MDPI

Article

Partial Exactness for the Penalty Function of Biconvex Programming

Min Jiang [1], **Zhiqing Meng** [1,*] and **Rui Shen** [2]

1   School of Management, Zhejiang University of Technology, Hangzhou 310023, China; jiangmin@zjut.edu.cn
2   School of Economics, Zhejiang University of Technology, Hangzhou 310023, China; shenrui@zjut.edu.cn
*   Correspondence: mengzhiqing@zjut.edu.cn

Abstract: Biconvex programming (or inequality constrained biconvex optimization) is an important model in solving many engineering optimization problems in areas like machine learning and signal and information processing. In this paper, the partial exactness of the partial optimum for the penalty function of biconvex programming is studied. The penalty function is partially exact if the partial Karush–Kuhn–Tucker (KKT) condition is true. The sufficient and necessary partially local stability condition used to determine whether the penalty function is partially exact for a partial optimum solution is also proven. Based on the penalty function, an algorithm is presented for finding a partial optimum solution to an inequality constrained biconvex optimization, and its convergence is proven under some conditions.

Keywords: biconvex programming; partial optimum; partially exact penalty function; partial exactness; partial local stability

1. Introduction

Multi-convex programming is a non-convex optimization problem [1,2], where biconvex programming is a special case. It is ubiquitous nowadays in fields such as control [3–5], machine learning [6,7], signal and information processing [8,9], communication [10,11], and also NP-hard problems [12]. The existing research on multi-convex programming mainly solves some very special models [12–19]. These studies all give specific methods for each special model. When problems change, new methods need to be studied. Biconvex programming is a simple case of multi-convex programming, and it is also studied in research [14] as a special case. Therefore, it is of great significance to study the algorithm of biconvex programming for solving practical engineering optimization problems.

Bilinear programming is the simplest case of biconvex programming and is the earliest and most studied as per [14]. There are two main ways to solve bilinear programming. One is the simplex algorithm based on sub-problems, and the other is the alternating direction method. For example, Liang and Bai [15] and Hajinezhad and Shi [16] both proposed the alternating direction method of multipliers (ADMM) algorithm for two special bilinear programming problems, where the extended Lagrangian penalty function uses a square penalty. Furthermore, Charkhgard et al. [17] presented a multi-linear programming algorithm using the linear programming algorithm.

In 2007, Gorski et al. [14] reviewed the development of the theory and algorithms of biconvex optimization. The biconvex functions have some good properties that are similar to convex functions, such as the biconvex separation theorem and an equivalence between local optimal solutions and stationary points. The biconvex programming algorithm is based on the idea that the solution to alternating sub-problems can converge to the stationary point of the original problem. For example, in 2015, Li et al. [18] studied an alternating convex search method to solve a stationary point problem of biconvex programming. In 2016, Shah et al. [19] presented an alternating search method with a square penalty function to solve biconvex programming, which had great effect in image

recognition. Al-Khayyaltt and Falk [20] discussed an algorithm of biconvex programming using the idea of branch-and-bound.

In short, based on the above, the mainstream method used to solve biconvex programming is an alternative to subproblem solving, because the bi-convexity of biconvex programming guarantees the convergence of alternative search methods and effectually cuts down the scale of problem solving via decomposition calculation. Hence, in view of the large scale of biconvex programming(BCP), alternative subproblem solving will be the main method adopted in further research regarding biconvex programming.

In this paper we consider biconvex programming with inequality constraints as follows:

$$\text{(BCP)} \quad \min \quad h(\boldsymbol{u}_1, \boldsymbol{u}_2)$$
$$\text{s.t.} \quad g_i(\boldsymbol{u}_1, \boldsymbol{u}_2) \leq 0, \ i = 1, 2, \cdots, m,$$

where $h, g_i : R^{n_1} \times R^{n_2} \to R(i = 1, 2, \cdots, m)$ are biconvex if $h(\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $g_i(\boldsymbol{u}_1, \boldsymbol{u}_2)$ are convex in $\boldsymbol{u}_1 \in R^{n_1}$ for every fixed $\boldsymbol{u}_2$ and in $\boldsymbol{u}_2 \in R^{n_2}$ for every fixed $\boldsymbol{u}_1$. Let $\mathbf{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$ and $I = \{1, 2, \cdots, m\}$. The feasible set of (BCP) is denoted by:

$$\mathbf{U} = \{(\boldsymbol{u}_1, \boldsymbol{u}_2) \in R^{n_1} \times R^{n_2} \mid g_i(\boldsymbol{u}_1, \boldsymbol{u}_2) \leq 0, \ i \in I\}.$$

Let

$$\mathbf{U}(\boldsymbol{u}_2) = \{\boldsymbol{u}_1 \in R^{n_1} \mid g_i(\boldsymbol{u}_1, \boldsymbol{u}_2) \leq 0, \ i \in I\}.$$

When $\boldsymbol{u}_2$ is fixed, define the suboptimal problem:

$$\text{(BCP)}(\boldsymbol{u}_2) \quad \min \quad h(\boldsymbol{u}_1, \boldsymbol{u}_2)$$
$$\text{s.t.} \quad \boldsymbol{u}_1 \in \mathbf{U}(\boldsymbol{u}_2).$$

Let

$$\mathbf{U}(\boldsymbol{u}_1) = \{\boldsymbol{u}_2 \in R^{n_2} \mid g_i(\boldsymbol{u}_1, \boldsymbol{u}_2) \leq 0, \ i \in I\}.$$

When $\boldsymbol{u}_1$ is fixed, define the suboptimal problem:

$$\text{(BCP)}(\boldsymbol{u}_1) \quad \min \quad h(\boldsymbol{u}_1, \boldsymbol{u}_2)$$
$$\text{s.t.} \quad \boldsymbol{u}_2 \in \mathbf{U}(\boldsymbol{u}_1).$$

Many practical engineering optimization problems can be transformed into biconvex programming, for example the two-cardinality sparse convex optimization problem and the nonconvex quadratic programming problem.

It is well known that an effective method to solve constrained optimization is by penalty function [21]. Its main idea is to transform a constrained optimization problem into a sequence of unconstrained optimization subproblems that are easier to solve. When the penalty function is not exact, many calculations are needed to solve unconstrained optimization subproblems so as to obtain an approximate solution to inequality constrained optimization. For example, it is proven that the exact penalty function method for solving constrained optimization is very efficient, as first proposed by Zangwill (1967) in [22]. In theory, if the penalty function is exact, to obtain an optimal solution to a constrained optimization problem only one unconstrained optimization subproblem is solved. Hence, the exact penalty function algorithm takes less time than the inexact penalty function algorithm. Additionally, this approach has recently received great attention from both theoretical and practical arenas. Many studies were later presented based on the exact penalty function algorithm, such as Rosenberg (1986) [23] and Di Pillo (1986) [24]. In addition, it is essential to determine the exactness of a penalty function under the stability condition [25,26]. Hence, this paper mainly focuses on the relationship between the partial exactness of the penalty function and the partial stability of biconvex programming. On the other hand, there exist other approaches to reduce constrained optimization problems

to unconstrained ones; for example, the index method presented in the monograph [27] or the method for computable boundaries presented in [28]. This shows the significance of the study of the partial exact penalty function.

In order to ensure exactness, we propose the following penalty function:

$$H_p(\boldsymbol{u}_1, \boldsymbol{u}_2, \rho) = h(\boldsymbol{u}_1, \boldsymbol{u}_2) + \rho \sum_{i \in I} \max\{g_i(\boldsymbol{u}_1, \boldsymbol{u}_2), 0\}^p,$$

where penalty parameter $\rho > 0$ and $0 < p \leq 1$. By the definition of exactness in [29], a penalty function is exact at $p = 0.5$, but is not exact at $p = 1$, such that $\min_{x \in R^1} f(x) = x$ $s.t.$ $x^2 \leq 0$. Hence, the exactness of an optimization problem depends on the structure of the problem. In this paper, we will study the exactness of a more extensive penalty function for biconvex programming than the one presented in [30].

The remainder of the paper is organized as follows. In Section 2, for a partial optimum solution, the partial exactness of the penalty function is proven under the partial Karush–Kuhn–Tucker (KKT) condition or the partial stableness condition. An algorithm is presented to find out a partial optimum solution to (BCP) with convergence.

2. Partial Exactness and a Penalty Function for (BCP)

According to Gorski et al. [14], defining the partial optimum of (BCP) is very meaningful. The concept of the partial optimum of (BCP) is given as follows:

Definition 1. *Let $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \in \boldsymbol{U}$. If:*

$$h(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \leq h(\boldsymbol{u}_1, \boldsymbol{u}_2^*), \forall \boldsymbol{u}_1 \in \boldsymbol{U}(\boldsymbol{u}_2^*), \tag{1}$$

$$h(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \leq h(\boldsymbol{u}_1^*, \boldsymbol{u}_2), \forall \boldsymbol{u}_2 \in \boldsymbol{U}(\boldsymbol{u}_1^*), \tag{2}$$

then $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*)$ is called a partial optimum of (BCP). A partial optimum $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*)$ of (BCP) means that $\boldsymbol{u}_1^*$ is an optimal solution to (BCP)$(\boldsymbol{u}_2^*)$ and $\boldsymbol{u}_2^*$ is an optimal solution to (BCP)$(\boldsymbol{u}_1^*)$.

Next, let us give the equivalence of a partial optimum of (BCP) to a partial KKT point under some conditions.

Let $h, g_i : R^{n_1} \times R^{n_2} \to R$, $i \in I = \{1, 2, \cdots, m\}$ be biconvex and differentiable. Let $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \in \boldsymbol{U}$. If there are $\alpha_i (i = 1, 2, \cdots, m)$ such that:

$$\nabla h(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) + \sum_{i=1}^{m} \alpha_i \nabla g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \tag{3}$$

$$\alpha_i g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \alpha_i \geq 0, i = 1, 2, \cdots, m, \tag{4}$$

then $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*)$ is a KKT point of (BCP).

Let $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \in \boldsymbol{U}$. If there are $\alpha_i^{(1)}, \alpha_i^{(2)} (i = 1, 2, \cdots, m)$ such that:

$$\nabla_{\boldsymbol{u}_1} h(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) + \sum_{i=1}^{m} \alpha_i^{(1)} \nabla_{\boldsymbol{u}_1} g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \tag{5}$$

$$\nabla_{\boldsymbol{u}_2} h(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) + \sum_{i=1}^{m} \alpha_i^{(2)} \nabla_{\boldsymbol{u}_2} g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \tag{6}$$

$$\alpha_i^{(1)} g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \alpha_i^{(2)} g_i(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) = 0, \alpha_i^1 \geq 0, \alpha_i^2 \geq 0, i = 1, 2, \cdots, m, \tag{7}$$

then $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*)$ is a partial KKT point of (BCP).

Let $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*) \in \boldsymbol{U}$. The constraint of (BCP) is called a partial Slater constraint qualification at $(\boldsymbol{u}_1^*, \boldsymbol{u}_2^*)$, if there is $(\bar{\boldsymbol{u}}_1, \bar{\boldsymbol{u}}_2) \in R^{n_1} \times R^{n_2}$ such that:

$$g_i(\boldsymbol{u}_1^*, \bar{\boldsymbol{u}}_2) < 0, g_i(\bar{\boldsymbol{u}}_1, \boldsymbol{u}_2^*) < 0, i = 1, 2, \cdots, m.$$

In fact, if the Slater constraint qualification is satisfied for a convex programming, an optimal solution of the convex programming is equal to a KKT condition. For biconvex programming, we have the results in [28] as follows.

Theorem 1. *Let $(u_1^*, u_2^*) \in u$. If (BCP) is satisfied with partial Slater constraint qualification at (u_1^*, u_2^*), then (u_1^*, u_2^*) is a partial optimum of (BCP) if and only if (u_1^*, u_2^*) is a partial KKT point of (BCP).*

Corollary 1. *Let $(u_1^*, u_2^*) \in u$ be a partial optimum of (BCP). If (BCP) is satisfied with partial Slater constraint qualification at (u_1^*, u_2^*), then (u_1^*, u_2^*) is a KKT point of (BCP) if and only if (5), (6), and (7) hold with $\alpha_i^1 = \alpha_i^2, i = 1, 2, \cdots, m$.*

Example 1. *Let the biconvex programming:*

$$(BCP1) \quad \min \quad h(x, y) = xy$$
$$s.t. \quad g_1(x, y) = 1 - x - y \leq 0,$$

where $h, g_1 : R \times R \to R$. For $\forall \theta \in (0, 1)$, it is clear that $(x, y) = (\theta, 1 - \theta)$ is the partial KKT point and the partial optimum of (BCP1). If $y = 1 - x$, then $h(x, y) = x - x^2$. We have $h(x, y) \to -\infty$ as $x \to \infty$. Hence, a local optimal solution to (BCP1) cannot be solved, such as in $(x, y) = (\frac{1}{2}, \frac{1}{2})$.

Example 1 means that if there is no optimal solution to biconvex programming, there may exist a partial KKT point or a partial optimum. It is obvious that an optimal solution to biconvex programming is the partial optimum. For biconvex programming, there is a partial optimum even if an optimal solution to biconvex programming is not obtained. Example 1 further indicates that the partial optimum is very important to biconvex programming.

Let $\rho > 0$ be given. Consider the following optimization problem:

$$(\text{BCP}(u_2; \rho)) \quad \min \quad H_p(u_1, u_2, \rho), \quad \text{s.t. } u_1 \in R^{n_1}.$$

where u_1 is a decision variable when u_2 is fixed, and

$$(\text{BCP}(u_1; \rho)) \quad \min \quad H_p(u_1, u_2, \rho), \quad \text{s.t. } u_2 \in R^{n_2}.$$

where u_2 is a decision variable when u_1 is fixed.

Definition 2. *Let $(\bar{u}_1^*, \bar{u}_2^*) \in R^{n_1} \times R^{n_2}$. If:*

$$H_p(\bar{u}_1^*, \bar{u}_2^*, \rho) \leq H_p(u_1, \bar{u}_2^*, \rho), \forall u_1 \in R^{n_1}, \tag{8}$$
$$H_p(\bar{u}_1^*, \bar{u}_2^*, \rho) \leq H_p(\bar{u}_1^*, u_2, \rho), \forall u_2 \in R^{n_2}, \tag{9}$$

then $(\bar{u}_1^*, \bar{u}_2^*)$ is called a partial optimum of $H_p(u_1, u_2, \rho)$.

Next, we define the concept of a partially exact penalty function for biconvex programming as follows.

Definition 3. *(1) Let (u_1^*, u_2^*) be a partial optimum of (BCP). If there is a ρ' such that (u_1^*, u_2^*) is a partial optimum of $H_p(u_1, u_2, \rho)$ for $\forall \rho > \rho'$, then $H_p(u_1, u_2, \rho)$ is called a partially exact penalty function.*
(2) Let (u_1^*, u_2^*) be a partial optimum of $H_p(u_1, u_2, \rho)$. If (u_1^*, u_2^*) of $H_p(u_1, u_2, \rho)$ is a partial optimum of (BCP), then ρ is called a partially exact value of the penalty parameter.

Example 2 below shows the partial exactness of a penalty function under different p values.

Example 2. *Let the biconvex programming:*

$$
(BCP2) \quad \min \quad h(x,y) = xy + x + y \\
\text{s.t.} \quad x^2 + y^2 \le 0.
$$

It is clear that $(0,0)$ is a partial optimum of (BCP2). The penalty function for (BCP2) is defined by:

$$
H_p(x,y;\rho) = xy + x + y + \rho \max\{x^2 + y^2, 0\}^p.
$$

When $p = 1$, a partial optimum of $\min_{x,y \in R} H_1(x,y;\rho)$ is $(x(\rho), y(\rho)) = (-\frac{1}{2\rho+1}, -\frac{1}{2\rho+1})$. Letting $\rho \to +\infty$, we have $(x(\rho), y(\rho)) \to (0,0)$. Hence, $H_1(x,y,\rho)$ is not a partially exact penalty function. When $p = 0.5$, we have:

$$
H_{0.5}(x,y;\rho) = xy + x + y + \rho \sqrt{x^2 + y^2}.
$$

A partial optimum of $\min_{x,y \in R} H_{0.5}(x,y;\rho)$ is $(0,0)$ for $\rho \ge 0$. Hence, $H_{0.5}(x,y,\rho)$ is a partially exact penalty function for $\rho \ge 1$. Example 2 means that the partial exactness of penalty function depends on the parameter p.

Example 3. *Let the biconvex programming:*

$$
(BCP3) \quad \min \quad h(x,y) = xy \\
\text{s.t.} \quad xy \ge 1.
$$

It is easy to verify that $(x^*, y^*) = (\theta, \frac{1}{\theta})^T$ is an optimal solution to (BCP3) for any $\theta \ne 0$. Thus, $(x^*, y^*) = (\theta, \frac{1}{\theta})^T$ is a partial optimum of (BCP3) too. The example illustrates that all partial optimums of (BCP) may be optimal solutions.

$(x^*, y^*) = (1,1)^T$ is a partial optimum of (BCP3). The square penalty function for (BCP3) is defined by:

$$
H_2(x,y;\rho) = xy + \rho \max\{0, 1 - xy\}^2.
$$

It is easy to check that $(x^*, y^*) = (\frac{1}{\delta}(1 - \frac{1}{2\rho}), \delta)$ is a partial optimum of $H_2(x,y,\rho)$ for $\delta \ne 0$ and $\rho \ge 1$. It is easy to check that $H_2(x,y;\rho)$ is not partially exact.

For $0 < p \le 1$, a penalty function for (BCP3) is defined by:

$$
H_p(x,y;\rho) = xy + \rho \max\{0, 1 - xy\}^p.
$$

We have that $(x^*, y^*) = (1,1)$ is a partial optimum of $H_p(x,y;\rho)$ for $\rho > 1$. It is easy to check that $H_p(x,y;\rho)$ is partially exact for $\rho > 1$. We easily check that $(1,1)$ is a KKT point.

Example 3 illustrates that the partially exact penalty function for a partial optimum can be as good as traditional exact penalty functions.

We prove the similarity of the partially exact penalty result to [28].

Theorem 2. *Let $h(u_1, u_2) \ge 0, \forall (u_1, u_2) \in R^{n_1} \times R^{n_2}$. Let $(u_1^*, u_2^*) \in U$ be a partial optimum of (BCP). If (u_1^*, u_2^*) is a KKT partial point of (BCP), i.e., there are $\alpha_i^{(1)}, \alpha_i^{(2)} (i = 1, 2, \cdots, m)$ such that (5), (6), and (7) are true, then $H_p(u_1, u_2, \rho)$ is a partially exact penalty function for $\rho > \rho^*$, where:*

$$
\rho^* = \max\{h(u_1^*, u_2^*), \alpha_i^{(1)}, \alpha_i^{(2)} | i = 1, 2, \cdots, m\}.
$$

Proof. If (u_1^*, u_2^*) is a partial KKT point of (BCP), there are $\alpha_i^{(1)}$ and $\alpha_i^{(2)}(i = 1, 2, \cdots, m)$ such that (5), (6) and (7) are true. Because $h(u_1, u_2), g_i(u_1, u_2)(\forall i \in I)$ are biconvex functions for $\forall u_1 \in R^{n_1}$ and $\forall u_2 \in R^{n_2}$, we have:

$$
\begin{align}
h(u_1, u_2^*) &\geq h(u_1^*, u_2^*) + \nabla_{u_1} h(u_1^*, u_2^*)^T (u_1 - u_1^*), \tag{10} \\
g_i(u_1, u_2^*) &\geq g_i(u_1^*, u_2^*) + \nabla_{u_1} g_i(u_1^*, u_2^*)^T (u_1 - u_1^*), i \in I, \tag{11} \\
h(u_1^*, u_2) &\geq h(u_1^*, u_2^*) + \nabla_{u_2} h(u_1^*, u_2^*)^T (u_2 - u_2^*), \tag{12} \\
g_i(u_1^*, u_2) &\geq g_i(u_1^*, u_2^*) + \nabla_{u_2} g_i(u_1^*, u_2^*)^T (u_2 - u_2^*), i \in I. \tag{13}
\end{align}
$$

By (5), (7), (10), and (11), we have:

$$
\begin{align}
h(u_1, u_2^*) &\geq h(u_1^*, u_2^*) - \sum_{i=1}^{m} \alpha_i^{(1)} \nabla_{u_1} g_i(u_1^*, u_2^*)^T (u_1 - u_1^*) \\
&\geq h(u_1^*, u_2^*) + \sum_{i=1}^{m} \alpha_i^{(1)} (g_i(u_1^*, u_2^*) - g_i(u_1, u_2^*)) \\
&= h(u_1^*, u_2^*) - \sum_{i=1}^{m} \alpha_i^{(1)} g_i(u_1, u_2^*). \tag{14}
\end{align}
$$

Let $P(t) = \max\{t, 0\}^p, t \in R$, and $0 < p \leq 1$. Let us take $\rho > \rho^*$ with $\rho^* = \max\{h(u_1^*, u_2^*), \alpha_i^{(1)}, \alpha_i^{(2)} | i = 1, 2, \cdots, m\}$. If there is a $g_i(u_1, u_2^*) > 1$, then $P(g_i(u_1, u_2^*)) > 1$. We have:

$$
h(u_1, u_2^*) + \rho \sum_{i=1}^{m} P(g_i(u_1, u_2^*)) \geq \rho^* \geq h(u_1^*, u_2^*),
$$

i.e.,

$$
H_p(u_1, u_2^*, \rho) \geq H_p(u_1^*, u_2^*, \rho).
$$

Otherwise, if $g_i(u_1, u_2^*) \leq 1$ for $\forall i \in I$, from (14) we have:

$$
\begin{align}
h(u_1, u_2^*) + \rho \sum_{i=1}^{m} P(g_i(u_1, u_2^*)) &\geq h(u_1^*, u_2^*) \\
&\quad + \sum_{i=1}^{m} (\rho \max\{g_i(u_1, u_2^*), 0\}^p - \alpha_i^{(1)} g_i(u_1, u_2^*)).
\end{align}
$$

Hence, for $\rho > \rho^*$, we have:

$$
H_p(u_1^*, u_2^*, \rho) \leq H_p(u_1, u_2^*, \rho). \tag{15}
$$

Similarly, for $\rho > \rho^*$, we have:

$$
H_p(u_1^*, u_2^*, \rho) \leq H_p(u_1^*, u_2, \rho).
$$

By Definition 2, we have that $H_p(u_1, u_2, \rho)$ is a partially exact penalty function for $\rho > \rho^*$. $\square$

Note. If there is a number A such that $h(u_1, u_2) \geq A$ for any $(u_1, u_2) \in R^{n_1} \times R^{n_2}$, then the conclusion of Theorem 2 holds too. If we let $h_1(u_1, u_2) = h(u_1, u_2) - A$, we have $h_1(u_1, u_2) \geq 0$ on $R^{n_1} \times R^{n_2}$. It is clear that the problem

$$
\begin{align}
\text{(BCP)}_1 \qquad \min \quad & h_1(u_1, u_2) \\
\text{s.t.} \quad & g_i(u_1, u_2) \leq 0, \ i = 1, 2, \cdots, m,
\end{align}
$$

is equal to the problem (BCP). Particularly, when $p = 1$, we have the following conclusion.

Theorem 3. *Let $(u_1^*, u_2^*) \in U$ be a partial optimum of (BCP). If (u_1^*, u_2^*) is a KKT partial point of (BCP), i.e., there are $\alpha_i^{(1)}$ and $\alpha_i^{(2)} (i = 1, 2, \cdots, m)$ such that (5), (6), and (7) are true, then $H_1(u_1, u_2, \rho)$ is a partially exact penalty function with $p = 1$ for $\rho > \rho^*$, where:*

$$\rho^* = \max\{\alpha_i^{(1)}, \alpha_i^{(2)} | i = 1, 2, \cdots, m\}.$$

Theorem 3 is consistent with Theorem 2 in [28].

Similar to that for a constrained penalty function presented in [25,26], the concept of stability for a penalty function of (BCP) is defined. Let $s_1 = (s_1^1, s_2^1, \cdots, s_m^1)$ and:

$$U(u_2, s_1) = \{u_1 \in R^{n_1} \mid g_i(u_1, u_2) \leq s_i^1, \ i \in I\}.$$

When u_2 is fixed, define a perturbed problem:

$$(\text{BCP})(u_2, s_1) \qquad \min \quad h(u_1, u_2)$$
$$\text{s.t.} \quad u_1 \in U(u_2, s_1).$$

Let $s_2 = (s_1^2, s_2^2, \cdots, s_m^2)$ and:

$$U(u_1, s_2) = \{u_2 \in R^{n_2} \mid g_i(u_1, u_2) \leq s_i^2, \ i \in I\}.$$

When u_1 is fixed, define a perturbed problem:

$$(\text{BCP})(u_1, s_2) \qquad \min \quad h(u_1, u_2)$$
$$\text{s.t.} \quad u_2 \in U(u_1, s_2).$$

Definition 4. *Let (u_1^*, u_2^*) be a partial optimum of (BCP), and $u_{1s_1}^*$ and $u_{2s_2}^*$ be optimal solutions to $(\text{BCP})(u_2^*, s_1)$ and $(\text{BCP})(u_1^*, s_2)$, respectively, for any $s_1, s_2 \in R^m$. If there is a $\rho' > 0$ such that:*

$$h(u_1^*, u_2^*) - h(u_{1s_1}^*, u_2^*) \leq \rho|s_1|, \quad \forall \rho > \rho', \tag{16}$$
$$h(u_1^*, u_2^*) - h(u_1^*, u_{2s_2}^*) \leq \rho|s_2|, \quad \forall \rho > \rho', \tag{17}$$

where $|s_1| = \sum\limits_{i=1}^{m} P(s_i^1)$ and $|s_2| = \sum\limits_{i=1}^{m} P(s_i^2)$, then $H_p(u_1, u_2, \rho)$ is partially stable at (u_1^*, u_2^*). Furthermore, if there are a $\gamma > 0$ and a $\rho' > 0$ such that (16) and (17) hold for $|s_1| \leq \gamma$ and $|s_2| \leq \gamma$, then $H_p(u_1, u_2, \rho)$ is partially locally stable at (u_1^*, u_2^*).

Theorem 4. *Let (u_1^*, u_2^*) be a partial optimum of (BCP). If $H_p(u_1, u_2, \rho)$ is partially stable, then $H_p(u_1, u_2, \rho)$ is a partially exact penalty function at (u_1^*, u_2^*).*

Proof. Let us prove that $H_p(u_1, u_2, \rho)$ is a partially exact penalty function when $H_p(u_1, u_2, \rho)$ is partially stable at (u_1^*, u_2^*). Suppose that $H_p(u_1, u_2, \rho)$ is not a partially exact penalty function. According to the definition of partial stability, for any s_1, s_2, we obtain that there is a ρ' satisfying that:

$$h(u_1^*, u_2^*) - h(u_{1s_1}^*, u_2^*) \leq \rho|s_1|, \forall \rho > \rho', \tag{18}$$
$$h(u_1^*, u_2^*) - h(u_2^*, u_{2s_2}^*) \leq \rho|s_2|, \forall \rho > \rho'. \tag{19}$$

Then, there always exists some $\rho > \rho'$ such that (u_1^*, u_2^*) is not a partial optimum of $H_p(u_1, u_2, \rho)$, i.e., there is some (u_1', u_2') such that:

$$H_p(u_1', u_2^*, \rho) < H_p(u_1^*, u_2^*, \rho) = h(u_1^*, u_2^*),$$

$$H_p(u_1^*, u_2', \rho) < H_p(u_1^*, u_2^*, \rho) = h(u_1^*, u_2^*).$$

Thus,

$$h(u_1', u_2^*) + \rho \sum_{i \in I} P(g_i(u_1', u_2^*)) < h(u_1^*, u_2^*),$$

$$h(u_1^*, u_2') + \rho \sum_{i \in I} P(g_i(u_1^*, u_2')) < h(u_1^*, u_2^*).$$

Suppose that $u_1' \in \mathbf{U}(u_2^*), u_2' \in \mathbf{U}(u_1^*)$. We have:

$$h(u_1', u_2^*) < h(u_1^*, u_2^*),$$

$$h(u_1^*, u_2') < h(u_1^*, u_2^*).$$

This implies that $h(u_1', u_2^*) < h(u_1^*, u_2^*) < h(u_1', u_2^*)$ and $h(u_1^*, u_2') < h(u_1^*, u_2^*) < h(u_1^*, u_2')$, which shows that (u_1^*, u_2^*) is not a partial optimum of (BCP). A contradiction occurs. Hence, $u_1' \in \mathbf{U}(u_2^*)$ and $u_2' \in \mathbf{U}(u_1^*)$ do not hold, and $\sum_{i \in I} P(g_i(u_1', u_2^*)) > 0$ or $\sum_{i \in I} P(g_i(u_1^*, u_2')) > 0$.

Let $s_1' = (s_1^{1'}, s_2^{1'}, \ldots, s_m^{1'})^\top$ with $s_i^{1'} = g_i(u_1', u_2^*)$ and $s_{2'} = (s_1^{2'}, s_2^{2'}, \ldots, s_m^{2'})^\top$ with $s_i^2 = g_i(u_1^*, u_2'), i = 1, 2, \cdots, m$, and $u_{1s_1'}^*$ and $u_{2s_2'}^*$ be optimal solutions to $(BCP)(u_2^*, s_1')$ and $(BCP)(u_1^*, s_2')$, respectively. Then, $h(u_{1s_1'}^*, u_2^*) \le h(u_1', u_2^*)$ and $h(u_1^*, u_{2s_2'}^*) \le h(u_1^*, u_2')$. Thus,

$$h(u_{1s_1'}^*, u_2^*) \le h(u_1', u_2^*),$$

$$h(u_1^*, u_{2s_2'}^*) \le h(u_1^*, x_2').$$

Therefore,

$$\begin{aligned} h(u_{1s_1'}^*, u_2^*) + \rho \sum_{i \in I} P(s_i^{1'}) &\le h(u_1', u_2^*) + \rho \sum_{i \in I} P(s_i^{1'}) \\ &= H_p(u_1', u_2^*, \rho) < h(u_1^*, u_2^*), \end{aligned}$$

and:

$$\begin{aligned} h(u_1^*, u_{2s_2'}^*) + \rho \sum_{i \in I} P(s_i^{2'}) &\le h(u_1^*, u_2') + \rho \sum_{i \in I} P(s_i^{2'}) \\ &= H_p(u_1^*, u_2', \rho) < h(u_1^*, u_2^*), \end{aligned}$$

which shows that:

$$h(u_1^*, u_2^*) - h(u_1^*, u_{2s_2'}^*) > \rho|s_1'|,$$

$$h(u_1^*, u_2^*) - h(u_{1s_1'}^*, u_2^*) > \rho|s_2'|,$$

where $|s_1'| = \sum_{i \in I} P(s_i^{1'})$ and $|s_2'| = \sum_{i \in I} P(s_i^{2'})$. These inequalities contradict (18) and (19). Hence, that $H_p(x, y, \rho)$ is not partially stable yields a contradiction with the assumption, which proves that $H_p(x, y, \rho)$ is a partially exact penalty function. $\square$

Theorem 5. *Let (u_1^*, u_2^*) be a partial optimum of (BCP). If $H_p(u_1, u_2, \rho)$ is a partially exact penalty function at (u_1^*, u_2^*), then $H_p(u_1, u_2, \rho)$ is partially locally stable. In particular, for $p = 1$, $H_1(u_1, u_2, \rho)$ is partially stable.*

Proof. Let us prove that $H_p(x, y, \rho)$ is partially locally stable when $H_p(x, y, \rho)$ is a partially exact penalty function. According to the definition of a partially exact penalty function, if (u_1^*, u_2^*) is a partial optimum of (BCP), there always exist some $\rho > \rho'$ such that:

$$H_p(u_1^*, u_2^*, \rho) \le H_p(u_1, u_2^*, \rho), \forall u_1 \in R^{n_1}, \tag{20}$$

$$H_p(u_1^*, u_2^*, \rho) \le H_p(u_1^*, u_2, \rho), \forall u_2 \in R^{n_2}. \tag{21}$$

Let $u_{1s_1}^*$ and $u_{2s_2}^*$ be optimal solutions to (BCP)(u_2^*, s_1) and (BCP)(u_1^*, s_2), respectively, for any $s_1, s_2 \in R^m$. By (20), and (21), we have:

$$h(u_1^*, u_2^*) = H_p(u_1^*, u_2^*, \rho) \leq H_p(u_{1s_1}^*, u_2^*, \rho), \tag{22}$$

$$h(u_1^*, u_2^*) = H_p(u_1^*, u_2^*, \rho) \leq H_p(u_1^*, u_{2s_2}^*, \rho). \tag{23}$$

It is clear that for $s_i^1 \leq 1, s_i^2 \leq 1 (i \in I)$, we have:

$$P(g_i(u_{1s_1}^*, u_2^*)) \leq P(s_i^1), \ P(g_i(u_1^*, u_{2s_2}^*)) \leq P(s_i^2), \ i \in I. \tag{24}$$

From (22), (23), and (24), for $|s_1| \leq m$ and $|s_2| \leq m$:

$$h(u_1^*, u_2^*) - h(u_{1s_1}^*, u_2^*) \leq \rho|s_1|, \quad \forall \rho > \rho',$$

$$h(u_1^*, u_2^*) - h(u_1^*, u_{2s_2}^*) \leq \rho|s_2|, \quad \forall \rho > \rho'.$$

It follows from the definition that $H_p(x, y, \rho)$ is partially locally stable. □

Theorem 4 and Theorem 5 mean that the stability condition is sufficient for the partial exact penalty function, but the necessary condition of the partial exact penalty function is partially locally stable, as shown in the following example.

Example 4. *Let the perturbed problem (BCP3), and its (BCP3(x, s_1)) and (BCP3(y, s_2)):*

$$
\begin{aligned}
(BCP3)(x, s_1) \qquad & \min_{y} \quad h(x, y) = xy \\
& s.t. \quad 1 - xy \leq s_1, \\
(BCP3)(y, s_2) \qquad & \min_{x} \quad h(x, y) = xy \\
& s.t. \quad 1 - xy \leq s_2.
\end{aligned}
$$

It is clear that $(x^*, y^*) = (1, 1)$ is a partial optimum of (MCP3) and its objective function value is 1. $x_{s_1}^* = 1 - s_1$ and $y_{s_2}^* = 1 - s_2$ are optimal solutions to (BCP3)(1, s_1) and (BCP3)(1, s_2), respectively, for any $s_1, s_2 \in R$. Let $P(t) = \max\{t, 0\}^p$ and $s_1, s_2 \leq 1$ with $0 < p < 1$. For $\rho > 1$,

$$h(x^*, y^*) - h(x^*, y_{s_1}^*) = 1 - (1 - s_1) \leq \rho P(s_1),$$

$$h(x^*, y^*) - h(x_{s_1}^*, y^*) = 1 - (1 - s_2) \leq \rho P(s_2),$$

then $H_p(x, y, \rho)$ is partially locally stable. When $p = 1$, $H_1(x, y, \rho)$ is partially stable.

3. Partial Optimum Penalty Function Algorithm for (BCP)

Now, we present an algorithm to solve a partial optimum of (BCP) by solving the penalty function problem BCP(ρ) as follows.

$$\text{BCP}(\rho) \quad \min \quad H_p(u_1, u_2, \rho) \ s.t. (u_1, u_2) \in R^{n_1} \times R^{n_2}.$$

Based on the above results, the Partial Optimum Penalty Function Algorithm (Algorithm 1) was designed to compute a partial optimum of (BCP).

We prove the convergence of the Algorithm 1 in Theorem 6. Let:

$$S(L, h) = \{(u_1, u_2) \mid L \geq h(u_1, u_2), \}$$

be a level set. If for any given $L > 0$, $S(L, h)$ is bounded, then $S(L, h)$ is called bounded.

Algorithm 1: POPFA Algorithm

Step 1: Choose $\rho_1 > 0, N > 1, 0 < p \leq 1$ and $k = 1$.
Step 2: Solve (u_1^k, u_2^k) to be a partial optimum of

$$\min H_p(u_1, u_2, \rho_k) \ \ s.t.(u_1, u_2) \in R^{n_1} \times R^{n_2}.$$

Step 3: If (u_1^k, u_2^k) is feasible to (BCP), stop and (u_1^k, u_2^k) is a partial optimum of (BCP).
Otherwise, $\rho_{k+1} = N\rho_k, k =: k + 1$ and go to Step 2.

Theorem 6. *Let $\{(u_1^k, u_2^k)\}$ be the sequence generated by the Algorithm 1 and $h(\cdot)$ be continuous.*
(i) If $\{(u_1^k, u_2^k)\}(k = 1, 2, \cdots, \bar{k})$ is a finite sequence (i.e., the Algorithm 1 stops at the $\bar{k}$-th iteration), then $(u_1^{\bar{k}}, u_2^{\bar{k}})$ is a partial optimum of (BCP).
(ii) Let $\{(u_1^k, u_2^k)\}$ be an infinite sequence, sequence $\{H_p(u_1^k, u_2^k, \rho_k)\}$ be bounded and the level set $S(L, h)$ be bounded. Then $\{(u_1^k, u_2^k)\}$ is bounded and any limit point (u_1^*, u_2^*) of it is a partial optimum of (BCP).

Proof. (i) The conclusion is clear.
(ii) By the Algorithm 1, since $\{H_p(u_1^k, u_2^k, \rho_k)\}$ is bounded as $k \to +\infty$, there must be some $L > 0$ such that:

$$
\begin{aligned}
L &> H_p(u_1^k, u_2^k, \rho_k) \\
&\geq h(u_1^k, u_2^k) + \rho_k \sum_{i=1}^{m} P(g_i(u_1^k, u_2^k)) \qquad (25) \\
&\geq h(u_1^k, u_2^k).
\end{aligned}
$$

Since h is continuous and the level set $S(L, h)$ is bounded and closed, $\{(u_1^k, u_2^k)\}$ and $\{h(u_1^k, u_2^k)\}$ are bounded. Without loss of generality, suppose $(u_1^k, u_2^k) \to (u_1^*, u_2^*)$. Hence, there is an $A > 0$ such that $h(u_1^k, u_2^k) > -A$. From (25), we have:

$$\sum_{i=1}^{m} P(g_i(u_1^k, u_2^k)) \leq \frac{1}{\rho_k}(L - h(u_1^k, u_2^k)) < \frac{L + A}{\rho_k}.$$

We have $\sum_{i=1}^{m} P(g_i(u_1^k, u_2^k)) \to 0$ as $\rho_k \to +\infty$. Hence, (u_1^*, u_2^*) is a feasible solution to (BCP).
Let any $\bar{u}_1 \in U(u_2^*)$ and $\bar{u}_2 \in U(u_1^*)$. Since (u_1^k, u_2^k) is a partial optimum of

$$\min H_p(u_1, u_2, \rho_k) \ \ s.t.(u_1, u_2) \in R^{n_1} \times R^{n_2},$$

we have:

$$h(u_1^k, u_2^k) \leq H_p(u_1^k, u_2^k, \rho_k) \leq H_p(\bar{u}_1, u_2^k, \rho_k),$$

$$h(u_1^k, u_2^k) \leq H_p(u_1^k, u_2^k, \rho_k) \leq H_p(u_1^k, \bar{u}_2, \rho_k).$$

Let $k \to +\infty$, and the above inequations are:

$$h(u_1^*, u_2^*) \leq h(\bar{u}_1, u_2^*),$$

$$h(u_1^*, u_2^*) \leq h(u_1^*, \bar{u}_2).$$

Hence, (u_1^*, u_2^*) is a partial optimum of (BCP). $\square$

Theorem 6 means that the Algorithm 1 has good convergence in theory. From Theorem 4, when the penalty function is partially stable, the Algorithm 1 solves a single unconstrained optimization problem for the smaller penalty parameter ρ. Since the penalty function $H_p(u_1, u_2, \rho)$ is nonsmooth for $0 < p < 1$, it is necessary to smooth the constrained

nonsmooth term to design an effective algorithm. Therefore, the smoothing algorithm of a partial exact penalty function is worthy of further study. In fact, for $p = 1$, we have published a paper regarding a smoothing partially exact penalty function algorithm [30], where two numerical examples show the proposed algorithm is effective for biconvex programming.

4. Conclusions

In this paper, we studied the partial optimum solution to biconvex programming using the penalty function, which is partially exact. The form of this penalty function is more general than that in [28]. We proved that the partial exactness of the penalty function for biconvex programming is equivalent to the partial KKT condition, and we proved that the partial exactness of the penalty function for biconvex programming is equivalent to partially local stability. Based on the penalty function, the Algorithm 1 was theoretically presented to solve a partial optimum solution to biconvex programming. The convergence of the algorithm was also proven. The Algorithm 1 may solve a partial optimum solution to biconvex programming under the smaller penalty parameter. In the future, we may study the smoothing problem of the partial exact penalty function and its algorithm.

References

1. Shen, X.; Diamond, S.; Udell, M.; Gu, Y. Disciplined multi-convex programming. In Proceedings of the Control And Decision Conference (CCDC), Chongqing, China, 28–30 May 2017.
2. Grant, M.; Boyd, S.; Ye, Y. Disciplined convex programming. In *Global Optimization: From Theory to Implementation, Nonconvex Optimization and its Applications*; Liberti, L., Maculan, N., Eds.; Springer: New York, NY, USA, 2006; pp. 155–210.
3. Chiu, W.Y. Method of reduction of variables for bilinear matrix inequality problems in system and control designs. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 1241–1256. [CrossRef]
4. Ichihara, H.; Nobuyama, E. Difference of multiconvex relaxation of parameterized LMIs: Control applications. In Proceedings of the SICE 2003 Annual Conference, Fukui, Japan, 4–6 August 2003.
5. Hours, J.; Jones, C. A parametric multiconvex splitting technique with application to real-time NMPC. In Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, USA, 15–17 December 2014.
6. Suh, S.; Shin, S.; Lee, J.; Reddy, C.K.; Choo, J. Localized user-driven topic discovery via boosted ensemble of nonnegative matrix factorization. *Knowl. Inf. Syst.* **2018**, *56*, 503–531. [CrossRef]
7. Udell, M.; Horn, C.; Zadeh, R.; Boyd, S. Generalized low rank models. *Found. Trends Mach. Learn.* **2016**, *9*, 1–118. [CrossRef]
8. Fu, X.; Huang, K.; Sidiropoulos, N.D. On identifiability of nonnegative matrix factorization. *IEEE Signal Process. Lett.* **2018**, *25*, 328–332. [CrossRef]
9. Hong, M.; Luo, Z. On the linear convergence of the alternating direction method of multipliers. *Math. Program.* **2017**, *162*, 165–199. [CrossRef]
10. Serbetli, S.; Yener, A. Transceiver optimization for multiuser MIMO systems. *IEEE Trans. Signal Process.* **2004**, *52*, 214–226. [CrossRef]
11. Al-Shatri, H.; Li, X.; Ganesa, R.S.S. Maximizing the Sum Rate in Cellular Networks Using Multiconvex Optimization. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 3199–3211. [CrossRef]
12. Wen, Z.; Yin, W.; Zhang, Y. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Math. Program. Comput.* **2012**, *4*, 333–361. [CrossRef]
13. Xu, Y.; Yin, W. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *Siam J. Imaging Sci.* **2013**, *6*, 1758–1789. [CrossRef]
14. Gorski, J.; Pfeuffer, F.; Klamroth, K. Biconvex sets and optimization with biconvex functions: A survey and extensions. *Math. Methods Oper. Res.* **2007**, *66*, 373–407. [CrossRef]

15. Liang, X.; Bai, J. Preconditioned ADMM for a class of bilinear programming problems. *Math. Probl. Eng.* **2018**, *2018*, 1–9. [CrossRef]
16. Hajinezhad, D.; Shi, Q.J. Alternating direction method of multipliers for a class of nonconvex bilinear optimization: Convergence analysis and applications. *J. Glob. Optim.* **2018**, *70*, 261–288. [CrossRef]
17. Charkhgard, H.; Savelsbergh, M.; Talebian, M. A linear programming based algorithm to solve a class of optimization problems with a multi-linear objective function and affine constraints. *Comput. Oper. Res.* **2018**, *89*, 17–30. [CrossRef]
18. Li, G.; Wen, C.; Zheng, W.X.; Zhao, G. Iterative identification of block-oriented nonlinear systems based on biconvex optimization. *Syst. Control. Lett.* **2015**, *79*, 68–75. [CrossRef]
19. Shah, S.; Yadav, A.K.; Castillo, C.D.; Jacobs, D.W.; Studer, C.; Goldstein, T. Biconvex Relaxation for Semidefinite Programming in Computer Vision. In Proceedings of the European Conference on Computer Vision: Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.
20. Al-Khayyaltt, F.A.; Falk, J.E. Jointly constrained biconvex programming. *Math. Oper. Res.* **1983**, *8*, 273–286. [CrossRef]
21. Pardalos, P.M.; Resende, M.G.C. *Handbook of Applied optimization*; Oxford University Press: Oxford, UK, 2002.
22. Zangwill, W.I. Nonlinear programming via penalty function. *Manang. Sci.* **1967**, *13*, 334–358 [CrossRef]
23. Rosenberg, E. Globally convergent algorithms for convex programming. *Math. Oper. Res.* **1981**, *6*, 437–443. [CrossRef]
24. Di Pillo, G.; Grippo, L. An exact penalty function method with global conergence properties for nonlinear programming problems. *Math. Program.* **1981**, *36*, 1–18. [CrossRef]
25. Clarke, F.H. *Optimization and Nonsmooth Analysis*; Wiley: New York, NY, USA, 1983.
26. Burke, J.V. Calmness and exact penalization. *Siam J. Control. Optim.* **1991**, *29*, 493–497. [CrossRef]
27. Strongin, R.G.; Sergeyev, Y.D. *Global Optimization with Non-Convex Constraints*; Science & Business: Dordrecht, The Netherlands, 2000.
28. Grishagin, V.; Israfilov, R. Multidimensional Constrained Global Optimization in Domains with Computable Boundaries. Available online: https://elar.urfu.ru/bitstream/10995/36002/1/ural_pdc-2015-09.pdf (accessed on 16 January 2021).
29. Yang, X.Q.; Meng, Z.Q. Lagrange Multipliers and Calmness Conditions of Order p. *Math. Oper. Res.* **2007**, *32*, 95–101. [CrossRef]
30. Shen, R.; Meng, Z.Q.; Jiang, M. Smoothing Partially Exact Penalty Function of Biconvex Programming. *Asia-Pac. J. Oper. Res.* **2020**, *37*, 27. [CrossRef]