Journal of King Saud University – Science 30 (2018) 561–567

Contents lists available at ScienceDirect

Journal of King Saud University – Science

journal homepage: www.sciencedirect.com

A new modified deflected subgradient method

Rachid Belgacem [a,*], Abdessamad Amir [b]

[a] Université Hassiba Benbouali de Chlef, Chlef 02000, Algeria
[b] Universié Abdelhamid Ibn Badis de Mostaganem, Mostaganem 02700, Algeria

ARTICLE INFO

Article history:
Received 19 April 2017
Accepted 13 August 2017
Available online 22 August 2017

MSC:
90C26
90C10
90C27
90C06

Keywords:
Integer linear programming
Subgradient method
Nonsmooth optimization
Travelling salesman problem

ABSTRACT

A new deflected subgradient algorithm is presented for computing a tighter lower bound of the dual problem. These bounds may be useful in nodes evaluation in a Branch and Bound algorithm to find the optimal solution of large-scale integer linear programming problems. The deflected direction search used in the present paper is a convex combination of the Modified Gradient Technique and the Average Direction Strategy. We identify the optimal convex combination parameter allowing the deflected subgradient vector direction to form a more acute angle with the best direction towards an optimal solution. The modified algorithm gives encouraging results for a selected symmetric travelling salesman problem (TSPs) instances taken from TSPLIB library.

1. Introduction

In this paper we consider the following integer linear program:

$$(IP)\begin{cases} & z^* = \min \top x \\ \text{s.t.} & A_1 x \leqslant b_1 \\ & x \in X = \{x \in \mathbb{Z}^n : A_2 x \leqslant b_2\}, \end{cases} \tag{1}$$

where x is an $n \times 1$ vector, $\mathbb{Z}^n$ is the set of integers, c, b_1, b_2, A_1 and A_2 are $n \times 1$, $m \times 1$, $k \times 1$, $m \times n$ and $k \times n$ matrices, respectively. We assume that the problem (IP) is feasible and that X is a bounded and finite set. The problem (IP) is called the "*primal problem*" and z^* is called the "*primal optimal value*". The constraints $A_2 x \leqslant b_2$ are generally called the easy constraints, in the sense that an integer linear program with only these constraints is easy to solve. Lagrangian duality (Bazaraa and Sherali, 1981) is the most computationally useful idea for solving hard integer programs. The Lagrangian dual problem is obtained via Lagrangian relaxation approach (Fisher, 1985), where the constraints $A_1 x \leqslant b_1$, which are called the "*complicated constraints*", are relaxed by introducing a multiplier vector $\lambda \in \mathbb{R}^m_+$, called "*Lagrangian multiplier*". The Lagrangian relaxation problem is formulated as follows:

$$(RP)\begin{cases} w(\lambda) = \min \ c^\top x + \lambda^\top (A_1 x - b_1) \\ \text{s.t.} \quad x \in X, \end{cases} \tag{2}$$

It is easy to prove that $w(\lambda) \leqslant z$ for all $\lambda \geqslant 0$ (weak duality (Bazaraa et al., 2006)). The best choice for λ would be the optimal solution of the following problem, called the dual problem:

$$(D)\begin{cases} w^* = \max \ w(\lambda) \\ \lambda \geqslant 0. \end{cases} \tag{3}$$

With some suitable assumptions, the dual optimal value w^* is equal to z^* (strong duality (Bazaraa et al., 2006)). In general, w^* provides a tighter lower bound of z^*. These bounds may be useful in nodes evaluation in exact methods such as Branch and Bound algorithm to find the optimal solution of (IP). The function $w(\lambda)$ is continuous and concave but non-smooth. The most widely adopted method for solving the dual problem is the subgradient optimization, see for instance Polyak (1967), Shor (1985), Nedic and Bertsekas (2010), Nesterov (2014) and Hu et al. (2015). The

* Corresponding author.
E-mail addresses: belgacemrachid02@yahoo.fr (R. Belgacem), amirabdessamad@yahoo.fr (A. Amir).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksus.2017.08.004
1018-3647/

pure subgradient optimization method is an iterative procedure that can be used to solve the problem of maximizing (minimizing) a non-smooth concave (convex) function $w(\lambda)$ on a closed convex set Ω. This procedure is summarized in Algorithm 1, and it is used in various fields in science and engineering (Sra et al., 2012).

Algorithm 1 (Based Subgradient Algorithm).

1. Choose an initial point $\lambda_0 \in \Omega$.
2. Construct a sequence of points $\left(\lambda^k\right)_k \subset \Omega$ which eventually converges to an optimal solution using the rule $\lambda^{k+1} = P_\Omega\left(\lambda^k + t_k s^k\right)$, where $P_\Omega(.)$ is a projection operator on the set Ω and t_k is a positive scalar called step length such that
$$t_k = \delta_k \frac{w^* - w^k}{\|s^k\|^2}, \tag{4}$$
where $w^k = w(\lambda^k)$ is the dual function at the current iteration, $\delta_k \in]0, 2[$ and s^k is a subgradient of the function w at λ^k.
3. Replace k by $k+1$ and repeat the process until some stopping criteria.

In the context of Lagrangian relaxation, computing the subgradient direction s^k and the projection $P_\Omega\left(\lambda^k + t_k s^k\right)$ $(\Omega = \mathbb{R}^m_+)$ is a relatively easy problem. Since the subgradient s^k is not necessarily a descent direction, the step-length rule (4) differs from those given in the area of descent methods. In fact, this choice assures the decreasing of the subsequence $\left(\left\|\lambda^k - \lambda^*\right\|\right)_k$ as well as the convergence of $\left(\lambda^k\right)_k$ to λ^*. However, it is impossible to know in advance the value of w^* for most problems. To this end, the most effective way is to use the variable target value methods developed in Kim et al. (1990), Fumero (2001) and Sherali et al. (2000).

Another challenge in subgradient optimization is the choice of direction search that affects the computational performance of the algorithm. It is known that choosing the subgradient direction s^k, leads to the zigzagging phenomenon that might cause slow the procedure to crawl towards optimality (Bazaraa et al., 2006). To overcome this situation, in the spirit of conjugate gradient method (Nocedal and Wright, 2006; Fletcher and Reeves, 1964), we can adopt a direction search that deflects the subgradient pure direction. Accordingly, the direction search d^k at λ^k is computed as:

$$d^k = s^k + \Psi_k d^{k-1}, \tag{5}$$

where $\Psi_k \geqslant 0$ is a deflection parameter, s^k is a subgradient of the function w at λ^k and d^{k-1} is the previous direction $(d^0 = 0)$. Then, the new iteration is computed as:

$$\lambda^{k+1} = P_\Omega\left(\lambda^k + t_k d^k\right). \tag{6}$$

Some promising deflection algorithms of this type are the Modified Gradient Technique (MGT) (Camerini et al., 1975) and the Average Direction Strategy (ADS) (Sherali and Ulular, 1989). The MGT method was found to be superior to the pure subgradient method when used in concert with a specially designed step-length selection rule. The deflection parameter Ψ_k^{MGT} is computed according to:

$$\Psi_k^{MGT} = \begin{cases} -\eta_k \frac{s^k d^{k-1}}{\|d^{k-1}\|^2} & \text{if } s^k d^{k-1} < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

where $0 < \eta_k \leqslant 2$. With this choice of the deflection parameter, the direction becomes:

$$d^k_{MGT} = s^k + \Psi_k^{MTG} d^{k-1}. \tag{8}$$

The ADS strategy recommends to make the deflection at each iteration point by choosing the direction search which simply bisects the angle between the current subgradient s^k and the previous direction search d^{k-1}. To get this direction, the deflection parameter is computed according to:

$$\Psi_k^{ADS} = \frac{\|s^k\|}{\|d^{k-1}\|}. \tag{9}$$

With this choice of the deflection parameter, the direction becomes:

$$d^k_{ADS} = s^k + \Psi_k^{ADS} d^{k-1}. \tag{10}$$

Nowadays, the deflected subgradient method remains an important tool for nonsmooth optimization problems, especially for linear integer programming, due to its simple formulation and low storage requirement. In this paper, we present a new deflected direction search as a convex combination of the direction d^k_{MGT} (8) and the direction d^k_{ADS} (10). Our main result is the identification of the convex combination parameter which forces the algorithm to have a better deflection search than those given in the pure subgradient, MGT and ADS. For a numerical comparison of our approach and the two concurrent techniques MGT and ADS, we opted for the Travelling Salesman Problem (TSP) where its importance comes from the richness of its application and the fact that it is a typical of other problems of combinatorial optimization (Diaby and Karwan, 2016; El-Sherbeny, 2010).

The remainder of the paper is organized as follows: in Section 2, we describe our deflected subgradient method with convergence analysis. The computational tests, conducted on the Lagrangian relaxation of TSP of different sizes are described in Section 3. In Section 4 we conclude the paper.

2. A new modified deflected subgradient method

In this section, we present a new modified deflected subgradient method (*NMDS*) which determines the direction search as follows:

$$d^k = (1 - \alpha_k) d^k_{MGT} + \alpha_k d^k_{ADS}, \quad \alpha_k \in (0, 1). \tag{11}$$

We then obtain the following deflection parameter:

$$\Psi_k = \begin{cases} \frac{-\eta_k(1-\alpha_k) s^k d^{k-1} + \alpha_k \|s^k\| \|d^{k-1}\|}{\|d^{k-1}\|^2} & \text{if } s^k d^{k-1} < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{12}$$

hence $d^k = s^k + \Psi_k d^{k-1}$.

Algorithm 2 (The Deflected Subgradient Algorithm).

1. (Initialization): Choose a starting point $\lambda^0 \in \Omega = \mathbb{R}^m_+$, let $d^0 = 0$ and $k = 0$.
2. Determine a subgradient $s^k \in \partial w(\lambda^k)$ and compute
$$d^k = s^k + \Psi_k d^{k-1},$$
$$\lambda^{k+1} = P_\Omega(\lambda^k + t_k d^k),$$
where Ψ_k is given by relation (12) and t_k will be specified later.
3. Replace k by $k+1$ if a stopping condition is not yet met and return to step 2.

Consider the deflected subgradient method algorithm given in Algorithm 2. The following proposition extends important properties of the subgradient vector s^k and the deflected subgradient direction d^k_{MGT} to the new deflected subgradient direction d^k (11). With a best choice of the parameter t_k, d^k make an acute angle with $\lambda^* - \lambda^k$ and d^{k-1}. We also get the decreasing of the subsequence $(\|\lambda^* - \lambda^k\|)_k$.

Proposition 1. *Let $s^k \in \partial w(\lambda^k)$, d^k be the new deflected subgradient direction given by (11) and (12) and let $\{\lambda^k\}$ be the sequence of iterations generated by the deflected subgradient scheme. If we take $0 < \eta_k \leqslant 2$ and the stepsize t_k to satisfy*

$$0 < t_k < \frac{w^* - w(\lambda^k)}{\left\|d^k\right\|^2}, \quad \forall k = 0, 1, 2, \dots \tag{13}$$

then,

1. $d^{k-1}\left(\lambda^* - \lambda^k\right) > 0, \quad (14)$
2. $\left\|\lambda^{k+1} - \lambda^*\right\| < \left\|\lambda^k - \lambda^*\right\|, \quad (15)$
3. $d^k d^{k-1} \geqslant 0.$

for all k where λ^k are non optimal points and λ^* is an optimal solution.

Proof.

1. The proof is established by induction on k. Since we start with $d^0 = 0$, the case $k = 1$ is trivial. Now, assume that we have

$$d^{k-2}\left(\lambda^* - \lambda^{k-1}\right) \geqslant 0, \ \forall k \geqslant 2, \tag{16}$$

and let us establish (14) at iteration k. Using the definition of d^{k-1}, we obtain that

$$\begin{aligned} d^{k-1}\left(\lambda^* - \lambda^k\right) &= d^{k-1}\left(\lambda^* - \lambda^{k-1} + \lambda^{k-1} - \lambda^k\right) \\ &= d^{k-1}\left(\lambda^* - \lambda^{k-1}\right) + d^{k-1}\left(\lambda^{k-1} - \lambda^k\right) \\ &= \left(s^{k-1} + \Psi_{k-1} d^{k-2}\right)\left(\lambda^* - \lambda^{k-1}\right) + d^{k-1}\left(\lambda^{k-1} - \lambda^k\right) \\ &= s^{k-1}\left(\lambda^* - \lambda^{k-1}\right) + \Psi_{k-1} d^{k-2}\left(\lambda^* - \lambda^{k-1}\right) + d^{k-1}\left(\lambda^{k-1} - \lambda^k\right). \end{aligned}$$

Furthermore, from the concavity of the function $w(\cdot)$, the induction hypothesis (16) and the inequalities in (13), we get

$$d^{k-1}\left(\lambda^* - \lambda^k\right) \geqslant \left(w^* - w(\lambda^{k-1})\right) - d^{k-1}\left(\lambda^k - \lambda^{k-1}\right) \tag{17}$$

Since the vector $\lambda^k - P_\Omega(\lambda^{k-1} + t_{k-1} d^{k-1})$ is perpendicular to the supporting hyperplane of $\Omega = \mathbb{R}^m_+$ at λ^k, the angle at λ^k is obtuse (see Fig. 1). We deduce that

$$d^{k-1}(\lambda^k - (\lambda^{k-1} + t_{k-1} d^{k-1})) \leqslant 0,$$

which is equivalent to

$$-d^{k-1}(\lambda^k - \lambda^{k-1}) \geqslant -t_{k-1}\|d^{k-1}\|^2 \tag{18}$$

Substituting (18) in (17) we obtain

$$d^{k-1}\left(\lambda^* - \lambda^k\right) \geqslant \left(w^* - w(\lambda^{k-1})\right) - t_{k-1}\|d^{k-1}\|^2 > 0. \tag{19}$$

2. We have

$$\begin{aligned} \left\|\lambda^* - \lambda^{k+1}\right\|^2 &= \left\|\lambda^* - P_\Omega(\lambda^k + t_k d^k)\right\|^2 \\ &\leqslant \left\|\lambda^* - \lambda^k - t_k d^k\right\|^2 \\ &= \left\|\lambda^* - \lambda^k\right\|^2 + t_k^2\left\|d^k\right\|^2 - 2t_k d^k\left(\lambda^* - \lambda^k\right) \\ &= \left\|\lambda^* - \lambda^k\right\|^2 + t_k\left[t_k\left\|d^k\right\|^2 - 2d^k\left(\lambda^* - \lambda^k\right)\right]. \end{aligned}$$

From the concavity of w, the inequalities in (13) and by applying (14) in Proposition 1, we get the following relations, respectively:

$$\begin{aligned} t_k\left\|d^k\right\|^2 &\leqslant w^* - w(\lambda^k) \leqslant 2\left(w^* - w(\lambda^k)\right) \\ &\leqslant 2s^k\left(\lambda^* - \lambda^k\right) \\ &\leqslant 2d^k_{MGT}\left(\lambda^* - \lambda^k\right) \\ &\leqslant 2d^k\left(\lambda^* - \lambda^k\right). \end{aligned}$$

It follows, that

$$t_k\left\|d^k\right\|^2 - 2d^k\left(\lambda^* - \lambda^k\right) \leqslant 0.$$

Consequently,

$$\left\|\lambda^* - \lambda^{k+1}\right\| < \left\|\lambda^* - \lambda^k\right\|.$$

3. If $s^k d^{k-1} \geqslant 0$ then $d^k = s^k$ and hence the claim follows. Thus, consider the case $s^k d^{k-1} < 0$. we have then

$$\begin{aligned} d^k d^{k-1} &= \left(s^k + \Psi_k d^{k-1}\right) d^{k-1} \\ &= s^k d^{k-1} + \Psi_k\left\|d^{k-1}\right\|^2 \\ &= s^k d^{k-1} - \eta_k(1 - \alpha_k) s^k d^{k-1} + \alpha_k\|s^k\|\left\|d^{k-1}\right\| \\ &= (-\alpha_k + \eta_k\alpha_k(1 - \alpha_k) + \alpha_k)\|s^k\|\left\|d^{k-1}\right\| \\ &= \eta_k\alpha_k(1 - \alpha_k)\|s^k\|\left\|d^{k-1}\right\| \\ &\geqslant 0. \end{aligned}$$

This completes the prove. □

The importance of Proposition 1 lies on the fact that choosing the deflection parameter Ψ_k using the rule (12) with $0 < \eta_k \leqslant 2$ forces the current deflected subgradient direction to form always an acute angle with the previous step direction and hence, this method eliminates the zigzagging of the pure subgradient procedure. Note that the choice of the vector of deflected direction d^k_{MGT} is always at least as good as the direction of the subgradient vector s^k. If $1 \leqslant \eta_k \leqslant 2$, then $d^k_{MGT} d^{k-1}_{MGT} \geqslant 0$ (Camerini et al., 1975).

The theorem below shows that with a particular choice of the convex combination parameter α_k and the parameter η_k, the deflected subgradient vector direction d^k is always at least as good as the direction d^k_{MGT} in a sense that d^k can form a more acute angle with the best direction towards an optimal solution than d^k_{MGT} does (see Fig. 2), which enhances the speed of convergence. The two lemmas below are necessary for the proof of our principal result.

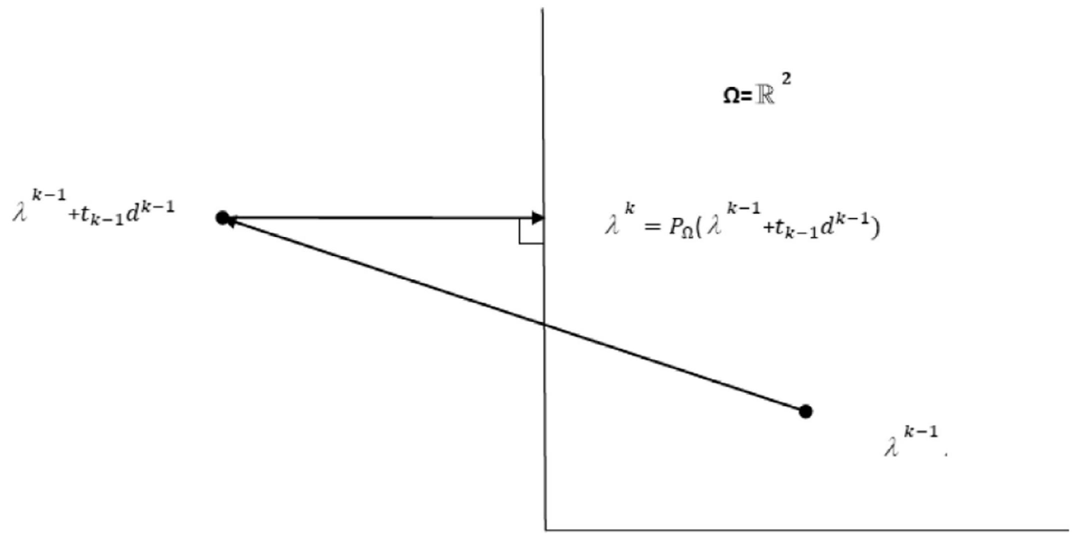

Fig. 1. Illustration in a two-dimensional case.

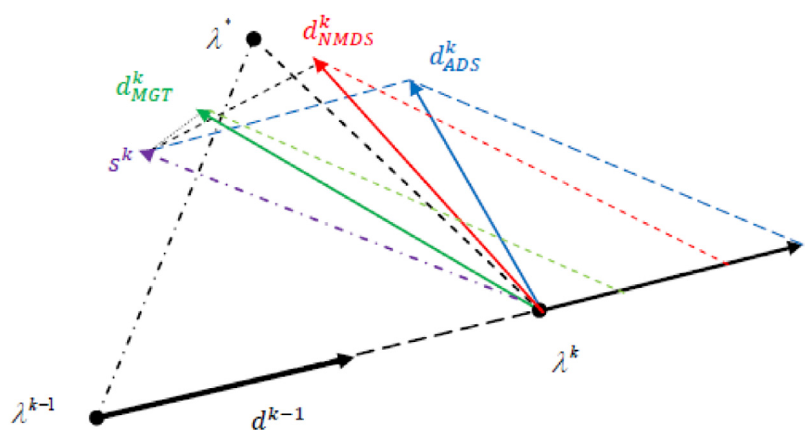

Fig. 2. Case where s^k is deflected since it has formed an obtuse angle with d^{k-1} and the direction d^k_{NMDS} is better as compared to other directions d^k_{ADS} and d^k_{MGT}.

Lemma 1. *Let $s^k \in \partial w(\lambda^k)$ and set $\alpha_k = -\cos(s^k, d^{k-1})$ if $s^k d^{k-1} < 0$. With the assumption in (13) and letting*

$$0 < \eta_k \leqslant \frac{1}{2-\alpha_k}, \tag{20}$$

then

$$d^k\left(\lambda^* - \lambda^k\right) \geqslant d^k_{MGT}\left(\lambda^* - \lambda^k\right) \text{for all} k. \tag{21}$$

Proof. Using (8), (10) and (11) we obtain the following relation:

$$\begin{aligned} & d^k\left(\lambda^* - \lambda^k\right) - d^k_{MGT}\left(\lambda^* - \lambda^k\right) \\ & = \alpha_k d^k_{MGT}\left(\lambda^* - \lambda^k\right) + (1-\alpha_k) d^k_{ADS}\left(\lambda^* - \lambda^k\right) - d^k_{MGT}\left(\lambda^* - \lambda^k\right) \\ & = (1-\alpha_k)\left[d^k_{ADS}\left(\lambda^* - \lambda^k\right) - d^k_{MGT}\left(\lambda^* - \lambda^k\right)\right] \\ & = (1-\alpha_k)\left[\left(s^k + \Psi^{ADS}_k d^{k-1}\right)\left(\lambda^* - \lambda^k\right) - \left(s^k + \Psi^{MGT}_k d^{k-1}\right)\left(\lambda^* - \lambda^k\right)\right] \\ & = (1-\alpha_k)\left(\Psi^{ADS}_k - \Psi^{MGT}_k\right) d^{k-1}\left(\lambda^* - \lambda^k\right). \end{aligned}$$

From (7) and (9) it follows that:

$$\Psi^{ADS}_k - \Psi^{MGT}_k = \frac{\|s^k\|\left\|d^{k-1}\right\|}{\left\|d^{k-1}\right\|^2}\left[1 + \eta_k \cos(s^k, d^{k-1})\right]. \tag{22}$$

Using the last equality and applying Proposition 1 we get (21). □

Lemma 2. *Under the same hypothesis of Lemma 1, we have*

$$\left\|d^k\right\| \leqslant \left\|d^k_{MGT}\right\| \quad \text{for all } k. \tag{23}$$

Proof. If $s^k d^{k-1} \geqslant 0$ then $\Psi_k = 0$ and hence (23) obviously holds and one simply has $d^k = d^k_{MGT}$. In the case where $s^k d^{k-1} < 0$, then:

$$\begin{aligned} \left\|d^k\right\|^2 - \left\|d^k_{MGT}\right\|^2 & = \left\|s^k + \Psi_k d^{k-1}\right\|^2 - \left\|s^k + \Psi^{MGT}_k d^{k-1}\right\|^2 \\ & = \left(\Psi_k^2 - \left(\Psi^{MGT}_k\right)^2\right)\left\|d^{k-1}\right\|^2 + 2\left(\Psi_k - \Psi^{MGT}_k\right) s^k d^{k-1} \\ & = \left(\Psi_k - \Psi^{MGT}_k\right)\left[\left(\Psi_k + \Psi^{MGT}_k\right)\left\|d^{k-1}\right\|^2 + 2 s^k d^{k-1}\right]. \end{aligned}$$

Since $s^k d^{k-1} = \|s^k\|\left\|d^{k-1}\right\| cos(s^k, d^{k-1})$ one finds:

$$\left\|d^k\right\|^2 - \left\|d^k_{MGT}\right\|^2 = \alpha_k^2 \|s^k\|^2(-\eta_k\alpha_k + 1)[\eta_k(2-\alpha_k) - 1].$$

By the choice of η_k such that we obtain

$$\left\|d^k\right\| \leqslant \left\|d^k_{MGT}\right\|. \quad \square$$

Theorem 1. *Under the same hypothesis of Lemma 1, we have*

$$(i)\ \frac{d^k(\lambda^*-\lambda^k)}{\|d^k\|} \geqslant \frac{d^k_{MGT}(\lambda^*-\lambda^k)}{\|d^k_{MGT}\|}. \tag{24}$$

(ii) If the vectors d^k and d^k_{MGT} form an angle θ_{d^k} and $\theta_{d^k_{MGT}}$, respectively, with the vector $\lambda^* - \lambda^k$, then

$$0 \leqslant \theta_{d^k} \leqslant \theta_{d^k_{MGT}} \leqslant 90^\circ.$$

Proof. Direct consequence of the previous lemmas. □

3. Computational results

The proposed algorithm has been applied to one of the standard integer linear programming problems in the field of operational research, namely the symmetric travelling salesman problem (TSPs). The travelling salesman problem is a classical NP-Hard combinatorial optimization problem (Garey and Johnson, 1990). It can be formulated as follows: giving a set of cities, and distances between them, the goal is to find the shortest tour visiting every city only once and returning to the starting city. More details on this problem may be found in Lawler et al., 1985. The TSPs can be stated as follows (where c_{ij} is the cost of link (i,j)):

$$\min \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} c_{ij}x_{ij}, \tag{25}$$

subject to

$$\sum_{j=1}^{m} x_{ij} = 1, \quad i = 1,\dots,m, \tag{26}$$

$$\sum_{i=1}^{m} x_{ij} = 1, \quad j = 1,\dots,m, \tag{27}$$

$$\sum_{i\in Q}^{m}\sum_{j\in Q}^{m} x_{ij} \leqslant |Q| - 1, \quad \forall Q : 2 \leqslant |Q| \leqslant m-2, \tag{28}$$

$$x_{ij} = 0 \text{ or } 1, \quad i,j = 1,\dots,m, \tag{29}$$

where $Q \subset \{1,\dots,m\}$. Letting X be the set of all 1-trees (Held and Karp, 1970), the subtour constrains (28) can be eliminated by insisting that a vector x satisfying the constraints (26), (27) and (29) must also belong to X. In particular for the symmetric case, constraints (26) and (27) can be replaced by (30) leading to the following equivalent formulation of the TSPs:

$$\min \sum_{i=1}^{m} \sum_{j=1 j\neq i}^{m} c_{ij}x_{ij},$$

subject to

$$\sum_{\substack{j=1 \\ j \neq i}}^{m} x_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^{m} x_{ji} = 2 \quad \text{for } i = 1,\dots,m, \tag{30}$$

$$x \in X.$$

From this, one obtains the following dual function, which has to be maximized:

$$w(\lambda) = \min\left\{ \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} (c_{ij} + \lambda_i + \lambda_j)x_{ij}, x \in X \right\} - 2\sum_{i=1}^{m}\lambda_i, \tag{31}$$

where $\lambda \in \mathbb{R}^m$ is the vector of Lagrangian multipliers.

Given a vector $\bar{\lambda}$, if $\bar{x}$ optimizes $w(\bar{\lambda})$, then a vector $\bar{s}$ whose ith component

$$\bar{s_i} = \left(\sum_{\substack{j=1 \\ j \neq i}}^{m} x_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^{m} x_{ji} - 2 \right) \tag{32}$$

is a subgradient of $w(\lambda)$ at $\bar{\lambda}$ (Bazaraa et al., 2006; Held et al., 1974).

To validate the feasibility and effectiveness of the proposed approach, we have applied it on some TSPs instances taken from TSPLIB.[1] The proposed algorithm, MGT and ADS were implemented in Matlab and executed on an Intel(R) Core(TM) i517U CPU @ 1.70 GHz 1.70 GHz RAM 4.00GO.

For all symmetric instances and for a fair comparison between the three algorithms, the following parameter settings were chosen:

- The same initial multiplier $\lambda^1 = (0,0,\dots,0)^T$ was used for the three algorithms.
- The stop conditions are the maximum number of iteration $iterMAX = 1000$, or $| w^* - w(\lambda^k) | \leqslant \varepsilon$, where ε is a small tolerance $(\varepsilon = 10^{-2})$.
- The step size t_k is defined according to formula (4).
- The parameter δ_k follows the Held et al. (1974) suggestion, that makes $0 < \delta_k \leqslant 2$, beginning with $\delta_k = 2$. If after 20iterations $w(\lambda^k)$ not increases, δ_k is updated to $\delta_k = \frac{\delta_k}{2}$.
- For MGT algorithm, as mentioned in Camerini et al. (1975), the use of $\eta_k = 1.5$ is recommended and its intuitive justification together with computational results are also given, which indicates that in practise, the performance of MGT strategy is superior to that of the pure subgradient algorithm.
- For our algorithm the value of η_k depends on the optimal convex combination parameter α_k as indicated in Lemma 1, where $\alpha_k = -\cos(s^k, d^{k-1})$ if $s^k d^{k-1} < 0$. We used $\eta_k = \frac{1}{2-\alpha_k} - \varepsilon$, where ε is an arbitrary small value.

Table 1 shows the experimental results obtained by: MGT strategy, ADS strategy and by applying our NMDS algorithm proposed in this paper with 11 symmetric benchmark instances between n = 6 and n = 101 vertices taken from TSPLIB. For the three strategies, the duality GAP for these 11 examples is null. However, always NMDS algorithm outperforms the others in number of iterations and execution time. Table 2 gives the computational results for 19 symmetric benchmark instances between 131 and 3056 vertices. This table also shows that our algorithm gives near optimal results for several instances. The column headers are as follows:

- Name: Indicate the instance name.
- n: Indicate the problem size.
- w^*: The best known optimal solution.
- LB: The best value (lower bound) obtained by each strategy.
- $Iter$: Number of iterations at which the best value $LB$is obtained (limited to 1000).

[1] http://comopt.ifi.uniidelberg.de/software/TSPLIB95/.

Table 1
Computational results for $6 \leqslant n \leqslant 101$.

Name	n	w^*	Strategy											
			Our method				ADS strategy				MGT strategy			
			LB	GAP	Iter	CPU	LB	GAP	Iter	CPU	LB	GAP	Iter	CPU
tsp6	6	207	207	0	2	0.027862 s	207	0	2	0.032690 s	207	0	2	0.028533 s
tsp7	7	106.4	106.4	0	3	0.028651 s	106.4	0	3	0.028819 s	106.4	0	3	0.029059 s
tsp8	8	100	100	0	3	0.032122 s	100	0	3	0.032360 s	100	0	3	0.033069 s
tsp10	10	378	10	0	3	0.077130 s	378	0	18	0.085784 s	378	0	15	0.099755 s
ulysses16	16	68.59	68.59	0	16	0.077255 s	68.59	0	30	0.099056 s	68.59	0	26	0.090043 s
gr21	21	2707	2707	0	19	0.077098 s	2707	0	26	0.086654 s	2707	0	22	0.080609 s
ulysses22	22	70.13	70.13	0	21	0.219532 s	70.13	0	70	0.467253 s	70.13	0	28	0.253197 s
tsp33	33	10861	10861	0	18	0.27955 s	10861	0	19	0.303524 s	10861	0	22	0.304997 s
eil76	76	538	538	0	12	0.728258 s	538	0	23	1.285156 s	538	0	18	1.071510 s
rat99	99	1211	1211	0	19	1.817552 s	1211	0	20	1.902072 s	1211	0	34	3.111573 s
eil101	101	629	629	0	13	1.253360 s	629	0	16	1.523116 s	629	0	15	1.422569 s

Table 2
Computational results for $131 \leqslant n \leqslant 3056$.

Name	n	w^*	Strategy											
			Our method				ADS strategy				MGT strategy			
			LB	GAP	Iter	CPU	LB	GAP	Iter	CPU	LB	GAP	Iter	CPU
xq131	131	564	555.96	0.0159	350	38.134934 s	555.52	0.0167	350	39.009219 s	555.64	0.0165	350	42.417376 s
ch150	150	6528	6498.3	0.0045	193	57.562539 s	6463.2	0.0099	200	58.301731 s	6490.4	0.0057	202	59.548258 s
tsp237	237	1019	1004.8	0.0139	190	76.942993 s	1002.4	0.0162	200	79.046667 s	1003.9	0.0148	197	77.288710 s
a280	280	2579	2569.8	0.0035	251	128.401733 s	2568.3	0.0041	251	134.245723 s	2569.8	0.0035	251	129.376815
linhp318	318	41345	41345	0	188	135.792252 s	41330	0.0003	200	148.125054 s	41337	0.0001	195	140.734334 s
pbk411	411	1343	1443	0	250	328.562376 s	1338.3	0.0034	250	376.626413 s	1338.6	0.0032	250	235.626147 s
pbn423	423	1365	1365	0	251	344.519801 s	1360.3	0.0034	251	349.933052 s	1361.5	0.0025	251	370.504702 s
pbm436	436	1443	1423.9	0.0132	207	241.663478 s	1422.3	0.0143	251	334.741944 s	1421.4	0.0149	207	255.055079 s
rat575	575	6773	6721.2	0.0076	350	1287.476161 s	6716	0.0084	350	1266.572875 s	6718.9	0.0079	350	1138.613357 s
rbx711	711	3115	3099.6	0.0049	500	1648.379114 s	3094.3	0.0066	500	1668.377875 s	3096.2	0.0060	500	1701.099452 s
rat783	783	8806	8652.2	0.0197	121	496.527482 s	8609.5	0.0223	180	744.089813 s	8642.8	0.0185	131	592.200505 s
dkg813	813	3199	3164.7	0.0107	200	126.290411 s	3147.5	0.0203	200	129.330427 s	3154.7	0.0138	200	138.698584 s
pbd984	984	2797	2797	0	500	2216.458302,s	2772	0.0089	500	4059.009584 s	2774.3	0.0081	500	3394.524016,s
xit1083	1083	3558	3551.1	0.0019	550	3394.524016 s	3525.1	0.0092	600	3394.524016 s	3549.6	0.0023	600	3726.7668 s
dka1376	1376	4666	4666	0	500	7321.695698 s	4659.6	0.0013	500	14632.080221 s	4662	0.0008	500	14704.797585 s
dja1436	1436	5257	5194.2	0.0437	500	10959.322745 s	5189.4	0.0128	500	12014.302471 s	5184.4	0.0138	500	11486.802618 s
dcc1911	1911	6396	6367.9	0.0043	500	10602.013548 s	6338.3	0.0097	500	11742.144423 s	6357.2	0.0060	500	11127.0759858 s
djb2103	2103	6197	6197	0	430	13566.845650 s	6133.8	0.0101	500	14954.211430 s	6169.3	0.0044	500	15955.063763 s
pia3056	3056	8285	8285	0	500	51124.513213 s	8192	0.0112	500	59123.013454 s	8228.2	0.0068	500	58072.083303 s

- $GAP = \frac{w^* - LB}{w^*}$.
- *CPU*: Total computer time, in second for calculating the best value *LB*obtained by each strategy.

4. Conclusion

By identifying the optimal convex combination parameter, a new deflected direction is given as convex combination of the deflected direction of MGT and ADS. This direction, at each iteration reduces the zigzagging phenomenon and hence getting closer and faster to the optimal solution. The analysis studies are consistent with the numerical experiments. Moreover, this method can be used to improve convergence in the area of deflected subgradient method using augmented Lagrangian duality (Burachik and Kaya, 2010) and dual subgradient methods (Gustavsson et al., 2015). One can also follow Lim and Sherali (2006) and combine this method with a variable target technique in order to have a good performance. Finally, the subgradient method is usually used as subroutine in exact, heuristic and metaheuristic optimization, which justifies the large spectrum of applications of our approach.

Acknowledgements

The authors wish to express their sincere thanks to the editor and the reviewers for their constructive comments, which are very useful for improving this paper.

References

Bazaraa, M.S., Sherali, H.D., 1981. On the choice of step sizes in subgradient optimization. Eur. J. Oper. Res. 7, 380–388.
Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 2006. Non Linear Programming: Theory and Algorithms. Wiley-Interscience Series in Discrete Mathematics and Optimization.
Burachik, R.S., Kaya, C.Y., 2010. A deflected subgradient method using a general augmented Lagrangian duality with implications on penalty methods. In: Burachik, R., Yao, J.C. (Eds.), Variational Analysis and Generalized Differentiation in Optimization and Control. Springer Optimization and Its Applications, vol. 47. Springer, New York, NY.
Camerini, P.M., Fratta, L., Maffioli, F., 1975. On improving relaxation methods by modified gradient techniques. In: Balinski, M.L., Wolfe, P. (Eds.), Nondifferentiable Optimization. Math. Program. Studies, vol. 3. Springer, Berlin, Heidelberg.
Diaby, M., Karwan, M.H., 2016. Advances in combinatorial optimization: linear programming formulation of the travelling salesman and other hard combinatorial optimization problems.
El-Sherbeny, N.A., 2010. Vehicle routing with time windows: An overview of exact, heuristic and metaheuristic methods. J. King Saud Univ. Sci. 22, 123–131.
Fletcher, R., Reeves, C.M., 1964. Function minimization by conjugate gradients. Comput. J. 7, 149–154.
Fisher, M.L., 1985. An application oriented guide to Lagrangian relaxation. Interfaces 15, 10–21.
Fumero, F., 2001. A modified subgradient algorithm for Lagrangian relaxation. Comput. Oper. Res. 28, 33–52.
Garey, M.R., Johnson, D.S., 1990. Computers and Intractability: A Guide to the Theory of NP Completeness. W. H. Freeman & Co., New York, NY, USA.
Gustavsson, E., Patriksson, M., Strömberg, 2015. Primal convergence from dual subgradient methods for convex optimization. Math. Program. 150, 365–390.
Held, M., Karp, R.M., 1970. The travelling salesman problem and minimum spanning trees. Oper. Res. 18, 1138–1162.

Held, M.H., Wolfe, P., Crowder, H.D., 1974. Validation of subgradient optimization. Math. Program. 6, 62–88.
Hu, Y., Yang, X., Sim, C.-K., 2015. Inexact subgradient methods for quasi-convex optimization problems. Eur. J. Oper. Res. 240, 315–327.
Kim, S., Ahn, H., Cho, S.-C., 1990. Variable target value subgradient method. Math. Program. 49, 359–369.
Lawler, E.L., Lenstra, Jan Karel, Rinnooy Kan, A.H.G., Shmoys, D.B., 1985. The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization. John Wiley & Sons Inc.
Lim, C., Sherali, H.D., 2006. Convergence and computational analyses for Some variable target value and subgradient deflection methods. Comput. Optim. Appl. 34, 409–428.
Nedic, A., Bertsekas, D.P., 2010. The effect of deterministic noise in subgradient methods. Math. Program. 125, 75–99.
Nesterov, Y., 2014. Subgradient methods for huge-scale optimizations problems. Math. Program. 146, 275–297.
Nocedal, J., Wright, S.J., 2006. Numerical Optimization. Springer Series in Operations Research.
Polyak, B.T., 1967. A general method of solving extremum problems. Sov. Math. Dokl. 8, 593–597.
Sherali, H.D., Ulular, O., 1989. A primal-dual conjugate subgradient algorithm for specially structured linear and convex programming problems. Appl. Math. Optim. 20, 193–221.
Sherali, H.D., Choi, G., Tuncbilek, C.H., 2000. A Variable target value method for nondifferentiable optimization. Oper. Res. Lett. 26, 1–8.
Shor, N.Z., 1985. Minimization Methods for Non-Differentiable Functions. Springer, Berlin.
Sra, S., Nowozin, S., Wright, S.J., 2012. Optimization For Machine Learning. MIT Press.