Sub-Direction Parallel Search Quasi-Newton Algorithm

Wang Bao and Wang Yanxin

***Abstract*—Advantages and disadvantages in a sequence of unconstrained optimization method are basically compared. The original constrained problem is replaced by a sequence of unconstrained sub-problems through the augmented Lagrangian multiplier method. The unconstrained sub-problems are solved by BFGS method and the sub-direction parallel search quasi-Newton algorithm. Efficiency of this method is compared. The results of numerical tests show that the calculation time of the sub-direction parallel search quasi-Newton algorithm is short and it can solve engineering optimization problems completely.**

***Index Terms*—Quasi-Newton equation, taylor series, global convergence, iteration formula.**

I. Introduction

The basic technical way to improve the effectiveness and efficiency of solving HEB optimization problem is to modify the calculation strategy of Hessian matrixes. The basic technical framework of quasi-Newton method is to change the high-dimensional function Hessian matrix by accurate calculation into a quasi-hessian matrix which has good approximate properties and moderate improvement of non-singularity. Based on the algorithm principle of standard quasi-Newton equation, this paper uses the fourth-order tensor model to construct a new quasi-Newton equation, which not only retains the property of constructing the symmetric positive qualitative correction matrix according to standard quasi-Newton equation, but also makes full use of function value information, and performs better than standard quasi-Newton equation in the process of optimization search. In this study, the new quasi-Newton equation is used to give the algorithm construction of four correction schemes, which can approach a Hessian matrix or its inverse matrix of high dimensional function more accurately, and also the proposed BFGS-T correction construction proves that the algorithm can guarantee the convergence of global optimization. Typical examples show the effectiveness of the proposed quasi-Newton construction scheme.

II. New quasi-Newton equation

The standard quasi-Newton equation, proposed by David on [1], is a lower convex quadratic polynomial function at the current point $\boldsymbol{X}_k$.

$$m_k(\boldsymbol{X}) = f_k + (\boldsymbol{X}-\boldsymbol{X}_k)^{\mathrm{T}}\boldsymbol{g}_k + (\boldsymbol{X}-\boldsymbol{X}_k)^{\mathrm{T}}\boldsymbol{B}_k(\boldsymbol{X}-\boldsymbol{X}_k)$$

to approximate the original objective function $f(\boldsymbol{X})$. In this case, f_k is the $f(\boldsymbol{X})$ value at the point X_k; $\boldsymbol{g}_k$ is the $f(\boldsymbol{X})$ gradient of the point $\boldsymbol{X}_k$; $\boldsymbol{B}_k$ is a symmetrical positive definite matrix which approaches the Hessian matrix, $\nabla^2 f(\boldsymbol{X}_k)$. In order to construct an approximate matrix, $\boldsymbol{B}_k$, we need to satisfy the condition of the Hessian matrix, $\nabla^2 f(\boldsymbol{X}_k)$. It can be obtained by the mean value theorem of multivariate differential

$$\boldsymbol{g}_{k+1} - \boldsymbol{g}_k = \left\lfloor \int_0^1 \nabla^2 f(\boldsymbol{X}_k + \theta \boldsymbol{s}_k) d\theta \right\rfloor \boldsymbol{s}_k \qquad (1)$$

where, $\boldsymbol{s}_k = \boldsymbol{X}_{k+1} - \boldsymbol{X}_k$.

In order to make $\boldsymbol{B}_{k+1}$ approach $\nabla^2 f(\boldsymbol{X}_{k+1})$,

let: $\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$, quasi-Newton equation can be obtained by

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k \qquad (2)$$

If $\boldsymbol{H}_{k+1}$ approximates $\left[\nabla^2 f(\boldsymbol{X}_{k+1})\right]^{-1}$, it can be obtained by

$$\boldsymbol{s}_k = \boldsymbol{H}_{k+1}\boldsymbol{y}_k \qquad (3)$$

By constructing a matrix $\boldsymbol{H}_{k+1}$ (or $\boldsymbol{B}_{k+1}$) that satisfies the quasi Newton equation (3) (or (2)), some correction formulas can be obtained.

Rank-correction formula

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{y}_k)^{\mathrm{T}}(\boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k)}$$

DFP correction formula

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{s}_k(\boldsymbol{s}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k)^{\mathrm{T}}\boldsymbol{y}_k} - \frac{\boldsymbol{H}_k\boldsymbol{y}_k(\boldsymbol{H}_k\boldsymbol{y}_k)^{\mathrm{T}}}{(\boldsymbol{y}_k)^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k}$$

BFGS correction formula

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \left(1 + \frac{\boldsymbol{y}_k\boldsymbol{H}_k\boldsymbol{y}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}\right)\frac{\boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k} - \frac{\boldsymbol{s}_k\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{H}_k + \boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{s}_k^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}$$

The quasi Newton algorithm produced by the standard quasi Newton equation can guarantee the n step iteration to converge [2] to the objective function of the lower convex quadratic function, but it only takes advantage of the gradient information and ignores the information of the function value, and the approximation order is the second order. The

Manuscript received December 5, 2017; revised January 5. 2018.

The authors are with Ningbo University of Technology, School of Science, China (e-mail: 864787886@ qq.com).

DOI: 10.7763/IJMO.2018.V8.637

comparison of equations (1) and (2) can see that $\boldsymbol{B}_{k+1}$ is only an approximation of the "mean value" between $\boldsymbol{X}_k$ and $\boldsymbol{X}_{k+1}$ at $\nabla^2 f(\boldsymbol{X})$. Therefore, it is reasonable to make $\boldsymbol{B}_{k+1}$ a more accurate approximation $\nabla^2 f(\boldsymbol{X})$. In recent years, a lot of research work has been carried out in this field. Wei and others [3] use the third order Taylor expansion of the target function to put forward the following quasi Newton equation:

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \tilde{\boldsymbol{y}}_k, \quad \tilde{\boldsymbol{y}}_k = \boldsymbol{y}_k + \frac{\tilde{\theta}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}}\boldsymbol{u} \tag{4}$$

Among them, $\boldsymbol{u} \in R^n$ makes $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u} \neq 0$ an arbitrary vector.

$$\tilde{\theta}_k = 2(f_k - f_{k+1}) + (\boldsymbol{g}_k + \boldsymbol{g}_{k+1})^{\mathrm{T}}\boldsymbol{s}_k$$

Fahimeh, Biglari and etc.[4], use the derivative and the value of the objective function to propose the following quasi Newton equation

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \tilde{\tilde{\boldsymbol{y}}}_k, \quad \tilde{\tilde{\boldsymbol{y}}}_k = \boldsymbol{y}_k + \frac{\tilde{\tilde{\theta}}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}}\boldsymbol{u} \tag{5}$$

Among them, $\boldsymbol{u} \in R^n$ makes $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u} \neq 0$ an arbitrary vector. $\tilde{\tilde{\theta}}_k = 4(f_k - f_{k+1}) + 2(\boldsymbol{g}_k + \boldsymbol{g}_{k+1})^{\mathrm{T}}\boldsymbol{s}_k$. Zhang and etc[5]. proposed the following quasi-Newton equation

$$\boldsymbol{B}_{k+1}s_k = \tilde{\tilde{\tilde{\boldsymbol{y}}}}_k, \quad \tilde{\tilde{\tilde{\boldsymbol{y}}}}_k = \boldsymbol{y}_k + \frac{\tilde{\tilde{\tilde{\theta}}}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}}\boldsymbol{u} \tag{6}$$

Among them, $\boldsymbol{u} \in R^n$ makes $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u} \neq 0$ an arbitrary vector. $\tilde{\tilde{\tilde{\theta}}}_k = 6(f_k - f_{k+1}) + 3(\boldsymbol{g}_k + \boldsymbol{g}_{k+1})^{\mathrm{T}}\boldsymbol{s}_k$

In order to unify (2) and (6) format, Hiroshi and Yabe, etc[6]., introduced a parameter to form the following quasi Newton equation.

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \tilde{\tilde{\tilde{\boldsymbol{y}}}}_k, \quad \tilde{\tilde{\tilde{\boldsymbol{y}}}}_k = \boldsymbol{y}_k + \rho_k\frac{\tilde{\tilde{\tilde{\theta}}}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}}\boldsymbol{u} \tag{7}$$

When ρ_k =0 or 1, it is corresponding to the quasi Newton equation (2) or quasi Newton equation (6).

These quasi-Newton equations not only take advantage of derivative information, but also make use of the information of function values, which is more accurate in theory than the standard quasi-Newton equation approximating the Hessian matrix of the real objective function.

For a higher-order approximation matrix, Hessian, this paper gives a more accurate approximation formula, $\nabla^2 f(\boldsymbol{X}_{k+1})\boldsymbol{s}_k$. A new quasi-Newton equation is established by use of the objective function value and derivative information. Assume that the objective function $f(\boldsymbol{X})$ is sufficiently smooth. According to the Taylor formula, along $-\boldsymbol{s}_k$ at the point $\boldsymbol{X}_{k+1}$, $f(\boldsymbol{X})$, $\boldsymbol{g}_k \otimes \boldsymbol{s}_k$ and $\boldsymbol{G}_k \otimes \boldsymbol{s}_k^2$ will be expanded and we can obtain :

$$f_k = f_{k+1} - \boldsymbol{g}_{k+1} \otimes \boldsymbol{s}_k + \frac{1}{2}\boldsymbol{G}_{k+1} \otimes \boldsymbol{s}_k^2 - \frac{1}{6}\boldsymbol{T}_{k+1} \otimes \boldsymbol{s}_k^3 + \frac{1}{24}\boldsymbol{V}_{k+1} \otimes \boldsymbol{s}_k^4 + O(\|\boldsymbol{s}_k\|^5)$$

$$\boldsymbol{g}_k \otimes \boldsymbol{s}_k = \boldsymbol{g}_{k+1} \otimes \boldsymbol{s}_k - \boldsymbol{G}_{k+1} \otimes \boldsymbol{s}_k^2 + \frac{1}{2}\boldsymbol{T}_{k+1} \otimes \boldsymbol{s}_k^3 - \frac{1}{6}\boldsymbol{V}_{k+1} \otimes \boldsymbol{s}_k^4 + O(\|\boldsymbol{s}_k\|^5)$$

$$\boldsymbol{G}_k \otimes \boldsymbol{s}_k^2 = \boldsymbol{G}_{k+1} \otimes \boldsymbol{s}_k^2 - \boldsymbol{T}_{k+1} \otimes \boldsymbol{s}_k^3 + \frac{1}{2}\boldsymbol{V}_{k+1} \otimes \boldsymbol{s}_k^4 + O(\|\boldsymbol{s}_k\|^5)$$

where $\otimes$ is the proper tensor product; $\boldsymbol{T}_{k+1}$ and $\boldsymbol{V}_{k+1}$ express the third and fourth order tensors, respectively; $\boldsymbol{G}_{k+1}$ denotes $\nabla^2 f(\boldsymbol{X}_{k+1})$; $\boldsymbol{G}_k \otimes \boldsymbol{s}_k^2$ denotes $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{G}_k\boldsymbol{s}_k$.

Eliminate the including $\boldsymbol{T}_{k+1}$ and $\boldsymbol{V}_{k+1}$ terms, and derive from the above three equations [6]:

$$\boldsymbol{G}_{k+1} \otimes \boldsymbol{s}_k^2 = \boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k + 12(f_k - f_{k+1}) + 5\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k + 7\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k + \boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{G}_k\boldsymbol{s}_k + O(\|\boldsymbol{s}_k\|^5) \tag{8}$$

In order to get $\boldsymbol{B}_{k+1}$ close to $\nabla^2 f(\boldsymbol{X}_{k+1})$, there are reasons to let $\boldsymbol{B}_{k+1}$ satisfy:

$$\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k + 12(f_k - f_{k+1}) + 5\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k + 7\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k + \boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k \tag{9}$$

In order to let $\boldsymbol{B}_{k+1}\boldsymbol{s}_k$ satisfy (8), the following form is one of the choices

$$\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \hat{\boldsymbol{y}}_k \quad \hat{\boldsymbol{y}}_k = \boldsymbol{y}_k + \frac{\hat{\theta}_k\boldsymbol{u}}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u}} \tag{10}$$

Among them, $\boldsymbol{u} \in R^n$ makes $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{u} \neq 0$ an arbitrary vector.

$$\hat{\theta}_k = 12(f_k - f_{k+1}) + 5\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k + 7\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k + \boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k$$

If $\boldsymbol{H}_{k+1}$ approximates the inverse of $\nabla^2 f(\boldsymbol{X}_{k+1})$, then there

$$\boldsymbol{H}_{k+1}\hat{\boldsymbol{y}}_k = \boldsymbol{s}_k$$

It is not hard to find when $k \to 0$, $\hat{\theta}_k \to 0$. New Quasi Newton Equation (10) approximates the Standard Quasi - Newton Equation (2). Among (10), in terms of the different selection of $\boldsymbol{u}$, the different Quasi-Newton equation can be obtained. Generally, take $\boldsymbol{u} = \boldsymbol{s}_k$, $\boldsymbol{B}_k\boldsymbol{s}_k$ (= $-\boldsymbol{g}_k$), or $-\boldsymbol{g}_{k+1}$ (if $\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{g}_{k+1} \neq 0$), theorems and numerical analyses show that $\boldsymbol{u} = \boldsymbol{y}_k$ is a quite good choice [7].

When some classical correction formulas are applied to this new quasi-Newton equation, a new correction formula is constructed. Because the new quasi-Newton equation formally replaces $\boldsymbol{y}_k$ to $\hat{\boldsymbol{y}}_k$ in the classical quasi-Newton equation, so the new calibration formula can be obtained by simply replacing the classical calibration formula.

The new quasi-Newton equation is applied to the BFGS correction formula, the following equation can be obtained

$$\boldsymbol{B}_{k+1}^{BFGS-T} = \left(\boldsymbol{I} - \frac{\hat{\boldsymbol{y}}_k \boldsymbol{s}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k}\right) \boldsymbol{B}_k^{BFGS-T} \left(\boldsymbol{I} - \frac{\boldsymbol{s}_k \hat{\boldsymbol{y}}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k}\right) + \frac{\hat{\boldsymbol{y}}_k \hat{\boldsymbol{y}}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k} \tag{11}$$

It is marked as BFGS-T correction and applied to inverse BFGS correction formula

$$\boldsymbol{H}_{k+1}^{BFGS-T} = \boldsymbol{H}_k^{BFGS-T} + \left(1 + \frac{\hat{\boldsymbol{y}}_k \boldsymbol{H}_k^{BFGS-T} \hat{\boldsymbol{y}}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k}\right) \frac{\boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k} - \frac{\boldsymbol{s}_k \hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{H}_k^{BFGS-T} + \boldsymbol{H}_k^{BFGS-T} \hat{\boldsymbol{y}}_k \boldsymbol{s}_k^{\mathrm{T}}}{\hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{s}_k} \tag{12}$$

It is recorded as BFGS-T inverse correction, and $\boldsymbol{H}_{k+1}$ is the approximation of $[\nabla^2 f(\boldsymbol{X}_{k+1})]^{-1}$, which is applied to the symmetry rank as a correction formula (SR1). We obtain

$$\boldsymbol{H}_k^{SR1-T} = \boldsymbol{H}_k^{SR1} + \frac{(\boldsymbol{s}_k - \boldsymbol{H}_k^{SR1-\mathrm{T}} \hat{\boldsymbol{y}}_k)(\boldsymbol{s}_k - \boldsymbol{H}_k^{SR1-T} \hat{\boldsymbol{y}}_k)^{\mathrm{T}}}{(\boldsymbol{s}_k - \boldsymbol{H}_k^{SR1-T} \hat{\boldsymbol{y}}_k)^{\mathrm{T}} \hat{\boldsymbol{y}}_k} \tag{13}$$

It is recorded as SR1-T correction and applied to the DFP correction formula. We obtain

$$\boldsymbol{H}_{k+1}^{DFP-T} = \boldsymbol{H}_k^{DFP-T} + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}}}{\boldsymbol{s}_k^{\mathrm{T}} \hat{\boldsymbol{y}}_k} - \frac{\boldsymbol{H}_k^{DFP-T} \hat{\boldsymbol{y}}_k \hat{\boldsymbol{y}}_k^{\mathrm{T}} \boldsymbol{H}_k^{DFP-T}}{\hat{\boldsymbol{y}}_k^{T} \boldsymbol{H}_k^{DFP-T} \hat{\boldsymbol{y}}_k} \tag{14}$$

It is recorded as DFP-T correction [7].

Both DFP-T and BFGS-T corrections are symmetric rank second corrections, and combining the weighted combination of DFP-T and BFGS-T, we obtain

$$\boldsymbol{H}_{k+1}^{\phi} = (1-\phi)\boldsymbol{H}_{k+1}^{DFP-T} + \phi \boldsymbol{H}_{k+1}^{BFGS-T} \tag{15}$$

where ϕ is a parameter; $\boldsymbol{H}_{k+1}^{DFP-T}$ is obtained from the correction formula (14); $\boldsymbol{H}_{k+1}^{BFGS-T}$ is obtained from the correction formula (14), and the corresponding equation of the above formula is called Broyden-T family correction. Obviously, the Broyden-T family correction satisfies the new quasi-Newton equation (10). When $\phi = 0$, we get DFP-T correction; when $\phi = 1$, we get inverse BFGS-T correction; when $\phi = \frac{\boldsymbol{s}_k^{\mathrm{T}} \hat{\boldsymbol{y}}_k}{(\boldsymbol{s}_k - \boldsymbol{H}_k \hat{\boldsymbol{y}}_k)^{\mathrm{T}} \hat{\boldsymbol{y}}_k}$, we get SR1-T correction. Four new corrections are constructed above. In order to ensure the symmetry positive definiteness of the correction formula and the computational stability of the algorithm, the properties of the quasi-Newton equation are discussed below.

III. New Quasi-Newton Equation Properties and Calculation Process

This section discusses some properties of the new quasi-Newton equations such as approximation accuracy, positive definiteness, etc., which is given in the form of a theorem.

Theorem 1 If $f(\boldsymbol{X})$ is sufficiently smooth and the Hessian matrix $\nabla^2 f(\boldsymbol{X})$ is continuous in a region Lipschitz, $\boldsymbol{X}_k$ and $\boldsymbol{X}_{k+1}$.

$$\boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{G}_{k+1} \boldsymbol{s}_k - \boldsymbol{s}_k^{\mathrm{T}} \hat{\boldsymbol{y}}_k = O(\|\boldsymbol{s}_k\|^5) \tag{16}$$

It is proved that the conclusions can be drawn from equations (8), (9) and (10).

Theorem 1 shows that the curvature of the objective function can be more accurately approximated by the new quasi-Newton equation (10) compared to the quasi-Newton equations (2), (4), (5) and (6)

$$\boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{y}_k > 0 \tag{17}$$

The approximate Hessian matrix produced by the standard quasi-Newton equation has symmetric positive definite genetic properties, If and only if $\boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{y}_k > 0$

Is valid, the DFP and BFGS updates produced by the standard quasi-Newton method have symmetry positivity [8]. Using the exact one-dimensional search or meeting Wolfe's conditions (18) and (19) in steps ensures that equation (17) holds.

$$f_{k+1} \le f_k + t\alpha_k \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{s}_k \tag{18}$$

$$\boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{s}_k \ge \sigma \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{s}_k \tag{19}$$

$$t \in (0, \frac{1}{2}) \quad \sigma \in (t, 1)$$

Our concern is whether the approximate Hessian matrix generated by the new quasi-Newton equation also has symmetric positive definite genetic properties.

For the new quasi-Newton equation, (17) will be replaced by the following formula

$$\boldsymbol{s}_k^{\mathrm{T}} \hat{\boldsymbol{y}}_k = \boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{y}_k + \hat{\theta}_k > 0 \tag{20}$$

This shows that the notation $\boldsymbol{s}_k^{\mathrm{T}} \hat{\boldsymbol{y}}_k$ is independent of the choice of vector u, but depends on the properties and steps of the objective function f(x). For a generally smoothing function, we can see from (16) that for a sufficiently small area $\boldsymbol{s}_k$, the formula (20) is satisfied in the area where the

Hessian matrix $G(X)$ is positive definite. Obviously we hope that the exact search and Wolfe conditions are not satisfied, $s_k^{\mathrm{T}}\hat{y}_k > 0$ is valid. We use the following strategy to ensure that this feature is met. Define the indicator set K as:

$$\mathrm{K} = \left\{ k: \frac{s_k^{\mathrm{T}}\hat{y}_k}{\|s_k\|^2} \geq \beta \|g_k\|^{\gamma} \right\}$$

where β and γ are positive constants. For BFGS-T update, the definition is listed as follows(the others are similar)

$$B_{k+1}^{BFGS-T} = \begin{cases} \left(I - \dfrac{\hat{y}_k s_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k}\right) B_k^{BFGS-T} \\ *\left(I - \dfrac{s_k \hat{y}_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k}\right) + \dfrac{\hat{y}_k \hat{y}_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k} & k \in \mathrm{K} \\ B_k^{BFGS-T} \end{cases} \tag{21}$$

$$H_{k+1}^{BFGS-T} = \begin{cases} H_{k+1}^{BFGS-T} = H_k^{BFGS-T} + \left(1 + \dfrac{\hat{y}_k H_k^{BFGS-T} \hat{y}_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k}\right) \dfrac{s_k s_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k} \\ \qquad - \dfrac{s_k \hat{y}_k^{\mathrm{T}} H_k^{BFGS-T} + H_k^{BFGS-T} \hat{y}_k s_k^{\mathrm{T}}}{\hat{y}_k^{\mathrm{T}} s_k} & k \in \mathrm{K} \\ H_k^{BFGS-T} \end{cases} \tag{22}$$

If B_0^{BFGS-T} (or H_0^{BFGS-T}) is a symmetric positive definite matrix, the symmetric positive definite matrix can be guaranteed for B_{k+1}^{BFGS-T} (or H_{k+1}^{BFGS-T}), so that the search direction of the structure is the descending direction, so as to ensure the convergence of the algorithm.

The matrix B in the above flow chart can be calculated by formula (11), and B^{-1} can be calculated by (12), (13), (14) or (15). In the classical quasi-Newton method, the matrix B (or B^{-1}) is calculated using the rank-1 correction formula, the DFP correction formula, the BFGS correction formula, or the Broyden-T family correction formula. In this algorithm, by constructing a correction formula, this new structure of the correction formula has not only symmetry positivity, but also the accuracy of approximation Hessian matrix is higher.

IV. Global Convergence Analysis

In this part, a condition that global convergence needs to be satisfied is given. Proved by the inverse proof, the algorithm constructed in this paper satisfies this condition. We prove that the proposed algorithm has global convergence. In the following, the global convergence of BFGS-T algorithm is given in the form of theorem. First, we give some basic assumptions:

Hypothesis A: The level set $\Omega = \{X : f(X) \leq f(X_0)\}$ is contained in a bounded convex set D.

Hypothesis B: The objective function $f(X)$ is secondary continuously differentiable, and there is a constant L> 0, making the following equation valid.

$g(X)$ indicates the gradient $f(X)$ at the point X.

Obviously, { $f(X_k)$ } generated by the algorithm is a descending sequence, and { X_k } is contained in the set Ω, so { $f(X_k)$ } is a bounded sequence, and then there is a constant

$$\lim_{k \to +\infty} f_k = f^* \tag{23}$$

The above equation is valid. Since { X_k } is bounded, by hypothesis B, there exists M> 0, so k

$$\|g_k\| \leq M$$

Is valid.

Lemma 1: If $f(X)$ satisfies the assumptions A and B, it is assumed that there are positive numbers a_1 and a_2, and makes it valid an infinite number of k

$$\|B_k s_k\| \leq a_1 \|s_k\|, \|s_k^{\mathrm{T}} B_k s_k\| \geq a_2 \|s_k\|^2 \tag{24}$$

There is

$$\liminf_{k \to \infty} g_k = 0$$

Proof: the expression E lets Equ.(24) valid for the set of k that holds (24). From $B_k d_k = -g_k$, $s_k = \alpha_k d_k$, and Equ. (24), there is

$$\|B_k d_k\| \leq a_1 \|d_k\|, \quad \|d_k^{\mathrm{T}} B_k d_k\| \geq a_2 \|d_k\|^2 \text{ and}$$
$$a_2 \|d_k\| \geq \|g_k\| \leq a_1 \|d_k\| \tag{25}$$

By Wolfe conditions (18), (19) and Hypothesis B, there is

$$L\alpha_k \|d_k\|^2 = L\|s_k\|\|d_k\| \geq \|g_{k+1} - g_k\|\|d_k\|$$
$$\geq (g_{k+1} - g_k)^{\mathrm{T}} d_k \geq (\sigma - 1) g_k^{\mathrm{T}} d_k \tag{26}$$

So there is

$$L\alpha_k \|d_k\|^2 \geq (\sigma - 1) g_k^{\mathrm{T}} d_k$$

This means that for any $k \in E$, there is

$$\alpha_k \geq \frac{(\sigma - 1) g_k^{\mathrm{T}} d_k}{L\|d_k\|^2} = \frac{(1-\sigma) d_k^{\mathrm{T}} B_k d_k}{L\|d_k\|^2} \geq \frac{(1-\sigma) a_2}{L} > 0 \tag{27}$$

On the other hand, from (23), there is

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) = \lim_{m\to+\infty}\sum_{k=1}^{m}(f_k - f_{k+1}) = f_1 - \lim_{m\to+\infty} f_{m+1} = f_1 - f^*$$

That is

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) < +\infty$$

By Wolfe condition (18), there is

$$\sum_{k=1}^{\infty} -t\alpha_k \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k < +\infty$$

$$\lim_{k\to+\infty} \alpha_k \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k = 0$$

Due to $\boldsymbol{B}_k\boldsymbol{d}_k = -\boldsymbol{g}_k$, there is

$$\lim_{k\to+\infty} \boldsymbol{d}_k^T\boldsymbol{B}_k\boldsymbol{d}_k = \lim_{k\to+\infty} -\boldsymbol{g}_k^T\boldsymbol{d}_k = 0$$

Because of the positive symmetric matrix $\boldsymbol{B}_k$, then

$$\lim_{k\to+\infty} \boldsymbol{d}_k = \boldsymbol{0} \text{ and } \lim_{k\to+\infty} \boldsymbol{g}_k = \boldsymbol{0}$$

From Lemma 1, in order to prove the global convergence of the algorithm, we must ensure an infinite number k. (24) holds. For making formula (24) valid, the following Lemma 2 is given. Its proof is similar to Theorem 2.1 of Ref. [119], so do not repeat.

Lemma 2: Let $\boldsymbol{B}_0^{BFGS-T}$ be a symmetric positive definite matrix, and $\boldsymbol{B}_{k+1}^{BFGS-T}$ is obtained by (21), if there exist normal numbers m and M such that

$$\frac{\boldsymbol{s}_k^{\mathrm{T}}\hat{\boldsymbol{y}}_k}{\|\boldsymbol{s}_k\|^2} \ge m,\ \frac{\|\hat{\boldsymbol{y}}_k\|^2}{\boldsymbol{s}_k^{\mathrm{T}}\hat{\boldsymbol{y}}_k} \le M \tag{28}$$

For all k, then $k \in \{1, 2, \cdots, t\}$. at least [t / 2] number of k makes Equ. (24) valid. If t is infinite, then there is an infinite number of k making Equ. (24) valid.

Theorem 2: Assuming that the objective function satisfies the assumptions A and B and the sequence { $\boldsymbol{X}_k$ } is generated by the BFGS-T algorithm, there is

$$\liminf_{k\to\infty} \boldsymbol{g}_k = \boldsymbol{0}$$

The above equation is valid.

Proof: From Lemma 1, we only need to prove that (24) is true for an infinite number of k. In two cases we prove it.

Case 1 K is a finite set. In this case, after a finite number of iterations, $\boldsymbol{B}_k^{BFGS-T}$ is a constant matrix that is independent of k. Thus there exist constants a_1 and a_2 for k so that Equ.(24) is valid. By Lemma 1, the conclusion holds.

Case 2 K is an infinite set. Inverse proofs prove that the conclusion established. If $\liminf_{k\to\infty} \boldsymbol{g}_k = 0$ discomfirm, then there is a constant $\varepsilon \in (0,1)$ making $\|g(\boldsymbol{X}_k)\| > \varepsilon$ valid that holds for all k. For all $k \in \mathrm{K}$, then there

$$\frac{\boldsymbol{s}_k^{\mathrm{T}}\hat{\boldsymbol{y}}_k}{\|\boldsymbol{s}_k\|^2} \ge \beta\|\boldsymbol{g}_k\|^{\gamma} > \beta\varepsilon^{\gamma} \tag{29}$$

Due to the symmetric positive definite matrix $\boldsymbol{B}_k^{BFGS-T}$, by the norm equivalence theorem, there exist L_1 and L_2 making the following formula valid

$$L_1\|\boldsymbol{s}_k\|^2 \le \|\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k\| \le L_2\|\boldsymbol{s}_k\|^2 \tag{30}$$

According to

$$|\hat{\theta}_k| = |12(f_k - f_{k+1}) + 5\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k + 7\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k + \boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k|$$

Then there is a constant $t \in (0,1)$ such that

$$|\hat{\theta}_k| = |12(-\boldsymbol{g}^{\mathrm{T}}(\boldsymbol{X}_k + t\boldsymbol{s}_k))\boldsymbol{s}_k + 5\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{s}_k + 7\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{s}_k + \boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k|$$
$$\le \|\boldsymbol{s}_k\|\|12(-\boldsymbol{g}(\boldsymbol{X}_k + t\boldsymbol{s}_k)) + 5\boldsymbol{g}_{k+1} + 7\boldsymbol{g}_k\| + \|\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{s}_k\|$$

By the assumption B and Equ. (30), then there is

$$|\hat{\theta}_k| \le (5L + 12Lt + L_2)\|\boldsymbol{s}_k\|^2$$

By your definition $\hat{y}_k$ and assumption B, there is

$$\|\hat{\boldsymbol{y}}_k\| = \left\|\boldsymbol{y}_k + \frac{\hat{\theta}_k\boldsymbol{s}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{s}_k}\right\| \le \|\boldsymbol{y}_k\| + \left\|\frac{\hat{\theta}_k\boldsymbol{s}_k}{\boldsymbol{s}_k^{\mathrm{T}}\boldsymbol{s}_k}\right\| \le \|\boldsymbol{y}_k\| + \frac{|\hat{\theta}_k|}{\|\boldsymbol{s}_k\|} \le (6L + 12Lt + L_2)\|\boldsymbol{s}_k\| \tag{31}$$

From Eqs. (29) and (31), there is

$$\frac{\|\hat{\boldsymbol{y}}_k\|^2}{\boldsymbol{s}_k^{\mathrm{T}}\hat{y}_k} \le \frac{\|\hat{\boldsymbol{y}}_k\|^2}{\beta\varepsilon^{\gamma}\|\boldsymbol{s}_k\|^2} \le \frac{(6L + 12Lt + L_2)^2\|\boldsymbol{s}_k\|^2}{\beta\varepsilon^{\gamma}\|\boldsymbol{s}_k\|^2} = \frac{(6L + 12Lt + L_2)^2}{\beta\varepsilon^{\gamma}} \tag{32}$$

From Eqs. (29) and (32), Lemma 2's conclusion is valid. By lemma 1, then $\lim_{k\to+\infty} \inf\,(g(\boldsymbol{X}_k)) = \boldsymbol{0}$ is a contradiction with the hypothesis, so $\lim_{k\to+\infty} \inf\,(g(\boldsymbol{X}_k)) = \boldsymbol{0}$ is valid.

V. Numerical Tests

In order to test the validity of the algorithm, some test functions are calculated in this paper. These test functions are from the reference [4]. $(1,1,\cdots,1)^{\mathrm{T}}$ and unit matrix as initial point and initial matrix respectively, the convergence condition is $\|\boldsymbol{g}_k\|_\infty \le 1e-5$.

Example 1: Find the minimum point of the Rosenbrock function, that is:

$$min\left[(2x_1-1)^2+\sum_{i=2}^{n} i(2x_{i-1}-x_i)^2\right]$$

The results are shown in Table 4.1.

Example 2: find the minimum point of the quadratic function, that is:

$$min\left[\sum_{i=1}^{n}(ix_i^2)\right]^2$$

The results are shown in Table II.

TABLE I: CALCULATION RESULTS OF EXAMPLE 1

M \ T n	500	800	1000	2000
BFGS-T	175	228	257	374
BFGS	204	266	301	440

TABLE II: CALCULATION RESULTS OF EXAMPLE 2

M \ T n	500	800	1000	2000
BFGS-T	468	699	908	1682
BFGS	1071	2003	2252	4038

T represents the number of iterations; M represents the calculation method; n represents the number of design variables.

Example 3 Find the minimum point of the Peak function. The Peak function and the contour image are shown in Fig. 1 and 2, respectively.

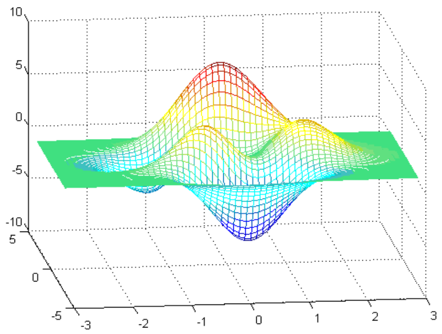

Fig. 1. Two-dimensional peak function image.

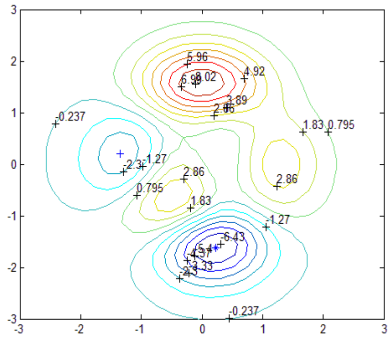

Fig. 2. Peak function contours.

The calculation results are shown in Table III.

From Numerical Examples 1 and 2, BFGS-T algorithm converges faster than BFGS, compared with BFGS. From Example 3, it can be seen that BFGS-T algorithm can find the global minimum more easily than BFGS.

Because the matrix constructed by the new quasi-Newton equation can approximate the Hessian matrix better than the standard quasi-Newton equation. The BFGS-T algorithm has global convergence, and the experimental data are basically consistent with the theoretical analysis.

TABLE III: COMPARISON OF BFGS-T AND BFGS ALGORITHMS

	Initial point	1 -2	0 0	-3 2	2 -3	-3 -3	2 2
	The initial point of the function value	-2.10235	0.98101	0.00002	-0.00431	-0.00007	0.13285
BFGS-T Algorithm calculation results	The best point	0.22828 -1.62553	23.83515 13.27876	0.22828 -1.62554	0.22828 -1.62554	-3.14080 -28.10305	6.91797 4.63040
	The best value of the function value	-6.55113	0	-6.55113	-6.55113	0	0
	The number of iterations	5	1	28	11	2	1
	$\|\nabla f(X_k)\|_\infty$	1.37e-07	1.00e-06	3.78e-06	4.39e-08	1.38e-06	2.71e-06
BFGS Algorithm	The best point	0.22689 -1.62506	23.83515 13.27876	-1.34740 0.20452	-1.34740 0.20452	-3.14080 -28.10350	6.91797 4.63040

calculation results	The best value of the function value	-6.55111	0	-3.04985	-3.04985	0	0
	The number of iterations	14	1	29	30	2	1
	$\|\nabla f(X_k)\|_\infty$	1.02e-06	1.01e-07	4.40e-08	1.69e-08	1.38e-06	2.71e-06

VI. Conclusion

Based on the fourth-order tensor expansion of the objective function, a new quasi-Newton equation is proposed. In terms of the new quasi-Newton equation, four quasi-Newton formats are put forward, which proves the global convergence of the proposed BFGS-T method. The new quasi-Newton equation not only retains the positive definite genetic property of some standard quasi-Newton equations, but also is the higher approximation accuracy than them. Numerical experiments also show the superiority of this algorithm.

References

[1] J. Z. Zhang, N. Y. Deng, and L. H. Chen, “New quasi-newton equation and related methods for unconstrained optimization,” *Journal of Optimization Theory and Applications*, 1999, vol. 102, pp. 147-167.

[2] H. Y. Huang, “Unified approach to quadratically convergent algorithms for function minimization,” *Journal of Optimization Theory and Applications*, 1970, vol. 5, pp. 405–423.

[3] Z. Wei, G. Li, and L. Qi, “New quasi-Newton methods for unconstrained optimization problems,” *Applied Mathematics and Computation*, 2006, vol. 175, pp. 1156–1188.

[4] F. Biglari, M. A. Hassan, and W. J. Leong, “New quasi-Newton methods via higher order tensor models,” *Journal of Computational and Applied Mathematics*, 2011, vol. 235, pp. 2412–2422.

[5] J. Z. Zhang and C. X. Xu, “Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations,” *Journal of Computational and Applied Mathematics*, 2001, vol. 137, pp. 269–278

[6] H. Yabe, H. Ogasawara, and M. Yoshino, “Local and super linear convergence of quasi-Newton methods based on modified quasi-Newton equations,” *Journal of Computational and Applied Mathematics*, 2007, vol. 205, pp. 617– 632.

[7] C. X. Xu and J. Z. Zhang, “Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations,” *Journal of Computational and Applied Mathematics*, 2001,vol. 137, no. 12,pp. 169-278

[8] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer New York, 1999.

Wang Bao works at Ningbo University of Technology in 2012. He received a Master's degree in Computational Mathematics and PhD in Aircraft design from Northwestern Polytechnical University in china. Research area is optimization algorithm design and data mining

Wang Yanxin joined Ningbo University of Technology in 2013. Previously He was a visiting scholar at the Department of Statistics, Warwick University. He received a Master's degree in Mathematics from Ningxia University and PhD in Mathematics from Wuhan University of China. He has published several technical papers in international journals. His current research includes Wavelet analysis, Numerical solution for differential and integral equations, Multi-scale method for time series and High dimensional data analysis.