A new spectral conjugate gradient method (SDYCG) is presented for solving unconstrained optimization problems in this paper. Our method provides a new expression of spectral parameter. This formula ensures that the sufficient descent condition holds. The search direction in the SDYCG can be viewed as a combination of the spectral gradient and the Dai-Yuan conjugate gradient. The global convergence of the SDYCG is also obtained. Numerical results show that the SDYCG may be capable of solving large-scale nonlinear unconstrained optimization problems.

1. Introduction

As well known, a great deal of issues, which are studied in scientific research fields, can be translated to unconstrained optimization problems. The spectral conjugate gradient (SCG) method does nice jobs among various algorithms for solving nonlinear optimization problems. The spectral conjugate gradient combines the spectral gradient and the conjugate gradient. To the SCG method, the choice of spectral parameter is crucially important. In this paper, we propose a new spectral conjugate gradient method based on the Dai-Yuan conjugate gradient method by providing a new spectral parameter. Our purpose is to obtain an efficient algorithm for the unconstrained optimization.

An unconstrained optimization problem is customarily expressed as

\[
\min_{x \in \mathbb{R}^n} f(x),
\]

where \( f \) is the objective function.

(P1) The function \( f \) is bounded below and is continuously differentiable in a neighborhood \( N \) of the level set \( \mathcal{L} = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \), where \( x_0 \) is the starting point.

(P2) The gradient \( g(x) \) of \( f \) is Lipschitz continuous in \( N \); that is, there exists a constant \( L > 0 \) such that \( \|g(x) - g(y)\| \leq L\|x - y\| \) for all \( x, y \in N \).

Generally, a sequence \( \{x_k\} \) is obtained in an algorithm for solving (1) and has the following format:

\[
x_{k+1} = x_k + \alpha_k d_k,
\]

where \( d_k \) is a search direction and \( \alpha_k \) is the step size. At each iterative point \( x_k \), we usually determine \( d_k \) firstly and then compute \( \alpha_k \) by some principles.

There are different ways to determine the direction \( d_k \). In the classical steepest-descent method, \( d_k = -g_k \). In the conjugate gradient (CG) method, \( d_k \) is of the form

\[
\begin{align*}
d_0 &= -g_0, \\
d_{k+1} &= -g_{k+1} + \beta_k d_k, \quad \text{if } k \geq 0,
\end{align*}
\]

where \( \beta_k \) is a scalar parameter characterizing the conjugate gradient method. The best-known expressions of \( \beta_k \) are
There are also some approaches to determine the step size \( \alpha_k \) in (2). Unfortunately, the steepest-descent method performs poorly. Barzilai and Borwein improved the steepest-descent method greatly by providing a spectral choice of step size in [6]. Their algorithm has the form

\[
x_{k+1} = x_k - \alpha_k g_k,
\]

where \( \alpha_k = s_k g_k / s_k y_k \) or \( \alpha_k = s_k y_k / g_k^T y_k \) with \( s_k = x_{k+1} - x_k \). Many algorithms are proved convergent under the Wolfe condition; that is, the step size \( \alpha_k \) satisfies

\[
f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k,
\]

\[
g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k
\]

with \( 0 < \delta < \sigma < 1 \).

In recent years, some scholars developed a new method—spectral conjugate gradient (SCG) method—for solving (1). For example, Raydan introduced the spectral gradient method for large-scale unconstrained optimization in [7]. He combined a nonmonotone line search strategy that guarantees global convergence with the Barzilai and Borwein method. Utilizing spectral gradient and conjugate gradient ideas, Birgin and Martinez proposed a spectral conjugate gradient method in [8]. In their algorithm, the search direction has the form

\[
d_0 = -g_0,
\]

\[
d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k d_k, \text{ if } k \geq 0.
\]

In [8], the best combination of this formula, the scaling, and the initial choice of step-length is also studied. Following [8], some papers discussed the various choices of the spectral parameter \( \theta_k \) based on different \( \beta_k \). For example, Du and Chen [9] gave a modified spectral FR conjugate gradient method with Wolfe-type line search based on FR formula. Their spectral parameters \( \theta_k \) and \( \beta_k \) are expressed as

\[
\theta_{k+1} = \frac{d_k^T y_k}{\|g_k\|^2},
\]

\[
\beta_k = \beta_k^F.
\]

Yu et al. [10] presented a modification of spectral Perry's conjugate gradient formula, which possessed the sufficient descent property independent of line search condition. Their search direction \( d_{k+1} \) is defined by (8) and \( \beta_k \) has the form

\[
\beta_k^{DSP} = \beta_k^F - \frac{C \|y_k - \delta_{k+1} s_k\|^2}{\delta_{k+1} (y_k^T d_k^k)^2} g_k^T d_k,
\]

where

\[
C \geq \frac{1}{4},
\]

\[
\delta_{k+1} = \frac{1}{\theta_{k+1}} = \frac{y_k^T d_k}{s_k^T y_k},
\]

\[
s_k = x_{k+1} - x_k.
\]

Liu and Li [11] proposed a spectral DY-type projection method for nonlinear monotone systems of equations. The direction \( d_k \) is also determined by (8) and the parameters are defined by

\[
\theta_{k+1} = \frac{s_k^T y_k}{s_k^T y_k},
\]

\[
\beta_k = \beta_k^{DY}.
\]

We will propose a new SCG method based on the Dai-Yuan-type conjugate gradient method in this paper. A new selection of \( \theta_{k+1} \) is introduced in our algorithm such that the sufficient descent condition holds. In addition, the global convergence of the new method is obtained.

The present paper is organized as follows. In Section 2, we outline our new method for unconstrained nonlinear optimization, and we show that the sufficient descent condition holds under mild assumptions. The global convergence is proved in Section 3, while the numerical results compared with CG-DESCENT are given in Section 4. At last section, we draw some conclusions about our new spectral conjugate gradient method.

### 2. Spectral Dai-Yuan-Type Conjugate Gradient Method

In this paper, we consider the spectral conjugate gradient method in which the search direction is of the form

\[
d_0 = -g_0,
\]

\[
d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k^{DY} d_k, \text{ if } k \geq 0,
\]

where

\[
\theta_{k+1} = \max \left\{ \frac{2 g_k^T d_k}{|d_k^T y_k|}, \frac{2 g_{k+1}^T d_k - g_k^T d_k}{d_k^T y_k} \right\}
\]

\[
\beta_k = \beta_k^{DY}.
\]
and the scalar parameter $\beta_{DY}^k$ is defined by (4). This is a new spectral conjugate gradient method for solving problem (1) because the expression (14) of spectral parameter $\theta_{k+1}$ is completely different from those in other papers. The search direction (13) is a combination of the spectral gradient and the Dai-Yuan conjugate gradient. We hope that (14) may be an efficient choice.

In order to obtain the global convergence of our method, we assume that the step size satisfies the strong Wolfe condition; that is, the step size $\alpha_k$ satisfies (6) and

$$\left| g^T (x_k + \alpha_k d_k) d_k \right| \leq -\sigma g^T d_k$$

with $0 < \delta < \sigma < 1$. It is easy to see that (7) holds if (15) holds.

The following is a detailed description of the spectral Dai-Yuan-type conjugate gradient (SDYCG) algorithm.

**SDYCG Algorithm**

Step 1 (initialization). Choose $x_0 \in \mathcal{R}^n$, set $k = 0$, and take $\epsilon > 0$.

Step 2 (check the convergence condition). Compute $g_k$; if $\|g_k\| \leq \epsilon$, then stop.

Step 3 (form the search direction). If $k = 0$, then $d_0 = -g_0$. Else, compute $\theta_{k+1}$ and $\beta_{DY}^k$ by formulas (14) and (4), respectively; then compute $d_k$ by (13) and (14).

Step 4 (line search). Find $\alpha_k$ which satisfies the strong Wolfe conditions (6) and (15).

Step 5 (compute the new point). Set $x_{k+1} = x_k + \alpha_k d_k$ and $k = k + 1$ and go to Step 2.

The framework of the SDYCG algorithm is similar to other spectral conjugate gradient algorithms. However, we choose a different spectral parameter $\theta_{k+1}$ (see (14)) which is the main difference between SDYCG and the others.

The global convergence of SDYCG algorithm will be given in the next section. Before that, we will show that the search direction (13) can ensure the sufficient descent condition.

**Lemma 1.** Suppose that the sequence $\{d_k\}$ is generated by the SDYCG algorithm; then

$$g^T d_k \leq -\|g_k\|^2$$

for all $k \geq 0$.

**Proof.** Since $d_k$ is calculated by formula (13), we can get, if $k = 0$, that $g^T d_0 = -\|g_0\|^2$ whereas when $k \geq 0$, from (14), we find

$$\theta_{k+1} \geq \frac{2g^T_{k+1}d_k - g^T_k d_k}{d_k^T y_k} = \frac{g^T_{k+1}d_k}{d_k^T y_k} + 1
\begin{align*}
= \frac{\|g_{k+1}\|^2}{\|d_k\|^2} - \frac{g^T_{k+1}d_k}{\|g_{k+1}\|} + 1 = \beta_{DY}^k \frac{g^T_{k+1}d_k}{\|g_{k+1}\|^2} + 1.
\end{align*}

Furthermore,

$$-\theta_{k+1} \|g_{k+1}\|^2 + \beta_{DY}^k g^T_{k+1}d_k \leq -\|g_{k+1}\|^2.$$

From the second formula of (13), we obtain

$$g^T_{k+1}d_{k+1} \leq -\|g_{k+1}\|^2.$$

The proof is completed.

This lemma gives the fact that the direction $d_k$ is a descent direction. Besides, $d_k$ possesses the following property.

**Lemma 2.** Suppose that the SDYCG algorithm is implemented with the step size $\alpha_k$ that satisfies the strong Wolfe conditions (6) and (15). If $g_k \neq 0$ for all $k \geq 0$, then

$$\frac{(\beta_{DY}^k)^2}{(g^T_{k+1}d_{k+1})^2} \leq \frac{1}{(g^T_k d_k)^2}.$$

**Proof.** By (14) and $0 < \sigma < 1$, we have

$$\theta_{k+1} \geq (\sigma + 1) \frac{g^T_{k+1}d_k}{d_k^T y_k} = (\sigma + 1) \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|} \frac{\|g_k d_k\|}{\|g_{k+1}\|^2}.$$

With (15) and (21), we get

$$\theta_{k+1} \|g_{k+1}\|^2 \leq \theta_{k} \|g_{k+1}\|^2 - \sigma \|\beta_{DY}^k g^T d_k\|$$

$$\leq \theta_{k} \|g_{k+1}\|^2 - \|\beta_{DY}^k g^T d_{k+1}\|$$

Therefore, inequality (20) follows. This completes the proof.

Inequality (20) gives the close relationship of the inner product of gradient and direction between the adjacent two iterations. It will play an important role in our global convergence analysis.

### 3. Convergence Analysis

Dai and Yuan stated in [5] that the following result had been essentially proved by Zoutendijk and Wolfe.

**Lemma 3.** Suppose that the function $f(x)$ has the properties (P1) and (P2). Assume that $d_k$ is a descent direction and $\alpha_k$ is obtained by the Wolfe conditions (6) and (7). Then

$$\sum_{k \geq 0} \left( \frac{g^T d_k}{\|d_k\|^2} \right)^2 < +\infty.$$
Theorem 4. Suppose that the function \( f(\mathbf{x}) \) has the properties (P1) and (P2). Sequences \( \{x_k\} \), \( \{g_k\} \), and \( \{d_k\} \) are generated by SDYCG algorithm. Then either \( g_k = 0 \) for some \( k \) or

\[
\lim \inf_{k \to \infty} \|g_k\| = 0. \tag{24}
\]

Proof. Suppose that \( g_k \neq 0 \) for all \( k \) and (24) is not true. Then there exists a constant \( c > 0 \), such that

\[
\|g_k\| \geq c \tag{25}
\]

for all \( k \) of the iterations.

The second equality of (13) implies

\[
d_{k+1} + \theta_{k+1}g_{k+1} = \beta_k^{\text{DY}} d_k; \tag{26}
\]

we get

\[
\|d_{k+1}\|^2 = \left(\beta_k^{\text{DY}}\right)^2 \|d_k\|^2 - 2\theta_{k+1} \|g_{k+1}\| \|d_k\| + \theta_{k+1}^2 \|g_{k+1}\|^2 - \left(\beta_k^{\text{DY}}\right)^2 \|d_k\|^2 \tag{27}
\]

Dividing both sides by \( g_k^T d_{k+1} \), we have

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} = \left(\frac{\beta_k^{\text{DY}}}{g_k^T d_{k+1}}\right)^2 \|d_k\|^2 - \left(\frac{\beta_k^{\text{DY}}}{g_k^T d_{k+1}}\right)^2 \|d_k\|^2 - \left(\frac{\theta_{k+1}}{\|g_{k+1}\|} + \frac{\theta_{k+1} \|g_{k+1}\|}{g_k^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}. \tag{28}
\]

Combining (20) in Lemma 2, we see the inequality

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} \leq \frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}. \tag{29}
\]

Summing both sides, we obtain

\[
\sum_{i=0}^{k} \frac{\|d_{i+1}\|^2}{(g_i^T d_{i+1})^2} \leq \sum_{i=0}^{k} \frac{\|d_i\|^2}{(g_i^T d_i)^2} + \frac{1}{\|g_{k+1}\|^2}. \tag{30}
\]

So, from (25) and (30), we get

\[
\frac{\|d_{k+1}\|^2}{(g_k^T d_{k+1})^2} \leq \frac{k+2}{c}. \tag{31}
\]

This relation is equivalent to

\[
\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c}{k+2}. \tag{32}
\]

Summing over \( k \), we obtain

\[
\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty. \tag{33}
\]

From the SDYCG algorithm, the step size satisfies the strong Wolfe condition, so the Wolfe condition (7) holds. And the directions obtained by the algorithm are descent from Lemma 1. But the last equality contradicts the Zoutendijk condition (23). Hence, our original assertion (25) must be false, giving that either \( g_k = 0 \) for some \( k \) or (24) holds. \( \square \)

### 4. Numerical Results

In order to test the numerical performance of the SDYCG algorithm, we choose some unconstrained problems with the initial points from CUTEr library [12, 13]. They are listed in Table 1.

The experiments are run on a personal computer with a 64-bit processor, 2.5 GHz of CPU, and 4 GB of RAM memory. All the codes are written in MATLAB language and are compiled with this software.

We would like to compare the SDYCG with the CG-DESCENT. The CG-DESCENT is a conjugate gradient algorithm with guaranteed descent proposed by Hager and Zhang in [14]. It has been proven an excellent algorithm in recent years.

To make the comparison as fair as possible, we use the criterion \( \|g\| < 10^{-5} \) to terminate the executions and impose restriction on the number of iterations less than 500 in both algorithms. All the step sizes satisfy the strong Wolfe conditions (6) and (15).

We use the performance profiles proposed by Dolan and More [15] to show the efficiency of comparisons. Performance profiles can be used as a tool for benchmarking and comparing optimization software. The performance profile for a solver is the (cumulative) distribution function for a performance metric. For example, if computing time is chosen as a metric, then we compute the ratio of the computing time of the solver versus the best time of all the solvers. That is, for each method, we plot the fraction \( P(y) \) of problems for which the method is within a factor \( y \) of the best time. The curve of a solver being above others means that it has the highest probability of being the optimal solver. We use a log scale for \( y \) to capture the performance of all the solvers.

In order to observe the numerical results of the SDYCG and the CG-DESCENT, we choose three different dimensions of each test function. The dimensions are \( n = 10^2, n = 10^3 \), and \( n = 10^4 \), respectively. According to the numerical results obtained in every dimension, we plot two figures based on CPU time and iterations, respectively.
Table 1: Test problems.

<table>
<thead>
<tr>
<th>Number</th>
<th>Function name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ENGVAL1</td>
</tr>
<tr>
<td>2</td>
<td>FLETCBV2</td>
</tr>
<tr>
<td>3</td>
<td>TOINTGSS</td>
</tr>
<tr>
<td>4</td>
<td>COSINE</td>
</tr>
<tr>
<td>5</td>
<td>ARWHEAD</td>
</tr>
<tr>
<td>6</td>
<td>EDENSCH</td>
</tr>
<tr>
<td>7</td>
<td>EG2</td>
</tr>
<tr>
<td>8</td>
<td>GENROSE</td>
</tr>
<tr>
<td>9</td>
<td>LIARWHD</td>
</tr>
<tr>
<td>10</td>
<td>Generalized White &amp; Holst</td>
</tr>
<tr>
<td>11</td>
<td>Extended Wood</td>
</tr>
<tr>
<td>12</td>
<td>Extended quadratic penalty QP1</td>
</tr>
<tr>
<td>13</td>
<td>BDEXP</td>
</tr>
<tr>
<td>14</td>
<td>HIMMELBG</td>
</tr>
<tr>
<td>15</td>
<td>Hager</td>
</tr>
<tr>
<td>16</td>
<td>Extended TET</td>
</tr>
<tr>
<td>17</td>
<td>Diagonal 5</td>
</tr>
<tr>
<td>18</td>
<td>Extended Himmelblau</td>
</tr>
<tr>
<td>19</td>
<td>Diagonal 6</td>
</tr>
<tr>
<td>20</td>
<td>Extended DENSCHNF</td>
</tr>
<tr>
<td>21</td>
<td>LIARWHD</td>
</tr>
<tr>
<td>22</td>
<td>Extended BDI</td>
</tr>
<tr>
<td>23</td>
<td>Extended Hiebert</td>
</tr>
<tr>
<td>24</td>
<td>Extended Tridiagonal 2</td>
</tr>
<tr>
<td>25</td>
<td>QUARTC</td>
</tr>
<tr>
<td>26</td>
<td>Extended DENSCHNB</td>
</tr>
<tr>
<td>27</td>
<td>Extended Rosenbrock</td>
</tr>
<tr>
<td>28</td>
<td>Raydan 2</td>
</tr>
<tr>
<td>29</td>
<td>Diagonal 2</td>
</tr>
<tr>
<td>30</td>
<td>Diagonal 4</td>
</tr>
<tr>
<td>31</td>
<td>Extended Maratos</td>
</tr>
<tr>
<td>32</td>
<td>Quadratic QF1</td>
</tr>
<tr>
<td>33</td>
<td>Extended quadratic exponential EPI</td>
</tr>
<tr>
<td>34</td>
<td>DQDRTIC</td>
</tr>
<tr>
<td>35</td>
<td>NONSCOMP</td>
</tr>
<tr>
<td>36</td>
<td>Extended Freudenstein &amp; Roth</td>
</tr>
<tr>
<td>37</td>
<td>Extended White &amp; Holst</td>
</tr>
<tr>
<td>38</td>
<td>Raydan 1</td>
</tr>
<tr>
<td>39</td>
<td>Extended Tridiagonal 1</td>
</tr>
<tr>
<td>40</td>
<td>Extended Cliff</td>
</tr>
<tr>
<td>41</td>
<td>Extended Trigonometric</td>
</tr>
<tr>
<td>42</td>
<td>Extended Beale</td>
</tr>
<tr>
<td>43</td>
<td>Generalized Tridiagonal 1</td>
</tr>
<tr>
<td>44</td>
<td>Generalized PSCI</td>
</tr>
<tr>
<td>45</td>
<td>Extended PSCI</td>
</tr>
<tr>
<td>46</td>
<td>Extended Powell</td>
</tr>
<tr>
<td>47</td>
<td>BDQRTIC</td>
</tr>
<tr>
<td>48</td>
<td>FLETCBV3</td>
</tr>
<tr>
<td>49</td>
<td>FLETCHCR</td>
</tr>
<tr>
<td>50</td>
<td>FREUROTH</td>
</tr>
<tr>
<td>51</td>
<td>GENHUMPS</td>
</tr>
</tbody>
</table>

Table 1: Continued.

<table>
<thead>
<tr>
<th>Number</th>
<th>Function name</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>NONDIA</td>
</tr>
<tr>
<td>53</td>
<td>NONDQUAR</td>
</tr>
<tr>
<td>54</td>
<td>SROSENBR</td>
</tr>
<tr>
<td>55</td>
<td>TQUARTIC</td>
</tr>
<tr>
<td>56</td>
<td>Extended Penalty</td>
</tr>
</tbody>
</table>

We can find from Figure 1 that the SDYCG is similar to the CG-DESCENT when \( n = 10^2 \) because their curves crosses each other. The predominance of the SDYCG appears in Figure 2 when \( n = 10^3 \). If the test dimension is chosen as \( n = 10^4 \), the SDYCG is better than the CG-DESCENT from the fact that its curve is almost completely above that of the CG-DESCENT in Figure 3.
Furthermore, we are interested in the robustness of our SDYCG algorithm. Ten problems are selected from CUTEr to be tested. The numerical results listed in Table 2 are obtained by changing the initial iteration point every time. “\(x_0\)” represents the standard initial iteration point of the problem; “iter.” and “time(s)” indicate the iterative number and time (in seconds).

The conclusion that can be drawn is that the SDYCG is a robust algorithm and it may be capable of solving large-scale nonlinear unconstrained optimization problem.

5. Conclusions

We propose a new spectral conjugate gradient method for nonlinear unconstrained optimization. This method, which we call the SDYCG, is built based on the Dai-Yuan conjugate gradient method. A new spectral choice is provided in the search direction. Numerical results show that the SDYCG is comparable with the CG-DESCENT. The SDYCG algorithm may be capable of solving large-scale nonlinear unconstrained optimization problems.
Table 2: Test of the robustness of SDYCG.

<table>
<thead>
<tr>
<th>Number</th>
<th>Function name</th>
<th>$x_0$</th>
<th>$10 \times x_0$</th>
<th>$100 \times x_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DQDRTIC</td>
<td>245</td>
<td>0.3436</td>
<td>290</td>
</tr>
<tr>
<td>2</td>
<td>QUARTC</td>
<td>1</td>
<td>0.0012</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>Diagonal 6</td>
<td>5</td>
<td>0.0176</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Extended DENSCHNB</td>
<td>13</td>
<td>0.0110</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Extended DENSCHNF</td>
<td>22</td>
<td>0.0636</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>LIARWHD</td>
<td>12</td>
<td>0.0186</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>EDENSCH</td>
<td>30</td>
<td>0.1897</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>EG2</td>
<td>9</td>
<td>0.0404</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>ENGVAL1</td>
<td>28</td>
<td>0.0893</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>FLETCBV2</td>
<td>53</td>
<td>0.1080</td>
<td>53</td>
</tr>
</tbody>
</table>

Conflict of Interests
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (11471159), the Natural Science Foundation of Jiangsu Province (BK20141409), and the Foundation of Education Department of Anhui Province (2014jyxm161).

References