (1976)
16. Powell, M.J.D.: Nonconvex minimization calculations and the conjugate gradient method. In: Griffiths, D.F. (Ed.) Numerical Analysis, Lecture Notes in Math. 1066, pp. 122–141. Springer, Berlin (1984)
17. Powell, M.J.D.: On the convergence of the DFP algorithm for unconstrained optimization when there are only two variables. Math. Program. Ser. B **87**, 281–301 (2000)
18. Shanno, D.F.: Conditioning of quasi-Newton methods for function minimization. Math. Comp. **24**, 647–650 (1970)
19. Shor, N.Z.: Quadratic optimization problems. Sov. J. Comput. Syst. Sci. **25**, 1–11 (1987)
20. Sun, W.Y., Yuan, Y.X.: Optimization Theory and Methods: Nonlinear Programming. Springer, New York (2006)
21. Yuan, Y.X.: Numerical Methods for Nonlinear Programming. Shanghai Scientific and Technical Publishers, Shanghai (1993); (in Chinese)