Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 20, No. 1, pp. 75–98, 2024

UNCONSTRAINED NONSMOOTH NONCONVEX OPTIMIZATION BASED ON CN FUNCTION*

Min Jiang, Zhiqing Meng†, Chuangyin Dang and Rui Shen

Abstract: Finding a global optimal solution to an unconstrained nonsmooth nonconvex optimization is not an easy job. To tackle the problem, we introduce in this paper the concept of CN function. Examples are given to show that some nonsmooth or nonconvex functions are actually CN functions. Furthermore, operations such as addition, subtraction, multiplication or division on CN functions still lead to CN functions. Sufficient conditions of optimal solution to unconstrained optimization based on CN function are presented, which are equivalent to Karush-Kuhn-Tucker (KKT) condition. Lagrange function and proper Lagrange function based on CN function are defined respectively, whose dual problems and strong duality properties show that the global optimal objective function value of the dual problem is equal to the global optimal objective function value of the CN optimization. Augmented Lagrangian penalty function and proper augmented Lagrangian penalty function of a CN function are also introduced respectively. With these augmented functions, we devise an augmented Lagrangian penalty function method to find the optimal solution to the CN optimization. Overall, this paper provides a new approach to solving unconstrained optimization problems. We obtain some new results on nonconvex nonsmooth unconstrained optimization problems without using subdifferential.

Key words: *nonsmooth nonconvex function, CN function, CN optimization, Lagrange dual, augmented Lagrangian penalty function*

Mathematics Subject Classification: *90C30*

1 Introduction

In recent years, nonsmooth or nonconvex optimization problems are becoming prominent in machine learning, artificial intelligence and other fields [30, 33, 44, 45]. Nevertheless, it is difficult to find an optimal solution to an unconstrained optimization problem $\min_{\boldsymbol{x}\in R^n} f(\boldsymbol{x})$ when $f(\boldsymbol{x})$ is a nonsmooth or nonconvex function. For a long time, there are few criterions to determine whether a solution is a global optimal solution to a nonsmooth or nonconvex optimization if the subdifferential theories of nonsmooth optimization are not used. In the past few decades, for the optimization problems with some special nonsmooth or generalized convex functions, a global optimal criterions, dualities and algorithms may be obtained by exploiting some subdifferential techniques in [3, 11, 37, 41]. So, to solve these problems, theoretical tools of nonsmooth and nonconvex functions are needed, such as the subdifferentiable, general convex, smoothing and so on [3, 9, 11, 16, 37]. A CN function is defined

*This research is funded by the National Natural Science Foundation of China under Grant No.11871434.
†Corresponding author

 DOI: https://doi.org/10.61208/pjo-2023-020

in Definition 2.1 in this paper, where the CN function is a nonconvex nonsmooth function form that can be transformed into a convex smooth function with convex equality constraints. The CN function somewhat relates to upper -UC^k function [5, 12, 19, 34, 37, 41] and factorable nonconvex function [6, 17, 25, 26, 31, 32, 43, 38].

The lower(upper)-C^k function was suggested by Professor Rockafellar [37]. The class of lower-C^1 functions was first introduced by Spingarn in [41]. In his work, Spingarn showed that these functions are (Mifflin) semi-smooth and Clarke regular, and are characterized by a generalized monotonicity property of their subgradients, called submonotonicity. The definition of lower(upper)-C^k function is given in [12]. The lower(upper)-C^k function is nonconvex or nondifferentiable, but it is locally Lipschitz approximate convex function in [12]. Research on lower(upper)-C^k function is done on subdifferentiation and optimization in [18, 19, 20, 21]. The Moreau envelopes erf is lower-C^2 [5, 19, 34] such that subdifferential of the lower(upper)-C^k functions can solve nonconvex optimization by prox-regularity and the proximal mapping(operator) [20]. Chieu et al. proved second-order necessary and sufficient conditions for lower-C^2 functions to be convex and strongly convex in [9].

Some methods for nonsmooth nonconvex optimization problems with lower(upper)-C^k functions have been studied in [13, 22, 23, 36]. Dao developed a nonconvex bundle method based on the downshift mechanism and a proximity control management technique to solve nonconvex nonsmooth constrained optimization problems, where he proved its global convergence in the sense of subsequences for both classes of lower-C^1 and upper-C^1 in [13]. Hare et al. studied two proximal bundle methods for nonsmooth nonconvex optimization in [22, 23] by proximal mapping on lower-C^2 functions. Noll defined a first-order model of f as an extended case of lower-C^k function and presented a bundle method in [36]. Clearly, if f is a first-order model, f is not necessarily lower-C^k, and the reverse is not necessarily true.

On the other hand, the branch-and-bound method in conjunction with underestimating convex problems had been proved to be an effective method to solve global nonconvex optimization problems [1, 4, 42]. Almost all the methods used to solve nonconvex optimization problems are to construct many convex relaxation subproblems with convex envelopes and convex underestimating, as in [4, 40, 38, 43]. Based on this idea, the factorable programming technique, one of the most popular approaches for constructing convex relaxations of nonconvex optimization problems including problems with convex-transformable functions, was given in [32]. Due to its simplicity, factorable programming technique is included in most global optimization packages such as Baron(1996), Antigone(2014), etc [35]. But, Nohra and Sahinidis(2018) pointed out in [35] that a main drawback of factorable programming technique is that it often results in large relaxation gaps.

In 1976, McCormick(1976) [32] first defined factorable nonconvex function, but factorable nonconvex function is not necessarily lower-C^1, such as $f(\boldsymbol{x}) = ||\boldsymbol{x}||^{0.1} + ||\boldsymbol{x}+1||_0$ on $\boldsymbol{x} \in R^n$, because $f(\boldsymbol{x}) = ||\boldsymbol{x}||^{0.1}+||\boldsymbol{x}+1||_0$ is not locally Lipschitz [7]. In fact, the factorable nonconvex functions in [26, 31, 32, 43] may be special CN functions (see Definition 2.1). In recent years, research on nonconvex factorable programming further shows its effectiveness in solving the global optimization, as shown in [6, 17, 25, 38]. There are many CN functions that are not upper-C^k functions or factorable functions, such as $|x|_0$ because upper-C^k functions are continuous as shown in [18]. So, a CN function is not necessarily an upper-C^k function or a factorable nonconvex function.

In summary, all studies of nonsmooth nonconvex optimization problems show that the subdifferential techniques are applied to the global optimal criterions, dualities and algorithms. Up to now, there is no published literature that gives the optimality conditions and dualities of nonsmooth nonconvex optimization by differential technique. However, in this paper, the optimality conditions and dualities of nonsmooth or nonconvex CN optimization problems are studied by differential technique.

In fact, some CN functions have special structure, such as a weak uniform CN function[27] and a strong CN function [28]. If a weak uniform CN function is decomposable, the decomposable algorithm can effectively reduce the scale in solving the unconstrained optimization problem with decomposable CN function which are [27]. In machine learning, there are many decomposable optimization problems in nonsmooth and nonconvex [44]. We have studied a vector feature extraction method based on SCN optimization,where the proposed is applied to high-dimensional vector sparse feature extraction experiments and compressed sensing reconstruction problems and is compared with other algorithms. It is proved that a better accuracy and sparsity is achieved and the important sparse features in the dataset are effectively selected in [46]. Furthermore, we propose a feature block decomposition algorithm(GDL-SVM) of sparse support vector machine based on SCN function. In the numerical experiment GDL-SVM is compared with other classical algorithms on six real datasets, which proves that GDL-SVM ensures similar classification accuracy, and is significantly better in sparse classification results than those of other algorithms [47].

The main contributions of this paper include: 1) a concept of CN function; 2) sufficient conditions for a global optimal solution to an unconstrained CN optimization problem; 3) the strong duality properties of unconstrained CN optimization; and 4) an augmented Lagrangian penalty function algorithm for a global optimal solution to CN optimization. The great value of this paper is that by CN function optimization existing differential theory is used to solve the non convex and non smooth optimization problems, avoiding the use of complex subdifferential theory. This makes it possible to design a fast algorithm with second order convergence.

The remainder of this paper is organized as follows. In Section 2, the definition, some examples and some properties of CN function are given. In Section 3, the formulation of unconstrained optimization with a CN function is given and the sufficient conditions of its global (local) optimal solution is proved. In Section 4, the (proper) Lagrange function of a CN function and its dual problems of unconstrained optimization are defined with their dual properties being discussed, and an augmented Lagrangian penalty function algorithm is proposed and its convergence is proved. In Section 5, the conclusion is given.

2 Definition and Properties

In this section, a new function is defined - called (exact) CN function. Examples of (exact) CN functions are given, where the exact CN function or CN function may be nonconvex or nonsmooth. Some properties of (exact) CN function are proved.

Definition 2.1. Let a nonsmooth or nonconvex function $f : R^n \to R^1$ be given. If there exist $r + 1$ differentiable convex functions, $g : R^n \times R^m \to R^1$ and $g_i : R^n \times R^m \to R^1$, $i = 1, 2, \ldots, r \geq 1$, such that for each $\boldsymbol{x} \in R^n$

$$f(\boldsymbol{x}) = \min_{\boldsymbol{y}}\{g(\boldsymbol{x}, \boldsymbol{y})|\forall(\boldsymbol{x}, \boldsymbol{y}) \in X(f)\} \tag{2.1}$$

holds, f is called a convertible nonsmooth or nonconvex function (abbreviated to CN function), where

$$X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \mid g_i(\boldsymbol{x}, \boldsymbol{y}) = 0,\ i = 1, 2, ..., r\}. \tag{2.2}$$

There is a special case of CN function: if for any $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$ one has $f(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{y})$, then f is called an exact CN function.

$[g : g_1, g_2, \ldots, g_r]$ is called a CN form of f, which is written as $f = [g : g_1, g_2, \ldots, g_r]$.

An exact CN function is a CN function. For example, for $x \in R$, the nonconvex function $f(x) = (x^2 - x)^2$ is an exact CN function, where a CN form of f is $f = [(y-x)^2 : x^2 - y]$ on $(x, y) \in R^2$.

By Definition 2.1, if f is a weak uniform CN function [27], f is a CN function. But, if f is a CN function, f may not necessarily be a weak uniform CN function. And if f is a strong CN function [28], f is a CN function. But, if f is a CN function, f may not necessarily be a strong CN function. Hence, the CN function is broader than weak uniform CN function and strong CN function.

Next, examples are given to show that some nonconvex functions are CN functions.

Example 2.2. Since the nonconvex function $f(\boldsymbol{x}) = \sum_{i=1}^{n} \sqrt{|x_i|}$ has a CN form function

$$f = [g(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} y_i : y_i^4 - y_{i+n}, x_i^2 - y_{i+n}, y_{i+2n}^2 - y_i, i = 1, 2, \dots, n],$$

it is a CN function, where $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^{3n}$ and

$$X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) | y_i^4 - y_{i+n} = 0, x_i^2 - y_{i+n} = 0, y_{i+2n}^2 - y_i = 0, i = 1, 2, \dots, n\}.$$

Since $y_i = y_{i+2n}^2 \geq 0$ and $y_i = \sqrt{|x_i|}(i = 1, 2, \dots, n)$, f is an exact CN function.

Example 2.3. The nonconvex function $f(\boldsymbol{x}) = (\sqrt{|\boldsymbol{a}_1^\top \boldsymbol{x}|} + b_1)^{\sqrt[3]{|\boldsymbol{a}_2^\top \boldsymbol{x}|} + b_2}$ is CN for $b_1, b_2 > 0$ and $\boldsymbol{a}_1, \boldsymbol{a}_2 \in R^n \backslash \{\boldsymbol{0}\}$, since it has a CN form

$$\begin{aligned} f(\boldsymbol{x}) = [y_{10} + 1 \quad : \quad & (\boldsymbol{a}_1^\top \boldsymbol{x})^2 - y_1, y_2^4 - y_1, y_3^2 - y_2, -\ln(y_2 + b_1) + y_8, \\ & (\boldsymbol{a}_2^\top \boldsymbol{x})^2 - y_4, y_5^6 - y_4, y_6^2 - y_5, -\text{In}(y_{10} + 1) + y_7, \\ & 0.5(y_5 + b_2 + y_8)^2 - 0.5y_9 - y_7, (y_5 + b_2)^2 + y_8^2 - y_9, y_{11}^2 - y_{10}], \end{aligned}$$

where $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^{11}$ and

$$\begin{aligned} X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \mid & (\boldsymbol{a}_1^\top \boldsymbol{x})^2 - y_1 = 0, y_2^4 - y_1 = 0, y_3^2 - y_2 = 0, -\text{In}(y_2 + b_1) + y_8 = 0, \\ & (\boldsymbol{a}_2^\top \boldsymbol{x})^2 - y_4 = 0, y_5^6 - y_4 = 0, y_6^2 - y_5 = 0, -\text{In}(y_{10} + 1) + y_7 = 0, \\ & 0.5(y_5 + b_2 + y_8)^2 - 0.5y_9 - y_7 = 0, (y_5 + b_2)^2 + y_8^2 - y_9 = 0, y_{11}^2 - y_{10} = 0\}. \end{aligned}$$

$X(f)$ is equivalent to

$$\begin{aligned} X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \quad | \quad & (\boldsymbol{a}_1^\top \boldsymbol{x})^2 = y_2^4, y_8 = \ln(y_2 + b_1), y_2 = y_3^2 \geq 0 \\ & (\boldsymbol{a}_2^\top \boldsymbol{x})^2 = y_5^6, y_7 = \ln(y_{10} + 1), y_5 = y_6^2 \geq 0, \\ & (y_5 + b_2)y_8 = y_7, y_{10} = y_{11}^2 \geq 0\}, \end{aligned}$$

i.e.,

$$\begin{aligned} X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \quad | \quad & y_8 = \ln(\sqrt{|\boldsymbol{a}_1^\top \boldsymbol{x}|} + b_1), y_5 = \sqrt[3]{|\boldsymbol{a}_2^\top \boldsymbol{x}|}, \\ & (y_5 + b_2)y_8 = \ln(y_{10} + 1), y_{10} = y_{11}^2 \geq 0\}, \end{aligned}$$

So, we have $y_{10} = (\sqrt{|\boldsymbol{a}_1^\top \boldsymbol{x}|} + b_1)^{y_5 + b_2} - 1$. Hence, $f(\boldsymbol{x})$ is an exact CN function.

Example 2.4. Let nonconvex function $f(\boldsymbol{x}) = \sqrt{||\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}| - b|}$ be given on $\boldsymbol{x} \in R^n$, where $\boldsymbol{a} \in R^n$ and $b \in R^1$. We have

$$f(\boldsymbol{x}) = [g(\boldsymbol{x}, y_1, y_2, y_3, y_4, y_5, y_6) = y_5 : (y_3 - b)^2 - y_4, y_5^4 - y_4, y_6^2 - y_5, y_3^2 - y_2, (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})^2 - y_2, y_1^2 - y_3],$$

where $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^6$ and

$$X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) | (y_3 - b)^2 = y_4, y_5^4 = y_4, y_6^2 = y_5, y_3^2 = y_2, (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})^2 = y_2, y_1^2 = y_3\}.$$

Since $y_5 = |y_3 - b|^{0.5}$ and $y_3 = |\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}|$, $f(\boldsymbol{x})$ is an exact CN function.

The following examples are given to show that some nondifferentiable or nonconvex functions are CN functions.

Example 2.5. For $(\boldsymbol{x}, \boldsymbol{y}) \in R^2 \times R^3$, the nondifferentiable nonconvex function

$$f(x_1, x_2) = (x_1^2 - x_2)^2 + |x_1^2 - x_2|_0 = \begin{cases} 0 & \text{if } x_1^2 - x_2 = 0, \\ (x_1^2 - x_2)^2 + 1 & \text{if } x_1^2 - x_2 \neq 0, \end{cases}$$

is a CN function since $f = [g : g_1, g_2, g_3] = [y_1^2 + y_2 : x_1^2 - x_2 - y_1, (y_1 - y_2 + 1)^2 - y_3, y_1^2 + (y_2 - 1)^2 - y_3, y_2^2 - y_2]$, where $|t|_0 = 1$ for $t \neq 0$ or $|t|_0 = 0$ for $t = 0$ and

$$\begin{aligned} X(f) = \{(x_1, x_2, y_1, y_2, y_3) \quad | \quad & x_1^2 - x_2 - y_1 = 0, (y_1 - y_2 + 1)^2 - y_3 = 0, \\ & y_1^2 + (y_2 - 1)^2 - y_3 = 0, y_2^2 - y_2 = 0\}. \end{aligned}$$

In fact, $X(f)$ is equivalent to

$$X(f) = \{(x_1, x_2, y_1, y_2, y_3) | x_1^2 - x_2 - y_1 = 0, y_1(y_2 - 1) = 0, y_2 \in \{0, 1\}\}.$$

We can easily verify that (1) is true. If $x_1^2 - x_2 = 0$ for given $(x_1, x_2) \in R^2$, then $y_1 = 0$, and $y_2, y_3 \in \{0, 1\}$. So, we have

$$f(x_1, x_2) = \min_{(y_1, y_2, y_3)} \{g(x_1, x_2, y_1, y_2, y_3) = y_1^2 + y_2 | y_1 = 0, y_2, y_3 \in \{0, 1\}\} = 0.$$

If $x_1^2 - x_2 \neq 0$ for given $(x_1, x_2) \in R^2$, then $x_1^2 - x_2 = y_1$, $y_2 = 1$ and $y_3 = y_1^2$. So, we have $f(x_1, x_2) = g(x_1, x_2, y_1, y_2, y_3) = y_1^2 + 1$.

Hence, f is not an exact CN function.

Example 2.6. Let a discontinuous function $f(\boldsymbol{x}) = \sqrt{\|A\boldsymbol{x} - \boldsymbol{b}\|} + \|\boldsymbol{x}\|_0$ be on $\boldsymbol{x} \in R^n$, where matrix $A \in R^{m \times n}$ and $\boldsymbol{b} \in R^m$ are given. $\|\boldsymbol{x}\|_0$ is 0-norm function in [7]. There are $g(\boldsymbol{x}, \boldsymbol{y})$ and $g_i(\boldsymbol{x}, \boldsymbol{y})$ with $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^{2n+3}$ such that

$$\begin{aligned} f(\boldsymbol{x}) = [y_{n+2} + \sum_{i=1}^{n} y_i \quad : \quad & \|A\boldsymbol{x} - \boldsymbol{b}\|^2 - y_{2n+1}, y_{2n+2}^4 - y_{2n+1}, y_{2n+3}^2 - y_{2n+2}, y_i^2 - y_i, \\ & (x_i + y_i - 1)^2 - y_{i+n}, x_i^2 + (y_i - 1)^2 - y_{i+n}, i = 1, 2, \ldots, n]. \end{aligned}$$

So, $f(\boldsymbol{x})$ is a CN funtion, but not exact.

Example 2.7. Let a nonconvex and nondifferentiable function:

$$f(\boldsymbol{x}) = 2\phi_1(\boldsymbol{x})\phi_2(\boldsymbol{x}) + |\phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x})|_0 = \begin{cases} 2\phi_1(\boldsymbol{x})\phi_2(\boldsymbol{x}) & \text{if } \phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x}) = 0, \\ 2\phi_1(\boldsymbol{x})\phi_2(\boldsymbol{x}) + 1 & \text{if } \phi_1(\boldsymbol{x}) - \phi_2(\boldsymbol{x}) \neq 0, \end{cases}$$

where $\phi_1, \phi_2 : R^n \to R$ are differentiable convex functions. Its CN form is defined by

$$f = [g(\boldsymbol{x}, \boldsymbol{y}) = (y_1 + y_2)^2 - y_3 + y_6 : \phi_1(\boldsymbol{x}) - y_1, \phi_2(\boldsymbol{x}) - y_2, y_1^2 + y_2^2 - y_3, y_6^2 - y_6,$$
$$(y_1 - y_2 + y_6 - 1)^2 - y_5, (y_1 - y_2)^2 + (y_6 - 1)^2 - y_5],$$

where $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^6$. So, $f(x_1)$ is a CN function.

For CN functions, we have the following properties.

Proposition 2.8. *If $f_1(\boldsymbol{x}), f_2(\boldsymbol{x}) : R^n \to R$ are exact CN functions, then the following conclusions hold,*

(i) $\alpha f_1(\boldsymbol{x})$ *is an exact CN function, where* $\alpha \neq 0$.

(ii) $\alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x})$ *is an exact CN function, where* $\alpha_1, \alpha_2 \neq 0$.

(iii) $\phi(f_1(\boldsymbol{x}))$ *is an exact CN function, where* $\phi : R \to R$ *is a monotone increasing convex function.*

(iv) $f_1(\boldsymbol{x}) f_2(\boldsymbol{x})$ *is an exact CN function.*

(v) $\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ *is an exact CN function, if* $f_2(\boldsymbol{x}) \neq 0$ *on* $\boldsymbol{x} \in R^n$.

(vi) $\min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ *is an exact CN function.*

(vii) $\log(f_1(\boldsymbol{x}))$ *is an exact CN function, if* $f_1(\boldsymbol{x}) > 0$ *on* $\boldsymbol{x} \in R^n$.

Proof. Since $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are exact CN functions, their CN forms are given respectively by

$$f_1(\boldsymbol{x}) = [g^1(\boldsymbol{x}, \boldsymbol{y}^1) : g_i^1(\boldsymbol{x}, \boldsymbol{y}^1),\ i = 1, 2, ..., r_1], \tag{2.3}$$
$$f_2(\boldsymbol{x}) = [g^2(\boldsymbol{x}, \boldsymbol{y}^2) : g_j^2(\boldsymbol{x}, \boldsymbol{y}^2),\ j = 1, 2, \ldots, r_2], \tag{2.4}$$

where $\boldsymbol{y}^1 \in R^{m_1}$,$\boldsymbol{y}^2 \in R^{m_2}$, $g^1(\boldsymbol{x}, \boldsymbol{y}^1)$ and $g_i^1(\boldsymbol{x}, \boldsymbol{y}^1)$ are differentiable convex functions on $(\boldsymbol{x}, \boldsymbol{y}^1)$,$i = 1, 2, ..., r_1$; $g^2(\boldsymbol{x}, \boldsymbol{y}^2)$ and $g_j^2(\boldsymbol{x}, \boldsymbol{y}^2)$ are differentiable convex functions on $(\boldsymbol{x}, \boldsymbol{y}^2)$,$j = 1, 2, \ldots, r_2$. Let $z_1 = g^1(\boldsymbol{x}, \boldsymbol{y}^1)$, $z_2 = g^2(\boldsymbol{x}, \boldsymbol{y}^2)$.

(i) By Definition 2.1 and (2.3), an exact CN form of $\alpha f_1(\boldsymbol{x})$ is defined by

$$\alpha f_1(\boldsymbol{x}) = [\alpha z_1 : g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), \qquad i = 1, 2, ..., r_1],$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, z_1) \in R^n \times R^{m_1} \times R^1$ is variable.

(ii) By (2.3) and (2.4), an exact CN form of $\alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x})$ is defined by

$$\alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x}) = [\alpha_1 z_1 + \alpha_2 z_2 \quad : \quad g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g^2(\boldsymbol{x}, \boldsymbol{y}^2) - z_2,$$
$$g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), i = 1, 2, ..., r_1, g_j^2(\boldsymbol{x}, \boldsymbol{y}^2), j = 1, 2, \ldots, r_2].$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2, \boldsymbol{z}) \in R^n \times R^{m_1} \times R^{m_2} \times R^2$ is variable.

(iii) By (2.3), an exact CN form of $\phi(f_1(\boldsymbol{x}))$ is defined by

$$\phi(f_1(\boldsymbol{x})) = [\phi(g^1(\boldsymbol{x}, \boldsymbol{y}^1)) :\ g_i^1(\boldsymbol{x}, \boldsymbol{y}^1),\ i = 1, 2, ..., r].$$

(iv) By (2.3) and (2.4), an exact CN form of $f_1(\boldsymbol{x}) f_2(\boldsymbol{x})$ is defined by

$$f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) = [0.5(z_1 + z_2)^2 - 0.5 z_3 \quad : \quad z_1^2 + z_2^2 - z_3, g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g^2(\boldsymbol{x}, \boldsymbol{y}^2) - z_2,$$
$$g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), i = 1, 2, ..., r_1, g_j^2(\boldsymbol{x}, \boldsymbol{y}^2), j = 1, 2, \ldots, r_2].$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2, \boldsymbol{z}) \in R^n \times R^{m_1} \times R^{m_2} \times R^3$ is variable.

(v) By (2.3) and (2.4), an exact CN form of $\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ is defined by

$$\begin{aligned}\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} = [z_3 \quad : \quad & (z_3+z_2)^2 - z_4 - 2z_1, z_2^2 + z_3^2 - z_4, g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g^2(\boldsymbol{x}, \boldsymbol{y}^2) - z_2, \\ & g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), i = 1, 2, ..., r_1, g_j^2(\boldsymbol{x}, \boldsymbol{y}^2), j = 1, 2, \ldots, r_2],\end{aligned}$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2, \boldsymbol{z}) \in R^n \times R^{m_1} \times R^{m_2} \times R^4$ is variable.

(vi) By (2.3) and (2.4), an exact CN form of $\min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is defined by

$$\begin{aligned}\min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} = [z_1 \quad : \quad & (z_2+z_3)^2 - z_4, z_2^2 + z_3^2 - z_4, z_5^2 - z_2, z_6^2 - z_3, \\ & g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1 + z_2, g^2(\boldsymbol{x}, \boldsymbol{y}^2) - z_1 + z_3, \\ & g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), i = 1, 2, ..., r_1, g_j^2(\boldsymbol{x}, \boldsymbol{y}^2), j = 1, 2, \ldots, r_2],\end{aligned}$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{y}^2, \boldsymbol{z}) \in R^n \times R^{m_1} \times R^{m_2} \times R^6$ is variable.

(vii) By Definition 2.1 and (2.3), an exact CN form of $\log(f_1(\boldsymbol{x}))$ is defined by

$$\log(f_1(\boldsymbol{x})) = [z_2 : -\log z_1 + z_2, g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), \qquad i = 1, 2, ..., r_1],$$

where $(\boldsymbol{x}, \boldsymbol{y}^1, \boldsymbol{z}) \in R^n \times R^{m_1} \times R^2$ is variable. □

Proposition 2.9. *If $f_1(\boldsymbol{x}), f_2(\boldsymbol{x}) : R^n \to R$ are CN functions, then the following conclusions hold,*

(i) $\alpha f_1(\boldsymbol{x})$ *is a CN function, where* $\alpha > 0$.

(ii) $\alpha_1 f_1(\boldsymbol{x}) + \alpha_2 f_2(\boldsymbol{x})$ *is a CN function, where* $\alpha_1, \alpha_2 > 0$.

(iii) $\phi(f_1(\boldsymbol{x}))$ *is a CN function, where* $\phi : R \to R$ *is a monotone increasing convex function.*

(iv) $f_1(\boldsymbol{x}) f_2(\boldsymbol{x})$ *is a CN function if* $f_1(\boldsymbol{x}) \geq 0$ *and* $f_2(\boldsymbol{x}) \geq 0$ *hold on* $\forall \boldsymbol{x} \in R^n$.

(v) $\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ *is a CN function, if* $f_2(\boldsymbol{x}) > 0$ *on* $\boldsymbol{x} \in R^n$ *is an exact CN function.*

Proof. The proof of (i,ii,iii) is completely similar to (i), (ii) and (iii) of Proposition 2.8.

Since $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are CN functions, their CN forms are given respectively by

$$f_1(\boldsymbol{x}) = [g^1(\boldsymbol{x}, \boldsymbol{y}^1) : g_i^1(\boldsymbol{x}, \boldsymbol{y}^1),\ i = 1, 2, ..., r_1], \tag{2.5}$$

$$f_2(\boldsymbol{x}) = [g^2(\boldsymbol{x}, \boldsymbol{y}^2) : g_j^2(\boldsymbol{x}, \boldsymbol{y}^2),\ j = 1, 2, \ldots, r_2], \tag{2.6}$$

where $\boldsymbol{y}^1 \in R^{m_1}$,$\boldsymbol{y}^2 \in R^{m_2}$, $g^1(\boldsymbol{x}, \boldsymbol{y}^1)$ and $g_i^1(\boldsymbol{x}, \boldsymbol{y}^1)$ are differentiable convex functions on $(\boldsymbol{x}, \boldsymbol{y}^1)$,$i = 1, 2, ..., r_1$; $g^2(\boldsymbol{x}, \boldsymbol{y}^2)$ and $g_j^2(\boldsymbol{x}, \boldsymbol{y}^2)$ are differentiable convex functions on $(\boldsymbol{x}, \boldsymbol{y}^2)$,$j = 1, 2, \ldots, r_2$. Let $z_1 = g^1(\boldsymbol{x}, \boldsymbol{y}^1)$, $z_2 = g^2(\boldsymbol{x}, \boldsymbol{y}^2)$.

(iv) By (2.5) and (2.6), an CN form of $f_1(\boldsymbol{x}) f_2(\boldsymbol{x})$ is defined by

$$\begin{aligned}f_1(\boldsymbol{x}) f_2(\boldsymbol{x}) = [0.5(z_1+z_2)^2 - 0.5z_3 \quad : \quad & z_1^2 + z_2^2 - z_3, g^1(\boldsymbol{x}, \boldsymbol{y}^1) - z_1, g^2(\boldsymbol{x}, \boldsymbol{y}^2) - z_2, \\ & g_i^1(\boldsymbol{x}, \boldsymbol{y}^1), i = 1, 2, ..., r_1, g_j^2(\boldsymbol{x}, \boldsymbol{y}^2), j = 1, 2, \ldots, r_2].\end{aligned}$$

where $(\boldsymbol{x},\boldsymbol{y}^1,\boldsymbol{y}^2,\boldsymbol{z})\in R^n\times R^{m_1}\times R^{m_2}\times R^3$ is variable with $\boldsymbol{z}=(z_1,z_2,z_3)$. For fixed $\boldsymbol{x}$, since

$$\begin{aligned}\min_{(\boldsymbol{y}^1,\boldsymbol{y}^2,\boldsymbol{z})}\{0.5(z_1+z_2)^2-0.5z_3 \quad | \quad & z_1^2+z_2^2-z_3=0, g^1(\boldsymbol{x},\boldsymbol{y}^1)-z_1=0,\\ & g^2(\boldsymbol{x},\boldsymbol{y}^2)-z_2=0, g_i^1(\boldsymbol{x},\boldsymbol{y}^1)=0, i=1,2,...,r_1,\\ & g_j^2(\boldsymbol{x},\boldsymbol{y}^2)=0, j=1,2,\ldots,r_2\}\\ =\min_{(\boldsymbol{y}^1,\boldsymbol{y}^2)}\{g^1(\boldsymbol{x},\boldsymbol{y}^1)g^2((\boldsymbol{x},\boldsymbol{y}^2) \quad | \quad & g_i^1(\boldsymbol{x},\boldsymbol{y}^1)=0, i=1,2,...,r_1, g_j^2(\boldsymbol{x},\boldsymbol{y}^2)=0, j=1,2,\ldots,r_2\}\\ =\min_{\boldsymbol{y}^1}\{g^1(\boldsymbol{x},\boldsymbol{y}^1)|g_i^1(\boldsymbol{x},\boldsymbol{y}^1)=0 \quad , \quad & i=1,2,...,r_1\}\min_{\boldsymbol{y}^2}\{g^2((\boldsymbol{x},\boldsymbol{y}^2)|g_j^2(\boldsymbol{x},\boldsymbol{y}^2)=0, j=1,2,\ldots,r_2\}\\ =f_1(\boldsymbol{x})f_2(\boldsymbol{x}),\end{aligned}$$

$f_1(\boldsymbol{x})f_2(\boldsymbol{x})$ is a CN function.

(v) By (2.3) and (2.4), an CN form of $\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ is defined by

$$\begin{aligned}\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}=[z_3 \quad : \quad & (z_3+z_2)^2-z_4-2z_1, z_2^2+z_3^2-z_4, g^1(\boldsymbol{x},\boldsymbol{y}^1)-z_1, g^2(\boldsymbol{x},\boldsymbol{y}^2)-z_2,\\ & g_i^1(\boldsymbol{x},\boldsymbol{y}^1), i=1,2,...,r_1, g_j^2(\boldsymbol{x},\boldsymbol{y}^2), j=1,2,\ldots,r_2],\end{aligned}$$

where $(\boldsymbol{x},\boldsymbol{y}^1,\boldsymbol{y}^2,\boldsymbol{z})\in R^n\times R^{m_1}\times R^{m_2}\times R^4$ is variable with $\boldsymbol{z}=(z_1,z_2,z_3,z_4)$. For fixed $\boldsymbol{x}$, since

$$\begin{aligned}&\min_{(\boldsymbol{y}^1,\boldsymbol{y}^2,\boldsymbol{z})}\{[z_3|(z_3+z_2)^2-z_4-2z_1=0, z_2^2+z_3^2-z_4=0, g^1(\boldsymbol{x},\boldsymbol{y}^1)-z_1=0, g^2(\boldsymbol{x},\boldsymbol{y}^2)-z_2=0,\\ &\qquad g_i^1(\boldsymbol{x},\boldsymbol{y}^1)=0, i=1,2,...,r_1, g_j^2(\boldsymbol{x},\boldsymbol{y}^2)=0, j=1,2,\ldots,r_2\}\\ &\qquad =\min_{(\boldsymbol{y}^1)}\{\frac{g^1(\boldsymbol{x},\boldsymbol{y}^1)}{g^2(\boldsymbol{x},\boldsymbol{y}^2)}|g_i^1(\boldsymbol{x},\boldsymbol{y}^1)=0, i=1,2,...,r_1\}=\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})},\end{aligned}$$

$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ is a CN function. □

The multi-convex function is a class of very important nonconvex functions [39](2017). Many types of multi-convex functions are CN funtions. For examples, the multi-convex function $f(\boldsymbol{x})=x_1x_2\ldots x_n$ is a CN function by Proposition 2.8. The DC-function is a CN function. Therefore, CN functions cover a wide range of nonconvex functions.

The above examples show that some nonsmooth, nonconvex or discontinuous functions are CN funtions. So, such nonsmooth and nonconvex optimization problems in machine learning can be converted to differentiable CN optimization problems. For example, the sparse optimization problem, $\min_{\boldsymbol{x}\in R^n} h(\boldsymbol{x})+\lambda\|\boldsymbol{x}\|_0$, where $\|\boldsymbol{x}\|_0$ is 0-norm and $h(\boldsymbol{x})$ is convex. Hence, it is worthwhile to study CN optimization theory, as shown in Section 3.

3 Unconstrained Optimization Problem of a CN function

Throughout this section, it is assumed that f is a CN function with $f=[g:g_1,g_2,\ldots,g_r]$. So, $f(\boldsymbol{x})$ is not necessarily differentiable on $\boldsymbol{x}\in R^n$. This paper is concerned with an unconstrained optimization problem with a CN function:

$$\text{(CNO)} \qquad \begin{aligned}\min \quad & f(\boldsymbol{x})\\ \text{s.t.} \quad & \boldsymbol{x}\in R^n,\end{aligned}$$

where $f : R^n \to R$ is a CN function. (CNO) is called a CN optimization problem. An application of a CN form of f brings us the following constrained optimization problem:

$$\text{(CNP)} \qquad \begin{array}{ll} \min & g(\boldsymbol{x}, \boldsymbol{y}) \\ \text{s.t.} & g_i(\boldsymbol{x}, \boldsymbol{y}) = 0, i = 1, 2, ..., r, \\ & \boldsymbol{x} \in R^n, \boldsymbol{y} \in R^m. \end{array}$$

By Definition 2.1, it is clear

$$\min_{\boldsymbol{x} \in R^n} f(\boldsymbol{x}) = \min_{(\boldsymbol{x},\boldsymbol{y}) \in X(f)} g(\boldsymbol{x}, \boldsymbol{y}).$$

Example 3.1. Consider an unconstrained optimization problem

$$\text{(EX3.1)} \quad \min \quad f(x_1) = |x_1 - 1|^{0.5} + |x_1^2 - x_1|_0 + x_1^2 \ \ s.t. \ \ x_1 \in R,$$

where $f(x_1)$ is a CN function. A CN form of f is defined by

$$\begin{array}{rl} f(x_1) = [g(\boldsymbol{x}, \boldsymbol{y}) = y_1 + y_6 + x_1^2 & : \ y_1^4 - y_2, (x_1 - 1)^2 - y_2, y_3^2 - y_1, x_1^2 - x_1 - y_4, \\ & (y_4 + y_6 - 1)^2 - y_5, y_4^2 + (y_6 - 1)^2 - y_5, y_6^2 - y_6], \end{array}$$

where $\boldsymbol{x} = x_1, \boldsymbol{y} = (y_1, y_2, y_3, y_4, y_5, y_6)^\top \in R^6$ and

$$\begin{array}{rl} X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \ | & y_1^4 - y_2 = 0, (x_1 - 1)^2 - y_2 = 0, y_3^2 - y_1 = 0, x_1^2 - x_1 - y_4 = 0, \\ & (y_4 + y_6 - 1)^2 - y_5 = 0, y_4^2 + (y_6 - 1)^2 - y_5 = 0, y_6^2 - y_6 = 0\}. \end{array}$$

$X(f)$ is equivalent to

$$X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) | y_1 = |x_1 - 1|^{0.5}, y_1 = y_3^2 \geq 0, (x_1^2 - x_1)(y_6 - 1) = 0, y_6 \in \{0, 1\}\}.$$

If $x_1 = 0$, then $y_1 = 1$ and $y_6 \in \{0, 1\}$. So, $f(x_1) = 1 = \min_{\boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y}) = y_1 + y_6 + x_1^2$.
If $x_1 = 1$, then $y_1 = 0$ and $y_6 \in \{0, 1\}$. So, $f(x_1) = 1 = \min_{\boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y}) = y_1 + y_6 + x_1^2$.
If $x_1^2 - x_1 \neq 0$, then $y_1 = |x_1 - 1|^{0.5}$ and $y_6 = 1$. So,

$$f(x_1) = |x_1 - 1|^{0.5} + 1 + x_1^2 = \min_{\boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y}) = y_1 + y_6 + x_1^2.$$

Hence, (EX3.1) is equivalent to

$$\begin{array}{lll} \text{(MEX3.1)} & \min & g(\boldsymbol{x}, \boldsymbol{y}) = y_1 + y_6 + x_1^2 \\ & s.t. & (\boldsymbol{x}, \boldsymbol{y}) \in X(f) = \{(\boldsymbol{x}, \boldsymbol{y}) | y_1^4 - y_2 = 0, (x_1 - 1)^2 - y_2 = 0, x_1^2 - x_1 - y_4 = 0, \\ & & y_3^2 - y_1 = 0, (y_4 + y_6 - 1)^2 - y_5 = 0, y_4^2 + (y_6 - 1)^2 - y_5 = 0, y_6^2 - y_6 = 0\}, \end{array}$$

where $\boldsymbol{x} = x_1, \boldsymbol{y} = (y_1, y_2, y_3, y_4, y_5, y_6)^\top \in R^6$. $x_1^* = 0$ and $x_1^* = 1$ are global optimal solutions to (EX3.1) with $f(x_1^*) = 1$. $f(x_1)$ is not subdifferentiable at $x_1^* = 0$ and $x_1^* = 1$. $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (0, 1, 1, 1, 0, 1, 0)$ and $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0, 1, 0)$ are global optimal solutions to (MEX3.1) with $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = 1$.

Example 3.1 shows that the optimal condition $\nabla f(\boldsymbol{x}^*) = 0$ of (CNO) cannot hold true if $\boldsymbol{x}^*$ is an optimal solution to (CNO) when $f(\boldsymbol{x})$ is not subdifferentiable at $\boldsymbol{x}^*$. Hence, it is valuable to study the optimal condition of an optimal solution to (CNP) as follows.

For a fixed $(\boldsymbol{x},\boldsymbol{y})$, a linear programming problem (CNP)$(\boldsymbol{x},\boldsymbol{y})$ is defined as follows, where $\boldsymbol{d}=(\boldsymbol{d}_1,\boldsymbol{d}_2)\in R^n\times R^m$ is variable.

$$\begin{array}{lll}(\text{CNP})(\boldsymbol{x},\boldsymbol{y}) & \min & \nabla g(\boldsymbol{x},\boldsymbol{y})^{\mathrm{T}}\boldsymbol{d}\\ & \text{s.t.} & \nabla g_i(\boldsymbol{x},\boldsymbol{y})^{\mathrm{T}}\boldsymbol{d}\le 0, i=1,2,...,r,\\ & & \boldsymbol{d}\in R^n\times R^m.\end{array}$$

Let $I=\{1,2,\dots,r\}$ and

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})=(g_1(\boldsymbol{x},\boldsymbol{y}),g_2(\boldsymbol{x},\boldsymbol{y}),\dots,g_r(\boldsymbol{x},\boldsymbol{y}))^{\mathrm{T}}.$$

When $f(\boldsymbol{x})$ is not differentiable on $\boldsymbol{x}\in R^n$, it is not easy to determine whether $\boldsymbol{x}$ is a local optimal solution to (CNO). So, it is very difficult to judge the global optimality to (CNP). So the following conclusions about how to judge the global optimality to (CNP) and (CNO) become important.

Theorem 3.2. *Let $(\boldsymbol{x}^*,\boldsymbol{y}^*)\in X(f)$. If there is an optimal solution $\boldsymbol{d}^*$ to (CNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^*\ge 0$, then $(\boldsymbol{x}^*,\boldsymbol{y}^*)$ is an optimal solution to (CNP), $\boldsymbol{x}^*$ is an optimal solution to (CNO) and there is $\boldsymbol{u}^*=(u_1^*,u_2^*,\dots,u_r^*)^{\mathrm{T}}\ge 0$ such that*

$$\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)+\nabla\boldsymbol{g}(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{u}^*=0. \tag{3.1}$$

Proof. For any $(\boldsymbol{x},\boldsymbol{y})\in X(f)$, we have

$$g(\boldsymbol{x},\boldsymbol{y})-g(\boldsymbol{x}^*,\boldsymbol{y}^*)\ge\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x},\boldsymbol{y})-(\boldsymbol{x}^*,\boldsymbol{y}^*)],$$
$$0=g_i(\boldsymbol{x},\boldsymbol{y})-g_i(\boldsymbol{x}^*,\boldsymbol{y}^*)\ge\nabla g_i(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x},\boldsymbol{y})-(\boldsymbol{x}^*,\boldsymbol{y}^*)],\qquad i=1,2,...,r.$$

So, $(\boldsymbol{x},\boldsymbol{y})-(\boldsymbol{x}^*,\boldsymbol{y}^*)$ is a feasible solution to (CNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$. Then

$$\begin{aligned}f(\boldsymbol{x})-f(\boldsymbol{x}^*) &= g(\boldsymbol{x},\boldsymbol{y})-g(\boldsymbol{x}^*,\boldsymbol{y}^*)\\ &\ge \nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x},\boldsymbol{y})-(\boldsymbol{x}^*,\boldsymbol{y}^*)]\\ &\ge \nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^*\ge 0.\end{aligned}$$

Hence, $(\boldsymbol{x}^*,\boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{x}^*$ is an optimal solution to (CNO). Because (CNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ is a linear programming problem, (CNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ can be rewritten as

$$\begin{array}{lll}(\text{CNP})(\boldsymbol{x}^*,\boldsymbol{y}^*) & \max & -\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}\\ & \text{s.t.} & \nabla g_i(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}\le 0, i=1,2,...,r,\\ & & \boldsymbol{d}\in R^n\times R^m.\end{array}$$

The dual problem (DCNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ of (CNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ is defined as follows:

$$\begin{array}{lll}(\text{DCNP})(\boldsymbol{x}^*,\boldsymbol{y}^*) & \min & \displaystyle\sum_{i=1}^{r}0\cdot u_i\\ & \text{s.t.} & \displaystyle -\sum_{i=1}^{r}u_i\nabla g_i(\boldsymbol{x}^*,\boldsymbol{y}^*)=\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)\\ & & u_i\ge 0, i=1,2,...,r;\end{array}$$

where $(u_1,u_2,\dots,u_r)$ is dual variable. By the strong duality theorem of linear programming, there is an optimal solution $\boldsymbol{u}^*=(u_1^*,u_2^*,\dots,u_r^*)^{\mathrm{T}}\ge 0$ to (DCNP)$(\boldsymbol{x}^*,\boldsymbol{y}^*)$ such that

$$0\le-\nabla g(\boldsymbol{x}^*,\boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^*\le 0.$$

Hence, (3.1) is true. $\square$

Corollary 3.3. *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. *Consider the problem,*

$$\begin{array}{lll} \text{(CNPP)}(\boldsymbol{x}^*, \boldsymbol{y}^*) & \min & \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] \\ & s.t. & \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] \le 0, i = 1, 2, ..., r, \\ & & (\boldsymbol{x}, \boldsymbol{y}) \in X(f). \end{array}$$

If $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNPP)*$(\boldsymbol{x}^*, \boldsymbol{y}^*)$, *then* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNP) and* $\boldsymbol{x}^*$ *is an optimal solution to (CNO), i.e. if* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is not an optimal solution to (CNP), then there is an* $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$ *such that* $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] < 0$.

Corollary 3.3 is very useful though (CNPP)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is not a linear programming problem. Theorem 3.4 tells us that $\boldsymbol{x}^*$ is an optimal solution to (CNO) if the optimal condition (3.1) holds for $\boldsymbol{u}^* \ge 0$.

By Theorem 3.2, the following conclusions are clear.

Theorem 3.4. *Let* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. *If there is* $\boldsymbol{u}^* = (u_1^*, u_2^*, \dots, u_r^*)^{\mathrm{T}} \ge 0$ *such that (3.1) holds, then* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNP) and* $\boldsymbol{x}^*$ *is an optimal solution to (CNO).*

Proof. Let any $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$. Since $g(\boldsymbol{x}, \boldsymbol{y})$ and $g_i(\boldsymbol{x}, \boldsymbol{y})(i = 1, 2, ..., r)$ are convex, we have

$$g(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{x}^*, \boldsymbol{y}^*) \ge \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)],$$
$$g_i(\boldsymbol{x}, \boldsymbol{y}) - g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) \ge \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)], i = 1, 2, ..., r.$$

From the above inequalities, we have

$$\begin{aligned} g(\boldsymbol{x}, \boldsymbol{y}) + \sum_{i=1}^{r} u_i^* g_i(\boldsymbol{x}, \boldsymbol{y}) \quad \ge \quad & g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)] \\ & + \sum_{i=1}^{r} u_i^* \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)]. \end{aligned}$$

Hence,

$$g(\boldsymbol{x}, \boldsymbol{y}) \ge g(\boldsymbol{x}, \boldsymbol{y}) + \sum_{i=1}^{r} u_i^* g_i(\boldsymbol{x}, \boldsymbol{y}) \ge g(\boldsymbol{x}^*, \boldsymbol{y}^*),$$

and $\boldsymbol{x}^*$ is an optimal solution to (CNO). □

Theorem 3.5. *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. *If* $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0$, *then* $\boldsymbol{x}^*$ *is an optimal solution to (CNO).*

Proof. Let any $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$. We have

$$g(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{x}^*, \boldsymbol{y}^*) \ge \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] = 0.$$

Hence, $\boldsymbol{x}^*$ is an optimal solution to (CNO). □

Example 3.6. Consider the optimization problem:

$$\begin{array}{lll} \text{(EX3.2)} & \min & f(x) = |x|^{\frac{1}{3}} \\ & s.t. & x \in R^1, \end{array}$$

where $f(x) = |x|^{\frac{1}{3}}$ is a nonsmooth and nonconvex function. The CN optimization of (EX3.2) is defined by

$$\text{(MEX3.2)} \qquad \begin{array}{ll} \min & g(x, y_1, y_2, y_3) = y_3 \\ \text{s.t.} & g_1(x, y_1, y_2, y_3) = y_3^6 - y_2 = 0, \\ & g_2(x, y_1, y_2, y_3) = x^2 - y_2 = 0, \\ & g_3(x, y_1, y_2, y_3) = y_1^2 - y_3 = 0, \\ & x, y_1, y_2, y_3 \in R^1. \end{array}$$

So, (3.1) at $(x, y_1, y_2, y_3) = (0, 0, 0, 0)$ holds by

$$\nabla g(0,0,0,0) + u_1 \nabla g_1(0,0,0,0) + u_2 \nabla g_2(0,0,0,0) + u_3 \nabla g_3(0,0,0,0) = 0,$$

where $(u_1, u_2, u_3) = (0, 0, 1)$.

Example 3.6 shows that the optimal condition in Theorem 3.2 holds. If the optimal condition in Theorem 3.2 or Theorem 3.4 does not hold, the global optimal solution to (CNP) is judged by solving a linear programming problem (CNP0)$(\boldsymbol{x}, \boldsymbol{y})$. For a fixed $(\boldsymbol{x}, \boldsymbol{y})$, the linear programming problem (CNP0)$(\boldsymbol{x}, \boldsymbol{y})$ of (CNP) is defined by

$$\text{(CNP0)}(\boldsymbol{x}, \boldsymbol{y}) \qquad \begin{array}{ll} \min & \nabla g(\boldsymbol{x}, \boldsymbol{y})^{\mathrm{T}} \boldsymbol{d} \\ \text{s.t.} & \nabla g_i(\boldsymbol{x}, \boldsymbol{y})^{\mathrm{T}} \boldsymbol{d} = 0, i = 1, 2, ..., r, \\ & \boldsymbol{d} \in R^n \times R^m. \end{array}$$

Let two feasible sets (or tangent cone) at a fixed $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ be defined respectively by

$$T(\boldsymbol{x}^*, \boldsymbol{y}^*) = \{(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m \quad | \quad \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] \le 0, i = 1, 2, ..., r\}$$

and

$$T_0(\boldsymbol{x}^*, \boldsymbol{y}^*) = \{(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m \quad | \quad \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] = 0, i = 1, 2, ..., r\}.$$

It is clear that $T_0(\boldsymbol{x}^*, \boldsymbol{y}^*) \subset T(\boldsymbol{x}^*, \boldsymbol{y}^*)$ and $X(f) \subset T(\boldsymbol{x}^*, \boldsymbol{y}^*)$ for each $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. Define

$$K_g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \{(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m \quad | \quad \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}[(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] < 0\}.$$

We have the following Lemmas.

Lemma 3.7. *Let* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. *If*

$$X(f) \cap K_g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \emptyset \tag{3.2}$$

holds, then $g(\boldsymbol{x}^*, \boldsymbol{y}^*) \le g(\boldsymbol{x}, \boldsymbol{y})$ *for all* $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$, *i.e.* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNP) and* $\boldsymbol{x}^*$ *is an optimal solution to (CNO).*

Lemma 3.8. *Let* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. *If*

$$X(f) \cap K_g(\boldsymbol{x}^*, \boldsymbol{y}^*) \cap T(\boldsymbol{x}^*, \boldsymbol{y}^*) \backslash T_0(\boldsymbol{x}^*, \boldsymbol{y}^*) = \emptyset \tag{3.3}$$

holds, then $g(\boldsymbol{x}^*, \boldsymbol{y}^*) \le g(\boldsymbol{x}, \boldsymbol{y})$ *for all* $(\boldsymbol{x}, \boldsymbol{y}) \in X(f) \cap T(\boldsymbol{x}^*, \boldsymbol{y}^*) \backslash T_0(\boldsymbol{x}^*, \boldsymbol{y}^*)$.

For $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$, let

$$X(f) \subset T_0(\boldsymbol{x}^*, \boldsymbol{y}^*). \tag{3.4}$$

It is clear that the condition (3.3) holds when the condition (3.4) is true. The condition (3.3) means if for $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ there is no $(\boldsymbol{x}, \boldsymbol{y}) \in X(f) \cap T(\boldsymbol{x}^*, \boldsymbol{y}^*) \backslash T_0(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $g(\boldsymbol{x}, \boldsymbol{y}) < g(\boldsymbol{x}^*, \boldsymbol{y}^*)$ holds. In other words, if the optimal condition (3.3) does not holds, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is not an optimal solution to (CNP).

The following conclusion is easily proved.

Theorem 3.9. *Suppose the condition (3.3) or (3.4) holds for $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. If there is an optimal solution $\boldsymbol{d}^*$ to (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d}^* \geq 0$, then $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP), $\boldsymbol{x}^*$ is an optimal solution to (CNO) and there is $\boldsymbol{u}^* = (u_1^*, u_2^*, \ldots, u_r^*)^{\mathrm{T}} \in R^r$ such that*

$$\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{u}^* = 0. \tag{3.5}$$

Proof. For any $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$, if (3.4) holds, $(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a feasible solution to (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$. So, we have

$$g(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} [(\boldsymbol{x}, \boldsymbol{y}) - (\boldsymbol{x}^*, \boldsymbol{y}^*)] \geq \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d}^* \geq 0.$$

If (3.3) holds, for $(\boldsymbol{x}, \boldsymbol{y}) \in T(\boldsymbol{x}^*, \boldsymbol{y}^*) \backslash T_0(\boldsymbol{x}^*, \boldsymbol{y}^*)$, we have $g(\boldsymbol{x}, \boldsymbol{y}) \geq g(\boldsymbol{x}^*, \boldsymbol{y}^*)$ by Lemma 3.8. For $(\boldsymbol{x}, \boldsymbol{y}) \in T_0(\boldsymbol{x}^*, \boldsymbol{y}^*)$, we have $g(\boldsymbol{x}, \boldsymbol{y}) \geq g(\boldsymbol{x}^*, \boldsymbol{y}^*)$ too. Hence, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{x}^*$ is an optimal solution to (CNO). Because (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is linear programming, (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ can be rewritten as

$$\begin{array}{lll} (\text{CNP0})(\boldsymbol{x}^*, \boldsymbol{y}^*) \qquad & \max & -\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d} \\ & \text{s.t.} & \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d} = 0, i = 1, 2, \ldots, r, \\ & & \boldsymbol{d} \in R^n \times R^m. \end{array}$$

Then, the dual problem (DCNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ of (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is defined as follows.

$$\begin{array}{lll} (\text{DCNP0})(\boldsymbol{x}^*, \boldsymbol{y}^*) \qquad & \min & \displaystyle\sum_{i=1}^{r} 0 \cdot u_i \\ & \text{s.t.} & \displaystyle -\sum_{i=1}^{r} u_i \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) = \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) \\ & & u_i \in R^1, i = 1, 2, \ldots, r, \end{array}$$

where $(u_1, u_2, \ldots, u_r)$ is dual variable. By the strong duality theorem of linear programming, there is an optimal solution $u^* = (u_1^*, u_2^*, \ldots, u_r^*)^{\mathrm{T}} \in R^r$ to (DCNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that

$$0 \leq -\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d}^* \leq 0.$$

Hence, (3.5) is true. □

By Theorem 3.9, we have the following conclusion.

Theorem 3.10. *Suppose that (3.3) or (3.4) holds for $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. If there is $\boldsymbol{u}^* = (u_1^*, u_2^*, \ldots, u_r^*)^{\mathrm{T}}$ such that (3.5) holds, then $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{x}^*$ is an optimal solution to (CNO).*

Proof. Because (3.5) holds for $\boldsymbol{u}^* = (u_1^*, u_2^*, \ldots, u_r^*)^{\mathrm{T}}$, $\boldsymbol{u}^*$ is a feasible solution to linear programming (DCNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$. It is clear that $\boldsymbol{d} = 0$ is a feasible solution to linear programming (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$. By the strong duality theorem, there is an optimal solution $\boldsymbol{d}^*$ to (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^* \geq 0$ and (3.5) holds. Hence, by Theorem 3.9, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{x}^*$ is an optimal solution to (CNO). □

When the conditions (3.3) and (3.4) do not hold, Theorem 3.9 and Theorem 3.10 can be combined into the following conclusion.

Theorem 3.11. *Let $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$. Then there is an optimal solution $\boldsymbol{d}^*$ to (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^* \geq 0$ if and only if there is $\boldsymbol{u}^* = (u_1^*, u_2^*, \ldots, u_r^*)^{\mathrm{T}}$ such that (3.5) holds.*

Theorem 3.11 means that $\boldsymbol{x}^*$ doesn't have to be an optimal solution to (CNO) if there is an optimal solution $\boldsymbol{d}^*$ to (CNP0)$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}}\boldsymbol{d}^* \geq 0$ for $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$.

Example 3.12. Consider the optimization problem:

$$\begin{array}{lll} \text{(EX3.3)} & \min & f(x) = \sqrt{|x^2 - x|} \\ & s.t. & x \in R^1, \end{array}$$

where $f(x) = \sqrt{|x^2 - x|}$ is a nonconvex function. $x^* = 1$ and $x^* = 0$ are optimal solutions to (EX3.3). The CN optimization of (EX3.3) is defined by

$$\begin{array}{lll} \text{(MEX3.3)} & \min & g(x, \boldsymbol{y}) = y_1 \\ & s.t. & g_1(x, \boldsymbol{y}) = y_1^4 - y_2 = 0, g_2(x, \boldsymbol{y}) = y_3^2 - y_2 = 0, \\ & & g_3(x, \boldsymbol{y}) = x^2 - x - y_3 = 0, g_4(x, \boldsymbol{y}) = y_4^2 - y_1 = 0, \end{array}$$

where $(x^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0)$, $(x^*, \boldsymbol{y}^*) = (0, 0, 0, 0, 0)$ and $(x^*, \boldsymbol{y}^*) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{16}, -\frac{1}{4}, \sqrt{\frac{1}{2}})$ are its CN points. When $(x^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0)$ and $(u_1^*, u_2^*, u_3^*, u_4^*) = (0, 0, 0, 1)$, the optimal condition (3.1) holds. When $(x^*, \boldsymbol{y}^*) = (0, 0, 0, 0, 0)$ and $(u_1^*, u_2^*, u_3^*, u_4^*) = (0, 0, 0, 1)$, the optimal condition (3.1) holds. When $(x^*, \boldsymbol{y}^*) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{16}, -\frac{1}{4}, \sqrt{\frac{1}{2}})$ and $(u_1^*, u_2^*, u_3^*, u_4^*) = (-\frac{1}{2}, \frac{1}{2}, \frac{1}{16}, 0)$, the optimal condition (3.5) holds. By Theorem 3.4, $(x^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0)$ and $(x^*, \boldsymbol{y}^*) = (0, 0, 0, 0, 0)$ are optimal solutions to (MEX3.3). But, $(x^*, \boldsymbol{y}^*) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{16}, -\frac{1}{4}, \sqrt{\frac{1}{2}})$ is not an local optimal solution to (MEX3.3). In fact, a stationary point $x = \frac{1}{2}$ of (EX3.3) is not an local optimal solution to (EX3.3).

Example 3.13. Consider the optimization problem (EX3.1) in Example 3.1. $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (0, 1, 1, 1, 0, 1, 0)$ and $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0, 1, 0)$ are global optimal solutions to (MEX3.1). When $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (0, 1, 1, 1, 0, 1, 0)$ and $(u_1^*, u_2^*, \ldots, u_7^*) = (-\frac{1}{4}, \frac{1}{4}, 0, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{4}, 1)$, the optimal condition (3.5) holds. When $(\boldsymbol{x}^*, \boldsymbol{y}^*) = (1, 0, 0, 0, 0, 1, 0)$ and $(u_1^*, u_2^*, \ldots, u_7^*) = (1, -1, 1, -2, 1, -1, 1)$, the optimal condition (3.5) holds.

Example 3.13 shows that the optimal condition (3.5) of CN optimization may be obtained for an optimal solution to (CNP) when f is not subdifferentiable. The optimal condition (3.5) holds at $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$, if $\boldsymbol{x}^*$ is an optimal solution to (CNO).

Next, the relationship between the problem (CNP) and the local optimal solution to (CNO) is discussed.

Theorem 3.14. *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *is a local optimal solution to (CNP). Then*

(i) $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d} \geq 0$ *for* $\boldsymbol{d} \in T_0(\boldsymbol{x}^*, \boldsymbol{y}^*)$.

(ii) *If* $\nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)(i = 1, 2, ..., r)$ *is linearly independent, then there are* $u_1^*, u_2^*, \ldots, u_r^*$ *such that*

$$\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \sum_{i=1}^{r} u_i^* \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0.$$

Furthermore, $\boldsymbol{x}^*$ *is a global optimal solution to (CNO) if* $u_1^*, u_2^*, \ldots, u_r^* \geq 0$.

Theorem 3.14 means that there is no feasible direction $\boldsymbol{d}$ of $X(f)$ at $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d} < 0$ if $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a local optimal solution to (CNP). In other words, if there is a feasible direction $\boldsymbol{d}$ of $X(f)$ at $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ such that $\nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{d} < 0$, then $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is not a local optimal solution to (CNP).

Theorem 3.15. *Suppose that* f *is an exact CN function and continuous on* R^n*. If* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *is a local optimal solution to (CNP), then* $\boldsymbol{x}^*$ *is a local optimal solution to (CNO).*

Proof. Suppose that $\boldsymbol{x}^*$ is not a local optimal solution to (CNO). Let a monotonically decreasing sequence $\varepsilon_k \to 0$ as $k \to +\infty$. For each neighborhood $B(\boldsymbol{x}^*, \varepsilon_k)$ of $\boldsymbol{x}^*$,$k = 1, 2, \ldots,$ there is a $\boldsymbol{x}_k$ such that

$$f(\boldsymbol{x}_k) < f(\boldsymbol{x}^*).$$

It is clear that $\boldsymbol{x}_k \to \boldsymbol{x}^*$ as $k \to +\infty$. There is $\boldsymbol{y}_k \in R^m$ such that $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in X(f)$ and $\boldsymbol{y}_k \to \boldsymbol{y}^*$ because the function f is continuous. Because $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ is a local optimal solution to (CNP), there is $k' > 0$ such that

$$f(\boldsymbol{x}^*) = g(\boldsymbol{x}^*, \boldsymbol{y}^*) \leq g(\boldsymbol{x}_k, \boldsymbol{y}_k) = f(\boldsymbol{x}_k), \qquad \forall k > k'.$$

A contradiction occurs. □

Theorem 3.16. *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$*. If* $\boldsymbol{x}^*$ *is a local optimal solution to (CNO), then* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is a local optimal solution to (CNP).*

Proof. Suppose that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is not a local optimal solution to (CNP). Let a monotonically decreasing sequence $\varepsilon_k \to 0$ as $k \to +\infty$. For each neighborhood $B(\boldsymbol{x}^*, \boldsymbol{y}^*, \varepsilon_k)$ of $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, $k = 1, 2, \ldots,$ there is a $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in B(\boldsymbol{x}^*, \boldsymbol{y}^*, \varepsilon_k) \cap X(f)$ such that

$$f(\boldsymbol{x}_k) = g(\boldsymbol{x}_k, \boldsymbol{y}_k) < g(\boldsymbol{x}^*, \boldsymbol{y}^*) = f(\boldsymbol{x}^*).$$

It is clear that $(\boldsymbol{x}_k, \boldsymbol{y}_k) \to (\boldsymbol{x}^*, \boldsymbol{y}^*)$ as $k \to +\infty$. Because $\boldsymbol{x}^*$ is a local optimal solution to (CNO), there is $k' > 0$ such that

$$g(\boldsymbol{x}^*, \boldsymbol{y}^*) = f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}_k) = g(\boldsymbol{x}_k, \boldsymbol{y}_k), \qquad \forall k > k'.$$

A contradiction occurs.

Theorem 3.15 and Theorem 3.16 tell us that a local optimal solution $\boldsymbol{x}^*$ to (CNO) is equivalent to a local optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ to (CNP) if the exact CN function f is continuous. The optimal solution to (CNO) can be obtained by solving the local optimal solution or the global optimal solution to (CNP). □

4 Lagrange Duality of (CNP)

The advantage that the gap between the optimal objective function value of the Lagrangian dual problem of the convex optimization problem and the optimal objective function value of the original problem is zero makes the utilization of the dual problem of convex optimization important in getting the global optimal solution. So, next the dual problem of CN optimization is studied.

For any $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m, \boldsymbol{u} \in R^r$, a Lagrange function of (CNP) is defined by

$$L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) = g(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}). \tag{4.1}$$

Let a dual function of $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})$ on $\boldsymbol{u} \in R^r$ be defined by

$$\theta(\boldsymbol{u}) = \min\{L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})|(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\}. \tag{4.2}$$

Let $R^r_+ = \{\boldsymbol{u} \in R^r|\boldsymbol{u} \geq 0\}$. For any $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m$ and $\boldsymbol{u} \in R^r_+$, a proper Lagrange function of (CNP) is defined by

$$L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) = g(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}). \tag{4.3}$$

Let a proper dual function of $L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})$ on $\boldsymbol{u} \in R^r_+$ be defined by

$$\theta_+(\boldsymbol{u}) = \min\{L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})|(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\}. \tag{4.4}$$

The following conclusions are clear.

Proposition 4.1. *If $\boldsymbol{u} \in R^r_+$, then $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) = L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})$ and $\theta(\boldsymbol{u}) = \theta_+(\boldsymbol{u})$.*

Proposition 4.2. *If $(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*)$ is a saddle point of $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})$ with $\boldsymbol{u}^* \in R^r_+$, then $(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*)$ is a proper saddle point of $L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u})$.*

A dual optimization problem of (CNP) is defined by

$$\text{(DLCNP)} \quad \max\ \theta(\boldsymbol{u}) \quad s.t.\ \boldsymbol{u} \in R^r.$$

It is clear that $\theta(\boldsymbol{u})$ is a concave function on R^r.

A proper dual optimization problem of (CNP) is defined by

$$\text{(PDLCNP)} \quad \max\ \theta_+(\boldsymbol{u}) \quad s.t.\ \boldsymbol{u} \in R^r_+.$$

It is clear that $\theta_+(\boldsymbol{u})$ is a concave function on R^r_+. By (4.1-4.4), the following weak duality is clear.

Proposition 4.3. (i) *For all $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m$ and $\boldsymbol{u} \in R^r$, $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) \geq \theta(\boldsymbol{u})$.*

(ii) *For all $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$ and $\boldsymbol{u} \in R^r$, $g(\boldsymbol{x}, \boldsymbol{y}) \geq \theta(\boldsymbol{u})$.*

(iii) *For all $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m$ and $\boldsymbol{u} \in R^r_+$, $L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) \geq \theta_+(\boldsymbol{u})$.*

(iv) *For all $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$ and $\boldsymbol{u} \in R^r_+$, $g(\boldsymbol{x}, \boldsymbol{y}) \geq \theta_+(\boldsymbol{u})$.*

The following strong duality theorem is proved.

Theorem 4.4. (i) *If* $L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = \theta(\boldsymbol{u}^*)$ *at* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in R^n \times R^m$ *and* $\boldsymbol{u}^* \in R^r$*, then*

$$\nabla L(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{u}^* = 0. \tag{4.5}$$

(ii) *If* $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta(\boldsymbol{u}^*)$ *at* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *and* $\boldsymbol{u}^* \in R^r$*. Then (4.5) holds,* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNP) and* $\boldsymbol{u}^*$ *is an optimal solution to (DLCNP).*

(iii) *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in R^n \times R^m$ *and* $\boldsymbol{u}^* \in R^r_+$*. Then*

$$\nabla L_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} \boldsymbol{u}^* = 0 \tag{4.6}$$

holds if and only if $L_+(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = \theta_+(\boldsymbol{u}^*)$.

(iv) *Suppose that* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *and* $\boldsymbol{u}^* \in R^r_+$*. Then (4.6) holds if and only if* $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta_+(\boldsymbol{u}^*)$*. Furthermore,* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to (CNP) and* $\boldsymbol{u}^*$ *is an optimal solution to (PDLCNP).*

Proof. (i) Because for fixed $\boldsymbol{u}^*$,

$$\theta(\boldsymbol{u}^*) = \min\{L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*) | (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\} = L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*),$$

$(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\{L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*) | (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\}$. Hence, $\nabla L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = 0$.

(ii) If $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta(\boldsymbol{u}^*)$ at $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ and $\boldsymbol{u}^* \in R^r$, $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = L(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = \theta(\boldsymbol{u}^*)$. So, $\nabla L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = 0$ holds. For all $(\boldsymbol{x}, \boldsymbol{y}) \in X(f)$, we have

$$g(\boldsymbol{x}, \boldsymbol{y}) \geq L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}^*) \geq L(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = g(\boldsymbol{x}^*, \boldsymbol{y}^*).$$

Let $\bar{\boldsymbol{u}}$ be an optimal solution to (DLCNP). We have

$$g(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq L(\boldsymbol{x}^*, \boldsymbol{y}^*, \bar{\boldsymbol{u}}) \geq \theta(\bar{\boldsymbol{u}}) \geq \theta(\boldsymbol{u}^*) = g(\boldsymbol{x}^*, \boldsymbol{y}^*).$$

Hence, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{u}^*$ is an optimal solution to (DLCNP).

(iii) Because for fixed $\boldsymbol{u}^*$,

$$\theta_+(\boldsymbol{u}^*) = \min\{L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*) | (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\} = L_+(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*),$$

then $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\{L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*) | (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\}$. Hence, $\nabla L_+(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = 0$.

Now, let us prove, if (4.6) is true, $L_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = \theta_+(\boldsymbol{u}^*)$. Because for a fixed $\boldsymbol{u}^*$, $L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*)$ is convex on $(\boldsymbol{x}, \boldsymbol{y})$, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\{L_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*) | (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m\}$. Hence,

$$\theta_+(\boldsymbol{u}^*) = L_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*).$$

(iv) For $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ and $\boldsymbol{u}^* \in R^r_+$, if (4.6) holds, $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta_+(\boldsymbol{u}^*)$ is true by (iii). On the other hand, if $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta_+(\boldsymbol{u}^*)$, then (4.6) holds, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to (CNP) and $\boldsymbol{u}^*$ is an optimal solution to (PDLCNP), similar to the proof of (ii) . □

By Theorem 4.4, the gap between the optimal objective value of the proper Lagrangian dual problem of (CNP) and the optimal objective value of the original problem (CNP) may be zero. When a Lagrange function $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*)$ is convex on $(\boldsymbol{x}, \boldsymbol{y})$ at $\boldsymbol{u}^* \in R^r$, the conclusion of zero gap holds by the proof of Theorem 4.4 (iii) and (iv). Theorem 4.4 tells us that there is not any CN form of f such that the conclusion of Theorem 4.4 is true if there is not any optimal solution to (CNP).

Corollary 4.5. (i) *Suppose that for* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in R^n \times R^m$ *and* $\boldsymbol{u}^* \in R^r$*, a Lagrange function* $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*)$ *is convex on* $(\boldsymbol{x}, \boldsymbol{y})$ *at* $\boldsymbol{u}^* \in R^r$*. If (4.5) holds, then* $L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{u}^*) = \theta(\boldsymbol{u}^*)$*.*

(ii) *Suppose that for* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *and* $\boldsymbol{u}^* \in R^r$*, a Lagrange function* $L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*)$ *is convex on* $(\boldsymbol{x}, \boldsymbol{y})$ *at* $\boldsymbol{u}^* \in R^r$*. If (4.5) hold, then* $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta(\boldsymbol{u}^*)$*.*

An augmented Lagrange penalty function of (CNP) is define by

$$A(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}, \rho) = g(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{u}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2, \quad (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m, \boldsymbol{u} \in R^r, \tag{4.7}$$

where $\rho > 0$ is a penalty parameter. A proper augmented Lagrange penalty function of (CNP) is defined by

$$A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}, \rho) = g(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{u}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2, \quad (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m, \boldsymbol{u} \in R^r_+, \tag{4.8}$$

where $\rho > 0$ is a penalty parameter.

Theorem 4.6. (i) *If* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *is an optimal solution to* $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*)$ *at* $\boldsymbol{u}^* \in R^r$*, then* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to* $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ *for all* $\rho > 0$*.*

(ii) *Let* $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$ *and* $\boldsymbol{u}^* \in R^r_+$*. Then (4.6) holds if and only if* $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ *is an optimal solution to* $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ *for all* $\rho > 0$*.*

Proof. (i) For any $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m$, we have

$$\begin{aligned} A(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho) &= g(\boldsymbol{x}, \boldsymbol{y}) + \sum_{i=1}^{r} u_i^* g_i(\boldsymbol{x}, \boldsymbol{y}) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2 \\ &\geq L(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \boldsymbol{v}^*) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2 \\ &\geq L(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = A(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*, \rho). \end{aligned}$$

Hence, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ for all $\rho > 0$.

(ii) If (4.6) holds, let's first prove that $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ for all $\rho > 0$. Let any $(\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m$. Since $g(\boldsymbol{x}, \boldsymbol{y})$ and $g_i(\boldsymbol{x}, \boldsymbol{y}) (i = 1, 2, ..., r)$ are convex, we have

$$g(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} [(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)], \tag{4.9}$$

$$g_i(\boldsymbol{x}, \boldsymbol{y}) - g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) \geq \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*)^{\mathrm{T}} [(\boldsymbol{x}, \boldsymbol{y}) - (x^*, \boldsymbol{y}^*)], i = 1, 2, ..., r. \tag{4.10}$$

From (4.9) and (4.10), we have

$$\begin{aligned} A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho) &= g(\boldsymbol{x}, \boldsymbol{y}) + \sum_{i=1}^{r} u_i^* g_i(\boldsymbol{x}, \boldsymbol{y}) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2 \\ &\geq g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \rho \sum_{i=1}^{r} g_i(\boldsymbol{x}, \boldsymbol{y})^2 \\ &\geq g(\boldsymbol{x}^*, \boldsymbol{y}^*) = A_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*, \rho). \end{aligned}$$

Hence, $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ for all $\rho > 0$.

If $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is an optimal solution to $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A_+(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^*, \rho)$ for all $\rho > 0$, it is clear that

$$\begin{aligned}\nabla A_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*, \rho) &= \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \sum_{i=1}^{r} u_i^* \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) + \rho \sum_{i=1}^{r} 2 g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*), \\ &= \nabla L_+(\boldsymbol{x}^*, \boldsymbol{y}^*; \boldsymbol{u}^*) = 0.\end{aligned}$$

Theorems 4.4 and 4.6 mean that $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta(\boldsymbol{u}^*)$ or $g(\boldsymbol{x}^*, \boldsymbol{y}^*) = \theta_+(\boldsymbol{u}^*)$ does not hold, if there is not $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in R^n \times R^m$ and $\boldsymbol{u}^* \in R_+^r$ such that (4.5) or (4.6) holds. This means that (4.5) or (4.6) is the necessary condition, if there is an global optimal solution to (CNP) by Theorem 4.4.

In order to find a solution to (CNO), by Theorem 4.6, the augmented Lagrange penalty optimization problem defined as the following is applied:

$$\begin{aligned}(\text{CNP})(\boldsymbol{u}, \rho) \quad &\min \quad A(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}, \rho), \\ &s.t. \quad (\boldsymbol{x}, \boldsymbol{y}) \in R^n \times R^m.\end{aligned}$$

To solve the problem (CNP)$(\boldsymbol{u}, \rho)$, an algorithm of augmented Lagrange penalty function of (CNP)(which is called ALPF Algorithm) is proposed. □

ALPF Algorithm.

Step 1: Let $\epsilon > 0, \rho_1 > 0, N > 1, (\boldsymbol{x}^1, \boldsymbol{y}^1) \in R^n \times R^m$, $\boldsymbol{u}^1 \in R^r$ and $k = 1$.

Step 2: Find $(\boldsymbol{x}^k, \boldsymbol{y}^k) \in R^n \times R^m$ to the problem $\min\limits_{(\boldsymbol{x},\boldsymbol{y})} A(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}^k, \rho_k)$ such that $\nabla A(\boldsymbol{x}^k, \boldsymbol{y}^k; \boldsymbol{u}^k, \rho_k) = 0$, and go to Step 3.

Step 3: If $(\boldsymbol{x}^k, \boldsymbol{y}^k) \in X(f)$ and $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}^k)$ is convex on $(\boldsymbol{x}, \boldsymbol{y})$, then stop and $\boldsymbol{x}^k$ is an optimal solution to (CNO). Otherwise, go to Step 4.

Step 4: If $|A(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{u}^k, \rho_k) - g(\boldsymbol{x}^k, \boldsymbol{y}^k)| < \epsilon$ and $\|\boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{y}^k)\| < \epsilon$, then stop and $\boldsymbol{x}^k$ is an approximate solution to (CNO). Otherwise, let $\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + 2\rho_k \boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{y}^k)$, $\rho_{k+1} = N\rho_k$, $k := k + 1$ and go to Step 2.

Note: By Theorems 4.4 and 4.6, if $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal solution to $\lim\limits_{(\boldsymbol{x},\boldsymbol{y})} A(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}^k, \rho_k)$, $A(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{u}^k, \rho_k) = g(\boldsymbol{x}^k, \boldsymbol{y}^k)$ and $(\boldsymbol{x}^k, \boldsymbol{y}^k) \in X(f)$ for some k, then $\theta(\boldsymbol{u}^k) = g(\boldsymbol{x}^k, \boldsymbol{y}^k)$ and $\boldsymbol{x}^k$ is an optimal solution to (CNO). Hence, $\boldsymbol{x}^k$ may be an approximate solution to (CNO) if $|A(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{u}^k, \rho_k) - g(\boldsymbol{x}^k, \boldsymbol{y}^k)| < \epsilon$ and $\|\boldsymbol{g}(\boldsymbol{x}^k, \boldsymbol{y}^k)\| < \epsilon$ hold. ALPF Algorithm may be able to find an approximate global optimal solution to (CNO). Under some conditions, it is proved that ALPF Algorithm can converge to a KKT point for $\epsilon = 0$.

Let

$$S(\pi, g) = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \pi \geq g(\boldsymbol{x}, \boldsymbol{y})\},$$

which is called a level set. If $S(\pi, g)$ is bounded for any given $\pi > 0$, then $S(\pi, g)$ is called to be bounded.

Theorem 4.7. *Let $\epsilon = 0$. Suppose that a sequence of $\{(\boldsymbol{x}^k, \boldsymbol{y}^k)\}$, $k = 1, 2, \ldots$, is obtained by ALPF Algorithm. Let the sequence of $\{H_k(\boldsymbol{x}^k, \boldsymbol{y}^k, \rho_k)\}$, $k = 1, 2, \ldots$, be bounded and the level set $S(\pi, g)$ be bounded, where*

$$H_k(\boldsymbol{x}^k, \boldsymbol{y}^k, \rho_k) = g(\boldsymbol{x}^k, \boldsymbol{y}^k) + \rho_k \sum_{i=1}^{r} g_i(\boldsymbol{x}^k, \boldsymbol{y}^k)^2.$$

(i) *If the algorithm stops in a finite number of steps k, then $\boldsymbol{x}^k$ is a global optimal solution to (CNO).*

(ii) *If the sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k)\}$ is an infinite sequence, then $\{(\boldsymbol{x}^k, \boldsymbol{y}^k)\}$ is bounded and any limit point $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ of the sequence belongs to $X(f)$, and there exist $\eta > 0$ and λ_i, $i = 1, 2, \ldots, r$, such that*

$$\eta \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \sum_{i=1}^{r} \lambda_i \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0. \tag{4.11}$$

If $\lambda_i \geq 0$, $i = 1, 2, \ldots, r$, then $\boldsymbol{x}^*$ is an optimal solution to (CNO)

Proof. (i) The conclusion is clear by Theorem 4.4 and Corollary 4.5.

(ii) By ALPF algorithm, as $k \to +\infty$, since $\{H_k(\boldsymbol{x}^k, \boldsymbol{y}^k, \rho_k)\}$ is bounded, there must be some $\pi > 0$ such that

$$\begin{aligned} \pi &> H_k(\boldsymbol{x}^k, \boldsymbol{y}^k, \rho_k) \\ &= g(\boldsymbol{x}^k, \boldsymbol{y}^k) + \rho_k \sum_{i=1}^{r} g_i(\boldsymbol{x}^k, \boldsymbol{y}^k)^2 \\ &\geq g(\boldsymbol{x}^k, \boldsymbol{y}^k). \end{aligned}$$

$\{(\boldsymbol{x}^k, \boldsymbol{y}^k)\}$ is bounded because the level set $S(\pi, f)$ is bounded. Without loss of generality, suppose $(\boldsymbol{x}^k, \boldsymbol{y}^k) \to (\boldsymbol{x}^*, \boldsymbol{y}^*)$. Since g is continuous, $S(\pi, g)$ is closed. So, $g(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is bounded and there is a $\sigma > 0$ such that $g(\boldsymbol{x}^k, \boldsymbol{y}^k > -\sigma$.

From the above inequality, we have that

$$\sum_{i=1}^{r} g_i(\boldsymbol{x}^k, \boldsymbol{y}^k)^2 \leq \frac{1}{\rho_k}(\pi - g(\boldsymbol{x}^k, \boldsymbol{y}^k)) < \frac{\pi + \sigma}{\rho_k}.$$

We have $\sum\limits_{i=1}^{r} (g_i(\boldsymbol{x}^k, \boldsymbol{y}^k))^2 \to 0$ as $\rho_k \to +\infty$. So, $(\boldsymbol{x}^*, \boldsymbol{y}^*) \in X(f)$.

By ALPF algorithm, there is an infinite sequence $\{(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{u}^k, \rho_k)\}$ such that $\nabla A(\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{u}^k, \rho_k) = 0$. We have

$$\nabla g(\boldsymbol{x}^k, \boldsymbol{y}^k) + \sum_{i=1}^{r} u_i^{k+1} \nabla g_i(\boldsymbol{x}^k, \boldsymbol{y}^k) = 0, \tag{4.12}$$

where $u_i^{k+1} = u_i^k + 2\rho_k g_i(\boldsymbol{x}^k, \boldsymbol{y}^k), i = 1, 2, ..., r$. Let

$$\gamma_k = 1 + \sum_{i=1}^{r} (\max\{u_i^{k+1}, 0\} + \max\{-u_i^{k+1}, 0\}) > 0.$$

Let $\eta^k = \frac{1}{\gamma_k} > 0$, $\mu_i^k = \frac{\max\{u_i^{k+1}, 0\}}{\gamma_k} \geq 0, i = 1, 2, ..., r$ and $\nu_i^k = \frac{\max\{-u_i^{k+1}, 0\}}{\gamma_k} \geq 0, i = 1, 2, ..., r$. Then,

$$\eta^k + \sum_{i=1}^{r} (\mu_i^k + \nu_i^k) = 1. \tag{4.13}$$

Clearly, as $k \to \infty$, we have $\eta^k \to \eta > 0, \mu_i^k \to \mu_i, \nu_i^k \to \nu_i, \forall i = 1, 2, ..., r$. By (4.12) and (4.13), we have

$$\eta \nabla g(\boldsymbol{x}^*, \boldsymbol{y}^*) + \sum_{i=1}^{r} (\mu_i - \nu_i) \nabla g_i(\boldsymbol{x}^*, \boldsymbol{y}^*) = 0. \tag{4.14}$$

By (4.14), let $\lambda^k = \mu_i^k - \nu_i^k \to \lambda$ as $k \to +\infty$, and we have (4.11). □

If (CNO) is a separable problem, such as

$$\min_{\mathbf{x}} f(\boldsymbol{x}) = f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + \cdots + f_m(\boldsymbol{x}_m),$$

where $\mathbf{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m) \in R^{m_1} \times R^{m_2} \times \cdots \times R^{m_m}$, by ALPF algorithm, we can solve these subproblems $\min\limits_{\boldsymbol{x}_k} f_k(\boldsymbol{x}_k)$ for $k = 1, 2, \ldots, m$ one by one. The above idea has been applied to literature [27, 46, 47]. It shows that CN optimization can solve some large-scale optimization separable problems with special structures.

5 Conclusion

In this paper, a new concept, CN function, is proposed, which covers many nonconvex, nonsmooth functions, and even discontinuous nonconvex functions. Some sufficient conditions of the global optimal solution are given for the unconstrained CN optimization problems. Lagrange function and the proper Lagrange function of the CN function are defined, as well as their dual problem and saddle point. Under some conditions, the strong duality theorem is proved. Moreover, the augmented Lagrangian penalty function and the proper augmented Lagrangian penalty function are introduced. It is shown that the optimal solution to the Lagrangian optimization problem is the optimal solution to the augmented Lagrangian penalty optimization problem for all positive penalty parameters. An algorithm is also proposed to solve the augmented Lagrangian penalty optimization problem, and its global convergence is attained.

This paper provides a feasible method for solving nonsmooth or nonconvex unconstrained optimization problems in theory, which shows its potential importance in solving nonsmooth nonconvex optimization problems in many application fields.

Acknowledgments

The authors would like to express their gratitudes to anonymous referees' detailed comments and remarks that help us improve our presentation of this paper considerably.

References

[1] F.A. Al-Khayyaltt and Falk, Jointly constrained biconvex programming, *Mathematics of Operations Research* 8 (1983) 273–286.

[2] A.M. Bagirov, A method for minimizing of quasidifferentiable functions, *Optim Methods Software* 17 (2002) 31–60.

[3] A.M. Bagirov, N. Karmitsa and M.M. Mäkelä, *Introduction to Nonsmooth Optimization: Theory, Practice and Software*. Springer, New York, 2014.

[4] X. Bao, A. Khajavirad, N.V. Sahinidis, and M. Tawarmalani, Global optimization of nonconvex problems with multilinear intermediates, *Mathematical Programming Computation* 7 (2015) 1–37.

[5] M.L. Bougeard, Morse theory for some lower-C-2 functions in finite dimension, *Mathematical Programming* 41 (1988) 141–159.

[6] G.A. Bunin, Extended reverse-convex programming: an approximate enumeration approach to global optimization, *Journal of Global Optimization* 65 (2016) 191–229.

[7] X.J. Chen, D.D. Ge, Z.Z. Wang and Y.Y. Ye, Complexity of unconstrained $L_2 - L_p$ minimization, *Mathematical Programming*, Ser. A 143 (2014) 371–383.

[8] X.J. Chen, F.M. Xu and Y.Y. Ye, Lower bound theory of nonzero entries in solutions of $L_2 - L_p$ minimization, *SIAM Journal Science Computing* 32 (2010) 2832–2852

[9] N.H. Chieu, G.M. Lee, B.S. Mordukhovich and T.T.A. Nghia, Coderivative characterizations of maximal monotonicity for set-valued mappings, *Journal of Convex Analysis* 23 (2016) 461–480.

[10] W.Y.Chiu, Method of reduction of variables for bilinear matrix inequality problems in system and control designs, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47 (2017) 1241–1256.

[11] F.H. Clarke, *Opimization and Nonsmooth Analysis*, John Wiley & Sons. Inc.(1983).

[12] A. Daniilidis and J. Malick, Filling the gap between lower-C-1 and lower-C-2 functions, *Journal of Convex Analysis* 12 (2005) 315–329.

[13] M.N.Dao, Bundle method for nonconvex nonsmooth constrained optimization, *Journal of Convex Analysis* 22 (2015) 1061–1090.

[14] G. Eichfelder, K. Klamroth and J. Niebling, Nonconvex constrained optimization by a filtering branch and bound, *Journal of Global Optimization* 80 (2021) 31–61.

[15] M. Frank and P. Wolfe, An algorithm for quadratic programming. *Naval Researach Logisics* 3 (1956) 95–110.

[16] M. Grant, S. Boyd and Y.Y. Ye, Disciplined convex programming, in: *Global Optimization: From Theory to Implementation, Nonconvex Optimization and its Applications*, L. Liberti and N. Maculan (eds), Springer, 2006, pp.155–210.

[17] L. Granvilliers, Filtering Domains of Factorable Functions Using Interval Contractors, in: *WCGO 2019*, AISC 991, 2020, pp. 99–108.

[18] W.L. Hare, Functions and sets of smooth substructure: relationships and examples, *Computational Optimization and Applications* 33 (2006) 249–270.

[19] W.L. Hare and R.A. Poliquin, Prox-Regularity and stability of the proximal mapping, *Journal of Convex Analysis* 14 (2007) 589–606.

[20] W.L. Hare and C. Sagastizábal, Computing proximal points of nonconvex functions, *Mathematical Programming, Ser. B* 116 (2009) 221–258.

[21] W.L. Hare and C. Planiden, Parametrically prox-regular functions, *Journal of Convex Analysis* 21 (2014) 901–923.

[22] W.L. Hare and C. Sagastizábal, A redistributed proximal bundle method for nonconvex optimization, *SIAM Journal Optimization* 20 (2010) 2442–2473.

[23] W.L. Hare, C. Sagastizábal and M. Solodov, A proximal bundle method for nonsmooth nonconvex functions with inexact information, *Computational Optimization Applcation* 63 (2016) 1–28.

[24] P. Hartman, On functions representable as a difference of convex functions, *Pacific Journal Mathematics* 9 (1959) 707–713

[25] T.T. He and M. Tawarmalani, A new framework to relax composite functions in nonlinear programs, *Mathematical Programming*190 (2021) 427–466.

[26] R.H.F. Jackson and G.P. Mccormick, Second-order sensitivity analysis in factorable programming - theory and applications, *Mathematical Programming* 41 (1988) 1–27.

[27] M. Jiang, R. Shen, Z.Q. Meng and C.Y. Dang, Optimization conditions and decomposable algorithms for convertible nonconvex optimization, *Journal of Nonlinear Variational Analysis* 7 (2023) 103–128.

[28] M. Jiang, R. Shen, Z.Q. Meng and C.Y. Dang, Exact penalty algorithm of strong convertible nonconvex optimization, 2022, https://doi.org/10.48550/arXiv.2202.07317

[29] A. Khajavirad, J.J. Michalek and N.K. Sahinidis, Relaxations of factorable functions with convex-transformable intermediates, *Mathematical Programming*144 (2014) 107–140.

[30] W.J. Li and W. Bian, Projection neural network for a class of sparse regression problems with cardinality penalty, *Neurocomting* 431 (2021) 188–200.

[31] A. Lundell, J. Westerlund and T. Westerlund, Some transformation techniques with applications in global optimization, *Jurnal of Globel Optimization* 43 (2009) 391–405.

[32] G.P. McCormick, Computability of global solutions to factorable nonconvex programs: part I-convex underestimating problems, *Mathematical Programming* 10 (1976) 147–175.

[33] M. Mohri, A. Rostamizadeh and A. Talwalkar, *Foundations of Machine Learning*,The MIT Press Cambridge, Massachusetts, London, England, 2012.

[34] H.N. Ngai and J.-P. Penot, Subdifferentiation of regularized functions, *Set-Valued Variational Analysis* 24 (2016) 167–189.

[35] J.C.Nohra and N.V. Sahinidis, Global optimization of nonconvex problems with convex-transformable intermediates, *Journal Globel Optimization* 72 (2018) 255–276.

[36] D. Noll, Cutting plane Oracles to minimize non-smooth non-convex functions, *Set-Valued Analysis*18 (2010) 531–568.

[37] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[38] F. Serrano, Intersection Cuts for Factorable MINLP, in: *IPCO 2019*, LNCS 11480, A. Lodi and V. Nagarajan (eds), 2019, pp. 385–398.

[39] X. Shen, S. Diamond, M. Udell and Y. Gu, Disciplined multi-convex programming. in: *Control And Decision Conference (CCDC)*, Chongqing, Chinese, ieeexplore.ieee.org, 2017. DOI: 10.1109/CCDC.2017.7978647.

[40] H.D. Sherali and H.J. Wang, Global optimization of nonconvex factorable programming problems, *Mathematical Programming* 89 (2001) 459–478.

[41] J.E.Spingarn, Submonotone subdifferentials of Lipschitz functions, *Transactions on Americal Mathematics Society* 264 (1981) 77–89.

[42] M. Tawarmalani and N.V. Sahinidis, Global optimization of mixed-integer nonlinear programs: A theoretical and computational study, *Mathematical Programming* 99 (2004) 563–591.

[43] M. Tawarmalani, J.P. Richard and C. Xiong, Explicit convex and concave envelopes through polyhedral subdivisions. *Mathematical Programming*138 (2013) 531–577.

[44] X.F. Wang, W.J. Zhang, J.C. Yana, X.M. Yuan and H.Y. Zha, On the flexibility of block coordinate descent for large-scale optimization, *Neurocomputing* 272 (2018) 471–480.

[45] H.M. Zhang and B. Dong, A review on deep learning in medical image reconstruction, *Journal Operational Researsch Society China* 8 (2020) 311–340.

[46] Z.Q. Meng, G.D. Wen, Y. Pan, R. Shen and M. Jiang, Sparse vector feature extraction method based on SCN Optimization, *PeerJ Computer Sciencs*, under review, 2022.

[47] Z.Q. Meng, Y. Pan, M. Jiang, G.D. Wen and R. Shen, Feature block decomposition algorithm of sparse support vector machine based on SCN function, in: *IEEE Transactions on Neural Networks and Learning Systems*, under review, 2022.

Manuscript received 6 January 2023
revised 11 April 2023
accepted for publication 14 June 2023

Min Jiang
School of Management, Zhejiang University of Technology
Hangzhou, Zhejiang, 310023, China
E-mail address: jiangmin@zjut.edu.cn

Zhiqing Meng
School of Management, Zhejiang University of Technology
Hangzhou, Zhejiang, 310023, China
E-mail address:mengzhiqing@zjut.edu.cn

Chuangyin Dang
Department of Advanced Design and System Engineering
City University of Hong Kong, Kowloon, Hong Kong
E-mail address: mecdang@cityu.edu.hk

Rui Shen
School of Economics, Zhejiang University of Technology
Hangzhou, Zhejiang, 310023, China
E-mail address: shenrui126@126.com