Biconvex Sets and Optimization with Biconvex Functions - A Survey and Extensions

Jochen Gorski*, Frank Pfeuffer, Kathrin Klamroth*

Institute for Applied Mathematics

Friedrich-Alexander-University Erlangen-Nuremberg

gorski@am.uni-erlangen.de, pfeuffer@am.uni-erlangen.de, klamroth@am.uni-erlangen.de

May 2007

Abstract

The problem of optimizing a biconvex function over a given (bi)convex or compact set frequently occurs in theory as well as in industrial applications, for example, in the field of multifacility location or medical image registration. Thereby, a function $f : X \times Y \to \mathbb{R}$ is called biconvex, if $f(x, y)$ is convex in y for fixed $x \in X$, and $f(x, y)$ is convex in x for fixed $y \in Y$. This paper presents a survey of existing results concerning the theory of biconvex sets and biconvex functions and gives some extensions. In particular, we focus on biconvex minimization problems and survey methods and algorithms for the constrained as well as for the unconstrained case. Furthermore, we state new theoretical results for the maximum of a biconvex function over biconvex sets.

Key Words: biconvex functions, biconvex sets, biconvex optimization, biconcave optimization, non-convex optimization, generalized convexity

1 Introduction

In practice, biconvex optimization problems frequently occur in industrial applications, for example, in the field of multifacility location or medical image registration. We review theoretical results for biconvex sets and biconvex functions and survey existing methods and results for general biconvex optimization problems.

We recall that a set $S \subseteq \mathbb{R}^k$ is said to be convex if for any two points $s_1, s_2 \in S$ the line segment joining s_1 and s_2 is completely contained in S. A function $f : S \to \mathbb{R}$ on a convex set S is called convex, if

$$f(\lambda s_1 + (1 - \lambda)s_2) \leq \lambda f(s_1) + (1 - \lambda) f(s_2)$$

is valid for all $\lambda \in [0, 1]$ and $s_1, s_2 \in S$.

*The authors were partially supported by a grant of the German Research Foundation (DFG)

For the definition of biconvex sets and biconvex functions, let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty, convex sets, and let $B \subseteq X \times Y$. We define x- and y-sections of B as follows:

$$\begin{aligned} B_x &:= \{y \in Y : (x, y) \in B\}, \\ B_y &:= \{x \in X : (x, y) \in B\}. \end{aligned}$$

Definition 1.1 *The set $B \subseteq X \times Y$ is called a* **biconvex set** *on $X \times Y$ or* **biconvex** *for short, if B_x is convex for every $x \in X$ and B_y is convex for every $y \in Y$.*

The most important results on biconvex sets are summarized in Section 2.

Definition 1.2 *A function $f : B \to \mathbb{R}$ on a biconvex set $B \subseteq X \times Y$ is called a* **biconvex function** *on B or* **biconvex** *for short, if*

$$f_x(\bullet) := f(x, \bullet) : B_x \to \mathbb{R}$$

is a convex function on B_x for every fixed $x \in X$ and

$$f_y(\bullet) := f(\bullet, y) : B_y \to \mathbb{R}$$

is a convex function on B_y for every fixed $y \in Y$.

From this definition, the definitions of **biconcave**, **bilinear** and **biaffine functions** are obtained by replacing the property of being convex for f_x and f_y by the property of being concave, linear, or affine, respectively. Since for a biconvex function $f : B \to \mathbb{R}$, the function $g := -f$ is biconcave on B, i.e., $g(x, y)$ is concave on B_x in y for fixed $x \in X$ and $g(x, y)$ is concave on B_y in x for fixed $y \in Y$, most of the results and methods mentioned in this paper can directly be transferred to the biconcave case, too.
In the first part of Section 3, we survey general properties of biconvex functions, like arithmetical properties or results on the continuity of such functions, which mostly result from the convex substructures of a biconvex function. In the second part, we discuss results on biconvex maximization problems and show that a biconvex function which attains its maximum in the relative interior of a given biconvex set B must be constant throughout B, assuming rather weak topological properties on B. Furthermore, we survey separation theorems for biconvex functions which are mostly applied in probability theory.

Definition 1.3 *An optimization problem of the form*

$$\min \{f(x, y) : (x, y) \in B\} \tag{1}$$

is said to be a **biconvex optimization problem** *or* **biconvex** *for short, if the feasible set B is biconvex on $X \times Y$, and the objective function f is biconvex on B.*

Different from convex optimization problems, biconvex problems are in general global optimization problems which may have a large number of local minima. However, the question arises whether the convex substructures of a biconvex optimization problem can be utilized more efficiently for the solution of such problems than in the case of general non-convex optimization problems. For this purpose, we discuss existing methods and algorithms, specially designed for biconvex minimization problems which primarily exploit the convex substructures of the problem and give examples for practical applications in

Section 4. We only briefly mention bilinear problems as there exist plenty of literature and methods, see, e.g., Horst and Tuy (1990) for a survey. We rather concentrate on minimization methods and algorithms which can be applied to general constrained as well as unconstrained biconvex minimization problems. In particular, we review the *Alternate Convex Search* method, stated, e.g., in Wendell and Hurter Jr. (1976), the *Global Optimization Algorithm*, developed by Floudas and Visweswaran (1990), and an algorithm for a special class of jointly constrained biconvex programming problems, given in Al-Khayyal and Falk (1983). Note that the above mentioned methods and algorithms can be and are applied to bilinear problems in practice, too (cf. Visweswaran and Floudas, 1993).

2 Biconvex Sets

The goal of this section is to recall the main definitions and results obtained for biconvex sets. Only a few papers exist in the literature where biconvex sets are investigated. The results presented here can be found in the papers of Aumann and Hart (1986) and Goh et al. (1994). In addition we give a short comparison between convex and biconvex sets.

2.1 Elementary Properties

In this first subsection we recall elementary properties of biconvex sets. We start with a characterization.

Theorem 2.1 (Aumann and Hart (1986)) *A set $B \subseteq X \times Y$ is biconvex if and only if for all quadruples (x_1, y_1), (x_1, y_2), (x_2, y_1), $(x_2, y_2) \in B$ it holds that for every $(\lambda, \mu) \in [0,1] \times [0,1]$*

$$(x_\lambda, y_\mu) := ((1-\lambda)x_1 + \lambda x_2, (1-\mu)y_1 + \mu y_2) \in B.$$

Obviously, a biconvex set is not convex in general. As an example we consider the letters "L" or "T" as a subset of $\mathbb{R} \times \mathbb{R}$, which are biconvex but not convex. Even worse, a biconvex set does not have to be connected in general, as the example

$$B^2 = \{(x,y) \in \mathbb{R}^2 : x, y > 0\} \cup \{(x,y) \in \mathbb{R}^2 : (-x), (-y) > 0\} \tag{2}$$

shows (see Figure 1). If in contrast B is convex, we derive the following result, which can easily be proven:

Theorem 2.2 *Let $k > 1$, let $B \subset \mathbb{R}^k$ be a convex set, and let (V_1, V_2) be an arbitrary partition of the variable set $V := \{x_1, \ldots, x_k\}$ into two non-empty subsets. Then B is biconvex on* $\operatorname{span}(V_1) \times \operatorname{span}(V_2)$, *where* $\operatorname{span}(V_i)$ *denotes the linear space generated by V_i $(i = 1, 2)$.*

The converse of the last theorem is obviously false. For a counter-example in $\mathbb{R}^2$ consider again the letters "L" or "T". For a more general counter-example in $\mathbb{R}^n$, we generalize the set B given in (2).

Example 2.1 Let $k \geq 2$, and let the set $B \subset \mathbb{R}^k$ be given by

$$B = \{z \in \mathbb{R}^k : z_i > 0, i = 1, \ldots, k\} \cup \{z \in \mathbb{R}^k : z_i < 0, i = 1, \ldots, k\}.$$

Since B is not connected, it cannot be convex. Now let (V_1, V_2) be an arbitrary, but fixed partition of the variable set $V := \{x_1, \ldots, x_k\}$ into two non-empty subsets. The given

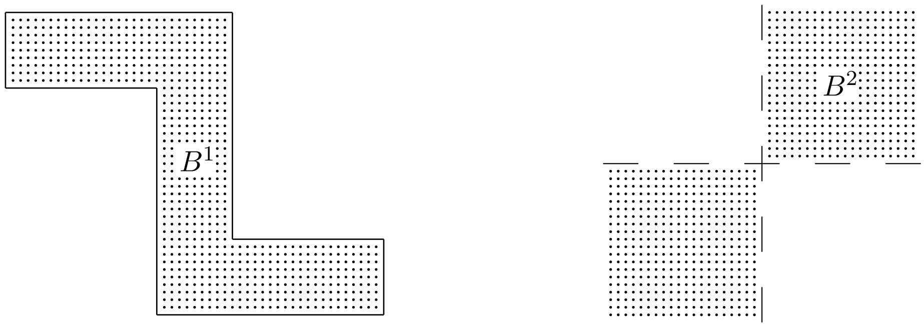

Figure 1: *Examples of biconvex sets which are non-convex (B^1) and non-convex and non-connected (B^2), respectively.*

set B is symmetric in all variables, thus we can rearrange the variables such that we can suppose without loss of generality that the partition of V is given by $V_1 = \{x_1, \ldots, x_\nu\}$ and $V_2 = \{x_{\nu+1}, \ldots, x_k\}$ with $1 \le \nu \le k-1$, i.e., $X := \text{span}(V_1) = \mathbb{R}^\nu$ and $Y := \text{span}(V_2) = \mathbb{R}^{k-\nu}$. Now choose $\hat{x} \in X$ arbitrary, but fixed. Then

$$B_{\hat{x}} = \begin{cases} \{y \in Y : y_j > 0,\, j = 1 \ldots, k-\nu\} & : \hat{x}_i > 0,\, i = 1, \ldots, \nu. \\ \emptyset & : \exists\, i, j \in \{1 \ldots, \nu\}, i \ne j : x_i \cdot x_j \le 0 \\ \{y \in Y : y_j < 0,\, j = 1 \ldots, k-\nu\} & : \hat{x}_i < 0,\, i = 1, \ldots, \nu. \end{cases}$$

Obviously, in all the three cases, $B_{\hat{x}}$ is convex. Similarly, it can be shown that $B_{\hat{y}}$ is convex for every fixed $\hat{y} \in Y$. Hence, B is biconvex for the chosen partitioning of V.

2.2 Biconvex Combinations and the Biconvex Hull

In convex analysis, the concept of convex combinations of k given points in $\mathbb{R}^n$ and their convex hull is well known and straight forward (see, e.g., Rockafellar, 1997, Section 2). In Aumann and Hart (1986) the concept of biconvex combinations as a special case of a convex combination of k given points is introduced and investigated. We recall the main ideas and results here.

Definition 2.1 *Let $(x_i, y_i) \in X \times Y$ for $i = 1, \ldots, k$. A convex combination*

$$(x, y) = \sum_{i=1}^{k} \lambda_i (x_i, y_i),$$

(with $\sum_{i=1}^{k} \lambda_i = 1$, $\lambda_i \ge 0$ for $i = 1, \ldots, k$) is called **biconvex combination** *or* **biconvex** *for short, if $x_1 = \cdots = x_m = x$ or $y_1 = \cdots = y_k = y$ holds.*

With the help of biconvex combinations another characterization for biconvex sets can be formulated:

Theorem 2.3 (Aumann and Hart (1986)) *A set $B \subseteq X \times Y$ is biconvex if and only if B contains all biconvex combinations of its elements.*

As in the convex case, it is possible to define the *biconvex hull* of a given set $A \subseteq X \times Y$. To do this, we proceed as in the convex case and denote by H the intersection of all biconvex sets that contain A.

Definition 2.2 *Let $A \subseteq X \times Y$ be a given set. The set*

$$H := \bigcap \{A_I : A \subseteq A_I,\ A_I \text{ is biconvex}\}$$

is called **biconvex hull** *of A and is denoted by* biconv(A).

Theorem 2.4 (Aumann and Hart (1986)) *The above defined set H is biconvex. Furthermore, H is the smallest biconvex set (in the sense of set inclusion), which contains A.*

As biconvex combinations are, by definition, a special case of convex combinations and the convex hull conv(A) of a given set A consists of all convex combinations of the elements of A (see, e.g., Rockafellar, 1997, Theorem 2.3), we have:

Lemma 2.5 *Let $A \subseteq X \times Y$ be a given set. Then*

$$\operatorname{biconv}(A) \subseteq \operatorname{conv}(A)$$

Aumann and Hart proposed in their paper another way to construct the biconvex hull of a given set A. They defined an inductively given sequence $\{A_n\}_{n\in\mathbb{N}}$ as follows:

$$\begin{aligned} A_1 &:= A \\ A_{n+1} &:= \{(x,y) \in A_n : (x,y) \text{ is a biconvex combination of elements of } A_n\}. \end{aligned}$$

Let $H' := \bigcup\limits_{n=1}^{\infty} A_n$ denote the limit of this sequence.

Theorem 2.6 (Aumann and Hart (1986)) *The above constructed set H' is biconvex and equals H, the biconvex hull of A.*

It is important to mention that when applying the above procedure to the convex case (i.e., for the construction of the convex hull of A), one iteration is sufficient as the convex hull consists exactly of all convex combinations of its elements. In general, there does not necessarily exist a finite number of sets A_n such that the union of these sets build the biconvex hull of the given set A. To see this, consider the following example.

Example 2.2 (Aumann and Hart (1986)) Let $X = Y = [0,1]$. For $m \in \mathbb{N}$ we define

$$\begin{aligned} z_1 &= (0,0), & w_1 &= (0,0) \\ z_{2m} &= (1-\frac{1}{2^{m-1}}, 1-\frac{3}{2^{m+2}}), & w_{2m} &= (1-\frac{1}{2^{m-1}}, 1-\frac{1}{2^m}), \\ z_{2m+1} &= (1-\frac{3}{2^{m+2}}, 1-\frac{1}{2^m}), & w_{2m+1} &= (1-\frac{1}{2^m}, 1-\frac{1}{2^m}). \end{aligned}$$

For $n \geq 2$, w_n is a biconvex combination of the points z_n and w_{n-1}, namely

$$w_n = \frac{4}{5}\, z_n + \frac{1}{5}\, w_{n-1}.$$

Now, let the set A be given by $\{z_n\}_{n\in\mathbb{N}}$. Then it is easy to see that $w_n \in A_n$, but $w_n \notin A_{n-1}$ for every $n \geq 2$ (see also Figure 2).
By adding the point $(1,1)$ to the set A, we obtain a closed and bounded set A with $A_n \subsetneq \operatorname{biconv}(A)$ for all $n \in \mathbb{N}$.

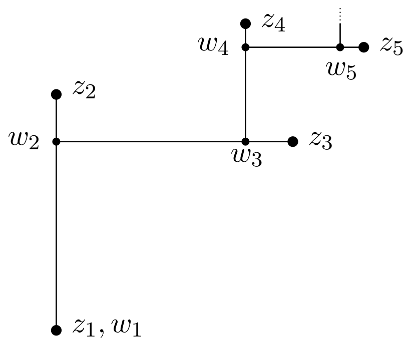

Figure 2: *Illustration of Example 2.2.*

3 Biconvex Functions

In this section we present important properties of biconvex functions. As these types of functions regularly appear in practice, biconvex functions and optimization problems are widely discussed in the literature. Since we are interested in optimization with biconvex functions f on subsets of $\mathbb{R}^{n+m}$ here, we focus on properties which are related to these optimization problems.

Note that biconvex functions are of importance in other mathematical contexts, too. For example, biconvex functions can be used to derive results on robust stability of control systems in practical control engineering. For further details see Geng and Huang (2000a) and Geng and Huang (2000b). Furthermore, biconvex functions play an important role in martingale theory and can be used to characterize whether a Banach space B is UMD (i.e., the space B has the unconditionality property for martingale differences), or whether B is a Hilbert space or not. Here, we refer to Burkholder (1981), Aumann and Hart (1986), Burkholder (1986) and Lee (1993). Finally, Thibault (1984), Jouak and Thibault (1985) and Borwein (1986) published results concerning the continuity and differentiability of (measurable) biconvex operators in topological vector spaces.

This section is organized as follows: The first subsection briefly reviews elementary properties of biconvex functions. We extend these properties by a comparison to convex functions. The next subsection summarizes results concerning continuity of biconvex functions given in Aumann and Hart (1986, Section 3). The last subsection deals with the maximization of biconvex functions. Several known and some new results are presented.

3.1 Elementary Properties

We start our summary with the most important elementary properties of biconvex functions. Note that as mentioned in Section 1, it is possible to transform a biconvex function to a biconcave one, and vice versa, by multiplying the given function by (-1). Similar to convex functions, biconvex functions can be characterized by an interpolation property:

Theorem 3.1 (Goh et al. (1994)) *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty, convex sets, and let f be a real-valued function on $X \times Y$. f is biconvex if and only if for all quadruples (x_1, y_1), (x_1, y_2), (x_2, y_1), $(x_2, y_2) \in X \times Y$ it holds, that for every $(\lambda, \mu) \in$*

$[0,1] \times [0,1]$

$$\begin{aligned} f(x_\lambda, y_\mu) \;\leq\; & (1-\lambda)\,(1-\mu)\,f(x_1,y_1) + (1-\lambda)\,\mu\,f(x_1,y_2) + \\ & +\lambda\,(1-\mu)\,f(x_2,y_1) + \lambda\,\mu\,f(x_2,y_2), \end{aligned}$$

where $(x_\lambda, y_\mu) := ((1-\lambda)x_1 + \lambda x_2, (1-\mu)y_1 + \mu y_2)$.

So, as one-dimensional interpolation always overestimates a convex function, two-dimensional interpolation always overestimates a biconvex function.
As convex functions have convex level sets, we state for the biconvex case:

Theorem 3.2 (Goh et al. (1994)) *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty, convex sets, and let f be a real-valued function on $X \times Y$. If f is biconvex on $X \times Y$, then its level sets*

$$\mathcal{L}_c := \{(x,y) \in X \times Y :\, f(x,y) \leq c\}$$

are biconvex for every $c \in \mathbb{R}$.

Like in the convex case, the converse of the last theorem is not true in general:

Example 3.1 Let the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $f(x,y) = x^3 + y^3$ be given, and let $c \in \mathbb{R}$. Then,

$$\begin{aligned} (\mathcal{L}_c)_{\bar{x}} &= \{y \in Y :\, y^3 \leq c - \bar{x}^3\} \; =]-\infty, \operatorname{sign}(c-\bar{x}^3)\,\sqrt[3]{|c-\bar{x}^3|}] \\ (\mathcal{L}_c)_{\bar{y}} &= \{x \in X :\, x^3 \leq c - \bar{y}^3\} \; =]-\infty, \operatorname{sign}(c-\bar{y}^3)\,\sqrt[3]{|c-\bar{y}^3|}] \end{aligned}$$

are convex sets and hence, the level set $\mathcal{L}_c$ of f is biconvex. But obviously f is not biconvex on $\mathbb{R} \times \mathbb{R}$, since $f_0(x) = f(x,0) = x^3$ is not a convex function on $\mathbb{R}$.

Also many arithmetic properties that are valid for convex functions can be transferred to the biconvex case.

Lemma 3.3 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty, convex sets, let $\mu \in \mathbb{R}^+$ be a non-negative scalar, and let $f, g : X \times Y \to \mathbb{R}$ be two biconvex functions. Then the functions $h, t : X \times Y \to \mathbb{R}$ with $h(x,y) := f(x,y) + g(x,y)$ and $t(x,y) := \mu f(x,y)$ are biconvex.*

Proof: The biconvexity of h and t follows immediately from Definition 1.2 for biconvex functions and the fact that the above stated lemma is valid for convex functions, hence for f_x and g_x (f_y and g_y, respectively), too, as they are convex for every fixed $x \in X$ ($y \in Y$) by definition. $\square$

For the composition of convex and biconvex functions we have:

Lemma 3.4 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty, convex sets, let $f : X \times Y \to \mathbb{R}$ be a biconvex function, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex, non-decreasing function. Then $h(x,y) := \varphi(f(x,y))$ is biconvex on $X \times Y$.*

Proof: For fixed $x \in X$ and fixed $y \in Y$ we consider the functions $h_x(y) := \varphi(f_x(y))$ and $h_y(x) := \varphi(f_y(x))$, respectively. Since Lemma 3.4 is valid for f convex (cf. Rockafellar, 1997, Theorem 5.1), f_x and f_y are both convex functions by definition and φ is convex and non-decreasing, h_x and h_y are convex, too. Hence, h is a biconvex function on $X \times Y$. □

Finally, we state a lemma concerning the pointwise supremum of biconvex functions.

Lemma 3.5 *The pointwise supremum of an arbitrary collection of biconvex functions is biconvex.*

Proof: Let I be an arbitrary index set, let $f^i : X \times Y \to \mathbb{R}$ be biconvex for all $i \in I$, and let $f(x, y) := \sup\{f^i(x, y),\ i \in I\}$ be the pointwise supremum of these functions. For fixed $\bar{y} \in Y$ and arbitrary $x \in X$ we have:

$$f_{\bar{y}}(x) \;=\; f(x, \bar{y}) \;=\; \sup_{i \in I}\{f^i(x, \bar{y})\} \;=\; \sup_{i \in I}\{f^i_{\bar{y}}(x)\}.$$

Since the functions $f^i_{\bar{y}}$ are convex for all $i \in I$ by assumption, $f_{\bar{y}}$, as pointwise supremum of convex functions, is convex by Rockafellar (1997, Theorem 5.5), too. Similarly it can be shown that $f_{\bar{x}}$ is convex on Y for every fixed $\bar{x} \in X$. Hence, f is biconvex. □

We close this subsection by a comparison between convex and biconvex functions. Obviously:

Theorem 3.6 *Let $f : \mathbb{R}^k \to \mathbb{R}$ ($k > 1$) be a convex function, and let (V_1, V_2) be an arbitrary partition of the variable set $V := \{x_1, \ldots, x_k\}$ into two non-empty subsets. Then f is biconvex on* $\operatorname{span}(V_1) \times \operatorname{span}(V_2)$.

As in the case of biconvex sets, if a given function f is biconvex for every arbitrary partition of the variable set, it has not to be convex in general. To see this, consider the following example:

Example 3.2 Let $n \geq 2$, let $b := \sqrt{\frac{2n-1}{2n(n-1)}}$, and let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be defined by

$$f(x_1, \ldots, x_{n+1}) \;=\; \frac{1}{2}(x_1^2 + \cdots + x_n^2 + x_{n+1}^2) + b \cdot x_{n+1} \cdot (x_1 + \cdots + x_n).$$

The partial derivatives of f are given by

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) \;=\; \begin{cases} x_i + b \cdot x_{n+1} & \text{, if } \quad i \neq n+1 \\ x_{n+1} + b \cdot (x_1 + \cdots + x_n) & \text{, if } \quad i = n+1 \end{cases}$$

and the Hessian matrix of f is

$$H(x) := \operatorname{Hess}(f)(x) = \begin{pmatrix} 1 & 0 & \ldots & 0 & b \\ 0 & 1 & \ldots & 0 & b \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & b \\ b & b & \ldots & b & 1 \end{pmatrix} \in M((n+1) \times (n+1), \mathbb{R}).$$

First, we show that f is biconvex for any partition of the variable set $\{x_1, \ldots, x_{n+1}\}$ into two non-empty disjoint subsets. So let (V_1, V_2) be such a partition and let $Y^i = \operatorname{span}(V_i)$,

$i = 1, 2$. We assume that $x_{n+1} \in V_2$. Let I^i denote the index set of the variables of V_i, and let $c_i := |I^i|$ be the cardinality of I^i $(i = 1, 2)$. Then the Hessian matrix of f_{Y^i} is given by

$$\text{Hess}(f_{Y^2})(y^1) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \text{ and } \text{Hess}(f_{Y^1})(y^2) = \begin{pmatrix} 1 & 0 & \dots & 0 & b \\ 0 & 1 & \dots & 0 & b \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & b \\ b & b & \dots & b & 1 \end{pmatrix},$$

where

$$\begin{aligned} y^1 &= (x_{i_1^1}, \dots, x_{i_{c_1}^1}) \in Y^1, & i_j^1 \in I^1 \ \forall\, j = 1, \dots, c_1, \\ y^2 &= (x_{i_1^2}, \dots, x_{i_{c_2-1}^2}, x_{n+1}) \in Y^2, & i_k^2 \in I^2 \ \forall\, k = 1, \dots, c_2 - 1, \end{aligned}$$

and $\text{Hess}(f_{Y^1}) \in M(c_2 \times c_2, \mathbb{R})$ and $\text{Hess}(f_{Y^2}) \in M(c_1 \times c_1, \mathbb{R})$. Obviously, $\text{Hess}(f_{Y^2})$ is positive definite for all $y^1 \in Y^1 = \mathbb{R}^{c_1}$ and hence, f_{Y^2} is convex (cf. Floudas, 2000).
To show the convexity of f_{Y^1}, we calculate the eigenvalues of $\text{Hess}(f_{Y^1})$. They are given by $\lambda_1 = 1$ and $\lambda_{2,3} = 1 \pm b\sqrt{c_2 - 1}$. Since $c_2 \leq n$, it holds that $\lambda_2 > 0$ and

$$\lambda_3 = 1 - b\sqrt{c_2 - 1} \geq 1 - b\sqrt{n-1} = 1 - \sqrt{1 - \frac{1}{2n}} > 0.$$

So, all eigenvalues of $\text{Hess}(f_{Y^1})$ are positive, i.e., $\text{Hess}(f_{Y^1})$ is positive definite, and hence, f_{Y^1} is convex, too.
Finally, we calculate the eigenvalues of $H(x)$. They are given by $\lambda_1 = 1$ and $\lambda_{2,3} = 1 \pm b\sqrt{n}$. Since

$$\lambda_3 = 1 - b\sqrt{n} = 1 - \sqrt{1 + \frac{1}{2\,(n-1)}} < 0,$$

$H(x)$ has a negative eigenvalue. Hence, $H(x)$ is indefinite for all $x \in \mathbb{R}^{n+1}$ and f is not convex on every open, convex set $X \subseteq \mathbb{R}^{n+1}$.
Note that for a counter-example for a function from $\mathbb{R}^2$ to $\mathbb{R}$, one can use the above given function with $b := 2$.

3.2 Continuity of Biconvex Functions

One of the central results in convex analysis is the fact that a finite, real-valued, convex function f is continuous throughout the interior of its domain $C \subseteq \mathbb{R}^n$ (cf. Rockafellar, 1997). Aumann and Hart (1986) transferred this result to the biconvex case.

Definition 3.1 *Let $B \subseteq X \times Y$ and let $z = (x, y) \in B$. The point z is called a* **bi-relatively interior point of** *B, if z is in the interior of B relative to* $\text{aff}(\text{proj}_X(B)) \times \text{aff}(\text{proj}_Y(B))$, *where* $\text{proj}_X(B)$ *and* $\text{proj}_Y(B)$ *denote the projection of B into the X- and Y-space, respectively, and* $\text{aff}(C)$ *is the affine space, generated by the set C.*

From Rockafellar (1997) we recall that an m**-dimensional simplex** is the convex hull of m affinely independent vectors $b_1 \dots, b_m \in \mathbb{R}^n$. A set $S \subseteq \mathbb{R}^n$ is called **locally simplicial**, if for each $x \in S$ there exists a finite collection of simplices $S_1, \dots, S_m$ such that, for some neighborhood U of x,

$$U \cap (S_1 \cup \dots \cup S_m) = U \cap S.$$

Examples of locally simplicial sets are line segments, polyhedral convex sets, or relatively open, convex sets. Note that a locally simplicial set does not need to be convex or closed in general.

Definition 3.2 *Let $B \subseteq X \times Y$ and let $z = (x, y) \in B$. We say that B is* **locally bi-simplicial at** *z, if there exists a neighborhood U of x in X and a neighborhood V of y in Y, a collection of simplices $S_1, \dots, S_k$ in X and a collection of simplices $T_1, \dots, T_l$ such that for $S := \bigcup_{i=1}^{k} S_i$ and $T := \bigcup_{i=1}^{l} T_i$, $S \times T \subseteq B$ and $(U \times V) \cap B = (U \times V) \cap (S \times T)$.*

It holds:

Theorem 3.7 (Aumann and Hart (1986)) *Let f be a biconvex function on a biconvex set B and let $z \in B$.*

1. *If z is a bi-relatively interior point of B then f is lower-semi-continuous at z.*
2. *If B is locally bi-simplicial at z then f is upper-semi-continuous at z.*

Since for all bi-relatively interior points z of B, B is locally bi-simplicial at z as well, it holds:

Corollary 3.8 *Let f be a biconvex function on a biconvex set B. Then f is continuous at all bi-relatively interior points $z \in B$.*

Note that only "directional continuity" (i.e., $f(x, \bullet) : Y \to \mathbb{R}$ and $f(\bullet, y) : x \to \mathbb{R}$ are continuous for all $x \in X$ and $y \in Y$) is not sufficient for a function f to be continuous on an open set $B \subseteq X \times Y$. A counter-example to this is given, for example, in the book of Gelbaum and Olmsted (2003, Chapter 9).

3.3 The Maximum of a Biconvex Function

This subsection deals with the problem of finding the maximum of a biconvex function over a given set contained in $X \times Y$. We recall known results for this problem and present a new result for the case that the maximum of a biconvex function is attained in the interior of a biconvex set B when B has some additional topological properties.

In the convex case, it is well-known that the set of all points where the supremum of a convex function relative to a given convex set C is attained, is given by a union of faces of C (cf. Rockafellar, 1997, Corollary 32.1.1), i.e., that the supremum of a convex function over a convex set C is attained at a boundary point of C if it exists. Al-Khayyal and Falk (1983) showed that this result is also valid for a continuous, biconvex function f over a compact and convex set $K \subseteq X \times Y$. (Actually, the result was proven for the minimum of a biconcave function, which is equivalent.)

Theorem 3.9 (Al-Khayyal and Falk (1983)) *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets, let $K \subseteq X \times Y$ be compact and convex, and let $f : K \to \mathbb{R}$ be a continuous, biconvex function. Then the problem*

$$\max \{ f(x, y) : (x, y) \in K \} \tag{3}$$

has always a solution on ∂K, the boundary of K.

If the given set K is a product of two polytopes in X and Y, respectively, Geng and Huang (2000b) stated:

Theorem 3.10 (Geng and Huang (2000b)) *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be biconvex and let $S \subset \mathbb{R}^n$, $T \subset \mathbb{R}^m$ be polytopes with vertex sets S^* and T^*, respectively. Then*

$$\max_{(x,y)\in S\times T} f(x,y) = \max_{(x,y)\in S^*\times T^*} f(x,y). \tag{4}$$

Note that in Geng and Huang (2000a) and Geng and Huang (2000b) the authors referred to a proof of the above theorem given in Barmish (1994). Another proof and an outer approximation algorithm for problem (4), based on the above theorem, can be found in Gao and Xu (2002).
Horst and Thoai (1996) presented a decomposition approach for the minimization of a biconcave function over a polytope $P \subset \mathbb{R}^{n+m}$ where P is not separable in the sense that it cannot be written as a product of two polytopes in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The authors used a combination of global optimization techniques such as branch and bound, polyhedral outer approximation and projection of polyhedral sets onto a subspace to design an algorithm for problems of the form

$$\min\{f(x,y) : x \in X,\, y \in Y,\, (x,y) \in D\}, \tag{5}$$

where $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are polytopes, $D \subset \mathbb{R}^{n+m}$ is a polyhedral set and f is biconcave. As special cases of problem (5), jointly constrained bilinear programming problems and separated jointly constrained biconcave programming problems of the form $f(x,y) = f_1(x) + f_2(y)$ are considered, among others.

Next we consider problems where the maximum of a biconvex function over a biconvex set B lies in the relative interior $\mathrm{ri}(B)$ of the set B. For the convex case we recall:

Theorem 3.11 (Rockafellar (1997)) *Let f be a convex function and let C be a convex set. If f attains its supremum relative to C at some point of the relative interior of C, then f is constant throughout C.*

Our aim is to prove that this result is also valid for the biconvex case if we make some more topological assumptions on the given biconvex set B. In order to derive a proof for this result we need some preliminary lemmas and definitions.

Definition 3.3 *Let $I = [a,b] \subseteq \mathbb{R}$ be an interval and let $\gamma : I \to M$ be a continuous function. Then γ is called a* **path** *with* **initial point** *$\gamma(a)$ and* **terminal point** *$\gamma(b)$.*

Definition 3.4 *Let $M \subseteq \mathbb{R}^n$ be a non-empty set. M is called* **path-connected** *if for any two points $m_1, m_2 \in M$ there exists a path $\gamma : [a,b] \to M$ with $\gamma(a) = m_1$, $\gamma(b) = m_2$ and $\gamma(t) \in M$ for all $t \in [a,b]$.*

Definition 3.5 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets, let $M \subseteq X \times Y$, let $m_1 := (x_1, y_1) \in M$ and $m_2 := (x_2, y_2) \in M$, and let γ be a path in M joining m_1 and m_2. We call γ* **L-shaped** *if we can partition γ into two subpaths γ_1 and γ_2 such that γ restricted to γ_1 consists of the line segment joining m_1 and the point $h_1 := (x_1, y_2)$ (or $h_2 := (x_2, y_1)$) and γ restricted to γ_2 consists of the line segment joining h_1 (or h_2)and m_2. The intermediate point h_1 (or h_2) is called* **inflection point** *of γ. An L-shaped path is said to be* **degenerate** *if $x_1 = x_2$ or $y_1 = y_2$.*

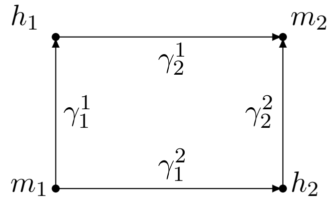

Figure 3: *Example of two* L*-shaped paths* γ^1 *and* γ^2 *joining* m_1 *and* m_2 *with inflection points* h_1 *and* h_2*, respectively.*

If $X, Y \subseteq \mathbb{R}$, an L-shaped path is a path of the form “L” or “¬” (cf. Figure 3). Furthermore, we define:

Definition 3.6 *Let* $X \subseteq \mathbb{R}^n$ *and* $Y \subseteq \mathbb{R}^m$ *be two non-empty sets, let* $M \subseteq X \times Y$*, and let* $m_1, m_2 \in M$*. If there exists a (finite) sequence of* L*-shaped paths joining* m_1 *and* m_2 *which is completely contained in* M*, we say that* m_1 *and* m_2 *are* **(finitely) L-connectable** *or* **(finitely) L-connected** *in* M*. The set* M *is* **(finitely) L-connected** *if any two points in* M *are (finitely)* L*-connectable.*

Due to the last definition it is obvious that every finitely L-connected set is path-connected, whereas the converse is not true in general. If we consider, for example, the line segment

$$M := \{(x, y) \in [0; 1] \times [0; 1] : (x, y) = \lambda\,(1, 0) + (1 - \lambda)\,(0, 1),\ \lambda \in [0; 1]\}$$

in $[0; 1] \times [0; 1]$, then M is path-connected, but any two points of M are not finitely L-connectable in M. Now, for $\epsilon > 0$ and $x \in \mathbb{R}^n$, let

$$K_\epsilon(x) := \{y \in \mathbb{R}^n : \|x - y\| < \epsilon\}$$

denote the open ball around x with radius ϵ. Then the following lemma can easily be proven by induction:

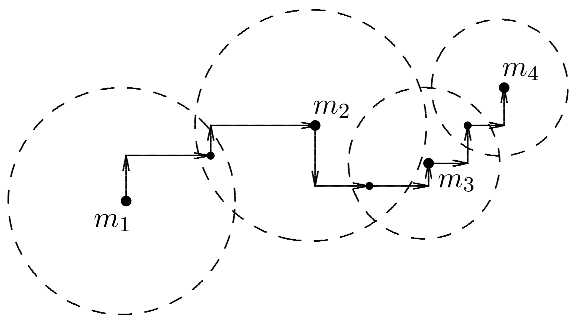

Figure 4: *Example of a sequence of* L*-shaped paths in Lemma 3.12.*

Lemma 3.12 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets and let $I := \{1 \ldots, k\}$ be an index set. Furthermore, let k points $m_i \in X \times Y$ be given and k positive numbers ϵ_i such that $K_{\epsilon_i}(m_i) \subseteq X \times Y$ for all $i \in I$ and the intersection $K_{\epsilon_i}(m_i) \cap K_{\epsilon_{i+1}}(m_{i+1})$ is not empty for all $i = 1, \ldots, k-1$. Then m_1 and m_k are finitely L-connectable in $\bigcup_{i \in I} K_{\epsilon_i}(m_i)$ such that the resulting path contains the points m_i, $i \in I$.*

The proof of this lemma is obvious and can be performed as indicated in Figure 4.

Theorem 3.13 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets and let $M \subseteq X \times Y$ be a non-empty, open, and path-connected set. Then M is finitely L-connected.*

Proof: Let $m_1, m_2 \in M \subseteq X \times Y$ be two arbitrary chosen points in M. Since by assumption M is path-connected, there exist $a, b \in \mathbb{R}$ and a path $\gamma : [a, b] \to M$ with $\gamma(a) = m_1$, $\gamma(b) = m_2$, and $\gamma(t) \in M$ for every $t \in I := [a, b]$.
Since M is an open set, for every point $\gamma(t) \in M$ on the curve there exists $\epsilon_t > 0$ such that $K_{\epsilon_t}(\gamma(t))$ is completely contained in M. Hence, $\{\bigcup_{t \in I} K_{\epsilon_t}(\gamma(t))\}$ builds an open covering of the image set $\gamma(I)$ of γ in M. Since $\gamma(I)$ is known to be compact, there exists $t_1, \ldots, t_n \in I$, such that $\gamma(I)$ is already covered by $\{\bigcup_{t=1}^{n} K_{\epsilon_{t_i}}(\gamma(t_i))\}$.
Without loss of generality we suppose that $t_1 = a$ and $t_n = b$, otherwise we add the two balls $K_{\epsilon_a}(\gamma(a))$ and $K_{\epsilon_b}(\gamma(b))$ to the finite open covering of $\gamma(I)$. By eventually deleting and rearranging the order of the open balls $K_{\epsilon_{t_i}}(\gamma(t_i))$ we can reorder the given finite covering in the way that the intersection of two consecutive open balls $K_{\epsilon_{t_i}}(\gamma(t_i))$ and $K_{\epsilon_{t_{i+1}}}(\gamma(t_{i+1}))$ is not empty.
Let the resulting covering be denoted again by $\{\bigcup_{t=1}^{n} K_{\epsilon_{t_i}}(\gamma(t_i))\}$. Then this covering satisfies all the assumptions made in Lemma 3.12. Hence, $\gamma(t_1) = m_1$ and $\gamma(t_n) = m_2$ are finitely L-connectable, which completes the proof. □

Now we can prove our main result:

Theorem 3.14 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets and let $B \subseteq X \times Y$ be a biconvex set such that the interior of B is non-empty and path-connected with $\partial(\mathrm{int}(B)) = \partial B$. Furthermore, let $f : B \to \mathbb{R}$ be a continuous, biconvex function. If the problem*

$$\max \{f(x, y) : (x, y) \in B\} \tag{6}$$

has an optimal solution $z^* := (x^*, y^*) \in \mathrm{int}(B)$, then f is constant throughout B.

Proof: We prove the theorem in two steps. First, we concentrate on points $z \in B$ lying in the interior of B and we show that $f(z^*) = f(z)$ holds for all points $z \in \mathrm{int}(B)$. In the second step we extend our results to points situated in $B \setminus \mathrm{int}(B)$.
So, let $z^* = (x^*, y^*) \in \mathrm{int}(B)$ denote the optimal solution of problem (6). First, consider the two functions

$$f_{x^*} : B_{x^*} \to \mathbb{R} \text{ and } f_{y^*} : B_{y^*} \to \mathbb{R},$$

where $B_{x^*} := \{y \in Y : (x^*, y) \in B\}$ and $B_{y^*} := \{x \in X : (x, y^*) \in B\}$, respectively. Since B is biconvex by assumption, the sets B_{x^*} and B_{y^*} are convex. Obviously, $y^* \in B_{x^*}$ and $x^* \in B_{y^*}$ hold. But since (x^*, y^*) is a point in $\mathrm{int}(B)$, y^* and x^* are elements of $\mathrm{ri}(B_{x^*})$ and $\mathrm{ri}(B_{y^*})$, respectively. Hence, by Theorem 3.11, f_{x^*} and f_{y^*} are constant on B_{x^*} and B_{y^*}, respectively. So we have that

$$f(z) = f(z^*) \quad \forall\, z \in B_{z^*} := \{(x^*, y) : y \in B_{x^*}\} \cup \{(x, y^*) : x \in B_{y^*}\}.$$

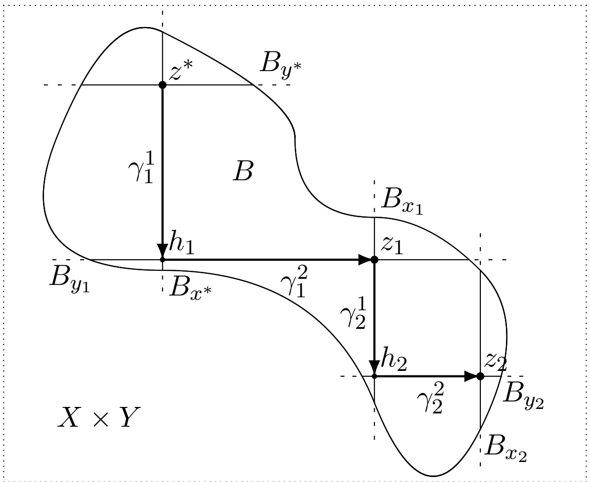

Figure 5: *Illustration of the proof of Theorem 3.14 with two intermediate points z_1 and z_2.*

Next, consider a point $z_1 = (x_1, y_1) \in \text{int}(B)$ which is L-connectable to z^* throughout $\text{int}(B)$ by exactly one L-shaped path γ, and let $h_1 := (x^*, y_1) \in \text{int}(B)$ denote the inflection point of γ. Since $h_1 \in B_{z^*}$, $f(h_1) = f(z^*)$. Since B is biconvex, the set $B_{y_1} := \{x \in X : (x, y_1) \in B\}$ is convex. Since $h_1 \in \text{int}(B)$, $x^* \in \text{ri}(B_{y_1})$. Hence, $f_{y_1}(x^*) \geq f_{y_1}(x)$ holds for all $x \in B_{y_1}$, and f_{y_1} is constant on B_{y_1} by Theorem 3.11. Since $x_1 \in B_{y_1}$, $f(z_1) = f(z^*)$. So we have proven that $f(z) = f(z^*)$ for all z which are L-connectable to z^* by exactly one L-shaped path.

Finally, let $z = (x, y) \in \text{int}(B)$ be arbitrarily chosen in $\text{int}(B)$. Since $\text{int}(B)$ is open by definition and non-empty and path-connected by assumption, z^* and z are finitely L-connectable in $\text{int}(B)$ by k L-shaped paths γ_k by Theorem 3.13 (see Figure 5).

Now, let $m_0 := z^*$, $m_k = z$, and let $m_i := (x_i, y_i) \in \text{int}(B)$ and $h_i := (x_{i-1}, y_i) \in \text{int}(B)$ $(i = 1, \ldots, k)$ denote the finite sequence of initial points and inflection points, respectively, obtained by the sequence of L-shaped paths from z^* to z.

Since $B_{x_i} := \{y \in Y : (x_i, y) \in B\}$ and $B_{y_i} := \{x \in X : (x, y_i) \in B\}$ are convex sets by assumption and $y_{i-1}, y_i \in \text{ri}(B_{x_{i-1}})$ and $x_{i-1}, x_i \in \text{ri}(B_{y_i})$ for $i = 1, \ldots, k$, respectively, we have, following the same argumentation as above, that f is subsequently constant on the L-shaped path γ_i joining m_{i-1} and m_i with inflection point h_i for $i = 1, \ldots, k$, i.e.,

$$f(z^*) = f(m_0) = f(m_1) = \ldots = f(m_{k-1}) = f(m_k) = f(z).$$

Hence, f is constant throughout $\text{int}(B)$. This completes the first step of the proof.

Now suppose that the point $z \in B$ is an element of $B \setminus \text{int}(B)$, i.e., $z \in \partial B \cap B$. Since, by assumption, $\partial(\text{int}(B)) = \partial B$, $z \in \partial\text{int}(B)$, i.e., there exists a sequence $\{z_n\}_{n \in \mathbb{N}}$ converging to z such that $z_n \in \text{int}(B)$ for all $n \in \mathbb{N}$. Since f is continuous on B and equal to the constant $f(z^*)$ on $\text{int}(B)$, we get that

$$f(z) = f(\lim_{n\to\infty} z_n) = \lim_{n\to\infty} f(z_n) = \lim_{n\to\infty} f(z^*) = f(z^*).$$

Hence, f is constant throughout B. $\square$

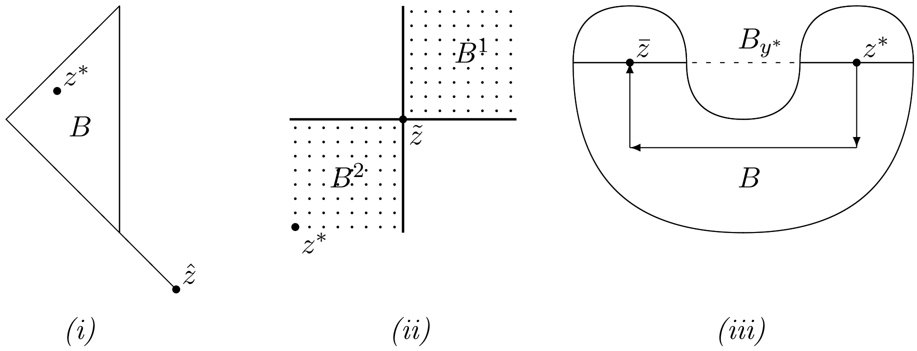

Figure 6: *Discussion of the assumptions made in Theorem 3.14.*

Before we conclude this subsection by reflecting on the assumptions made in Theorem 3.14, we remark that for a set $A \subseteq \mathbb{R}^n$ it holds that

$$\partial(\mathrm{int}(A)) = \partial A \quad \Longleftrightarrow \quad \mathrm{cl}(\mathrm{int}(A)) = \mathrm{cl}(A).$$

So, the assumption $\partial(\mathrm{int}(A)) = \partial A$ in the last theorem could be replaced by $\mathrm{cl}(\mathrm{int}(A)) = \mathrm{cl}(A)$. Note that in set-theoretical topology a set which equals the closure of its interior is called a regular (closed) set.

Now consider Figure 6. Figure 6(i) shows a set B where the assumption $\partial(\mathrm{int}(B)) = \partial B$ is not valid since the point $\hat{z}$ is an element of $\partial B \setminus \partial(\mathrm{int}(B))$, and $f(\hat{z})$ can be chosen arbitrarily without effecting the biconvexity of f. In this case, a biconvex function having a global maximum in $z^* \in \mathrm{int}(B)$ does not need to be constant throughout B since the point $\hat{z}$ is not L-connectable to z^* within B.

Figure 6(ii) shows the biconvex set $B := \mathbb{R}^2_+ \cup (-\mathbb{R}^2_+)$ where the interior of B is not path-connected any more. If a biconvex function f takes its maximum in $z^* \in \mathrm{int}(B^2)$ ($B^2 := -\mathbb{R}^2_+$), then f has to be constant on $B^2 \cup \partial B^1$, where $B^1 := \mathbb{R}^2_+$, but not necessarily on $\mathrm{int}(B^1)$. For a counter-example consider the function f given on B as follows:

$$f(z) = \begin{cases} 1 & , \text{ if } z \in B^2 \cup \partial B^1, \\ 1 - \min\{x, y\}, & \text{ if } x \in [0,1] \text{ or } y \in [0,1], \\ 0 & , \text{ if } x > 1 \text{ and } y > 1. \end{cases}$$

Obviously, f is continuous and biconvex on B but not constant throughout B, although B (but not $\mathrm{int}(B)$) is path-connected, using $\tilde{z}$ passing from B^1 to B^2 and the other way round.

Figure 6(iii) shows a set B which is not biconvex since the y^*-cut B_{y^*} is not convex. Hence, Theorem 3.14 is not applicable directly. Nevertheless, a biconvex function f taking its global maximum in $z^* \in \mathrm{int}(B)$ is constant throughout the given set since every point $\bar{z} \in \mathrm{int}(B)$ is still L-connectable to z^*. Hence, the line of argumentation of Theorem 3.14 is still valid, provided that the given set B can be partitioned into appropriate biconvex subsets such that Theorem 3.14 is applicable in these subsets. So, the biconvexity-assumption for the set B might be weakened.

3.4 Biconvexity and Separation

Aumann and Hart (1986) stated several separation theorems for biconvex functions. In this context, separation does not mean that we separate two biconvex sets by a biconvex or bilinear function, but we determine the set of all points $z \in B$, B biconvex, which cannot be separated from a subset $A \subset B$ of B by a biconvex function f. We give the main results and ideas here. For further details, we refer to the original paper. The results for the convex case, which we state next, can also be found there.

Definition 3.7 *Let $C \subseteq \mathbb{R}^n$ be a convex set and $A \subseteq C$. Then a point $z \in C$ is* **convex separated from A with respect to C** *if there exists a bounded convex function f on C such that $f(z) > \sup f(A) := \sup\{f(a) : a \in A\}$. Furthermore, let $\mathrm{ncs}(C)$ $(= \mathrm{ncs}_A(C))$ denote the set of all points $z \in C$ that cannot be convex separated from A.*

For the set $\mathrm{ncs}_A(C)$ we have:

Theorem 3.15 (Aumann and Hart (1986)) *Let $C \subseteq \mathbb{R}^n$ be a convex set and let $A \subseteq C$, then $\mathrm{ncs}_A(C)$ is a convex set and*

$$\mathrm{conv}(A) \subseteq \mathrm{ncs}_A(C) \subseteq \overline{\mathrm{conv}(A)},$$

where $\overline{\mathrm{conv}(A)}$ denotes the closure of the convex hull of A.

For biconvex sets this is as follows:

Definition 3.8 *Let $B \subseteq \mathbb{R}^n \times \mathbb{R}^m$ be a biconvex set and $A \subseteq B$. Then a point $z \in B$ is* **biconvex separated from A with respect to B** *if there exists a bounded biconvex function f on B such that $f(z) > \sup f(A) := \sup\{f(a) : a \in A\}$. Furthermore, let $\mathrm{nbs}(B)$ $(= \mathrm{nbs}_A(B))$ denote the set of all points $z \in B$ that cannot be biconvex separated from A.*

Obviously we have:

Lemma 3.16 (Aumann and Hart (1986)) *Let $B \subseteq \mathbb{R}^n \times \mathbb{R}^m$ be a biconvex set and let $A \subseteq B$. Then $z \in \mathrm{nbs}_A(B)$ if and only if $z \in B$ and, for all biconvex functions f defined on B, we have $f(z) \leq \sup f(A)$.*

As level sets of biconvex functions are biconvex by Theorem 3.2, for the set $\mathrm{nbs}_A(B)$ we have:

Theorem 3.17 (Aumann and Hart (1986)) *Let B be a biconvex set and let $A \subseteq B$. Then the set $\mathrm{nbs}_A(B)$ is biconvex and*

$$\mathrm{biconv}(A) \subseteq \mathrm{nbs}_A(B).$$

Different to the convex case, for biconvex separation we have $\mathrm{nbs}_A(B) \not\subseteq \overline{\mathrm{biconv}(A)}$ in general. For an example see Aumann and Hart (1986), Example 3.3.
Furthermore, the set $\mathrm{nbs}_A(B)$ depends on the given domain B, i.e., if $A \subset B^* \subset B$ and B and B^* are biconvex sets, then $\mathrm{nbs}_A(B^*) \subsetneq \mathrm{nbs}_A(B)$ in general (cf. Aumann and Hart, 1986, Example 3.5).
For more theorems dealing with the concept of biconvex separability of a point $z \in B$ from a given set $A \subseteq B$, we refer again to Aumann and Hart (1986). For example, one can find results for the case that the separating biconvex function f additionally has to be continuous on A.

4 Biconvex Minimization Problems

In the following we discuss biconvex minimization problems of the form given in Definition 1.3. As mentioned in Section 1, biconvex optimization problems may have a large number of local minima as they are global optimization problems in general. Nevertheless, there exist a couple of methods and algorithms which exploit the convex substructures of a biconvex optimization problem in order to solve such problems more efficiently than general global optimization methods do. Of course, such methods can be used to solve biconvex problems, too, as proposed, for example, in Goh et al. (1994), where subgradient descent methods or interior point methods were suggested to solve a special class of non-smooth, biconvex minimization problems. Since we are especially interested in biconvex optimization methods, we survey only those algorithms and methods which utilize the biconvex structure of the given problem.
This section is organized as follows: In the first subsection we discuss the notion of partial optimality and recall a necessary optimality condition for biconvex problems with separable constraints while the second subsection gives a review of existing solution methods for biconvex minimization problems and their practical applications.

4.1 Partial Optimality

In the following let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be two non-empty sets, let $B \subseteq X \times Y$, and let B_x and B_y denote the x-sections and y-sections of B, respectively.

Definition 4.1 *Let $f : B \to \mathbb{R}$ be a given function and let $(x^*, y^*) \in B$. Then, (x^*, y^*) is called a* **partial optimum** *of f on B, if*

$$f(x^*, y^*) \le f(x, y^*) \, \forall \, x \in B_{y^*} \text{ and } f(x^*, y^*) \le f(x^*, y) \, \forall \, y \in B_{x^*}.$$

We recall:

Definition 4.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a given function, let $\zeta \in \mathbb{R}^n$, and let the partial derivatives of f in ζ exist. If $\nabla f(\zeta) = 0$, then ζ is called a* **stationary point** *of f.*

Obviously we have:

Theorem 4.1 *Let $f : B \to \mathbb{R}$ be partial differentiable at $z^* \in \operatorname{int}(B)$ and let z^* be a partial optimum. Then, z^* is a stationary point of f in B.*

Note that the converse of Theorem 4.1 is not true in general.

Example 4.1 Let $z^* := (0, 0) \in \mathbb{R}^2$ and let the function $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = x^3 \cdot (x - 2) + y^2.$$

Then $\nabla f(z^*) = 0$ holds true, but for fixed $y^* = 0$ we have:

$$f(1, y^*) = f(1, 0) = -1 < 0 = f(0, 0) = f(z^*)$$

Hence, z^* is not a partial optimum.

However, if f is biconvex it is easy to prove that:

Theorem 4.2 *Let B be a biconvex set and let $f : B \to \mathbb{R}$ be a differentiable, biconvex function. Then, each stationary point of f is a partial optimum.*

Proof: Let $z^* := (x^*, y^*)$ be a stationary point of f in B. For fixed y^*, the function $f_{y^*} : B_{y^*} \to \mathbb{R}$ is convex, so

$$f_{y^*}(x) \geq f_{y^*}(x^*) + \left(\frac{\partial}{\partial x_1} f_{y^*}(x^*), \ldots, \frac{\partial}{\partial x_n} f_{y^*}(x^*)\right)^t (x - x^*)$$

is valid for all $x \in B_{y^*}$ (cf. Rockafellar, 1997). Since x^* is also a stationary point of f_{y^*}, the second summand equals zero and hence

$$f_{y^*}(x) \geq f_{y^*}(x^*) \quad \forall x \in B_{y^*}.$$

By symmetry of the problem we also have that

$$f_{x^*}(y) \geq f_{x^*}(y^*) \quad \forall y \in B_{x^*}$$

So, z^* is a partial optimum. □

Corollary 4.3 *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a differentiable, biconvex function. Then a point $z \in \mathbb{R}^{n+m}$ is stationary if and only if z is a partial optimum.*

Finally, we shortly review a necessary local optimality condition for the biconvex minimization problem with separable constraints

$$\min\{f(x, y) : x \in X \subseteq \mathbb{R}^n, y \in Y \subseteq \mathbb{R}^m\}. \tag{7}$$

In the case of separable constraints the notion of partial optimality of a point $(x^*, y^*) \in X \times Y$ simplifies to

$$f(x^*, y^*) \leq f(x, y^*) \, \forall x \in X \text{ and } f(x^*, y^*) \leq f(x^*, y) \, \forall y \in Y.$$

Theorem 4.4 (Wendell and Hurter Jr. (1976)) *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be convex sets and let $f : X \times Y \to \mathbb{R}$ be a biconvex function with a partial optimum in $(x^*, y^*) \in X \times Y$. Furthermore, let $U(y^*)$ denote the set of all optimal solutions to (7) with $y = y^*$ and let $U(x^*)$ be the set of optimal solutions to (7) with $x = x^*$. If (x^*, y^*) is a local optimal solution to (7), then it necessarily holds that*

$$f(x^*, y^*) \leq f(x, y) \quad \forall x \in U(x^*) \; \forall y \in U(y^*). \tag{8}$$

Note that the given local optimality condition is in general not sufficient.

Example 4.2 (Luenberger (1989), mod.) Consider the biconvex minimization problem

$$\min\{x^3 - x^2 y + 2y^2 : x \geq 4, y \in [0; 10]\}.$$

This problem has a partial optimum at $(6, 9)$ that satisfies the condition (8) of the last theorem, but that is not a local optimum.

4.2 Algorithms

In this subsection we discuss methods and algorithms for solving biconvex minimization problems of the form (1) which exploit the biconvex structure of the problem. We give short algorithmic descriptions for every solution approach and discuss convergence results and limitations of the considered methods. In detail, we present the *Alternate Convex Search* method as a special case of *Block-Relaxation Methods*, the *Global Optimization Algorithm*, developed in Floudas and Visweswaran (1990), and an algorithm for solving jointly constrained biconvex programming problems using the so called convex envelope of a function f.

4.2.1 Alternate Convex Search

Alternate Convex Search (ACS) is a minimization method which is a special case of the *Block-Relaxation Methods* where the variable set is divided into disjoint blocks (de Leeuw, 1994). In every step, only the variables of an active block are optimized while those of the other blocks are fixed. For ACS we only consider the two blocks of variables defined by the convex subproblems that are activated in cycles. Since the resulting subproblems are convex, efficient convex minimization methods can be used to solve these subproblems. In the case that $n = m = 1$, i.e., $f : B \subseteq \mathbb{R}^2 \to \mathbb{R}$, ACS can be seen as a special case of the *Cyclic Coordinate Method (CCM)* which is stated, e.g., in Bazaraa et al. (1993). A survey on the ACS approach for convex as well as for biconvex objective functions can be found, e.g., in Wendell and Hurter Jr. (1976).
In the following we will show that under weak assumptions the set of all accumulation points generated by ACS form a connected, compact set C and that each of these points is a stationary point of f but that no better convergence results (like local or global optimality properties) can be obtained in general.

Algorithm 4.1 (Alternate Convex Search)
Let a biconvex optimization problem in the sense of Definition 1.3 be given.

Step 1: Choose an arbitrary starting point $z_0 = (x_0, y_0) \in B$ and set $i = 0$.

Step 2: Solve for fixed y_i the convex optimization problem

$$\min \{f(x, y_i),\, x \in B_{y_i}\}. \tag{9}$$

If there exists an optimal solution $x^* \in B_{y_i}$ to this problem, set $x_{i+1} = x^*$, otherwise STOP.

Step 3: Solve for fixed x_{i+1} the convex optimization problem

$$\min \{f(x_{i+1}, y),\, y \in B_{x_{i+1}}\}. \tag{10}$$

If there exists an optimal solution $y^* \in B_{x_{i+1}}$ to this problem, set $y_{i+1} = y^*$, otherwise STOP.

Step 4: Set $z_{i+1} = (x_{i+1}, y_{i+1})$. If a stopping criterion is satisfied, then STOP, otherwise augment i by 1 and go back to Step 2.

Remarks:

1. The order of the optimization problems in Step 2 and Step 3 can be permuted, i.e., it is possible first to optimize in the y-variables, followed by an optimization in the x-variables.

2. There are several ways to define the stopping criterion in Step 4 of the algorithm. For example, one can consider the absolute value of the difference of z_{i-1} and z_i (or the difference in their function values) or the relative increase in the z-variable compared to the last iteration. The stopping criterion may also depend on the special structure of the given biconvex objective function.

The following convergence properties of ACS are motivated by the results of Zangwill (1969), Meyer (1976) and de Leeuw (1994). The results stated in these papers cannot be applied directly to ACS, since in these papers it is assumed that the algorithmic map A (see Definition 4.3 below) is uniformly compact on B (i.e., there exists $B_0 \subseteq B$, compact, such that $A(z) \subseteq B_0$ for all $z \in B$), which is not true for ACS in general. Note that for most of the following results only continuity of f is needed.

Theorem 4.5 *Let $B \subseteq \mathbb{R}^n \times \mathbb{R}^m$, let $f : B \to \mathbb{R}$ be bounded from below, and let the optimization problems* (9) *and* (10) *be solvable. Then the sequence $\{f(z_i)\}_{i\in\mathbb{N}}$ generated by ACS converges monotonically.*

Proof: Since the sequence of function values $\{f(z_i)\}_{i\in\mathbb{N}}$, generated by Algorithm 4.1, is monotonically decreasing and f is bounded from below, the sequence $\{f(z_i)\}_{i\in\mathbb{N}}$ converges to a limit value $a \in \mathbb{R}$. □

The statement of Theorem 4.5 is relatively weak. The boundedness of the objective function f only ensures the convergence of the sequence $\{f(z_i)\}_{i\in\mathbb{N}}$ but not automatically the convergence of the sequence $\{z_i\}_{i\in\mathbb{N}}$. Indeed, there exist biconvex functions where the sequence $\{f(z_i)\}_{i\in\mathbb{N}}$ generated by ACS converges while the sequence $\{z_i\}_{i\in\mathbb{N}}$ diverges. To see this, we consider the following example:

Example 4.3 Let the biconvex function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by:

$$f(x,y) := \begin{cases} (x-y)^2 + \frac{1}{x+y+1} & , \text{ if } \quad x \geq -y \\ (x-y)^2 + 1 - x - y, & \text{if } \quad x < -y. \end{cases}$$

It is easy to check that for any starting point $(x_0, y_0) \in \mathbb{R}^2$ the generated sequence $\{f(z_i)\}_{i\in\mathbb{N}}$ converges to 0 while the sequence $\{z_i\}_{i\in\mathbb{N}}$ diverges to infinity.

To give convergence results for the generated sequence $\{z_i\}_{i\in\mathbb{N}}$ we introduce the *algorithmic map* of the ACS algorithm. For a general definition of algorithmic maps we refer to Bazaraa et al. (1993).

Definition 4.3 *Let $B \subseteq \mathbb{R}^n \times \mathbb{R}^m$, let $z_k = (x_k, y_k) \in B$ for $k = 1, 2$, and let $f : B \to \mathbb{R}$ be given. The map $A : B \to \mathcal{P}(B)$ from B onto the power set $\mathcal{P}(B)$ of B defined by $z_2 \in A(z_1)$ if and only if*

$$f(x_2, y_1) \leq f(x, y_1) \quad \forall x \in B_{y_1} \text{ and } f(x_2, y_2) \leq f(x_2, y) \quad \forall y \in B_{x_2}$$

is called the **algorithmic map** *of the ACS algorithm.*

Using the algorithmic map, the ACS algorithm can be described as the iterative selection of a $z_{i+1} \in A(z_i)$. This means that z_{i+1} is a possible outcome of the algorithm with starting point z_i after one complete iteration.

Lemma 4.6 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be closed sets and let $f : X \times Y \to \mathbb{R}$ be continuous. Then the algorithmic map A is closed, i.e., it holds:*

$$\left.\begin{array}{ll} z_i := (x_i, y_i) \in X \times Y, & \lim\limits_{i\to\infty} (x_i, y_i) = (x^*, y^*) =: z^* \\ z_i' := (x_i', y_i') \in A(z_i), & \lim\limits_{i\to\infty} (x_i', y_i') = (x', y') =: z' \end{array}\right\} \Longrightarrow z' \in A(z^*).$$

Proof: Since $z_i' \in A(z_i)$ for all $i \in \mathbb{N}$ we have that

$$f(x_i', y_i) \le f(x, y_i)\, \forall x \in X \quad \text{and} \quad f(x_i', y_i') \le f(x_i', y)\, \forall y \in Y.$$

Since f is continuous by assumption we get that

$$f(x', y^*) = \lim_{i\to\infty} f(x_i', y_i) \le \lim_{i\to\infty} f(x, y_i) = f(x, y^*) \quad \forall x \in X$$

and

$$f(x', y') = \lim_{i\to\infty} f(x_i', y_i') \le \lim_{i\to\infty} f(x_i', y) = f(x', y) \quad \forall y \in Y.$$

Hence, $z' \in A(z^*)$. □

The following theorem states a condition for the limit of the sequence of points generated by ACS.

Theorem 4.7 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be closed sets and let $f : X \times Y \to \mathbb{R}$ be continuous. Let the sequence $\{z_i\}_{i\in\mathbb{N}}$ generated by ACS converge to $z^* \in X \times Y$. Then z^* is a partial optimum.*

Proof: The sequence $\{z_{i+1}\}_{i\in\mathbb{N}}$ is convergent with limit point z^*. Since the algorithmic map A is closed by Lemma 4.6 and $z_{i+1} \in A(z_i)$ for all $i \in \mathbb{N}$, also z^* is contained in $A(z^*)$. Hence,

$$f(x^*, y^*) \le f(x, y^*) \quad \forall x \in X \quad \text{and} \quad f(x^*, y^*) \le f(x^*, y) \quad \forall y \in Y$$

and z^* is a partial optimum. □

Note that a similar result is mentioned in Wendell and Hurter Jr. (1976) for X and Y being compact sets. The next lemma ensures that, as long the algorithm generates new points that are no partial optima, a descent in the function values can be achieved during one iteration.

Lemma 4.8 *Let $B \subseteq \mathbb{R}^n \times \mathbb{R}^m$ and $f : B \to \mathbb{R}$ be given. Let the optimization problems* (9) *and* (10) *be solvable and let $z_1 := (x_1, y_1) \in B$ and $z_2 := (x_2, y_2) \in A(z_1)$.*

1. *If the optimal solution of* (9) *with $y = y_1$ is unique, then*

$$z_1 \textit{ is not a partial optimum} \quad \Longrightarrow \quad f(z_2) < f(z_1).$$

2. *If the optimal solution of* (10) *with $x = x_2$ is unique, then*

$$z_2 \textit{ is not a partial optimum} \quad \Longrightarrow \quad f(z_2) < f(z_1).$$

3. *If the optimal solutions of both* (9) *with* $y = y_1$ *and* (10) *with* $x = x_2$ *are unique, then*

$$z_1 \neq z_2 \quad \Longrightarrow \quad f(z_2) < f(z_1).$$

Proof: Obviously, it holds true that

$$f(z_2) \;=\; f(x_2, y_2) \;\leq\; f(x_2, y_1) \;\leq\; f(x_1, y_1) \;=\; f(z_1).$$

We assume that $f(x_2, y_2) = f(x_2, y_1) = f(x_1, y_1)$ and show the reversed statements. Since $z_2 \in A(z_1)$,

$$f(x_2, y_1) \;\leq\; f(x, y_1)\, \forall x \in B_{y_1} \quad \text{and} \quad f(x_2, y_2) \;\leq\; f(x_2, y)\, \forall y \in B_{x_2}.$$

If the optimal solution of (9) with $y = y_1$ is unique, then $x_1 = x_2$ and z_1 is a partial optimum. If the optimal solution of (10) with $x = x_2$ is unique, then $y_1 = y_2$ and z_2 is a partial optimum. If the optimal solutions of both (9) with $y = y_1$ and (10) with $x = x_2$ are unique, $x_1 = x_2$ and $y_1 = y_2$, hence $z_1 = z_2$. □

Now the following theorem about the convergence of the sequence $\{z_i\}_{i\in\mathbb{N}}$ can be stated and proven.

Theorem 4.9 *Let* $X \subseteq \mathbb{R}^n$ *and* $Y \subseteq \mathbb{R}^m$ *be closed sets and let* $f : X \times Y \to \mathbb{R}$ *be continuous. Let the optimization problems* (9) *and* (10) *be solvable.*

1. *If the sequence* $\{z_i\}_{i\in\mathbb{N}}$ *generated by the ACS algorithm is contained in a compact set, then the sequence has at least one accumulation point.*

2. *In addition suppose that for each accumulation point* $z^* = (x^*, y^*)$ *of the sequence* $\{z_i\}_{i\in\mathbb{N}}$ *the optimal solution of* (9) *with* $y = y^*$ *or the optimal solution of* (10) *with* $x = x^*$ *is unique, then all accumulation points are partial optima and have the same function value.*

3. *Furthermore, if for each accumulation point* $z^* = (x^*, y^*)$ *of the sequence* $\{z_i\}_{i\in\mathbb{N}}$ *the optimal solutions of both* (9) *with* $y = y^*$ *and* (10) *with* $x = x^*$ *are unique, then*

$$\lim_{i\to\infty} \|z_{i+1} - z_i\| = 0$$

and the accumulation points form a compact continuum C *(i.e.,* C *is a connected, compact set).*

Proof: By condition 1, the sequence $\{z_i\}_{i\in\mathbb{N}}$ has at least one accumulation point $z^* := (x^*, y^*)$. Thus we have a convergent subsequence $\{z_k\}_{k\in\mathcal{K}}$ with $\mathcal{K} \subseteq \mathbb{N}$ that converges to z^*. Similarly, $\{z_{k+1}\}_{k\in\mathcal{K}}$ has an accumulation point $z^+ := (x^+, y^+)$ to which a subsequence $(z_{l+1})_{l\in\mathcal{K}}$ with $\mathcal{L} \subseteq \mathcal{K}$ converges. By Lemma 4.6 and Theorem 4.5 it follows that $z^+ \in A(z^*)$ and $f(z^+) = f(z^*)$. In the same manner we see that the sequence $\{z_{k-1}\}_{k\in\mathcal{K}}$ has an accumulation point $z^- := (x^-, y^-)$ with $z^* \in A(z^-)$ and $f(z^*) = f(z^-)$.
Now suppose that z^* is not a partial optimum even though condition 2 is fulfilled. Thus one of the optimization problems (9) with $y = y^*$ or (10) with $x = x^*$ has a unique solution, and by Lemma 4.8, $f(z^+) < f(z^*)$ or $f(z^*) < f(z^-)$, which gives a contradiction. Therefore, z^* must be a partial optimum. If there exist further accumulation points, their

function values must equal $f(z^*)$ due to Theorem 4.5.
Suppose that additionally condition 3 is satisfied, but $\|z_{i+1} - z_i\| > \delta$ for infinitely many $i \in \mathbb{N}$ and $\delta > 0$. Then the sequences $\{z_i\}_{i\in\mathbb{N}}$ and $\{z_{i+1}\}_{i\in\mathbb{N}}$ again have accumulation points z^* and z^+ with $\|z^+ - z^*\| \geq \delta$. In particular, $z^+ \neq z^*$. As above we see that $z^+ \in A(z^*)$ and $f(z^+) = f(z^*)$. But by Lemma 4.8 it follows that $f(z^+) < f(z^*)$ which gives a contradiction. Thus the sequence $\{\|z_{i+1} - z_i\|\}_{i\in\mathbb{N}}$ converges to 0, and since $\{z_i\}_{i\in\mathbb{N}}$ is bounded the accumulation points form a compact continuum (cf. Ostrowski, 1966, p. 203). □

Note that in Theorem 4.9 the problems (9) and (10) must be uniquely solvable only for the set of accumulation points but not for an arbitrary element of the sequence $\{z_i\}_{i\in\mathbb{N}}$. For a biconvex function f uniqueness of the solutions is automatically guaranteed in practice if, for example, f is strictly convex as a function of y for fixed x and vice versa.
Unfortunately, Theorem 4.9 still does not guarantee the convergence of the sequence $\{z_i\}_{i\in\mathbb{N}}$ but is close enough for all practical purposes. Note that statements similar to Theorem 4.9 can be found in the literature, e.g., for CCM in Bazaraa et al. (1993). But for ACS the assumptions the Theorem 4.9 are weaker since the biconvex structure is used.

Corollary 4.10 *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be closed sets and let $f : X \times Y \to \mathbb{R}$ be a differentiable function. Furthermore, let the sequence $\{z_i\}_{i\in\mathbb{N}}$ generated by the ACS algorithm be contained in a compact set, and for each accumulation point $z^* = (x^*, y^*)$ of the sequence $\{z_i\}_{i\in\mathbb{N}}$ let the optimal solutions of both (9) with $y = y^*$ and (10) with $x = x^*$ be unique. Then all accumulation points z^* which lie in the interior of $X \times Y$ are stationary points of f.*

Proof: This is an immediate consequence of Theorem 4.2 and Theorem 4.9. □

It is obviously clear that for every stationary point $z^* := (x^*, y^*) \in B$ of a differentiable, biconvex function f there exists a non-empty set S of starting points such that z^* is an outcome of the ACS algorithm when the optimal solutions of both (9) with $y = y^*$ and (10) with $x = x^*$ are unique, since all points of the form $(x, y^*) \in B$ will lead to z^* within one iteration. So theoretically, all stationary points of f can be generated by ACS.
Furthermore, it can be shown that if the assumptions of Theorem 4.9 are satisfied and X and Y are subsets of $\mathbb{R}$ (i.e., ACS simplifies to CCM), the generated compact continuum C simplifies to a singleton, i.e., the sequence $\{z_i\}_{i\in\mathbb{N}}$ is actually convergent.
Although an accumulation or limit point z^*, generated by Algorithm 4.1, might be a partial optimum, it neither has to be a global nor a local optimum to the given biconvex optimization problem even if z^* is stationary, as stationary points can be saddle points of the given function. To see this for the case when f is not everywhere differentiable over its whole domain we consider:

Example 4.4 (Goh et al. (1994)) Let the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by

$$f(x, y) := \max\{y - 2x,\, x - 2y,\, \frac{1}{4}(x^2 + y^2 - 16)\}.$$

As the pointwise maximum of three convex functions f is convex and thus also biconvex (cf. Lemma 3.5 and Theorem 3.6). Let $M^{(1)}$ and $M^{(2)}$ denote the points $(-4, 2)$ and $(2, -4)$, respectively, and define two sets $C^{(1)}$ and $C^{(2)}$ by

$$\begin{aligned} C^{(1)} &= \{z \in \mathbb{R}^2 : \|z - M^{(1)}\| \leq 6\} \\ C^{(2)} &= \{z \in \mathbb{R}^2 : \|z - M^{(2)}\| \leq 6\}. \end{aligned}$$

A calculation shows that

$$f(x,y) = \begin{cases} y - 2x & \text{for } (x,y) \in C^{(1)} \cap \{(x,y) \in \mathbb{R}^2 : x \leq y\} \\ x - 2y & \text{for } (x,y) \in C^{(2)} \cap \{(x,y) \in \mathbb{R}^2 : x \geq y\} \\ \frac{1}{4}(x^2 + y^2 - 16) & \text{for } (x,y) \in \mathbb{R}^2 \setminus (C^{(1)} \cup C^{(2)}). \end{cases}$$

Furthermore, f is continuous but not everywhere differentiable, since it has non-smooth transitions between the three regions defined above. Nevertheless, f has a global minimum at $z^* = (2,2)$ and the two convex optimization problems (9) and (10) given by the ACS algorithm are well-defined and always have unique solutions.
If this procedure is applied to f with a starting-point $z_0 = (x_0, y_0) \in \mathbb{R} \times [2, -4]$, the algorithm will converge to the point $\zeta := (y_0, y_0)$ within one iteration. But ζ is clearly not a minimum of f for $y_0 \in]2, -4]$.
Due to the symmetry of the problem, the result remains true if the first optimization in Step 2 of the algorithm is performed over the y-variables and x_0 is given in the interval $[2, -4]$.

What might happen in the cases when the domain of f is not of the form $X \times Y$ or the set of accumulation points is not a part of the interior of $X \times Y$? As the next example shows it is possible that, depending on the starting point, the resulting limit point of the ACS algorithm need not to be the global or a local minimum in most cases.

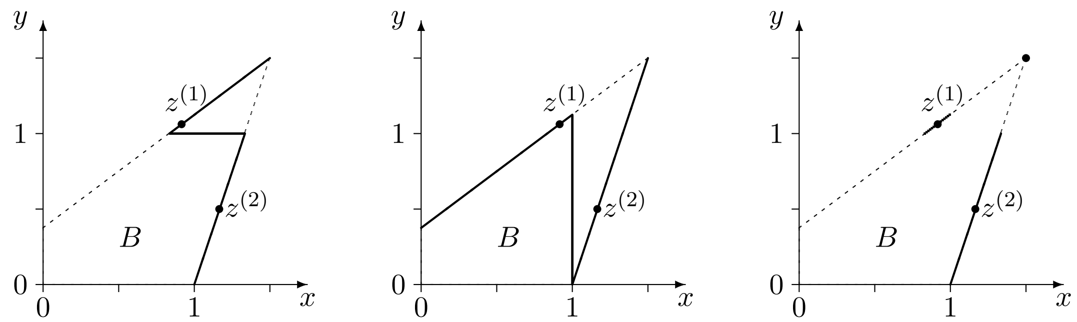

Figure 7: *The dashed lines mark the feasible set B of the problem in Example 4.5. The bold line in the first picture illustrates the set of optimal solutions for varying $y \in [0; 1.5]$ with f optimized in x-direction. The bold line in the second picture shows the optimal solutions for varying $x \in [0; 1.5]$ with f optimized in y-direction, while in the third picture the bold lines mark the set of possible outcomes M of the ACS algorithm depending on the chosen starting point.*

Example 4.5 (Floudas and Visweswaran (1990), mod.) Consider the biaffine, constrained minimization problem

$$\begin{array}{rrcl} \min\limits_{x,y} & -x + xy - y & & \\ \text{s.t.} & -6x + 8y & \leq & 3 \\ & 3x - y & \leq & 3 \\ & x, y & \in & [0; 1.5] \end{array}$$

which has a local minimum at the boundary point $z^{(1)} = (0.916, 1.062)$ and a global one at the boundary point $z^{(2)} = (1.167, 0.5)$. Denote the feasible set by B and the objective function by f. Since the objective function is affine for fixed x or y in $[0; 1.5]$, the minimal value of f for fixed x or y is attained in a boundary point of B. If we apply a constrained version of the ACS method to solve the problem given above, a calculation shows that every point of the set

$$M := \{(x, y) : 3x - y = 3, \, y \in [0; 1]\} \cup \{(x, y) : -6x + 8y = 3, \, y \in]1; 1.125]\} \cup \{(1.5, 1.5)\}$$

is a possible outcome of the algorithm, depending on the chosen starting point (see Figure 7). Furthermore, the set of starting points which lead to the global as well as to the local optimum is a discrete point set, and for a starting point z_0 with $y_0 \in [0; 1[$, only the choice $y_0 = 0.5$ results in the global optimum. For $y_0 \in [0; 1[$, the local minimum $z^{(1)}$ is never obtained. Hence the ACS algorithm, applied to the problem given above, can provide a point which is far away from being a global or local minimum of the problem.

To find the global optimum of a biconvex minimization problem by ACS, a multistart version of ACS can be used (like, e.g., suggested in Goh et al. (1994)). But as we have seen in the last example, there is no guarantee to find the global optimum within a reasonable amount of time or to be sure that the actual best minimum is the global one.

We conclude this subsection with a survey of classes of optimization problems where variants of the ACS method are frequently used to solve biconvex minimization problems in practice. One of these classes are the *Location-Allocation Problems (LAP)* in Location Theory. Examples for these types of problems are the *Multisource Weber Problem (MWP)*, first formally stated by Cooper (1963), or the *K-Connection Location Problem (KCLP)* in the plane (see Huang et al., 2005) which can be seen as an extension of the classical MWP. In these problems, m new facilities ($m > 1$) have to be placed in the plane and allocated to a set of n given existing facilities (or demand points) such that given demands of the existing facilities are satisfied and the total transportation cost between the existing and the new facilities is minimized. Note that the definition of the total transportation cost depends on the specific problem. This class of optimization problems has a biconvex objective function which is neither convex nor concave (cf. Cooper (1963) for the MWP). A well-known heuristic approach to the classical MWP which can also be applied to general LAP's is the alternate location and allocation algorithm developed by Cooper (1964) that alternates between a location and an allocation phase until no further improvement can be achieved. This corresponds to the ACS approach applied to the given LAP. A general survey on the application of Location-Allocation methods in Location Theory can be found, for example, in Hodgson et al. (1993) and Plastria (1995). If we apply the above developed convergence results of the ACS algorithm to the special cases of MWPs and KCLAPs, respectively, we can state that Theorem 4.5 holds true in both cases since the objective function is always non-negative. So, the generated sequence of function values always converges. Furthermore, since the m new facilities lie within the convex hull of all existing facilities if the distance function is chosen appropriately (this is true, e.g., for Euclidean distances) and the decision variables are restricted to $\{0, 1\}$, also item (1.) in Theorem 4.9 applies in both cases. But since neither the position of the new locations nor the partition of the decision variables need to be unique in general, no further results in the decision space can be given in general.
In medical image analysis an ACS approach can be used to register two medical images.

In general, a registration problem is a problem where two given data sets have to be rendered in a joint coordinate system such that corresponding features of the two data sets are aligned. In medical image registration the two data sets normally correspond to 2- or 3-dimensional images, the *template image* T which has to be mapped onto the *reference image* R by an appropriate transformation f which can be a rigid function, i.e., a combination of rotations and translations, or a non-rigid function. In practice, rigid transformations are used for registration when it is known that both images T and R show the same part of the body but from a different perspective. Furthermore, they are used to detect morphological changes of an organ (e.g., the growth of a tumor) while non-rigid transformations are normally applied to compensate those changes.

One way to formulate the described registration problem is to select a set of characteristical points $X = \{x_1, \ldots, x_I\}$ in the template image T and a set of corresponding characteristical points $Y = \{y_1, \ldots, y_J\}$ in the reference image R. Then the transformation $f : T \to R$ is chosen from a set $\mathcal{F}$ of feasible transformations such that the sum of the distances between each image point $f(x_i) \in R$ ($x_i \in X$) and its closest point $y_j \in Y$ in the reference set is minimized. This approach leads to the following generalized biconvex assignment problem

$$
\begin{array}{lrcll}
\min\limits_{f,z} & \sum\limits_{i=1}^{I} \sum\limits_{j=1}^{J} z_{ij} \|f(x_i) - y_j\|^2 & & & \\
\text{s.t.} & \sum\limits_{j=1}^{J} z_{ij} & = & 1, & i = 1, \ldots, I, \\
& z_{ij} & \in & \{0,1\}, & i = 1, \ldots, I,\ j = 1, \ldots, J, \\
& f & \in & \mathcal{F}, &
\end{array}
\tag{11}
$$

where z_{ij} equals 1 if the point $x_i \in X$ is assigned to the point $y_j \in Y$, i.e., y_j is the closest point to $f(x_i)$ in Y. Otherwise, z_{ij} is set to 0. Note that the binary constraints on the assignment variables z_{ij} can be relaxed to $z_{ij} \in [0, 1]$ since the assignment matrix is totally unimodular.

A common solution approach to problem (11) is the *Iterative Closest Point (ICP)* algorithm which was developed by Besl and McKay (1992) and corresponds to the ACS approach. The algorithm alternates between an assignment step in which the points $f(x_i) \in R$ are assigned to their closest neighbor in Y and a step in which a new transformation function f is chosen until no further improvement is achieved. For further details we refer to Zitová and Flusser (2003) where a survey on image registration can be found, and to Besl and McKay (1992) for information on the ICP algorithm. From the theoretical point of view, we get the same results as in the case of the LAPs. The given objective function is always non-negative, so the sequence of function values produced by the ICP algorithm is convergent by Theorem 4.5. Usually, the set of all feasible transformations $\mathcal{F}$ can be restricted such that the transformations are determined by only a finite number of parameters which are contained in a compact subset of $\mathbb{R}^n$. In this case, the feasible set of the problem is compact and the sequence generated by the ICP algorithm has an accumulation point in the decision space by Theorem 4.9. Since neither the chosen transformation nor the assignment variables need to be unique, no further results can be obtained in general.

Another field where ACS is frequently used as a standard approach is the field of (robust) control theory. For example, the *Biaffine Matrix Inequality (BMI) Feasibility Problem*, stated, e.g., in Goh et al. (1994), can be solved by the ACS method. But since the BMI problem has a non-smooth objective function, in general no global or local minimum can be determined by using the ACS approach (cf. Example 4.4). So, other non-convex optimization methods have to be considered to obtain an optimal solution for the BMI problem. For further details see, e.g., Goh et al. (1994) and Goh et al. (1995).

4.2.2 The Global Optimization Algorithm

In this subsection we review an algorithm for constrained biconvex minimization problems which exploits the convex substructure of the problem by a primal-relaxed dual approach. The algorithm is called *Global Optimization Algorithm (GOP)* and was developed by Floudas and Visweswaran (1990). The method follows decomposition ideas introduced by Benders (1962) and Geoffrion (1972). Like in the second step of the ACS method, the constrained problem is firstly solved for a fixed value of the y-variables which leads to an upper bound on the solution of the biconvex problem. This problem is called *primal problem.* To get a lower bound to the solution, duality theory and linear relaxation are applied. The resulting *relaxed dual problem* is solved for every possible combination of bounds in a subset of the x-variables, the set of *connected* x-variables. By iterating between the primal and the relaxed dual problem a finite ϵ-convergence to the global optimum can be shown.
In the following, we focus on the assumptions that have to be satisfied by the given biconvex minimization problem so that the GOP algorithm can be applied to this problem. A deeper description of the mathematical background and a detailed outline of the algorithm are given in Floudas and Visweswaran (1990), Floudas and Visweswaran (1993) and the books Floudas (1995) and Floudas (2000). We shortly review convergence results and give a short survey on the optimization fields in which the algorithm is used.
We consider an optimization problem of the form

$$
\begin{array}{rrcl}
 & \min\limits_{x,y} f(x,y) & & \\
\text{s.t.} & g(x,y) & \leq & 0 \\
 & h(x,y) & = & 0 \\
 & x \in X,\, y \in Y, & &
\end{array}
\tag{12}
$$

where X and Y are compact convex sets, and $g(x,y)$ and $h(x,y)$ are vectors of inequality and equality constraints. The functions must be differentiable and they must be given in explicit form. Furthermore, the following conditions, denoted by *Conditions (A)*, need to be satisfied (cf. Floudas, 2000, Chapter 3.1)

1. f is biconvex on $X \times Y$.
2. g is biconvex on $X \times Y$.
3. h is biaffine on $X \times Y$.
4. An appropriate first order constraint qualification is satisfied for fixed y.

Note that, for example, in Floudas (2000) partitioning and transformation methods for the variable set of quadratic programming problems are suggested so that it is possible to transform this class of problems into a problem of type (12) where Condition (A) is

satisfied automatically.
Now, let

$$V := \{y : h(x, y) = 0,\ g(x, y) \leq 0 \text{ for some } x \in X\},$$

then the following ϵ-convergence result for the GOP algorithm holds:

Theorem 4.11 (Floudas (2000)) *If X and Y are non-empty compact convex sets satisfying that $Y \subset V$, f, g, and h are continuous on $X \times Y$, the set $U(y)$ of optimal multipliers for the primal problem is non-empty for all $y \in Y$ and uniformly bounded in some neighborhood of each such point and Condition (A) is satisfied, then the GOP algorithm terminates in a finite number of steps for any given $\epsilon > 0$.*

For the resulting solution it holds:

Corollary 4.12 (Floudas (2000)) *If the conditions stated in Theorem 4.11 hold, the GOP algorithm will terminate at the global optimum of the biconvex minimization problem.*

What are the advantages and drawbacks of the GOP algorithm? As mentioned above, one of the advantages of the algorithm is the fact that the primal problem which has to be solved in the first step of each iteration is a convex problem. Hence, every local optimum is the global minimum of the subproblem. Furthermore, the set of constraints for the convex subproblem often simplifies to linear or quadratic constraints in the x-variables so that the primal problem can be solved by any conventional non-linear local optimization solver. As another advantage of this approach can be seen that the relaxed dual problem has only to be solved in the connected x-variables. This might reduce the number of variables for which the relaxed dual problem has to be solved. For further details see, e.g., Floudas (2000).
The main drawback of the GOP algorithm is the fact that in each iteration of the algorithm, $2^{|I|}$ in general non-linear subproblems have to be solved to obtain a new lower bound to the problem, where I denotes the set of the connected x-variables. In fact, in each iteration a total enumeration of all possible assignments of the connected x-variables to their lower and upper bounds is done and the relaxed dual problem is solved for every combination of these bounds. In Floudas (2000), several improvements of the algorithm, depending on the structure of the given biconvex problem, are given to reduce the number of relaxed dual problems.
The GOP algorithm is a useful tool for different classes of practical optimization problems. Visweswaran and Floudas (1990), Visweswaran and Floudas (1993), and Floudas (2000) discuss, among others, quadratic problems with linear constraints, quadratically constrained problems, and univariate polynomial problems. Furthermore, an application to bilevel linear and quadratic problems, a practical approach to phase and chemical equilibrium problems as well as an implementation and computational studies of the GOP algorithm can be found there.
In Barmish et al. (1995) a solution algorithm for some open robustness problems including matrix polytope stability is stated which was influenced by the ideas of the GOP approach. There the optimization on the x-variables is carried out in the usual way (i.e., for fixed y) to get a valid upper bound. The optimization on the y-variables is done by a "relaxation" process where the relaxation is refined at each subsequent iteration step. By an accumulation of the resulting constraints better lower bounds on the problem are obtained in each step of the iteration and an ϵ-convergence to the optimum can be proved. Note that for the problems stated Barmish et al. (1995), only a finite number of linear

programs have to be solved to get an ϵ-global optimum.
A convex minimization problem with an additional biconvex constraint is considered in the paper of Tuyen and Muu (2001). There, a convex criterion function of a multiple objective affine fractional problem has to be minimized over the set of all weakly efficient solutions of the fractional problem. As in the GOP algorithm, Lagrangian duality and a simplicial subdivision is used to develop a branch and bound algorithm which is proven to converge to a global ϵ-optimal solution of the problem.

4.2.3 Jointly Constrained Biconvex Programming

In this subsection we concentrate on a special case of a jointly constrained biconvex programming problem, firstly considered in Al-Khayyal and Falk (1983). The specific problem is given by

$$\begin{aligned} \min_{(x,y)} \quad \Phi(x,y) &= f(x) + x^t y + g(y) \\ \text{s.t. } (x,y) &\in S \cap \Omega, \end{aligned} \tag{13}$$

where

1. f and g are convex over $S \cap \Omega$,
2. S is a closed, convex set, and
3. $\Omega = \{(x,y) : l \leq x \leq L,\, m \leq y \leq M\}$.

Since the functions f and g are convex, the objective function Φ is biconvex on $S \cap \Omega$. Problem (13) can be seen as a generalization of an ordinary bilinear programming problem which is of the form

$$\begin{aligned} \min_{(x,y)} \quad & c^t x + x^t A y + d^t y \\ \text{s.t.} \quad & x \in X,\, y \in Y, \end{aligned} \tag{14}$$

where $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$ are given vectors, A is an $(n \times n)$-matrix and X and Y are polytopes in $\mathbb{R}^n$. The above given biconvex problem is more general since it allows joint constraints in (x,y) and non-linear, convex subfunctions f and g in Φ. The bilinear problem (14) can be transformed into the biconvex problem (13) by replacing the term $x^t\,(Ay)$ by $x^t\,z$ and including the linear constraint $z = Ay$ among the constraints defining the feasible set.
While bilinear problems of the form (14) always have extreme-point solutions in $X^* \times Y^*$ (cf. Horst and Tuy, 1990), where X^* and Y^* denote the set of extreme-points of the polytopes X and Y, respectively, this is no longer the case for biconvex problems of the form (13) (cf. Al-Khayyal and Falk, 1983). Nevertheless, if the objective function Φ is also a biconcave function which is optimized over a compact, convex set $C \subset \mathbb{R}^n \times \mathbb{R}^n$, then it can be shown that if the minimum of Φ over C exists, it is achieved in at least one boundary point of C (cf. Al-Khayyal and Falk, 1983).
Al-Khayyal and Falk (1983) used a pattern of Falk and Soland (1969) to develop a branch and bound algorithm for solving the jointly constrained biconvex problem (13). The necessary bounds are obtained by employing convex envelopes.

Definition 4.4 (Floudas (2000)) *Let f be a lower semicontinuous function defined on a non-empty convex set $C \subseteq \mathbb{R}^n$. Then the* **convex envelope** *of f on C is a function $\Psi_C(f) : C \to \mathbb{R}$ that satisfies:*

1. *$\Psi_C(f)$ is convex on C.*
2. *$\Psi_C(f(x)) \leq f(x)$ for all $x \in C$.*
3. *If h is any convex function defined on C such that $h(x) \leq f(x)$ for all $x \in C$, then $h(x) \leq \Psi_C(f(x))$ for all $x \in C$.*

Note that the convex envelope of f is obtained by taking the pointwise supremum of all convex (or linear) functions which underestimate f over C (cf. Al-Khayyal and Falk, 1983).
Since

$$\begin{aligned} \Psi_\Omega(x^t y) &= \sum_{i=1}^{n} \Psi_{\Omega_i}(x_i y_i) && \forall\, (x,y) \in \Omega, \\ \Psi_\Omega(x^t y) &= x^t y && \forall\, (x,y) \in \partial\Omega, \end{aligned}$$

and $\Psi_{\Omega_i}(x_i y_i)$ can easily be calculated (cf. Al-Khayyal and Falk, 1983),

$$F(x,y) := f(x) + \Psi_\Omega(x^t y) + g(y)$$

is a convex underestimator of Φ on $S \cap \Omega$ that coincides with Φ on $\partial\Omega$ and is used to calculate lower bounds of the objective functions.
Now the algorithm works as follows: In the first step the minimization problem (13) is solved with F instead of Φ as objective function which leads to an optimal point $z^1 = (x^1, y^1) \in S \cap \Omega$ and valid lower and upper bounds $F(z^1)$ and $\Phi(z^1)$. If $F(z^1) = \Phi(z^1)$, then z^1 is optimal. Otherwise, there exists at least one $i \in \{1, \ldots, n\}$ such that $\Phi_{\Omega_i}(x_i y_i) < x_i y_i$. So, the index i that leads to the largest difference between $x_i y_i$ and $\Phi_{\Omega_i}(x_i y_i)$ is chosen, and the i^{th} rectangle Ω_i is split up into four subrectangles. Then new bounds are calculated in each of the resulting four new hyper-rectangles. This leads to a point z^2 and new lower and upper bounds for f which can be shown to be tighter than the bounds of the last iteration. If $F(z^2) = \Phi(z^2)$ the algorithm stops, otherwise a new refinement is performed. By iteratively applying this procedure, it can be shown that the algorithm converges to a global optimum of problem (13).
Horst and Tuy (1990) presented in their book a modified version of the algorithm which differs from the original one in the choice of the new iterate z^k and the subdivision rule which is based on bisection there. In Al-Khayyal (1990) the author strengthened the algorithm by also evaluating the concave envelope of the problem. In Audet et al. (2000) a short overview of papers which concentrate on the application of the algorithm to bilinear programming problems and quadratically constrained quadratic programming problems is given.
Another algorithm for a special type of functions f and g of problem (13) is developed in Sherali et al. (1994) and Sherali et al. (1995) for risk management problems. Instead of working with the convex envelope, the authors used a specialized implementation of Geoffrion's Generalized Benders' decomposition (see Geoffrion, 1972). With the help of a projection method and dual theory, an alternative graphical solution scheme is proposed that enables the decision maker to interact subjectively with the optimization process.

5 Conclusion

In this paper we gave a survey on optimization problems with biconvex sets and biconvex functions and reviewed properties of these sets and functions given in the literature. We

stated a new result for the case that the maximum of a biconvex function f is attained in the relative interior of a biconvex set B by assuming further, rather weak topological properties on B. We showed that under these assumptions f must be constant throughout B.
Existing methods and algorithms, specially designed for biconvex minimization problems which primarily exploit the convex substructures of the problem, were discussed for the constrained as well as for the unconstrained case. In particular, we showed that an alternating convex search approach, a primal-relaxed dual approach, as well as an approach that uses the convex envelope of parts of the biconvex objective function are suitable for solving biconvex optimization problems using the special properties of these problems. For each of these methods different practical applications as well as applications to the bilinear and biaffine case were discussed. We recalled that under appropriate assumptions the primal-relaxed dual approach as well as the approach that uses the convex envelope lead to a global optimum while the alternating approach in general only finds partial optima and stationary points of the objective function, respectively. The advantage of this approach is that it can be applied to any biconvex minimization problem while for the other approaches additional properties for the given objective function as well as for the feasible set must be satisfied.
Further fields of research related to biconvex sets and functions are separation theorems of disjoint biconvex sets with biconvex functions (Aumann and Hart, 1986). Also improvements of the given minimization algorithms are of interest, especially of the ACS method.

References

Al-Khayyal, F. (1990). Jointly constrained bilinear programs and related problems: An overview. *Computers in Mathematical Applications 19*(11), 53–62.

Al-Khayyal, F. and J. Falk (1983). Jointly constrained biconvex programming. *Mathematics of Operations Research 8*(2), 273–286.

Audet, C., P. Hansen, B. Jaumard, and G. Savard (2000, January). A branch and cut algorithm for non-convex quadratically constrained quadratic programming. *Mathematical Programming, Series A 87*(1), 131–152.

Aumann, R. and S. Hart (1986). Bi-convexity and bi-martingales. *Israel Journal of Mathematics 54*(2), 159–180.

Barmish, B. (1994). *New Tools for Robustness of Linear Systems.* (New York): Maxwell, Macmillan International.

Barmish, B., C. Floudas, C. Hollot, and R. Tempo (1995, June). A global programming solution to some open robustness problems including matrix polytope stability. In *Proceedings of the American Control Conference Seattle, Washington*, pp. 3871–3877.

Bazaraa, M., H. Sherali, and C. Shetty (1993). *Nonlinear Programming - Theory and Algorithms* (Second ed.). (New York): John Wiley & Sons, Inc.

Benders, J. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik 4*, 238–252.

Besl, P. and N. McKay (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 239–256.

Borwein, J. (1986). Partially monotone operators and the generic differentiability of convex-concave and biconvex mappings. *Israel Journal of Mathematics 54*(1), 42–50.

Burkholder, D. (1981). A geometrical characterization of Banach spaces in which martingale difference sequences are unconditional. *The Annals of Probability 9*(6), 997–1011.

Burkholder, D. (1986). *Lecture Notes in Mathematics*, Volume 1206 of *Probability and Analysis (Varenna, 1985)*, Chapter Martingales and Fourier analysis in Banach spaces, pp. 61–108. Springer-Verlag.

Cooper, L. (1963). Location-allocation problems. *Operations Research 11*, 331 – 343.

Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM Review 6*, 37–53.

de Leeuw, J. (1994). Block relaxation algorithms in statistics. In H. Bock, W. Lenski, and M. Richter (Eds.), *Information Systems and Data Analysis*, pp. 308–325. Springer.

Falk, J. and R. Soland (1969, May). An algorithm for separable nonconvex programming problems. *Management Science 15*(9), 550–569.

Floudas, C. (1995). *Nonlinear and Mixed Integer Optimization: Fundamentals and Applications*. New York: Oxford Press.

Floudas, C. (2000). *Deterministic Global Optimization* (First ed.). (Dordrecht, Netherlands): Kluwer Academic Publishers.

Floudas, C. and V. Visweswaran (1990). A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: I. Theory. *Computers and Chemical Engineering 14*(12), 1397 – 1417.

Floudas, C. and V. Visweswaran (1993). A primal-relaxed dual global optimization approach. *Journal of Optimization Theory and Applications 78*(2), 187–225.

Gao, Y. and C. Xu (2002). An outer approximation method for solving the biconcave programs with separable linear constraints. *Mathematica Applicata 15*(3), 42–46.

Gelbaum, B. and J. Olmsted (2003). *Counterexamples in Analysis*. (Mineola, New York): Dover Publications, Inc.

Geng, Z. and L. Huang (2000a). Robust stability of systems with both parametric and dynamic uncertainties. *Systems & Control Letters 39*, 87–96.

Geng, Z. and L. Huang (2000b). Robust stability of the systems with mixed uncertainties under the IQC descriptions. *International Journal of Control 73*(9), 776–786.

Geoffrion, A. (1972). Generalized Benders decomposition. *Journal of Optimization Theory and Applications 10*(4), 237–260.

Goh, K., M. Safonov, and G. Papavassilopoulos (1995). Global optimization for the biaffine matrix inequality problem. *Journal of Global Optimization 7*, 365–380.

Goh, K., L. Turan, M. Safonov, G. Papavassilopoulos, and J. Ly (1994, June). Biaffine matrix inequality properties and computational methods. In *Proceedings of the American Control Conference Baltimore, Maryland*, pp. 850–855.

Hodgson, M., K. Rosing, and Shmulevitz (1993). A review of location-allocation applications literature. *Studies in Locational Analysis 5*, 3–29.

Horst, R. and N. Thoai (1996, March). Decomposition approach for the global minimization of biconcave functions over polytopes. *Journal of Optimization Theory and Applications 88*(3), 561–583.

Horst, R. and H. Tuy (1990). *Global Optimization, Deterministic Approaches.* (Berlin, Heidelberg): Springer-Verlag.

Huang, S., R. Batta, K. Klamroth, and R. Nagi (2005). K-Connection location problem in a plane. *Annals of Operations Research 136*, 193–209.

Jouak, M. and L. Thibault (1985). Directional derivatives and almost everywhere differentiability of biconvex and concave-convex operators. *Mathematica Scandinavica 57*, 215–224.

Lee, J. (1993, June). On Burkholder's biconvex-function characterisation of Hilbert spaces. In *Proceedings of the American Mathematical Society*, Volume 118, pp. 555–559.

Luenberger, D. (1989). *Linear and Nonlinear Programming* (Second ed.). Reading, Massachusetts: Addison-Wesley.

Meyer, R. (1976). Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *Journal of Computer and System Sciences 12*, 108–121.

Ostrowski, A. (1966). *Solution of Equations and Systems of Equations* (Second ed.). (New York and London): Academic Press.

Plastria, F. (1995). Continuous location problems. In Z. Drezner (Ed.), *Facility Location*, pp. 225–262. Springer Series in Operations Research.

Rockafellar, R. (1997). *Convex Analysis* (First ed.). (Princton, New Jersey): Princeton University Press.

Sherali, H., A. Alameddine, and T. Glickman (1994). Biconvex models and algorithms for risk management problems. *American Journal of Mathematical and Management Sciences 14*(3 & 4), 197–228.

Sherali, H., A. Alameddine, and T. Glickman (1995). Biconvex models and algorithms for risk management problems. *Operations Research / Management Science 35*(4), 405–408.

Thibault, L. (1984, February). Continuity of measurable convex and biconvex operators. In *Proceedings of the American Mathematical Society*, Volume 90, pp. 281–284.

Tuyen, H. and L. Muu (2001). Biconvex programming approach to optimization over the weakly efficient set of a multiple objective affine fractional problem. *Operations Research Letters 28*, 81–92.

Visweswaran, V. and C. Floudas (1990). A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: II. Application of theory and test problems. *Computers and Chemical Engineering 14*(12), 1419 – 1434.

Visweswaran, V. and C. Floudas (1993). New properties and computational improvement of the GOP algorithm for problems with quadratic objective function and constraints. *Journal of Global Optimization 3*(3), 439–462.

Wendell, R. and A. Hurter Jr. (1976). Minimization of non-separable objective function subject to disjoint constraints. *Operations Research 24*(4), 643–657.

Zangwill, W. (1969). Convergence conditions for nonlinear programming algorithms. *Management Science 16*(1), 1–13.

Zitová, B. and J. Flusser (2003). Image registration methods: A survey. *Image and Vision Computing 21*, 977–1000.