Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 20, No. 4, pp. 781–808, 2024

DP OPTIMIZATION: OPTIMALITY CONDITIONS AND NUMERICAL METHODS

Adil M. Bagirov* and Julien Ugon†

Abstract: The aim of this paper is to study necessary and sufficient local and global optimality conditions and to develop local and global search algorithms for unconstrained and constrained piecewise linear programming problems. This aim is achieved using a difference of convex polyhedral (DP) functions representations of continuous piecewise linear functions. First, optimality conditions and numerical algorithms are studied for unconstrained DP programming problems. Then these results are extended to problems on minimizing piecewise linear functions subject to linear equalities and inequalities. Convergence of algorithms are studied and illustrative examples are presented.

Key words: *nonsmooth optimization, nonconvex optimization, DP optimization, optimality conditions, numerical methods*

Mathematics Subject Classification: *65K05, 90C25*

---

1 Introduction

Piecewise linear programming is about minimizing (or maximizing) piecewise linear functions subject to linear equalities and inequalities. Such problems arise in many applications. Piecewise linear (PWL) functions (including discontinuous) have been widely used for modeling different processes in economics [6, 21], circuit [49] and oil production systems [27]. They have been applied in solving network flow [3, 47], data mining [4, 22, 25, 35] and regression analysis [5, 48] problems. PWL functions have been used to design algorithms for solving various optimization problems such as mixed integer nonlinear programming [31], nonsmooth optimization [20, 24] and global optimization problems [47, 50]. Convex and concave PWL functions arise when one tries to use Lagrangian relaxation for solving optimization problems [18]. In addition, PWL functions might be present as a part of objective and/or constraint functions in more general optimization models.

Nonconvex PWL functions are used to approximate nonlinearities arising from factors such as economies of scale or complex technological processes. They naturally appear as cost functions of supply chain problems to model discounts for high volume and fixed charges. Other applications include production planning [19], operation planning of gas networks [38], and network flow problems [12].

*Corresponding author. This research by Dr. A. Bagirov was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (Project No. DP190100580).

†The research by Dr. J. Ugon was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (Project No. DP180100602).

 DOI: https://doi.org/10.61208/pjo-2024-008

The aim of this paper is to study necessary and sufficient local and global optimality conditions for nonconvex piecewise linear programming problems as well as to design numerical algorithms to locally and globally solve such problems. Optimality conditions and algorithms are studied for both unconstrained and constrained piecewise linear programming problems.

Results obtained in this paper are based on the notion of codifferentials described in [15]. These results can be obtained in more compact form using the notion of exhausters (see, for example, [2]). However, the use of codifferentials allows one to design ready to use numerical methods.

The paper is organized as follows. A brief overview of relevant work is given in Section 2. Section 3 contains some preliminaries on DP functions and their subdifferentials. Necessary and sufficient conditions for local optimality in unconstrained PWL optimization are given in Section 4 and such conditions for global optimality are presented in Section 5. Local search algorithms for minimizing DP functions are described in Section 6 and global search algorithms for such problems are given in Section 7. Illustrative examples are given in Section 8 and Section 9 concludes the paper.

2 Related work

First, we provide a brief overview of existing results on optimality conditions and numerical methods for PWL programming.

Convex PWL programming problems have been extensively investigated. Necessary and sufficient conditions for such problems can be obtained from optimality conditions for general nonsmooth convex programming problems using the subdifferential calculus. There exist several algorithms for minimization of convex PWL functions. Finite convergent algorithms are presented, for example, in [7, 10, 34]. An interior point method for convex PWL programming problems is introduced in [8].

Problems with nonconvex PWL objective and/or constraint functions have attracted less attention despite the fact that such problems have more practical applications than their convex counterparts. Although some results (for example, global optimality conditions from [28]) can be applied to these problems, necessary and sufficient global optimality conditions specifically for nonconvex PWL programming problems have not been studied extensively. Such conditions have been studied for problems where PWL functions are represented as a difference of two convex polyhedral (DP) functions. The paper [39] uses conjugate functions as well as codifferentials [15] to derive necessary and sufficient conditions for the unboundedness and the boundedness of DP functions and necessary and sufficient global optimality conditions for unconstrained DP problems. Global optimality conditions for DP functions in terms of codifferentials are also given in [17] and using these conditions the finite convergent modified codifferential method is designed which does not use the line search. In [45], the authors investigate various generalized subdifferentials for DP functions. The set of global minimizers of PWL functions is described using exhausters in [1].

Specialized algorithms for solving optimization problems involving nonconvex PWL functions were introduced in [14, 33, 43]. PWL problems can also be modeled as mixed integer programming (MIP) problems [11, 13, 32, 36, 37] and solved with a general purpose MIP solver. In [46], the authors formulate nonconvex PWL programming problems as MIP problems. In addition, they extend these formulations to problems with lower semicontinuous PWL objective functions.

MIP models for nonconvex PWL problems have been extensively studied, but existing comparisons [11, 32] only concentrate on the case in which the functions are separable (i.e. can be written as the sum of univariate functions). When a non-separable function is known analytically it can sometimes be converted into a separable one by algebraic manipulations [43]. However, this conversion might be undesirable for numerical reasons [12, 38]. Furthermore, in many applications the functions come from complicated simulation models and are not known analytically.

3 Preliminaries

In this section we briefly describe main concepts and definitions used in the paper. In what follows we denote by $\mathbb{R}^n$ n-dimensional Euclidean space, by $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$ the inner product of vectors u and v in $\mathbb{R}^n$ and by $\|\cdot\|$ the associated norm in $\mathbb{R}^n$. $S_1 = \{v \in \mathbb{R}^n : \ \|v\| = 1\}$ is the unit sphere centered at $0_n \in \mathbb{R}^n$, $B_\varepsilon(x) = \{y \in \mathbb{R}^n : \ \|y - x\| < \varepsilon\}$ is the open ball centered at $x \in \mathbb{R}^n$ with the radius $\varepsilon > 0$, "conv" denotes convex hull of a set.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. The set

$$\partial_c f(x) = \Big\{ v \in \mathbb{R}^n : \ f(y) \geq f(x) + \langle v, y - x\rangle \ \forall y \in \mathbb{R}^n \Big\}$$

is the subdifferential of f at $x \in \mathbb{R}^n$ [40]. Each vector $v \in \partial_c f(x)$ is called a subgradient of f at x. Given $\varepsilon > 0$, the ε-subdifferential of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ is defined as [40]:

$$\partial_\varepsilon f(x) = \Big\{ v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x\rangle - \varepsilon \ \forall y \in \mathbb{R}^n \Big\}.$$

A generalized directional derivative of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ with respect to a direction $d \in \mathbb{R}^n$ is:

$$f^0(x, d) = \limsup_{y \to x, \alpha \downarrow 0} \frac{f(y + \alpha d) - f(y)}{\alpha}.$$

The generalized subdifferential (or Clarke subdifferential) of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is defined as [9]:

$$\partial f(x) = \Big\{ v \in \mathbb{R}^n : \ f^0(x, d) \geq \langle v, d\rangle \ \ \forall d \in \mathbb{R}^n \Big\}.$$

For a convex function $f : \mathbb{R}^n \to \mathbb{R}$ one has $\partial f(x) = \partial_c f(x)$. Therefore from now on we use the notation $\partial f(x)$ for subdifferentials. The directional derivative of a function f at $x \in \mathbb{R}^n$ with respect to a direction $d \in \mathbb{R}^n$ is

$$f'(x, d) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha},$$

if this limit exists. A locally Lipschitz function f is called Clarke regular at $x \in \mathbb{R}^n$ if it is directionally differentiable and $f'(x, d) = f^0(x, d)$ for all $d \in \mathbb{R}^n$.

3.1 Piecewise linear functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is piecewise linear (PWL) if there is a finite set $f_1, \ldots, f_m$ of linear functions such that $f_i : \mathbb{R}^n \to \mathbb{R}$ and for each $x \in \mathbb{R}^n$, $f(x) = f_i(x)$ for some $i \in \{1, \ldots, m\}$. PWL functions appearing in applications can be continuous and discontinuous and in general, represented using IF-statements. Continuous PWL functions admit alternative representations such as difference of convex polyhedral (DP) functions and a max-min of affine functions [26].

In this paper, we assume that DP representations of PWL functions are known. Thus, we study the following optimization problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n, \end{cases} \tag{3.1}$$

where the objective function f is a continuous PWL function and is represented as $f(x) = f_1(x) - f_2(x)$, where

$$f_1(x) = \max_{i \in I_1} \langle a^{1i}, x \rangle - b_{1i}, \quad f_2(x) = \max_{j \in I_2} \langle a^{2j}, x \rangle - b_{2j}.$$

Here $I_k = \{1, \ldots, l_k\},\ l_k \geq 1,\ k = 1,2$. Functions f_1 and f_2 are convex PWL functions, which are also called convex polyhedral functions. The function f is a DP function. The class of DP functions is a subset of difference of convex (DC) functions. Results on DC functions and DC programming can be found, for example, in [29, 41, 44].

It is obvious that functions $f_k,\ k = 1,2$ are Lipschitz continuous on $\mathbb{R}^n$ with the Lipschitz constant

$$L_k = \max_{i \in I_k} \|a^{ki}\| < \infty$$

and the function f is Lipschitz continuous on $\mathbb{R}^n$ with the constant $L = L_1 + L_2$. The subdifferential of the function $f_k, k = 1,2$ at x is:

$$\partial f_k(x) = \text{conv}\left\{a^{ki} :\ i \in R_k(x)\right\}$$

where

$$R_k(x) = \left\{i \in I_k :\ \langle a^{ki}, x \rangle - b_{ki} = f_k(x)\right\},\ k = 1,2. \tag{3.2}$$

Denote $\varphi_{ki}(x) = \langle a^{ki}, x \rangle - b_{ki},\ i \in I_k,\ k = 1,2$. At a point $x \in \mathbb{R}^n$ for functions f_1 and f_2 define the following sets:

$$df_k(x) = \text{conv}\left\{(\xi_{ki}, a^{ki}),\ \xi_{ki} = \varphi_{ki}(x) - f_k(x),\ i \in I_k\right\},\ k = 1,2.$$

The sets $df_1(x), df_2(x)$ are defined in $\mathbb{R}^{n+1}$. These sets are components of a codifferential of the function f at x, introduced in [15]. It is obvious that $\xi_{ki} = 0, i \in R_k(x)$ and $\xi_{ki} < 0, i \in I_k \setminus R_k(x),\ k = 1,2$, that is

$$\max_{(\xi,a) \in df_k(x)} \xi = 0,\ k = 1,2. \tag{3.3}$$

Furthermore,

$$\partial f_1(x) = \{u \in \mathbb{R}^n :\ (0,u) \in df_1(x)\},\ \partial f_2(x) = \{v \in \mathbb{R}^n :\ (0,v) \in df_2(x)\}.$$

Proposition 3.1. *The mappings $x \mapsto df_k(x),\ k = 1,2$ are Lipschitz continuous on $\mathbb{R}^n$ in the Hausdorff metrics that is there exists $L_H > 0$ such that*

$$H(df_k(x), df_k(y)) \leq L_H \|x - y\|, \quad \forall\ x, y \in \mathbb{R}^n,\ k = 1,2.$$

Proof. Take any $x, y \in \mathbb{R}^n$. For any $(\xi, a) \in df_k(x)$, where $\xi = \varphi_{ki}(x) - f_k(x)$ and $a = a^{ki}$ for some $i \in I_k$, choose $(\eta, c) \in df_k(y)$ such that $\eta = \varphi_{ki}(y) - f_k(y),\ c = a^{ki}$. Then

$$\begin{aligned} \|(\xi, a) - (\eta, c)\| &= |\varphi_{ki}(x) - f_k(x) - [\varphi_{ki}(y) - f_k(y)]| \\ &\leq |\varphi_{ki}(x) - \varphi_{ki}(y)| + |f_k(x) - f_k(y)|. \end{aligned}$$

Then it follows form the definition of functions $\varphi_{ki}, i \in I_k, k = 1,2$ that $\|(\xi, a) - (\eta, c)\| \leq 2L_k\|x - y\|$. This means that for $L_H = 2L_k$ the following holds:

$$\max_{(\xi,a)\in df_k(x)} \min_{(\eta,c)\in df_k(y)} \|(\xi, a) - (\eta, c)\| \leq L_H\|x - y\|.$$

In the same way we can show that

$$\max_{(\eta,c)\in df_k(y)} \min_{(\xi,a)\in df_k(x)} \|(\xi, a) - (\eta, c)\| \leq L_H\|x - y\|.$$

The proof is completed. □

Proposition 3.2. *For functions $f_k,\ k = 1, 2$ the following holds at any $x \in \mathbb{R}^n$:*

$$f_k(x + h) = f_k(x) + \max_{(\xi,u)\in df_k(x)} [\xi + \langle u, h\rangle], \quad \forall h \in \mathbb{R}^n. \tag{3.4}$$

Proof. Consider the following set at the point $y = x + h$

$$R_k(y) = \left\{i \in I_k : \ \langle a^{ki}, y\rangle - b_{ki} = f_k(y)\right\}.$$

For any $j \in R_k(y)$

$$\begin{aligned} f_k(y) - f_k(x) &= \varphi_{kj}(y) - f_k(x) \\ &= \langle a^{kj}, x\rangle - b_{kj} - f_k(x) + \langle a^{kj}, y - x\rangle \\ &= \xi_{kj} + \langle a^{kj}, y - x\rangle \end{aligned}$$

and for any other $j \notin R_k(y)$

$$f_k(y) - f_k(x) > \varphi_{kj}(y) - f_k(x) = \xi_{kj} + \langle a^{kj}, y - x\rangle.$$

This completes the proof. □

Corollary 3.3. *Let $x \in \mathbb{R}^n$ be a given point. Then there exists $\varepsilon > 0$ such that*

$$f_k(x + h) = f_k(x) + \max_{(0,u)\in df_k(x)} \langle u, h\rangle, \ k = 1, 2$$

for all $h \in B_\varepsilon(0_n)$.

Proof. Consider the set $R_k(x)$ at the point x given by (3.2). Define the following function

$$\hat{f}_k(y) = \max_{j\in R_k(x)} \varphi_{kj}(y).$$

Continuity of functions $\varphi_{kj},\ j \in I_k$ implies that there exists $\varepsilon > 0$ such that $f_k(x + h) = \hat{f}_k(x + h)$ for all $h \in B_\varepsilon(0_n)$. It is clear that

$$d\hat{f}_k(x) = \Big\{(\xi, u) \in df_k(x) : \quad \xi = 0\Big\}.$$

Then the rest of the proof follows from Proposition 3.2. □

4 Necessary and Sufficient Conditions for Local Optimality

In this section we study local optimality conditions for the PWL optimization problem (3.1). The following result is true for general DC functions.

Theorem 4.1 ([42]). *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be convex functions. If $x^* \in \mathbb{R}^n$ is a local minimizer of $f = f_1 - f_2$, then*

$$\partial f_2(x^*) \subseteq \partial f_1(x^*). \tag{4.1}$$

Points satisfying (4.1) are called inf-stationary points. The condition (4.1) is sufficient for local optimality if f_2 is a polyhedral convex function. The following proposition follows from Theorem 4.1.

Proposition 4.2. *The condition* (4.1) *is necessary and sufficient for the point x^* to be local minimizer of the problem* (3.1).

One can define two other types of stationary points for unconstrained PWL optimization problems. A point $x^* \in \mathbb{R}^n$ is called a Clarke stationary point of the problem (3.1) if $0_n \in \partial f(x^*)$. A point $x^* \in \mathbb{R}^n$ is called a critical point of the problem (3.1) if $\partial f_1(x^*) \cap \partial f_2(x^*) \neq \emptyset$. Any inf-stationary point is also Clarke stationary and critical and any Clarke stationary point is also critical point. However the reverse is not always true. Examples confirming this are given in [23, 30].

Next we formulate necessary and sufficient local optimality conditions using sets $df_1(x)$ and $df_2(x)$. At a point x for a given $(0, w) \in df_2(x)$ consider the set

$$L_{0w}(x) = -(0, w) + df_1(x).$$

Proposition 4.3. *If a point $x^* \in \mathbb{R}^n$ is a local minimizer of the function f then*

$$0_{n+1} \in L_{0w}(x^*) \quad \forall (0, w) \in df_2(x). \tag{4.2}$$

Proof. Assume the contrary, that is x^* is a local minimizer, however there exists $(0, w) \in df_2(x)$ such that $0_{n+1} \notin L_w(x^*)$. Since the set $L_w(x^*)$ is compact and convex it follows that

$$\|z^0\| = \min\{\|z\| : \quad z \in L_{0w}(x^*)\} > 0 \text{ with } \ z^0 = (\xi_0, u^0).$$

The necessary condition for a minimum implies that

$$\langle z^0, z - z^0 \rangle \geq 0 \quad \forall z = (\xi, u) \in L_{0w}(x^*). \tag{4.3}$$

First we will show that $u^0 \neq 0_n$. Assume the contrary, that is $u^0 = 0_n$. Since $z^0 \neq 0_{n+1}$ we get that $\xi_0 < 0$. Then it follows from (4.3) that $\xi_0(\xi - \xi_0) \geq 0$ or $\xi \leq \xi_0 < 0$. In other words

$$\max_{(\xi,u) \in L_{0w}(x^*)} \xi < 0.$$

It follows from the definition of the set $L_{0w}(x^*)$ that

$$\max_{(\xi,u) \in df_1(x^*)} \xi = \max_{(\xi,u) \in L_{0w}(x^*)} \xi < 0$$

which contradicts (3.3).

Dividing both sides of (4.3) by $-\|z^0\|$ we get

$$-\frac{\xi_0 \xi}{\|z^0\|} + \langle u, d^0 \rangle \leq -\|z^0\|. \tag{4.4}$$

Here $d^0 = -\|z^0\|^{-1}u^0 \in \mathbb{R}^n$. It is clear that $\|z^0\|^{-1}\xi_0 \in (-1, 0)$. It is also obvious that for sufficiently small $\alpha > 0$

$$\mu = -\frac{\alpha\xi_0}{\|z^0\|} \in (0, 1).$$

Therefore taking into account that $\xi \leq 0$ and (4.4) we get

$$\xi + \alpha\langle u, d^0\rangle \leq \mu\xi + \alpha\langle u, d^0\rangle = -\frac{\alpha\xi_0}{\|z^0\|}\xi + \alpha\langle u, d^0\rangle \leq -\alpha\|z^0\|.$$

Thus $\xi + \alpha\langle u, d^0\rangle \leq -\alpha\|z^0\|$ for all $z = (\xi, u) \in L_{0w}(x^*)$. It follows from Proposition 3.2 that

$$\begin{aligned} f(x^* + \alpha d^0) &= f(x^*) + \max_{(\xi,u)\in df_1(x^*)} \left[\xi + \alpha\langle u, d^0\rangle\right] - \max_{(\eta,v)\in df_2(x^*)} \left[\eta + \alpha\langle v, d^0\rangle\right] \\ &\leq f(x^*) + \max_{(\xi,u)\in L_{0w}(x^*)} \left[\xi + \alpha\langle v, d^0\rangle\right] \\ &= f(x^*) - \alpha\|z^0\|. \end{aligned}$$

Then $f(x^* + \alpha d^0) < f(x^*)$ for all sufficiently small $\alpha > 0$, which contradicts to the fact that x^* is a local minimizer. □

Remark 4.4. The necessary condition (4.2) for general codifferentiable functions was also formulated in [15]. However our proof differs from that of [15].

Next we prove that the necessary local optimality condition (4.2) is also sufficient.

Proposition 4.5. *A point $x^* \in \mathbb{R}^n$ is a local minimizer of Problem (3.1) if and only if the condition (4.2) holds.*

Proof. The necessity is given by Proposition 4.3. Therefore we prove only sufficiency. Assume that the condition (4.2) holds. It follows from Proposition 3.2 and Corollary 3.3 that there exists $\varepsilon > 0$ such that

$$f(y) = f(x^*) + \max_{(\xi,u)\in df_1(x^*)} [\xi + \langle u, y - x^*\rangle] - \max_{(0,w)\in df_2(x^*)} \langle w, y - x^*\rangle$$

for all $y \in B_\varepsilon(x^*)$. For $y \in B_\varepsilon(x^*)$ consider

$$w(y) = \operatorname*{argmax}_{(0,w)\in df_2(x^*)} \langle w, y - x^*\rangle.$$

Then $f(y) - f(x^*) \geq \eta + \langle v - w(y), y - x^*\rangle$ for all $(\eta, v) \in df_1(x^*)$. This can be rewritten as $f(y) - f(x^*) \geq \eta + \langle u, y - x^*\rangle$, $(\eta, u) \in L_{0w(y)}(x^*)$. In particular, this is true for $(\eta, u) = (0, 0_n) \in L_{0w(y)}(x^*)$. Since $y \in B_\varepsilon(x^*)$ is arbitrary we get $f(x^*) \leq f(y)$ for all $y \in B_\varepsilon(x^*)$ that is x^* is a local minimizer of Problem (3.1). □

Remark 4.6. The optimality condition (4.1) is tighter than the condition (4.2). However, the latter can be extended to global optimality conditions for Problem (3.1). Such conditions are studied in the next section.

5 Necessary and Sufficient Conditions for Global Optimality

The following optimality condition is applicable also to DP functions, however it is rather difficult to utilize in solution algorithms as it requires availability of the whole ε-subdifferential which is not easy to calculate even for DP functions.

Theorem 5.1 ([28]). *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be convex functions. A DC function $f = f_1 - f_2$ attains its global minimum at a point $x^* \in \mathbb{R}^n$, if and only if $\partial_\varepsilon f_2(x^*) \subset \partial_\varepsilon f_1(x^*)$ for all $\varepsilon \geq 0$.*

We study global optimality conditions for Problem (3.1) using mappings $x \mapsto df_1(x)$ and $x \mapsto df_2(x)$. Consider the set

$$L_{\theta w}(x) = -(\theta, w) + df_1(x), \quad (\theta, w) \in df_2(x).$$

Proposition 5.2. *Suppose that at a point $x^* \in \mathbb{R}^n$*

$$0_{n+1} \in L_{\theta w}(x^*) \quad \forall (\theta, w) \in df_2(x^*). \tag{5.1}$$

Then x^* is a global minimizer of Problem (3.1).

Proof. Take any $x \in \mathbb{R}^n$ and define

$$(\theta_*, w^*) \in \underset{(\theta,w)\in df_2(x^*)}{\text{Argmax}} \left\{\theta + \langle w, x - x^*\rangle\right\}.$$

Then it follows from Proposition 3.2 that

$$\begin{aligned} f(x) - f(x^*) &= \max_{(\eta,v)\in df_1(x^*)} [\eta + \langle v, x - x^*\rangle] - \max_{(\theta,w)\in df_2(x^*)} [\theta + \langle w, x - x^*\rangle] \\ &= \max_{(\eta,v)\in df_1(x^*)} [\eta + \langle v, x - x^*\rangle] - \theta_* - \langle w^*, x - x^*\rangle \\ &= \max_{(\eta,v)\in -(\theta_*,w^*)+df_1(x^*)} [\eta + \langle v, x - x^*\rangle]. \end{aligned}$$

Since $0_{n+1} \in (\theta_*, w^*) + df_1(x^*)$ we get $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^n$, that is x^* is a global minimizer of Problem (3.1). □

The following example demonstrates that the condition (5.1) is not always necessary for global minimizers of Problem (3.1).

Example 5.3. Consider the function $f(x) = \max(-2x, 0) - \max(-x, -1)$, depicted in Figure 1. This function attains its global minimum at the point $x^* = 0$, and

$$df_1(x^*) = \text{conv}\{(0, -2), (0, 0)\}, \quad df_2(x^*) = \text{conv}\{(0, -1), (-1, 0)\}.$$

It is easy to see that $0_2 \notin L_{(-1,0)}(x^*) = \text{conv}\{(1, -2), (1, 0)\}$.

Next we formulate a necessary and sufficient condition for global minimizers of Problem (3.1). First, we will prove the following useful proposition.

Proposition 5.4. *Let us define a polytope $U = \text{conv}\{(\mu_i, u^i) \in \mathbb{R} \times \mathbb{R}^n,\ i = 1, \cdots, m\} \subset \mathbb{R}^{n+1}$. Then*

$$\max_{(\mu,u)\in U} \left[\mu + \langle u, d\rangle\right] \geq 0, \ \forall d \in \mathbb{R}^n$$

if and only if there exists $\mu \geq 0$ such that $(\mu, 0) \in U$.

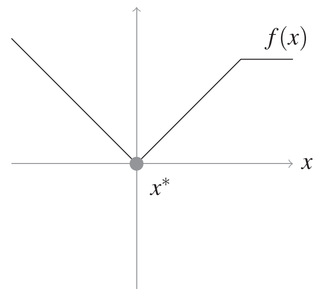

Figure 1: Example showing that the sufficient condition (5.1) is not always necessary.

Proof. Define the function

$$h(d) = \max_{(\mu,u)\in U} \Big[\mu + \langle u, d\rangle\Big],\ d \in \mathbb{R}^n.$$

The sufficiency is straightforward. Indeed, if $(\mu, 0) \in U$ with $\mu \geq 0$, then by the definition of h, we have $h(d) \geq \mu \geq 0$ for any $d \in \mathbb{R}^n$.

Now assume that $h(d) \geq 0$ for any $d \in \mathbb{R}^n$. Then, as a continuous piecewise linear function bounded from below, it attains a minimum h^* at a point $d^* \in \mathbb{R}^n$. Furthermore at any point d, we have:

$$h(d) - h(d^*) = \max_{(\mu,u)\in U} \Big[(\mu - h^*) + \langle u, d\rangle\Big]\ \forall d \in \mathbb{R}^n,$$

which means that

$$dh(d^*) = \Big\{(\mu - h^*, u) : (\mu, u) \in U\Big\}.$$

Applying the necessary condition for a minimum of the convex function h we get that $0_{n+1} \in dh(d^*)$. Therefore there exists $(\mu, 0) \in U$ such that $\mu = h^* \geq 0$. □

At a given point $x \in \mathbb{R}^n$ take any $(\theta, w) \in df_2(x)$ and define the following set:

$$d_\theta f(x) = \{(\eta, v) \in df_1(x) :\ \eta - \theta \geq 0\}.$$

It is clear that $d_\theta f(x) \neq \emptyset$ for any $(\theta, w) \in df_2(x)$. Define the set

$$L^+_{\theta w}(x) = -(\theta, w) + d_\theta f(x).$$

Proposition 5.5. *A point $x^* \in \mathbb{R}^n$ is a global minimizer of Problem (3.1) if and only if $0_n \in \{v : (\eta, v) \in L^+_{\theta w}(x^*)\}$ for any $(\theta, w) \in df_2(x^*)$.*

Proof. It follows from Proposition 3.2 that

$$f(x) - f(x^*) = \max_{(\eta,v)\in df_1(x^*)} \Big[\eta + \langle v, x - x^*\rangle\Big] - \max_{(\theta,w)\in df_2(x^*)} \Big[\theta + \langle w, x - x^*\rangle\Big].$$

Thus, $f(x) \geq f(x^*)$ for any $x \in \mathbb{R}^n$ if and only if

$$\max_{(\eta,v)\in df_1(x^*)} \Big[\eta + \langle v, x - x^*\rangle\Big] - \Big[\theta + \langle w, x - x^*\rangle\Big] \geq 0\ \ \forall(\theta, w) \in df_2(x^*),$$

that is:

$$\max_{(\mu,u)\in L^+_{\theta w}(x)} \Big[\mu + \langle u, x - x^*\rangle\Big] \geq 0\ \ \forall(\theta, w) \in df_2(x^*).$$

The rest of the proof is a direct application of Proposition 5.4. □

Remark 5.6. Although the condition (5.1) is not always necessary for global minimizers it is much easier to check than the condition in Proposition 5.5. Since the sets $df_1(x)$ and $df_2(x)$ are polytopes the condition (5.1) is checked only for extreme points of the set $df_2(x)$ which is equivalent to checking the solvability of a finite number (which is the number of extreme points of $df_2(x)$) of systems of linear equations.

We now propose the following:

Proposition 5.7. *Let* $A = \text{conv}\{(\alpha_1, a_1), \ldots, (\alpha_m, a_m)\} \subset \mathbb{R} \times \mathbb{R}^n$*. Define* $A^+ = \{(\alpha, a) \in A : \alpha \geq 0\}$ *and* $A_+ = \text{conv}(A \cup \{(-1, 0_n)\})$*. Then,* $0_n \in \{a : (\alpha, a) \in A^+\}$ *if and only if* $0_{n+1} \in A_+$*.*

Proof. We know that $0_n \in \{a : (\alpha, a) \in A^+\}$ if and only if there exist $\lambda_1, \ldots, \lambda_m$ such that $\lambda_i \in [0, 1], i = 1, \ldots, m$, $\sum_{i=1}^m \lambda_i = 1$ and

$$\sum_{i=1}^m \lambda_i a_i = 0, \quad \sum_{i=1}^m \lambda_i \alpha_i \quad \geq 0 \tag{5.2}$$

and similarly, $0_{n+1} \in A_+$ if and only if there exist $\lambda_1, \ldots, \lambda_{m+1}$ such that $\lambda_i \in [0, 1], i = 1, \ldots, m+1$, $\sum_{i=1}^{m+1} \lambda_i = 1$ and

$$\sum_{i=1}^m \lambda_i a_i = 0, \quad \sum_{i=1}^m \lambda_i \alpha_i - \lambda_{m+1} \quad = 0. \tag{5.3}$$

It suffices to show that the two sets of equations are equivalent. It is obvious that (5.3) implies (5.2). Assume that (5.2) is true and let $\beta = \sum_{i=1}^m \lambda_i \alpha_i$. Define

$$\gamma_i = \frac{\lambda_i}{(\beta + 1)}, i = 1, \ldots, m, \ \gamma_{m+1} = \frac{\beta}{\beta + 1}.$$

It is clear that $\sum_{i=1}^{m+1} \gamma_i = 1$ and that $\gamma_i \in [0, 1], i = 1 \ldots, m+1$. Furthermore,

$$\sum_{i=1}^m \gamma_i a_i = \frac{1}{\beta + 1} \sum_{i=1}^m \lambda_i a_i = 0$$

and

$$\sum_{i=1}^m \gamma_i \alpha_i - \gamma_{m+1} = \frac{1}{\beta + 1} \left(\sum_{i=1}^m \lambda_i \alpha_i - \beta \right) = 0.$$

This completes the proof. □

Corollary 5.8. *The point* x^* *is a global minimizer of Problem (3.1) if and only if*

$$0_{n+1} \in \text{conv}\left(L_{\theta w}(x^*) \cup \{(-1, 0)\}\right) \quad \forall (\theta, w) \in df_2(x^*). \tag{5.4}$$

Proof. The proof derives directly from Propositions 5.5 and 5.7. □

5.1 Constrained piecewise linear optimization problems

In this subsection we propose necessary and sufficient conditions for solutions of the following problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & \langle g^i, x\rangle \le p_i, i = 1, \dots, l, \end{cases} \tag{5.5}$$

where f is defined as in (3.1) and $g^i \in \mathbb{R}^n,\ p_i \in \mathbb{R},\ i = 1 \dots, l$.

For a given $x \in \mathbb{R}^n$, let $\alpha_i(x) = \langle g^i, x\rangle - p_i$.

Proposition 5.9. *A point $x^* \in \mathbb{R}^n$ is a global minimizer of Problem (5.5) if and only if*

$$0_n \in \text{conv}\left\{v : (\eta, v) \in L^+_{\theta w}(x^*)\right\} \cup \left\{(\alpha_i(x^*), g^i)\right\}$$

for any $(\theta, w) \in df_2(x^*)$.

Proof. The point x^* is a solution to the problem (5.5) if and only if for any $x \in \mathbb{R}^n$, either $f(x) - f(x^*) \ge 0$ or, $\langle g^i, x\rangle > p_i$ for some $i \in \{1, \dots, l\}$. Thus, x^* is a solution if and only if

$$\max_{(\mu,u)\in L_{\theta w}(x^*)} \left[\mu + \langle u, x - x^*\rangle\right] \ge 0 \quad \forall(\theta, w) \in df_2(x^*)$$

or

$$\max_{i=1,\dots,l} \left[\langle g^i, x - x^*\rangle + (\langle g^i, x^*\rangle - p_i)\right] \ge 0,$$

that is, if and only if

$$\max\Big(\max_{(\mu,u)\in L_{\theta w}(x^*)} \left[\mu + \langle u, x - x^*\rangle\right], \max_{1,\dots,l} \left[\alpha_i(x) + \langle g^i, x - x^*\rangle\right]\Big) \ge 0.$$

This completes the proof. □

Applying Proposition 5.4 we obtain the following result.

Corollary 5.10. *The point x^* is a global minimizer of Problem (3.1) if and only if*

$$0_{n+1} \in \text{conv}\Big(L_{\theta w}(x^*) \cup \{(\alpha_i(x^*), g^i)\} \cup \{(-1, 0)\}\Big) \quad \forall(\theta, w) \in df_2(x^*). \tag{5.6}$$

The proof derives directly from Propositions 5.7 and 5.9.

5.2 Examples

In this subsection we present some examples to demonstrate necessary and sufficient global optimality conditions.

Example 5.11. For the function f from Example 5.3 elements of the set $df_2(0)$ can be represented as $(1 - \alpha, \alpha)$ where $\alpha \in [0, 1]$. Then we get that:

$$L^+_{(1-\alpha,\alpha)}(0) = [-2 + \alpha, \alpha].$$

It is clear that $0 \in L^+_{(1-\alpha,\alpha)}(0)$ for any $\alpha \in [0, 1]$. Then it follows from Proposition 5.5 that the point $x^* = 0$ is a global minimizer of f.

Example 5.12. Consider the PWL function $f(x) = f_1(x) - f_2(x)$, where

$$f_1(x) = \max\Big\{3x_1 + 3x_2 - 6, -6x_1 + 3x_2 - 12, 3x_1 - 9x_2 - 6, -9x_1 + 3x_2 - 6\Big\},$$

$$f_2(x) = \max\Big\{-x_1 - x_2 - 3, -3x_1 - 3x_2 - 3, -4x_1 + 3x_2 - 5\Big\}.$$

The graph of this function is depicted in Figure 2. Take a point $x = (0, 0)$. Then

$$df_1(0,0) = \text{conv}\Big\{(0,3,3), (-6,-6,3), (0,3,-9), (0,-9,3)\Big\},$$

$$df_2(0,0) = \text{conv}\Big\{(0,-1,-1), (0,-3,-3), (-2,-4,3)\Big\}.$$

It is clear that $0_{n+1} \in -(\theta, w) + df_1(0,0)$ for all $(\theta, w) \in df_2(0,0)$. Then it follows from Proposition 5.2 that the point $x = (0,0)$ is a global minimizer of the function f and $f(0,0) = -3$.

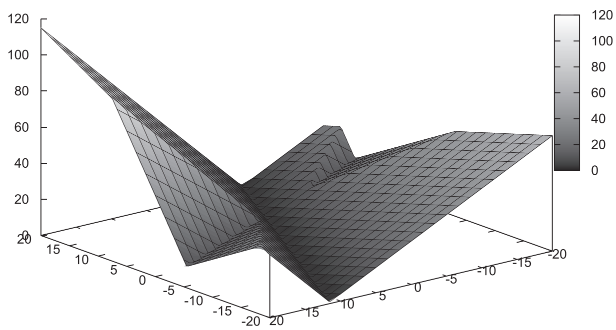

Figure 2: Plot of the function from Example 5.12.

Example 5.13. Consider the following function [16]:

$$f(x) = \min\Big\{\max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5\},\ \max\{-2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\}\Big\}.$$

Its graph is depicted in Figure 3.

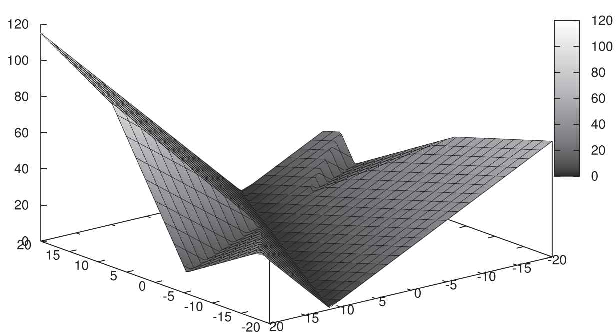

Figure 3: Plot of the function from Example 5.13.

This function is DP: $f(x) = f_1(x) - f_2(x)$ where

$$f_1(x) = \max\Big\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5\Big\} + \max\Big\{-2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\Big\},$$

$$f_2(x) = \max\Big\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5, -2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\Big\}.$$

The function f has two sets of local minimizers:

$$X_1^* = \Big\{x \in \mathbb{R}^2 : \ x_1 + 2x_2 = 3\Big\}, \quad X_2^* = \Big\{x \in \mathbb{R}^2 : \ 2x_1 + x_2 = 9\Big\}.$$

$f(x) = 1$ for all $x \in X_1^*$ and $f(x) = 12$ for all $x \in X_2^*$. Thus the set X_1^* is the set of global minimizers of the function f over $\mathbb{R}^2$. Take the point $x^0 = (1,1) \in X_1^*$. Then

$$df_1(x^0) = \text{conv}\Big\{(0,-3,-3), (0,0,3), (-24,5,1), (-24,8,7)\Big\},$$

$$df_2(x^0) = \text{conv}\Big\{(-17,-1,-2), (-17,2,4), (0,-2,-1), (-24,6,3)\Big\}.$$

It is easy to show that $(0,-2,-1),\ (-24,6,3) \in df_1(x^0)$, however $(-17,-2,-1),\ (-17,2,4) \notin df_1(x^0)$. Again this example demonstrates that the condition (5.1) is not always necessary for global minimizers.

The sets $L^+_{\theta w}(x^0)$ for extreme points $(0,-2,-1),\ (-24,6,3)$ are as follows:

$$L^+_{(0,-2,-1)}(x^0) = \text{conv}\Big\{(-1,-2), (2,4)\Big\},$$

$$L^+_{(-24,6,3)}(x^0) = \text{conv}\Big\{(-9,-6), (-6,0), (-1,-2), (2,4)\Big\}.$$

It is clear that $0_2 \in L^+_{(0,-2,-1)}(x^0)$ and $0_2 \in L^+_{(-24,6,3)}(x^0)$. For other two points sets are as follows:

$$L^+_{(-17,-1,-2)}(x^0) = (1,2) + \Big\{v \in \mathbb{R}^n : \ v = \alpha_1(-3,-3) + \alpha_2(0,3) + \alpha_3(5,1) + \alpha_4(8,7),$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1,\ \alpha_3 + \alpha_4 \le 17/24,\ \alpha_i \ge 0,\ i = 1,2,3,4\Big\},$$

$$L^+_{(-17,2,4)}(x^0) = (2,4) + \Big\{v \in \mathbb{R}^n : \ v = \alpha_1(-3,-3) + \alpha_2(0,3) + \alpha_3(5,1) + \alpha_4(8,7),$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1,\ \alpha_3 + \alpha_4 \le 17/24,\ \alpha_i \ge 0,\ i = 1,2,3,4\Big\}.$$

One can show that $0_2 \in L^+_{(-17,-1,-2)}(x^0)$ and $0_2 \in L^+_{(-17,2,4)}(x^0)$. Then it follows from Proposition 5.5 that the point $x^0 = (1,1)$ is a global minimizer of the function f.

6 Numerical Methods for Local Minimization of DP Functions

In this section we describe algorithms to locally solve Problem (3.1). We will design both exact and inexact algorithms for finding local minimizers of DP functions.

6.1 An exact algorithm for local minimization of DP functions

Before the description of the exact algorithm we present some results on the properties of steepest descent directions of DP functions. Assume that $x \in \mathbb{R}^n$ is not a stationary point which means that $\partial f_2(x) \not\subset \partial f_1(x)$. Let

$$R(x) = \Big\{(i,j) \in I \times J : \ i \in R_1(x), \ j \in R_2(x)\Big\},$$

where the sets $R_k(x), k = 1,2$ are defined by (3.2). Compute $\bar{v} \in \partial f_1(x)$ and $\bar{w} \in \partial f_2(x)$ such that

$$\|\bar{v} - \bar{w}\| = \max_{w \in \partial f_2(x)} \ \min_{v \in \partial f_1(x)} \|v - w\|. \tag{6.1}$$

It is clear that $\|\bar{v} - \bar{w}\| > 0$. Define the direction

$$\bar{d} = -(\bar{v} - \bar{w}). \tag{6.2}$$

A direction $d^0 = \|\bar{d}\|^{-1}\bar{d}$ is the steepest descent direction of f at x [15]. Compute $\bar{\lambda}$ by

$$\bar{\lambda} = \sup\{\lambda : \ R(x + \lambda\bar{d}) \subset R(x)\}.$$

If $\bar{\lambda} = \infty$, then the function f is unbounded along the ray $\{x + \lambda\bar{d} : \lambda \geq 0\}$. Main properties of the direction $\bar{d}$ are summarized in the following proposition.

Proposition 6.1. *Assume that $x \in \mathbb{R}^n$ is not a stationary point and the direction $\bar{d}$ is defined by (6.2). Then the following hold:*

1. $\bar{\lambda} > 0$;
2. $f(x + \lambda\bar{d}) \leq f(x) - \lambda\|\bar{d}\|^2$ *for* $\lambda \in [0, \bar{\lambda})$.

Proof. 1) Since both functions f_1 and f_2 are piecewise linear the upper semicontinuity of the subdifferential mappings $\partial f_1(x)$ and $\partial f_2(x)$ implies that there exists $\delta > 0$ such that $\partial f_i(y) \subseteq \partial f_i(x), \ i = 1,2$ for all $y \in B_\delta(x)$. This means that $\bar{\lambda} \geq \delta > 0$.

2) For any $\lambda \in [0, \bar{\lambda})$, we have:

$$\begin{aligned} f_1(x + \lambda\bar{d}) &= \max_{i \in R_1(x)} \ \big[\langle a^{1i}, x + \lambda\bar{d}\rangle - b_{1i}\big] \\ &\leq \max_{i \in R_1(x)} \ \big[\langle a^{1i}, x\rangle - b_{1i}\big] + \lambda \max_{i \in R_1(x)} \ \langle a^{1i}, \bar{d}\rangle \\ &= f_1(x) + \lambda \max_{i \in R_1(x)} \ \langle a^{1i}, \bar{d}\rangle. \end{aligned} \tag{6.3}$$

Necessary condition for a minimum implies that

$$\langle v - \bar{v}, \bar{v} - \bar{w}\rangle \geq 0 \quad \forall v \in \partial f_1(x). \tag{6.4}$$

Since $a^{1i} \in \partial f_1(x), \ i \in R_1(x)$ it follows from (6.4) that $\langle a^{1i}, \bar{d}\rangle \leq \langle \bar{v}, \bar{d}\rangle$ and therefore applying (6.3) we have $f_1(x + \lambda\bar{d}) \leq f_1(x) + \lambda\langle \bar{v}, \bar{d}\rangle$. On the other hand the subgradient inequality implies that $f_1(x + \lambda\bar{d}) \geq f_1(x) + \lambda\langle \bar{v}, \bar{d}\rangle$ and so combining the above two inequalities, we obtain

$$f_1(x + \lambda\bar{d}) = f_1(x) + \lambda\langle \bar{v}, \bar{d}\rangle. \tag{6.5}$$

Applying the subgradient inequality to the function f_2 we get $f_2(x + \lambda\bar{d}) \geq f_2(x) + \lambda\langle \bar{w}, \bar{d}\rangle$, and therefore

$$f(x + \lambda\bar{d}) \leq f(x) + \lambda\langle \bar{v} - \bar{w}, \bar{d}\rangle = f(x) - \lambda\langle \bar{d}, \bar{d}\rangle = f(x) - \lambda\|\bar{d}\|^2.$$

This completes the proof. □

Remark 6.2. In order to find $\bar{v}$ and $\bar{w}$ in (6.1) it is sufficient to find the distance between each extreme point of the set $\partial f_2(x)$ and the polytope $\partial f_1(x)$. Since the set $\partial f_2(x)$ is also a polytope finding $\bar{v}$ and $\bar{w}$ is equivalent to solving a finite number of quadratic programming problems. The number of these quadratic programming problems is the number of extreme points of the set $\partial f_2(x)$.

The exact algorithm uses subdifferentials $\partial f_1(x), \partial f_2(x)$ of component functions of the DP function f and proceeds as follows.

Algorithm 6.3. An exact algorithm for local minimization of DP functions.

Step 1. Select any starting point x^1. Set $k := 1$.

Step 2. If $\partial f_2(x^k) \subseteq \partial f_1(x^k)$, then stop. x^k is a local minimizer. Otherwise, select $j_k \in R_2(x^k)$ such that

$$0_n \neq a^{2j_k} \in \operatorname*{Argmax}_{w \in \partial f_2(x^k)} \min_{v \in \partial f_1(x^k)} \|v - w\| \tag{6.6}$$

and solve the following convex piecewise linear minimization problem:

$$\begin{cases} \text{minimize} & g_{j_k}(x) = f_1(x) - \langle a^{1j_k}, x\rangle + b_{2j_k} \\ \text{subject to} & x \in \mathbb{R}^n. \end{cases} \tag{6.7}$$

If this problem is unbounded, then stop. Otherwise, let x_*^k be a solution.

Step 3. Set $x^{k+1} := x_*^k,\ k := k + 1$ and go to Step 2.

Remark 6.4. The problem (6.7) is convex and there exist finite convergent algorithms for its solution. For example, it has been shown that an algorithm proposed in [7] converges to the solution of the problem (6.7) in finite iterations.

Remark 6.5. It should be noted that solving (6.1) and (6.6) are equivalent. In fact the optimal value of Problem (6.1) is the deviation of $\partial f_2(x)$ from the set $\partial f_1(x)$. Since both sets are polytopes it is sufficient to find the maximum distance between extreme points of $\partial f_2(x)$ and the set $\partial f_1(x)$. Thus, a^{2j_k} in (6.6) is found among extreme points of the set $\partial f_2(x)$.

Next we prove that Algorithm 6.3 is finite convergent.

Proposition 6.6. *At any iteration k we have $f(x^{k+1}) < f(x^k)$.*

Proof. The proof follows immediately from Item 2) of Proposition 6.1 and the fact that at the first step of the k-th iteration the descent direction coincides with the descent direction from (6.1). □

Proposition 6.7. *For any iteration $m > k$, $j_m \neq j_k$.*

Proof. The point x^{m+1} is a minimizer of the function g_{j_m}, while according to (6.6), x^m is not. Therefore,

$$g_{j_m}(x^{m+1}) < g_{j_m}(x^m) = f(x^m) \leq f(x^{k+1}) \leq g_{j_k}(x^{k+1}) \leq g_{j_k}(x^{m+1}).$$

We can conclude that $j_m \neq j_k$. □

Proposition 6.8. *Algorithm 6.3 converges in a finite number of iterations, and either concludes that the problem* (3.1) *is unbounded, or attains a local minimizer of the problem.*

Proof. By Proposition 6.7 and the fact that the set J is finite the algorithm terminates in at most $|J|$ iterations, where $|\cdot|$ stands for the cardinality of a set. It is clear from the algorithm that either it returns a point x^* such that $\partial f_2(x^*) \subseteq \partial f_1(x^*)$, or there exists $j \in J$ such that g_j is unbounded. In such a case, since $f(x) \leq g_j(x)$ for any $x \in \mathbb{R}^n$, the function f is also unbounded. Then we conclude that the maximum number of iterations of Algorithm 6.3 is finite. ☐

Remark 6.9. It is clear that the function f in (3.1) can be represented as follows:

$$f(x) = \min_{j \in J} \max_{i \in I} \left[\langle a^{1i} - a^{1j}, x \rangle - (b_{1i} - b_{2j}) \right]$$

which means that minimization of the function f is equivalent to the minimization of $|J|$ convex piecewise linear functions:

$$\varphi(x) = \max_{i \in I} \left[\langle a^{1i} - a^{2j}, x \rangle - (b_{1i} - b_{2j}) \right], \quad j \in J.$$

However, Algorithm 6.3 allows one to avoid solving of most of these problems and to consider only those which contribute to local minimizers.

6.2 Inexact method for local minimization of DP functions

In this subsection we will design an algorithm for finding the so-called ε-approximate solutions to the problem (3.1). This algorithm uses the sets $df_1(x)$ and $df_2(x)$. We start with the definition of the ε-approximate stationary point.

Definition 6.10. A point x^* is called an ε-approximate stationary point of Problem (3.1) iff:

$$0_{n+1} \in L_{0w}(x^*) + B_\varepsilon(0_{n+1}) \quad \forall (0, w) \in df_2(x^*).$$

Proposition 6.11. *Assume that the point x is not a stationary point of the problem (3.1) and for some $(0, w) \in df_2(x)$*

$$\|\bar{z}\|^2 \equiv \|(\bar{\eta}, \bar{v})\|^2 = \min \left\{ \|z\|^2 : \ z \in L_{0w}(x) \right\} > 0.$$

Then $\bar{v} \neq 0_n$. Moreover, if $\bar{\eta} = 0$ then the function f is unbounded from below and if $\bar{\eta} < 0$ then for $\bar{\alpha} = -1/\bar{\eta} > 0$

$$f(x - \bar{\alpha}\bar{v}) - f(x) \leq -\|\bar{z}\| \tag{6.8}$$

and

$$f(x - \bar{\alpha}\bar{v}) - f(x) \leq \bar{\eta}. \tag{6.9}$$

Proof. If x is not a stationary point then it follows from (4.2) that there exists $(0, w) \in df_2(x)$ such that $0_{n+1} \notin L_{0w}(x)$. This means that

$$\min_{z \in L_{0w}(x)} \|z\|^2 > 0.$$

Let $\bar{z} = (\bar{\eta}, \bar{v}) = \text{argmin}\{\|z\|^2 : \ z \in L_{0w}(x)\}$. Then it follows from the necessary condition for a minimum that

$$\bar{\eta}(\eta - \bar{\eta}) + \langle \bar{v}, v - \bar{v} \rangle \geq 0, \quad \forall (\eta, v) \in L_{0w}(x)$$

which means that

$$-\langle \bar{v}, v \rangle \leq -\|\bar{v}\|^2 + \bar{\eta}(\eta - \bar{\eta}). \tag{6.10}$$

It can be proved that if $\|\bar{z}\| > 0$ then $\bar{v} \neq 0_n$. Indeed if $\bar{v} = 0_n$ then it follows from (3.3) that $\bar{\eta} < 0$. In this case (6.10) implies that $\eta \leq \bar{\eta}$ for all $(\eta, v) \in df_1(x)$ which contradicts (3.3). Thus $\bar{v} \neq 0_n$. It follows from (3.4) and (6.10) that for any $\alpha > 0$

$$\begin{aligned}
f(x - \alpha\bar{v}) &= f(x) + \max_{(\eta,v)\in df_1(x)} [\eta - \alpha\langle\bar{v}, v\rangle] - \max_{(\mu,u)\in df_2(x)} [\mu - \alpha\langle\bar{v}, u\rangle] \\
&\leq f(x) + \max_{(\eta,v)\in df_1(x)} [\eta - \alpha\langle\bar{v}, v\rangle] - \alpha\langle\bar{v}, w\rangle \\
&= f(x) + \max_{(\eta,v)\in L_{0w}(x)} [\eta - \alpha\langle\bar{v}, v\rangle] \\
&\leq f(x) + \max_{(\eta,v)\in L_{0w}(x)} \left[\eta - \alpha\|\bar{v}\|^2 + \alpha\bar{\eta}(\eta - \bar{\eta})\right] \\
&= f(x) + \max_{(\eta,v)\in L_{0w}(x)} \left[\eta(1 + \alpha\bar{\eta}) - \alpha(\|\bar{v}\|^2 + \bar{\eta}^2)\right] \\
&= f(x) + \max_{(\eta,v)\in L_{0w}(x)} \eta(1 + \alpha\bar{\eta}) - \alpha\|\bar{z}\|^2.
\end{aligned}$$

If $\bar{\eta} = 0$ then taking into account (3.3) we have

$$\begin{aligned}
f(x - \alpha\bar{v}) &\leq f(x) + \max_{(\eta,v)\in df_1(x)} \eta - \alpha\|\bar{z}\|^2 \\
&= f(x) - \alpha\|\bar{z}\|^2.
\end{aligned}$$

In this case $f(x - \alpha\bar{v}) \to -\infty$ as $\alpha \to +\infty$ and therefore f is unbounded from below on $\mathbb{R}^n$. If $\bar{\eta} < 0$ then one can take $\bar{\alpha} = -1/\bar{\eta} > 0$. In this case $1 + \bar{\alpha}\bar{\eta} = 0$. Then

$$f(x - \bar{\alpha}\bar{v}) - f(x) \leq \frac{1}{\bar{\eta}}\|\bar{z}\|^2.$$

Since $-\bar{\eta} < \|\bar{z}\|$ we have

$$f(x - \bar{\alpha}\bar{v}) - f(x) \leq -\|\bar{z}\| \quad \text{and} \quad f(x - \bar{\alpha}\bar{v}) - f(x) \leq \bar{\eta}.$$

The proof is completed. □

Based on Proposition 6.11 we propose the following algorithm for solving Problem (3.1). Let $\varepsilon > 0$ be a given sufficiently small number.

Algorithm 6.12. An approximate algorithm for local minimization of DP functions.

Step 1. Select a starting point $x^1 \in \mathbb{R}^n$ and set $k := 1$.

Step 2. Compute $\bar{z}^k = (\bar{\eta}_k, \bar{v}^k) \in \mathbb{R}^{n+1}$ such that

$$\|\bar{z}^k\|^2 = \max_{(0,w)\in df_2(x^k)} \min_{z\in L_{0w}(x^k)} \|z\|^2.$$

Step 3. If $\bar{\eta}_k = 0$ and $\bar{v}_k = 0_n$ then the algorithm terminates. x^k is a local minimizer.

Step 4. If $\bar{\eta}_k = 0$ and $\bar{v}_k \neq 0_n$ then $f^* = -\infty$. The algorithm terminates. The objective function f is unbounded from below.

Step 5. If $\|\bar{z}^k\| < \varepsilon$, then the algorithm terminates. x^k is an ε-approximate solution.

Step 6. Compute $\bar{\alpha}_k = -1/\eta_k$ and $x^{k+1} = x^k - \bar{\alpha}_k\bar{v}^k$, set $k := k + 1$ and go to Step 2.

Finite convergence of Algorithm 6.12 is proved in the next proposition.

Theorem 6.13. *Suppose that the function f is bounded from below, that is $f_* > -\infty$. Then Algorithm 6.12 finds an ε-approximate solution of Problem (3.1) in finitely many iterations $m > 0$ where*

$$m \leq \left\lfloor \frac{f(x^1) - f_*}{\varepsilon} \right\rfloor + 1.$$

Proof. Since $f_* > -\infty$ the stopping criterion in Step 4 will never happen. It is sufficient to show that the stopping criterion in Step 5 will happen in finitely many iterations. Assume the contrary that is the algorithm is infinite convergent. Then $\|\bar{z}^k\| \geq \varepsilon$ for all $k \geq 1$ and therefore it follows from (6.8) that

$$f(x^k - \bar{\alpha_k}\bar{v}^k) - f(x^1) \leq -k\varepsilon, \quad k \geq 2.$$

This means that $f(x^k) \to -\infty$ as $k \to \infty$. We get the contradiction. It is clear that the maximum number m of iterations necessary for the stopping criterion in Step 5 to happen is:

$$m \leq \left\lfloor \frac{f(x^1) - f_*}{\varepsilon} \right\rfloor + 1.$$

The proof is complete. □

Remark 6.14. Even though Algorithm 6.12 reaches only an approximate solution, its convergence rate does not depend on the number of extreme points in $df_2(x)$, and so it may reach a solution faster than the exact Algorithm 6.3 in practice. It is possible to combine these two algorithms by using the approximate solution obtained by Algorithm 6.12 as a starting point for Algorithm 6.3.

7 Numerical Methods for Global Minimization of DP Functions

In this section we design an algorithm for finding global minimizers of Problem (3.1). This algorithm consists of two main steps. First we apply an algorithm for local minimization of DP functions. This algorithm either finds a local minimizer or determine that the objective function is unbounded from below, in which case the global minimization algorithm terminates. In the former case we apply an algorithm to escape from the local minimizer that finds better starting point for a local minimization algorithm and repeat the process as many times as necessary. Therefore we start by presenting results on computation of global search directions of DP functions.

In order to compute a descent direction at local minimizers (which are not global ones) of DP functions we will apply the sufficient condition given in Proposition 5.2 and the necessary and sufficient condition given in Proposition 5.5.

Proposition 7.1. *Suppose that a point $x \in \mathbb{R}^n$ is not a global minimizer of the function f and $0_{n+1} \notin L_{\theta w}(x)$ for some $(\theta, w) \in df_2(x)$. Let*

$$\|\bar{z}\|^2 \equiv \|(\bar{\mu}, \bar{u})\|^2 = \min_{z \in L_{\theta w}(x)} \|z\|^2 > 0.$$

We have:

1) *If $\bar{\mu} < 0$, then $\bar{u} \neq 0_n$ and $f(x - \alpha\bar{u}) < f(x) - \alpha\|\bar{z}\|^2$ for $\alpha = -1/\bar{\mu}$;*
2) *If $\bar{\mu} = 0$, then $\bar{u} \neq 0_n$ and $f(x - \alpha\bar{u}) < f(x)$ for all $\alpha > \theta/\|\bar{z}\|^2$. Moreover, in this case the function f is unbounded from below;*

3) *If* $\bar\mu > 0,\ 0 < \theta < \|\bar z\|$ *and* $\bar u \neq 0_n$*, then* $f(x - \alpha\bar u) < f(x)$ *for all* $\alpha > \theta/(\|\bar z\|^2 - \theta^2)$*. In this case the function* f *is unbounded from below.*

Proof. The necessary condition for a minimum implies that

$$\langle \bar z, z - \bar z\rangle \geq 0 \quad \forall z \in L_{\theta w}(x). \tag{7.1}$$

It is clear that $\bar z = (\bar\mu, \bar u) = (-\theta + \bar\eta, -w + \bar v)$ where $(\bar\eta, \bar v) \in df_1(x)$. First we will consider the case when $\bar\mu \leq 0$ and show that in this case $\bar u = -w + \bar v \neq 0_n$. Since $\bar z \neq 0_{n+1}$ if $\bar\mu = 0$ we have that $\bar u \neq 0_n$. Consider the case $\bar\mu < 0$ and assume the contrary, that is $\bar u = 0_n$. It follows from (7.1) that $\bar\mu(\mu - \bar\mu) \geq 0$ for all $(\mu, u) \in L_{\theta w}(x)$. Since for any $(\mu, u) \in L_{\theta w}(x)$ one can write $\mu = -\theta + \eta, (\eta, v) \in df_1(x)$ we get that

$$\bar\mu(\eta - \bar\eta) \geq 0 \ \text{ for all } \ (\eta, v) \in df_1(x). \tag{7.2}$$

On the other hand $\bar\mu = -\theta + \bar\eta < 0$ and therefore $\bar\eta < \theta$. It follows from (3.3) that $\theta \leq 0$ and consequently $\bar\eta < 0$. Since $\bar\mu < 0$, (7.2) implies that $\eta \leq \bar\eta < 0$ for all $(\eta, v) \in \underline{d}f(x)$. This contradicts (3.3). Therefore $\bar u \neq 0_n$.

It follows from (7.1) that

$$\langle u, -\bar u\rangle \leq \bar\mu\mu - \|\bar z\|^2 \quad \forall (\mu, u) \in L_{\theta w}(x).$$

The formula (3.3) implies that

$$\theta = \max_{(\mu,u)\in L_{\theta w}(x)} \mu. \tag{7.3}$$

Then for any $\alpha \geq 0$ we have

$$\begin{aligned}
f(x - \alpha\bar u) &= f(x) + \max_{(\eta,v)\in df_1(x)} [\eta - \alpha\langle v, \bar u\rangle] - \max_{(\omega,c)\in df_2(x)} [\omega - \alpha\langle c, \bar u\rangle] \\
&\leq f(x) + \max_{(\eta,v)\in df_1(x)} [\eta - \alpha\langle v, \bar u\rangle] + \theta - \alpha\langle w, \bar u\rangle \\
&= f(x) + \max_{(\mu,u)\in L_{\theta w}(x)} [\mu - \alpha\langle u, \bar u\rangle] \\
&\leq f(x) + \max_{(\mu,u)\in L_{\theta w}(x)} \left[\mu(1 + \alpha\bar\mu) - \alpha\|\bar z\|^2\right] \\
&= f(x) - \alpha\|\bar z\|^2 + \max_{(\mu,u)\in L_{\theta w}(x)} \mu(1 + \alpha\bar\mu).
\end{aligned}$$

If $\bar\mu < 0$ then for $\alpha = -1/\bar\mu$ we have

$$\max_{(\mu,u)\in L_{\theta w}(x)} \mu(1 + \alpha\bar\mu) = 0$$

and therefore $f(x - \alpha\bar u) \leq f(x) - \alpha\|\bar z\|^2$.

If $\bar\mu = 0$ then (7.3) implies that $f(x - \alpha\bar u) \leq f(x) - \alpha\|\bar z\|^2 + \theta$. Then $f(x - \alpha\bar u) < f(x)$ for any $\alpha > \theta/\|\bar z\|^2 > 0$. If we take $\alpha_m = m\theta/\|\bar z\|^2 > \theta/\|\bar z\|^2,\ m = 2, 3, \ldots$ then

$$f(x - \alpha_m\bar u) < f(x) - (m-1)\theta,\ m = 2, 3, \ldots.$$

It is obvious that $f(x - \alpha_m\bar u) \to -\infty$ as $m \to +\infty$, that is in this case the function f is unbounded from below.

Finally, assume that $\bar\mu > 0,\ \theta < \|\bar z\|$ and $\bar u \neq 0_n$. It follows from (7.3) that $\bar\mu \leq \theta$. Then for any $\alpha \geq 0$

$$f(x - \alpha\bar u) \leq f(x) - \alpha\|\bar z\|^2 + \theta(1 + \alpha\theta)$$

and $f(x-\alpha\bar{u}) < f(x)$ for all $\alpha > \theta/(\|\bar{z}\|^2-\theta^2)$. If we take $\alpha_m = m\theta/(\|\bar{z}\|^2-\theta^2),\ m=2,3,\dots$ then

$$f(x-\alpha_m\bar{u}) < f(x) - (m-1)\theta,\ m=2,3,\dots.$$

Therefore $f(x-\alpha_m\bar{u}) \to -\infty$ as $m \to +\infty$, that is in this case the function f is unbounded from below. □

Remark 7.2. Proposition 7.1 shows how one can compute global descent direction at a point x which is not a global minimizer. It is sufficient to compute an element $\bar{z}$ only for extreme points of the set $df_2(x)$. The element $\bar{z}$ is found by solving a quadratic programming problem. The number of such problems solved at the point x is no greater than the number of extreme points of $df_2(x)$. In many cases this number is small because one stops as soon as a global descent direction is found and therefore not considering all extreme points of the set $df_2(x)$.

The number of quadratic programming problems to be solved is especially small when the point x is a local minimizer which is not a global one. In this case all extreme points of the set $df_2(x)$ of the form $(0,w)$ are excluded from consideration because $0_{n+1} \in L_{0w}(x)$ for all $(0,w) \in df_2(x)$. This means that finding the global descent directions at local minimizers is easier than at any other points.

Proposition 7.3. *Suppose that a point $x \in \mathbb{R}^n$ is not a global minimizer of the function f. Then there exist $(\bar{\theta},\bar{w}) \in df_2(x)$ and $y \in \mathbb{R}^n$ such that*

$$\max_{(\mu,u)\in L_{\bar{\theta}\bar{w}}(x)} \{\mu + \langle u, y-x\rangle\} < 0 \tag{7.4}$$

and $f(y) < f(x)$.

Proof. It follows from (3.4) that for any $y \in \mathbb{R}^n$

$$f(y) - f(x) = \max_{(\eta,v)\in df_1(x)} \{\eta + \langle v, y-x\rangle\} - \max_{(\theta,w)\in df_2(x)} \{\theta + \langle w, y-x\rangle\}.$$

Since x is not a global minimizer there exists $y \in \mathbb{R}^n$ such that $f(y) < f(x)$ and therefore

$$\max_{(\eta,v)\in df_1(x)} \{\eta + \langle v, y-x\rangle\} - \max_{(\theta,w)\in df_2(x)} \{\theta + \langle w, y-x\rangle\} < 0.$$

Let

$$(\bar{\theta},\bar{w}) = \operatorname*{argmax}_{(\theta,w)\in df_2(x)} \{\theta + \langle w, y-x\rangle\}.$$

Then (7.4) is satisfied for $(\bar{\theta},\bar{w}) \in df_2(x)$ and in this case $f(y) < f(x)$. □

Remark 7.4. Proposition 7.3 implies that if a point $x \in \mathbb{R}^n$ is not a global minimizer then

$$\min_{y\in\mathbb{R}^n} \min_{(\theta,w)\in df_2(x)} \max_{(\mu,u)\in L_{\theta w}(x)} \{\mu + \langle u, y-x\rangle\} < 0$$

or

$$\min_{(\theta,w)\in df_2(x)} \min_{y\in\mathbb{R}^n} \max_{(\mu,u)\in L_{\theta w}(x)} \{\mu + \langle u, y-x\rangle\} < 0.$$

This means that in order to find a global descent direction at a point x one can minimize a convex piecewise linear function

$$\psi(y) = \max_{(\mu,u)\in L_{\theta w}(x)} \{\mu + \langle u, y-x\rangle\}$$

for each $(\theta,w) \in df_2(x)$.

Corollary 7.5. *Let x be a local minimizer of the function f which is not a global one. Then for any $(0, w) \in df_2(x)$ and $y \in \mathbb{R}^n$*

$$\max_{(\mu,u)\in L_{0w}(x)} \{\mu + \langle u, y - x\rangle\} \geq 0.$$

Proof. Since x is a local minimizer then the necessary condition for a minimum implies that $0_{n+1} \in L_{0w}(x)$ for all $(0, w) \in df_2(x)$ which completes the proof. □

Remark 7.6. Corollary 7.5 implies that if the point x is a local minimizer but not a global one all elements of the set $df_2(x)$ of the form $(0, w)$ should be excluded when the global descent direction is computed at this point. Therefore the complexity of computation of descent directions at local minimizers can be reduced significantly.

Based on results from Propositions 7.1 and 7.3 we propose the following algorithm for solving Problem (3.1). Let $\varepsilon > 0$ be a sufficiently small number.

Algorithm 7.7. Global minimization of DP functions.

Step 1. Select a starting point $x^1 \in \mathbb{R}^n$, set $\bar{x} = x^1$ and $k := 1$.

Step 2. Starting from the point $\bar{x}$ apply either Algorithm 6.3 or Algorithm 6.12 to find the local minimizer of Problem (3.1). As a result these algorithms either compute a local minimizer $\bar{y}$ or find that Problem (3.1) is unbounded from below.

Step 3. (The first stopping criterion). If the function f is unbounded from below, then stop. Problem (3.1) has no solution. Otherwise set $x^{k+1} = \bar{y}$ and $k := k + 1$.

Step 4. (The second stopping criterion). If $0_{n+1} \in L_{\theta w}(x^k)$ for all $(\theta, w) \in df_2(x^k)$, then stop. x^k is a global minimizer.

Step 5. (The third stopping criterion). If $0_n \in \{v : (\eta, v) \in L^+_{\theta w}(x^k)\}$ for any $(\theta, w) \in df_2(x^k)$, then stop. x^k is a global minimizer.

Step 6. Compute $(\theta, w) \in df_2(x^k)$ and $\bar{z}^k = (\bar{\mu}_k, \bar{u}^k) \in L_{\theta w}(x^k)$ such that

$$\|\bar{z}^k\| = \min_{z\in L_{\theta w}(x^k)} \|z\| > 0.$$

Step 7. If $\bar{\mu}_k < 0$, then set $\alpha_k = -1/\bar{\mu}_k$ and go to Step 9. If $\bar{\mu}_k = 0$ or $\bar{\mu}_k > 0, 0 < \theta < \|\bar{z}^k\|$ and $\bar{u}^k \neq 0_n$, then stop. The objective function f is unbounded from below and Problem (3.1) has no solution. Otherwise go to Step 8.

Step 8. Compute $(\theta, w) \in \overline{d}f(x^k)$ and $y \in \mathbb{R}^n$ such that

$$\max_{(\mu,u)\in L_{\theta w}(x)} \left\{\mu + \langle u, y - x^k\rangle\right\} < 0.$$

Set $\alpha_k := 1,\ \bar{u}^k := x^k - y$.

Step 9. Compute $\bar{x} = x^k - \alpha_k \bar{u}^k$ and go to Step 2.

Remark 7.8. Some explanation on Algorithm 7.7 follows. The algorithm starts with the choice of the starting point which can be any point from $\mathbb{R}^n$ (Step 1). Then starting from this point we apply one of the local search algorithms (either Algorithm 6.3 or Algorithm 6.12) and find a local minimizer (Step 2). Steps 3-5 contain three stopping criteria. Both Algorithm 6.3 or Algorithm 6.12 may find that the objective function is not bounded from below. In this case the problem has no solution and the algorithm terminates. This is done in Step 3. In Step 4 we check the sufficient condition (5.1) for global optimality. If it is satisfied then the global minimizer has been found and the algorithm terminates. Finally, in Step 5 we check the necessary and sufficient condition from Proposition 5.5. If it is satisfied then the global minimizer has been found and the algorithm terminates.

Note that the sufficient condition (5.1) is easier to check than the necessary and sufficient condition from Proposition 5.5. If none of these stopping criteria are satisfied then in Steps 6-8 we compute the global search direction $-\bar{u}^k$ and step length $\alpha_k > 0$. In Steps 6 and 7 we apply results from Proposition 7.1 to do so. In these two steps we either compute the global search direction and the step length or find that Problem (3.1) has no solution. If the conditions in Step 7 are not satisfied then in Step 8 we apply results from Proposition 7.3 to find the global search direction and the step length. Again we have to point out that Steps 6 and 7 are easier to implement than Step 8. We follow the global descent in Step 9 and find a better starting point $\bar{x}$ for the local search.

The proof of the following theorem about the complexity of Algorithm 7.7 is straightforward.

Theorem 7.9. *Algorithm 7.7 terminates after finite number $K > 0$ of iterations and either determines that the objective function is unbounded from below or finds the global solution to Problem (3.1). Here*

$$K \le 2^l Np \quad \textit{if one applies Algorithm 6.3 in Step 2}$$

and

$$K \le N \left[\frac{f(x) - f_*}{\varepsilon} + 1 \right] \quad \textit{if one applies Algorithm 6.12 in Step 2.}$$

Here f_* is the value of the global minimum and N is the number of local minimizers of Problem (3.1).

8 Illustrative Examples

In this section we present three examples to demonstrate how Algorithm 7.7 works.

Example 8.1. Consider the following DP function

$$f(x) = \min\{1, |x|\} = \max\{1 + x, 1 - x\} - \max\{1, x, -x\}, \quad x \in \mathbb{R}.$$

Here $f_1(x) = \max\{1+x, 1-x\}$, $f_2(x) = \max\{1, x, -x\}$. The global minimizer of this function is $x = 0$ with $f(0) = 0$. Any point $y \in \mathbb{R}$ belonging to the set $U = \{x \in \mathbb{R} : \ x \le -1 \text{ or } x \ge 1\}$ is a local minimizer with $f(y) = 1$. Take the point $x = 2$, which is a local minimizer. We have

$$df_1(2) = \text{conv}\{(-4, -1), (0, 1)\} \ \text{ and } \ df_2(2) = -\,\text{conv}\{(-4, -1), (0, 1), (-1, 0)\}.$$

Since $0_2 \in -(0, 1) + df_1(2)$ the necessary and sufficient condition for local optimality is satisfied. It is obvious that $0_2 \in -(-4, -1) + df_2(2)$, however $0_2 \notin -(-1, 0) + df_1(2)$. We define $L_{(-1,0)}(2) = \text{conv}\{(-3, -1), (1, 1)\}$ and then compute

$$\|(-0.2, 0.4)\| = \min \left\{ \|(\mu, u)\| : \ (\mu, u) \in L_{(-1,0)}(2) \right\}.$$

In order to find the global descent direction we apply Proposition 7.1. Since $\bar{\mu} = -0.2$ it follows from Proposition 7.1, 1) that the descent direction is $\bar{u} = -0.4$ and $\alpha = -1/\bar{\mu} = 5$. Then we find $\bar{x} = x + \alpha \bar{u} = 0$ which is the global minimizer of the function. This example demonstrates that Proposition 7.1 can be applied to find global descent directions from local minimizers.

Example 8.2. Consider the following DP function

$$f(x) = \max\{0, 2(x-1)\} - \max\{-x, x\}, \ x \in \mathbb{R}.$$

Here $f_1(x) = \max\{0, 2(x-1)\}$, $f_2(x) = -\max\{-x, x\}$. This function is unbounded from below that is $f_* = -\infty$. The point $x = 1$ is a local minimizer of f and $f(1) = -1$. Take the point $x = 2$. We have $f(2) = 0$ and

$$df_1(2) = \text{conv}\{(-2, 0), (0, 2)\}, \quad df_2(2) = \text{conv}\{(0, -1), (4, 1)\}.$$

Compute

$$L_{(0,-1)}(2) = \text{conv}\left\{(-2, -1), (0, 1)\right\} \quad \text{and} \quad L_{(4,1)}(2) = \text{conv}\left\{(2, 1), (4, 3)\right\}.$$

Since $0_2 \notin L_{(0,-1)}(2)$ it follows that $x = 2$ is not a local minimizer. It is easy to see that

$$\|(-1/2, 1/2)\| = \min\{\|(\mu, u)\| : (\mu, u) \in L_{(0,-1)}(2)\}.$$

Then $\bar{\mu} = -1/2$ and applying Algorithm 6.12 we find the step length $\alpha = -1/\bar{\mu} = 2$ and descent direction $\bar{u} = -1/2$ in Step 2 of Algorithm 7.7. We have $\bar{x} = x + \alpha\bar{u} = 1$ which is the local minimizer. For $\bar{x} = 1$ we have

$$df_1(1) = \text{conv}\{(0, 0), (0, 2)\} \ \text{ and } \ df_2(1) = \text{conv}\{(0, -1), (2, 1)\}.$$

Then we compute

$$L_{(-1,0)}(1) = \text{conv}\{(0, -1), (0, 1)\} \quad \text{and} \quad L_{(2,1)}(1) = \text{conv}\{(2, 1), (2, 3)\}.$$

Since $x = 1$ is a local minimizer the necessary condition for a minimum implies that $0_2 \in L_{(-1,0)}(1)$. It is easy to see that $0_2 \notin L_{(2,1)}(1)$ and

$$\|\bar{z}\| \equiv \|(2, 1)\| = \min_{(\mu,u) \in L_{(2,1)}(1)} \|(\mu, u)\|.$$

In this case $\bar{\mu} > 0$, $\bar{u} = -1 \neq 0$ and $\theta = 2 < \|\bar{z}\|$. Therefore we can apply Proposition 7.1, 3) which implies that the function f is unbounded from below.

Example 8.3. Consider the following function [16]:

$$f(x) = f_1(x) - f_2(x)$$

where

$$f_1(x) = \max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5\} + \max\{-2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\},$$

$$f_2(x) = \max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5, -2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\}.$$

To minimize it we take $x^0 = (4, 4)$ as a starting point. We have $f(x^0) = 19$ and

$$df_1(x^0) = \text{conv}\{(-39, -3, -3), \ (-27, 5, 1), (-12, 0, 3), (0, 8, 7)\},$$

$$df_2(x^0) = \text{conv}\{(29, 1, 2), \ (2, -2, -4), (12, 2, 1), (0, -6, -3)\}.$$

In order to find the local descent direction we consider the set:

$$L_{(0,-6,-3)}(x^0) = \text{conv}\{(-39, -9, -6), \ (-27, -1, -2), (-12, -6, 0), (0, 2, 4)\}.$$

Then

$$\|(-1.71429, 0.85714, 3.42857)\| = \min\{\|(\mu, u)\| : (\mu, u) \in L_{(0,-6,-3)}(x^0)\}.$$

Since $\bar{\mu} < 0$ the step length $\alpha > 0$ is defined as $\alpha = -1/\bar{\mu}$. The descent direction is $\bar{u} = (-0.85714, -3.42857)$. Applying Algorithm 6.12 we have $x^1 = x^0 + \alpha\bar{u} = (3.5, 2)$ and $f(x^1) = 12$. Next we compute:

$$df_1(x^1) = \text{conv}\{(-13.5, -3, -3),\ (-13.5, 5, 1), (0, 0, 3), (0, 8, 7)\},$$

$$df_2(x^1) = \text{conv}\{(15.5, 1, 2),\ (2, -2, -4),\ (0, 2, 1),\ (0, -6, -3)\}.$$

We have

$$L_{(0,2,1)}(x^1) = \text{conv}\{(-13.5, -1, -2),\ (-13.5, 7, 2), (0, 2, 4), (0, 10, 8)\},$$

$$L_{(0,-6,-3)}(x^1) = \text{conv}\{(-13.5, -9, -6),\ (-13.5, -1, -2), (0, -6, 0), (0, 2, 4)\}.$$

For the set $L_{(0,-6,-3)}(x^1)$ we get

$$\|(-0.73469, -1.10204, 2.20408)\| = \min\{\|(\mu, u)\| : (\mu, u) \in L_{(0,-6,-3)}(x^1)\}.$$

Therefore $\bar{\mu} = -0.73469 < 0$ and the descent direction $\bar{u} = (1.10204, -2.20408)$. Applying Algorithm 6.12 we have $\bar{x}^2 = (5, -1)$ and $f(\bar{x}^2) = 1$. For the set $L_{(0,2,1)}(x^1)$ we get

$$\|(-1.78218, 1.60396, 3.20792)\| = \min\{\|(\mu, u)\| : (\mu, u) \in L_{(0,2,1)}(x^1)\}.$$

Then $\bar{\mu} = -1.78218 < 0$ and the descent direction $\bar{u} = (-1.60396, -3.20792)$. Applying Algorithm 6.12 we have $\tilde{x}^2 = (2.6, 0.2)$ and $f(\tilde{x}^2) = 1$. We take $x^2 = \bar{x}^2$. Then

$$df_1(x^2) = \text{conv}\{(0, -3, -3),\ (0, 5, 1), (0, 0, 3), (0, 8, 7)\},$$

$$df_2(x^2) = \text{conv}\{(11, 1, 2),\ (11, -2, -4),\ (0, 2, 1),\ (0, -6, -3)\}$$

and

$$L_{(0,2,1)}(x^2) = \text{conv}\{(0, -1, -2),\ (0, 7, 2), (0, 2, 4), (0, 10, 8)\},$$

$$L_{(0,-6,-3)}(x^2) = \text{conv}\{(0, -9, -6),\ (0, -1, -2), (0, -6, 0), (0, 2, 4)\}.$$

$0_2 \in L_{(0,2,1)}(x^2)$ and $0_2 \in L_{(0,-6,-3)}(x^2)$. This means that x^2 is a local minimizer of the function f. In order to check global optimality conditions at the point x^2 we consider the following two sets:

$$L_{(11,1,2)}(x^2) = \text{conv}\{(11, -2, -1),\ (11, 6, 3), (11, 1, 5), (11, 9, 9)\},$$

$$L_{(11,-2,-4)}(x^2) = \text{conv}\{(11, -5, -7),\ (11, 3, -3), (11, -2, -1), (11, 6, 3)\}.$$

Here $(11, 0, 0) \in L_{(11,1,2)}(x^2)$ and $(11, 0, 0) \in L_{(11,-2,-4)}(x^2)$. Since $\eta = 0$ for any $(\eta, v) \in df_1(x^2)$ it follows from the necessary and sufficient condition for global optimality of Proposition 5.5 that the point x^2 is a global minimizer of f.

9 Conclusions

In this paper, we develop new algorithms to both locally and globally minimize continuous piecewise linear functions represented as a difference of polyhedral functions. We present two algorithms to locally minimize DP functions. One of these algorithms is based on the subdifferential of the component functions, it is exact and finite convergent. The second algorithm is based on the concept of codifferential, it converges to approximate local minimizers in a finite number of iterations. We then develop an algorithm to globally minimize DP functions. This algorithm consists of two main steps. In the first, we apply one of the local search algorithms to find local minimizers of DP functions and then apply an algorithm based on the codifferential to escape from those local minimizers and find better starting points for local search algorithms. We prove that this algorithm is finite convergent.

Acknowledgment

The authors would like to thank two anonymous referees for their comments that helped to improve the quality of the paper.

References

[1] M. Abbasov, Finding the set of global minimizers of a piecewise affine function, *J. Glob. Optim.* 85 (2023) 1–13.

[2] M. Abbasov and V. Demyanov, Proper and adjoint exhausters in nonsmooth analysis: optimality conditions, *J. Glob. Optim.* 56 (2013) 569–585.

[3] E. Aghezzaf and L. Wolsey, Modeling piecewise linear concave costs in a tree partitioning problem, *Discrete Appl. Math.* 50 (1994) 101–109.

[4] A. Bagirov, Max-min separability, *Optimization Methods and Software* 20 (2005) 271–290.

[5] A. Bagirov, C. Clausen and M. Kohler, Estimation of a regression function by maxima of minima of linear functions, *IEEE Transactions on Information Theory* 55 (2009) 833–845.

[6] Z. Bayindir, S. Birbil and J. Frenk, A deterministic inventory/production model with general inventory cost rate function and piecewise linear concave production costs,*European Journal of Operational Research* 179 (2007) 114–123.

[7] M. Bazaraa, J. Goode and R. Rardin, A finite steepest ascent algorithm for maximizing piecewise linear concave functions, *Journal of Optimization Theory and Applications* 26 (1978) 437–442.

[8] M. Cavichia and M. Arenales, Piecewise linear programming via interior points, *Computers & Operations Research* 27 (2000) 1303–1324.

[9] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[10] D. Clark and M. Osborne, A descent algorithm for minimizing polyhedral convex functions, *SIAM J. Scientific and Statistical Computing* 4 (1983) 757–786.

[11] K. Croxton, B. Gendron and T. Magnanti, A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems, *Management Sci.* 49 (2003) 1268–1273.

[12] K. Croxton, B. Gendron and T. Magnanti, Variable disaggregation in network flow problems with piecewise linear costs, *Oper. Res.* 55 (2007) 146–157.

[13] G. Dantzig, On the significance of solving linear-programming problems with some integer variables, *Econometrica* 28 (1960) 30–44.

[14] I.J. de Farias, M. Zhao and H. Zhao, A special ordered set approach for optimizing a discontinuous separable piecewise linear function, *Operations Research Letters* 36 (2008) 234–238.

[15] V. Demyanov, *A. Rubinov, Constructive Nonsmooth Analysis*, Peter Lang, Frankfurt am Main, 1995.

[16] V. Demyanov, A. Bagirov and A. Rubinov, A method of truncated codifferential with application to some problems of cluster analysis, *J.Glob. Optim.* 23 (2002) 63–80.

[17] M. Dolgopolik, The method of codifferential descent for convex and global piecewise affine optimization, *Optim. Methods and Software* 35 (2020) 1191–1222.

[18] M. Fisher, W. Northup and J. Shapiro, Using duality to solve discrete optimization problems: theory and computational experiences, *Mathematical Programming Study* 3 (1975) 56–94.

[19] R. Fourer, D.M. Gay and B.W. Kernighan, *AMPL-A Modeling Language for Mathematical Programming*, The Scientific Press, 1993.

[20] A. Fuduli, M. Gaudioso and G. Giallombardo, Minimizing nonconvex nonsmooth functions via cutting planes and proximity control, *SIAM Journal on Optimization* 14 (2004) 743–756.

[21] D. Fullerton and L. Gan, A simulation-based welfare loss calculation for labor taxes with piecewise linear budgets, *J. Public Economics* 88 (2004) 2339–2359.

[22] E. Gada, A. Atiya, S. Shaheen and A. El-Dessouki, A new algorithm for learning in piecewise-linear neural networks, *Neural Networks* 13 (2000) 485–505.

[23] M. Gaudioso, G. Giallombardo, G. Miglionico and A. Bagirov, Minimizing nonsmooth DC functions via successive DC piecewise-affine approximations, *J. Glob. Optim.* 71 (2018) 37–55.

[24] M. Gaudioso, E. Gorgone and M. Monaco, Piecewise linear approximations in nonconvex nonsmooth optimization, Numeriche Mathematik 113 (2009) 73–88.

[25] J. Glen, Mathematical programming models for piecewise-linear discriminant analysis, *Journal of the Opeartional Research Society* 56 (2005) 331–341.

[26] V. Gorokhovik, O. Zorko and G. Birkhoff, Piecewise affine functions and polyhedral sets, *Optimization* 31 (1994) 209–221.

[27] V. Gunnerud, B. Foss, Oil production optimization - a piecewise linear model, solved with two decomposition strategies, *Comp. Chem. Engr.* 34 (2010) 1803–1812.

[28] J.-B. Hiriart-Urruty, From convex minimization to nonconvex minimization: necessary and sufficient conditions for global optimality, in: *Nonsmooth Optimization and Related Topics*, F. Clarke, V. Demyanov, F. Giannessi (Eds.), Plenum, New York, 1989, pp. 147–175.

[29] R. Horst and N. Thoai, DC programming: Overview, *J. Optim. Theory Appl.* 103 (1999) 1–43.

[30] K. Joki, A. Bagirov, N. Karmitsa and M. Mäkelä, A proximal bundle method for nonsmooth DC optimization utilizing nonconvex cutting planes, *J. Glob. Optim.* 68 (2017) 501–535.

[31] S. Kameshwaran and Y. Narahari, Nonconvex piecewise linear knapsack problems, *European Journal of Operational Research* 192 (2009) 56–68.

[32] A. Keha, I. de Farias and G. Nemhauser, Models for representing piecewise linear cost functions, *Operations Research Letters* 32 (2004) 44–48.

[33] A. Keha, I. de Farias and G. Nemhauser, A branch-and-cut algorithm without binary variables for nonconvex piecewise linear optimization, *Operations Research* 54 (2006) 847–858.

[34] K. Kiwiel, A Cholesky dual method for proximal piecewise-linear programming, *Numeriche Mathematik* 68 (1994) 325–340.

[35] K. Lam and J. Moy, A piecewise linear programming approach to the two-group discriminant problem - an adaptation to Fisher's linear discriminant function model, *European Journal of Operational Research* 145 (2003) 471–481.

[36] J. Lee and D. Wilson, Polyhedral methods for piecewise-linear functions I: the lambda method, *Discrete Applied Mathematics* 108 (2001) 269–285.

[37] H. Markowitz and A. Manne, On the solution of discrete programming-problems, *Econometrica* 25 (1957) 84–110.

[38] A. Martin, M. Moller and S. Moritz, Mixed integer models for the stationary case of gas network optimization, *Mathematical Programming* 105 (2006) 563–582.

[39] L. Polyakova, On global unconstrained minimization of the difference of polyhedral functions, *J. Glob. Optim.* 50 (2011) 179–195.

[40] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[41] P. Tao and L. An, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, *J. Global Optim.* 133 (2005) 23–46.

[42] J. Toland, On subdifferential calculus and duality in nonconvex optimization, *Bull. Soc. Math. France, M'emoire* 60 (1979) 173–180.

[43] J. Tomlin, A suggested extension of special ordered sets to non-separable non-convex programming problems, in: *Annals of Discrete Mathematics vol. 11: Studies on Graphs and Discrete Programming*, P. Hansen (Ed.), North Holland, 1981, pp. 359–370.

[44] H. Tuy, *Convex Analysis and Global Optimization*, Kluwer Academic Publishers, Dordrecht, 1998.

[45] N. Van Hang and N. Yen, On the problem of minimizing a difference of polyhedral convex functions under linear constraints, *J. Optim. Theory Appl.* 171 (2016) 617–642.

[46] J. Vielmaa, S. Ahmed and G. Nemhauser, Mixed-integer models for nonseparable piecewise-linear optimization: unifying framework and extensions, *Operations Research* 58 (2010) 303–315.

[47] J. Vielmaa, A. Keha and G. Nemhauser, Nonconvex lower semicontinuous piecewise linear optimization, *Discrete Optimization* 5 (2008) 467–488.

[48] C. Wen and X. Ma, A max-piecewise-linear neural network for function approximation, *Neurocomputing* 71 (2008) 843–852.

[49] K. Yamamura and A. Machida, An efficient algorithm for finding all DC solutions of piecewise-linear circuits, *Int. J. Circ. Theor. Appl.* 36 (2008) 3989–1000.

[50] H. Zhang and S. Wang, Global optimization of separable objective functions on convex polyhedra via piecewise-linear approximation, *Journal of Computational and Applied Mathematics* 197 (2006) 212–217.

Manuscript received 20 July 2023
revised 30 December 2023
accepted for publication 17 January 2024

Adil M. Bagirov
Centre for Smart Analytics, Institute of Innovation
Science and Sustainability
Federation University Australia
Ballarat, Australia E-mail address: a.bagirov@federation.edu.au

Julien Ugon
School of Information Technology, Faculty of Science
Engineering and Built Environment, Deakin University
Melbourne, Australia
E-mail address: julien.ugon@deakin.edu.au