NEW QUASI-NEWTON (DFP) WITH LOGISTIC MAPPING

Salah Gazi Shareef and Bayda Ghanim Fathi
Dept. of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region-Iraq.
(Accepted for publication: February 25, 2016)

Abstract:
In this paper, we propose a modification of the self-scaling quasi-Newton (DFP) method for unconstrained optimization using logistic mapping. We show that it produces a positive definite matrix. Numerical results demonstrate that the new algorithm is superior to standard DFP method with respect to the NOI and NOF.

Keywords: Unconstrained optimization, Quasi-Newton methods, DFP method, Logistic mapping.

1- Introduction

The quasi-Newton algorithms for minimizing a function \( f(x), x \in \mathbb{R}^n \), are iterative accelerated gradient methods which use past positions and functional values rather than an analytically or numerically calculated one to approximate the inverse of the Hessian matrix \( H \) of the function. This is accomplished by selecting an initial approximation \( H_0 \) to the inverse Hessian, as well as an initial approximation \( x_0 \) to the minimum of \( f(x) \), and then finding at each step \( \alpha_k \), the scalar parameter which minimizes \( f(x_k - \alpha_k H_k g_k) \) where \( g_k = g(x_k) = \nabla f(x) \).

It is known that the search direction of the quasi-Newton algorithms is

\[
d_k = -H_k g_k, \quad (1.1)
\]

and the approximate matrix \( H_k \) is updated by

\[
H_{k+1} = H_k + D_k, \quad (1.2)
\]

where \( D_k \) is the correction matrix.

The Davidon Fletcher Powell (DFP) algorithm was the first quasi-Newton algorithm created (Shanno and Kettler, 1970). In this technique, substituting \( v_k = \frac{H_{k+1} v_k}{v_k^T H_{k+1} v_k} \) where \( v_k = x_{k+1} - x_k \) and \( y_k = g_{k+1} - g_k \) for \( D_k \) and giving

\[
H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{y_k v_k^T}{v_k^T y_k} \quad (1.3)
\]

The following theorem will be used later.

Theorem (1.1). (Edwin and Stanislaw, 2001). Let a function \( f \in C, x_k \in \mathbb{R}^n, g_k = 0 \), and \( H_k \) is an \( n \times n \) real symmetric positive definite matrix. If we set \( x_{k+1} = x_k - \alpha_k H_k g_k \), where \( \alpha_k = \arg\min_{\alpha} f(x_k - \alpha H_k g_k) \), then \( \alpha_k > 0 \), and \( f(x_{k+1}) < f(x_k) \).

2- A new self-scaling quasi-Newton (DFP) formula

For a control parameter, \( \mu \), the logistic mapping (Lu et al., 2006) is defined by

\[
z_{k+1} = \mu z_k (1 - z_k) \quad (2.1)
\]

Let us consider the quasi-Newton condition

\[
H_{k+1} y_k = v_k, \quad (2.2)
\]

where \( v_k = \alpha_k d_k = x_{k+1} - x_k, \mu, \nu \in (0,1) \) and \( y_k = \Delta g_k = g_{k+1} - g_k \). A new self-scaling quasi-Newton (DFP) formula can be defined as

\[
H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{\nu (1 - \mu) y_k v_k^T}{v_k^T y_k} \quad (2.3)
\]
Algorithm: A New DFP Algorithm

Step (1):- Set $k = 0$; select $x_0$, and a real symmetric positive definite $H_0 (H_0 = I)$.
Step (2):- If $g_k = 0$, stop; else $d_k = -H_k g_k$, where $g(x) = \nabla f(x)$
Step (3):- Compute $\alpha_k = \arg \min f(x_k + \alpha d_k)$

\[ x_{k+1} = x_k + \alpha_k d_k. \]
Step (4):- Compute $v_k = \Delta x_k = \alpha_k d_k$

\[ y_k = \Delta g_k = g_{k+1} - g_k = g_{k+1} - g_k = Gv_k. \]

Step (5):- Set $k = k + 1$; go to step 2.

Theorem (2.1). If the new self-scaling quasi-Newton (DFP) formula (2.3) applied to the quadratic function with Hessian $G = G^T$, then $H_{k+1}\Delta g_i = \mu(1-\gamma)\Delta x_i$ for $0 \leq i \leq k$ where $v_k = \Delta x_k = x_{k+1} - x_k$ and $y_k = \Delta g_k = g_{k+1} - g_k = Gv_k$.

Note: $d_k^T Gd_i = 0$.

Proof. We prove this theorem by using induction criteria. For $k = 0$, we have

\[ H_1 y_0 = H_0 y_0 - \frac{H_0 y_0 y_0^T H_0}{y_0^T H_0 y_0} y_0 + \frac{\mu(1-\gamma)v_0 v_0^T}{v_0^T y_0} y_0 = \mu(1-\gamma)v_0. \]

Assume the result is true for $k - 1$; that is $H_k \Delta g_i = \mu(1-\gamma)\Delta x_i$, $0 \leq i \leq k - 1$.

We now show that $H_{k+1} \Delta g_i = \mu(1-\gamma)\Delta x_i$, $0 \leq i \leq k$. First consider $i = k$, we have

\[ H_{k+1} y_k = H_k y_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} y_k + \frac{\mu(1-\gamma)v_k v_k^T}{v_k^T y_k} y_k, \]

implies that

\[ H_{k+1} y_k = \mu(1-\gamma)v_k. \]

It remains to consider the case $i < k$. Using the hypothesis, we have

\[ H_{k+1} y_i = H_k y_i - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} y_i + \frac{\mu(1-\gamma)v_k v_k^T}{v_k^T y_k} y_i = \mu(1-\gamma)v_i - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} (y_k^T v_i) + \frac{\mu(1-\gamma)v_k v_k^T}{v_k^T y_k} (v_k^T y_i). \]

Since

\[ v_k^T y_i = v_k^T Gv_i = \alpha_k a_i d_k^T Gd_i = 0 \]

and

\[ y_k^T v_i = v_k^T Gv_i = \alpha_k a_i d_k^T Gd_i = 0. \]

Hence,

\[ H_{k+1} y_i = \mu(1-\gamma)v_i. \]

The proof is completed.
Theorem (2.2). Suppose that $g_k \neq 0$. In the new self-scaling quasi-Newton (DFP) formula (2.3), if $H_k$ is positive definite, then so is $H_{k+1}$.

**Proof.** Multiply both sides of (2.3) by $x^T$ from left and by $x$ from right, we get

\[
x^T H_{k+1} x = x^T H_k x - x^T H_k y_k y_k^T H_k x + \frac{\mu(1 - \gamma)}{y_k^T H_k y_k} v_k y_k^T x
\]

\[
= x^T H_k x - \frac{(x^T H_k y_k)^2}{y_k^T H_k y_k} + \frac{\mu(1 - \gamma)(x^T v_k)^2}{v_k^T y_k}.
\]

We can define

\[
a = H_k^{1/2} x \quad \text{and} \quad b = H_k^{1/2} y_k,
\]

where $H_k = H_k^{1/2} H_k^{1/2}$.

Now, using the definition of $a$ and $b$, we obtain

\[
x^T H_k x = x^T H_k^{1/2} H_k^{1/2} x = a^T a,
\]

\[
x^T H_k y_k = x^T H_k^{1/2} H_k^{1/2} y_k = a^T b,
\]

and

\[
y_k^T H_k y_k = y_k^T H_k^{1/2} H_k^{1/2} y_k = b^T b.
\]

Hence

\[
x^T H_{k+1} x = a^T a - \frac{(a^T b)^2}{b^T b} + \frac{\mu(1 - \gamma)(x^T v_k)^2}{v_k^T y_k}
\]

\[
= \frac{\|a\|^2 \|b\|^2 - (a^T b)^2}{\|b\|^2} + \frac{\mu(1 - \gamma)(x^T v_k)^2}{v_k^T y_k}.
\]

We know that $\mu(1 - \gamma)$ is positive and we have $v_k^T y_k = v_k^T (g_{k+1} - g_k) = -v_k^T g_k$ because $v_k^T g_{k+1} = \alpha_k d_k^T g_{k+1} = 0$ by (In the conjugate direction algorithm, $g_{k+1}^T d_i = 0$ for all $k$, $0 \leq k \leq n - 1$, and $0 \leq i \leq k$ (Edwin and Stanislaw, 2001)).

Since $v_k = \alpha_k d_k = -\alpha_k H_k g_k$, we get

\[
v_k^T y_k = -v_k^T g_k = \alpha_k g_k^T H_k g_k.
\]

The above yields

\[
x^T H_{k+1} x = \frac{\|a\|^2 \|b\|^2 - (a^T b)^2}{\|b\|^2} + \frac{\mu(1 - \gamma)(x^T v_k)^2}{\alpha_k g_k^T H_k g_k} \quad (2.4)
\]

The fractional terms on the right-hand side of (2.4) are nonnegative, the first term is nonnegative because of the Cauchy-Schwarz inequality, and the second term is nonnegative because $H_k$, $\alpha_k > 0$ by Theorem (1.1) and $\mu(1 - \gamma) > 0$. Therefore, to show that $x^T H_{k+1} x > 0$ for $x \neq 0$, we only need to demonstrate that these terms do not vanish simultaneously. The first term vanishes only if $a$ and $b$ are proportional, that is if $a = \beta b$ for a scalar $\beta$.
To complete the proof it is enough to show that if $a = \beta b$, then $rac{\mu y(1-\gamma)(x^Tv_k)}{a_k g_k^T H_k g_k} > 0$.

First observe that

$$H_k^{1/2} x = a = \beta b = \beta H_k^{1/2} y_k = H_k^{1/2}(\beta y_k).$$

Hence,

$$x = \beta y_k$$

Using the above expression for $x$ and $v_k^T y_k = -\alpha_k g_k^T H_k g_k$, we obtain

$$\frac{\mu y(1-\gamma)(x^Tv_k)^2}{a_k g_k^T H_k g_k} = \frac{\mu y(1-\gamma)\beta^2(y_k^Tv_k)^2}{a_k g_k^T H_k g_k} = \frac{\mu y(1-\gamma)\beta^2(a_k g_k^T H_k g_k)^2}{a_k g_k^T H_k g_k} = \mu y(1-\gamma)\beta^2 a_k g_k^T H_k g_k > 0.$$  

Thus, for all $x \neq 0$

$$x^T H_{k+1} x > 0.$$  

Then the proof is completed.

3- Numerical Results

This section is devoted to test the implementation of the new method. We compare standard formula of DFP and new formula of self-scaling Q-N (DFF), the comparative tests involve well-known nonlinear problems (standard test function) with different dimensions $4 \leq n \leq 100$, all programs are written in FORTRAN95 language and for all cases the stopping condition is $\|g_{k+1}\|_\infty \leq 10^{-5}$. Efficiency of the new DFP algorithm has been tested by means of 10 standard problems. Experimental results in Table (1) represent the number of function evaluations NOF and the number of iterations NOI. Table (2) shows the percentage of improving the new algorithm and confirms that the new method is superior to standard method with respect to the NOI and NOF.
Table (1): Comparison between the performance of the standard DFP update and new DFP update.

<table>
<thead>
<tr>
<th>Test fun.</th>
<th>n</th>
<th>NOI</th>
<th>NOF</th>
<th>NOI</th>
<th>NOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powell</td>
<td>4</td>
<td>23</td>
<td>126</td>
<td>17</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>80</td>
<td>467</td>
<td>35</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>60</td>
<td>328</td>
<td>35</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>44</td>
<td>230</td>
<td>38</td>
<td>151</td>
</tr>
<tr>
<td>Wood</td>
<td>4</td>
<td>39</td>
<td>250</td>
<td>37</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>243</td>
<td>1380</td>
<td>251</td>
<td>1178</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>751</td>
<td>3439</td>
<td>700</td>
<td>2575</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1192</td>
<td>4758</td>
<td>1106</td>
<td>3548</td>
</tr>
<tr>
<td>Wolfe</td>
<td>4</td>
<td>7</td>
<td>18</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>72</td>
<td>145</td>
<td>55</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>82</td>
<td>165</td>
<td>61</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>95</td>
<td>191</td>
<td>68</td>
<td>137</td>
</tr>
<tr>
<td>Cubic</td>
<td>4</td>
<td>18</td>
<td>76</td>
<td>17</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>34</td>
<td>114</td>
<td>46</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>54</td>
<td>166</td>
<td>41</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>60</td>
<td>183</td>
<td>47</td>
<td>135</td>
</tr>
<tr>
<td>Rosen</td>
<td>4</td>
<td>36</td>
<td>145</td>
<td>34</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>247</td>
<td>1017</td>
<td>219</td>
<td>767</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>605</td>
<td>2240</td>
<td>348</td>
<td>1038</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>984</td>
<td>3570</td>
<td>459</td>
<td>1347</td>
</tr>
<tr>
<td>Mile</td>
<td>4</td>
<td>26</td>
<td>119</td>
<td>24</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>38</td>
<td>174</td>
<td>30</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>34</td>
<td>152</td>
<td>31</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>44</td>
<td>193</td>
<td>41</td>
<td>164</td>
</tr>
<tr>
<td>Beale</td>
<td>4</td>
<td>8</td>
<td>22</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>10</td>
<td>27</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>10</td>
<td>27</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>10</td>
<td>27</td>
<td>10</td>
<td>26</td>
</tr>
<tr>
<td>Gedger</td>
<td>4</td>
<td>6</td>
<td>18</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>shallow</td>
<td>4</td>
<td>13</td>
<td>40</td>
<td>13</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>45</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>16</td>
<td>46</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>16</td>
<td>46</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>G. central</td>
<td>4</td>
<td>21</td>
<td>146</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>21</td>
<td>146</td>
<td>16</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>22</td>
<td>154</td>
<td>16</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>22</td>
<td>154</td>
<td>16</td>
<td>97</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>5076</td>
<td>20598</td>
<td>3925</td>
<td>13286</td>
</tr>
</tbody>
</table>

Table (2): Percentage of improving the new algorithm

<table>
<thead>
<tr>
<th>Tools</th>
<th>Standard formula</th>
<th>New formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOI</td>
<td>100 %</td>
<td>77.3 %</td>
</tr>
<tr>
<td>NOF</td>
<td>100 %</td>
<td>64.5 %</td>
</tr>
</tbody>
</table>
4- Conclusion

A new formula for updating quasi-Newton matrices based on DFP and which uses logistic mapping is presented. It is shown that the new algorithm produces positive definite matrices. Numerical experiments indicate that our algorithm is better than the original DFP with respect to the NOI and NOF.

5- References

