A guided tour through the zoo of paired optimization problems

Dedicated to Marco Antonio López Cerdá

Oliver Stein

Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

Corresponding author(s). E-mail(s): stein@kit.edu;

Abstract

Many mathematical models base on the coupling of two or more optimization problems. This paper surveys possibilities to couple two optimization problems and discusses how solutions of the different models are interrelated with each other. The considered pairs stem from the fields of standard and generalized Nash equilibrium problems, optimistic and pessimistic bilevel problems, saddle problems, standard and generalized semi-infinite problems, robust optimization, Lagrange duality, bicriteria optimization problems, minimax problems, decomposition, and two-stage stochastic optimization. Connections to vector optimization and variational inequalities are discussed as well.

1 Introduction

Mathematical models from application fields like economics, finance, engineering, operations research, and machine learning often base on the coupling of two or more optimization problems. This paper surveys, in a structured manner, possibilities to couple two optimization problems and discusses how solutions of the different models are interrelated with each other. The focus is on these structural relations, but less on applications, optimality conditions, stability, or solution methods.

In Section 2 we propose general models for symmetrically (Section 2.1) and asymmetrically (Section 2.2) coupled pairs of optimization problems, provide their interpretations as Nash equilibrium problems and Stackelberg games, respectively, and briefly introduce bicriteria and vector optimization problems in Section 2.3.

Section 3 discusses special symmetrically coupled pairs of optimization problems, partly supported by the variational inequality formulation (Section 3.1). We discuss problems with a potential function (Section 3.2), saddle problems (Section 3.3) and generalized saddle problems (Section 3.4). Motivated by the latter, also two concepts of local saddle points are briefly compared (Section 3.5).

Special asymmetrically coupled pairs are considered in Section 4, in particular standard and generalized semi-infinite problems with applications to robust optimization (Section 4.1), minimax problems (Section 4.2), decomposition (Section 4.3), and two-stage stochastic optimization (Section 4.4).

Interconnections between models based on symmetrically and asymmetrically coupled pairs, and also some relations to vector optimization, are discussed in Section 5. A fundamental connection between symmetrically and asymmetrically coupled pairs of optimization problems is the characterization of saddle points by solutions of minimax problems (Section 5.1), which lies at the heart of Lagrange duality in nonlinear optimization (Section 5.2). A connection between Nash equilibrium and bilevel problems under far more general assumptions in discussed in Section 5.3. Section 5.4 considers relations between Nash equilibrium and vector optimization problems, before Section 5.5 provides results on relations between bilevel and vector optimization problems.

Some final remarks close this survey in Section 6.

2 General models

The coupling of two optimization problem may either be symmetric (Section 2.1) or asymmetric (Section 2.2). In both cases we will assume one of the two problems to be of the form

$$(Q(x)) \qquad \min_{y\in\mathcal{Y}} f(x,y) \quad \text{s.t.} \quad y \in Y(x)$$

with parameter vectors x from a host set $\mathcal{X} \subseteq \mathbb{R}^n$ and vectors y of decision variables from a host set $\mathcal{Y} \subseteq \mathbb{R}^m$. The parameter-dependent feasible sets are defined via the set-valued mapping

$$Y : \mathcal{X} \rightrightarrows \mathcal{Y},\ x \mapsto \{y \in \mathcal{Y} \mid g(x,y) \leq 0,\ h(x,y) = 0\}$$

with vector-valued constraint functions g, h defined on $\mathcal{X} \times \mathcal{Y}$. The objective function f is real-valued and also defined on $\mathcal{X} \times \mathcal{Y}$. The functional description of $Y(x)$ will play a minor role in the following. Integrality constraints on (some of) the variables in the vector y or parameters in the vector x may be modeled by appropriate choices of the host sets $\mathcal{Y}$ and $\mathcal{X}$, respectively.

With the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ of extended real numbers, we denote the minimal value function of $Q(x)$ by

$$\varphi : \mathcal{X} \to \overline{\mathbb{R}},\ x \mapsto \inf_{y\in Y(x)} f(x,y), \tag{2.1}$$

where we adopt the usual convention $\varphi(x) = -\infty$ if $f(x,\cdot)$ is not bounded from below on $Y(x)$, and $\varphi(x) = +\infty$ for $Y(x) = \emptyset$. The (possibly empty) set of minimal points of $Q(x)$ may then be written as

$$S(x) = \operatorname*{arg\,min}_{y \in Y(x)} f(x,y) = \{y \in Y(x) \mid f(x,y) = \varphi(x)\}.$$

When required, the equality constraint in the definition of $S(x)$ may equivalently be replaced by $f(x,y) \leq \varphi(x)$.

To avoid confusion, we remark that the 'min' in the problem statement of $(Q(x))$ refers to the task of minimizing the objective function, but does not imply that a minimal point of $(Q(x))$ exists. Therefore the minimal value function φ is defined with the 'inf' notation, which takes the several types of unsolvability of $(Q(x))$ into account. In the solvability case this infimum is attained as a minimum and we shall write $\varphi(x) = \min_{y \in Y(x)} f(x,y)$. The 'max' and 'sup' notations are handled analogously.

Example 2.1. *For the choices $\mathcal{X} = \mathcal{Y} = [-4,4]$, $f(x,y) = y$ and $Y(x) = \{y \in \mathcal{Y} \mid x+y \geq 2\}$ with $x \in \mathcal{X}$ one obtains $\varphi(x) = 2-x$ and $S(x) = \{2-x\}$ for all $x \in [-2,4]$ as well as $\varphi(x) = +\infty$ and $S(x) = \emptyset$ for all $x \in [-4,-2)$.*

2.1 Symmetric coupling

2.1.1 Model

For symmetrically coupled optimization problems, the second problem possesses essentially the same format as the problem $(Q(x))$, but now with y playing the role of a parameter vector, while x is the vector of decision variables. The host sets $\mathcal{X}$ and $\mathcal{Y}$ are identical to the ones in $(Q(x))$, but the defining functions may differ. More explicitly, we consider the problem

$$(P(y)) \qquad \min_{x \in \mathcal{X}} F(x,y) \quad \text{s.t.} \quad x \in X(y)$$

with parameter $y \in \mathcal{Y}$. The feasible sets are defined via the set-valued mapping

$$X : \mathcal{Y} \rightrightarrows \mathcal{X},\ y \mapsto \{x \in \mathcal{X} \mid G(x,y) \leq 0,\ H(x,y) = 0\}$$

with vector-valued constraint functions G, H defined on $\mathcal{X} \times \mathcal{Y}$. The objective function F is real-valued on $\mathcal{X} \times \mathcal{Y}$. We denote the minimal value function of $(P(y))$ by

$$\Phi : \mathcal{Y} \to \overline{\mathbb{R}},\ y \mapsto \inf_{x \in X(y)} F(x,y)$$

and the set of minimal points by

$$R(y) = \operatorname*{arg\,min}_{x \in X(y)} F(x,y) = \{x \in X(y) \mid F(x,y) = \Phi(y)\}.$$

We will refer to $(P(y), Q(x))$ with $(x, y) \in \mathcal{X} \times \mathcal{Y}$ as a pair of symmetrically coupled optimization problems.

Note that not all decision variables $x_1, \ldots, x_n$ of $P(y)$ need to enter $Q(x)$ as parameters, and likewise not all decision variables $y_1, \ldots, y_m$ of $Q(x)$ need to enter $P(y)$. The variables x_i and y_j which do are called linking variables. Furthermore, not all constraints in the system $G(x, y) \leq 0$, $H(x, y) = 0$ need to depend on the parameter y, and not all constraints in the system $g(x, y) \leq 0$, $h(x, y) = 0$ need to depend on the parameter x. The ones which do are called coupling constraints.

2.1.2 Solution concept

The basic solution concept for pairs of symmetrically coupled optimization problems is that of an equilibrium. Indeed, a point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ is called equilibrium of the pair $(P(y), Q(x))$ if $\bar{x} \in R(\bar{y})$ and $\bar{y} \in S(\bar{x})$ hold. This notion may be restated using the graphs

$$\operatorname{gph} R \;=\; \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid x \in R(y)\}$$

and

$$\operatorname{gph} S \;=\; \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid y \in S(x)\}$$

of the set-valued mappings $R : \mathcal{Y} \rightrightarrows \mathcal{X}$ and $S : \mathcal{X} \rightrightarrows \mathcal{Y}$, respectively. The set E of equilibria of the pair $(P(y), Q(x))$ can thus be written succinctly as

$$E = \operatorname{gph} R \cap \operatorname{gph} S. \tag{2.2}$$

The generalization of this equilibrium concept to more than two coupled optimization problems is straightforward.

We remark that the identity (2.2) relies on a slight abuse of notation with respect to the set $\operatorname{gph} R$. The standard definition of this graph is $\{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid x \in R(y)\}$ [62]. Indeed, in standard notation our set $\operatorname{gph} R$ is the graph of the inverse mapping of R. However, this non-standard definition only re-orders the variable groups and does not affect the development of the subsequent results. The same remark applies to the graph of X, which will appear in Section 2.2.1.

Example 2.2. *In addition to the choices from Example 2.1 let $F(x, y) = x^2 + y^2$, and $X(y) \equiv \overline{X} = [-4, 4]$ for all $y \in \mathcal{Y}$. This yields $\Phi(y) \equiv 0$ and $R(y) \equiv \{0\}$ for all $y \in \mathcal{Y}$. Since the graphs of R and S are $\operatorname{gph} R = \{(x, y) \in [-4, 4]^2 \mid x = 0\}$ and $\operatorname{gph} S = \{(x, y) \in [-4, 4]^2 \mid x + y = 2\}$, the pair $(P(y), Q(x))$ possesses the unique equilibrium point $(\bar{x}, \bar{y}) = (0, 2)$.*

One can easily construct symmetrically coupled pairs of optimization problems which do not possess equilibrium points, just by choosing $(P(y))$ and $(Q(x))$ with $(x, y) \in \mathcal{X} \times \mathcal{Y}$ such that the graphs $\operatorname{gph} R$ and $\operatorname{gph} S$ of their minimal point mappings do not intersect. Sufficient conditions for the existence of equilibrium points are presented in, e.g., [25, Sec. 4.1], [26, Prop. 2.2.9], and in the references therein. Their basic versions rely on the observation that the set E coincides with the set of fixed points of the set-valued mapping $T : \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{X} \times \mathcal{Y}, (x, y) \to R(y) \times S(x)$. In particular, under certain compactness and convexity assumptions as well as the closedness of the graphs $\operatorname{gph} R$ and $\operatorname{gph} S$, Kakutani's fixed point theorem [41] guarantees that T possesses a fixed point on $\mathcal{X} \times \mathcal{Y}$. The closedness of $\operatorname{gph} S$ will be further discussed in Section 2.2.2.

2.1.3 Game theoretic interpretation

A main, but not exclusive (see Section 3), source of coupled optimization problems are applications from non-cooperative game theory. There one interprets the problems $(P(y))$ and $(Q(x))$ as to be owned by two respective decision makers, also named players or agents, who do not cooperate. The decision of either player enters the other player's problem as a parameter. Let player 1 and player 2 own the problem $(P(y))$ and $(Q(x))$, respectively. In this context the set-valued mapping S is termed response mapping of player 2, given the decision x of player 1, and vice versa R is the response mapping of player 1, given the decision y of player 2. Again, the generalization of such a two-person game to an N-person game with $N \geq 2$ is straightforward.

Equilibria $(\bar{x}, \bar{y})$ of $(P(y), Q(x))$ are exactly the points in which none of the two players possess a rational incentive to deviate from their decision, since both players have chosen a minimal point of their respective problem, given the other player's decision. In the game theoretic context, such points are called generalized Nash equilibria. The task to find a point in $E = \operatorname{gph} R \cap \operatorname{gph} S$ is called a (two-person) generalized Nash equilibrium problem (GNEP). GNEPs go back to [1, 19]. We refer to [25, 27] for comprehensive surveys of theory, applications, methods, and history of GNEPs.

In the game theoretic context the feasible sets $X(y)$ and $Y(x)$ are called strategy sets. In the case of constant strategy sets $X(y) \equiv \overline{X} \subseteq \mathcal{X}$ and $Y(x) \equiv \overline{Y} \subseteq \mathcal{Y}$ one speaks of a (two-person) standard Nash equilibrium problem (NEP), and the points in $E = \operatorname{gph} R \cap \operatorname{gph} S$ are called (standard) Nash equilibria. Standard NEPs were introduced by Nash in [60].

In the following we will frequently return to the assumption of constant feasible sets, since it simplifies theory as well as algorithmic approaches across different couplings of optimization problems.

2.2 Asymmetric coupling

2.2.1 Models

In pairs of asymmetrically coupled optimization problems, there exists a hierarchy in which the problem $(Q(x))$ is subordinate to a higher level problem. Therefore, such problems are also called bilevel problems. To distinguish symmetric from asymmetric coupling the former is sometimes called horizontal, and the latter vertical coupling.

To derive the two main formats of bilevel problems, we start with the equivalent [70] epigraph reformulation

$$(P_{\text{epi}}(y)) \qquad \min_{(x,z) \in \mathcal{X} \times \mathbb{R}} z \quad \text{s.t.} \quad F(x,y) \leq z, \; x \in X(y)$$

of $P(y)$. Both bilevel approaches turn its parameter y into a decision variable which is constrained to lie in the minimal point set $S(x)$ of $Q(x)$. To ensure that the resulting problem is well-defined, y is equipped with a quantifier. More specifically, the strong version of a bilevel problem in epigraph formulation is

$$(P_{s,\text{epi}}) \qquad \min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \forall\, y \in S(x):\; F(x,y) \le z,\; x \in X(y),$$

whereas the weak version has the form

$$(P_{w,\text{epi}}) \qquad \min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \exists\, y \in S(x):\; F(x,y) \le z,\; x \in X(y).$$

In the well-posed case [75] when $S(x)$ is a singleton $\{y(x)\}$ for all $x \in \mathcal{X}$ (in other words, when $Q(x)$ possesses a unique minimal point for each $x \in \mathcal{X}$, so that the set-valued mapping S is single-valued, like in Example 2.1), the strong and the weak versions both reduce to the problem

$$(P_{\text{epi}}) \qquad \min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad F(x,y(x)) \le z,\; x \in X(y(x)),$$

and reversing the epigraph reformulation results in the problem

$$(P_{wp}) \qquad \min_{x\in\mathcal{X}} F(x,y(x)) \quad \text{s.t.} \quad x \in X(y(x)).$$

Since single-valuedness of S is often not guaranteed in applications, in the following we shall not consider the well-posed case, but rather discuss the more challenging general setting.

We remark that the attribution of 'strong' and 'weak' to the problems $(P_{s,\text{epi}})$ and $(P_{w,\text{epi}})$ is not consistently used in the literature. While, e.g., [49] switches the two attributes, we adapt the seemingly more established choice from, e.g., [20].

The strong version $(P_{s,\text{epi}})$ may be rewritten as

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \sup_{y\in S(x)} F(x,y) \le z,\; x \in \bigcap_{y\in S(x)} X(y),$$

and reversing the epigraph reformulation leads to the general format

$$(P_s) \qquad \min_{x\in\mathcal{X}} \sup_{y\in S(x)} F(x,y) \quad \text{s.t.} \quad x \in \bigcap_{y\in S(x)} X(y)$$

of the strong version of a bilevel problem.

For the weak formulation one cannot argue analogously, since the problem

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \inf_{y\in S(x)} F(x,y) \le z,\; x \in \bigcup_{y\in S(x)} X(y)$$

is only a possibly proper relaxation of $(P_{w,\text{epi}})$ without further assumptions (see below).

However, by a projection argument one can see that with respect to global minimal points and the global minimal value the problem $(P_{w,\mathrm{epi}})$ is equivalent to

$$\min_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathbb{R}} z \quad \text{s.t.} \quad F(x,y)\le z,\ x\in X(y),\ y\in S(x). \tag{2.3}$$

We emphasize that, in the terminology of [10, 55], this reformulation makes the implicit variable y explicit. Consequently one must expect that (2.3) possesses more local minimal points than $(P_{w,\mathrm{epi}})$. Such spurious local minimizers may occur as the output of methods from local nonlinear optimization, although they do not correspond to local minimizers of the original problem. A sufficient condition for the absence of spurious local minimizers is given in [55, Th. 4.2].

Reversing the epigraph formulation in (2.3) results in the format

$$(P_{sw}) \qquad \min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} F(x,y) \quad \text{s.t.} \quad x\in X(y),\ y\in S(x),$$

called the standard weak version of a bilevel problem. Observe that in (P_{sw}) F is minimized in the pair of variables (x,y) over the set

$$\operatorname{gph} X\cap\operatorname{gph} S=\{(x,y)\in\operatorname{gph} X\cap\operatorname{gph} Y \mid f(x,y)=\varphi(x)\}, \tag{2.4}$$

which is referred to as the optimal-value-function (or briefly value-function) formulation. As mentioned above, the equality constraint in (2.4) may equivalently be replaced by $f(x,y)\le\varphi(x)$. The set $\operatorname{gph} X\cap\operatorname{gph} S$ is also called inducible region of (P_{sw}).

The set

$$\Omega=\operatorname{gph} X\cap\operatorname{gph} Y \tag{2.5}$$

from (2.4) is called shared constraint set (or shared feasible set) of (P_{sw}), and the minimization of F over Ω is referred to as the high point relaxation of (P_{sw}). While in a general model the shared constraint set Ω it is derived from the set-valued mappings X and Y by (2.5), vice versa, as a structural assumption one can also derive X and Y from a given shared constraint set Ω by taking its fibers

$$X(y):=\{x\in\mathcal{X} \mid (x,y)\in\Omega\}, \quad y\in\mathcal{Y}, \tag{2.6}$$

$$Y(x):=\{y\in\mathcal{Y} \mid (x,y)\in\Omega\}, \quad x\in\mathcal{X}. \tag{2.7}$$

This yields especially $\Omega=\operatorname{gph} X=\operatorname{gph} Y$, so that (2.5) is trivially satisfied. That X and Y are derived from this set Ω is actually equivalent to the requirement that Y is the inverse mapping of the set-valued mapping X [62, Ch. 5A].

Also GNEPs, as pairs of symmetrically coupled optimization problems, are often studied under this assumption. If Ω is also closed and convex, and F and f are convex on Ω, then a GNEP is said to enjoy joint convexity [25]. We remark that in the literature it is less common to call Ω from (2.5) the shared constraint set of a general GNEP, but the formulation that a GNEP 'possesses a shared (or joint) constraint set' rather refers to the fact that X and Y are derived from a given set Ω via (2.6) and (2.7). To avoid confusion, we adopt the latter terminology in this paper. Observe that also here one may equivalently require that Y is the inverse mapping of X, but that this view does not generalize to N-player GNEPs with $N > 2$.

Returning to bilevel problems, other equivalent reformulations of the problem $(P_{w,\text{epi}})$ are possible for a y-independent function $F(x)$ or for a y-independent feasible set $X(y) \equiv \overline{X} \subseteq \mathcal{X}$. The first case will be discussed in detail in Section 4.1. In the second case $(P_{w,\text{epi}})$ becomes

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \exists\, y \in S(x) : \; F(x,y) \leq z, \; x \in \overline{X}. \tag{2.8}$$

Under the assumption that finite infima of $F(x,\cdot)$ over $S(x)$ are attained for all $x \in \overline{X}$ (e.g., for polyhedral problems), problem (2.8) may be rewritten as

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad \inf_{y\in S(x)} F(x,y) \leq z, \; x \in \overline{X},$$

and reversing the epigraph reformulation yields the format

$$(P_{ow}) \qquad \min_{x\in\mathcal{X}} \inf_{y\in S(x)} F(x,y) \quad \text{s.t.} \quad x \in \overline{X},$$

called the original weak version of a bilevel problem. Since going from the original weak version (P_{ow}) to the standard weak version

$$\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} F(x,y) \quad \text{s.t.} \quad x \in \overline{X}, \; y \in S(x) \tag{2.9}$$

still makes the implicit variable y explicit, the two problems are in general only equivalent with respect to global minimal points and the global minimal value, but not with respect to the local ones. This effect is investigated in detail in [75]. Despite this drawback of the standard weak version, it is often studied, since it indeed possesses a two-level structure. In contrast, the original weak version exhibits a three-level structure, since it interrelates three minimization problems in a hierarchical fashion. Its theoretical and algorithmic treatment is thus more intricate than in the standard setting.

Analogously, one may call (P_s) the original strong version of a bilevel problem [73]. A corresponding standard strong version is introduced in [47]. It also makes the implicit variable y explicit, and it allows a reformulation of the strong version by a bilevel problem in standard weak version with a parametric two-player GNEP as the lower level problem. Roughly speaking, this replaces the vertical three-level structure of the original strong version by a vertical two-level structure with a horizontally (i.e. symmetrically) coupled pair of optimization problems in the lower level.

Linking variables and coupling constraints in asymmetrically coupled pairs of optimization problems are defined as for symmetrically coupled pairs (Section 2.1.1). Comprehensive introductions to bilevel optimization are given in [5, 20, 64], and for a bibliography with more than 1,500 entries see [21].

2.2.2 Solution concept

In contrast to symmetrically coupled problems, no special solution concept (in analogy to equilibria) needs to be introduced for asymmetrically coupled problems. In fact, one needs to find minimal points of (P_{sw}), (P_{ow}) or (P_s).

Example 2.3. *With the specifications from Example 2.2 the feasible set of* (P_{sw}) *is* $\operatorname{gph} X \cap \operatorname{gph} S = \{(x,y) \in [-4,4]^2 \mid x+y=2\}$*, so that the unique minimal point of* (P_{sw}) *is* $(\bar{x},\bar{y}) = (1,1)$*. In comparison to Example 2.2 this illustrates that the symmetric and asymmetric couplings of the same problems in general possess different solutions.*

The statement of sufficient conditions for the existence of such minimal points is an intricate issue, even under continuity and compactness assumptions on the involved functions and sets. Indeed, semi-continuity properties of the set-valued mapping S play a crucial role for solvability.

For example, proving the existence of a minimal point of (P_{sw}) by the Weierstrass theorem requires, besides lower semi-continuity of F and a non-empty and bounded feasible set $\operatorname{gph} X \cap \operatorname{gph} S$, the closedness of the latter intersection. The set $\operatorname{gph} X$ is easily seen to be closed for lower semi-continuous and continuous entries of the functions G and H, respectively, but the closedness of $\operatorname{gph} S$ may fail even for a linear function f and non-empty compact sets $Y(x)$, $x \in \mathcal{X}$, if Y is not inner semi-continuous [3, 62] on $\mathcal{X}$. The following mixed-integer example with even single-valued S illustrates this issue.

Example 2.4 ([45, Ex. 1.1])**.** *The choices* $\mathcal{X} = \mathbb{R}$*,* $\mathcal{Y} = \mathbb{Z}$*,* $F(x,y) = x - y$*,* $X(y) \equiv [0,1]$*,* $f(x,y) = y$*, and* $Y(x) = \{y \in \mathcal{Y} \mid x \le y \le 1\}$ *lead to* $\operatorname{gph} S = \{(0,0)\} \cup \big((0,1] \times \{1\}\big)$*, so that the infimum* -1 *of* (P_{sw}) *is not attained.*

The closedness of $\operatorname{gph} S$ actually corresponds to a semi-continuity property of S, namely its outer semi-continuity [3, 62]. A proof of the solvability of (P_s) by the Weierstrass theorem requires the lower semi-continuity of its objective function $\sup_{y \in S(x)} F(x,y)$. However, this does not follow from continuity of F and outer semi-continuity of S, but under the rather strong assumption of inner semi-continuity of S [3]. This is illustrated by the next example, inspired by [50, Ex. 4.2].

Example 2.5 ([47, Ex. 3.1])**.** *The choices* $\mathcal{X} = \mathcal{Y} = \mathbb{R}$*,* $F(x,y) = x^2 + y$*,* $f(x,y) = xy$*,* $X(y) = Y(x) \equiv [-1,1]$*, result in the closed graph* $\operatorname{gph} S = \big([-1,0) \times \{1\}\big) \cup \big(\{0\} \times [-1,1]\big) \cup \big((0,1]) \times \{-1\}\big)$ *and*

$$\sup_{y \in S(x)} F(x,y) = \begin{cases} x^2+1, & x \in [-1,0], \\ x^2-1, & x \in (0,1]. \end{cases}$$

Therefore the infimum -1 *of* (P_s) *is not attained. Observe that* S *is not inner semi-continuous at* $\bar{x} = 0$ *and, indeed, the above function is not lower semi-continuous* *at $\bar{x}$. On the other hand, the infimum -1 of the original weak formulation (P_{ow}) is attained, since its objective function $\inf_{y \in S(x)} F(x,y)$ is lower semi-continuous (but not upper semi-continuous) at $\bar{x}$.*

More detailed discussions of solvability in bilevel optimization may be found in [20, 44, 75] and the references therein.

2.2.3 Game theoretic interpretation

Like symmetrically coupled problems, also asymmetrically coupled problems may be motivated from a game theoretic perspective. Again, two players are involved, namely an upper level player and a lower level player, also called leader and follower, respectively. Unlike in GNEPs, these players do not decide simultaneously, but sequentially. Indeed, the leader passes a decision vector x to the follower, who then returns an element y of $S(x)$ to the leader. The leader anticipates the choice of y and thus can solve an optimization problem in the variable x. This structure is called a Stackelberg game [67]. The minimal point $\bar{x}$ of the leader together with the selected minimal point $\bar{y} \in S(\bar{x})$ of the follower's problem is called Stackelberg equilibrium (although 'only' optimization problems are solved). Under appropriate conditions, for the leader it is advantageous to move first, which is a natural motivation to consider a Stackelberg game rather than a Nash game (e.g., the Nash equilibrium $(\bar{x}, \bar{y}) = (0, 2)$ from Example 2.2 possesses the objective value $F(\bar{x}, \bar{y}) = 4$, whereas the Stackelberg equilibrium $(\bar{x}, \bar{y}) = (1, 1)$ from Example 2.3 has the lower value $F(\bar{x}, \bar{y}) = 2$). For a succinct discussion of the first mover advantage see [54].

Stackelberg games are considered under the assumption of a constant upper level feasible set $X(y) \equiv \overline{X} \subseteq \mathcal{X}$. Let us also assume that $S(x)$ is non-empty and compact for each $x \in \overline{X}$. Then the original strong version (P_s) of the Stackelberg game is

$$(SG_{os}) \qquad \min_{x \in \mathcal{X}} \max_{y \in S(x)} F(x,y) \quad \text{s.t.} \quad x \in \overline{X},$$

while the original weak version (P_{ow}) is

$$(SG_{ow}) \qquad \min_{x \in \mathcal{X}} \min_{y \in S(x)} F(x,y) \quad \text{s.t.} \quad x \in \overline{X}.$$

Hence, if the leader solves the strong version (SG_{os}), she expects that the follower returns a point y from $S(x)$ with the worst possible value of her upper level objective function $F(x, \cdot)$. This makes sense if the follower is an adversary of the leader. Of course this approach only has an effect for points $x \in \overline{X}$ with a non-unique lower level minimal point set $S(x)$.

On the other hand, solving the weak version (SG_{ow}) means that the leader expects a cooperative follower who returns some $y \in S(x)$ with the best possible value of $F(x, \cdot)$. For these reasons the strong version (SG_{os}) of the Stackelberg game is also called pessimistic, while the weak version (SG_{ow}) is called optimistic. More generally, for the two versions of any bilevel optimization problem the terms pessimistic bilevel problem and optimistic bilevel problem are used as synonyms for their strong and weak versions, respectively.

In case that the follower's problem does not depend on x (since $F(x, y)$ and $Y(x)$ are constant with respect to x), also the set-valued mapping S is constant with $S(x) \equiv \overline{S}$. Then the Stackelberg game is called purely hierarchical or simple. The distinction between the pessimistic and the optimistic version then still makes sense, as long as $\overline{S}$ is not a singleton.

We mention that the recent paper [17] calls a point $(\bar{x}, \bar{y})$ a lower Stackelberg equilibrium if it satisfies $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ and $F(\bar{x}, \bar{y}) \leq \min_{x \in \overline{X}} \max_{y \in S(x)} F(x, y)$. Each such point satisfies

$$\min_{x \in \overline{X}} \min_{y \in S(x)} F(x, y) \leq F(\bar{x}, \bar{y}) \leq \min_{x \in \overline{X}} \max_{y \in S(x)} F(x, y)$$

[17, Rem. 3.2] and, thus, covers the notions of, both, optimistic and pessimistic solutions. In [17] it is shown that lower Stackelberg equilibria are stable under perturbations of the definition functions, as opposed to merely optimistic or pessimistic solutions.

2.3 Bicriteria and vector optimization

This section briefly discusses optimization problems with two objective functions, also called bicriteria problems, and their generalization to vector optimization problems. Bicriteria problems are typically not considered as coupled pairs of optimization problems, since they are not used to model interrelated decisions of two agents, but rather to model the decision of a single agent who is confronted with two conflicting objectives. Still, in Section 5.4 we will report a connection between pairs of symmetrically coupled optimization problems and vector optimization problems.

A general bicriteria optimization is given by a host set $\mathcal{X} \subseteq \mathbb{R}^n$, a feasible set $M \subseteq \mathcal{X}$ and two objective functions $F, f : M \to \mathbb{R}$. Its statement

$$\min_{x \in \mathcal{X}} \begin{pmatrix} F(x) \\ f(x) \end{pmatrix} \quad \text{s.t.} \quad x \in M$$

requires an appropriate definition of minimality. Since the image space $\mathbb{R}^2$ of the two-dimensional objective function cannot be equipped with a total order (respecting its vector space structure), this is not straightforward.

Appropriate generalizations from the single-objective case are based on the negative formulation of global minimality, namely the lack of feasible points with strictly better objective values. Indeed, a point $\bar{x} \in M$ is called weakly efficient (or weakly Pareto optimal) if no $x \in M$ satisfies $(F(x), f(x)) < (F(\bar{x}), f(\bar{x}))$, where the inequality sign is meant componentwise. This notion covers points $\bar{x} \in M$ for which still $x \in M$ with, e.g., $F(x) = F(\bar{x})$ and $f(x) < f(\bar{x})$ exist. Since a decision maker would prefer any such x over $\bar{x}$, some weakly efficient points may not be interesting in practice.

Therefore, as a second, more appropriate generalization from the single-objective setting, a point $\bar{x} \in M$ is called efficient (or Pareto optimal) if no $x \in M$ satisfies $(F(x), f(x)) \leq (F(\bar{x}), f(\bar{x}))$ and $(F(x), f(x)) \neq (F(\bar{x}), f(\bar{x}))$, where the inequality sign is again meant componentwise. In contrast to single-objective optimization, the set of efficient points is not even a singleton under regularity assumptions, but in general an infinite set. From this the decision maker may choose according to her subjective preferences (if any).

To mention a pair of optimization problems arising in bicriteria optimization, let F and f both be bounded from below on M, let y_1 be the optimal value of the minimization of F over M, and let y_2 be the optimal value of the minimization of f over M. Then the point $y = (y_1, y_2)$ in the image space $\mathbb{R}^2$ is called the ideal point of the bicriteria problem. The ideal point may be attained as $(F(\bar{x}), f(\bar{x}))$ with some $\bar{x} \in M$ only for consenting objectives. In this case, however, $\bar{x}$ can also be found by minimizing either one of the two objectives, that is, by solving some single-objective problem. Such situations are not relevant for bicriteria optimization. For conflicting objectives, on the other hand, the ideal point cannot be attained. It still plays an important role in the formulation of some scalarization approaches for the determination of efficient points, e.g., in the compromise approach [24, 56].

For more than two objective functions, say $\theta_1, \ldots, \theta_k$, (weak) efficiency of $\bar{x} \in M$ can be defined analogously. With the function vector $\theta = (\theta_1, \ldots, \theta_k)$ an equivalent formulation for efficiency of $\bar{x} \in M$ requires $\theta(\bar{x}) \leq \theta(x)$ for all $x \in M$ with $\theta(x) \leq \theta(\bar{x})$. Since both of these two conditions involve the same binary relation between vectors in $\mathbb{R}^k$, it can be generalized to other binary relations. Indeed, in Section 5.4 we will use that every cone $C \subseteq \mathbb{R}^k$ induces a binary relation on $\mathbb{R}^k$ by virtue of the definition

$$y^1 \leq_C y^2 \;:\Leftrightarrow\; y^2 - y^1 \in C.$$

A partial order on $\mathbb{R}^k$ is defined by a binary relation which is reflexive, transitive and antisymmetric. The binary relation $\leq_C$ is reflexive if and only if $0 \in C$ holds, transitive if and only if C is convex, and antisymmetric if and only if C is pointed [24, Th. 1.20]. Hence, with the k-dimensional standard ordering cone $\mathbb{R}^k_{\geq} = \{y \in \mathbb{R}^k \mid y \geq 0\}$ the binary relation $\leq_{\mathbb{R}^k_{\geq}}$ induces a partial order on $\mathbb{R}^k$.

Efficiency can be defined with respect to any binary relation, not only ones leading to a partial order. In fact, in Section 5.4 we will use a nonconvex ordering cone C with $0 \in C$ and call $\bar{x} \in M$ efficient with respect to C if $\theta(\bar{x}) \leq_C \theta(x)$ holds for all $x \in M$ with $\theta(x) \leq_C \theta(\bar{x})$. In the case $C = \mathbb{R}^k_{\geq}$ one speaks of multi-objective or multicriteria optimization, whereas optimization with respect to a general ordering cone on a possibly infinite-dimensional host set $\mathcal{X}$ is referred to as vector optimization.

A generalized definition of weak efficiency requires an ordering cone C with nonempty interior $\operatorname{int} C$ and defines $y^1 <_C y^2$ by $y^2 - y^1 \in \operatorname{int} C$. Then $\bar{x} \in M$ is called weakly efficient if no $x \in M$ with $\theta(x) <_C \theta(\bar{x})$ exists.

3 Special symmetric couplings

3.1 Equilibria and variational inequalities

In this section we assume that the problems $P(y)$, $y \in \mathcal{Y}$, and $Q(x)$, $x \in \mathcal{X}$, are smooth and convex in the following sense.

Assumption 3.1. *The following two properties hold:*

(a) For all $y \in \mathcal{Y}$ the function $F(\cdot, y)$ lies in $C^1(\mathcal{X}, \mathbb{R})$ and is convex, and the set $X(y)$ is nonempty and convex.
(b) For all $x \in \mathcal{X}$ the function $f(x, \cdot)$ lies in $C^1(\mathcal{Y}, \mathbb{R})$ and is convex, and the set $Y(x)$ is nonempty and convex.

In the game theoretic setting, the convexity requirements from Assumption 3.1 are referred to as player convexity. It follows, in particular, if the pair $(P(y), Q(x))$ enjoys joint convexity, that is, X and Y are derived from a closed and convex shared constraint set Ω via (2.6) and (2.7), respectively, and F and f are convex on Ω.

By the variational formulation of convex optimization problems (cf., e.g., [69, Th. 2.8.2]), also known as the minimum principle [26], Assumption 3.1 guarantees for each parameter $y \in \mathcal{Y}$ that the elements of the optimal point set $R(y)$ of $P(y)$ are exactly the points which solve the variational inequality (VI)

$$\text{find } \bar{x} \in X(y) \text{ with } \langle \nabla_x F(\bar{x}, y), x - \bar{x} \rangle \geq 0 \quad \forall\, x \in X(y) \tag{3.1}$$

and, analogously, for each $x \in \mathcal{X}$ the elements of the optimal point set $S(x)$ of $Q(x)$ are characterized as the solutions of the VI

$$\text{find } \bar{y} \in Y(x) \text{ with } \langle \nabla_y f(x, \bar{y}), y - \bar{y} \rangle \geq 0 \quad \forall\, y \in Y(x). \tag{3.2}$$

The equilibria $(\bar{x}, \bar{y})$ of $(P(y), Q(x))$ are thus characterized as the simultaneous solutions of (3.1) with parameter $y := \bar{y}$ and (3.2) with parameter $x := \bar{x}$. With the set-valued mapping $\Omega : \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{X} \times \mathcal{Y}, z := (x, y) \to X(y) \times Y(x)$ these two VIs may be aggregated to

$$\text{find } \bar{z} \in \Omega(\bar{z}) \text{ with } \left\langle \begin{pmatrix} \nabla_x F(\bar{z}) \\ \nabla_y f(\bar{z}) \end{pmatrix}, z - \bar{z} \right\rangle \geq 0 \quad \forall\, z \in \Omega(\bar{z}). \tag{3.3}$$

While, like X in (3.1) and Y in (3.2), the feasible set Ω in (3.3) is not constant, a crucial difference is that it depends on the decision variable $\bar{z}$, rather than on a parameter. Therefore, (3.3) is not a parameter-dependent standard VI, but a quasi-variational inequality (QVI).

Conversely, the QVI (3.3) can also be de-aggregated into (3.1) and (3.2) in the sense that for each solution $\bar{z} = (\bar{x}, \bar{y})$ of (3.3) the point $\bar{x}$ solves (3.1), and $\bar{y}$ solves (3.2). This readily follows from the particular choices $z = (x, \bar{y})$ and $z = (\bar{x}, y)$ in (3.3).

This shows the following result.

Theorem 3.2 ([26, Prop. 1.4.2])**.** *Under Assumption 3.1 the set E of equilibrium points of the pair $(P(y), Q(x))$ coincides with the solutions of the QVI* (3.3).

A more general definition of equilibrium problems is studied in [28] (see also [15, 39]) by generalizing the structure of a standard VI to

$$(EP) \qquad \text{find } \bar{z} \in Z \text{ with } T(z, \bar{z}) \geq 0 \quad \forall\, z \in Z$$

for some some closed convex set Z and a function $T : Z \times Z \to \mathbb{R}$ with $T(z, z) = 0$ for all $z \in Z$. Beyond variational inequalities, optimization problems, and Nash equilibrium problems, (EP) also covers complementarity problems, fixed point problems, and vector optimization problems [39]. The latter is briefly elaborated in Appendix A.

3.2 Symmetrically coupled pairs with a potential function

In [59] Monderer and Shapley transferred the idea of the potential of a vector field from physics to Nash equilibrium problems. We formulate this concept for a pair $(P(x), Q(y))$ of symmetrically coupled optimization problems with $(x, y) \in \mathcal{X} \times \mathcal{Y} =: \mathcal{Z} \subseteq \mathbb{R}^{n+m}$.

Indeed, from the physics perspective a vector field $G : \mathcal{Z} \to \mathbb{R}^{n+m}$ is said to be conservative if a function $\Psi \in C^1(\mathcal{Z}, \mathbb{R})$ with $G = \nabla\Psi$ exists, and Ψ is then called a potential function for G on $\mathcal{Z}$. For a simply connected domain $\mathcal{Z}$ a vector field $G \in C^1(\mathcal{Z}, \mathbb{R}^n)$ is conservative if and only if the Jacobian DG is symmetric on $\mathcal{Z}$ [63, Ths. 10.39,10.40], [26, Th. 1.3.1].

In application to a pair $(P(x), Q(y))$ of symmetrically coupled optimization problems with $(x, y) \in \mathcal{Z}$, let $F, f \in C^2(\mathcal{Z}, \mathbb{R})$, and let $\mathcal{Z}$ be simply connected. Then a potential function $\Psi : \mathcal{Z} \to \mathbb{R}$ for the vector field

$$G(z) := \begin{pmatrix} \nabla_x F(z) \\ \nabla_y f(z) \end{pmatrix}$$

on $\mathcal{Z}$ satisfies

$$\forall\, (x, y) \in \mathcal{Z} : \quad \nabla_x F(x, y) = \nabla_x \Psi(x, y), \tag{3.4}$$

$$\forall\, (x, y) \in \mathcal{Z} : \quad \nabla_y f(x, y) = \nabla_y \Psi(x, y), \tag{3.5}$$

and such a function exists if and only if the matrix

$$DG(z) = \begin{pmatrix} D_x^2 F(z) & D_y \nabla_x F(z) \\ D_x \nabla_y f(z) & D_y^2 f(z) \end{pmatrix}$$

is symmetric for all $z \in \mathcal{Z}$. The latter is characterized by

$$D_y \nabla_x F(z) = (D_x \nabla_y f(z))^\intercal \;\text{ for all }\; z \in \mathcal{Z}. \tag{3.6}$$

By the fundamental theorem of calculus, (3.4) and(3.5) are equivalent to

$$\forall\, \bar{x}, x \in \mathcal{X}, \bar{y} \in \mathcal{Y} : \quad F(\bar{x}, \bar{y}) - F(x, \bar{y}) = \Psi(\bar{x}, \bar{y}) - \Psi(x, \bar{y}),$$

$$\forall\, \bar{x} \in \mathcal{X}, \bar{y}, y \in \mathcal{Y} : \quad f(\bar{x}, \bar{y}) - f(\bar{x}, y) = \Psi(\bar{x}, \bar{y}) - \Psi(\bar{x}, y),$$

which is the formulation presented in [59] for a merely continuous potential function Ψ. Thus, in a game theoretic setting, the existence of a potential function may be interpreted as a situation in which both players minimize the same objective function in their respective decision variables. Observe that a potential function, if it exists, is uniquely determined up to an additive constant.

To render the existence of a potential function useful for the computation of equilibria, one may combine the first order information from (3.4) and (3.5) with the convexity requirements from Assumption 3.1 and employ Theorem 3.2. It then remains to find conditions under which (3.3) becomes a standard VI which characterizes the optimal points of some convex optimization problem.

Indeed, to obtain a standard VI we assume constant feasible sets $X(y) \equiv \overline{X} \subseteq \mathcal{X}$ and $Y(x) \equiv \overline{Y} \subseteq \mathcal{Y}$ like in a standard NEP (as opposed to a GNEP). This yields $\Omega(z) \equiv \overline{\Omega} := \overline{X} \times \overline{Y}$ with $\overline{\Omega}$ being nonempty and convex, so that (3.3) reduces to the standard VI

$$\text{find} \quad \bar{z} \in \overline{\Omega} \quad \text{with} \quad \left\langle \begin{pmatrix} \nabla_x F(\bar{z}) \\ \nabla_y f(\bar{z}) \end{pmatrix}, z - \bar{z} \right\rangle \geq 0 \quad \forall\, z \in \overline{\Omega}. \tag{3.7}$$

Since $\overline{X}$ and $\overline{Y}$ are constant feasible sets, for simplicity we may put $\mathcal{X} := \overline{X}$ and $\mathcal{Y} := \overline{Y}$. The convexity of $\overline{\Omega}$ then implies that $\mathcal{Z} = \overline{\Omega}$ is simply connected, so that a potential function Ψ exists if and only if (3.6) is satisfied. In this case the VI (3.7) becomes

$$\text{find} \quad \bar{z} \in \overline{\Omega} \quad \text{with} \quad \langle \nabla \Psi(\bar{z}), z - \bar{z} \rangle \geq 0 \quad \forall\, z \in \overline{\Omega}. \tag{3.8}$$

To render (3.8) the variational formulation of a convex optimization problem

$$\min_z \ \Psi(z) \quad \text{s.t.} \quad z \in \overline{\Omega},$$

one additionally needs that Ψ is convex. Since with F and f also Ψ is a C^2-function, it is convex if its Hessian is positive semi-definite on $\overline{\Omega}$. This shows the following result.

Theorem 3.3. *For a pair $(P(x), Q(y))$ of symmetrically coupled optimization problems with constant feasible sets $\mathcal{X} = \overline{X}$, $\mathcal{Y} = \overline{Y}$ and $(x, y) \in \overline{\Omega} = \overline{X} \times \overline{Y}$, let Assumption 3.1 hold with $F, f \in C^2(\overline{\Omega}, \mathbb{R})$, and let*

$$\begin{pmatrix} D_x^2 F(z) & D_y \nabla_x F(z) \\ D_x \nabla_y f(z) & D_y^2 f(z) \end{pmatrix}$$

be symmetric and positive semi-definite on $\overline{\Omega}$. Then the set E of equilibria coincides with the minimal point set of Ψ on $\overline{\Omega}$.

Theorem 3.3 justifies to call Ψ a potential function for the symmetrically coupled pair $(P(y), Q(x))$. In the game theoretic setting, one speaks of a potential game in this case. In view of Theorem 3.3, equilibria of symmetrically coupled pairs of optimization can be determined using the well-developed algorithmic tools from convex optimization [16], provided that Ψ can be easily computed from F and f by integration. Otherwise, one may apply solution methods for monotone VIs [26, Ch. 10] directly to (3.8).

3.3 Saddle problems

Saddle points are a key concept in minimax theory and for duality results. Their standard definition assumes constant sets $\overline{X} \subseteq \mathcal{X}$ and $\overline{Y} \subseteq \mathcal{Y}$. A point $\bar{z} = (\bar{x}, \bar{y}) \in \overline{\Omega} = \overline{X} \times \overline{Y}$ is called a saddle point of F on $\overline{\Omega}$ if

$$\forall\, x \in \overline{X}, y \in \overline{Y} :\; F(\bar{x}, y) \le F(\bar{x}, \bar{y}) \le F(x, \bar{y}) \tag{3.9}$$

holds, and the task to find a saddle point of F on $\overline{\Omega}$ is called a saddle problem. The condition (3.9) is equivalent to

$$\max_{y \in \overline{Y}} F(\bar{x}, y) = F(\bar{x}, \bar{y}) = \min_{x \in \overline{X}} F(x, \bar{y}). \tag{3.10}$$

In particular the maximal and minimal values are attained at $\bar{y}$ and $\bar{x}$, respectively. Thus, $\bar{z}$ is a saddle point of F on $\overline{\Omega}$ if and only if it is an equilibrium point of the pair of symmetrically coupled optimization problems $(P(y), Q(x))$ with constant feasible sets $X(y) \equiv \overline{X} \subseteq \mathcal{X}$ and $Y(x) \equiv \overline{Y} \subseteq \mathcal{Y}$ as well as the special choice $f := -F$. Due to $f + F = 0$, in a game theoretic setting this situation is called a two-person zero sum game.

The symmetric structure of saddle problems admits an important characterization of their equilibrium points. The main idea is study the connection of the equilibrium points of the underlying pair of symmetrically coupled optimization problems to the solutions of two asymmetrically coupled pairs. We postpone the discussion of this connection to Section 5.1.

3.4 Generalized saddle problems

In a generalized version, one may also consider saddle points of a function F on a nonempty set $\Omega \subseteq \mathcal{X} \times \mathcal{Y}$ which does not necessarily possess the Cartesian product structure $\overline{X} \times \overline{Y}$. Note, however, that the decomposition of the variable vector z into the variable groups x and y is prescribed in this concept. Following the definition from, e.g., [22], $(\bar{x}, \bar{y}) \in \Omega$ is called a saddle point of F on Ω if

$$\forall (\bar{x}, y), (x, \bar{y}) \in \Omega :\; F(\bar{x}, y) \le F(\bar{x}, \bar{y}) \le F(x, \bar{y}) \tag{3.11}$$

holds. With the fibers $X(y)$ and $Y(x)$ of Ω from (2.6) and (2.7), respectively, (3.11) can be rewritten as

$$\max_{y \in Y(\bar{x})} F(\bar{x}, y) = F(\bar{x}, \bar{y}) = \min_{x \in X(\bar{y})} F(x, \bar{y}). \tag{3.12}$$

Therefore $(\bar{x}, \bar{y}) \in \Omega$ is a saddle point in this generalized setting if and only if it is an equilibrium of the corresponding pair $(P(y), Q(x))$ of symmetrically coupled optimization problems with $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where X and Y are derived from the shared constraint set Ω.

We remark that in the literature (cf., e.g., [74]) one also finds a different definition of saddle points for sets Ω not possessing Cartesian product structure. There $(\bar{x}, \bar{y}) \in \Omega$

is called a saddle point of F on Ω if, rather than (3.11),

$$\forall (x,y) \in \Omega: \; F(\bar{x},y) \leq F(\bar{x},\bar{y}) \leq F(x,\bar{y}) \tag{3.13}$$

holds. Simple examples, even with compact convex sets Ω, show that the saddle point definitions based on (3.11) and (3.13) are unrelated. In particular, the one using (3.13) is not consistent with the equilibrium concept for pairs of symmetrically coupled optimization problems.

From the game theoretic perspective, the saddle problem with $\overline{\Omega} = \overline{X} \times \overline{Y}$ corresponds to a two-person standard NEP, whereas the generalized saddle problem with $\Omega \subseteq \mathcal{X} \times \mathcal{Y}$ and (3.11) possesses the structure of a two-person GNEP, both with the special choice $f = -F$.

A straightforward characterization of generalized saddle points in analogy to the one discussed in Section 5.1 for standard saddle points is not possible, as will be explained there.

3.5 Local saddle points

In addition to the global versions of saddle points discussed so far, there exist at least two concepts for local versions. Firstly, in (3.10) and (3.12) the global minimality of $\bar{x}$ and global maximality of $\bar{y}$ can straightforwardly be replaced by local minimality and maximality, respectively, to define a (generalized) local saddle point. This admits, for example, the development of a local Lagrange duality theory as in [51]. Let us call this a local saddle point of type 1.

Secondly, by (3.12), a generalized saddle point $\bar{z} = (\bar{x}, \bar{y})$ of F on Ω is a global minimal point in x, a global maximal point in y, but (except for trivial cases) neither a global minimal nor a global maximal point of F in z. Thus, an interior point $\bar{z}$ of $\Omega \subseteq \mathbb{R}^N$ can be called a local saddle point of a C^1-function F if it satisfies $\nabla F(\bar{z}) = 0$, but is neither a local minimal nor a local maximal point. For the sake of comparison, we call this concept a local saddle point of type 2. It generalizes the notion of a saddle point from elementary calculus, where for $N = 1$ and $\Omega = \mathbb{R}$ the function $F(z) = z^3$ possesses a saddle point of type 2 at $\bar{z} = 0$. For $N > 1$ local saddle points of type 2 play a prominent role in the global structural analysis of functions [40].

As a major deviation from the concept of local saddle points of type 1, in addition to the smoothness assumption and the employment of a first order optimality condition, for local saddle points of type 2 no variable groups are prescribed. In particular, as opposed to local saddle points of type 1, this concept makes sense for univariate functions F. For a more detailed comparison of the two concepts see Appendix B.

4 Special asymmetric couplings

4.1 Semi-infinite and robust optimization

In this section we specialize the problem $(P(y))$ with parameter $y \in \mathcal{Y}$ to a y-independent objective function $F : \mathcal{X} \to \mathbb{R}$ and to the feasible set mapping

$$X : \mathcal{Y} \rightrightarrows \mathcal{X}, \; y \mapsto \{x \in \overline{X} \mid f(x,y) \geq 0\},$$

which employs some set $\overline{X} \subseteq \mathcal{X}$ and the objective function f of the problem $(Q(x))$. Then the strong and weak versions of the asymmetrically coupled problems from Section 2.2.1 are

$$\min_{x \in \mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ \forall\, y \in S(x):\ f(x,y) \geq 0 \tag{4.1}$$

and

$$\min_{x \in \mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ \exists\, y \in S(x):\ f(x,y) \geq 0, \tag{4.2}$$

respectively. For each $y \in S(x)$ one has $f(x,y) = \varphi(x)$ with $\varphi(x) = \inf_{y \in Y(x)} f(x,y)$ from (2.1). For simplicity let us assume $\mathcal{X} \subseteq \operatorname{dom} S := \{x \in \mathcal{X} \mid S(x) \neq \emptyset\}$ (the effective domain of S), that is, $Q(x)$ is solvable for all $x \in \mathcal{X}$ (e.g., under the assumptions of the Weierstrass theorem). Then φ is real-valued, and (4.1), (4.2) are both equivalent to

$$\min_{x \in \mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ \varphi(x) \geq 0. \tag{4.3}$$

In particular, the strong and weak versions of this bilevel problem are identical. In a more explicit form, (4.3) can be stated as

$$(GSIP) \qquad \min_{x \in \mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ f(x,y) \geq 0 \ \forall\, y \in Y(x).$$

Observe that under the assumption $\mathcal{X} \subseteq \operatorname{dom} S$ the set $Y(x)$ is nonempty for all $x \in \mathcal{X}$.

The problem $(GSIP)$ employs potentially infinitely many inequality constraints for a finite-dimensional decision variable x, which explains that it is called a generalized semi-infinite optimization problem [68]. The feasible set $Y(x)$ of $Q(x)$ takes the role of the index set of the inequality constraints. In the case of a constant index set $\overline{Y} \subseteq \mathcal{Y}$ one speaks of a standard semi-infinite optimization problem (SIP) [33, 34, 36, 48].

Since the standard weak version (P_{sw}) corresponding to (4.2) is

$$\min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ f(x,y) \geq 0,\ y \in S(x),$$

semi-infinite problems can be studied from the point of view of bilevel optimization [68].

One application of semi-infinite optimization (among many others, [23, 68]) is the treatment of uncertainty in an optimization problem [14]. While (two-stage) stochastic optimization uses information about the underlying distribution of the uncertainty (Section 4.4), the worst-case approach is called robust optimization, and it is strongly related to semi-infinite optimization.

Indeed, assume that a nominal parameter $y_0 \in \mathbb{R}^m$ in the inequality constraint of the optimization problem

$$\min_{x \in \mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ f(x,y_0) \geq 0 \tag{4.4}$$

is uncertain, but that at least an uncertainty set $\overline{Y} \subseteq \mathbb{R}^m$ containing y_0 is known. Then the robust counterpart of (4.4) is

$$\min_{x\in\mathcal{X}} F(x) \quad \text{s.t.} \quad x \in \overline{X},\ f(x,y) \geq 0 \ \forall\, y \in \overline{Y}, \tag{4.5}$$

that is, a standard SIP. As above, the infinitely many inequality constraints in (4.5) may be rewritten as the the worst-case constraint $\varphi(x) = \inf_{y\in\overline{Y}} f(x,y) \geq 0$. Likewise, x-dependent uncertainty sets $Y(x)$ give rise to a GSIP.

Hence, robust counterparts of optimization problems with uncertainties in inequality constraints coincide with semi-infinite optimization models. Nevertheless, considerable attention has been paid to robust optimization in its own right [8, 13], often with a focus on algorithmically beneficial properties of the optimal value function φ (e.g., for linear functions $f(x,\cdot)$, $x \in \mathcal{X}$, and an ellipsoidal set $\overline{Y}$ [7]).

4.2 Minimax problems

With a constant set $\overline{X} \subseteq \mathcal{X}$ as well as the objective function F and the set-valued mapping Y from Section 2 the problem

$$(MM) \qquad \min_{x\in\mathcal{X}} \sup_{y\in Y(x)} F(x,y) \quad \text{s.t.} \quad x \in \overline{X}$$

is called a minimax problem. Observe that the inner maximization of the objective function is performed with respect to $y \in Y(x)$, and not $y \in S(x)$ as in the original strong version (P_s) of a bilevel optimization problem. Therefore, while (P_s) possesses a three-level structure, (MM) is a bilevel problem. Indeed, it is equivalent [70] to its epigraph reformulation

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad x \in \overline{X},\ \sup_{y\in Y(x)} F(x,y) \leq z$$

which, more explicitly, turns out to be the GSIP (Section 4.1)

$$\min_{(x,z)\in\mathcal{X}\times\mathbb{R}} z \quad \text{s.t.} \quad x \in \overline{X},\ z - F(x,y) \geq 0 \ \forall\, y \in Y(x).$$

We emphasize that minimax problems (MM) belong to the class of asymmetrically coupled pairs of optimization problems and must, hence, not be mistaken for saddle problems. Nevertheless, an important connection between the two classes exists and is discussed in Section 5.1.

From the point of view of robust optimization, the arguments from Section 4.1 regarding an uncertain nominal parameter $y_0 \in Y(x)$ in an inequality constraint readily transfer to an uncertain parameter in the objective function. Indeed, in terms of minimization, $\sup_{y\in Y(x)} F(x,y)$ is then the worst-case objective function. Hence (MM) is the robust counterpart for this case.

4.3 Decomposition

We return to the setting from Section 3.4, that is, a real-valued function f on a set $\Omega \subseteq \mathcal{X} \times \mathcal{Y}$ with the fibers $Y(x) = \{y \in \mathcal{Y} \mid (x, y) \in \Omega\}$, $x \in \mathcal{X}$, from (2.7). The (effective) domain

$$\operatorname{dom} Y = \{x \in \mathcal{X} \mid Y(x) \neq \emptyset\}$$

of the set-valued mapping Y allows the description

$$\Omega = \bigcup_{x \in \operatorname{dom} Y} (\{x\} \times Y(x)),$$

which yields the decomposition of the minimal value of f on Ω into

$$\inf_{(x,y)\in\Omega} f(x, y) = \inf_{x\in\operatorname{dom} Y} \inf_{y\in Y(x)} f(x, y).$$

This means that $\inf_{(x,y)\in\Omega} F(x, y)$ may be determined as the minimal value of the problem

$$\min_{x\in\mathcal{X}} \varphi(x) \quad \text{s.t.} \quad x \in \operatorname{dom} Y \tag{4.6}$$

with the minimal value $\varphi(x) = \inf_{y\in Y(x)} f(x, y)$ of $Q(x)$ from (2.1). This can be useful if the determination of $\varphi(x)$ for each $x \in \operatorname{dom} Y$ is easier to handle algorithmically than the direct minimization of F on Ω, e.g., if $Q(x)$ is a linear or a (smooth) convex problem for each $x \in \operatorname{dom} Y$. In addition, explicit descriptions or approximations of φ and $\operatorname{dom} Y$ need to be available for the solution of (4.6).

Indeed, one main idea of (generalized) Benders decomposition [9, 32] is to secure assumptions under which strong duality holds between $Q(x)$ and its dual problem (Section 5.2). This reformulates (4.6) into a minimax problem (Section 4.2), which can be treated as a semi-infinite problem (Section 4.1).

The historically first application of Benders decomposition [9] focused on linear mixed-integer optimization problems with discrete variables x and continuous variables y. Then $Q(x)$ is a linear optimization problem for each $x \in \operatorname{dom} Y$, and dual descriptions of both $\operatorname{dom} Y$ and φ can be derived. In the resulting semi-infinite reformulation of the minimax problem the polyhedral index set can be replaced by its vertex set, due to the vertex theorem of linear optimization. This reduces the semi-infinite to a finite problem with, however, a possibly vast number of constraints (corresponding to the size of the vertex set of the dual feasible set). These are handled by considering a relaxed master problem with few constraints and, depending on its solution, successively adding new constraints in the fashion of a cutting plane method.

It turns out that the same ideas can also be used for the solution of two-stage stochastic optimization problems (Section 4.4), where they lead to the L-shaped method.

4.4 Two-stage stochastic optimization

A standard model in stochastic optimization [14] assumes that the decision variables can be split into a vector x which does not depend on stochastic information and a vector y whose optimal choice can only be determined after the realization of a random variable ξ has been revealed. Therefore a first-stage decision about $x \in \overline{X}$ (the here-and-now decision) can be furnished independently of the outcome of ξ, and a second-stage decision on $y \in Y_\xi(x)$ (the recourse decision) after the realization of ξ, and given the first-stage decision x.

For the formulation of the second-stage problem we introduce the random variable ξ into the objective function and feasible set of the problem $(Q(x))$, and thus obtain

$$(Q_\xi(x)) \qquad \min_{y \in \mathcal{Y}} f_\xi(x, y) \quad \text{s.t.} \quad y \in Y_\xi(x)$$

with the optimal value function $\varphi_\xi(x)$.

For the statement of a deterministic first-stage problem, the expected value of $\varphi_\xi(x)$ is employed. To avoid integrability issues and to promote the algorithmic tractability, one usually assumes a discrete and finite random variable with probabilities p_s of the scenarios ξ_s, $s \in [\sigma] := \{1, \ldots, \sigma\}$ (or a continuous random variable which is approximated by sampling these finitely many scenarios). This yields the expected value function

$$\mathbb{E}_\xi \varphi_\xi(x) = \sum_{s \in [\sigma]} p_s \varphi_{\xi_s}(x)$$

of the second-stage problem with the deterministic optimal values $\varphi_{\xi_s}(x)$ of the deterministic problems $Q_{\xi_s}(x)$, $s \in [\sigma]$. The choice $x \in D := \bigcap_{s \in [\sigma]} \operatorname{dom} Y_{\xi_s}$ with $\operatorname{dom} Y_{\xi_s} = \{x \in \overline{X} \mid Y_{\xi_s}(x) \neq \emptyset\}$, $s \in [\sigma]$, secures $\mathbb{E}_\xi \varphi_\xi(x) < +\infty$.

The resulting first-stage problem

$$(FS) \qquad \min_{x \in \mathcal{X}} \sum_{s \in [\sigma]} p_s \varphi_{\xi_s}(x) \quad \text{s.t.} \quad x \in D$$

is a bilevel problem with finitely many lower level problems, namely the scenario-based problems $Q_{\xi_s}(x)$, $s \in [\sigma]$. For the algorithmic treatment of (FS) this entails the introduction of one lower level decision variable y_s for each scenario s, so that for a large number σ of scenarios the problem size may become vast.

Fortunately, under the assumption that the problems $Q_{\xi_s}(x)$, $s \in [\sigma]$, are linear optimization problems, the dualization and relaxation ideas of Benders decomposition (Section 4.3) allow to solve the two-stage stochastic optimization problem by a cutting plane method (the L-shaped method), controlled by solutions of an appropriate master problem [14].

One may wonder whether (FS) also results from decomposition (Section 4.3) of some single-level optimization problem with the large number of decision variables

$x, y_1, \ldots, y_\sigma$. Indeed, with the σ-fold Cartesian product $\mathcal{Y}^\sigma := \mathcal{Y} \times \ldots \times \mathcal{Y}$ the problem

$$(EF) \qquad \min_{(x,y_1,\ldots,y_\sigma)\in\mathcal{X}\times\mathcal{Y}^\sigma} \sum_{s\in[\sigma]} p_s f_{\xi_s}(x, y_s) \quad \text{s.t.} \quad x \in \overline{X},\ y_s \in Y_{\xi_s}(x),\ s \in [\sigma],$$

is called the extensive form of the scenario-based two-stage stochastic optimization problem. With the graphs

$$\operatorname{gph} Y_{\xi_s} = \{(x, y_1, \ldots, y_\sigma) \in \overline{X} \times \mathcal{Y}^\sigma \mid y_s \in Y_{\xi_s}(x)\},\ s \in [\sigma],$$

of the scenario-wise feasible set mappings $Y_{\xi_s} : \overline{X} \to \mathcal{Y}^\sigma$, the set Ω from Section 4.3 is $\Omega = \bigcap_{s\in[\sigma]} \operatorname{gph} Y_{\xi_s}$. For $x \in \overline{X}$ its fibers are

$$\begin{aligned} Y(x) &= \{(y_1, \ldots, y_\sigma) \in \mathcal{Y}^\sigma \mid (x, y_1, \ldots, y_\sigma) \in \Omega\} \\ &= \{(y_1, \ldots, y_\sigma) \in \mathcal{Y}^\sigma \mid y_s \in Y_{\xi_s}(x),\ s \in [\sigma]\}, \\ &= Y_{\xi_1}(x) \times \ldots \times Y_{\xi_\sigma}(x), \end{aligned}$$

which yields $\operatorname{dom} Y = \bigcap_{s\in[\sigma]} \operatorname{dom} Y_{\xi_s} = D$. Decomposition thus allows to write the minimal value of (EF) as

$$\begin{aligned} \inf_{x\in\operatorname{dom} Y} \inf_{(y_1,\ldots,y_\sigma)\in Y(x)} \sum_{s\in[\sigma]} p_s f_{\xi_s}(x, y_s) &= \inf_{x\in D} \sum_{s\in[\sigma]} p_s \inf_{y_s\in Y_{\xi_s}(x)} f_{\xi_s}(x, y_s) \\ &= \inf_{x\in D} \sum_{s\in[\sigma]} p_s \varphi_{\xi_s}(x), \end{aligned} \tag{4.7}$$

that is, as the minimal value of the first-stage problem (FS). Hence, decomposition of the extensive form (EF) recovers the two-stage structure of the above recourse problem.

A crucial step in this derivation is the identity (4.7), where a rule of calculus for optimal values of separable functions on Cartesian products (cf., e.g., [69, Ex. 1.3.2]) is applied to the inner optimal value. This is only possible after the decomposition into the variable groups x and $(y_1, \ldots, y_\sigma)$, since Ω itself does not necessarily exhibit a Cartesian product structure. Indeed, this construction decouples the linking variable x from the large number of variables y_s, $s \in [\sigma]$, and the latter only appear in the independent lower level optimization problems $Q_{\xi_s}(x)$, $s \in [\sigma]$. This decomposition makes the possibly large-scale optimization problem (EF) algorithmically tractable. As a conceptually different single-level idea for two-stage stochastic optimization we mention the wait-and-see approach (see Appendix C for a brief explanation).

We emphasize that many other stochastic versions of bilevel problems exist (for references see [21, Sec. 20.4.1.5]). In particular, the one presented in [53] assumes that the leader of a Stackelberg game possesses stochastic information about the follower's choice from her non-unique optimal point set $S(x)$. In addition to the pessimistic and optimistic solutions, this allows the introduction of an intermediate Stackelberg strategy.

5 Interconnections between models

In Sections 3 and 4 we have seen that several models involving pairs of optimization problems may be interpreted as special cases of the general symmetrically and asymmetrically coupled pairs, respectively. In contrast, the focus of the present section is on connections between symmetrically and asymmetrically coupled models, and on their relations to vector optimization.

5.1 Characterization of saddle points

A fundamental connection between symmetrically and asymmetrically coupled pairs of optimization problems arises in a basic characterization of saddle points $(\bar{x}, \bar{y})$ of a function F on a set $\overline{X} \times \overline{Y}$ (Section 3.3). Indeed, the symmetric structure of a saddle problem raises the question whether the saddle value $F(\bar{x}, \bar{y})$ may be found by successive minimization of F over x and maximization over y, and even regardless of the order of these optimizations. More precisely, one asks whether the identities $\sup_{y \in \overline{Y}} \inf_{x \in \overline{X}} F(x, y) = F(\bar{x}, \bar{y}) = \inf_{x \in \overline{X}} \sup_{y \in \overline{Y}} F(x, y)$ are true. The two underlying asymmetrically coupled optimization problems are maximin and minimax problems, respectively (Section 4.2).

To formulate the connection between the symmetrically coupled and the two asymmetrically coupled pairs of optimization problems, recall that Φ and φ denote the minimal value functions of $P(y)$ and $Q(x)$, respectively. In the present case of constant feasible sets it suffices to consider $\overline{X}$ and $\overline{Y}$, respectively, as their domains, and we obtain

$$\begin{aligned} \Phi &: \overline{X} \to \overline{\mathbb{R}}, \ y \mapsto \inf_{x \in \overline{X}} F(x, y), \\ \varphi &: \overline{Y} \to \overline{\mathbb{R}}, \ x \mapsto \inf_{y \in \overline{Y}} (-F(x, y)) = -\sup_{y \in \overline{Y}} F(x, y). \end{aligned}$$

With $\psi := -\varphi$ the main idea of said characterization is to 'counter optimize' Φ and ψ, that is, to consider

$$\begin{aligned} \sup_{y \in \overline{Y}} \Phi(y) &= \sup_{y \in \overline{Y}} \inf_{x \in \overline{X}} F(x, y), \\ \inf_{x \in \overline{X}} \psi(x) &= \inf_{x \in \overline{X}} \sup_{y \in \overline{Y}} F(x, y), \end{aligned}$$

and to compare these (extended) real values. The following theorem formulates the central connection between these symmetrically and asymmetrically coupled pairs of optimization problems.

Theorem 5.1 ([26, Th. 1.4.1]). *For $F : \overline{X} \times \overline{Y} \to \mathbb{R}$ the minimax (or maximin) inequality*

$$\sup_{y \in \overline{Y}} \inf_{x \in \overline{X}} F(x, y) \leq \inf_{x \in \overline{X}} \sup_{y \in \overline{Y}} F(x, y) \tag{5.1}$$

holds, and for $(\bar{x}, \bar{y}) \in \overline{X} \times \overline{Y}$ *the following statements are equivalent.*

(a) $(\bar{x}, \bar{y})$ *is a saddle point of* F *on* $\overline{X} \times \overline{Y}$*.*
(b) $\bar{x}$ *is a global minimal point of* ψ *on* $\overline{X}$*,* $\bar{y}$ *is a global maximal point of* Φ *on* $\overline{Y}$*, and equality holds in* (5.1)*.*
(c) The identities $\psi(\bar{x}) = \Phi(\bar{y}) = F(\bar{x}, \bar{y})$ *are true.*

The assumptions of the following existence result for saddle points strengthen Assumption 3.1.

Theorem 5.2 ([26, Cor. 2.2.10])**.** *For nonempty, convex and compact sets* $\overline{X}$ *and* $\overline{Y}$ *let* F *be continuously differentiable on* $\overline{\Omega} = \overline{X} \times \overline{Y}$*, let* $F(\cdot, y)$ *be convex on* $\overline{X}$ *for each* $y \in \overline{Y}$*, and let* $F(x, \cdot)$ *be concave on* $\overline{Y}$ *for each* $x \in \overline{X}$*. Then the set of saddle points of* F *on* $\overline{\Omega}$ *is nonempty and compact, and the identity*

$$\min_{x \in \overline{X}} \max_{y \in \overline{Y}} F(x, y) = \max_{y \in \overline{Y}} \min_{x \in \overline{X}} F(x, y) \tag{5.2}$$

is true.

One may expect that the characterization of standard saddle points from Theorem 5.1 can be extended to the generalized case from Section 3.4 based on the generalized minimax inequality

$$\sup_{y \in \operatorname{dom} X} \inf_{x \in X(y)} F(x, y) \leq \inf_{x \in \operatorname{dom} Y} \sup_{y \in Y(x)} F(x, y). \tag{5.3}$$

However, the following example shows that (5.3) fails to hold in general, even for a convex-concave function F on a nonempty, polyhedral and compact set Ω. In cases where it does hold, the proof of Theorem 5.1 can be extended verbatim to the generalized case.

Example 5.3. *For* $F(x, y) = x^2 - y^2$ *on* $\Omega = \{(x, y) \in [0, 1]^2 \mid x + y \geq 1\}$ *one obtains* $\operatorname{dom} X = [0, 1]$*,* $X(y) = [1-y, 1]$ *and* $\inf_{x \in X(y)} F(x, y) = 1-2y$ *for all* $y \in \operatorname{dom} X$*, thus* $\sup_{y \in \operatorname{dom} X} \inf_{x \in X(y)} F(x, y) = 1$*. Analogously,* $\operatorname{dom} Y = [0, 1]$*,* $Y(x) = [1 - x, 1]$ *and* $\sup_{y \in Y(x)} F(x, y) = 2x - 1$ *for all* $x \in \operatorname{dom} Y$ *result in* $\inf_{x \in \operatorname{dom} Y} \sup_{y \in Y(x)} F(x, y) = -1$*. Therefore* (5.3) *is violated.*

5.2 Lagrange duality

A prominent example of a saddle problem lies at the heart of Lagrange duality in nonlinear optimization. Indeed, for the minimization of a function $\theta : \mathbb{R}^n \to \mathbb{R}$ over a feasible set $M = \{x \in \overline{X} \mid c(x) \leq 0,\ c_{eq}(x) = 0\}$, described by vector-valued constraint functions $c : \mathbb{R}^n \to \mathbb{R}^p$ and $c_{eq} : \mathbb{R}^n \to \mathbb{R}^q$, the Lagrange function aggregates all defining functions to

$$L :\ \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R},\ (x, \lambda, \mu) \mapsto \theta(x) + \langle \lambda, c(x) \rangle + \langle \mu, c_{eq}(x) \rangle.$$

With the combined multiplier vector $y := (\lambda, \mu)$ and the set $\overline{Y} := \mathbb{R}^p_\geq \times \mathbb{R}^q$ one verifies $v_P = \inf_{x\in\overline{X}} \sup_{y\in\overline{Y}} L(x,y)$ for the primal optimal value $v_P := \inf_{x\in M} \theta(x)$. This motivates to study saddle points of L on $\overline{X}\times\overline{Y}$.

Since, in the above notation, the primal optimal value satisfies $v_P = \inf_{x\in\overline{X}} \psi(x)$ with $\psi(x) = \sup_{y\in\overline{Y}} L(x,y)$, consequentially one also considers the maximization of $\Phi(y) = \inf_{x\in\overline{X}} L(x,y)$ over $\overline{Y}$ and its maximal value $v_D := \sup_{y\in\overline{Y}} \Phi(y)$. The latter maximization problem is called the Lagrange dual to the minimization of f over M, and v_D the dual optimal value.

The minimax inequality (5.1) becomes $v_D \leq v_P$ and is referred to as weak duality between the primal and dual problems. The thus non-negative term $v_P - v_D$ is called duality gap. In view of Theorem 5.1, $(\bar{x},\bar{y}) \in \overline{X}\times\overline{Y}$ is a saddle point of L on $\overline{X}\times\overline{Y}$ if and only if $\bar{x}$ is a global minimal point of the primal problem, $\bar{y}$ is a global maximal point of the dual problem, and $v_P = v_D$ holds, that is, the duality gap vanishes. The latter is referred to as strong duality. In view of the unboundedness of $\overline{Y}$, the existence of saddle points of L on $\overline{X}\times\overline{Y}$ cannot be deduced from Theorem 5.2. For appropriate existence results see, e.g., [12, Secs. 2.6, 6.2], [61, §36], [72].

Under certain convexity assumptions one can show that $(\bar{x},\bar{y})$ is a saddle point of L on $\overline{X}\times\overline{Y}$ if and only if $\bar{x}$ is a Karush-Kuhn-Tucker (KKT) point of the primal optimization problem with multiplier vector $\bar{y}$ (cf., e.g., [69] for details). Hence, KKT points can be interpreted as equilibrium points of a certain symmetrically coupled pair of optimization problems.

5.3 Nash equilibria and bilevel optimization

The article [46] provides assumptions under which the solutions of the standard weak (i.e., optimistic) version (P_{sw}) of a bilevel problem can be related to solutions of an appropriately formulated Nash equilibrium problem, that is, some symmetrically coupled pair of optimization problems. As in [46], for simplicity let us assume a constant feasible set $X(y) \equiv \overline{X} \subseteq \mathcal{X}$ in $(P(y))$ and consider the bilevel problem

$$\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} F(x,y) \quad \text{s.t.} \quad x\in\overline{X},\ y\in S(x)$$

from (2.9), where $S(x)$ denotes the optimal point set of $(Q(x))$. The relation $\overline{X} \subseteq \operatorname{dom} S$ is assumed to hold.

In [46] it is suggested to construct a Nash equilibrium problem from these data by coupling $(Q(x))$ symmetrically with the problem

$$(P'(y)) \qquad \min_{(x,\eta)\in\mathcal{X}\times\mathcal{Y}} F(x,\eta) \quad \text{s.t.} \quad x\in\overline{X},\ f(x,\eta)\leq f(x,y),\ \eta\in Y(x).$$

Observe that $(Q(x))$ parametrically depends only on the x-part of the decision variables of $(P'(y))$, but not on η.

To see the main idea of this insertion of the objective function f of $(Q(x))$ into the feasible set of $(P'(y))$, with the shared constraint set $\operatorname{gph} Y = \{(x,y)\in\overline{X}\times\mathcal{Y} \mid y\in Y(x)\}$ of (2.9) we write the feasible set of $(P'(y))$ succinctly as $X'(y) := \{(x,\eta)\in \operatorname{gph} Y\} \mid f(x,\eta)\leq f(x,y)\}$ and the one of (2.9) as $\{(x,\eta)\in\operatorname{gph} Y \mid f(x,\eta)\leq\varphi(x)\}$, where we have renamed the decision variable y of (2.9) to η. This shows that for each minimal point $\bar{y} \in S(x)$ with $x \in \overline{X}$ the feasible set of $(P'(\bar{y}))$ reduces to that of (2.9). Based on this observation, in [46] the following results are proved.

Proposition 5.4 ([46, Prop. 3.1])**.** *Let $(\bar{x}, \bar{\eta})$ be a global minimal point of* (2.9) *which minimizes F even over the whole set* $\operatorname{gph} Y$ *(i.e., $(\bar{x}, \bar{\eta})$ even solves the high point relaxation of* (2.9)*). Then for each $\bar{y} \in S(\bar{x})$ the triple $(\bar{x}, \bar{\eta}, \bar{y})$ is an equilibrium point of the symmetrically coupled pair of optimization problems* $(P'(y), Q(x))$.

Theorem 5.5 ([46, Cor. 3.1])**.** *Let $(Q(x))$ possess a constant feasible set mapping $Y(x) \equiv \overline{Y} \subseteq \mathcal{Y}$ on $\overline{X}$, and let $(\bar{x}, \bar{\eta}, \bar{y})$ be an equilibrium point of the pair $(P'(y), Q(x))$. Then $(\bar{x}, \bar{\eta})$ is a global minimal point of* (2.9).

The following result covers purely hierarchical bilevel problems.

Theorem 5.6 ([46, Th. 3.2])**.** *Let the optimal point set $S(x) \equiv \overline{S}$ be constant on $\overline{X} \cap \operatorname{dom} Y$. Then the following assertions hold.*

(a) For each equilibrium point $(\bar{x}, \bar{\eta}, \bar{y})$ of the pair $(P'(y), Q(x))$, the point $(\bar{x}, \bar{\eta})$ is globally minimal for (2.9).

(b) For each global minimal point $(\bar{x}, \bar{\eta})$ of (2.9) *and each $\bar{y} \in S$ the point $(\bar{x}, \bar{\eta}, \bar{y})$ is an equilibrium of the pair $(P'(y), Q(x))$.*

The relation between local minimal points of (2.9) and certain equilibria of the pair $(P'(y), Q(x))$ is more intricate, and elaborated in [46, Th. 3.3]. We remark that Theorem 5.6 is crucial for the mentioned reformulation (Section 2.2.1) of the standard strong version of a bilevel optimization problem to a bilevel problem in standard weak version with a parametric two-player GNEP in the lower level [47].

As another connection between Nash equilibria and bilevel optimization we mention that by [17, Th. 4.2], under weak assumptions, lower Stackelberg equilibria coincide with the outcomes of subgame perfect Nash equilibria. However, the latter concept interprets the two-step sequential nature of a Stackelberg game as a dynamic Nash game, which goes beyond the scope of the present survey.

5.4 Nash equilibria and vector optimization

In this section we consider a pair of symmetrically coupled optimization problems $(P(y), Q(x))$ with x and y from linear spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and with X and Y being derived from a shared constraint set $\Omega \subseteq \mathcal{X} \times \mathcal{Y}$ via (2.6) and (2.7). One may then ask whether the equilibrium points of $(P(y), Q(x))$ are related to the efficient points of the bicriteria optimization problem

$$\min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \begin{pmatrix} F(x,y) \\ f(x,y) \end{pmatrix} \quad \text{s.t.} \quad (x,y) \in \Omega \tag{5.4}$$

with respect to the standard ordering cone $C = \mathbb{R}^2_\geq$. The following example, inspired by [4, Ex. 1], shows that the two concepts are unrelated, even in the polyhedral case (see also [71]).

Example 5.7. *For $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ let*

$$\Omega = \{(x,y) \in \mathcal{X} \times \mathcal{Y} \mid x + y/2 \geq 2,\ -x/4 + y \leq 2,\\ x + y/2 \leq 8,\ x - 2y \geq 4,\ y \geq 0\},$$

$F(x,y) = x + y$ and $f(x,y) = -5x - y$. The boundary of the polytope Ω consists of five edges $E_1, \dots, E_5$, corresponding to the activities of the five inequality constraints, in the order of their above listing (see [4, Fig. 1] for a visualization). This yields $\operatorname{gph} R = E_1 \cup E_2$, $\operatorname{gph} S = E_2 \cup E_3$, hence the equilibrium set $E = E_2$. On the other hand, the set of efficient points of (5.4) *is $E_4 \cup E_5$. This shows that in general the set E and the set of efficient points can be disjoint.*

Nevertheless, a technique presented in [29] shows that the set of equilibria of $(P(y), Q(x))$ may be written as the efficient set of some different vector optimization problem, with $n + m + 2$ objective functions and an appropriately defined ordering cone C.

Theorem 5.8 ([29, Th. 2.6])**.** *The equilibrium set E of $(P(y), Q(x))$ coincides with the set of efficient points of the function $(x, y, F(x,y), f(x,y))$ on Ω with respect to the ordering cone*

$$C = \big(\{0_n\} \times \mathcal{Y} \times \mathbb{R} \times \mathbb{R}_\geq\big) \cup \big(\mathcal{X} \times \{0_m\} \times \mathbb{R}_\geq \times \mathbb{R}\big). \tag{5.5}$$

Indeed, for $(\bar{x}, \bar{y}), (x,y) \in \Omega$ the relation

$$(\bar{x}, \bar{y}, F(\bar{x}, \bar{y}), f(\bar{x}, \bar{y})) \leq_C (x, y, F(x,y), f(x,y))$$

holds if and only if $\bar{x} = x$, $f(\bar{x}, \bar{y}) \leq f(x,y)$ or $\bar{y} = y$, $F(\bar{x}, \bar{y}) \leq F(x,y)$ are true. This means $f(\bar{x}, \bar{y}) \leq f(\bar{x}, y)$ for $y \in Y(\bar{x})$ or $F(\bar{x}, \bar{y}) \leq F(x, \bar{y})$ for $x \in X(\bar{y})$, which motivates the relation to equilibrium points. The assumption of X and Y being derived from a shared constraint set is for simplicity only and may also be dropped (see [29, Cor. 3.2]). The following result is immediate from our considerations in Section 3.4.

Corollary 5.9. *For $\Omega \subseteq \mathcal{X} \times Y$ the set of saddle points of a real-valued function F on Ω coincides with the set of efficient points of the function $(x, y, F(x,y), -F(x,y))$ on Ω with respect to the ordering cone C from* (5.5)*.*

In view of the results in Section 5.2, Corollary 5.9 also provides a relation between Lagrange duality and vector optimization.

Observe that the cone C from (5.5) contains the origin, but is neither convex nor pointed. Therefore the binary relation $\leq_C$ is reflexive, but neither transitive nor antisymmetric, which limits the algorithmic applicability of this reformulation to special cases. In [30] the set of all Nash equilibria of low-rank bi-matrix games is determined using weakly efficient sets of vector linear optimization problems.

Taking a different point of view, since efficient points of (F, f) with respect to the standard ordering cone are unrelated to equilibria (Example 5.7), one may want to determine the efficient ones among all equilibria. In their seminal work [35] Harsanyi and Selten introduce the notion of payoff dominance to explain how players select some solution of a Nash equilibrium problem from a set of nonunique equilibria. In [6] this concept is formulated for generalized Nash equilibrium problems, and payoff dominance is relaxed to the more widely applicable requirement of payoff nondominatedness. In fact, this amounts to the selection of efficient ones among all elements of the equilibrium set. In [6] it is shown how different characterizations of generalized Nash equilibria yield different semi-infinite optimization problems (Section 4.1) for the computation of payoff nondominated equilibria.

5.5 Bilevel and vector optimization

As discussed in Section 2.2.1, the optimal-value-function formulation of the standard weak version (P_{sw}) of a bilevel problem can be written as

$$\min_{(x,y)\in\mathcal{X}\times\mathcal{Y}} F(x,y) \quad \text{s.t.} \quad (x,y)\in\Omega,\ f(x,y)=\varphi(x) \tag{5.6}$$

with the shared constraint set $\Omega = \operatorname{gph} X \cap \operatorname{gph} Y$. In analogy to Section 5.4, this raises the question whether the minimal points of (5.6) are related to the efficient points of the bicriteria optimization problem (5.4) with respect to the standard ordering cone $C = \mathbb{R}^2_{\geq}$. Like in Section 5.4 the answer is negative, and a counterexample results from the same data as in Example 5.7 (which was, in fact, the original intention of [4, Ex. 1]). Note that in this example X and Y are derived from a given shared constraint set Ω.

Example 5.10. *Consider $\mathcal{X}$, $\mathcal{Y}$, Ω, F and f from Example 5.7. As seen there, with the five edges $E_1, \ldots, E_5$ of the polytope Ω one obtains $\operatorname{gph} S = E_2 \cup E_3$. Hence the unique minimal point of (5.6) is $(\bar{x}, \bar{y}) = (8/9, 20/9) \in E_2 \cap E_1$. Since, on the other hand, the set of efficient points of (5.4) is $E_4 \cup E_5$, the point $(\bar{x}, \bar{y})$ is not efficient. This illustrates that in general the set of minimal points of (5.6) and the set of efficient points of (5.4) can be disjoint.*

Under rather strong assumptions the picture changes.

Theorem 5.11 ([57, Ths. 3,4])**.** *For $\overline{X} \subseteq \mathcal{X}$ and $\overline{Y} \subseteq \mathcal{Y}$ let $\Omega = \overline{X} \times \overline{Y}$, let S be single-valued on $\overline{X}$ with $S(x) = \{y(x)\}$, $x \in \overline{X}$, and let*

$$\begin{aligned} &\{(x,\xi,y,\eta)\in \overline{X}\times\overline{X}\times\overline{Y}\times\overline{Y} \mid f(\xi,\eta)\le f(x,y)\} \\ &\quad \subseteq \{(x,\xi,y,\eta)\in \overline{X}\times\overline{X}\times\overline{Y}\times\overline{Y} \mid F(\xi,\eta)< F(x,y)\} \end{aligned}$$

be satisfied. Then the following assertions hold.

(a) *For each minimal point* $\bar{x}$ *of* (5.6) *the pair* $(\bar{x}, y(\bar{x}))$ *is efficient for* (5.4).
(b) *For each efficient point* $(\bar{x}, \bar{y})$ *of* (5.4) *with* $\bar{y} = y(\bar{x})$ *the point* $\bar{x}$ *is minimal for* (5.6).

Finally we mention that the set of weakly efficient points of minimizing a vector function $\theta : \mathbb{R}^m \to \mathbb{R}^n$ with convex entries over a convex compact set $\overline{Y} \subseteq \mathbb{R}^m$ with respect to the standard ordering cone $\mathbb{R}^n_{\geq}$ may be characterized as the collection of all minimal points $S(x)$, $x \in \overline{X} := \{x \in \mathbb{R}^n_{\geq} \mid \sum_{i=1}^n x_i = 1\}$, of $(Q(x))$ with $f(x, y) = \langle x, \theta(y)\rangle$ (i.e., by the weighted sum scalarization, [24, 56]). Hence, the minimization of some objective function F over the set of weakly efficient points may be written as

$$\min_{y \in \mathcal{Y}} F(y) \quad \text{s.t.} \quad \exists\, x \in \overline{X} : \; y \in S(x).$$

By making the implicit variable x explicit, one obtains the bilevel problem

$$\min_{(x,y) \in \mathbb{R}^n \times \mathcal{Y}} F(y) \quad \text{s.t.} \quad x \in \overline{X}, \; y \in S(x)$$

in a standard weak version (P_{sw}). For details on this approach see [11, 66] and [21, Sec. 20.4.1.2].

6 Final remarks

This paper intends to provide a high-level survey about different possibilities to couple pairs of optimization problems. It does not cover, though, the important aspects of optimality conditions for smooth and nonsmooth problems, stability considerations, and algorithmic approaches.

Since pairs of symmetrically coupled optimization problems correspond to two-player GNEPs, surveys of appropriate optimality conditions and solution methods can be found in [25, 31]. For tractability reasons, often convexity assumptions are imposed on the problems $(P(y))$ and $(Q(x))$, and algorithms are mostly designed to compute or approximate a single equilibrium point. In particular, under convexity assumptions and constraint qualifications the optimal points of $(P(y))$ and $(Q(x))$ coincide with their respective KKT points, so that the equilibria coincide with the solutions of the corresponding joint KKT system. However, since the KKT conditions comprise the notoriously hard to handle complementarity constraints, the latter systems require specialized solution methods [25]. A branch-and-bound approach for the approximation of all solutions of a standard NEP without any convexity assumptions is presented in [43].

Optimality conditions and solution methods for pairs of asymmetrically coupled problems may be found in the vast literature on bilevel and (generalized) semi-infinite optimization [20, 21, 36, 37, 42, 58, 65, 68, 76] and the references therein. Again, under convexity assumptions and a lower level constraint qualification, the optimal points of $(Q(x))$ coincide with its KKT points. Thus, in a pair of asymmetrically coupled optimization problems, the lower level may be rewritten as a KKT system, which gives rise to an equivalent single-level optimization problem (with potentially spurious local minimizers). Again in view of the appearing complementarity constraints, the algorithmic treatment of such problems is challenging, and they are referred to as the class of mathematical programs with complementarity constraints. For specialized solution methods see, e.g., [52].

The concept of pairs of symmetrically coupled optimization problems is easily generalized to finitely many symmetrically coupled optimization problems, corresponding to N-person GNEPs. Also the concept of pairs of asymmetrically coupled optimization problems can be generalized to finitely many asymmetrically coupled optimization problems. This can be in the form of multi-level optimization problems [21, Sec. 20.3], for example the three-level structures of the original weak and strong versions (P_{ow}) and (P_s) of a bilevel optimization problem. Alternatively, in multi-follower bilevel problems [21, Sec. 20.7] there are only two levels, but multiple follower problems which are not coupled symmetrically among each other, but only via the upper level. An example is the first-stage problem (FS) in two-stage stochastic optimization (Section 4.4).

There also exist many models which mix both symmetric and asymmetric structures. For example, the standard weak reformulation of the standard strong version of a bilevel optimization problem from [47] couples a single leader with a two-player GNEP in the lower level. Such structures are called multi-follower games. In multi-leader-follower games the upper as well as the lower level can be GNEPs [2, 18, 38].

Acknowledgments. The author wishes to thank Alain Zemkoho for pointing out the benefit of comparing the several notions of saddle points.

Statements and declarations. The author has no competing interests to declare that are relevant to the content of this article.

References

[1] Arrow K, Debreu G (1954) Existence of an equilibrium for a competitive economy. Econometrica 22:265–290

[2] Aussel D, Svensson A (2020) A short state of the art on multi-leader-follower games. In: Bilevel Optimization: Advances and Next Challenges. Springer, p 53–76

[3] Bank B, Guddat J, Klatte D, et al (1982) Non-Linear Parametric Optimization. Walter de Gruyter

[4] Bard JF (1984) Optimality conditions for the bilevel programming problem. Naval Research Logistics Quarterly 31(1):13–26

[5] Bard JF (1998) Practical Bilevel Optimization: Algorithms and Applications. Springer Science & Business Media

[6] Beck M, Stein O (2023) Semi-infinite models for equilibrium selection. Minimax Theory Appl 9:1–18

[7] Ben-Tal A, Nemirovski A (1999) Robust solutions of uncertain linear programs. Operations Research Letters 25(1):1–13

[8] Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust Optimization. Princeton Series in Applied Mathematics, Princeton University Press

[9] Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. Numerische Mathematik 4(3):238–252

[10] Benko M, Mehlitz P (2021) On implicit variables in optimization theory. Journal of Nonsmooth Analysis and Optimization 2

[11] Benson HP (1991) An all-linear programming relaxation algorithm for optimizing over the efficient set. Journal of Global Optimization 1(1):83–104

[12] Bertsekas D, Nedic A, Ozdaglar A (2003) Convex Analysis and Optimization. Athena Scientific

[13] Bertsimas D, Hertog Dd (2022) Robust and adaptive optimization. Dynamic Ideas LLC

[14] Birge JR, Louveaux F (1997) Introduction to Stochastic Programming. Springer-Verlag

[15] Blum E, Oettli W (1994) From optimization and variational inequalities to equilibrium problems. Math Student 63:123–145

[16] Boyd SP, Vandenberghe L (2004) Convex Optimization. Cambridge University Press

[17] Caruso F, Ceparano MC, Morgan J (2025) Lower Stackelberg equilibria: from bilevel optimization to Stackelberg games. Optimization 74(12):2857–2883

[18] Carvalho M, Dragotto G, Feijoo F, et al (2024) When Nash meets Stackelberg. Management Science 70(10):7308–7324

[19] Debreu G (1952) A social equilibrium existence theorem. Proceedings of the National Academy of Sciences 38:886–893

[20] Dempe S (2002) Foundations of Bilevel Programming. Springer

[21] Dempe S (2020) Bilevel optimization: theory, algorithms, applications and a bibliography. In: Bilevel Optimization: Advances and Next Challenges. Springer, p 581–672

[22] Dem'yanov VF, Pevnyi AB (1972) Numerical methods for finding saddle points. USSR Computational Mathematics and Mathematical Physics 12(5):11–52

[23] Djelassi H, Mitsos A, Stein O (2021) Recent advances in nonconvex semi-infinite programming: applications and algorithms. EURO Journal on Computational Optimization 9:100006

[24] Ehrgott M (2005) Multicriteria Optimization. Springer

[25] Facchinei F, Kanzow C (2010) Generalized Nash equilibrium problems. Annals of Operations Research 175(1):177–211

[26] Facchinei F, Pang JS (2003) Finite-dimensional Variational Inequalities and Complementarity Problems, Vol. 1&2. Springer

[27] Facchinei F, Pang JS (2009) Nash equilibria: the variational approach. In: Palomar D, Eldar Y (eds) Convex Optimization in Signal Processing and Communications. Cambridge University Press, p 443–493

[28] Fan K (1972) A minimax inequality and applications. In: Inequalities III. Academic Press, New York, p 103–113

[29] Feinstein Z, Rudloff B (2024) Characterizing and computing the set of Nash equilibria via vector optimization. Operations Research 72(5):2082–2096

[30] Feinstein Z, Löhne A, Rudloff B (2025) Computing all Nash equilibria of low-rank bi-matrix games. arXiv preprint arXiv:250911718

[31] Fischer A, Herrich M, Schönefeld K (2014) Generalized Nash equilibrium problems - recent advances and challenges. Pesquisa Operacional 34(3):521–558

[32] Geoffrion AM (1972) Generalized Benders decomposition. Journal of Optimization Theory and Applications 10(4):237–260

[33] Goberna MA, López MA (1998) Linear Semi-infinite Optimization. Wiley, Chichester

[34] Goberna MA, López MA (2018) Recent contributions to linear semi-infinite optimization: an update. Annals of Operations Research 271(1):237–278

[35] Harsanyi JC, Selten R (1988) A General Theory of Equilibrium Selection in Games. The MIT Press

[36] Hettich R, Kortanek KO (1993) Semi-infinite programming: theory, methods, and applications. SIAM Review 35(3):380–429

[37] Hettich R, Zencke P (1982) Numerische Methoden der Approximation und semi-infiniten Optimierung. Teubner, Stuttgart

[38] Hori A, Tsuyuguchi D, Fukuda EH (2024) A method for multi-leader–multi-follower games by smoothing the followers' response function. Journal of Optimization Theory and Applications 203(1):305–335

[39] Iusem AN, Sosa W (2003) New existence results for equilibrium problems. Nonlinear Analysis: Theory, Methods & Applications 52(2):621–635

[40] Jongen HT, Jonker P, Twilt F (2013) Nonlinear optimization in finite dimensions: Morse theory, Chebyshev approximation, transversality, flows, parametric aspects. Springer Science & Business Media

[41] Kakutani S (1941) A generalization of Brouwer's fixed point theorem. Duke Math J 8:457–459

[42] Kirst P, Stein O (2019) Global optimization of generalized semi-infinite programs using disjunctive programming. Journal of Global Optimization 73(1):1–25

[43] Kirst P, Schwarze S, Stein O (2024) A branch-and-bound algorithm for nonconvex Nash equilibrium problems. SIAM Journal on Optimization 34(4):3371–3398

[44] Kleinert T, Labbé M, Ljubić I, et al (2021) A survey on mixed-integer programming techniques in bilevel optimization. EURO Journal on Computational Optimization 9:100007

[45] Köppe M, Queyranne M, Ryan CT (2010) Parametric integer programming algorithm for bilevel mixed integer programs. Journal of Optimization Theory and Applications 146(1):137–150

[46] Lampariello L, Sagratella S (2017) A bridge between bilevel programs and Nash games. Journal of Optimization Theory and Applications 174(2):613–635

[47] Lampariello L, Sagratella S, Stein O (2019) The standard pessimistic bilevel problem. SIAM Journal on Optimization 29(2):1634–1656

[48] López MA (to appear) A brief tutorial on semi-infinite optimization. The model, applications and basic mathematical constructions. TOP

[49] Loridan P, Morgan J (1996) Weak via strong Stackelberg problem: new results. Journal of Global Optimization 8(3):263–287

[50] Lucchetti R, Mignanego F, Pieri G (1987) Existence theorems of equilibrium points in Stackelberg games with constraints. Optimization 18(6):857–866

[51] Luenberger DG, Ye Y (1984) Linear and Nonlinear Programming. Springer

[52] Luo ZQ, Pang JS, Ralph D (1996) Mathematical Programs with Equilibrium Constraints. Cambridge University Press

[53] Mallozzi L, Morgan J (2001) Mixed strategies for hierarchical zero-sum games. In: Advances in Dynamic Games and Applications. Springer, p 65–77

[54] Mallozzi L, Messalli R, Patrì S, et al (2018) Some aspects of the Stackelberg leader/follower model. In: Open Problems in Optimization and Data Analysis. Springer, p 171–181

[55] Mehlitz P (2025) Revisiting implicit variables in mathematical optimization: simplified modeling and a numerical evidence. arXiv:2503.16902

[56] Miettinen K (1999) Nonlinear Multiobjective Optimization, vol 12. Springer Science & Business Media

[57] Migdalas A (1995) When is a Stackelberg equilibrium Pareto optimum? In: Advances in Multicriteria Analysis. Springer, p 175–181

[58] Mitsos A, Tsoukalas A (2015) Global optimization of generalized semi-infinite programs via restriction of the right hand side. Journal of Global Optimization 61(1):1–17

[59] Monderer D, Shapley LS (1996) Potential games. Games and Economic Behavior 14(1):124–143

[60] Nash JF (1951) Non-cooperative games. Annals of Mathematics 54:286–295

[61] Rockafellar RT (1970) Convex Analysis. Princeton University Press, Princeton, NJ, USA

[62] Rockafellar RT, Wets RJ (1998) Variational Analysis. Springer

[63] Rudin W (1976) Principles of Mathematical Analysis, 3rd ed. McGraw-Hill, Inc.

[64] Schmidt M, Beck Y (2023) A gentle and incomplete introduction to bilevel optimization. Tech. rep., Optimization Online

[65] Seidel T, Küfer KH (2022) An adaptive discretization method solving semi-infinite optimization problems with quadratic rate of convergence. Optimization 71(8):2211–2239

[66] Sosa W, Raupp F (2007) On minimization over weakly efficient sets. Optimization 56(1-2):207–219

[67] von Stackelberg H (1934) Marktform und Gleichgewicht. Julius Springer, Vienna

[68] Stein O (2003) Bi-Level Strategies in Semi-Infinite Programming. Kluwer Academic Publishers

[69] Stein O (2024) Basic Concepts of Global Optimization. Springer

[70] Stein O (2025) A tutorial on properties of the epigraph reformulation. EURO Journal on Computational Optimization p 100109

[71] Tadelis S (2013) Game Theory: An Introduction. Princeton University Press

[72] Tuy H (2008) Minimax: existence and stability. In: Pareto Optimality, Game Theory and Equilibria. Springer, p 3–21

[73] Wiesemann W, Tsoukalas A, Kleniati P, et al (2013) Pessimistic bilevel optimization. SIAM Journal on Optimization 23(1):353–380

[74] Žaković S, Pantelides C, Rustem B (2000) An interior point algorithm for computing saddle points of constrained continuous minimax. Annals of Operations Research 99(1):59–77

[75] Zemkoho AB (2016) Solving ill-posed bilevel programs. Set-Valued and Variational Analysis 24(3):423–448

[76] Zingler A, Mitsos A (2025) Derivative-enhanced lower-bounding in adaptive discretization for the global solution of semi-infinite optimization problems. Optimization 0(0):1–34

Appendix A Vector optimization and variational inequalities

Consider a vector optimization problem (Section 2.3) with objective function $\theta : M \to \mathbb{R}^k$ on a closed set M as well as a closed convex ordering cone $C \subseteq \mathbb{R}^k$ such that both C and its polar cone $C^+ := \{\zeta \in \mathbb{R}^k \mid \langle y, \zeta \rangle \geq 0 \, \forall \, y \in C\}$ possess a nonempty interior. Then the weakly efficient points $\bar{x}$ of θ on M with respect to C agree with the solutions of the equilibrium problem (EP) with $Z := M$ and

$$T(x, \bar{x}) := \max_{\zeta \in C^+, \, \|\zeta\|=1} \langle \zeta, \theta(\bar{x}) - \theta(x) \rangle$$

if M is convex and θ is C-convex on M [39]. The latter means that $\theta((1-\lambda)x^1 + \lambda x^2) \leq_C (1-\lambda)\theta(x^1) + \lambda\theta(x^2)$ holds for all $x^1, x^2 \in M$ and $\lambda \in (0, 1)$.

Appendix B Local saddle points of type 1 and 2

This section provides a detailed comparison of local saddle points of type 1 and type 2 from Section 3.5. For brevity this comparison is confined to interior points of Ω, which locally corresponds to the unconstrained case.

Definition B.1. *A point $\bar{z} = (\bar{x}, \bar{y})$ is called a local saddle point of type 1 for $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ if there exist neighborhoods U and V of $\bar{x}$ and $\bar{y}$, respectively, such that $\bar{z}$ is a saddle point of F on $U \times V$, that is,* (3.9) *holds with $\overline{X} := U$ and $\overline{Y} := V$.*

By (3.10), $(\bar{x}, \bar{y})$ is a local saddle point of type 1 for F if and only if $\bar{x}$ is a local minimal point of $F(\cdot, \bar{y})$, and $\bar{y}$ is a local maximal point of $F(\bar{x}, \cdot)$.

Definition B.2. *Let $F : \mathbb{R}^N \to \mathbb{R}$ be differentiable at $\bar{z}$. Then $\bar{z}$ is called a local saddle point of type 2 for F if $\nabla F(\bar{z}) = 0$ holds and $\bar{z}$ is neither a local minimal nor a local maximal point of F.*

Both Definition B.1 and B.2 can be extended to the constrained case. While this is straightforward for Definition B.1, in Definition B.2 one needs a criticality concept which covers, both, local minimal and local maximal points of constrained optimization problems. Even if a description of Ω by smooth constraints is assumed, this is not the case for the KKT conditions, but they need to be weakened by dropping the sign constraints on the Lagrange multipliers corresponding to active inequality constraints. A feasible point satisfying this relaxed concept of a KKT point is called a critical point of F on Ω [40]. The subsequent discussion can be extended to this case, but we omit it for the ease of presentation.

Since no variable groups are prescribed in Definition B.2, local saddle points of type 2 cover situations like at the critical point $\bar{z} = 0$ of the function $F(z) = z_1 z_2$. The Hessian $D^2F(0)$ is indefinite, but not a diagonal matrix. Indeed, F possesses a local minimizer at $\bar{z} = 0$ along range$((1,1)^\top)$ and a local maximizer along range$((1,-1)^\top)$. Also observe that $D^2F(0)$ is non-singular, while for $N = 1$ each type 2 local saddle point $\bar{z}$ of a C^2-function F must possess the singular Hessian $F''(\bar{z}) = 0$ (making it an inflection point), since otherwise, by the standard second-order sufficient optimality conditions, it would be a local minimal or local maximal point. As a consequence, type 2 local saddle points for $N = 1$ exist, but vanish under small perturbations of F, while type 2 local saddle points for $N > 1$ with non-singular Hessian are stable and, therefore, must e.g. be considered when studying the convergence behavior of optimization algorithms (see [40] for detailed stability considerations). This motivates the following definition.

Definition B.3. *Let F be twice continuously differentiable at $\bar{z}$. Then a type 2 local saddle point $\bar{z}$ of F is called non-degenerate if the Hessian $D^2F(\bar{z})$ is non-singular.*

In this terminology, all type 2 local saddle points of univariate C^2-functions are degenerate, and $F(z) = z_1 z_2$ possesses a non-degenerate type 2 local saddle point at $\bar{z} = 0$.

Even if local saddle points of type 1 are considered under smoothness assumptions, and variable groups are prescribed for local saddle points of type 2, without additional assumptions the two concepts are unrelated. Indeed, for the function $F(x, y) = x^2$ the point $\bar{z} = (\bar{x}, \bar{y}) = (0, 0)$ is a local saddle point of type 1, but not of type 2 since, while it is a critical point of F, it is also a local minimizer of F in both variables (x, y). Furthermore, the function $F(x, y) = x^2 + y^2 \sin(1/y)$ is differentiable at $\bar{z} = (0, 0)$ and possesses $\bar{z}$ as a critical point which is neither a local minimizer nor a local maximizer of F, hence a local saddle point of type 2. However, while $\bar{x}$ is a local minimizer of

$F(\cdot,\bar{y})$, $\bar{y}$ is not a local maximizer of $F(\bar{x},\cdot)$, so that $\bar{z}$ is not a local saddle point of type 1.

Additional non-degeneracy assumptions rule out such examples. Indeed, in the following result we let F be twice continuously differentiable at $\bar{z}$. Then any local maximizer $\bar{y}$ of $F(\bar{x},\cdot)$ satisfies $\nabla_y F(\bar{z}) = 0$, and the Hessian $D_y^2 F(\bar{z})$ is negative semi-definite. We shall use that such a matrix is not the zero matrix iff it possesses at least one negative eigenvalue.

Proposition B.4. *Let F be twice continuously differentiable at $\bar{z} = (\bar{x},\bar{y})$. Then the following assertions hold.*

(a) Let $(\bar{x},\bar{y})$ be a local saddle point of type 1 for F, and let none of the matrices $D_x^2 F(\bar{z})$ and $D_y^2 F(\bar{z})$ be a zero matrix. Then $\bar{z}$ is a local saddle point of type 2 for F.

(b) Let $\bar{z}$ be a non-degenerate local saddle point of type 2 for F, let $D^2 F(\bar{z})$ possess n positive and m negative eigenvalues, collect pairwise orthogonal eigenvectors corresponding to the positive eigenvalues of $D^2 F(\bar{z})$ as columns of the matrix V^+, and collect pairwise orthogonal eigenvectors corresponding to the negative eigenvalues of $D^2 F(\bar{z})$ as columns of the matrix V^-. If $\operatorname{range} V^+ = \mathbb{R}^n \times \{0_m\}$ and $\operatorname{range} V^- = \{0_n\} \times \mathbb{R}^m$ hold (where actually each one of the two conditions implies the other one), then $\bar{x}$ is a local saddle point of type 1 for F.

Proof Part (a): Fermat's rule yields $\nabla_x F(\bar{x},\bar{y}) = 0$ for the local minimal point $\bar{x}$ of $F(\cdot,\bar{y})$ and $\nabla_y F(\bar{x},\bar{y}) = 0$ for the local maximal point $\bar{y}$ of $F(\bar{x},\cdot)$, so that altogether $\bar{z}$ is a critical point of F. Assume that $\bar{z}$ is even a local minimal point of F. Then the Hessian $D^2 F(\bar{z})$ is positive semi-definite, which contradicts the fact that its lower right block $D_y^2 F(\bar{z})$ possesses at least one negative eigenvalue. Likewise, $\bar{z}$ cannot be a local maximal point of F. Hence, $\bar{z}$ is a local saddle point of type 2 for F.

Part (b): Since $\bar{z}$ is a local saddle point of type 2 of F, $\bar{x}$ is a critical point of $F(\cdot,\bar{y})$. Moreover the Hessian $D_x^2 F(\bar{x},\bar{y})$ satisfies for all $d \in \mathbb{R}^n \setminus \{0\}$

$$\begin{aligned} d^\top D_x^2 F(\bar{z}) d &= \begin{pmatrix} d \\ 0 \end{pmatrix}^\top D^2 F(\bar{z}) \begin{pmatrix} d \\ 0 \end{pmatrix} = (V^+\eta)^\top D^2 F(\bar{z})(V^+\eta) \\ &= \eta^\top ((V^+)^\top D^2 F(\bar{z}) V^+)\eta > 0 \end{aligned}$$

with some $\eta \in \mathbb{R}^n \setminus \{0\}$. Hence $D_x^2 F(\bar{z})$ is positive definite, which shows that $\bar{x}$ is a local minimal point of $F(\cdot,\bar{y})$. Analogously one demonstrates that $\bar{y}$ is a local maximal point of $F(\bar{x},\cdot)$. This shows the assertion. □

Appendix C The wait-and-see approach in stochastic optimization

Based on the notation from Section 4.4, the wait-and-see approach assumes that for every scenario ξ_s one can decide simultaneously over x and y. One then determines the expected value $\sum_{s\in[\sigma]} p_s \varphi_s$ of the optimal values φ_s of the problems

$$\min_{(x_s,y_s)\in\mathcal{X}\times\mathcal{Y}} f_{\xi_s}(x_s,y_s) \quad \text{s.t.} \quad (x_s,y_s) \in \operatorname{gph} Y_{\xi_s}$$

with $s \in [\sigma]$. In contrast to the here-and-now solution of the two-stage problem with recourse from Section 4.4, not a single decision x is determined as an optimal point of the second-stage expected value function, but for each scenario a different x_s is computed, before the expected value is taken. This amounts to the assumption of complete information about the scenarios already in the first-stage. Therefore the resulting value typically turns out to be an unrealistic underestimate of the expected minimal value, but it serves as a benchmark for the more realistic result from the two-stage approach [14].