ON GLOBAL OPTIMALITY CONDITIONS FOR D.C. MINIMIZATION PROBLEMS WITH D.C. CONSTRAINTS

ALEXANDER S. STREKALOVSKY

Matrosov Institute for System Dynamics & Control Theory SB RAS, Lermontov st., 134, Irkutsk, 664033, Russia

Abstract. The paper addresses the nonconvex nonsmooth optimization problem with the cost function, and equality and inequality constraints given by d.c. functions, i.e. represented as a difference of convex functions. The original problem is reduced to a problem without constraints with the help of the exact penalization theory. After that, the penalized problem is represented as a d.c. minimization problem without constraints, for which the new mathematical tools under the form of global optimality conditions (GOCs) are developed. The GOCs reduce the nonconvex problem in question to a family of convex (linearized with respect to the basic nonconvexities) problems. In addition, the GOCs are related to some nonsmooth form of the KKT-theorem for the original problem. Besides, the GOCs possess the constructive (algorithmic) property, which, when the GOCs are broken down, implies the producing of a feasible point that is better (in the original problem) than the one in question. The effectiveness of the GOCs is demonstrated by examples.

Keywords. D.C. functions; Exact penalty; Global optimality conditions; Lagrange function; KKT point.

1. INTRODUCTION

The majority of applied problems turn out to be nonconvex [1, 2] and, as mentioned in [3, 4, 5], result in the so-called d.c. optimization problems, where, for example, equality and inequality constraints, objective functionals are given by (d.c.) functions, represented as a difference of two convex functions. As well known the d.c. functions possess a number of popular properties. For example, they are quasidifferentiables [5, 6, 7, 8, 9].

The first publications on the simplest optimization problems with d.c. functions were appeared, in particular, in the former USSR [10, 11], see also the references from [4, 9].

On the other hand, from the theoretical view-point, the class $DC(\mathbb{R}^n)$ of the d.c. functions generated by the convex cone of convex functions forms a very attractive linear space which is closed with respect to standard operations arising in optimization, variational analysis [4, 6, 8, 12, 13], and applications [3, 4, 5, 9, 14, 15, 16, 17, 18].

Moreover, the class $DC(\mathbb{R}^n)$ includes the well-known spaces of twice differentiable functions, power and trigonometric polynomials, etc. In addition, any continuous function on a compact (for instance, in $\mathbb{R}^n$) can be approximated by a d.c. function with any desired accuracy. Consequently, any continuous optimization problem (on a compact set) can be approximated by an optimization problem with the d.c. data [4, 5, 9, 14, 16, 17].

Furthermore, the main focus in modern optimization has been shifted to numerical methods designed to address problems of practical interest.

However, it is important to highlight that the field of optimization problems should be separated into two parts: convex and nonconvex [1, 2, 19]. If we look at the problems from the viewpoint of numerical treatment and take into consideration the available information, it is easy to observe that under minimal computability assumptions convex optimization problems are computationally tractable [1, 2, 15, 19]. It means that computational efforts required for finding a solution at a given accuracy grows moderately with the dimension of the problem in question.

In contrast to this, in general, nonconvex problems of general form have a lot (often a huge number) of local extrema and stationary points located rather far from a global solution [1, 2, 3, 4, 5, 14, 19, 20, 21, 22, 23, 24, 25, 26]. As a consequence, the classical optimization methods [1, 2, 19] often turn out to be inoperative and ineffective when it comes to finding just a global solution in nonconvex problems. This is due to the fact that most of the classical algorithms are based on the contemporary optimality condition theory, in particular, the famous KKT theorem, and therefore those methods fail to escape local pitfalls and stationary (critical) vectors while improving a current value of objective functional at the same time. Besides, a direct application of classical methods might yield unpredictable results, sometimes distracting researchers away from a desired solution. Hence, nowadays it is important to equip the modern extremum theory with new mathematical tools that help not only escape a local pitfall but also reach a global solution in nonconvex problems.

In our early papers [9, 27] such tools were proposed for the standard (canonical) d.c. optimization problems such as d.c. minimization, convex maximization, reverse-convex problems etc. Recently, we have developed the Global Optimality Conditions (GOC) for more general d.c. optimization program with d.c. inequality constraints and the smooth data, showing that the classical optimality conditions can follow from the GOCs, which, in turn, can help jump out of local pits and stationary (KKT) vectors [9, 18, 27, 28]. It is worth noting that the present paper continues and develops further the results of the previous investigations [9, 29, 30, 31].

In this paper, following the statement of the nonconvex nonsmooth d.c. optimization Problem ($\mathscr{P}$) with equality and inequality constraints, we reduce ($\mathscr{P}$) to a problem without constraints with the help of the Exact Penalization Theory [1, 2, 6, 14, 15, 19, 20, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41]. The latter problem also turns out to be a d.c. minimization program as we show in Section 4. Further, the necessary GOCs are given in Theorem 4.1, after which various properties of the GOCs are discussed. In Section 5, we investigate the interrelations between the GOCs and the KKT conditions in nonsmooth original Problem ($\mathscr{P}$), in particular, we highlight connections between their Lagrange multipliers. In Section 6, we study situations when the GOCs become sufficient. Besides, Theorems 4.1, 5.1, and 6.1 impose conditions on the value of the penalty parameter. The effectiveness of the GOCs developed is demonstrated by examples.

2. PROBLEM STATEMENT

Consider the following problem:

$$(\mathscr{P}): \quad \begin{aligned} f_0(x) &:= g_0(x) - h_0(x) \downarrow \min_x, \quad x \in S, \\ f_i(x) &:= g_i(x) - h_i(x) \le 0, \quad i \in I = \{1, \dots, m\}, \\ f_i(x) &:= g_i(x) - h_i(x) = 0, \quad i \in \mathscr{E} = \{m+1, \dots, l\}; \end{aligned} \Bigg\}$$

where the functions $g_i(\cdot)$, $h_i(\cdot)$, $i \in \{0\} \cup I \cup \mathscr{E}$, are convex on $\mathbb{R}^n$, so that the functions $f_i(\cdot)$, $i \in \{0\} \cup I \cup \mathscr{E}$, are the d.c. functions [3, 4, 5, 9, 14, 17]. Recall that any continuous function can be approximated by a d.c. function with any desirable accuracy. Besides, assume that the set $S \subset \mathbb{R}^n$ is convex and compact.

Furthermore, suppose that the set $Sol(\mathscr{P})$ of global solutions to Problem $(\mathscr{P})$, $Sol(\mathscr{P}) := \{x \in \mathscr{F} \mid f_0(x) = \mathscr{V}(\mathscr{P})\}$ and the feasible set $\mathscr{F}$ of Problem $(\mathscr{P})$, $\mathscr{F} := \{x \in S \mid f_i(x) \le 0, i \in I, f_i(x) = 0, i \in \mathscr{E}\}$, are non-empty. Additionally, in what follows the optimal value $\mathscr{V}(\mathscr{P})$ of Problem $(\mathscr{P})$ is supposed to be finite:

$$\mathscr{V}(\mathscr{P}) := \inf(f_0, \mathscr{F}) := \inf_x \{f_0(x) \mid x \in \mathscr{F})\} > -\infty.$$

3. THE EXACT PENALTY

In this section, we introduce the following penalty function $W(\cdot)$ for Problem $(\mathscr{P})$

$$W(x) := \max\{0, f_1(x), \dots, f_m(x)\} + \sum_{j \in \mathscr{E}} |f_j(x)|, \tag{3.1}$$

and, along with Problem $(\mathscr{P})$, consider the penalized problem without the inequality and equality constraints:

$$(\mathscr{P}_\sigma): \qquad F_\sigma(x) \downarrow \min_x, \quad x \in S, \tag{3.2}$$

where $\sigma \ge 0$ is a penalty parameter, and the penalized function

$$F_\sigma(x) := f_0(x) + \sigma W(x), \tag{3.3}$$

is the cost function of the auxiliary problem $(\mathscr{P}_\sigma)$–(3.2).

It is well-known that if $z \in Sol(\mathscr{P}_\sigma)$, and z is feasible in $(\mathscr{P})$, i.e. $z \in \mathscr{F}$, then z is a global solution to $(\mathscr{P})$: $z \in Sol(\mathscr{P})$ [1, 2, 6, 14, 15, 19, 20, 32, 33, 34, 35, 41]. It is worth mentioning that, generally, the inverse proposition does not hold.

Hence, the key feature of the exact penalization (EP) theory is the existence of a threshold value $\sigma_* \ge 0$ of the penalty parameter $\sigma \ge 0$ for which $Sol(\mathscr{P}_\sigma) \subset Sol(\mathscr{P})$, $\forall \sigma \ge \sigma_*$. The latter means that, for $\sigma \ge \sigma_*$, Problems $(\mathscr{P})$ and $(\mathscr{P}_\sigma)$ are equivalent: $Sol(\mathscr{P}) = Sol(\mathscr{P}_\sigma)$ (see [6, 33], and [14, Chapt. VII, Lemma 1.2.1]).

On the other hand, the existence of the threshold exact penalty parameter $\sigma_* \ge 0$ implies that instead of solving a sequence of unconstrained problems with $\sigma_k \to \infty$ [19, 1, 6, 2], we need to solve only a single unconstrained problem.

Hence, the proof of existence of the exact penalty threshold $\sigma_* \ge 0$ is a key moment in the investigation of relations between Problems $(\mathscr{P})$ and $(\mathscr{P}_\sigma)$ [6, 15, 32, 33, 34, 35, 41].

Recall that under various constraint qualification (CQ) conditions (MFCQ, etc. [6, 15, 32, 33, 34, 35, 40, 41, 42, 43]), the error bound properties [1, 2, 6, 15, 20, 32, 33, 34, 35, 40, 41, 42, 43], the metric sub-regularity conditions, calmness of constraints systems can help prove the

existence of the exact penalty threshold $\sigma_* \geq 0$ for the local solution as well as for the global one [6, 15, 32, 33, 34, 35, 41].

Assume that some regularity conditions, which ensure the existence of such threshold value $\sigma_* \geq 0$ of the penalty parameter, are fulfilled.

4. GLOBAL OPTIMALITY CONDITIONS (GOC)

Before all, we will show that the cost function $F_\sigma(\cdot)$ of Problem $(\mathscr{P}_\sigma)$ is a d.c. function, i.e., it can be represented as a difference of convex functions. Indeed, since [4, 5, 9, 17]

$$|f_i(x)| = 2\max\{g_i(x), h_i(x)\} - [g_i(x) + h_i(x)],$$

it can be readily seen that

$$F_\sigma(x) \overset{\triangle}{=} f_0(x) + \sigma \max\{0, f_i(x), i \in I\} + \sigma \sum_{i \in \mathscr{E}} |f_i(x)| = G_\sigma(x) - H_\sigma(x), \qquad (4.1)$$

where

$$H_\sigma(x) := h_0(x) + \sigma \left[\sum_{i \in I} h_i(x) + \sum_{j \in \mathscr{E}} (g_j(x) + h_j(x)) \right], \qquad (4.2)$$

$$G_\sigma(x) := F_\sigma(x) + H_\sigma(x)$$
$$= g_0(x) + \sigma \max \left\{ \sum_{j \in I} h_j(x); \left[g_i(x) + \sum_{\substack{j \in I \\ j \neq i}}^{j \neq i} h_j(x) \right], i \in I \right\} \qquad (4.3)$$
$$+ 2\sigma \sum_{i \in \mathscr{E}} \max\{g_i(x); h_i(x)\}.$$

It is easy to see that $G_\sigma(\cdot)$ and $H_\sigma(\cdot)$ are both convex functions [8, 12, 14] so that the function $F_\sigma(\cdot)$ is a d.c. function, as claimed. Besides, it is clear that, for a feasible (in $(\mathscr{P})$) point $z \in S$, we have

$$W(z) \overset{\triangle}{=} \max\{0, f_1(z), \ldots, f_m(z)\} + \sum_{i \in \mathscr{E}} |f_i(z)| = 0.$$

Therefore, for $\zeta := f_0(z)$, we obtain

$$F_\sigma(z) = f_0(z) + \sigma W(z) = f_0(z) = \zeta. \qquad (4.4)$$

Theorem 4.1. *Let a feasible point $z \in \mathscr{F}$, $\zeta := f_0(z)$, be a solution to Problem $(\mathscr{P})$ and $\sigma \geq \sigma_* > 0$, where $\sigma_* \geq 0$ is a threshold value of the penalty parameter such that $Sol(\mathscr{P}) = Sol(\mathscr{P}_\sigma) \ \forall \sigma \geq \sigma_*$. Then, for every pair $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ such that*

$$H_\sigma(y) = \beta - \zeta, \qquad (4.5)$$

the following inequality holds

$$G_\sigma(x) - \beta \geq \langle H'_\sigma(y), x - y \rangle \quad \forall x \in S, \qquad (4.6)$$

for every subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the function $H_\sigma(\cdot)$ at the point y.

Remark 4.1. In view of the Moreau-Rockafellar Theorem [8, 12, 14] and (4.2), we have

$$H'_\sigma(y) = h'_0(y) + \sigma \left[\sum_{i \in I} h'_i(y) + \sum_{j \in \mathscr{E}} (g'_j(y) + h'_j(y)) \right], \qquad (4.7)$$

where $h'_i(y) \in \partial h_i(y)$, $i \in \{0\} \cup I \cup \mathscr{E}$, $g'_j(y) \in \partial g_j(y)$, $j \in \mathscr{E}$, are some subgradients of the corresponding functions at the point y in the classical (convex) sense.

Proof. Since $z \in Sol(\mathscr{P})$ and $\sigma \geq \sigma_* \geq 0$, then, according to [14, 15, 32, 33, 34, 35, 41] (see [6] and[14, Chapt. VII, Lemma 1.2.1]), $z \in Sol(\mathscr{P}_\sigma)$, $\forall \sigma \geq \sigma_* > 0$. It means that (see (4.1)–(4.4)) $\zeta = f_0(z) = F_\sigma(z) \leq F_\sigma(x) = G_\sigma(x) - H_\sigma(x)$, $\forall x \in S$. Hence, $z \in Sol(\mathscr{P}_\sigma)$. Then, for every tuple (y, β) satisfying (4.5), we have

$$0 \leq G_\sigma(x) - H_\sigma(x) - \zeta = G_\sigma(x) - H_\sigma(x) + H_\sigma(y) - \beta \quad \forall x \in S,$$

whence,

$$G_\sigma(x) - \beta \geq H_\sigma(x) - H_\sigma(y) \geq \langle H'_\sigma(y), x - y \rangle, \quad \forall x \in S$$

for any subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the function $H_\sigma(\cdot)$ at the point y thanks to the convexity of $H_\sigma(\cdot)$. The latter completes the proof. $\square$

Remark 4.2. It is not difficult to notice that Theorem 4.1 reduces the solution of the nonconvex Problem $(\mathscr{P}_\sigma)$ to an investigation of the family of the convex (linearized) problems

$$(\mathscr{P}_\sigma L(y)): \qquad \Phi_{\sigma y}(x) := G_\sigma(x) - \langle H'_\sigma(y), x \rangle \downarrow \min_x, \quad x \in S, \qquad (4.8)$$

depending on the pairs $(y, \beta) \in I\!R^{n+1}$, which fulfill equation (4.5) or, which is the same, due to (4.7),

$$(\mathscr{P}_\sigma L(y)): \qquad \Phi_{\sigma y}(x) := G_\sigma(x) - \langle h'_0(y) + \sigma \Big[\sum_{i \in I} h'_i(y) \\ + \sum_{j \in \mathscr{E}} (g'_j(y) + h'_j(y)) \Big], x \rangle \downarrow \min_x, \quad x \in S; \qquad (4.8')$$

for every collection of subgradients $h'_i(y) \in \partial h_i(y), i \in \{0\} \cup I \cup \mathscr{E}$, $g'_j(y) \in \partial g_j(y), j \in \mathscr{E}$, of the corresponding functions $h_i(\cdot)$, $g_j(\cdot)$ at the vector $y \in I\!R^n$.

It is worth noting that the linearization is carried out with respect to the "unified" nonconvexity of Problem $(\mathscr{P})$ accumulated by the function $H_\sigma(\cdot)$ (see $(\mathscr{P})$–(3.1) and (4.2)) that includes all the functions $h_i(\cdot), i \in \{0\} \cup I \cup \mathscr{E}$, $g_j(\cdot), j \in \mathscr{E}$, which generate all nonconvexities in Problems $(\mathscr{P})$ and $(\mathscr{P}_\sigma)$ (according to the representations (4.1)–(4.3)).

Hence, the verification of the principal inequality (4.6) can be performed by solving the linearized problems $(\mathscr{P}_\sigma L(y))$ and varying the parameters (y, β) satisfying (4.5). Besides, we have to verify (4.6), which can be rewritten as follows

$$\mathscr{V}(\mathscr{P}_\sigma L(y)) \geq \beta - \langle H'_\sigma(y), y \rangle, \qquad (4.6')$$

where $\mathscr{V}(\mathscr{P}_\sigma L(y))$ is the optimal value of the linearized problem $(\mathscr{P}_\sigma L(y))$ (see [29, 30, 31]).

Remark 4.3. Suppose that we found a triple (y, β, u) and only one subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$, where $(y, \beta) \in I\!R^n \times I\!R$, $H_\sigma(y) = \beta - \zeta$, $u \in S$, such that the principal inequality (4.6) is violated, i.e.,

$$0 > G_\sigma(u) - \beta - \langle H'_\sigma(y), u - y \rangle,$$

where $H'_\sigma(y)$ is presented by some collection of subgradients $h'_i(y) \in \partial h_i(y), i \in \{0\} \cup I \cup \mathscr{E}$, $g'_j(y) \in \partial g_j(y), j \in \mathscr{E}$, of the corresponding functions $h_i(\cdot), g_j(\cdot)$, according to (4.7).

Then, using equation (4.5) and the convexity of function $H_\sigma(\cdot)$, we derive

$$0 > G_\sigma(u) - \beta - H_\sigma(u) + H_\sigma(y) = F_\sigma(u) - \zeta = F_\sigma(u) - F_\sigma(z),$$

or, $F_\sigma(z) > F_\sigma(u)$, $z \in \mathscr{F}$, $u \in S$. Hence, the point z can not be a solution to $(\mathscr{P}_\sigma)$.

Moreover, if z and u are feasible in $(\mathscr{P})$, $z, u \in \mathscr{F}$, and $W(u) = 0 = W(z)$, we obtain $f_0(z) = F_\sigma(z) > F_\sigma(u) = f_0(u)$. It means that $z \notin Sol(\mathscr{P})$ and $u \in \mathscr{F}$ is a vector better than $z \in \mathscr{F}$.

Hence, the conditions (4.5)–(4.6) of Theorem 4.1 possess the classical constructive (algorithmic) property: once the conditions are violated, one can find a feasible vector which is better than the point under investigation.

Let us use an example to demonstrate the effectiveness of this property.

Example 4.1. Consider the problem ([29, Example 4.3])

$$\left.\begin{array}{l} f_0(x) = \frac{1}{2}(x_1 - 4)^2 + (x_2 + 2)^2 \downarrow \min_x, \\ f_1(x) = (x_1 - 1)^2 - (x_2 + 1)^2 \leq 0, \\ f_2(x) = (x_2 - 2)^2 - (x_1 + 2)^2 \leq 0. \end{array}\right\} \tag{4.9}$$

One can see that the point $z_* = (4, -2)^\top$ is the global minimizer of the strongly convex function $f_0(\cdot)$ on $\mathbb{R}^2$, and $f_0(z_*) = 0$ is a lower bound for $\mathcal{V}(4.9) = \inf(f_0, \mathcal{F}) \geq 0$. Here, the feasible set of (4.9)

$$\mathcal{F} = \{x \in \mathbb{R}^2 \mid f_1(x) \leq 0, \quad f_2(x) \leq 0\}, \quad z_* \notin \mathcal{F},$$

is nonconvex and coincides with the union $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ of two convex parts:

$$\mathcal{F}_1 = \{x \in \mathbb{R}^2 \mid x_1 + x_2 = 0\},$$
$$\mathcal{F}_2 = \{x \in \mathbb{R}^2 \mid x_1 + x_2 \geq 0, \ x_2 - x_1 - 4 \leq 0, \ x_1 - x_2 - 2 \leq 0\},$$

since $f_1(x) = (x_1 - x_2 - 2)(x_1 + x_2) \leq 0$, $f_2(x) = (x_2 - x_1 - 4)(x_2 + x_1) \leq 0$.

Consider a feasible point $z = (\frac{4}{3}, -\frac{2}{3})^\top \in \mathcal{F}$, which satisfies the KKT-conditions with $\lambda_0 = 1$, $\lambda_1 = 4 > 0$, $\lambda_2 = 0$, $\zeta := f_0(z) = 5\frac{1}{3}$, $f_1(z) = 0$ and $f_2(z) = -4 < 0$.

Let us show that point z is not a global solution to (4.9) with the help of Theorem 4.1. To this end, let us choose the following d.c. representation of the data of the problem (4.9)

$$\left.\begin{array}{l} g_0(x) = f_0(x), \ h_0(x) \equiv 0, \ f_i(x) = g_i(x) - h_i(x), \ i = 1, 2, \\ g_1(x) = (x_1 - 1)^2, \ h_1(x) = (x_2 + 1)^2, \\ g_2(x) = (x_2 - 2)^2, \ h_2(x) = (x_1 + 2)^2. \end{array}\right\} \tag{4.10}$$

Then, according to (4.1)–(4.3), we get

$$\left.\begin{array}{l} W(x) \overset{\triangle}{=} \max\{0, f_1(x), f_2(x)\}, \\ F_\sigma(x) = f_0(x) + \sigma W(x) = G_\sigma(x) - H_\sigma(x); \end{array}\right\} \tag{4.11}$$

$$H_\sigma(x) \overset{\triangle}{=} h_0(x) + \sigma[h_1(x) + h_2(x)] = \sigma[(x_2 + 1)^2 + (x_1 + 2)^2]; \tag{4.12}$$

$$\begin{aligned} G_\sigma(x) &= g_0(x) + \sigma \max\{h_1(x) + h_2(x); g_1(x) + h_2(x); g_2(x) + h_1(x)\} \\ &= \frac{1}{2}(x_1 - 4)^2 + (x_2 + 2)^2 + \sigma \max\left\{(x_2 + 1)^2 + (x_1 + 2)^2; \right. \\ &\quad \left. (x_1 - 1)^2 + (x_1 + 2)^2; (x_2 - 2)^2 + (x_2 + 1)^2 \right\} \end{aligned} \tag{4.13}$$

Now, set $\sigma := 4 = \lambda_1 + \lambda_2 = \lambda_1(z) + \lambda_2(z)$ and $y := (\frac{3}{2}, -1)^\top \notin \mathcal{F}$. Then, we obtain

$$H_\sigma(y) = 4[(y_2 + 1)^2 + (y_1 + 2)^2] = 4 \cdot (3\frac{1}{2})^2 = 49;$$
$$\beta = H_\sigma(y) + \zeta = 49 + 5\frac{1}{3} = 54\frac{1}{3}.$$

Furthermore, we merely compute

$$\nabla H_\sigma(x) = 4(2(x_1 + 2), 2(x_2 + 1))^\top = 8(x_1 + 2, x_2 + 1)^\top,$$
$$\nabla H_\sigma(y) = 8\left(\frac{3}{2} + 2, 0\right)^\top = (28, 0)^\top.$$

Finally, choose $u = \left(\frac{4}{3}, -\frac{4}{3}\right)^\top \in \mathscr{F}$, and compute $u - y = -\left(\frac{1}{9}, \frac{1}{3}\right)^\top$ and

$$\beta + \langle \nabla H_\sigma(y), u - y \rangle = 54\frac{1}{3} + \left\langle (28, 0)^\top, \left(-\frac{1}{9}, -\frac{1}{3}\right)^\top \right\rangle = 54\frac{1}{3} - 3\frac{1}{9} = 51\frac{2}{9}.$$

Thus, in order to verify the principal inequality (4.6) of Theorem 4.1, we need the value of $G_\sigma(u)$. Therefore, using the representation (4.13), we get

$$\begin{aligned}
G_\sigma(u) &= \tfrac{1}{2}(u_1 - 4)^2 + (u_2 + 2)^2 + 4\max\left\{(u_2 + 1)^2 + (u_1 + 2)^2;\right. \\
&\quad \left.(u_1 - 1)^2 + (u_1 + 2)^2; (u_2 - 2)^2 + (u_2 + 1)^2\right\} \\
&= \tfrac{32}{9} + \tfrac{4}{9} + 4\max\left\{\left(-\tfrac{1}{3}\right)^2 + \left(\tfrac{10}{3}\right)^2; \left(\tfrac{1}{3}\right)^2 + \left(\tfrac{10}{3}\right)^2; \left(-\tfrac{10}{3}\right)^2 + \left(-\tfrac{1}{3}\right)^2\right\} \\
&= 4 + 4 \cdot \tfrac{101}{9} = 48\tfrac{8}{9}.
\end{aligned}$$

So, we see that $G_\sigma(u) = 48\frac{8}{9} < 51\frac{2}{9} = \beta + \langle \nabla H_\sigma(y), u - y \rangle$.

In other words, the KKT point $z = \left(\frac{4}{3}, -\frac{2}{3}\right)^\top$ is not a global solution to (4.9) in virtue of Theorem 4.1. In addition, $f_0(u) = 4 < f_0(z) = 5\frac{1}{3}$.

Remember that (see [29, Example 4.3]), the max-merit function

$$\Theta(x, \zeta) := \max\{f_0(x) - \zeta; f_1(x), \ldots, f_m(x)\}$$

does not distinguish the points z and u. Moreover, the GOCs of Theorems 4.1 and 4.2 in [29] fail to improve the point $z = \left(\frac{4}{3}, -\frac{2}{3}\right)^\top$. On the other hand, the GOCs constructed on the base of the classical Lagrange function [29, Theorems 5.1, 5.2] helped us conclude that the pair (z, λ), $(\lambda = (\lambda_0, \lambda_1, \lambda_2)^\top = (1, 4, 0)^\top)$ is not a saddle point of the Lagrange function. Besides, those GOCs provided the point $\hat{u} = \left(\frac{4}{3}, 0\right)^\top$, which improves the value of the Lagrange function: $\mathscr{L}(\hat{u}, \lambda) < f_0(z) = \mathscr{L}(z, \lambda)$, but not the value of the goal function $f_0(\cdot)$ to the original problem $(\mathscr{P})$: $f_0(u) = 7\frac{5}{9} > 5\frac{1}{3} = f_0(z)$ (see [29, Example 5.2]).

Thus, we see that the cost function $F_\sigma(\cdot)$ of Problem $(\mathscr{P}_\sigma)$ has some advantages with respect to $\Theta(x, \zeta)$ and $\mathscr{L}(x, \lambda)$ (see also [30]).

5. THE LAGRANGE MULTIPLIERS

In this section, we address possible relations between the conditions (4.5)–(4.6) of Theorem 4.1 and the classical optimality conditions [1, 2, 5, 8, 9, 12, 13, 14, 15, 16, 17, 19], in particular, the KKT theorem for Problem $(\mathscr{P})$.

For this purpose, suppose that a feasible (in Problem $(\mathscr{P})$) point z ($z \in \mathscr{F}$) satisfies the conditions (4.5)–(4.6) of Theorem 4.1. First, set $y = z$ in (4.5)–(4.6). Then we immediately derive that $\beta := H_\sigma(z) + \zeta = G_\sigma(z)$.

Further, it follows from (4.6) that

$$G_\sigma(x) - G_\sigma(z) \geq \langle H'_\sigma(z), x - z \rangle, \quad \forall x \in S,$$

for each subgradient $H'_\sigma(z) \in \partial H_\sigma(z)$ (see (4.7)) of $H_\sigma(\cdot)$ at z. It implies that the point z (satisfying (4.5)–(4.6)) is a solution to the linearized (convex) problem

$$(\mathscr{P}_\sigma L(z)): G_\sigma(x) - \langle H'_\sigma(z), x \rangle \downarrow \min_x, \quad x \in S,$$

for every subgradient $H'_\sigma(z)$ of $H_\sigma(\cdot)$ at z. Since $(\mathscr{P}_\sigma L(z))$ is a convex optimization problem, it is well-known that the following inclusion is the necessary and sufficient optimality condition

for z to be a solution to $(\mathscr{P}_\sigma L(z))$:

$$0_n \in \partial G_\sigma(z) - H'_\sigma(z) + N(z \mid S).$$

In addition, taking into consideration the inclusion $\forall H'_\sigma(z) \in \partial H_\sigma(z)$, we obtain

$$\partial H_\sigma(z) \subset \partial G_\sigma(z) + N(z \mid S). \tag{5.1}$$

When $S = \mathbb{R}^n$, the inclusion (5.1) implies that

$$\partial H_\sigma(z) \subset \partial G_\sigma(z), \tag{5.1'}$$

which is the necessary optimality condition for Problem $(\mathscr{P}_\sigma)$ with $S = \mathbb{R}^n$ [1, 9, 14, 17]. Thus, the conditions (4.5)–(4.6) of Theorem 4.1 entail the well-known optimality conditions (5.1) and (5.1') [19, 1, 15, 14, 2, 12, 13, 8] for Problem $(\mathscr{P}_\sigma)$.

On the other hand, it is easy to see the drawback of (5.1) and (5.1'), where the function $G_\sigma(\cdot)$ is not smooth (see (4.3)), while the entries of $(\mathscr{P})$ can be differentiable.

In order to overcome this shortcoming, let us apply Lemma 4.1 from [29] (see also [1, 16, 2]), according to which Problem $(\mathscr{P}_\sigma L(z))$ amounts to the following problem (see (4.3) and (4.7)):

$$\left.\begin{array}{l} g_0(x) - \langle H'_\sigma(z), x \rangle + \sigma\gamma + 2\sigma \sum\limits_{j \in \mathscr{E}} t_j \downarrow \min\limits_{(x,\gamma,t)}, \ x \in S, \\[2mm] g_i(x) + \sum\limits_{\substack{j \in I \\ j \neq i}} h_j(x) \leq \gamma, \ i \in I; \quad \sum\limits_{j \in I} h_j(x) \leq \gamma, \ \gamma \in \mathbb{R}; \\[2mm] g_k(x) \leq t_k, \ h_k(x) \leq t_k, \ k \in \mathscr{E}, \ t = (t_{m+1}, \ldots, t_l)^\top \in \mathbb{R}^{l-m}. \end{array}\right\} \tag{5.2}$$

Note, at first, that problem (5.2) is a convex optimization problem, where the variables are $(x, \gamma, t) \in \mathbb{R}^{n+1} \times \mathbb{R}^{l-m}$. Secondary, problem (5.2) has only inequality constraints that distinguishes it from the original Problem $(\mathscr{P})$.

In addition, if the entries of Problem $(\mathscr{P})$ are smooth, auxiliary problem (5.2) remains smooth as well. Besides, the number $(l - m + 1)$ of supplementary variables $(\gamma, t_{m+1}, \ldots, t_l)$ does not exceed $(n+1)$. Otherwise, Problem $(\mathscr{P})$ would be overdetermined or inconsistent.

Finally, it can be readily seen that $\forall x \in S$ there exist a number $\gamma = \gamma(x) \in \mathbb{R}$ and a vector $t = t(x) = (t_{m+1}, \ldots, t_l)^\top \in \mathbb{R}^{l-m}$ such that the inequality constraints in (5.2) are strongly satisfied (i.e., in particular, $g_k(x) < t_k(x)$, $h_k(x) < t_k(x)$, $k \in \mathscr{E}$). It means that, in problem (5.2), the Slater's condition holds.

Therefore, in the Lagrange function of problem (5.2), we have $\mu_0 = 1$ and, besides, the corresponding KKT-conditions become necessary and sufficient for the tuple (z, γ_*, t_*) to be a solution to the convex problem (5.2). In particular, the point z, satisfying (4.5)–(4.6), must be such a solution, since $z \in Sol(\mathscr{P}_\sigma L(z))$.

For simplicity of presentation, we suppose $S = \mathbb{R}^n$. Then the Lagrange function for problem (5.2) takes the following form

$$\mathscr{L}(x, \gamma, t; \mu_1, \ldots, \mu_m, \mu_{m+1}, \eta_{m+1}, \ldots, \eta_l, \nu_{m+1}, \ldots, \nu_l)$$
$$= g_0(x) - \langle H'_\sigma(z), x \rangle + \sigma\gamma + 2\sigma \sum_{k \in \mathscr{E}} t_k + \sum_{i \in I} \mu_i \left[g_i(x) + \sum_{\substack{j \in I \\ j \neq i}} h_j(x) - \gamma \right] \tag{5.3}$$
$$+ \mu_{m+1} \left[\sum_{j \in I} h_j(x) - \gamma \right] + \sum_{k \in \mathscr{E}} \eta_k(g_k(x) - t_k) + \sum_{k \in \mathscr{E}} \nu_k(h_k(x) - t_k).$$

Then the KKT system for $z \in Sol(\mathscr{P}_\sigma L(z))$, i.e. for $(z, \gamma_*, t_*) \in Sol(5.2)$ with

$$\gamma_* = \max\left\{ \sum_{i \in I} h_i(z); \left[g_i(z) + \sum_{j \in I}^{j \neq i} h_j(z) \right], i \in I \right\}, \\ t_{k*} = \max\{ g_k(z); h_k(z) \}, \ k \in \mathscr{E}; \tag{5.4}$$

can be written as follows. There exists a vector $(\mu_1, \dots, \mu_m, \mu_{m+1}, \eta_{m+1}, \dots, \eta_l, \ v_{m+1}, \dots, v_l)$ $\in \mathbb{R}_+^{m+1} \times \mathbb{R}_+^{2(l-m)}$, such that the linear complementarity conditions hold true:

$$\mu_i \left[g_i(z) + \sum_{j \in I}^{j \neq i} h_j(z) - \gamma_* \right] = 0, \ \mu_i \geq 0, \quad i \in I; \\ \mu_{m+1} \left[\sum_{j \in I} h_j(z) - \gamma_* \right] = 0, \ \mu_{m+1} \geq 0, \\ \eta_k[g_k(z) - t_{k*}] = 0 = v_k[h_k(z) - t_{k*}], \ \eta_k \geq 0, \ v_k \geq 0, \ k \in \mathscr{E}. \tag{5.5}$$

In addition, the following equations take place

$$\frac{\partial \mathscr{L}(z, \gamma_*, t_*)}{\partial \gamma} = \sigma - \mu_{m+1} - \sum_{i \in I} \mu_i = 0, \quad \text{i.e.} \ \mu_{m+1} + \sum_{i \in I} \mu_i = \sigma; \tag{5.6}$$

$$\frac{\partial \mathscr{L}(z, \gamma_*, t_*)}{\partial t_k} = 2\sigma - \eta_k - v_k = 0, \quad \text{i.e.} \ \eta_k + v_k = 2\sigma, \quad k \in \mathscr{E}. \tag{5.7}$$

Finally, from the inclusion $0_n \in \partial_x \mathscr{L}(z, \gamma_*, t_*)$ with the help of the Moreaux-Rockafellar's theorem, we derive

$$g_0'(z) - H_\sigma'(z) + \mu_{m+1} \sum_{i \in I} h_i'(z) + \sum_{i \in I} \mu_i \left[g_i'(z) + \sum_{j \in I}^{j \neq i} h_j'(z) \right] \\ + \sum_{k \in \mathscr{E}} \eta_k g_k'(z) + \sum_{k \in \mathscr{E}} v_k h_k'(z) = 0_n \in \mathbb{R}^n, \tag{5.8}$$

for some collection of subgradients $g_i'(z) \in \partial g_i(z)$, $h_i'(z) \in \partial h_i(z)$, $i \in \{0\} \cup I \cup \mathscr{E}$, of the corresponding functions $h_i(\cdot)$, $g_j(\cdot)$ at the point z. Further, on account of (4.7) with $y = z$, we obtain

$$H_\sigma'(z) = h_0'(z) + \sigma \left[\sum_{i \in I} h_i'(z) + \sum_{k \in \mathscr{E}} (g_k'(z) + h_k'(z)) \right]. \tag{4.7'}$$

Then, (5.8) entails the equality

$$0_n = g_0'(z) - h_0'(z) + (\mu_{m+1} - \sigma) \sum_{i \in I} h_i'(z) + \sum_{i \in I} \mu_i \left[g_i'(z) + \sum_{j \in I}^{j \neq i} h_j'(z) \right] \\ + \sum_{k \in \mathscr{E}} [\eta_k g_k'(z) + v_k h_k'(z)] - \sigma \sum_{k \in \mathscr{E}} [g_k'(z) + h_k'(z)]. \tag{5.8'}$$

Furthermore, let us recall that a d.c. function $f(\cdot) = g(\cdot) - h(\cdot)$, where $g(\cdot)$ and $h(\cdot)$ are convex functions, possesses the directional derivative at any point $x \in int\ dom\ g \cap int\ dom\ h$.

For our local goals, we need another mathematical object, namely, the difference $g'(x) - h'(x)$ of two subgradients $g'(x) \in \partial g(x)$ and $h'(x) \in \partial h(x)$, which will be referred to as the d.c. subgradient of the function $f(\cdot)$ at a point x and denoted by $f'(x) = g'(x) - h'(x)$. Note that in the smooth case from the definition of d.c. subgradient of $f(\cdot)$ it follows the classical relation with gradients $\nabla f(x) = \nabla g(x) - \nabla h(x)$, so that the new definition is in a harmonic relation with the classical analysis.

Now return to the equality $(5.8')$, which, thanks to (5.6)–(5.7), $v_k = 2\sigma - \eta_k$ and the definition of d.c. subgradient can be rewritten in the following form:

$$0 = f_0'(z) + \left(\mu_{m+1} + \sum_{i \in I} \mu_i - \sigma\right) \sum_{i \in I} h_i'(z) + \sum_{i \in I} \mu_i \left[g_i'(z) - h'(z)\right]$$
$$+ \sum_{k \in \mathscr{E}} (\eta_k - \sigma) g_k'(z) + \sum_{k \in \mathscr{E}} (\sigma - \eta_k) h_k'(z).$$

Due to (5.6), the latter equality amounts to the relation

$$0 = f_0'(z) + \sum_{i \in I} \mu_i f_i'(z) + \sum_{k \in \mathscr{E}} (\eta_k - \sigma) f_k'(z), \tag{5.9}$$

which is clearly the principal equation of the KKT system for the original Problem $(\mathscr{P})$ at the point z with the Lagrange multipliers $\lambda_0 = 1$ and λ_i, $i \in I \cup \mathscr{E}$, satisfying the condition

$$\lambda_i = \mu_i \geq 0, \ i \in I, \ \lambda_k = \eta_k - \sigma, \ \eta_k \geq 0, \ k \in \mathscr{E}. \tag{5.10}$$

If the entries of Problem $(\mathscr{P})$ are smooth, the KKT equation (5.9) can be naturally transformed into the classical form [1, 2, 8, 13, 14, 19], where $f_i'(z) = \nabla f_i(z)$, $i \in I \cup \{0\} \cup \mathscr{E}$.

Now, let us show that, in Problem $(\mathscr{P})$, the complementarity conditions with $\lambda_i \geq 0$, $i \in I$, are also fulfilled.

Indeed, on account of (5.4) and (5.5), we have

$$\mu_i \left[g_i(z) + \sum_{j \in I}^{j \neq i} h_j(z) - \max\left\{\sum_{s \in I} h_s(z); [g_p(z) + \sum_{j \in I}^{j \neq p} h_j(z)], \ p \in I\right\}\right] = 0, \ i \in I,$$

which is equivalent to

$$\mu_i[f_i(z) - \max\{0; f_p(z), \ p \in I\}] = 0, \ i \in I.$$

Since the point z is feasible in $(\mathscr{P})$, the latter equalities amount to

$$\mu_i f_i(z) = 0, \ i \in I, \tag{5.11}$$

as was claimed above.

Hence, in the nonconvex Problem $(\mathscr{P})$, the Lagrange multipliers at the point z are completely defined by the Lagrange multipliers $(\mu_1, \ldots, \mu_m, \mu_{m+1}, \eta_{m+1}, \ldots, \eta_l, v_{m+1}, \ldots, v_l)$ of the auxiliary convex problem (5.2) at the point z and the penalty parameter $\sigma > 0$.

To sum up, we just proved a rather unexpected result, which at first was difficult to predict.

Proposition 5.1. *Let a feasible point z in Problem $(\mathscr{P})$ satisfy the conditions* (4.5)–(4.6) *of Theorem* 4.1. *Then, z is a KKT point in $(\mathscr{P})$ (fulfills (5.9) and (5.11)) with the Lagrange multipliers $\lambda_0 = 1$, λ_i, $i \in I \cup \mathscr{E}$, satisfying (5.6), (5.7), (5.10), where the Lagrange multipliers μ_i, $i \in I$, η_k, v_k, $k \in \mathscr{E}$, are provided by the auxiliary convex problem (5.2).*

In addition, we have now the supplementary information about relations between the vectors

$$\lambda = (\lambda_1, \ldots, \lambda_l)^\top \in \mathbb{R}^l, \ (\mu, \eta, v) \in \mathbb{R}_+^{m+1} \times \mathbb{R}_+^{2(l-m)},$$

and the penalty parameter $\sigma > 0$, provided by (5.5)–(5.7)

$$\sum_{i \in I} \lambda_i \leq \mu_{m+1} + \sum_{i \in I} \mu_i = \sigma, \quad \eta_k + v_k = 2\sigma \ \forall k \in \mathscr{E}. \tag{5.12}$$

It is clear that (5.12) is rather informative and helpful for computational treatment of Problems $(\mathscr{P})$, $(\mathscr{P}_\sigma)$, $(\mathscr{P}_\sigma L(z))$ and (5.2). Nevertheless, there arises a natural question whether we can

find a triple $(y, \beta, u) \in \mathbb{R}^{2n+1}$, satisfying (4.5), that violates the inequality (4.6). The answer is given by the next result.

Theorem 5.1. *Assume that a feasible in Problem* $(\mathscr{P})$ *point z is not an ε-solution to* $(\mathscr{P})$, *i.e.,*

$$\inf(f_0, \mathscr{F}) + \varepsilon = \mathscr{V}(\mathscr{P}) + \varepsilon < \zeta := f_0(z). \tag{5.13}$$

In addition, let a vector $w \in \mathbb{R}^n$ *satisfy the following inequality*

$$(\mathscr{H}): \qquad f_0(w) > \zeta - \varepsilon. \tag{5.14}$$

Then, for any penalty parameter $\sigma > 0$, *there exists a tuple* (y, β, u), $(y, \beta) \in \mathbb{R}^{n+1}$, $u \in \mathscr{F}$ *such that, for every subgradient* $H'_\sigma(y) \in \partial H_\sigma(y)$ *of the function* $H_\sigma(\cdot)$ *at the point y, the following conditions take place*

$$\left. \begin{array}{l} (a)\ H_\sigma(y) = \beta - \zeta + \varepsilon; \quad (b)\ G_\sigma(y) \leq \beta, \\ (c)\ G_\sigma(u) - \beta < \langle H'_\sigma(y), u - y \rangle. \end{array} \right\} \tag{5.15}$$

Proof. (A) It follows from (5.13) that there is a feasible in $(\mathscr{P})$ vector u $(u \in \mathscr{F})$ such that $f_0(u) + \varepsilon < \zeta = f_0(z)$. Indeed, if we had $\zeta \leq f_0(x) + \varepsilon, \forall x \in \mathscr{F}$, i.e., $\zeta \leq \mathscr{V}(\mathscr{P}) + \varepsilon$, then we would get the contradiction to (5.13). Furthermore, thanks to the equalities $W(z) = 0 = W(u)$, we derive that, for any $\sigma > 0$,

$$F_\sigma(u) + \varepsilon = f_0(u) + \varepsilon < \zeta = F_\sigma(z), \ \text{ or }\ G_\sigma(u) < H_\sigma(u) + \zeta - \varepsilon. \tag{5.16}$$

Now, define the convex set $C \subset \mathbb{R}^{n+1}$ as follows

$$C := epi(H_\sigma(\cdot) + \zeta - \varepsilon) \triangleq \{(x, \gamma) \mid H_\sigma(x) + \zeta - \varepsilon \leq \gamma\}.$$

Then, (5.16) is equivalent to the relation

$$(u, G_\sigma(u)) \notin C. \tag{5.16'}$$

On the other hand, from $z \in \mathscr{F}$, and $(\mathscr{H})$–(5.14), we obtain

$$\begin{array}{c} F_\sigma(z) - \varepsilon = f_0(z) - \varepsilon = \zeta - \varepsilon < f_0(w) \leq f_0(w) + \sigma W(w) = F_\sigma(w), \\ \text{or} \quad F_\sigma(w) > \zeta - \varepsilon \end{array} \tag{5.14'}$$

which can be rewritten in the form

$$G_\sigma(w) > H_\sigma(w) + \zeta - \varepsilon. \tag{5.17}$$

The latter inequality, in turn, amounts to the inclusion

$$(w, G_\sigma(w)) \in \text{int}\, C = \{(x, \gamma) \mid H_\sigma(x) + \zeta - \varepsilon < \gamma\}. \tag{5.17'}$$

(B) Now, taking into account the convexity of $C = epi[H_\sigma(\cdot) + \zeta - \varepsilon]$, and also (5.16') and (5.17'), it can be readily seen that there exists a vector $(y, \beta) \in \mathbb{R}^{n+1}$, which belongs to the open interval $](u, G_\sigma(u)); (w, G_\sigma(w))[\subset \mathbb{R}^{n+1}$ and to the boundary of C, $(y, \beta) \in \text{bd}\, C = \{(x, \gamma) \mid H_\sigma(x) + \zeta - \varepsilon = \gamma\}$.

In other words, there exists a number α, $0 < \alpha < 1$, such that $(y, \beta) = \alpha(u, G_\sigma(u)) + (1 - \alpha)(w, G_\sigma(w)) \in \text{bd}\, C$. More precisely, we have

$$y = \alpha u + (1 - \alpha)w, \quad \beta = \alpha G_\sigma(u) + (1 - \alpha)G_\sigma(w) = H_\sigma(y) + \zeta - \varepsilon. \tag{5.18}$$

Hence, (5.15) (a) is proven. Furthermore, with the help of convexity of $G_\sigma(\cdot)$ and (5.18), we derive

$$G_\sigma(y) \leq \alpha G_\sigma(u) + (1 - \alpha)G_\sigma(w) = \beta,$$

whence (5.15) (b) follows. Besides, (5.18) can be rewritten in the form

$$u = \alpha^{-1}[y - (1 - \alpha)w], \quad G_\sigma(u) = \alpha^{-1}[\beta - (1 - \alpha)G_\sigma(w)]. \tag{5.18$'$}$$

(C) Suppose now that the inequality (5.15) (c) is not satisfied for some subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the convex function $H_\sigma(\cdot)$ at the point y, i.e., $G_\sigma(u) - \beta \geq \langle H'_\sigma(y), u - y \rangle$. Then, with the help of (5.18$'$), we get

$$\begin{aligned}
0 &\geq \beta - G_\sigma(u) + \langle H'_\sigma(y), u - y \rangle \\
&= \beta - \alpha^{-1}[\beta - (1 - \alpha)G_\sigma(w)] + \langle H'_\sigma(y), \alpha^{-1}[y - (1 - \alpha)w] - y \rangle \\
&= \frac{1 - \alpha}{\alpha}[G_\sigma(w) - \beta] + \frac{1 - \alpha}{\alpha}\langle H'_\sigma(y), y - w \rangle.
\end{aligned}$$

Finally, using the convexity of $H_\sigma(\cdot)$, the equality (5.15) (a) (already proven), and also the assumption $(\mathcal{H})$–(5.14), we obtain the following chain

$$0 \geq \frac{1 - \alpha}{\alpha}[G_\sigma(w) - \beta + H_\sigma(y) - H_\sigma(w)] = \frac{1 - \alpha}{\alpha}[F_\sigma(w) - \zeta + \varepsilon] > 0,$$

which is impossible. Hence, the conjecture of the part (C) of the proof led us to the absurdity and, therefore, it is incorrect. Therefore, (5.15) (c) is proven, as well as Theorem 5.1. □

Remark 5.1. According to Remark 4.3, it is sufficient to find only one subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ violating inequality (4.6) in order to obtain the point u improving the point z in question: $F_\sigma(u) < F_\sigma(z)$. However, Theorem 5.1 states that it is possible to find a triple (y, β, u) violating (4.6) for every $H'_\sigma(y) \in \partial H_\sigma(y)$.

Remark 5.2. It is clear that Theorems 4.1 and 5.1 of this paper generalize the results of Theorems 1 and 2 from [30], respectively, (see also [29]) for the nonsmooth case and the presence of d.c. equality constraints. Besides, in Theorem 5.1, the point z is not an ε-solution to $(\mathcal{P})$ instead of a simple solution as it was in Theorem 2 in [30].

In addition, in Theorem 4.1, the inequality (4.6) takes place for every subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$. As a consequence, the vector z satisfying (4.4)–(4.6) turns out to be a solution to the linearized problem $(\mathcal{P}_\sigma L(z))$ for every subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$. This implies, in turn, that the KKT equation (5.9) has to be satisfied for all $H'_\sigma(z) \in \partial H_\sigma(z)$ (having the form (4.7$'$)), i.e., the equation

$$0 = g'_0(z) - h'_0(z) + \sum_{i \in I} \mu_i[g'_i(z) - h'_i(z)]) + \sum_{k \in \mathcal{E}} (\eta_k - \sigma)[g'_k(z) - h'_k(z)], \tag{5.9$'$}$$

must hold for all collections of subgradients $h'_i(z) \in \partial h_i(z)$, $i \in \{0\} \cup I$, $g'_k(z) \in \partial g_k(z)$, $h'_k(z) \in \partial h_k(z)$, $k \in \mathcal{E}$.

Remark 5.3. Now let us point out that the penalty parameter $\sigma \geq 0$ plays different roles in Theorems 4.1 and 5.1. Theorem 4.1 implies that value of penalty parameter should be greater than the threshold value $\sigma > \sigma_* \geq 0$, which provides the equivalence between $(\mathcal{P})$ and $(\mathcal{P}_\sigma)$, $Sol(\mathcal{P}) = Sol(\mathcal{P}_\sigma)$.

Meanwhile, in Theorem 5.1, the penalty parameter value can be arbitrary, but should remain positive: $\sigma > 0$. Regardless of that, one can find a triple (y, β, u) satisfying (5.15) and improving the current value $F_\sigma(z)$ of the goal function of the penalized program $(\mathcal{P}_\sigma)$ in question (see Remark 4.3). It is very important from the view-point of computational implementations of the methods based on Theorems 4.1 and 5.1.

Now, let us demonstrate the effectiveness of the GOCs of Theorems 4.1 and 5.1 by another example.

Example 5.1. Consider the problem

$$\left. \begin{array}{l} f_0(x) = x_1^2 - 2x_2^2 + x_3^2 \downarrow \min_x, x \in \mathbb{R}^3, \\ f_1(x) = x_3^2 - x_1^2 - x_2^2 = 0, \quad f_2(x) = 4x_1x_3 = 0, \quad -2 \le x_2 \le 1. \end{array} \right\} \tag{5.19}$$

It can be readily seen that the point $z = (0,0,0)^\top$, $\zeta := f_0(z) = 0$ is a degenerate KKT point in problem (5.19) since $f_1(z) = f_2(z) = 0$ and $\nabla f_0(z) = \nabla f_1(z) = \nabla f_2(z) = (0,0,0)^\top$. It is a difficult task to escape this point with the help of classical optimization methods [1, 14, 2]. However, it is not clear whether the KKT vector z is a global solution to (5.19) or not. In addition, note that $I = \varnothing$, $\mathcal{E} = \{1,2\}$.

Therefore, let us apply Theorems 4.1 and 5.1 to clarify the situation. It is easy to see that in problem (5.19) we have $g_0(x) = x_1^2 + x_3^2$, $h_0(x) = 2x_2^2$, $g_1(x) = x_3^2$, $h_1(x) = x_1^2 + x_2^2$. In addition, using the d.c. representation $f_2(x) = 4x_1x_3 = (x_1+x_3)^2 - (x_1-x_3)^2$, we obtain $g_2(x) = (x_1+x_3)^2$, $h_2(x) = (x_1-x_3)^2$.

For simplicity of presentation, we will apply the denotation $S = [-2,1]$ for bounding the variable $x_2 \in \mathbb{R}$, but in the investigation of the linearized problems, we use two inequality constraints $x_2 \le 1$, $x_2 + 2 \ge 0$.

Hence, according to (4.1)–(4.3), we have

$$\begin{array}{rl} H_\sigma(x) = & h_0(x) + \sigma \sum_{j \in \mathcal{E}} [g_j(x) + h_j(x)] \\ = & 2x_2^2 + \sigma[(x_3^2 + x_1^2 + x_2^2) + (x_1+x_3)^2 + (x_1-x_3)^2] \\ = & 2x_2^2 + \sigma[3x_1^2 + 3x_3^2 + x_2^2]; \end{array} \tag{5.20}$$

$$\begin{array}{rl} G_\sigma(x) = & g_0(x) + 2\sigma \sum_{j \in \mathcal{E}} \max\{g_j(x); h_j(x)\} \\ = & x_1^2 + x_3^2 + 2\sigma[\max\{x_3^2; x_1^2 + x_2^2\} + \max\{(x_1+x_3)^2; (x_1-x_3)^2\}]. \end{array} \tag{5.21}$$

Set $\sigma := 1$, $y = (\frac{1}{6}, 1, \frac{7}{6})^\top \notin \mathcal{F}$. Then, we obtain $\nabla H_\sigma(x) = (0, 4x_2, 0)^\top + \sigma(6x_1, 2x_2, 6x_3)^\top = 6(x_1, x_2, x_3)^\top$, and $\nabla H_\sigma(y) = (1, 6, 7)^\top$.

In order to find a suitable point $u \in \mathcal{F}$, we consider the linearized problem

$$(\mathscr{P}_\sigma L(y)): \quad \begin{array}{l} G_\sigma(x) - \langle \nabla H_\sigma(y), x \rangle = x_1^2 + x_3^2 + 2\max\{x_3^2; x_1^2 + x_2^2\} \\ +2\max\{(x_1+x_3)^2; (x_1-x_3)^2\} - \langle (1,6,7)^\top, x \rangle \downarrow \min_x, \\ x \in \mathbb{R}^3, \ -2 \le x_2 \le 1. \end{array} \tag{5.22}$$

As above in (5.2), it is clear that problem (5.22) amounts to the following one

$$\left. \begin{array}{l} x_1^2 + x_3^2 + 2t_1 + 2t_2 - x_1 - 6x_2 - 7x_3 \downarrow \min_{(x,t)}, \\ x_3^2 \le t_1, \quad x_1^2 + x_2^2 \le t_1, \quad t = (t_1, t_2)^\top \in \mathbb{R}^2, \\ (x_1+x_3)^2 \le t_2, \quad (x_1-x_3)^2 \le t_2, \\ x_2 \le 1, \quad x_2 + 2 \ge 0, \quad x \in \mathbb{R}^3. \end{array} \right\} \tag{5.22'}$$

Besides, as above, it can be readily seen that the Slater condition holds in $(5.22')$. Furthermore, the solution vector $(u, t_*) \in \mathbb{R}^5$ to problem $(5.22')$ satisfies the complementarity conditions

$$\left. \begin{array}{c} \eta_1(x_3^2 - t_1) = 0 = \nu_1(x_1^2 + x_2^2 - t_1), \\ \eta_2\left[(x_1 + x_3)^2 - t_2\right] = 0 = \nu_2\left[(x_1 - x_3)^2 - t_2\right], \\ \mu_1(x_2 - 1) = 0 = \mu_2(x_2 + 2). \end{array} \right\} \tag{5.23}$$

In addition, we have, for $t_* = (t_{1*}, t_{2*})^\top$,

$$t_{1*} = \max\{u_3^2; u_1^2 + u_2^2\}; \quad t_{2*} = \max\{(u_1 + u_3)^2; (u_1 - u_3)^2\}. \tag{5.24}$$

Since the Lagrange function for $(5.22')$ has the following form

$$\begin{aligned} \mathscr{L}(x, t; \eta_1, \nu_1, \eta_2, \nu_2, \mu_1, \mu_2) = {} & x_1^2 + x_3^2 + 2t_1 + 2t_2 - x_1 - 6x_2 - 7x_3 \\ & + \eta_1(x_3^2 - t_1) + \nu_1(x_1^2 + x_2^2 - t_1) + \eta_2\left[(x_1 + x_3)^2 - t_2\right] \quad (5.25) \\ & + \nu_2\left[(x_1 - x_3)^2 - t_2\right] + \mu_1(x_2 - 1) - \mu_2(x_2 + 2), \end{aligned}$$

$((\eta_1, \nu_1, \eta_2, \nu_2, \mu_1, \mu_2) \in \mathbb{R}_+^6)$, then the KKT system is written as follows

$$\left. \begin{array}{l} \dfrac{\partial \mathscr{L}(u, t_*)}{\partial t_1} = 2 - \eta_1 - \nu_1 = 0, \quad \text{i.e. } \eta_1 + \nu_1 = 2, \\[2mm] \dfrac{\partial \mathscr{L}(u, t_*)}{\partial t_2} = 2 - \eta_2 - \nu_2 = 0, \quad \text{i.e. } \eta_2 + \nu_2 = 2; \end{array} \right\} \tag{5.26}$$

$$\begin{array}{ll} (a) & \dfrac{\partial \mathscr{L}(u, t_*)}{\partial x_1} = 2u_1 - 1 + 2\nu_1 u_1 + 2\eta_2(u_1 + u_3) + 2\nu_2(u_1 - u_3) = 0, \\[2mm] (b) & \dfrac{\partial \mathscr{L}(u, t_*)}{\partial x_2} = -6 + 2\nu_1 u_2 + \mu_1 - \mu_2 = 0, \\[2mm] (c) & \dfrac{\partial \mathscr{L}(u, t_*)}{\partial x_3} = 2u_3 - 7 + 2\eta_1 u_3 + 2\eta_2(u_1 + u_3) + 2\nu_2(u_3 - u_1) = 0. \end{array} \left. \right\} \tag{5.27}$$

It can be readily seen that the point $u = (0, 1, 1)^\top$ satisfies the KKT conditions (5.23), (5.26), (5.27) with $t_* = (t_{1*}, t_{2*})^\top = (1, 1)^\top$ (see (5.24)). Indeed, the equations (5.27) at $u = (0, 1, 1)^\top$ take the form

$$\begin{array}{lll} (a) & 2\eta_2 - 1 - 2\nu_2 = 0, \quad \text{or} \quad \eta_2 - \nu_2 = \frac{1}{2}, \\ (b) & 2\nu_1 - 6 + \mu_1 - \mu_2 = 0, \\ (c) & 2 - 7 + 2\eta_1 + 2\eta_2 + 2\nu_2 = 0. \end{array} \left. \right\} \tag{5.27$'$}$$

Then, from $(5.27')$ (a) we derive $\eta_2 = \frac{5}{4}$, $\nu_2 = \frac{3}{4}$. Further, from $(5.27')$ (c) with the help of (5.26) it follows that $2\eta_1 = 5 - 2(\eta_2 + \nu_2) = 1$, i.e. $\eta_1 = \frac{1}{2}$, $\nu_1 = \frac{3}{2}$.

On the other hand, thanks to (5.23), we see that $\mu_2 = \mu_2(u) = 0$. Then $(5.27')$ (b) provides that $3 - 6 + \mu_1 = 0$, i.e. $\mu_1 = 3$. Hence, the vector $(u, t_*)^\top$ is really a KKT point in $(5.22')$, and, due to convexity of problem $(5.22')$, u is also a solution to (5.22) $((u, t_*)^\top$ is a solution to $(5.22'))$.

Now let us verify whether the principal inequality (4.6) of Theorem 4.1 holds with (y, β, u) where $\beta = H_\sigma(y) + \zeta$, $\zeta = f_0(z) = 0$. First, compute $H_\sigma(y)$ with $y = \left(\dfrac{1}{6}, 1, \dfrac{7}{6}\right)$, $H_\sigma(y) = 3(y_1^2 + y_2^2 + y_3^2) = 3\left(\dfrac{1}{36} + 1 + \dfrac{49}{36}\right) = 7\dfrac{1}{6}$. Since $\zeta = 0$, $\beta = H_\sigma(y) + \zeta = 7\dfrac{1}{6}$. Furthermore,

$\nabla H(y) = (1, 6, 7)^\top$, $(u - y) = (0, 1, 1)^\top - \left(\dfrac{1}{6}, 1, \dfrac{7}{6}\right)^\top = \left(-\dfrac{1}{6}, 0, -\dfrac{1}{6}\right)^\top$; $\langle \nabla H(y), (u - y)\rangle =$

$-\dfrac{1}{6} - \dfrac{7}{6} = -\dfrac{8}{6} = -1\dfrac{1}{3}$; $\beta + \langle \nabla H(y), (u-y) \rangle = 7\dfrac{1}{6} - \dfrac{4}{3} = 5\dfrac{5}{6}$. On the other hand, it can be readily calculated that $G_\sigma(u) = u_1^2 + u_3^2 + 2\gamma_{*1} + 2\gamma_{*2} = 1 + 4 = 5$.

Therefore, we have $G_\sigma(u) = 5 < 5\dfrac{5}{6} = \beta + \langle \nabla H(y), (u-y) \rangle$.

Hence, the principal inequality (4.6) of Theorem 4.1 is violated. As a consequence, the degenerate KKT point $z = (0,0,0)^\top$ is not a global solution to problem (5.19).

Moreover, by solving the linearized problem $(\mathscr{P}L(y))$ with the infeasible point $y = \left(\dfrac{1}{6}, 1, \dfrac{7}{6} \right)^\top$ (but satisfying the equation (4.5)), we constructed the feasible in (5.19) point $u = (0,1,1)^\top$, which is better than z, since $f_0(u) = -1 < \zeta_0 = f_0(z) = 0$.

Furthermore, it can be readily seen, as above, that the point $u = (0,1,1)^\top$ is also a KKT point in the original problem (5.19), but not a global solution to (5.19). Moreover, we can show this fact by repeating the same procedure of finding another pair (y_1, β_1), such that $H_\sigma(y_1) = \beta_1 - \zeta_1$, where $\zeta_1 := f_0(u) = f_0(z_1) = -1$, $z_1 := u$, and by solving the linearized problem $(\mathscr{P}L_1) := (\mathscr{P}L(y_1))$, which provides the point u_1 such that

$$f_0(u_1) =: \zeta_2 < \zeta_1 = f_0(u) = f_0(z_1).$$

So, the procedure described above provides us a hint on how to construct one of the simplest global search procedures which is able to escape stationary points and local solutions in non-convex Problem $(\mathscr{P})$.

6. Sufficient optimality conditions

Now, we turn to the question of when the conditions $(4.5) - (4.6)$ of Theorem 4.1 become sufficient for a feasible point to be a global solution to nonconvex Problem $(\mathscr{P})$.

Theorem 6.1. *Suppose that, in Problem $(\mathscr{P})$, for a feasible point z, $\zeta := f_0(z)$, the condition $(\mathscr{H})$–(5.14) is fulfilled. In addition, let some penalty parameter $\sigma > 0$ be given. Finally, assume that, for every pair $(y, \beta) \in \mathbb{R}^n \times \mathbb{R}$ satisfying the relation*

$$(a) \quad H_\sigma(y) = \beta - \zeta + \varepsilon, \qquad (b) \quad G_\sigma(y) \le \beta, \qquad (6.1)$$

there is a subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the function $H_\sigma(\cdot)$ at y such that the following inequality holds

$$G_\sigma(x) - \beta \ge \langle H'_\sigma(y), x - y \rangle \quad \forall x \in S. \qquad (6.2)$$

Then, the point $z \in \mathscr{F}$ turns out to be an ε-global solution to Problem $(\mathscr{P}_\sigma)$ as well as to Problem $(\mathscr{P})$.

Remark 6.1. Recall that the third assumption of Theorem 6.1, according to (4.7), can be precised as follows. For every pair (y, β) satisfying (6.1) (a) (b), there exists only one collection of subgradients $h'_i(y) \in \partial h_i(y)$, $i \in \{0\} \cup I \cup \mathscr{E}$, and $g'_j(y) \in \partial g_j(y)$, $j \in \mathscr{E}$, of the corresponding functions $h_i(\cdot)$ and $g_j(\cdot)$ at the point y such that the following inequality holds

$$G_\sigma(x) - \beta \ge \left\langle h'_0(y) + \sigma \left[\sum_{i \in I} h'_i(y) + \sum_{j \in \mathscr{E}} (g'_j(y) + h'_j(y)) \right], x - y \right\rangle \quad \forall x \in S. \qquad (6.2')$$

Proof. (A) Suppose that, despite of the assertion of Theorem 6.1, we found a vector $u \in S$ such that $F_\sigma(u) + \varepsilon < F_\sigma(z)$, regardless of the fact that the conditions (6.1)–(6.2) are satisfied. Further on, the proof coincides with the proof of Theorem 5.1 till the part (C).

(C) Since the pair $(y, \beta) \in \mathbb{R}^{n+1}$ (constructed in the same way as in parts (A) and (B) of the proof of Theorem 5.1) satisfies the conditions (6.1) (a) and (b), and besides, $u \in S$, the inequality (6.2) must hold with u, i.e., $0 \geq \beta - G_\sigma(u) + \langle H'_\sigma(y), u - y \rangle$.

As in the proof on Theorem 5.1, the latter inequality, with the help of the presentation (5.18′), leads us to the absurdity

$$0 \geq \alpha^{-1}(1 - \alpha)[F_\sigma(w) - \zeta + \varepsilon] > 0.$$

Hence, the conjecture made in the beginning of part (A) is incorrect. As a consequence, we obtain

$$F_\sigma(z) \leq F_\sigma(x) + \varepsilon \quad \forall x \in S.$$

It means that the point z is an ε-solution to $(\mathscr{P}_\sigma)$. Finally, since $z \in \mathscr{F}$, we have

$$f_0(z) = F_\sigma(z) \leq F_\sigma(x) + \varepsilon = f_0(x) + \varepsilon \quad \forall x \in \mathscr{F},$$

so that z is also an ε-solution to $(\mathscr{P})$. $\square$

Remark 6.2. Recall (see Remark 6 in [30]) that, in Theorem 6.1, the value of penalty parameter $\sigma > 0$ is not precise, but fixed. Nevertheless, according to the assumptions of Theorem 6.1, the point z is feasible, so that the value $\sigma > 0$ must be sufficiently large to ensure the feasibility of z. On the other hand, the inequality (6.2) is satisfied for only one subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$. This is different to Theorems 4.1 and 5.1, where the corresponding inequalities are fulfilled for all $H'_\sigma(y)$ in the subdifferential $\partial H_\sigma(y)$. Thus, Theorem 6.1 provides better options for constructing numerical methods capable of escaping local pitfalls and stationary points in Problem $(\mathscr{P})$.

Remark 6.3. It is worth noting that, similar to Remarks 3, 5 and 7 in [30], one can conclude that, from the view-point of the optimization theory [1, 2, 5, 6, 7, 8, 12, 13, 14, 15, 16, 17, 19] Theorems 4.1, 5.1, and 6.1 look considerably better and are more relevant than the corresponding results from [29, 30]. This is because these theorems are closer to the original Problem $(\mathscr{P})$ dealing, for instance, with a feasible (in $(\mathscr{P})$) point z that can be a solution to $(\mathscr{P}_\sigma)$, as well as to $(\mathscr{P})$.

Meanwhile, in [29], we investigated the case when the pair $(z, \lambda) \in S \times \mathbb{R}^m_+$ can be a saddle point for $\mathscr{L}(x, \mu) = f_0 + \sum_{i \in I} \lambda_i f_i(x)$ on $S \times \mathbb{R}^m_+$, that entails $z \in Sol(\mathscr{P})$. It often happens that a saddle point does not exist, while an ε-solution always does (see example 5.3 in [29]).

Hence, not all questions have been answered, but, apparently, Theorems 4.1, 5.1, and 6.1 provide a starting point and initiate the development of new algorithms of nonconvex optimization, as it was done, for instance, in [9, 18, 27, 28, 44, 45, 46].

Now let us demonstrate the effectiveness of the GOCs of Theorems 4.1–6.1 using the following example.

Example 6.1. Here we address a well-known combinatorial problem, the so-called Channel Communication Problem (ChCP)$(x \in \mathbb{R}^n)$[19] in its continuous form

$$
\left.
\begin{aligned}
f_0(x) &:= -\sum_{i=1}^{n} c_i x_i^2 \downarrow \min_x, \ x \in \mathbb{R}^n, \\
f_i(x) &= x_i^2 - x_i = 0, \ i \in \mathscr{E} = \{1,\dots,n\}, \\
f_{ij}(x) &:= x_i x_j = 0, \ (i,j) \in \Gamma \subset \Omega.
\end{aligned}
\right\}
\tag{6.3}
$$

Let us consider the case, where $n = 3$, $\Gamma = \{(1,2),(2,3)\}$, $\Omega = \{(1,2),(1,3),(2,3)\}$, $c = (1,3,3)^\top \in \mathbb{R}^3$, and

$$
\left.
\begin{aligned}
f_i(x) &= x_i^2 - x_i = 0, \ i = 1,2,3, \\
f_4(x) &= 2x_1 x_2 = 0, \ f_5(x) = 2x_2 x_3 = 0.
\end{aligned}
\right\}
\tag{6.4}
$$

It is clear, that $f_0(x) = g_0(x) - h_0(x)$, where

$$
g_0(x) \equiv 0, \ h_0(x) = x_1^2 + 3x_2^2 + 3x_3^2.
\tag{6.5}
$$

For $i = 1,2,3$, we use the following d.c. representation

$$
\left.
\begin{aligned}
f_i(x) &\overset{\triangle}{=} x_i^2 - x_i = (2x_i^2 - x_i) - x_i^2, \\
g_i(x) &:= 2x_i^2 - x_i, \ h_i(x) := x_i^2, \ i = 1,2,3.
\end{aligned}
\right\}
\tag{6.6}
$$

In addition, for $i = 4,5$, we employ the d.c. decomposition as follows

$$
\left.
\begin{aligned}
f_4(x) &\overset{\triangle}{=} 2x_1 x_2 = (x_1 + x_2)^2 - (x_1^2 + x_2^2), \\
f_5(x) &\overset{\triangle}{=} 2x_2 x_3 = (x_2 + x_3)^2 - (x_2^2 + x_3^2);
\end{aligned}
\right\}
\tag{6.7}
$$

$$
g_4(x) = (x_1 + x_2)^2, h_4(x) = x_1^2 + x_2^2, g_5(x) = (x_2 + x_3)^2, h_5(x) = x_2^2 + x_3^2.
\tag{6.7$'$}
$$

Furthermore, since $|f_k(x)| = \max\{f_k(x); -f_k(x)\} = |g_k(x) - h_k(x)| = \max\{g_k(x) - h_k(x); h_k(x) - g_k(x)\} = 2\max\{g_k(x); h_k(x)\} - [g_k(x) + h_k(x)], \ k = 1,\dots,5$, it can be readily seen that (6.6) entails

$$
|f_i(x)| = 2\max\{2x_i^2 - x_i; x_i^2\} - (3x_i^2 - x_i), \ i = 1,2,3.
\tag{6.8}
$$

Besides, for $f_4(\cdot)$ and $f_5(\cdot)$, thanks to (6.7), (6.7$'$), we got

$$
\left.
\begin{aligned}
|f_4(x)| &= 2\max\{(x_1 + x_2)^2; x_1^2 + x_2^2\} - [(x_1 + x_2)^2 + x_1^2 + x_2^2], \\
|f_5(x)| &= 2\max\{(x_2 + x_3)^2; x_2^2 + x_3^2\} - [(x_2 + x_3)^2 + x_2^2 + x_3^2].
\end{aligned}
\right\}
\tag{6.9}
$$

Now, let us build the penalized cost function

$$
\begin{aligned}
F_\sigma(x) &:= f_0(x) + \sigma \sum_{i=1}^{5} |f_1(x)| = G_\sigma(x) - H_\sigma(x) \\
&= -\sum_{i=1}^{3} c_i x_i^2 + \sigma \left\{ \sum_{i=1}^{3} \left[2\max\{2x_i^2 - x_i; x_i^2\} - (3x_i^2 - x_i) \right] \right. \\
&\quad + 2\max\{(x_1 + x_2)^2; x_1^2 + x_2^2\} - 2[x_1^2 + x_2^2 + x_1 x_2] \\
&\quad \left. + 2\max\{(x_2 + x_3)^2; x_2^2 + x_3^2\} - 2[x_2^2 + x_3^2 + x_2 x_3] \right\}.
\end{aligned}
\tag{6.10}
$$

According to (6.10), we obtain

$$
\begin{aligned}
H_\sigma(x) &:= \sum_{i=1}^{3} c_i x_i^2 + \sigma \left\{ \sum_{i=1}^{3} (3x_i^2 - x_i) + (x_1 + x_2)^2 + x_1^2 + x_2^2 \right. \\
&\quad \left. + (x_2 + x_3)^2 + x_2^2 + x_3^2 \right\}.
\end{aligned}
\tag{6.11}
$$

$$G_\sigma(x) := 2\sigma \left[\sum_{i=1}^{3} \max\{2x_i^2 - x_i; x_i^2\} + \max\{(x_1 + x_2)^2; x_1^2 + x_2^2\} \right.$$
$$\left. + \max\{(x_2 + x_3)^2; x_2^2 + x_3^2\} \right]. \tag{6.12}$$

It can be readily seen that $H_\sigma(\cdot)$ can be transformed as follows

$$H_\sigma(x) = x_1^2 + 3x_2^2 + 3x_3^2 + \sigma \left[5x_1^2 + 7x_2^2 + 5x_3^2 - x_1 - x_2 - x_3 \right.$$
$$\left. + 2x_1 x_2 + 2x_2 x_3 \right], \tag{6.13}$$

and, besides, the gradient $\nabla H_\sigma(x)$ has the form

$$\nabla H_\sigma(x) = \begin{pmatrix} 2x_1 + 10\sigma x_1 - \sigma + 2\sigma x_2 \\ 6x_2 + 14\sigma x_2 - \sigma + 2\sigma x_1 + 2\sigma x_3 \\ 6x_3 + 10\sigma x_3 - \sigma + 2\sigma x_2 \end{pmatrix}. \tag{6.14}$$

(a) Let us now show that the worst feasible in (6.3) point $z_0 = (0,0,0)^\top$ is, nonetheless, the degenerate KKT point to the ChCP ($n = 3$). Indeed, it is easy to see that z_0 is feasible in ChCP ($n = 3$) and $\zeta_0 := f_0(z_0) = 0$.

Furthermore, since $\nabla f_1(x) = (2x_1 - 1, 0, 0)^\top$, $\nabla f_2(x) = (0, 2x_2 - 1, 0)^\top$, $\nabla f_3(x) = (0, 0, 2x_3 - 1)^\top$, $\nabla f_4(x) = (2x_2, 2x_1, 0)^\top$, $\nabla f_5(x) = (0, 2x_3, 2x_2)^\top$, it is clear that $\nabla f_i(z_0) = -e^i$, $i = 1, 2, 3$, $\nabla f_0(z_0) = (0,0,0)^\top = \nabla f_4(z_0) = \nabla f_5(z_0)$. Therefore, the KKT equation

$$\lambda_0 \nabla f_0(z_0) + \sum_{i=1}^{3} \lambda_i \nabla f_i(z_0) + \lambda_4 \nabla f_4(z_0) + \lambda_5 \nabla f_5(z_0) = 0_3$$

holds at the vector z_0 with $\lambda_i = 0$, $i = 1, 2, 3$ and $\lambda_0 = 1 = \lambda_4 = \lambda_5$.

Passing ahead of the regularity conditions (CQ), let us find out if we can escape this pitfall with the help of Theorem 4.1 and 5.1.

(b) Set $\sigma := 1$ and consider the convex linearized (at $y \in \mathbb{R}^3$) problem

$$(\mathscr{P}L(y)): \qquad G_\sigma(x) - \langle \nabla H_\sigma(y), x \rangle \downarrow \min_x, \ x \in \mathbb{R}^3.$$

On account of (6.12), we have the problem

$$2 \left[\sum_{i=1}^{3} \max\{2x_i^2 - x_i; x_i^2\} + \max\{(x_1 + x_2)^2; x_1^2 + x_2^2\} \right.$$
$$\left. + \max\{(x_2 + x_3)^2; x_2^2 + x_3^2\} \right] - \langle \nabla H_\sigma(y), x \rangle \downarrow \min_x, \ x \in \mathbb{R}^3. \tag{6.15}$$

Thanks to Lemma 4.1 from [29], the problem (6.15) amounts to the following one

$$\left. \begin{array}{c} 2 \sum_{j=1}^{5} t_j - \langle \nabla H_\sigma(y), x \rangle \downarrow \min_{(x,t)}, \ x \in \mathbb{R}^3, \ t \in \mathbb{R}^5, \\ 2x_i^2 - x_i \le t_i, \ x_i^2 \le t_i, \ i = 1, 2, 3, \\ (x_1 + x_2)^2 \le t_4, \ x_1^2 + x_2^2 \le t_4, \\ (x_2 + x_3)^2 \le t_5, \ x_2^2 + x_3^2 \le t_5. \end{array} \right\} \tag{6.15$'$}$$

It is worth noting that in (6.15$'$) we have only the inequality constraints, meanwhile in the original statement (6.3) includes only the equality constraints. Moreover, the problem (6.15$'$) is a convex optimization problem, in which, as shown above in the comments that follow (4.8),

the Slater condition holds and, therefore, the Lagrange function for $(6.15')$ takes the form

$$\mathcal{L}(x,t) = 2\sum_{j=1}^{5} t_j - \langle \nabla H_\sigma(y), x \rangle + \sum_{i=1}^{3} [\eta_i(2x_i^2 - x_i - t_i) + v_i(x_i^2 - t_i)] + \\ + \eta_4[(x_1+x_2)^2 - t_4] + v_4(x_1^2 + x_2^2 - t_4) + \\ + \eta_5[(x_2+x_3)^2 - t_5] + v_5(x_2^2 + x_3^2 - t_5). \tag{6.16}$$

Whence, due to the KKT equations, we derive

$$\frac{\partial \mathcal{L}(x,t)}{\partial t_j} = 2 - \eta_j - v_j = 0, \; i.e. \; \eta_j + v_j = 2, \; \eta_j \geq 0, \; v_j \geq 0, \; j = 1, \ldots, 5. \tag{6.17}$$

Moreover, employing the denotations $\nabla H_\sigma(y) = (A(y), B(y), C(y))^\top$, we have

$$\frac{\partial \mathcal{L}(x,t)}{\partial x_1} = -A(y) + \eta_1(4x_1 - 1) + 2v_1x_1 + 2\eta_4(x_1+x_2) + 2v_4x_1 = 0; \tag{6.18}$$

$$\frac{\partial \mathcal{L}(x,t)}{\partial x_2} = -B(y) + \eta_2(4x_2 - 1) + 2v_2x_2 + 2\eta_4(x_1+x_2) + 2v_4x_2 + \\ + 2\eta_5(x_2+x_3) + 2v_5x_2 = 0; \tag{6.19}$$

$$\frac{\partial \mathcal{L}(x,t)}{\partial x_3} = -C(y) + \eta_3(4x_3 - 1) + 2v_3x_3 + 2\eta_5(x_2+x_3) + 2v_5x_3 = 0. \tag{6.20}$$

(c) Now let us choose $y := (\frac{5}{6}, 0, \frac{5}{8})$, which is infeasible in the ChCP $(n=3)$, and compute $\nabla H_\sigma(y)$. From (6.14) with $\sigma = 1$, it follows

$$\nabla H_\sigma(y) = \begin{pmatrix} 12y_1 + 2y_2 - 1 \\ 2y_1 + 20y_2 + 2y_3 - 1 \\ 2y_2 + 16y_3 - 1 \end{pmatrix} = \left(9, \frac{23}{12}, 9 \right)^\top, \tag{6.21}$$

so that in $(6.18)-(6.20)$ we have $A(y) = C(y) = 9$, $B(y) = \frac{23}{12}$.

Let us show that the vector $u = (1,0,1)^\top$, which is feasible in the ChCP (6.3) $(n=3)$, satisfies the KKT system (6.18)–(6.20). Then, the pair (u,t_*), $t_* = (t_{1*}, \ldots, t_{j*}, \ldots, t_{5*})^\top \in \mathbb{R}^5$ would be a solution to the auxiliary problem $(6.15')$, i.e., the vector u would be a solution to the linearized problem (6.15).

Indeed, from (6.18), it follows $(x = u)$ $9 = \eta_1(4u_1 - 1) + 2v_1u_1 + 2\eta_4(u_1 + u_2) + 2v_4u_1 = 3\eta_1 + 2v_1 + 2(\eta_4 + v_4) = \eta_1 + 2(\eta_1 + v_1) + 2(\eta_4 + v_4)$, whence, thanks to (6.17), we derive $9 = \eta_1 + 8$, $\eta_1 = v_1 = 1$.

Similarly, using (6.20) and (6.17), we get

$$9 = 3\eta_3 + 2v_3 + 2(\eta_5 + v_5), \; i.e. \; 9 = \eta_3 + 8, \; \eta_3 = v_3 = 1.$$

On the other hand, it follows from (6.19) with $x = u$ that

$$\frac{23}{12} = 2(\eta_4 + \eta_5) - \eta_2, \; (\eta_2, v_2, \eta_4, v_4, \eta_5, v_5) \in \mathbb{R}_+^6,$$

which can be satisfied together with (6.17) by a family of the Lagrange multipliers $(\eta, v) \in int\, \mathbb{R}_+^8$. Hence, the vector $u = (1,0,1)^\top$ is really a solution to the convex linearized problem $(\mathscr{P}L(y)) - (6.15)$.

(d) Finally, let us verify the principal inequality (4.6) of the GOCs of Theorem 4.1 at the tuple (y, β, u). To do it, calculate the number $\beta = H_\sigma(y) + \zeta_0$. Using (6.13), we compute

$$H_\sigma(y) = 6y_1^2 + 8y_3^2 - y_1 - y_3 = 6\left(\frac{5}{6}\right)^2 + 8\left(\frac{5}{8}\right)^2 - \frac{5}{6} - \frac{5}{8} = \frac{10}{3} + \frac{5}{2} = 5\frac{5}{6}.$$

Since $\zeta_0 = f_0(z_0) = 0$, we get $\beta = H_\sigma(y) = 5\frac{5}{6}$. Further, with the help of (6.16), we obtain

$$u - y = \left(\frac{1}{6}; 0; \frac{3}{8}\right)^\top, \langle \nabla H_\sigma(y), u - y\rangle = 9\left(\frac{1}{6} + \frac{3}{8}\right) = 1\frac{1}{2} + 3\frac{3}{8} = 4\frac{7}{8}.$$

Thus, $\beta + \langle \nabla H_\sigma(y), u - y\rangle = 5\frac{5}{6} + 4\frac{7}{8} = 10\frac{17}{24}$, meanwhile, by virtue of (6.12), $G_\sigma(u) = 2[2\max\{2-1;1\} + 2\max\{1;1\}] = 8$. Finally, we see that the principal inequality (4.6) of Theorem 4.1 is violated

$$G_\sigma(u) = 8 < 10\frac{17}{24} = \beta + \langle \nabla H_\sigma(y), u - y\rangle.$$

Hence, in virtue of Theorem 4.1, the point $z_0 = (0, 0, 0)^\top$ can not be a solution to ChCP $(n = 3)$. In addition, by solving the convex linearized problem $(\mathscr{P}L(y))$–(6.15), we produced a feasible in ChCP (6.3) $(n = 3)$ point $u = (1, 0, 1)^\top$, which is better than z_0, because $f_0(z_0) = 0 > -4 = f_0(u)$.

We can verify that $u \in Sol(\text{ChCP})$ $(n = 3)$. Hence, we arrive at the conclusion that by applying the GOCs of Theorem 4.1 we escaped the worst feasible vector z_0 and also jumped right to the global solution u. And it was done in the integer combinatorial problem — the Channel Communication Problem.

7. CONCLUSION

In this paper, we continued the investigations started in [4, 10, 11, 17]. We investigated the nonconvex nonsmooth Problem $(\mathscr{P})$ with equality and inequality constraints given by d.c. functions. This problem extends the smooth statements with only inequality constraints considered before [29, 30].

The principal objectives of the paper were the Global Optimality Conditions (GOCs). However, before obtaining them, we managed to use the exact penalty procedure proposed in [36, 38] to perform reduction of the original problem to a d.c. minimization problem without equality and inequality constraints.

Only after that it became possible to develop the GOCs which reduce the nonconvex penalized problem to a family of convex (linearized) problems. It is worth noticing that the linearization was applied to the function which accumulates all the nonconvexities of the original Problem $(\mathscr{P})$.

Moreover, it was shown that the GOCs are connected with the KKT theorem, and, consequently, with the Classical Optimization Theory.

In addition, the GOCs possess the so-called constructive (algorithmic) property which allows one to escape local pitfalls provided, for instance, by the classical [1, 2, 19] and the special [3, 4, 5] Optimization Methods. This property was effectively demonstrated by examples.

We also obtained a rather unexpected result on the Lagrange multipliers provided by the linearized (at the point z in question) problem, which turned out to be the Lagrange multipliers for the point z in the original Problem $(\mathscr{P})$.

To sum up, we developed new mathematical tools that help not only to escape local and stationary pitfalls, but also to reach a global solution in nonconvex optimization problems with equality and inequality constraints defined by continuous functions.

References

[1] J.-F. Bonnans, J.C. Gilbert, C. Lemaréchal, C.A. Sagastizábal, Numerical Optimization: Theoretical and Practical aspects, 2nd edn. Springer-Verlag, Berlin, Heidelberg, 2006.

[2] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 2006.

[3] C.A. Floudas, P.M. Pardalos (eds.), Frontiers in Global Optimization, Kluwer Academic Publishers, Dordrecht, 2004.

[4] R. Horst, H. Tuy, Global Optimization. Deterministic Approaches, Springer-Verlag, Berlin, 1993.

[5] H. Tuy, D.C. Optimization: Theory, Methods and Algorithms, In: Horst, R., Pardalos, P.M. (eds.) Handbook of Global optimization, pp. 149-216, Kluwer Academic Publisher, Dordrecht, 1995.

[6] V.F. Demyanov, Extremum's Conditions and Variational Calculus, High school edition, Moscow, 2005.

[7] V.F. Demyanov, L.V. Vasiliev, Nondifferentiable Optimization, Translation Series in Mathematics and Engineering, Springer-Optimization Software, New York, 1985.

[8] R.T. Rockafellar, R.J.-.B. Wets, Variational Analysis, Springer, New York, 1998.

[9] A.S. Strekalovsky, Elements of Nonconvex Optimization, Nauka, Novosibirsk, 2003.

[10] H. Tuy, Concave programming under linear constraints, Dokl. Akad. Nauk SSSR 159 (1964), 32-35.

[11] S.I. Zukhovitskii, R.A. Polyak, M.E. Primak, A problem of nonconvex programming, Dokl. Akad. Nauk SSSR 179 (1968), 786-789.

[12] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.

[13] R.T. Rockafellar, Lagrange multipliers and optimality, SIAM Rev. 35 (1993), 183-238.

[14] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms, Springer-Verlag, Berlin 1993.

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

[16] J.-B. Hiriart-Urruty, Optimisation et Analyse Convex, Presses Universitaires de France, Paris, 1998.

[17] J.-B. Hiriart-Urruty, Generalized differentiability, duality and optimization for problems dealing with difference of convex functions, In: Ponstein, J. (ed.) Convexity and Duality in Optimization, Lect. Notes Econ. Math. Syst. 256 (1985), 37-69.

[18] A.S. Strekalovsky, I.M. Minarchenko, A local search method for optimization problem with d.c. inequality constraints, Appl. Math. Model. 58 (2018), 229-244.

[19] A. Ben-Tal, A.S. Nemirovski, Lectures on modern convex optimization: analysis, algorithms, and engineering applications, MOS-SIAM Series on Optimization, Philadelphia, 2001. https://doi.org/10.1137/1.9780898718829

[20] R. Byrd, G. Lopez-Calva, J. Nocedal, A line search exact penalty method using steering rules, Math. Program. Ser. A 133 (2012), 39-73.

[21] I. Griva, R.A. Polyak, Primal-dual nonlinear rescaling method with dynamic scaling parameter update, Math. Program., Ser. A 106 (2006), 237-259.

[22] L.A. Kirievskii, R.A. Polyak, On the acceleration of convergence of convex programming methods, Dokl. Akad. Nauk SSSR 209 (1973), 1024-1027.

[23] B.T. Polyak, Introduction to optimization, Optimization Software, Publications Division, New York, 1987.

[24] R.A. Polyak, Methods of control sequences in extremal problems in the presence of constraints, U.S.S.R. Comput. Math. Math. Phys. 18 (1978), 14-31.

[25] S.I. Zukhovitskii, R.A. Polyak, M.E. Primak, Two methods of finding equilibrium points of concave n-person games, Dokl. Akad. Nauk SSSR 185 (1969), 24-27.

[26] S.I. Zukhovitskii, R.A. Polyak, M.E. Primak, The concave n-person game and a certain production model, Dokl. Akad. Nauk SSSR 191 (1970), 1220-1223.

[27] A.S. Strekalovsky, On solving optimization problems with hidden nonconvex structures, In: Rassias, T.M., Floudas, C.A., Butenko, S. (eds.) Optimization in Science and Engineering, 465-502, Springer, New York, 2014.

[28] A.S. Strekalovsky, Local search for nonsmooth DC optimization with DC equality and inequality, In: Bagirov, A.M., Gaudioso, M., Karmitsa, N., Makela, M.M., Taheri, S. (Eds.) Constraints Numerical Nonsmooth Optimization: State of the Art Algorithms, pp. 229-261, Springer, 2020,

[29] A.S. Strekalovsky, Global optimality conditions in nonconvex optimization, J. Optim. Theory Appl. 173 (2017), 770-792.

[30] A.S. Strekalovsky, Global optimality conditions and exact penalization, Optim. Lett. 13 (2019), 597-615.

[31] A.S. Strekalovsky, New global optimality conditions in a problem with d.c. constraints, Trudy Inst. Mat. i Mekh. UrO RAN 25 (2019), 245-261.

[32] S. Han, O. Mangasarian, Exact penalty functions in nonlinear programming, Math. Program. 17 (1979), 251-269.

[33] J. Burke, An exact penalization viewpoint of constrained optimization, SIAM J. Control Optim. 29 (1991), 968-998.

[34] G. Di Pillo, S. Lucidi, F. Rinaldi, An approach to constrained global optimization based on exact penalty functions, J. Global Optim. 54 (2012), 251-260.

[35] G. Di Pillo, S. Lucidi, F. Rinaldi, A derivative-free algorithm for constrained global optimization based on exact penalty functions, J. Optim. Theory Appl. 164 (2015), 862-882.

[36] I. Eremin, The penalty method in convex programming, Soviet Math. Dokl. 8 (1966), 459-462.

[37] A.V. Fiacco, G.P. McCormick, Nonlinear programming: sequential unconstrained minimization techniques, SIAM, 1990.

[38] W. Zangwill, Non-linear programming via penalty functions, Manage. Sci. 13 (1967), 344-358.

[39] W. Zangwill, Nonlinear programming: a unified approach, Prentice-Hall, 1969.

[40] A.J. Zaslavski, Optimization on Metric and Normed Spaces, Springer Optimization and Its Applications Book 44, 2010.

[41] A.J. Zaslavski, Exact penalty property in optimization with mixed constraints via variational analysis, SIAM J. Optim. 23 (2013), 170-187.

[42] A. Kruger, L. Minchenko, J. Outrata, On relaxing the Mangasarian-Fromovitz constraint qualiffcation, Positivity 18 (2014), 171-189.

[43] A. Kruger, Error bounds and metric subregularity, Optimization 64 (2015), 49-79.

[44] A.S. Strekalovsky, Global and local search methods for d.c. constrained problems, Lecture Notes in Comput. Sci. 12095 (2020), 3-18.

[45] A.S. Strekalovsky, On a global search in d.c. optimization problems, Commun. Comput. Inf. Sci. 1145 (2020), 222-236.

[46] A.S. Strekalovsky, A.V. Orlov, Global search for bilevel optimization with quadratic data, In: Dempe, S., Zemkoho, A. (Eds.) Bilevel optimization: advances and next challenges, Springer Optimization and Its Applications, 161, 2020.