

# Two-Order Approximate and Large Stepsize Numerical Direction Based on the Quadratic Hypothesis and Fitting Method

Xiaoli Yin, Chunming Li, and Yuan Zhang

**Abstract**—Many effective optimization algorithms require partial derivatives of objective functions, while some optimization problems' objective functions have no derivatives. According to former research studies, some search directions are obtained using the quadratic hypothesis of objective functions. Based on derivatives, quadratic function assumptions, and directional derivatives, the computational formulas of numerical first-order partial derivatives, second-order partial derivatives, and numerical second-order mixed partial derivatives were constructed. Based on the coordinate transformation relation, a set of orthogonal vectors in the fixed coordinate system was established according to the optimization direction. A numerical algorithm was proposed, taking the second order approximation direction as an example. A large stepsize numerical algorithm based on coordinate transformation was proposed. Several algorithms were validated by an unconstrained optimization of the two-dimensional Rosenbrock objective function. The numerical second order approximation direction with the numerical mixed partial derivatives showed good results. Its calculated amount is 0.2843% of that of without second-order mixed partial derivative. In the process of rotating the local coordinate system 360°, because the objective function is more complex than the quadratic function, if the numerical direction derivative is used instead of the analytic partial derivative, the optimization direction varies with a range of 103.05°. Because theoretical error is in the numerical negative gradient direction, the calculation with the coordinate transformation is 94.71% less than the calculation without coordinate transformation. If there is no theoretical error in the numerical negative gradient direction or in the large-stepsize numerical optimization algorithm based on the coordinate transformation, the sawtooth phenomenon occurs. When each numerical mixed partial derivative takes more than one point, the optimization results

cannot be improved. The numerical direction based on the quadratic hypothesis only requires the objective function to be obtained, but does not require derivability and does not take into account truncation error and rounding error. Thus, the application scopes of many optimization methods are extended.

**Index Terms**—Directional derivative, numerical differential, optimization method, quadratic function hypothesis.

## I. INTRODUCTION

OPTIMIZATION methods are widely used in various research fields and play an important role in scientific and technological problems [1]–[7]. At present, the research of optimization methods is mainly divided into multi-objective [8]–[10], intelligence, genetic algorithm [10], [11], clustering [12] methods, and so on. Classical multidimensional optimization methods are also studied [13].

Many efficient optimization methods need derivative and partial derivative information of the objective function, such as the linear fitting derivative method, high-order second approximate fixed-point method, tangent intersection point method, negative gradient direction method, zigzag line negative gradient direction method (blind- walking method), second approximate direction method [14], linear fitting gradient method [15], various conjugate direction methods [16]–[18], the optimal available direction method (feasible direction method), half step method [19], and so on. These methods are not useful in optimization problems where the objective function is not differentiable. Numerical differentiation algorithms [20] may allow these methods to work.

Comparing the above algorithms, the second approximation direction is the best, and the direction of the negative gradient is most commonly used. The negative gradient direction is the fastest local descent direction, and usually deviates from the extreme value point in the whole design space. The conjugate directions constructed on the basis of the negative gradient direction have better optimization efficiency [21], [22].

The numerical differential algorithm can not only determine the numerical partial derivative, but also obtain the numerical directional derivative in the general direction.

Looking at the teaching contents of the course “mechanical optimum design” or “optimization methods” and the research literature of optimization methods, and at classical optimization methods, it can be determined that they are all proposed under the quadratic objective function hypothesis and then applied to the general objective function. Although

Manuscript received November 9, 2018; revised March 11, 2019; accepted May 14, 2019. This work was supported in part by the Teaching Reform Research Foundation of Shengli College in China University of Petroleum (East China) (JG201725), the Natural Science Foundation Shandong Province of China (ZR2018PEE009), and the Project of Science and Technology of Shandong Universities in China (J17KA044, J17KB061). Recommended by Associate Editor Xin Luo. (Corresponding author: Chunming Li.)

Citation: X. L. Yin, C. M. Li, and Y. Zhang, “Two-order approximate and large stepsize numerical direction based on the quadratic hypothesis and fitting method,” *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 3, pp. 901–909, May 2020.

X. L. Yin and C. M. Li are with the School of Mechanical and Electrical Engineering, China University of Petroleum (East China), Qingdao 266580, and also with the Shengli College in China University of Petroleum (East China), Dongying 257061, China (e-mail: yinxl@slcupc.edu.cn; lchming@126.com).

Y. Zhang is with the School of Pipeline and Civil Engineering, China University of Petroleum (East China), Qingdao 266580, and also with the Shengli College in China University of Petroleum (East China), Dongying 257061, China (e-mail: 40208235@qq.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2019.1911735

there is an optimization direction based on finite differences [23], the truncation error and rounding error are emphasized while the quadratic hypothesis is neglected due to the fitting effect of the objective function. By representing each derivative in the system of differential equations by its finite difference approximation [24], the quadratic hypothesis is ignored. The research based on the quadratic hypothesis and fitting method is more general.

Classical optimization methods are not proposed to deal with large-scale matrices. However, some large-scale sparse matrix problems borrow optimization ideas, such as the graph coloring problem [25] and nonlinear equations solving method [26]. Although document [25] deals with the optimization direction based on finite differences, it is far from the research in this paper.

According to the characteristics of the optimization method, the numerical partial derivative and the numerical directional derivative of the objective function are constructed by using the numerical differentiation algorithm under the quadratic function assumption. On this basis, the optimization effect of the numerical second approximate direction and the local coordinate negative gradient direction is studied.

## II. CONSTRUCTION METHOD OF NUMERICAL FIRST AND TWO ORDER PARTIAL DERIVATIVES BASED ON THE QUADRATIC HYPOTHESIS

The optimization methods with a clear direction of optimization are put forward under the unimodal assumption. The optimization method based on the quadratic function is universal. There is no universal significance in the study of higher order objective functions. In the design space, the objective function is assumed to be quadratic in any direction, so there are three undetermined coefficients. Take the current point  $\mathbf{x}^{(k)}$  as the center and taking point  $\mathbf{x}^{(k-1)}$  and point  $\mathbf{x}^{(k+1)}$  by a microstep  $\Delta s$  at two sides. Assume that the objective function values of the three points are  $f^{(k-1)}$ ,  $f^{(k)}$  and  $f^{(k+1)}$ , respectively.  $\Delta s$  shall be reasonably determined according to the computer's expressive ability and termination condition value. Three equations can be obtained from the above three points. Thus the above three parameters are obtained so that the quadratic function of the hypothesis is known. The partial derivatives at points  $\mathbf{x}^{(k-1)}$ ,  $\mathbf{x}^{(k)}$  and  $\mathbf{x}^{(k+1)}$  are shown as follows:

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^{(k-1)}} = \frac{-f^{(k+1)} - 3f^{(k-1)} + 4f^{(k)}}{2\Delta s} \quad (1)$$

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^{(k)}} = \frac{f^{(k+1)} - f^{(k-1)}}{2\Delta s} \quad (2)$$

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}^{(k+1)}} = \frac{3f^{(k+1)} + f^{(k-1)} - 4f^{(k)}}{2\Delta s} \quad (3)$$

The above three formulas are consistent with the results obtained by the Lagrange interpolation function method. The second partial derivatives at the three points are equal.

$$\left. \frac{\partial^2 f}{\partial x_i^2} \right|_{\mathbf{x}^{(k)}} = \frac{f^{(k+1)} + f^{(k-1)} - 2f^{(k)}}{(\Delta s)^2} \quad (4)$$

Equation (4) is consistent with the Newton center difference formula. The numerical gradient and the numerical second partial derivative at the point  $\mathbf{x}^{(k)}$  can be obtained from the above 4 formulas, but the second mixed partial derivative cannot be obtained.

## III. CONSTRUCTION METHOD OF NUMERICAL DIRECTIONAL DERIVATIVE

The directional derivative is the rate of change of a function along a certain direction, which is a generalization of the concept of a partial derivative. Let  $s$  be the direction from  $\mathbf{x}^{(k)}$  to  $\mathbf{x}^{(k+1)}$ , where  $s$  can be expressed as follows:

$$s = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \quad (5)$$

The gradient function of the objective function at point  $\mathbf{x}^{(k)}$  is as follows:

$$\nabla f(\mathbf{x}^{(k)}) = \left[ \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \right]_{\mathbf{x}^{(k)}}^T \quad (6)$$

The directional derivative can be regarded as the point product of the gradient vector and the unit directional vector [27], equal to the length of the vector  $\nabla f(\mathbf{x})$  projected on  $s$ .

$$\begin{aligned} \left. \frac{\partial f}{\partial s} \right|_{\mathbf{x}^{(k)}} &= [\nabla f(\mathbf{x}^{(k)})]^T \cdot s \\ &= \|\nabla f(\mathbf{x}^{(k)})\| \cdot \|s\| \cdot \cos(\nabla f(\mathbf{x}^{(k)}), s) \\ &= \frac{\nabla f(\mathbf{x}^{(k)})^T [\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}]}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|} \end{aligned} \quad (7)$$

where  $\cos(\nabla f(\mathbf{x}), s)$  is the cosine of the angle between  $\nabla f(\mathbf{x}^{(k)})$  and  $s$ .  $k$  is the optimal serial number. The minute step along  $s$  is decomposed into the minute step along  $N$  coordinate axes. Then, according to the definition of the partial derivative (7) can be obtained.

According to the definition of directional derivative change rate, the numerical directional derivative at  $\mathbf{x}^{(m)}$ , the midpoint of  $\mathbf{x}^{(k+1)}$  and  $\mathbf{x}^{(k)}$  is as follows:

$$\left. \frac{\partial f}{\partial s} \right|_{\mathbf{x}^{(m)}} = \frac{f^{(k+1)} - f^{(k)}}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|} \quad (8)$$

If  $\mathbf{x}^{(k+1)}$  is sufficiently close to  $\mathbf{x}^{(k)}$ , the numerical directional derivative at point  $\mathbf{x}^{(k)}$  can be expressed approximately by (8). Calculate the objective function value  $f^{(m)}$  of point  $\mathbf{x}^{(m)}$ ; then the numerical directional derivatives at point  $\mathbf{x}^{(k)}$  are as follows:

$$\left. \frac{\partial f}{\partial s} \right|_{\mathbf{x}^{(k)}} = \frac{-f^{(k+1)} - 3f^{(k)} + 4f^{(m)}}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|} \quad (9)$$

A linear equation of gradient vectors can be obtained from (7) and (8). By searching for  $N$  points near the current point  $\mathbf{x}^{(0)}$  and establishing the above equations, a set of  $N$ -ary linear equations can be obtained, and the numerical gradient of the objective function can be solved.

$$\begin{cases} A\nabla f(\mathbf{x}^{(0)}) = B \\ A = \begin{bmatrix} x_1^{(1)} - x_1^{(0)} & \dots & x_N^{(1)} - x_N^{(0)} \\ \dots & \dots & \dots \\ x_1^{(N)} - x_1^{(0)} & \dots & x_N^{(N)} - x_N^{(0)} \end{bmatrix} \\ B = \begin{bmatrix} f^{(1)} - f^{(0)} \\ \dots \\ f^{(N)} - f^{(0)} \end{bmatrix} \end{cases} \quad (10)$$

The numerical negative gradient direction obtained from (10) has some theoretical errors due to the approximate directional derivative of (8). According to the quadratic function hypothesis, the linear equations can be obtained from (7) and (9).

$$\begin{cases} A\nabla f(\mathbf{x}^{(0)}) = B \\ A = \begin{bmatrix} x_1^{(1)} - x_1^{(0)} & \dots & x_N^{(1)} - x_N^{(0)} \\ \dots & \dots & \dots \\ x_1^{(N)} - x_1^{(0)} & \dots & x_N^{(N)} - x_N^{(0)} \end{bmatrix} \\ B = \begin{bmatrix} -f^{(1)} - 3f^{(0)} + 4f^{(m1)} \\ \dots \\ -f^{(N)} - 3f^{(0)} + 4f^{(mN)} \end{bmatrix} \end{cases} \quad (11)$$

According to the method of taking points above, the coefficient matrix (10) and (11) is a real diagonal matrix, is non-singular, and its rank is  $N$ . It is almost impossible to find the same objective function at each point as that at point  $\mathbf{x}^{(0)}$ , so (10) and (11) have a unique nonzero solution. In the same way, if the first partial derivative of the objective function is known,  $N$  points can be found near the current point  $\mathbf{x}^{(0)}$ , and the above  $N$ -ary linear equations can be obtained, so that the second partial derivative matrix of the objective function can be solved by the least square method.

#### IV. CONSTRUCTION METHOD OF THE NUMERICAL DIRECTIONAL DERIVATIVE

Suppose that the points along the axis of  $x_i$  and  $x_j$  at the current point  $\mathbf{x}^{(0)}$  are taken as shown in Fig. 1,  $o$ ,  $a$ ,  $b$ ,  $c$ ,  $d$ . Referring to (10), eight equations for the second mixed partial derivative  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  can be listed, but there are only four valid equations and five unknowns. For this reason, assume that the difference between the objective function values of point  $c$  and point  $o$  is the smallest, and that the difference between the

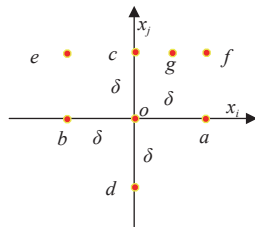


Fig. 1. The point drawing for obtaining mixed partial derivative by numerical differentiation.

objective function values of point  $a$  and point  $o$  is greater than that of point  $b$ . Then, the following assumptions are made

$$\left. \frac{\partial f}{\partial x_i} \right|_c = \left. \frac{\partial f}{\partial x_i} \right|_o.$$

Then the second mixed partial derivative is obtained from the following equation.

$$-\delta \frac{\partial^2 f}{\partial x_i^2} + \delta \frac{\partial^2 f}{\partial x_i \partial x_j} = \left. \frac{\partial f}{\partial x_i} \right|_c - \left. \frac{\partial f}{\partial x_i} \right|_a. \quad (12)$$

According to the quadratic function hypothesis, the relationship between the two coordinates has six undetermined coefficients, and the five points in Fig. 1 can not determine these coefficients, so a numerical point must be added. If we take point  $f$  on the bisector of two coordinate axes, then according to (2) at the midpoint  $g$  of  $cf$ , the partial derivative of  $x_i$  is as follows:

$$\left. \frac{\partial f}{\partial x_i} \right|_g = \frac{f(f) - f(c)}{\delta}.$$

The directional derivative of  $\frac{\partial f}{\partial x_i}$  is obtained in the direction of  $ag$ . For the middle point of  $a$  and  $g$ , the following equation can be obtained.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2} + \left( \left. \frac{\partial f}{\partial x_i} \right|_g - \left. \frac{\partial f}{\partial x_i} \right|_a \right) / \delta. \quad (13)$$

The directional derivative of  $\frac{\partial f}{\partial x_i}$  is obtained in the direction of  $og$ . For the middle point of  $o$  and  $g$ , the following equation can be obtained.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = -\frac{1}{2} \frac{\partial^2 f}{\partial x_i^2} + \left( \left. \frac{\partial f}{\partial x_i} \right|_g - \left. \frac{\partial f}{\partial x_i} \right|_o \right) / \delta. \quad (14)$$

The numerical second mixed partial derivative is derived from the assumption that the objective function is quadratic, so the results of (13) and (14) should be the same.

In fact, the objective function is not necessarily a quadratic function. Both (13) and (14) have some errors with the  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  of point  $o$ . Therefore, it is advisable to calculate the parameter more precisely by taking another symmetry point.

The parameters (3), (4), (13) and (14) are not set according to the specific objective function, but determined, accurate and unchanged. Without the value of theoretical analysis, there is no need to discuss truncation error and rounding error.

#### V. THE CONSTRUCTION METHOD OF A LOCAL ORTHOGONAL COORDINATE SYSTEM FOR OPTIMIZATION

If a local coordinate system is established in the design space, the partial derivative of the objective function can hardly be obtained, but the numerical partial derivative can be obtained using the numerical differential algorithm as in the global coordinate system. After optimizing in a certain direction, it is better to search in the direction of orthogonality again. An orthogonal direction vector group is set up whose last direction is the optimization direction. The orthogonal vector group can be regarded as the coordinate vector group of the local coordinate system. The coordinate vector group is

then the coordinate transformation matrix  $A$  between the two coordinate systems. If the origin of the local coordinate system is  $\mathbf{x}^{(0)}$  in the fixed coordinate system, then a point  $\mathbf{x}'$  in the local coordinate system is shown in the global coordinate system as follows:

$$\mathbf{x} = A\mathbf{x}' + \mathbf{x}^{(0)}. \quad (15)$$

The vector  $\mathbf{s}'$  in the local coordinate system is shown in the global coordinate system as shown below:

$$\mathbf{s} = A\mathbf{s}'. \quad (16)$$

Set the previous optimization direction  $\mathbf{S}$  as the last vector  $\alpha_N$ . Let the largest element of  $\mathbf{S}$  (the element with the largest absolute value) be the element  $i$ . Then in the  $N \times N$  unit matrix, it corresponds to the column vector  $\mathbf{i}$ . In order, the column vector of the unit matrix acts as other vectors; thus we can obtain the following linear independent vectors

$$\alpha_1, \alpha_2, \dots, \alpha_N.$$

The modified Schmitt orthogonalization method is used to perform orthogonalization [28], [29]. We use the following definitions and change vector to unitization vectors with two norm

$$\beta_N = \frac{\alpha_N}{\|\alpha_N\|} \quad (17)$$

$$\beta_{N-1} = \alpha_{N-1} + k\beta_N. \quad (18)$$

Among them,  $k$  is an undetermined coefficient. According to the orthogonality between vectors, the following formula can be obtained. We then change the vectors to unitization vectors with two norm

$$\beta_{N-1} = \alpha_{N-1} - \frac{\alpha_{N-1}^T \beta_N}{\beta_N^T \beta_N} \beta_N \quad (19)$$

$$\beta_{N-1} = \frac{\beta_{N-1}}{\|\beta_{N-1}\|}$$

$$\beta_1 = \alpha_1 - \sum_{i=2}^N \frac{\alpha_1^T \beta_i}{\beta_i^T \beta_i} \beta_i \quad (20)$$

$$\beta_1 = \frac{\beta_1}{\|\beta_1\|}.$$

The square matrix composed of unitized vectors  $\beta_1, \beta_2, \dots, \beta_N$  is the coordinate transformation matrix of (15) and (16).

The orthogonal vector group can also make  $\mathbf{S}$  the last coordinate axis of the local coordinate system after coordinate transformation, so as to obtain the last column elements of  $A$  according to (16). Then, according to the type of coordinate transformation, other elements of  $A$  can be obtained.

For a non-quadratic objective function, the optimization direction obtained from the numerical gradient and the second partial derivative matrix constructed by 1, 2, 3 in different local coordinate systems is different.

## VI. LARGE STEPSIZE NUMERICAL OPTIMIZATION ALGORITHM BASED ON COORDINATE TRANSFORMATION

The second approximate direction directly points to the

extreme point of the quadratic objective function. Such a numerical direction is good. Another new algorithm is as follows:

In order to make full use of the optimization effect obtained by the detection point, after each detection of the best point, the best point is taken as the current point  $\mathbf{x}$ , and the original  $\mathbf{x}$  is taken as the reserve point  $\mathbf{x}^{(i)}$  ( $i$  is the coordinate axis serial number). If neither probe point is good enough, halve the step size and continue to probe until the step size is less than the termination condition value. The next best point is taken as the reserved point  $\mathbf{x}^{(i)}$ , and then, from the current point (the best point) detect along the next coordinate axis. After sequential detection along  $N$  coordinate axes, the numerical negative gradient or numerical second approximation direction is constructed. A step optimization is achieved after one dimensional optimization until the current point  $\mathbf{x}$  is close enough to the reserve point  $\mathbf{x}^{(i)}$ .

Due to the large distance between the detection points, the numerical first partial derivatives and second partial derivatives have large errors. Instead of calculating the mixed partial derivative and one-dimensional optimization in the new direction, the backward half-step searching algorithm [30] based on the blind path-finding idea is used to find better points according to the current step size. The optimization scheme is shown in Fig. 2.

Because of the continuity of the feasible region, the initial point can be chosen arbitrarily. If the step size is too large, the optimization direction may deviate too much from the negative gradient direction and the optimization effect would not be good; if the step size is too small, it causes a sawtooth phenomenon because it is too close to the negative gradient direction. Therefore, the initial step size should be determined according to the specific optimization problem, and a better value can be determined by trial calculation.

## VII. EXAMPLE VERIFICATION

### A. The Two-Dimensional Example

The two-dimensional Rosenbrock function is the four square function, and it is one of the most difficult examples to test the optimization algorithm with. The unconstrained optimization problem whose objective function is as follows:

$$\min f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2. \quad (21)$$

The gradient of the objective function is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 400x_1^3 - 400x_2x_1 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{bmatrix}. \quad (22)$$

The second partial derivative matrix of the objective function is as follows:

$$H_{f(\mathbf{x})} = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}. \quad (23)$$

The objective function consists of two components, each of which has a minimum value of zero. A system of two variables and quadratic equations can be obtained, and the extreme points and extremes can be obtained as follows:

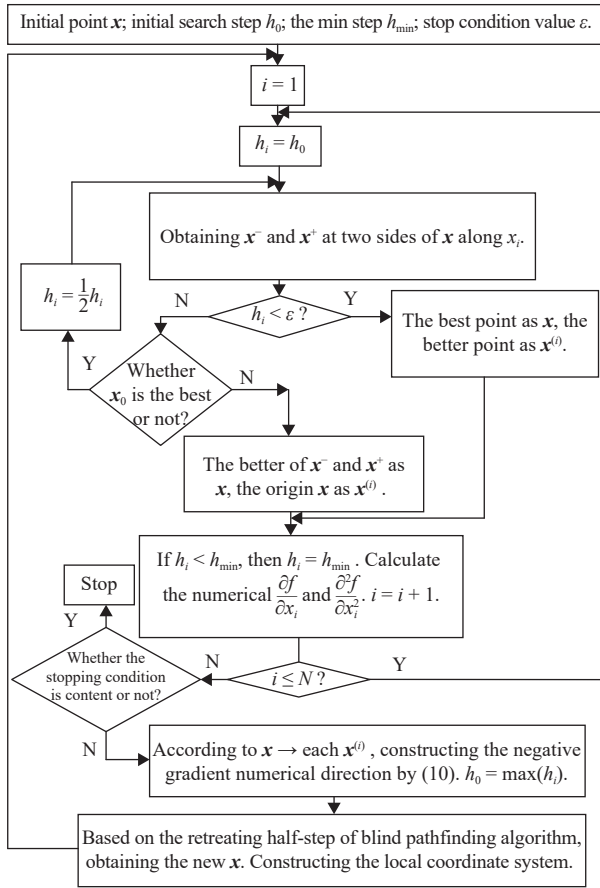


Fig. 2. The program diagram of large step numerical algorithm based on coordinate transformation.

$$\begin{cases} \mathbf{x}^* = \begin{bmatrix} 1 & 1 \end{bmatrix}^T \\ f(\mathbf{x}^*) = 0. \end{cases}$$

Take the initial point and the micro step length as follows:

$$\begin{cases} \mathbf{x}^{(0)} = \begin{bmatrix} -0.4 & 0.6 \end{bmatrix}^T \\ h = 1.0 \times 10^{-5}. \end{cases}$$

The objective function value, numerical gradient and second partial derivative matrix at this point are as follows:

$$\begin{cases} f(\mathbf{x}^{(0)}) = 21.32 \\ \nabla f(\mathbf{x}^{(0)}) = \begin{bmatrix} 67.6 \\ 88 \end{bmatrix} \\ H_f(\mathbf{x}^{(0)}) = \begin{bmatrix} -46 & 160 \\ 160 & 200 \end{bmatrix}. \end{cases}$$

The relative errors between the numerical solutions obtained from (2), (4), (13), and (14) and the analytical solutions are all less than  $10^{-5}$ .

In this paper, the numerical second approximate direction algorithm and the large step numerical algorithm based on coordinate transformation are realized by the C language. The blind searching method [31] is used in one dimensional optimization.

### B. The Test of Numerical Second Approximate Direction Method

The optimal direction is obtained by constructing the numerical diagonal second partial derivative matrix from (4). After one-dimensional optimization is performed 3955 times, the objective function is calculated 156 517 times. The optimal point and optimal solution are obtained as follows:

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 0.99971 & 0.99942 \end{bmatrix}^T \\ f(\mathbf{x}) = 8.2970 \times 10^{-8}. \end{cases}$$

The optimization process is shown in Fig. 3, solid line 1.

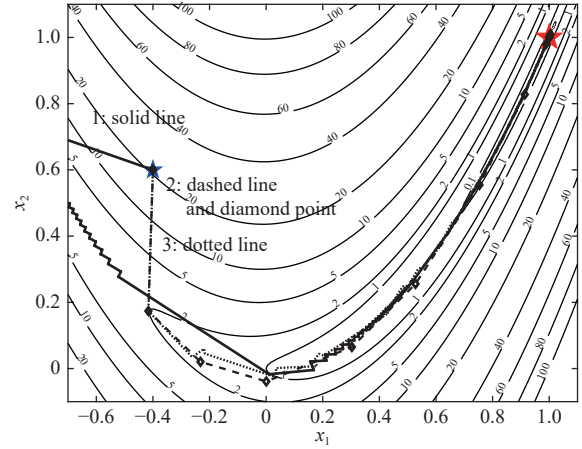


Fig. 3. The seeking process of the numerical second-order approximate direction method.

The optimal direction is obtained by constructing the numerical symmetric second partial derivative matrix from (4), (13), (14). After one-dimensional optimization is performed 11 times, the objective function is calculated 445 times, and the optimal point and solution are obtained as follows:

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 1.00000 & 0.99999 \end{bmatrix}^T \\ f(\mathbf{x}) = 2.49348 \times 10^{-13}. \end{cases}$$

With the numerical second mixed partial derivative, the optimization calculation amount is 0.2843% of without it. The optimization process is shown in the gem-shaped point and dashed line 2 in Fig. 3. The result is similar to that of the analytical second approximate direction method.

A local coordinate system is established for each optimization. After one-dimensional optimization is performed 244 times, the objective function is called 9884 times, and the optimal point and solutions are obtained as follows:

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 1.00000 & 1.00000 \end{bmatrix}^T \\ f(\mathbf{x}) = 4.6722 \times 10^{-10}. \end{cases}$$

Its optimization process is shown in Fig. 3, dotted line 3. Coordinate transformation is not suitable.

### C. The Test of the Constructing Search Direction in Local Coordinate System

In order to test the influence of coordinate transformation on

the numerical optimization direction, a local coordinate system is established with  $\mathbf{x}^{(0)}$  as the origin, and the local coordinate system is rotated counterclockwise  $360^\circ$ .

The numerical direction derivative of the fixed coordinate system is used as the numerical partial derivative of the local coordinate system to construct the optimization direction. In the above rotation process, the trajectory of the numerical optimization direction in the local coordinate system is shown in the plus point and dashed line 1 in Fig. 4. It is not evenly distributed, indicating that it is not a fixed direction. In the fixed coordinate system, the trajectory is not a fixed direction, but varies within the range  $103.05^\circ$ , as shown by the asterisk point and dotted line 2 in Fig. 4. It is clear that the calculated results are consistent with those of the 5 analyses.

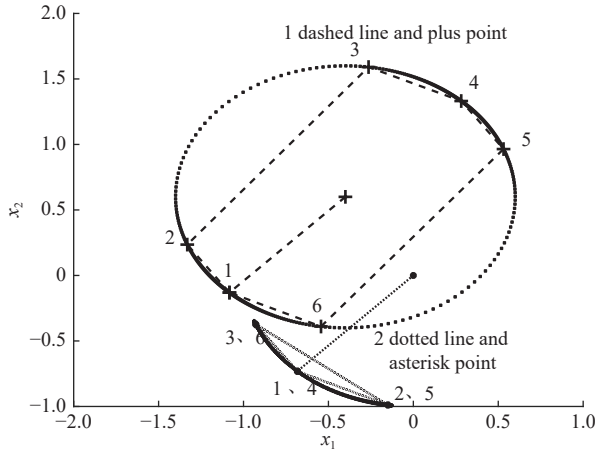


Fig. 4. Numerical direction trajectories in the rotation process of local coordinate system.

#### D. Test of the Large Step Numerical Optimization Algorithm

The initial step size, the minimum initial step length, and the termination condition value are as follows:

$$\begin{cases} h_0 = 0.1 \\ h_{\min} = 10^{-2} \\ \varepsilon = 10^{-8} \end{cases}$$

In the first optimization round, the following points are detected along the  $N$  axis:

$$\mathbf{x} = \begin{bmatrix} -0.5 & 0.5 \end{bmatrix}^T.$$

According to (27), the analytic gradient of the objective function at this point is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 47 & 50 \end{bmatrix}^T.$$

The numerical gradient obtained by (9) is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 68.2 & 60.0 \end{bmatrix}^T.$$

The difference between them is quite large. Then we obtain the  $N$  (2) midpoint, and the numerical gradient obtained by (11) is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 48.925 & 50.000 \end{bmatrix}^T.$$

Using (9), after 1437 detection rounds, the objective

function is called 89 155 times, and the optimal point and solutions are obtained as follows:

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 0.99999 & 0.99999 \end{bmatrix}^T \\ f(\mathbf{x}) = 2.7059 \times 10^{-11}. \end{cases}$$

The optimization process is shown in Fig. 5, dotted line 1. Adopting (11), the objective function must be called 87 065 times in 1726 detection rounds.

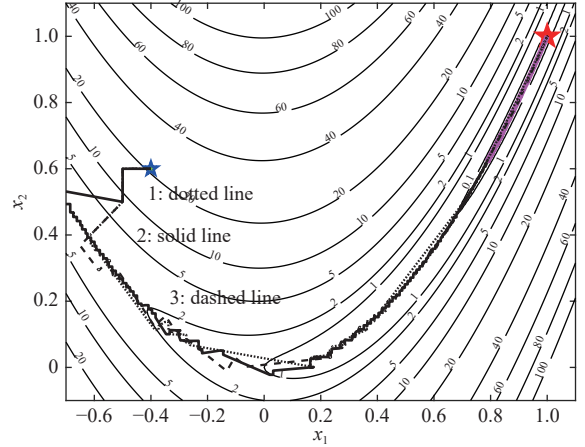


Fig. 5. The search process of the large step numerical algorithm based on coordinate transformation.

At the current point  $\begin{bmatrix} -0.5 & 0.5 \end{bmatrix}^T$ , the numerical gradient obtained by (1)–(3) is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 70.4 & 50.0 \end{bmatrix}^T.$$

Then the numerical optimization direction of (4) and (21) is as follows:

$$\mathbf{s} = \begin{bmatrix} -14080 & 2200 \end{bmatrix}^T.$$

After 1430 detection rounds, we call the objective function 70 621 times to get the optimal point and solution.

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 1.00000 & 1.00000 \end{bmatrix}^T \\ f(\mathbf{x}) = 6.4038 \times 10^{-12}. \end{cases}$$

The optimization process is shown in Fig. 5, solid line 2.

The numerical gradient is obtained by means of 5 coordinate transformation and (9). After 144 detection rounds, we call the objective function 4717 times to get the optimal point and solution

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 1.00000 & 1.00000 \end{bmatrix}^T \\ f(\mathbf{x}) = 2.1238 \times 10^{-10}. \end{cases}$$

The optimization process is shown in the dashed line 3 of Fig. 5. Compared with the black dotted line, the computation amount is reduced by 94.71%. Due to the large step size, the numerical negative gradient direction obtained by (9) has a large error. If the numerical direction is rotated  $90^\circ$ , after 243 rounds of detection, the objective function is called 7872 times, and the optimal point and solution are obtained as

follows:

$$\begin{cases} \mathbf{x} = \begin{bmatrix} 1.0000 & 1.0000 \end{bmatrix}^T \\ f(\mathbf{x}) = 2.7255 \times 10^{-13}. \end{cases}$$

The calculation amount is not large, which is consistent with the analysis of 7.3. In the process of approaching, the extremum point is skipped over for several times, but not arrived at step by step. After the numerical negative gradient direction is constructed by (11) and compared with the dotted line 1, the optimal solution is  $2.4764 \times 10^{-15}$  after 14 975 detection rounds and 944 941 calls to the objective function. There is a very serious zigzag phenomenon because it is too close to the analytic negative gradient direction.

#### E. The Three-Dimensional Example

The three-dimensional Rosenbrock function is one of the most difficult multi-dimensional examples to test the optimization algorithm with. The unconstrained optimization problem whose objective function is as follows:

$$\min f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 100(x_3 - x_2^2)^2 + (1 - x_2)^2. \quad (24)$$

The gradient of the objective function is as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) - 400x_2(x_3 - x_2^2) - 2(1 - x_2) \\ 200(x_3 - x_2^2) \end{bmatrix}. \quad (25)$$

The second partial derivative matrix of the objective function is as follows:

$$H_{f(\mathbf{x})} = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 & 0 \\ -400x_1 & 200 + 1200x_2^2 - 400x_3 + 2 & -400x_2 \\ 0 & -400x_2 & 200 \end{bmatrix}. \quad (26)$$

The objective function consists of 4 components, each of which has a minimum value of 0. A system of three variable quadratic equations can be obtained, and the extreme points and extremes can be obtained as follows:

we take the initial point as follows:

$$\begin{cases} \mathbf{x}^{(0)} = \begin{bmatrix} -0.5 & 0.5 & 0.5 \end{bmatrix}^T \\ f(\mathbf{x}^{(0)}) = 15. \end{cases} \quad (27)$$

Adopting the analytic 2 order approximation direction, after 16 rounds of one-dimensional optimization, the objective function is calculated 614 times, and the optimal point and solution are obtained as follows:

$$\begin{cases} \mathbf{x}^{(*)} = \begin{bmatrix} 1.0001 & 1.0002 & 1.0003 \end{bmatrix}^T \\ f(\mathbf{x}^{(*)}) = 2.3394 \times 10^{-7}. \end{cases} \quad (28)$$

Adopting the numerical two order approximation direction, if we do not calculate the numerical mixed partial derivative, after 446 rounds of one-dimensional optimization, the objective function is calculated 20 038 times, and the optimal point and solution are obtained as follows:

$$\begin{cases} \mathbf{x}^{(*)} = \begin{bmatrix} 0.96740 & 0.93554 & 0.87553 \end{bmatrix}^T \\ f(\mathbf{x}^{(*)}) = 5.2367 \times 10^{-3}. \end{cases} \quad (29)$$

Using (13) or (14) to calculate the numerical mixed partial derivative, after 16 rounds of one-dimensional optimization, the objective function is calculated 614 times, and the optimal point and solution are obtained as follows:

$$\begin{cases} \mathbf{x}^{(*)} = \begin{bmatrix} 1.0002 & 1.0002 & 1.0004 \end{bmatrix}^T \\ f(\mathbf{x}^{(*)}) = 1.5794 \times 10^{-6}. \end{cases} \quad (30)$$

Taking more than one point for each mixed partial derivative, it is then computed by using (13) or (14). After 16 rounds of one-dimensional optimization, the objective function is calculated 660 times, and the optimal point and solution are obtained as follows:

$$\begin{cases} \mathbf{x}^{(*)} = \begin{bmatrix} 1.0001 & 1.0002 & 1.0003 \end{bmatrix}^T, \\ f(\mathbf{x}^{(*)}) = 2.9050 \times 10^{-7}. \end{cases} \quad (31)$$

#### VIII. DISCUSSION

The mathematical deduction in this paper is concise and easy to understand. A lot of tedious deduction processes are omitted. The key derivation steps and principles are explained in the text and it is similar to many classical numerical differentiation formulas (difference formula).

The purpose of this paper is to broaden the scope of existing optimization methods. It should test the approximation between numerical and analytical solutions. If the two approaches are close enough, the optimization effect of the original optimization method will not be affected. The Rosenbrock function is four square, with a curved canyon. At the bottom of the canyon, the objective function contour is close to the parallel, and there are no characteristics of the quadratic function at all. The optimization methods with the optimal direction are all based on the quadratic objective function hypothesis. Some optimization algorithms cannot obtain the effective direction at the bottom of the valley. Therefore, this example is one of the most difficult examples to test the optimization algorithm with. The successful verification of the numerical example is sufficient to prove the optimization effect of numerical optimization direction.

Conjugate directions need  $N$  to expand enough in optimization space, and more than three-dimensional examples are needed to verify the optimization effect. The research in this paper does not involve the verification of conjugate directions. Two-dimensional examples can not only illustrate the problem, but also visually represent the optimization process. Even if the numerical algorithm of the conjugate direction is studied, a three-dimensional numerical example is not necessary, because if the analytical direction is not good, the numerical direction will not be good.

For multimodal objective functions, different initial points may have different optimization results. For a single peak objective function, the optimization algorithm with a certain



direction will not yield different optimal points because of the different initial points. The initial point of the example in this paper is taken on a flat slope on the side of the canyon, which is far from the extreme value point. It not only checks the speed of searching to the canyon bottom, but also checks the effect of approaching the extreme value point at the canyon bottom.

In this paper, we compare different algorithms using computation and optimization results, and we show that it is not necessary to use an elegant intelligent algorithm. This algorithm is not comparable with an evolutionary algorithm or swarm intelligence optimization algorithm. The latter is an optimization method based on other theories, and does not require gradient information such as in the Powell method [32]. However, there is no concept of optimizing direction; it is an algorithm that seeks order in disorder; it is an algorithm that explores necessity in occasionality. The algorithm has obvious advantages for multi-modal feasible regions, but for single-modal feasible regions, it is not as effective as an algorithm with a definite direction. The numerical direction constructed by the numerical method is not more accurate than that obtained by the analytical method, and both are suitable for a single-peak feasible region.

#### IX. CONCLUSION

Through the verification of several numerical algorithms with two-dimensional Rosenbrock objective function optimization, the following conclusions can be drawn:

1) The numerical formulas for calculating the first, second and second mixed partial derivatives have higher accuracy. The numerical first partial derivatives are in agreement with the results obtained by the method of constructing Lagrange interpolation functions. Thus, these numerical formulas are correct. Because it is difficult to obtain second partial derivatives, research in optimization methods that only need the gradient of objective function have been explored. This paper continues the research in this area. At present, there are some research directions involving using numerical direction to improve the accuracy of optimization. The research based on the quadratic hypothesis also ends this direction.

2) The numerical second approximate direction method with numerical second mixed partial derivatives is the best.

3) The optimization direction obtained in a local coordinate system has strong randomness, and the large step numerical algorithm requires it.

4) Based on the quadratic hypothesis and fitting method, the numerical algorithm proposed in this paper allows the derivatives, gradients, second partial derivatives and second mixed partial derivatives of the objective function to remain unresolved, which makes optimization methods that require this information more applicable. There is no derivable objective function for optimization problems in many engineering fields. This study provides more optimization methods for solving these optimization problems.

5) Because constrained optimization problems are often encountered in control processes, future research on constrained optimization methods may become the focus of our research.

#### REFERENCES

- [1] H. Mu, Z. Li, W. Huo, and Z. Jin, "Hash map optimization and its application in column-oriented database query," *Chinese J. Frontiers of Computer Science and Technology*, vol. 10, no. 9, pp. 1250–1261, 2016.
- [2] Z. Jin, H. Yang, and G. Yin, "A numerical approach to optimal dividend policies with capital injections and transaction costs," *Acta Mathematicae Applicatae Sinica (English Serie)*, vol. 33, no. 1, pp. 221–238, 2017.
- [3] J. Qian, D. Wang, X. Miao, and G. Su, "Numerical simulation and optimization of automobile outflow field based on CFD," *Chinese Manufacturing Automation*, vol. 38, no. 4, pp. 74–76, 2016.
- [4] J. Lin, L. Guan, J. Han, and M. Lu, "Design and optimization of the parallel elliptical vibration cutting mechanism," *Chinese J. Machine Design*, vol. 35, no. 5, pp. 53–59, 2018.
- [5] J. Shi, L. Wang, C. Luo, G. Cai, C. Shen, and Z. Zhu, "Sparse representation for gearbox compound fault feature extraction based on majorization-minimization algorithm," *Chinese J. Vibration Engineering*, vol. 30, no. 6, pp. 1045–1055, 2017.
- [6] S. Zou, I. Kokyu, I. Norio, and F. Yuji, "Design optimization of friction damper with coupling mechanism for seismic response of base isolated structure," *Chinese J. Vibration Engineering*, vol. 29, no. 2, pp. 201–206, 2016.
- [7] Y. Hu and X. Liu, "Optimization of grouping evacuation strategy in high-rise building fires based on graph theory and computational experiments," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 6, pp. 1104–1112, 2018.
- [8] X. Bi, L. Zhang, and J. Xiao, "Constrained multi-objective optimization algorithm based on dual populations," *Chinese J. Computer Research and Development*, vol. 52, no. 12, pp. 2813–2823, 2015.
- [9] X. Guo, L. Lu, and C. Zhu, "Two phase many-objective optimization algorithm based on Pareto dominance relationship," *Chinese J. Frontiers of Computer Science and Technology*, vol. 12, no. 8, pp. 1350–1360, 2018.
- [10] J. Li, H. L. i, and X. Zhao, "Multi-objective optimization of hybrid electrical vehicle based on immune genetic algorithm," *Computer Engineering and Applications in China*, vol. 54, no. 4, pp. 237–262, 2018.
- [11] L. Wang, L. Qiao, and L. Wei, "Optimal convolutional neural networks learning method combined with genetic algorithm," *Computer Engineering and Design in China*, vol. 38, no. 7, pp. 1945–1950, 2017.
- [12] J. Xie, S. Su, and J. Wang, "Search strategy of artificial bee colony algorithm guided by approximate gradient," *Chinese J. Frontiers of Computer Science and Technology*, vol. 10, no. 12, pp. 1773–1782, 2016.
- [13] Y. Ke and C. Ma, "A preconditioner for elliptic PDE-constrained optimization problems," *Mathematic Numerica Sinica in China*, vol. 39, no. 1, pp. 70–80, 2017.
- [14] C. L. Zhang, "The improvement of Newton method and the validation of optimization idea of blind walking repeatedly," *Advanced Materials Research*, vol. 1270, no. 503, pp. 4061–4064, 2011.
- [15] C. Li, H. Wang, W. Xu, and X. Li, "Multidimensional two-time fitting function optimization method," *J. Gansu Sciences*, vol. 29, no. 5, pp. 26–28, 2017.
- [16] Q. Liu, X. Liu, X. Yin, and C. Li, "Conjugate method of adjacent directions based on gradient vector of objective function," *J. Gansu Sciences*, vol. 29, no. 5, pp. 15–21, 2017.
- [17] X. Jiang and J. Jian, "A self-adjusting Polak-Ribiere-Polyak type conjugate gradient method," *Acta Mathematicae Applicatae Sinica*, vol. 40, no. 3, pp. 449–460, 2017.
- [18] X. L. Yin, F. Sun, and C. M. Li, "Six search optimization method on obtaining conjugate direction after continuous negative gradient directions," *Chinese J. Frontiers of Computer Science and Technology*, vol. 13, no. 9, pp. 1604–1612, 2019.
- [19] X. L. Yin, F. Sun, and C. M. Li, "Half-step constrained mechanical



- optimization method based on conjugate direction,” *J. Gansu Sciences*, vol. 30, no. 5, pp. 29–36, 2018.
- [20] J. Chen and L. Zhang, “Numerical approximation of solution for the coupled nonlinear Schrödinger equations,” *Acta Mathematicae Applicatae Sinica (English Serie)*, vol. 33, no. 2, pp. 435–450, 2017.
- [21] X. Li, Q. Han, J. Wang, Q. Wang, and X. Duan, “ERT image reconstruction based on improved CG method,” *Chinese J. Scientific Instrument*, vol. 37, no. 7, pp. 1673–1679, 2016.
- [22] J. Li, X. Lv, T. Tao, and M. Lai, “Conditional gradient algorithm for solving operator equation least squares problem under the bound constraints,” *Mathematica Numerica Sinica*, vol. 38, no. 4, pp. 372–390, 2016.
- [23] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, 1983. DOI: 10.1016/0377-2217(85)90293-0.
- [24] O. A. Arqub and Z. Abo-Hammour, “Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm,” *Information Sciences*, vol. 2014, no. 279, pp. 396–415, 2014.
- [25] T. F. Coleman and J. G. More, “Estimation of sparse Jacobean matrices and graph coloring problems”, Report No. ANL-81-39, Argonne National Laboratory, Argonne, Ill., 1981.
- [26] P. N. Brown and Y. Saad, “Hybrid Krylov methods for nonlinear systems of equations,” *SIAM J. Scientific and Statistical Computing*, vol. 11, no. 3, pp. 450–481, 1990.
- [27] W. Cheney and D. Kincaid, *Linear Algebra: Theory and Applications*, Sudbury, MA: Jones and Bartlett. 2009, pp. 544–558. ISBN 978-0-7637-5020-6.
- [28] L. Pursell and S. Y. Trimble, “Gram-Schmidt orthogonalization by Gauss elimination,” *The American Mathematical Monthly*, vol. 98, no. 6, pp. 544–549, 1991.
- [29] E. D. Solomentsev, *Euclidean Space*, Encyclopedia of Mathematics, 2011, Springer Retrieved, 2014.
- [30] C. Li, “Negative gradient direction method for blind person exploring the way,” *J. Gansu Sciences*, vol. 28, no. 5, pp. 116–122, 2016.
- [31] C. Li, “Blind-walking optimization method,” *J. Networks*, vol. 5, no. 12, pp. 1458–1466, 2010.
- [32] X. Liu, X. Yin, and C. Li, “On the improvement of the Powell mechanical optimization method,” *Chinese J. Machine Design*, vol. 36, no. 6, pp. 80–86, 2019.



**Xiaoli Yin** is a Ph.D. candidate at China University of Petroleum (East China). She is currently an Associate Professor at Shengli College in China University of Petroleum (East China). Her research interest is optimization design method of mechanical oil production systems.



**Chunming Li** received the Ph.D. degree in flight vehicle design from Nanjing University of Aeronautics & Astronautics in 2001. He is currently a Professor at Shengli College in China University of Petroleum (East China). His research interests include optimization methods, mechanical design and theory, and mechanical kinetics.



**Yuan Zhang** received the Ph. D. degree in structural engineering from Tianjin University in 2007. She is currently a Professor at Shengli College in China University of Petroleum (East China). Her research interests include structural control, vehicle-bridge coupled vibration, and damage detection.