

ON THE CONVERGENCE OF COLUMN-AND-CONSTRAINT GENERATION ALGORITHMS IN TWO-STAGE ROBUST OPTIMIZATION

HENRI LEFEBVRE

ABSTRACT. We study the well-posedness and convergence of column-and-constraint generation algorithms for general two-stage robust optimization problems. The analysis is formulated in terms of regularity properties of the objective, the second-stage feasible region mapping, and the separation value function, without relying on a particular algebraic representation of the second-stage problem. We give sufficient conditions for the existence of solutions and the well-posedness of the algorithm, prove finite termination for every fixed positive tolerance, and characterize accumulation points when zero tolerance is used. We then introduce pointwise, graph, and violation tolerance properties, which translate the stopping condition into guarantees on the objective and approximate second-stage feasibility. The resulting theory is applied to canonical, residual-based, and graph-distance separation value functions. Finally, we consider robust bilevel optimization with a wait-and-see convex follower and obtain finite-termination and convergence guarantees for the corresponding algorithm under explicit regularity assumptions.

1. INTRODUCTION

Two-stage robust optimization problems are optimization problems in which a set of decisions must be made before the realization of an uncertain parameter, while another set of decisions can be made after uncertainty has been revealed. Formally, such problems take the form

$$\inf_{x \in X} \sup_{u \in U} \inf_{y \in Y(x, u)} f(x, y, u), \quad (1)$$

in which $X \subseteq \mathbb{R}^{n_x}$ denotes the set of feasible first-stage decisions, $U \subseteq \mathbb{R}^{n_u}$ is the uncertainty set, and $Y(x, u) \subseteq \bar{Y}$ is the second-stage feasible region, for some $\bar{Y} \subseteq \mathbb{R}^{n_y}$. Finally, $f : \mathbb{R}^{n_x + n_u + n_y} \rightarrow \mathbb{R}$ is the objective function.

The nested min-max-min structure of Problem (1) makes two-stage robust optimization considerably more challenging than its single-stage counterpart. A widely used solution approach is the column-and-constraint generation (CCG) algorithm, originally proposed by Zeng and Zhao (2013) for linearly constrained problems. CCG exploits the decomposable structure of the problem by iteratively solving two optimization problems. First, the master problem considers only a finite subset of uncertainty scenarios and enforces the recourse constraints for those scenarios only. Then, the separation problem searches the entire uncertainty set U for a scenario under which the current master solution either fails to be feasible or fails to attain the reported objective value, i.e., for which the second-stage problem has no feasible point, or its optimal value exceeds the estimation returned by the master problem. If such a scenario is found, it is added to the master problem, and the process is repeated until no such scenario remains.

Date: September 23, 2026.

2020 Mathematics Subject Classification. 90-XX, 90C47, 90C31, 49J53, 90C15.

Key words and phrases. Column-and-constraint generation, Two-stage robust optimization, Min-max-min optimization, Convergence analysis, Finite termination, Set-valued analysis, Robust bilevel optimization, Wait-and-see follower.

Since its introduction, the method has been extended along several directions. Zhao and Zeng (2012) and Lefebvre et al. (2026) consider settings in which the second-stage recourse decisions are mixed-integer. Subramanyam (2022) and Lefebvre and Subramanyam (2025) study the case of discrete uncertainty sets. Beyond linear constraints, Lefebvre et al. (2025) extend the method to general convex constraints. Recently, Hu et al. (2026), Xu (2023), and Lefebvre and Thuerlauf (2026) adapt the method to robust linear bilevel optimization problems. In the latter, the second-stage problem is a bilevel problem that is parameterized by the first-stage decisions and the uncertain parameters. Finally, Zeng and Wang (2022) consider decision-dependent uncertainty sets, where the set U itself depends on the first-stage decision x , so that the master and separation problems must be coupled more tightly than in the standard, decision-independent setting.

Surprisingly, to the best of our knowledge, there has been no systematic study of the convergence properties of the method, nor of the conditions under which it is well-posed, i.e., under which the master and separation subproblems are guaranteed to have solutions, beyond linear constraints. Instead, convergence and finite-termination arguments are typically derived from scratch for each new problem class and each new separation scheme, tailored to the specific structure at hand. This is in contrast to the main motivation of this paper that addresses convergence from a purely topological standpoint.

In this paper, we close this gap by giving general sufficient conditions for the well-posedness and convergence analysis of the CCG method. Our main contributions are as follows.

- (i) We derive sufficient conditions for the existence of solutions to the two-stage robust problem and for the well-posedness of the master and separation problems generated by the CCG method.
- (ii) We prove finite termination for every fixed positive tolerance and convergence of accumulation points when the tolerance is zero. We also obtain an iteration bound under a Hölder continuity condition and a finite-termination result for polyhedral uncertainty sets.
- (iii) We introduce pointwise, graph, and violation tolerance properties to relate the stopping condition of the algorithm to objective and second-stage feasibility guarantees. We establish relations between these properties and verify them for canonical, residual-based, and graph-distance separation value functions.
- (iv) We apply the general framework to robust bilevel optimization with a wait-and-see convex follower. Sufficient conditions for the required regularity of the follower solution map are obtained from either a Slater condition and strict convexity or a right-hand-side parameterization with weakly analytic functions.

Figure 1 summarizes the role of the main assumptions in the existence and convergence results.

2. PRELIMINARIES

In this section, we recall some preliminary results that are well-known from variational analysis and define the main notation used throughout the paper. To that end, let $\varphi : \mathbb{R}^n \rightarrow \bar{\mathbb{R}} =: \mathbb{R} \cup \{\pm\infty\}$ be a real extended function. We call it proper if and only if the following holds: $\varphi(x) > -\infty$ for all x and there exists $\hat{x}$ such that $\varphi(\hat{x}) < \infty$ is satisfied. Its domain is denoted by $\text{dom}(\varphi) := \{x \in \mathbb{R}^n : \varphi(x) < \infty\}$. We also recall that φ is lower (resp. upper) semicontinuous if and only if its epigraph (resp. hypograph) is closed. It is level-bounded if and only if all of its sublevel sets are bounded. Next, we recall the well-known Weierstrass theorem on the

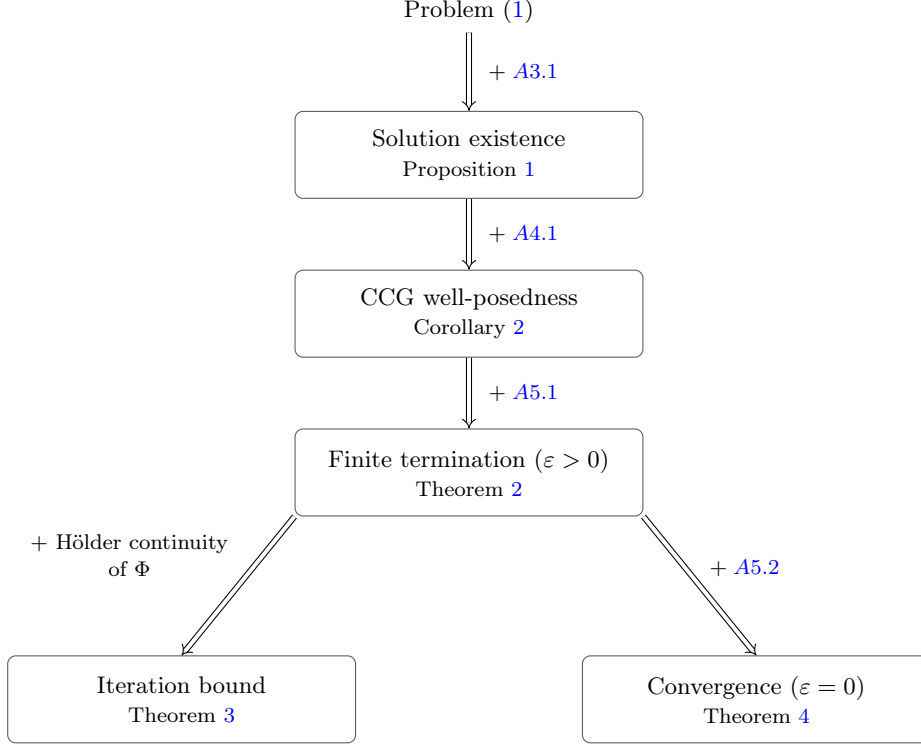


FIGURE 1. Main assumptions and their implications.

existence of solutions to optimization problems; see, e.g., Rockafellar and Wets (1998, Theorem 1.9).

Theorem 1 (Weierstrass). *Assume that $\varphi : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is lower semicontinuous, level-bounded and proper. Then the value $\inf \varphi$ is finite and the set $\arg \min \varphi$ is non-empty and compact.*

The last theorem only deals with “unconstrained” optimization problems. However, it is well-known that constrained optimization can be incorporated using the indicator function of a set C defined by

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Consequently, every lower semicontinuous real-valued function attains its (finite) infimum on every non-empty compact set. This also justifies the convention $\inf_{\emptyset} \varphi = \infty$ and $\sup_{\emptyset} \varphi = -\infty$.

Given the nested structure of Problem (1), it will also be useful to talk about set-valued maps. Thus, let $\Gamma : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{n_2}$ be given. We denote its graph by $\text{graph}(\Gamma) := \{(x, y) : y \in \Gamma(x)\}$ and its domain by $\text{dom}(\Gamma) := \{x : \exists y \in \Gamma(x)\}$. We now briefly discuss continuity properties of Γ . Let $t \in \mathbb{R}^{n_1}$ be given. We say that Γ is outer semicontinuous at t if, for any sequence $(t_k, \gamma_k) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ such that $\gamma_k \in \Gamma(t_k)$ and $(t_k, \gamma_k) \rightarrow (t, \gamma)$ for some $\gamma \in \mathbb{R}^{n_2}$, it holds $\gamma \in \Gamma(t)$.

By definition, it follows that if Γ is outer semicontinuous at each $t \in \mathbb{R}^{n_1}$, then its graph is closed. Moreover, if there exists a bounded set $K \subseteq \mathbb{R}^{n_2}$ such that $\Gamma(t) \subseteq K$ holds for all $t \in \mathbb{R}^{n_1}$, then also its domain is closed.

Note that we use the notion of outer semicontinuity from Rockafellar and Wets (1998) that is equivalent, in finite dimensions, to the sequential definition above.

We highlight that, for locally bounded set-valued maps with closed values, outer semicontinuity is equivalent to upper hemicontinuity in the sense of Aliprantis and Border (2006); see, e.g., Rockafellar and Wets (1998, Theorem 5.19).

We say that Γ is inner semicontinuous at t if, for any $\gamma \in \Gamma(t)$ and any sequence $t_k \rightarrow t$, there exists a sequence $\gamma_k \in \Gamma(t_k)$ such that $\gamma_k \rightarrow \gamma$. Again, in finite-dimensional spaces, inner semicontinuity is equivalent to lower hemicontinuity in the sense of Aliprantis and Border (2006); see, e.g., Theorem 17.21 in cited book.

Several of the continuity assumptions below may alternatively be formulated relative to the sets on which the corresponding maps are used, such as X and U . We retain the ambient-space formulation for notational convenience.

Using the aforementioned continuity notions, we now state two well-known results regarding continuity properties of optimal-value functions of parameterized optimization problems; see, e.g., Aliprantis and Border (2006, Lemma 17.29 and 17.30).

Lemma 1. *Let $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ and $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$ be endowed with their relative Euclidean topologies. The following statements are true.*

- i. *Let $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a non-empty-valued inner semicontinuous set-valued map, and let $\varphi : \text{graph}(\Gamma) \rightarrow \mathbb{R}$ be upper semicontinuous. Then, the extended-real-valued function $v : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ defined by*

$$v(x) := \inf_{y \in \Gamma(x)} \varphi(x, y)$$

is upper semicontinuous on $\mathcal{X}$.

- ii. *Let $\Gamma : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a non-empty-valued, locally bounded, outer semicontinuous set-valued map, and let $\varphi : \text{graph}(\Gamma) \rightarrow \mathbb{R}$ be lower semicontinuous. Then, the function $v : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ defined by*

$$v(x) := \min_{y \in \Gamma(x)} \varphi(x, y)$$

is lower semicontinuous on $\mathcal{X}$.

3. EXISTENCE OF SOLUTIONS

In this section, we give sufficient conditions for the existence of solutions to Problem (1) itself. Arguably, any meaningful set of conditions ensuring the convergence of CCG algorithms must, in particular, guarantee that Problem (1) itself has a solution. The conditions derived in this section therefore serve as a natural starting point, and will be reused, and strengthened whenever necessary. Before we proceed, we make the following standing assumption.

Standing Assumption. *The sets X , U and $\bar{Y}$ are non-empty compact sets.*

The standing assumption is natural in many applications such as the one considered in Section 7. We note, however, that we do not assume that Problem (1) is feasible, as we define next.

Definition 1 (Feasibility). *We say that Problem (1) is feasible if and only if there exists $x \in X$ such that, for every $u \in U$, there exists $y \in Y(x, u)$.*

To ease our exposition, we introduce the optimal-value functions

$$\begin{aligned} \xi(x) &:= \sup_{u \in U} \inf_{y \in Y(x, u)} f(x, y, u), \\ Q(x, u) &:= \inf_{y \in Y(x, u)} f(x, y, u). \end{aligned}$$

Hence, ξ denotes the (worst-case) objective value of Problem (1) as a function of the first-stage decision x , while Q denotes the optimal value of the second-stage problem, for a fixed first-stage decision x and a given scenario u . With this notation, Problem (1) may be written compactly as $\inf_x \{\xi(x) : x \in X\}$.

The Weierstrass theorem recalled in Theorem 1 implies that Problem (1) is either infeasible or has a finite solution if the function ξ is lower semicontinuous and the set $X \cap \text{dom}(\xi)$ is compact. Our first result states that the following assumption is sufficient for this to be satisfied.

A3.1. *The following statements are true.*

- i. *The function $f : \mathbb{R}^{n_x+n_y+n_u} \rightarrow \mathbb{R}$ is real valued.*
- ii. *For all fixed $u \in U$, $f(\cdot, \cdot, u)$ is (jointly) lower semicontinuous.*
- iii. *For all fixed $x \in X$, $f(x, \cdot, \cdot)$ is (jointly) upper semicontinuous.*
- iv. *For all fixed $u \in U$, $Y(\cdot, u)$ is outer semicontinuous.*

Note that A3.1.i, A3.1.ii and A3.1.iv are natural assumptions for the deterministic problem given by

$$\inf_{x \in X, y \in Y(x, u)} f(x, y, u)$$

to be either infeasible or attain a finite optimal value for any $u \in U$. The only additional assumption in this sense is A3.1.iii. This asymmetry reflects the min–max structure: lower semicontinuity is natural for the minimization problems, whereas upper semicontinuity in the uncertainty parameter controls the outer supremum. We now give an example showing that Problem (1) may be feasible but have infinite optimal value if A3.1.iii does not hold.

Example 1 (Infinite optimal value). *Let $X = \{0\}$, $U = [0, 1]$, and $Y(x, u) := \{0\}$ for all $u \in U$. Moreover, consider*

$$f(x, y, u) := \begin{cases} \frac{1}{u}, & \text{if } u > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, Problem (1) is feasible, since the unique first-stage decision $x = 0$ admits the second-stage decision $y = 0$ for every $u \in U$. Moreover, f is real-valued and, for every fixed $u \in U$, $f(\cdot, \cdot, u)$ is even (jointly) continuous. The set-valued map $Y(\cdot, u)$ is also outer semicontinuous for every fixed $u \in U$. Hence, A3.1.i, A3.1.ii, and A3.1.iv are satisfied.

On the other hand, $f(0, \cdot, \cdot)$ is not upper semicontinuous at $(0, 0)$. Indeed,

$$\limsup_{u \downarrow 0} f(0, 0, u) = \infty > f(0, 0, 0) = 0.$$

Using the fact that $Q(0, u) = f(0, 0, u)$ holds for every $u \in U$, we obtain

$$\xi(0) = \sup_{u \in [0, 1]} Q(0, u) = \infty.$$

Hence, Problem (1) is feasible but has infinite optimal value.

We now state the existence theorem regarding solutions to Problem (1).

Proposition 1. *Let A3.1 hold. Then, the following statements are true.*

- i. *For all $(x, u) \in \mathbb{R}^{n_x} \times U$, the problem defining $Q(x, u)$ is either infeasible or attains a finite optimal value.*
- ii. *For all fixed $u \in U$, the extended real function $Q(\cdot, u) : \mathbb{R}^{n_x} \rightarrow \bar{\mathbb{R}}$ is lower semicontinuous, and its domain is closed.*
- iii. *The extended real function $\xi : \mathbb{R}^{n_x} \rightarrow \bar{\mathbb{R}}$ is lower semicontinuous, and the set $X \cap \text{dom}(\xi)$ is compact.*
- iv. *Problem (1) is either infeasible or attains a finite optimal value.*

Proof. First, let $(x, u) \in \mathbb{R}^{n_x} \times U$ be fixed. We first show that $Y(x, u)$ is compact. By A3.1.iv, $Y(\cdot, u)$ is outer semicontinuous and, by the standing assumption, $\bar{Y}$ is bounded. Hence, it follows that $\text{graph}(Y(\cdot, u))$ and $\text{dom}(Y(\cdot, u))$ are closed. Thus, $Y(x, u) \subseteq \bar{Y}$ gives that $\text{graph}(Y(\cdot, u)) \cap (\{x\} \times \mathbb{R}^{n_y})$ is compact, and since the

projection of a compact set is compact, we also have that $Y(x, u)$ is compact for all $(x, u) \in \mathbb{R}^{n_x} \times U$. Next, since $f(x, \cdot, u)$ is lower semicontinuous, as per [A3.1.ii](#), and the feasible region $Y(x, u)$ compact, the Weierstrass theorem (Theorem 1) implies item [i](#).

Second, let $u \in U$ be fixed and consider the restriction $Y(\cdot, u) : \text{dom}(Y(\cdot, u)) \rightrightarrows \bar{Y}$. This set-valued map is non-empty-valued by definition and compact-valued as discussed above. Moreover, it is locally bounded since all of its values are contained in the bounded set $\bar{Y}$. By [A3.1.iv](#), $Y(\cdot, u)$ is outer semicontinuous on $\text{dom}(Y(\cdot, u))$ and, by [A3.1.ii](#), $f(\cdot, \cdot, u)$ is lower semicontinuous. Hence, Lemma [1.ii](#) implies that $Q(\cdot, u)$ is lower semicontinuous on $\text{dom}(Y(\cdot, u))$. Moreover, by item [i](#),

$$\text{dom}(Q(\cdot, u)) = \text{dom}(Y(\cdot, u)),$$

and this set is closed as discussed above. Since $Q(x, u) = \infty$ for every $x \notin \text{dom}(Q(\cdot, u))$, it follows that the extended-real-valued function $Q(\cdot, u) : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ is lower semicontinuous on $\mathbb{R}^{n_x}$. This proves item [ii](#).

Third, item [iii](#) follows from the fact that $\xi(x)$ is the pointwise supremum of lower semicontinuous extended-real-valued functions, and is therefore lower semicontinuous. Moreover, we show that, for every $x \in X$, we have

$$x \in \text{dom}(\xi) \iff x \in \bigcap_{u \in U} \text{dom}(Q(\cdot, u)).$$

The forward implication is immediate. Conversely, let $x \in X \cap \bigcap_{u \in U} \text{dom}(Q(\cdot, u))$. By [A3.1.iii](#), the function $f(x, \cdot, \cdot)$ is upper semicontinuous on the compact set $\bar{Y} \times U$. Hence, there exists $M_x < \infty$ such that $f(x, y, u) \leq M_x$ holds for all $(y, u) \in \bar{Y} \times U$. Since $x \in \text{dom}(Q(\cdot, u))$ holds for every $u \in U$, we have that $Y(x, u) \neq \emptyset$ holds for every $u \in U$. Therefore, we have that

$$Q(x, u) = \min_{y \in Y(x, u)} f(x, y, u) \leq M_x$$

holds for all $u \in U$, and thus,

$$\xi(x) = \sup_{u \in U} Q(x, u) \leq M_x < \infty$$

also holds.

All in all, we obtain that

$$X \cap \text{dom}(\xi) = X \cap \bigcap_{u \in U} \text{dom}(Q(\cdot, u))$$

holds. Since each set $\text{dom}(Q(\cdot, u))$ is closed by item [ii](#) and X is compact, it follows that $X \cap \text{dom}(\xi)$ is compact.

Finally, if Problem [\(1\)](#) is feasible, then $X \cap \text{dom}(\xi) \neq \emptyset$. Since ξ is lower semicontinuous and finite-valued on the non-empty compact set $X \cap \text{dom}(\xi)$, the Weierstrass theorem implies that ξ attains a finite minimum on this set. This proves item [iv](#). $\square$

Proposition [1](#) shows that [A3.1](#) is sufficient to ensure that Problem [\(1\)](#) is feasible if and only if $X \cap \text{dom}(\xi) \neq \emptyset$ holds. However, we highlight that $\text{dom}(\xi)$ need not be closed. Next, we give further sufficient conditions for [A3.1](#) to be satisfied that may be more natural to practitioners.

Example 2. Let f be as required by [A3.1.i–A3.1.iii](#). Assume that, for all $u \in U$, there exists $g(\cdot, \cdot, u) : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{m_y}$ such that $Y(x, u) := \{y \in \mathbb{R}^{n_y} : g(x, y, u) \leq 0\}$. Moreover, let $g(\cdot, \cdot, u) : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{m_y}$ be (jointly) lower semicontinuous. Then, [A3.1](#) holds. Indeed, let $x_n \rightarrow x$ and $y_n \in Y(x_n, u) \rightarrow y$ be given, then $g_k(x_n, y_n, u) \leq 0$ holds for every component k and every n . In turn, lower semicontinuity of $g_k(\cdot, \cdot, u)$

gives $g_k(x, y, u) \leq \liminf_n g_k(x_n, y_n, u) \leq 0$. Thus, $y \in Y(x, u)$ holds and $Y(\cdot, u)$ is outer semicontinuous.

While Proposition 1 ensures that Problem (1) is well-posed, it does not necessarily imply that the inner supremum, i.e., the optimization problem defining $\xi(x)$, attains its optimal value. A natural sufficient condition for this is the upper semicontinuity of $Q(x, \cdot)$. However, this additional property is not needed to develop a convergence theory for CCG methods that still yields solutions to Problem (1). For the sake of completeness, however, we now give an example where $Q(x, \cdot)$ fails to be upper semicontinuous and the inner supremum is not attained, then derive sufficient conditions ensuring its attainment.

Example 3. For simplicity, let us choose $X = \{0\}$ and consider the problem

$$\sup_{u \in [0, 1]} \min_y \{y - u : yu \geq u, y \in [-1, 1]\} =: Q(u).$$

When $u = 0$, the constraint “ $yu \geq u$ ” vanishes and the solution to the inner problem is given by $y = -1$ and evaluates to -1 . Otherwise, $u > 0$ and the constraint forces to choose $y \geq 1$. Since y is minimized, the solution is $y = 1$ and has a value of $y - u$. Thus, we have that the function

$$Q(u) = \begin{cases} -1 & \text{if } u = 0, \\ 1 - u & \text{otherwise,} \end{cases}$$

is not upper semicontinuous. Moreover, the supremum equal to 1, obtained by letting $u \rightarrow 0$, is never attained by any point in $[0, 1]$.

We now state the sufficient conditions.

Proposition 2. Let A3.1 hold and assume that, for all fixed $x \in X$, $Y(x, \cdot)$ is inner semicontinuous relative to its domain. Then, for all $x \in X$, the following statements are true.

- i. The extended-real-valued function $Q(x, \cdot) : \mathbb{R}^{n_u} \rightarrow \bar{\mathbb{R}}$ is upper semicontinuous on its domain.
- ii. The supremum defining $\xi(x)$ is attained. Moreover, its value is equal to ∞ if and only if there exists $u \in U$ such that $Y(x, u) = \emptyset$.

Proof. First, let $x \in X$ be fixed and consider the restriction $Y(x, \cdot) : \text{dom}(Y(x, \cdot)) \rightrightarrows \bar{Y}$. This set-valued map is non-empty-valued by definition and inner semicontinuous by assumption. Moreover, $f(x, \cdot, \cdot)$ is upper semicontinuous by A3.1.iii. Hence, Lemma 1.i implies that $Q(x, \cdot)$ is upper semicontinuous on $\text{dom}(Y(x, \cdot))$, proving the first item.

Second, we distinguish two cases. Suppose first that $U \subseteq \text{dom}(Y(x, \cdot))$ holds. Then $Q(x, \cdot)$ is real-valued and upper semicontinuous on the non-empty compact set U . Hence, it attains its maximum on U , and this maximum is finite. Otherwise, there exists $u \in U \setminus \text{dom}(Y(x, \cdot))$. For such a point, $Y(x, u) = \emptyset$ and therefore $Q(x, u) = \infty$ holds. Hence, the supremum defining $\xi(x)$ is equal to ∞ and is attained at u . $\square$

Combining Proposition 1 and Proposition 2, we obtain the following corollary.

Corollary 1. Let the assumptions of Proposition 2 hold. Then, Problem (1) can be written as

$$\min_{x \in X} \max_{u \in U} \min_{y \in Y(x, u)} f(x, y, u).$$

We conclude this section with an example of a class of two-stage robust optimization problems with convex constraints which can be written as a min-max-min problem. To the best of our knowledge, this result is new in the literature.

Example 4. For simplicity, assume that f is jointly continuous, but not necessarily convex. Moreover, assume that there exists a continuous function $T : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{m_y \times n_u}$ and a jointly lower semicontinuous function $\tilde{g} : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{m_y}$ such that

$$Y(x, u) := \{y \in \mathbb{R}^{n_y} : T(x)u + \tilde{g}(x, y) \leq 0\}$$

holds. In addition, assume that each component of $\tilde{g}$ is weakly analytic in y , for every fixed $x \in X$. Then, the assumptions of Proposition 2 are satisfied.

Indeed, continuity of f gives A3.1.i, A3.1.ii, and A3.1.iii. Moreover, since T is continuous and $\tilde{g}$ is jointly lower semicontinuous, the function

$$g(x, y, u) := T(x)u + \tilde{g}(x, y)$$

is jointly lower semicontinuous in (x, y, u) . Hence, for every fixed $u \in U$, the set-valued map $Y(\cdot, u)$ is outer semicontinuous, as in Example 2, which gives A3.1.iv.

It remains to show that $Y(x, \cdot)$ is inner semicontinuous relative to its domain for every fixed $x \in X$. To this end, fix $x \in X$ and define $g_i^x(y) := \tilde{g}_i(x, y)$, for each $i = 1, \dots, m_y$, together with $\lambda^x(u) := -T(x)u$. Then, we may write $Y(x, u) = \{y : g_i^x(y) \leq \lambda_i^x(u), i = 1, \dots, m_y\}$. With this notation at hand, we can apply Theorem 3.3.3 of Bank et al. (1982) with $s = m_y$ and with parameter space

$$\Lambda_x := \{\lambda^x(u) : u \in \text{dom}(Y(x, \cdot))\}.$$

Since $s = m_y$, the strict-feasibility condition appearing in that theorem is vacuous. Moreover, by construction, the corresponding feasible set is non-empty for every $\lambda \in \Lambda_x$. Since each function g_i^x is weakly analytic, Theorem 3.3.3 implies that the feasible-set map is inner semicontinuous with respect to the parameter λ on Λ_x . Finally, since the map $u \mapsto \lambda^x(u) = -T(x)u$ is continuous, it follows that $Y(x, \cdot)$ is inner semicontinuous relative to $\text{dom}(Y(x, \cdot))$.

4. THE CCG ALGORITHM AND ITS WELL-POSEDNESS

The idea behind the CCG algorithm for Problem (1) is conceptually simple. First, a finite subset of uncertainty realizations is considered, and the corresponding master problem is solved to obtain a lower bound on the optimal objective value of Problem (1). Solving the master problem also yields a candidate first-stage decision, which must then be tested against the entire uncertainty set U . To this end, a separation problem, often referred to as the adversarial problem, is solved. Its purpose is to identify a realization of the uncertain parameter in U for which the candidate first-stage decision either renders the second-stage problem infeasible or demonstrates that the computed lower bound is not tight enough.

Thus, at each iteration t , the master problem considers a non-empty but finite set $U^t \subseteq U$ and asks to solve

$$\inf_{x_0, x, y_1, \dots, y_{|U^t|}} x_0 \tag{2a}$$

$$\text{s.t. } x_0 \geq f(x, y_k, u_k), \quad \text{for all } k = 1, \dots, |U^t|, \tag{2b}$$

$$y_k \in Y(x, u_k), \quad \text{for all } k = 1, \dots, |U^t|, \tag{2c}$$

$$x \in X. \tag{2d}$$

We may also view Problem (2) as Problem (1) in which the set U has been replaced by U^t , i.e., Problem (2) may be written as

$$\min_{x \in X} \max_{u \in U^t} \min_{y \in Y(x, u)} f(x, y, u).$$

With this in mind, the following proposition follows directly from Proposition 1.

Proposition 3. Let A3.1 hold and let $U^t := \{u_1, \dots, u_{|U^t|}\} \subseteq U$ be given. Then Problem (2) is either infeasible or has a solution.

Let (x_0^t, x^t) denote a solution to Problem (2). A separation problem is then solved to assert that x^t is feasible for the entire uncertainty set, i.e., $\forall u \in U, \exists y \in Y(x^t, u)$ holds, and that x_0^t matches the optimal objective function value of Problem (1). We model the separation problem through a given extended-real-valued function

$$\Phi : \mathbb{R} \times X \times U \rightarrow \bar{\mathbb{R}},$$

which we call the *separation value function*. At iteration t , the separation problem is

$$\sup_{u \in U} \Phi(x_0^t, x^t, u). \quad (3)$$

To fix ideas, one may think of the canonical separation value function given by

$$\Phi^{\text{can}}(x_0, x, u) := Q(x, u) - x_0 = \inf_{y \in Y(x, u)} f(x, y, u) - x_0.$$

However, we postpone the discussion of concrete choices of Φ to Section 6.

Now, let v^t denote the optimal objective function value of Problem (3). Following the classic CCG algorithm, whenever $v^t > \varepsilon$, where $\varepsilon \geq 0$ is some given tolerance, a scenario $u \in U$ that attains this supremum must be added to the master problem in (2). Thus, this requires that the separation problem is solvable. The following assumption is sufficient for this to be true, as we show next.

A4.1. *The following statements are true.*

- i. *For all $(x_0, x, u) \in \mathbb{R} \times X \times U$, $\Phi(x_0, x, u) > -\infty$ holds.*
- ii. *For all $(x_0, x) \in \mathbb{R} \times X$, $\Phi(x_0, x, \cdot)$ is upper semicontinuous on U .*

We now formally prove this result.

Proposition 4. *Let A4.1 hold. Then, for any $(x_0, x) \in \mathbb{R} \times X$, the supremum in Problem (3) is attained (possibly equal to ∞). Moreover, it is finite if $\Phi(x_0, x, u) < \infty$ holds for every $u \in U$.*

Proof. We distinguish two cases. Suppose first that $\Phi(x_0, x, u) < \infty$ holds for every $u \in U$. By A4.1.i, $\Phi(x_0, x, \cdot)$ is real-valued and, by A4.1.ii, it is upper semicontinuous on the non-empty compact set U . Hence, it attains a finite maximum on U . Otherwise, there exists $u \in U$ such that $\Phi(x_0, x, u) = \infty$, in which case the supremum in Problem (3) equals ∞ and is attained at that u . $\square$

All in all, the complete CCG algorithm is formally presented in Algorithm 1. Combining Proposition 3 and Proposition 4, we obtain the next corollary giving sufficient conditions for the algorithm to be well-posed.

Corollary 2 (Well-posedness of CCG). *Let A3.1 and A4.1 hold. Then, Algorithm 1 is well-posed. More precisely, at every iteration t , the master problem is either infeasible, in which case the algorithm terminates, or it has a solution, in which case the corresponding separation problem also has a solution.*

5. CONVERGENCE AND FINITE TERMINATION

In this section, we derive sufficient conditions for the finite termination and convergence of Algorithm 1. We start with finite termination and show that the following assumption, along with A3.1 and A4.1 that ensure that the algorithm is well-posed, are sufficient if a fixed positive tolerance is used.

A5.1. *The following statements hold.*

- i. *For all $x \in X$, $f(x, \cdot, \cdot)$ is jointly lower semicontinuous.*
- ii. *For all $x \in X$, $Y(x, \cdot)$ is outer semicontinuous.*
- iii. *There exists $M < \infty$ such that $f(x, y, u) \leq M$ holds for all $(x, y, u) \in X \times \bar{Y} \times U$.*

Algorithm 1 CCG Algorithm for Problem (1)

```

1: Input: An initial non-empty finite set  $U^1 \subseteq U$  and a tolerance  $\varepsilon \geq 0$ .
2: Set  $t \leftarrow 1$ .
3: Solve the master problem (2).
4: if it is infeasible then
5:   return "Problem (1) is infeasible.".
6: end if
7: Let  $(x_0^t, x^t, y_1^t, \dots, y_{|U^t|}^t)$  denote a solution to the master problem (2).
8: Solve the separation problem (3).
9: Let  $u^t$  denote one of its solutions and  $v^t$  its value.
10: if  $v^t > \varepsilon$  then
11:   Set  $U^{t+1} \leftarrow U^t \cup \{u^t\}$  and  $t \leftarrow t + 1$ .
12:   go to Line 3.
13: end if
14: return  $(x_0^t, x^t)$ .

```

- iv. The separation value function Φ is upper semicontinuous on $\mathbb{R} \times X \times U$.
- v. For all $(x_0, x, u) \in \mathbb{R} \times X \times U$ such that $Q(x, u) \leq x_0$ holds, we have that $\Phi(x_0, x, u) \leq 0$ also holds.

It is easy to see that (joint) continuity of f and g when $Y(x, u) := \{y \in \mathbb{R}^{n_y} : g(x, y, u) \leq 0\}$ is enough to ensure A5.1.i–A5.1.iii. Indeed, lower semicontinuity is a special case of continuity and outer semicontinuity follows from the closedness of $\text{graph}(Y)$. As to A5.1.iii, it is ensured by the Weierstrass theorem which implies that, e.g., $\max \{f(x, y, u) : x \in X, y \in \bar{Y}, u \in U\} < \infty$ holds, since X , $\bar{Y}$ and U are non-empty and compact by the standing assumption. The last point, namely A5.1.v, is also natural and is satisfied by all the separation value functions discussed in Section 6. The remaining condition is the upper semicontinuity of Φ in A5.1.iv.

Before we prove finite termination, we first derive some preliminary results. The next lemma shows that the function $Q(x, \cdot)$ is lower semicontinuous if A5.1 holds.

Lemma 2. *Let A5.1.i and A5.1.ii hold. Then, for all fixed $x \in X$, the extended real function $Q(x, \cdot) : \mathbb{R}^{n_u} \rightarrow \mathbb{R}$ is lower semicontinuous, and its domain is closed.*

Proof. The proof is identical to that of item ii in Proposition 1, with the roles of x and u exchanged. $\square$

Using this result, the following lemma provides further details on the accumulation points of the iterates generated by Algorithm 1 in the case where it does not terminate.

Lemma 3. *Let A3.1, A4.1, and A5.1.i–A5.1.iii hold and assume that Algorithm 1 does not terminate. Then, any accumulation point (x_0^*, x^*, u^*) of the iterates (x_0^t, x^t, u^t) generated by Algorithm 1 satisfies $Q(x^*, u^*) \leq x_0^* < \infty$.*

Proof. Note that A3.1 and A4.1 are needed to ensure that Algorithm 1 is well-posed. Now, assume that Algorithm 1 never terminates. Then, at every iteration t , the master problem must be feasible (otherwise, it would have stopped at Line 5) and therefore has a solution; see Proposition 3. By A5.1.iii, the sequence $(x_0^t)_{t \geq 1}$ is bounded from above by some $M < \infty$. Moreover, since $U^t \subseteq U^{t+1}$, it is also non-decreasing. Thus, $(x_0^t)_{t \geq 1}$ converges to some finite value. Now, let (x_0^*, x^*, u^*) be any accumulation point of (x_0^t, x^t, u^t) . Then, there exists a sub-sequence $(t_i)_{i \in \mathbb{N}}$ such that $(x_0^{t_i}, x^{t_i}, u^{t_i}) \rightarrow (x_0^*, x^*, u^*)$ holds. In particular, $x_0^{t_i} \rightarrow x_0^*$.

Now, fix $i \in \mathbb{N}$ and note that $u^{t_i} \in U^{t_i+1} \subseteq U^{t_j}$ holds for all $j > i$. Hence, $Q(x^{t_j}, u^{t_i}) \leq x_0^{t_j}$ holds for all $j > i$. In particular, $Q(x^{t_j}, u^{t_i}) < \infty$ holds, showing that $x^{t_j} \in \text{dom}(Q(\cdot, u^{t_i}))$ for all $j > i$. Since $\text{dom}(Q(\cdot, u^{t_i}))$ is closed (Proposition 1.ii) and $x^{t_j} \rightarrow x^*$, we obtain $x^* \in \text{dom}(Q(\cdot, u^{t_i}))$. Next, recall that $Q(\cdot, u^{t_i})$ is lower semicontinuous by Proposition 1. Hence, letting $j \rightarrow \infty$ gives that $Q(x^*, u^{t_i}) \leq x_0^*$ holds for all $i \in \mathbb{N}$. Again, in particular, we have that $Q(x^*, u^{t_i}) < \infty$, showing that $u^{t_i} \in \text{dom}(Q(x^*, \cdot))$. Since $\text{dom}(Q(x^*, \cdot))$ is closed and $u^{t_i} \rightarrow u^*$, we obtain $u^* \in \text{dom}(Q(x^*, \cdot))$. By Lemma 2, $Q(x^*, \cdot)$ is lower semicontinuous. Thus, letting $i \rightarrow \infty$ leads to $Q(x^*, u^*) \leq x_0^*$. $\square$

We are now ready to state the finite termination theorem for any positive tolerance $\varepsilon > 0$. For that purpose, the proof requires A5.1.iv and A5.1.v.

Theorem 2 (Finite termination). *Let A3.1, A4.1 and A5.1 hold and let $\varepsilon > 0$ be fixed. Then, Algorithm 1 terminates after a finite number of iterations, with a valid proof that either Problem (1) is infeasible, or, the returned point (x_0^t, x^t) satisfies*

$$\max_{u \in U} \Phi(x_0^t, x^t, u) \leq \varepsilon. \quad (4)$$

Proof. Assume, for contradiction, that the algorithm does not terminate. Let $(t_i)_{i \in \mathbb{N}}$, (x_0^*, x^*, u^*) be as in the proof of Lemma 3, so that $Q(x^*, u^*) \leq x_0^*$. By A5.1.v, we have that $\Phi(x_0^*, x^*, u^*) \leq 0$ holds. Moreover, $v^{t_i} = \Phi(x_0^{t_i}, x^{t_i}, u^{t_i})$ for all i . Hence, by upper semicontinuity of Φ ,

$$\limsup_{i \rightarrow \infty} v^{t_i} \leq \Phi(x_0^*, x^*, u^*) \leq 0.$$

Thus, there exists i_0 with $v^{t_{i_0}} \leq \varepsilon$, contradicting the assumption that the algorithm never stops.

Finally, if Algorithm 1 terminates because the master problem is infeasible, then Problem (1) is infeasible since $U^t \subseteq U$. Otherwise, it terminates with $v^t \leq \varepsilon$. Since u^t solves the separation problem, we obtain that (4) holds. $\square$

Theorem 2 therefore shows that the stated assumptions are enough to get the finite termination of Algorithm 1 when $\varepsilon > 0$. However, it says nothing about the quality of the returned solution (x_0^t, x^t) w.r.t. Problem (1). In fact, it only guarantees that (4) holds. Whether this property can be transferred to something relevant for Problem 1 heavily depends on the choice of Φ . This is the subject of Section 6, where different separation value functions are studied. Next, we show that if $\Phi(x_0, x, \cdot)$ is Hölder continuous uniformly in (x_0, x) , one can bound the number of iterations of Algorithm 1. Before we do, we need a lemma.

Lemma 4 (Vershynin 2026, Proposition 4.2.12 and Corollary 4.2.13). *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be non-empty and bounded and let $\delta > 0$. Then any subset of $\mathcal{D}$ whose points are pairwise more than δ apart has at most*

$$\left(\frac{2 \cdot \text{diam}(\mathcal{D})}{\delta} + 1 \right)^n$$

elements, where $\text{diam}(\mathcal{D}) := \sup_{u, u' \in \mathcal{D}} \{\|u - u'\| : u, u' \in \mathcal{D}\}$.

We now state the theorem.

Theorem 3. *Let $\varepsilon > 0$ be given and let A3.1, A4.1, A5.1.v hold. In addition, assume that $\Phi(x_0, x, \cdot)$ is Hölder continuous for any $(x_0, x) \in \mathbb{R} \times X$, i.e., there exists $L > 0$ and $\alpha \in (0, 1]$ such that for any $u, u' \in U$ and all $(x_0, x) \in \mathbb{R} \times X$, it holds*

$$|\Phi(x_0, x, u) - \Phi(x_0, x, u')| \leq L \|u - u'\|^\alpha.$$

Then, Algorithm 1 terminates after at most

$$T := \left\lceil \left(2 \operatorname{diam}(U) \left(\frac{L}{\varepsilon} \right)^{1/\alpha} + 1 \right)^{n_u} \right\rceil + 1 < \infty$$

iterations.

Proof. Suppose, for contradiction, that Algorithm 1 has not terminated by iteration T . Thus, for all $t \in \{1, \dots, T\}$, $v^t > \varepsilon$ holds. Now, take any two iterations $t_i < t_j$. Since $u^{t_i} \in U^{t_i+1} \subseteq U^{t_j}$ by construction, we have, at iteration t_j , that the master problem (2) yields a second-stage decision $y^{t_j}(u^{t_i})$ such that $y^{t_j}(u^{t_i}) \in Y(x^{t_j}, u^{t_i})$ and $f(x^{t_j}, y^{t_j}(u^{t_i}), u^{t_i}) \leq x_0^{t_j}$ holds. Thus, $Q(x^{t_j}, u^{t_i}) \leq x_0^{t_j}$ and, by A5.1.v, $\Phi(x_0^{t_j}, x^{t_j}, u^{t_i}) \leq 0$ holds.

Thus, using Hölder continuity of $\Phi(x_0^{t_j}, x^{t_j}, \cdot)$, it follows that

$$\varepsilon < v^{t_j} = \Phi(x_0^{t_j}, x^{t_j}, u^{t_j}) \leq \Phi(x_0^{t_j}, x^{t_j}, u^{t_i}) + L \|u^{t_i} - u^{t_j}\|^\alpha \leq L \|u^{t_i} - u^{t_j}\|^\alpha.$$

In turn, we have, for any two t_i, t_j with $t_i < t_j$, that

$$\|u^{t_j} - u^{t_i}\| > \left(\frac{\varepsilon}{L} \right)^{1/\alpha} =: \delta.$$

Thus, $u^1, \dots, u^T$ are T pairwise δ -separated points in the bounded set U . By Lemma 4, we get

$$T \leq \left(\frac{2 \operatorname{diam}(U)}{\delta} + 1 \right)^{n_u} = \left(2 \operatorname{diam}(U) \left(\frac{L}{\varepsilon} \right)^{1/\alpha} + 1 \right)^{n_u} < T,$$

where the last inequality follows from the definition of T . This is a contradiction. Hence, Algorithm 1 terminates within T iterations. $\square$

The argument above is closely related to the finite-termination analysis of cutting-set methods in Mutapcic and Boyd (2009), where Lipschitz regularity prevents successive cuts from being generated arbitrarily close to one another. In the present setting, the same geometric idea acts directly in the uncertainty space. In fact, Hölder continuity of the separation value function implies that any two scenarios generated before termination are separated by more than $(\varepsilon/L)^{1/\alpha}$. Lemma 4 then turns this separation property into an explicit packing bound on the number of iterations. Thus, Theorem 3 extends the underlying Lipschitz-type argument to Hölder-continuous separation value functions while making the dependence of the iteration bound on ε , L , α , $\operatorname{diam}(U)$, and n_u explicit.

While choosing a positive tolerance $\varepsilon > 0$ ensures finite termination of the CCG method, we now consider the case where $\varepsilon = 0$. In this setting, Algorithm 1 may not terminate. However, we show that, under some additional assumptions, accumulation points of the iterates generated by Algorithm 1 can still be related to solutions of Problem (1). To that end, we first derive sufficient conditions ensuring that an accumulation point (x_0^*, x^*, u^*) satisfies

$$u^* \in \arg \max_{u \in U} \Phi(x_0^*, x^*, u). \quad (5)$$

Indeed, for every iteration t ,

$$u^t \in \arg \max_{u \in U} \Phi(x_0^t, x^t, u) \quad (6)$$

holds by definition. Yet, this property need not be preserved at accumulation points. To ensure that it is, we make the following assumption.

A5.2. *The separation value function Φ is finite-valued and jointly continuous on $\mathbb{R} \times X \times U$.*

We now give the convergence theorem.

Theorem 4. *Let A3.1, A4.1, A5.1 and A5.2 hold and fix $\varepsilon = 0$. Then, any accumulation point (x_0^*, x^*, u^*) of the iterates (x_0^t, x^t, u^t) generated by Algorithm 1 satisfies (5). Moreover, we readily have*

$$\max_{u \in U} \Phi(x_0^*, x^*, u) \leq 0.$$

Proof. Since Φ is jointly continuous and U is non-empty and compact, Berge maximum theorem gives that the set-valued map

$$(x_0, x) \mapsto \arg \max_{u \in U} \Phi(x_0, x, u)$$

is outer semicontinuous. Let (x_0^*, x^*, u^*) be any accumulation point of the iterates and let $(t_i)_{i \in \mathbb{N}}$ be a subsequence such that $(x_0^{t_i}, x^{t_i}, u^{t_i}) \rightarrow (x_0^*, x^*, u^*)$. Since

$$u^{t_i} \in \arg \max_{u \in U} \Phi(x_0^{t_i}, x^{t_i}, u)$$

holds for every i by construction of Algorithm 1, outer semicontinuity yields

$$u^* \in \arg \max_{u \in U} \Phi(x_0^*, x^*, u). \quad (7)$$

By Lemma 3, $Q(x^*, u^*) \leq x_0^*$. Hence, by A5.1.v, $\Phi(x_0^*, x^*, u^*) \leq 0$ holds. Combining the last display with (7) gives

$$\max_{u \in U} \Phi(x_0^*, x^*, u) = \Phi(x_0^*, x^*, u^*) \leq 0. \quad \square$$

Here again, Theorem 4 can only guarantee a property related to the separation value function, absent any concrete choice of Φ . Yet, the following corollary leverages Theorem 4 to a complete description of the outcome of Algorithm 1 in the case $\varepsilon = 0$.

Corollary 3. *Let the assumptions of Theorem 4 hold and further assume that, for any $(x_0, x) \in \mathbb{R} \times X$, if*

$$\max_{u \in U} \Phi(x_0, x, u) \leq 0$$

holds, then, for all $u \in U$, there exists $y \in Y(x, u)$ such that $f(x, y, u) \leq x_0$. Then, one of the following three cases occurs.

- *Algorithm 1 correctly identifies Problem (1) as infeasible after a finite number of iterations.*
- *Algorithm 1 terminates after a finite number of iterations and returns a solution to Problem (1).*
- *Algorithm 1 does not terminate and, for any accumulation point (x_0^*, x^*, u^*) of its iterates (x_0^t, x^t, u^t) , x^* is a solution to Problem (1).*

We note that the additional condition in Corollary 3 is naturally satisfied by two out of the three separation value functions considered in Section 6.

We now illustrate Theorem 2, Theorem 4 and Corollary 3 on a simple two-stage robust problem.

Example 5. *We consider again the problem defined in Example 3. As previously discussed, $Q(x, \cdot)$ fails to be upper semicontinuous. Thus, the canonical separation value function $\Phi^{\text{can}}(x_0, x, u) = Q(x, u) - x_0$ does not satisfy the continuity assumptions used in the preceding theorems. Instead, we consider the separation value function induced by the constraint-violation-based representation*

$$\Phi(x_0, x, u) := \min_{y \in \mathbb{R}^{n_y}} [y - u - x_0]^+ + [u - yu]^+ + [-1 - y]^+ + [y - 1]^+.$$

To ease our exposition, we let Ψ denote the objective function of the separation problem defined above. Note that the standing assumption as well as A3.1, A4.1,

A5.1 and A5.2 are easily checked to be satisfied. Hence, Theorem 2, Theorem 4, and Corollary 3 apply.

We now trace the execution of Algorithm 1 in three steps: we first solve the master problem in closed form (as a function of the tested scenarios U^t), then the separation problem in closed form (as a function of x_0^t), and finally combine both into an explicit formula for every iterate.

The master problem. We assume that $U^1 = \{0\}$ holds. At iteration 1, the master problem reads

$$\min_{x_0, y_1} \quad x_0 \quad \text{s.t.} \quad x_0 \geq y_1, \quad y_1 \in [-1, 1],$$

where, for simplicity, we omit the trivial first-stage variable $x = 0$ from the model. Also note that the constraint “ $yu \geq u$ ” vanishes because of the initial scenario being $u = 0$. By inspection, $x_0^1 = -1$. Now, consider any later iteration, i.e., $t > 1$, and assume that there exists $u \in U^t$ such that $u \neq 0$ (this will be verified below when we study the separation problem). In this setting, the constraint “ $uy \geq u$ ” is equivalent to “ $y \geq 1$ ” and we conclude that

$$x_0^t = \begin{cases} -1 & \text{if } t = 1, \\ 1 - \min_{u \in U^t \setminus \{0\}} u & \text{otherwise,} \end{cases} \quad (8)$$

holds for any $t \geq 1$.

The separation problem. We first consider the minimization of $\Psi(x_0, x, \cdot, u)$ over $\mathbb{R}^{n_y}$. We distinguish two cases.

- Case $u = 0$. By (8), $x_0^t \geq -1$ always holds. Hence, the point $y = -1$ satisfies $[y - x_0]^+ = [-1 - x_0]^+ = 0$, $[-1 - y]^+ = [-1 + 1]^+ = 0$, and $[y - 1]^+ = [1 - 1]^+ = 0$, i.e., it attains the trivial lower bound 0 of $\Psi(x_0, x, \cdot, 0)$, hence it is a minimizer.
- Case $u > 0$. Observe that $\Psi(x_0, x, \cdot, u)$ is a convex piecewise linear function with breakpoints: $u + x_0$, -1 and 1 . Thus, its minimum occurs at the breakpoint where the slope changes sign. If $x_0 \leq 1 - u$ (so that $-1 \leq u + x_0 \leq 1$), the slopes across the three resulting intervals are $-u$, $1 - u$, and 2 (moving left to right from -1), and the sign change from $-u \leq 0$ to $1 - u \geq 0$ occurs at $y = u + x_0$, giving minimum value $u(1 - u - x_0)$. If instead $x_0 > 1 - u$ (so $-1 \leq 1 \leq u + x_0$), the same reasoning gives a sign change at $y = 1$ instead, with minimum value 0.

Combining the two cases above, we may write compactly

$$\Phi(x_0, x, u) = \min_{y \in \mathbb{R}} \Psi(x_0, x, y, u) = u[1 - u - x_0]^+.$$

Recall that the separation problem maximizes the above function over U and observe that, if $v^t > 0$, u^t must be a maximizer of $u \mapsto u(1 - u - x_0)$. The unique maximizer is given by

$$\tilde{u}^t = \frac{1 - x_0^t}{2}, \quad \text{with value} \quad \tilde{v}^t = \frac{(1 - x_0^t)^2}{4}.$$

For any $x_0^t \in [-1, 1]$, we have that $\tilde{u}^t \in U$. Therefore, it is optimal and it holds $u^t = \tilde{u}^t$ and $v^t = \tilde{v}^t$ for all $t > 1$.

Closed-form trajectory. We show by induction that $x_0^t = 1 - u^{t-1}$ for all $t > 1$. First, this holds for $t = 2$ by direct computation. Next, assuming it holds up to $t - 1$, the scenario u^{t-1} is the minimum of $U^t \setminus \{0\}$, so that (8) gives $x_0^t = 1 - u^{t-1}$, which completes the induction. Combining this with the closed-form formulas for

TABLE 1. Iterations of Algorithm 1 on Example 5.

t	x_0^t	u^t	v^t
1	-1	1	1
2	0	0.5	0.25
3	0.5	0.25	0.0625
4	0.75	0.125	0.015625
5	0.875	0.0625	0.00390625
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	$1 - 2^{2-t}$	2^{1-t}	4^{1-t}

the separation problem finally gives the recursion $u^t = u^{t-1}/2$. With $u^1 = 1$, this leads to

$$u^t = 2^{1-t}, \quad x_0^t = 1 - 2^{2-t}, \quad v^t = 4^{1-t}, \quad t \geq 1.$$

The first 5 iterations are reported in Table 1.

Conclusion. Consistently with Theorem 2, Algorithm 1 terminates after finitely many iterations for any fixed $\varepsilon > 0$, since $v^t \rightarrow 0$. Consistently with Theorem 4 and Corollary 3, we further have $x_0^t \rightarrow 1$, so that, even if $\varepsilon = 0$ were used instead (in which case Algorithm 1 never terminates, since $v^t > 0$ holds for every t), the sequence of iterates would still converge to a solution of Problem (1). This illustrates that Problem (1) can be addressed, in a well-defined sense, even though the inner supremum defining ξ is not attained.

Remark 1. We note that Assumption A5.2 may be weakened if the algorithm separately detects second-stage infeasibility before solving the separation problem. This strategy was, for instance, used by Lefebvre and Thuerauf (2026) in the context of robust bilevel optimization with wait-and-see follower.

5.1. Inner Representation of the Separation Value Function. Although the convergence analysis is stated directly in terms of the separation value function Φ , it is sometimes convenient to obtain Φ from an inner representation, say, of the form

$$\Phi(x_0, x, u) := \inf_{y \in \mathbb{R}^{n_y}} \Psi(x_0, x, y, u), \quad (9)$$

for some extended-real-valued function Ψ . We collect here sufficient conditions on such an inner representation ensuring the assumptions on Φ used above, while the convergence results themselves require only the corresponding assumptions on Φ .

Proposition 5. Assume that Φ admits the representation (9), where $\Psi : \mathbb{R} \times X \times \mathbb{R}^{n_y} \times U \rightarrow \mathbb{R}$ satisfies the following statements.

- i. For all $(x_0, x, y, u) \in \mathbb{R} \times X \times \mathbb{R}^{n_y} \times U$, $\Psi(x_0, x, y, u) > -\infty$ holds.
- ii. For all $(x_0, x, u) \in \mathbb{R} \times X \times U$, $\Psi(x_0, x, \cdot, u)$ is lower semicontinuous and level-bounded on $\mathbb{R}^{n_y}$.
- iii. For all $(x_0, x, y) \in \mathbb{R} \times X \times \mathbb{R}^{n_y}$, $\Psi(x_0, x, y, \cdot)$ is upper semicontinuous on U .

Then A4.1 holds. In addition, for every $(x_0, x, u) \in \mathbb{R} \times X \times U$, the infimum in (9) is attained (possibly equal to ∞). Moreover, if, for every $u \in U$, there exists $y \in \mathbb{R}^{n_y}$ such that $\Psi(x_0, x, y, u) \neq \infty$ holds, then $\Phi(x_0, x, \cdot)$ is finite-valued.

Proof. First, fix $(x_0, x, u) \in \mathbb{R} \times X \times U$. By the first hypothesis, $\Psi(x_0, x, y, u) > -\infty$ holds for every y and, by the second hypothesis, $\Psi(x_0, x, \cdot, u)$ is lower semicontinuous and level-bounded. Thus, it is proper whenever there exists y with $\Psi(x_0, x, y, u) < \infty$. In that case, Theorem 1 gives that its infimum is attained and finite. Otherwise,

$\Psi(x_0, x, \cdot, u) = \infty$, and the infimum, which then equals ∞ , is trivially attained by any y . In particular, $\Phi(x_0, x, u) > -\infty$ always holds.

Second, by the third hypothesis, $\Psi(x_0, x, y, \cdot)$ is upper semicontinuous on U for every fixed y . Hence, $\Phi(x_0, x, \cdot)$, being the pointwise infimum of upper semicontinuous functions, is upper semicontinuous on U . Thus, A4.1 holds. The last statement follows directly from the first part of the proof. $\square$

Proposition 6. *Let A3.1 hold. Assume that Φ admits the representation (9), where $\Psi : \mathbb{R} \times X \times \mathbb{R}^{n_y} \times U \rightarrow \bar{\mathbb{R}}$ satisfies the following statements.*

- i. *For all $y \in \mathbb{R}^{n_y}$, $\Psi(\cdot, \cdot, y, \cdot)$ is upper semicontinuous on $\mathbb{R} \times X \times U$.*
- ii. *for all $(x_0, x, u) \in \mathbb{R} \times X \times U$ and all $y \in Y(x, u)$ satisfying $f(x, y, u) \leq x_0$, we have $\Psi(x_0, x, y, u) \leq 0$.*

Then A5.1.iv and A5.1.v hold.

Proof. Since Φ is the pointwise infimum over y of the upper semicontinuous functions $\Psi(\cdot, \cdot, y, \cdot)$, A5.1.iv follows directly.

Now, let $(x_0, x, u) \in \mathbb{R} \times X \times U$ satisfy $Q(x, u) \leq x_0$. In particular, $Y(x, u)$ is non-empty. By Proposition 1.i, there exists $\hat{y} \in Y(x, u)$ such that $f(x, \hat{y}, u) = Q(x, u) \leq x_0$. Hence, $\Phi(x_0, x, u) \leq \Psi(x_0, x, \hat{y}, u) \leq 0$ holds, proving A5.1.v. $\square$

Proposition 7. *Assume that Φ admits the representation (9), where $\Psi : \mathbb{R} \times X \times \mathbb{R}^{n_y} \times U \rightarrow \bar{\mathbb{R}}$ satisfies the following statements.*

- i. *Ψ is jointly continuous on $\mathbb{R} \times X \times \mathbb{R}^{n_y} \times U$.*
- ii. *$\Psi(x_0, x, \cdot, u)$ is level-bounded, locally uniformly in (x_0, x, u) .*

Then A5.2 holds.

Proof. By the hypotheses, Ψ is jointly continuous and level-bounded in y , locally uniformly in (x_0, x, u) . Hence, Theorem 1.17 in Rockafellar and Wets (1998) gives that Φ is finite-valued and jointly continuous on $\mathbb{R} \times X \times U$. $\square$

5.2. A Special Note on Polyhedral Uncertainty Sets. Theorems 2 and 4 establish a general convergence theory for Algorithm 1 under topological assumptions on the problem data and the separation value function Φ . These assumptions, however, are not necessary for finite termination. When the uncertainty set U is a polytope and the separation value function satisfies an appropriate quasi-convexity property, a purely combinatorial proof of finite termination is available, requiring no continuity assumptions on Φ . Indeed, in this setting the separation problem always admits a solution at an extreme point of U , so only finitely many distinct scenarios can be generated before the algorithm terminates. This observation underlies the classical finite-termination analysis of CCG for polyhedral uncertainty sets; see, e.g., Zeng and Zhao (2013). The following theorem states this result.

Theorem 5. *Let A3.1 and A5.1.v as well as the following hold.*

- *U is a polytope.*
- *For all $(x_0, x) \in \mathbb{R} \times X$, $\Phi(x_0, x, \cdot)$ is quasi-convex on U .*
- *In Line 11, u^t is chosen such that $u^t \in \{u \in \text{ext}(U) : v^t = \Phi(x_0^t, x^t, u)\}$.*

Then, for any $\varepsilon \geq 0$, Algorithm 1 terminates after at most $|\text{ext}(U)| + 1$ iterations, with a valid proof that either Problem (1) is infeasible, or, the returned point (x_0^t, x^t) satisfies

$$\max_{u \in U} \Phi(x_0^t, x^t, u) \leq \varepsilon.$$

Proof. By Proposition 3 and A3.1, at every iteration t , the master problem is either infeasible or has a solution. By quasi-convexity of $\Phi(x_0, x, \cdot)$, we have

$$\sup_{u \in U} \Phi(x_0, x, u) = \max_{u \in \text{ext}(U)} \Phi(x_0, x, u). \quad (10)$$

Hence, the separation problem is solvable at every iteration t and Algorithm 1 is well-posed.

Assume that Algorithm 1 does not terminate. Then, the master problem must be feasible at every iteration and $v^t > \varepsilon$ holds for all t . Now, let $t_j > t_i$ for any iteration $t_i \geq 1$. Since, $U^{t_i} \subseteq U^{t_i+1} \subseteq U^{t_j}$ by construction, we have that the master problem at iteration t_j gives $y^{t_j}(u^{t_i}) \in Y(x^{t_j}, u^{t_i})$ such that $x_0^{t_j} \geq f(x^{t_j}, y^{t_j}(u^{t_i}), u^{t_i})$. By A5.1.v, we have $\Phi(x_0^{t_j}, x^{t_j}, u^{t_i}) \leq 0 \leq \varepsilon$, so that $u^{t_j} \neq u^{t_i}$. Since each u^t is an extreme point of U by construction (which exists due to (10)) and $\text{ext}(U)$ is finite, this contradicts non-termination after $|\text{ext}(U)| + 1$ iterations. $\square$

We note that the assumptions of Theorem 5 are naturally satisfied by a large class of convex two-stage robust optimization problems when using the canonical separation value functions. We illustrate it with an example.

Example 6. *For simplicity, assume that f and g are jointly continuous convex functions and that*

$$Y(x, u) := \{y \in \mathbb{R}^{n_y} : g(x, y, u) \leq 0\}$$

holds. Further assume that $Y(x, u) \neq \emptyset$ for all $(x, u) \in X \times U$. Then, choosing the canonical separation value function $\Phi(x_0, x, u) = Q(x, u) - x_0$ implies that $\Phi(x_0, x, \cdot)$ is convex, and hence quasi-convex, for every $(x_0, x) \in \mathbb{R} \times X$.

6. SEPARATION VALUE FUNCTIONS

In this section, we study concrete choices of the separation value function Φ and derive the guarantees they provide on the point returned by Algorithm 1. Indeed, by construction of Algorithm 1, if the algorithm terminates at iteration t , then either Problem (1) is infeasible, or it returns (x_0^t, x^t) satisfying

$$\max_{u \in U} \Phi(x_0^t, x^t, u) \leq \varepsilon.$$

Yet, this inequality by itself says nothing about the true recourse problem. The purpose of this section is to show that, whenever Φ satisfies one of the tolerance properties introduced in Section 6.1, this bound can be unpacked into a guarantee on the true objective and on an appropriate notion of approximate second-stage feasibility.

6.1. Tolerance Properties. We define three notions of tolerance guarantees that a general separation value function may satisfy. Before we do so, we introduce the following assumptions that we invoke whenever needed.

A6.1. *For all $x \in X$ and all $u \in U$, $Y(x, u)$ is nonempty.*

A6.2. *There exists $x \in X$ and $u \in U$ such that $Y(x, u)$ is nonempty.*

A6.3. *There exists a function $g : \mathbb{R}^{n_x+n_u+n_y} \rightarrow \mathbb{R}^{m_y}$ such that the second-stage feasible region is given by $Y(x, u) := \{y \in \mathbb{R}^{n_y} : g(x, y, u) \leq 0\}$.*

A6.1 is the standard complete-recourse assumption commonly used in the two-stage robust optimization literature. It rules out second-stage infeasibility for every first-stage decision and every uncertainty realization. On the other hand, A6.2 is considerably weaker. It only requires that $\text{graph}(Y)$ be nonempty. In particular, this assumption is sufficient to derive guarantees based on the graph of Y instead of $Y(x, u)$, thereby allowing us to consider points $x \in X$ such that $Y(x, u)$ is empty for some $u \in U$. Finally, A6.3 is a structural assumption used when feasibility is measured through constraint violations. It allows approximate feasibility to be quantified through residuals such as $[g(x, y, u)]^+$ and relaxed second-stage feasible regions obtained by perturbing the right-hand side of the defining inequalities.

Accordingly, we now give three tolerance properties tied to each of these assumptions.

Definition 2 (Pointwise Tolerance Property). *Let A6.1 hold and let $\tau \geq 0$ and $\gamma \geq 0$ be given. A separation value function $\Phi : \mathbb{R} \times X \times U \rightarrow \bar{\mathbb{R}}$ satisfies the (τ, γ) -pointwise tolerance property if and only if there exists $\bar{\varepsilon} \geq 0$ such that, for every $\varepsilon \in [0, \bar{\varepsilon}]$ and every $(x_0, x, u) \in \mathbb{R} \times X \times U$, it holds*

$$\Phi(x_0, x, u) \leq \varepsilon \implies \exists y \in \mathbb{R}^{n_y}, \begin{cases} f(x, y, u) - x_0 \leq \tau, \\ y \in Y(x, u) + B_\gamma. \end{cases}$$

Definition 3 (Graph Tolerance Property). *Let A6.2 hold and let $\tau \geq 0$ and $\gamma \geq 0$ be given. A separation value function $\Phi : \mathbb{R} \times X \times U$ satisfies the (τ, γ) -graph tolerance property if and only if there exists $\bar{\varepsilon} \geq 0$ such that, for every $\varepsilon \in [0, \bar{\varepsilon}]$ and every $(x_0, x, u) \in \mathbb{R} \times X \times U$, it holds*

$$\Phi(x_0, x, u) \leq \varepsilon \implies \exists y \in \mathbb{R}^{n_y}, \begin{cases} f(x, y, u) - x_0 \leq \tau, \\ \text{dist}((x, u, y), \text{graph}_{X,U}(Y)) \leq \gamma, \end{cases}$$

where $\text{graph}_{X,U}(Y) := \{(x, u, y) \in X \times U \times \bar{Y} : y \in Y(x, u)\}$ denotes the graph of Y restricted to $X \times U$. Moreover, the last condition is equivalent to

$$y \in \text{proj}_y((\text{graph}_{X,U}(Y) + B_\gamma) \cap (\{x\} \times \{u\} \times \mathbb{R}^{n_y})).$$

Definition 4 (Violation Tolerance Property). *Let A6.3 hold and let $\tau \geq 0$ and $\gamma \geq 0$ be given. A separation value function $\Phi : \mathbb{R} \times X \times U$ satisfies the (τ, γ) -violation tolerance property if and only if there exists $\bar{\varepsilon} \geq 0$ such that, for every $\varepsilon \in [0, \bar{\varepsilon}]$ and every $(x_0, x, u) \in \mathbb{R} \times X \times U$, it holds*

$$\Phi(x_0, x, u) \leq \varepsilon \implies \exists y \in \mathbb{R}^{n_y}, \begin{cases} f(x, y, u) - x_0 \leq \tau, \\ y \in Y_\gamma(x, u), \end{cases}$$

where $Y_\gamma(x, u) = \{y \in \mathbb{R}^{n_y} : g(x, y, u) \leq \gamma e\}$.

The three tolerance properties capture different notions of approximation and are generally distinct. Pointwise tolerance measures proximity to the feasible set $Y(x, u)$ for the prescribed pair (x, u) , while graph tolerance allows comparison with nearby feasible triples in $\text{graph}_{X,U}(Y)$ and therefore remains meaningful even when $Y(x, u) = \emptyset$. Violation tolerance is instead tied to a particular constraint representation of Y . The corresponding unrelaxed feasibility notions, i.e., with $\varepsilon = 0$, all coincide with exact feasibility. Next, we give an example illustrating how these relaxed feasibility notions lead to different guarantees.

Example 7. *Let $\gamma = 1/2$ be a desired feasibility tolerance and consider the second-stage feasible region given by*

$$Y(u) := \{y \in \mathbb{R} : |u| + 2|y| \leq 1\},$$

as well as $U = [-3/2, 3/2]$. Note that Y does not depend on the first-stage decisions for simplicity. The three different tolerance properties are illustrated in Figure 2. Whenever needed, we consider the ℓ_2 -norm ball of the respective space.

The pointwise tolerance property keeps u fixed and only perturbs the second-stage variables. Thus, whenever $Y(u) \neq \emptyset$, we have

$$Y(u) + B_\gamma = \left\{ y : |y| \leq \frac{1 - |u|}{2} + \frac{1}{2} \right\}.$$

Note that those u such that $Y(u) = \emptyset$ holds are excluded by definition.

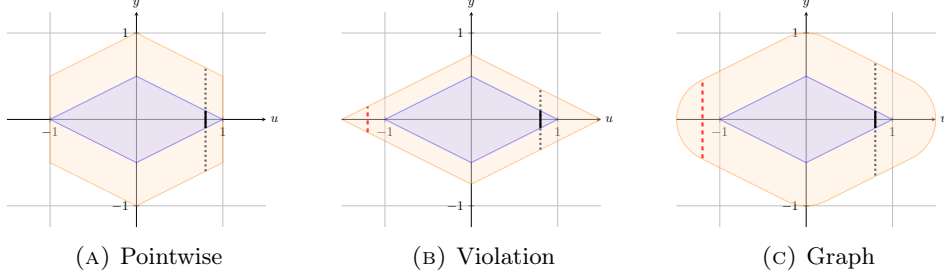


FIGURE 2. The different tolerance properties in Example 7.

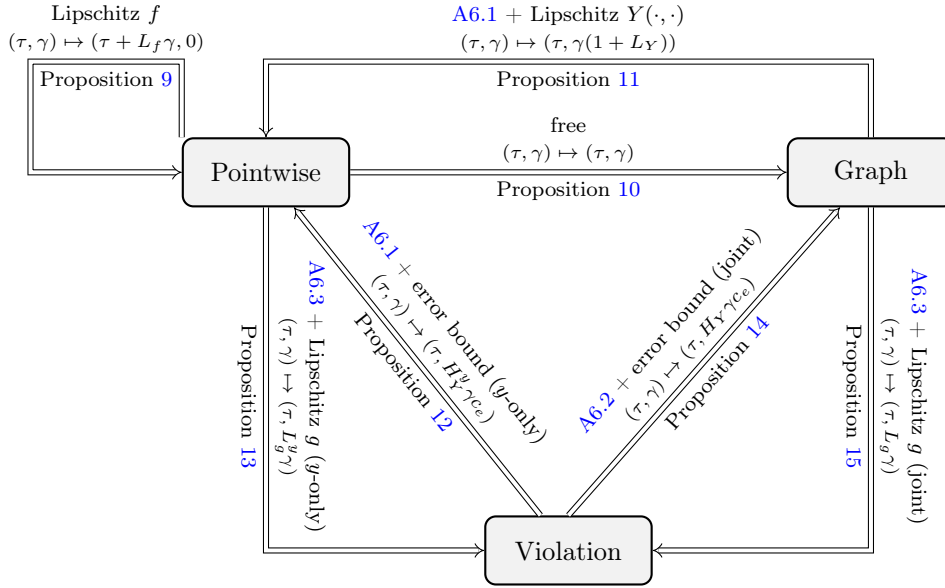


FIGURE 3. Relations between the parameterized tolerance properties

In contrast, the violation tolerance property relaxes the defining inequality and gives

$$Y_\gamma(u) = \left\{ y : |u| + 2|y| \leq \frac{3}{2} \right\}.$$

Hence, we may have that $Y_\gamma(u)$ is nonempty even for some u with $Y(u) = \emptyset$.

Finally, graph tolerance enlarges the whole graph of Y , allowing both u and y to move. Consequently, a point corresponding to an infeasible realization u may still be close to a feasible triple associated with some nearby $\bar{u}$. We see that the graph tolerance property is close in spirit to the violation tolerance property without requiring a functional representation of the set-valued map Y .

The example above also shows that these notions should not, in general, be identified with one another. Nevertheless, under additional regularity conditions, they may be related. Figure 3 summarizes such relations, while the corresponding proofs are deferred to the appendix. Each arrow indicates that one tolerance property can be transferred to another under the additional assumptions displayed along the arrow. The parameters (τ, γ) may change under such a transfer; the resulting parameter transformation is indicated below the corresponding arrow.

6.2. Canonical Separation Value Function. One of the most commonly used separation value functions in the literature is obtained by simply solving the second-stage problem. We call this the canonical separation value function. Formally, it is defined by

$$\Phi^{\text{can}}(x_0, x, u) := Q(x, u) - x_0, \quad (11)$$

and leads to the separation problem

$$\max_{u \in U} \Phi^{\text{can}}(x_0, x, u) = \max_{u \in U} \min_{y \in Y(x, u)} f(x, y, u) - x_0,$$

where the right-hand side is understood in the extended-real sense when $Y(x, u) = \emptyset$. Hence, Φ^{can} tests the true recourse value without any approximation. Under the attainment conditions below, it satisfies the requested pointwise tolerance properties. This is formalized in the next lemma.

Lemma 5. *Let the following statements hold.*

- i. *For all $(x, u) \in X \times U$, $Y(x, u)$ is non-empty and compact.*
- ii. *For all $(x, u) \in X \times U$, $f(x, \cdot, u)$ is lower semicontinuous.*

Then, for every $\tau \geq 0$ and $\gamma \geq 0$, Φ^{can} satisfies the (τ, γ) -pointwise tolerance property with $\bar{\varepsilon} = \tau$.

Proof. Let $\tau \geq 0$, $\gamma \geq 0$, $0 \leq \varepsilon \leq \bar{\varepsilon} := \tau$, and $(x_0, x, u) \in \mathbb{R} \times X \times U$ be such that $\Phi^{\text{can}}(x_0, x, u) \leq \varepsilon$ holds. By compactness of $Y(x, u)$ and lower semicontinuity of $f(x, \cdot, u)$, there exists $y \in Y(x, u)$ such that $f(x, y, u) = Q(x, u)$ holds. Hence, by (11), we have

$$f(x, y, u) - x_0 = Q(x, u) - x_0 = \Phi^{\text{can}}(x_0, x, u) \leq \varepsilon \leq \tau.$$

Trivially, $y \in Y(x, u) \subseteq Y(x, u) + B_\gamma$, which proves the (τ, γ) -pointwise tolerance property. $\square$

We now give sufficient conditions for the finite termination of Algorithm 1 with $\Phi = \Phi^{\text{can}}$ and a sufficiently small positive stopping tolerance.

Theorem 6. *Let A6.1 hold. Suppose that the following statements hold.*

- i. *f is jointly continuous.*
- ii. *For all $u \in U$, $Y(\cdot, u)$ is outer semicontinuous.*
- iii. *For all $x \in X$, $Y(x, \cdot)$ is outer semicontinuous.*
- iv. *Y is jointly inner semicontinuous relative to $X \times U$.*

Let $\tau \geq 0$ and $\gamma \geq 0$ be given and set $\bar{\varepsilon} = \tau$. Then Φ^{can} satisfies the (τ, γ) -pointwise tolerance property, and A3.1, A4.1, and A5.1 hold for $\Phi = \Phi^{\text{can}}$.

In particular, for any $\varepsilon \in (0, \bar{\varepsilon}]$ and $\Phi = \Phi^{\text{can}}$, Algorithm 1 is well-posed and terminates after a finite number of iterations, returning (x_0^t, x^t) such that, for every $u \in U$, there exists $y \in Y(x^t, u)$ that satisfies $f(x^t, y, u) \leq x_0^t + \tau$.

Proof. Joint continuity of f and outer semicontinuity of $Y(\cdot, u)$ for every fixed $u \in U$ imply A3.1. Moreover, joint continuity of f gives A5.1.i, while outer semicontinuity of $Y(x, \cdot)$ for every fixed $x \in X$ gives A5.1.ii. Since X , $\bar{Y}$, and U are compact, joint continuity of f also gives A5.1.iii.

By joint inner semicontinuity of Y relative to $X \times U$ and Lemma 1.i, Q is jointly upper semicontinuous on $X \times U$. Moreover, Q is finite-valued on $X \times U$ since $Y(x, u)$ is non-empty for every $(x, u) \in X \times U$. Hence, Φ^{can} satisfies A4.1. Joint upper semicontinuity of Q gives A5.1.iv, while A5.1.v follows directly from (11). Consequently, A3.1, A4.1, and A5.1 hold, and Theorem 2 applies.

Algorithm 1 returns (x_0^t, x^t) satisfying $\max_u \{\Phi^{\text{can}}(x_0^t, x^t, u) : u \in U\} \leq \varepsilon$. Hence, for every $u \in U$, $\Phi^{\text{can}}(x_0^t, x^t, u) \leq \varepsilon$ holds. The result then follows from Lemma 5. $\square$

We end this section on the canonical separation value function with a convergence result.

Theorem 7. *Let A6.1 hold. Suppose that the following statements hold.*

- i. f is jointly continuous.*
- ii. Y is jointly continuous on $X \times U$.*

Then A3.1, A4.1, A5.1, and A5.2 hold for $\Phi = \Phi^{\text{can}}$.

In particular, for $\varepsilon = 0$ and $\Phi = \Phi^{\text{can}}$, Algorithm 1 is well-posed and one of the following two cases occurs.

- Algorithm 1 terminates after a finite number of iterations and returns a solution to Problem (1).*
- Algorithm 1 does not terminate and, for any accumulation point (x_0^*, x^*, u^*) of its iterates (x_0^t, x^t, u^t) , x^* is a solution to Problem (1).*

Proof. Joint continuity of f and outer semicontinuity of $Y(\cdot, u)$ for every fixed $u \in U$ imply A3.1. Moreover, joint continuity of f gives A5.1.i. Joint outer semicontinuity of Y on $X \times U$ implies outer semicontinuity of $Y(x, \cdot)$ for every fixed $x \in X$, and therefore gives A5.1.ii. Since X , $\bar{Y}$, and U are compact, joint continuity of f also gives A5.1.iii.

By joint inner and outer semicontinuity of Y on $X \times U$ and Lemma 1, Q is jointly upper and lower semicontinuous on $X \times U$. Moreover, Q is finite-valued since $Y(x, u)$ is non-empty for every $(x, u) \in X \times U$. Hence, Φ^{can} is finite-valued and jointly continuous on $\mathbb{R} \times X \times U$. Furthermore, A5.1.v follows directly from (11). Consequently, A3.1, A4.1, A5.1, and A5.2 hold. Thus, Theorem 4 applies.

Finally, if $\max_u \{\Phi^{\text{can}}(x_0, x, u) : u \in U\} \leq 0$ holds, then $Q(x, u) \leq x_0$ for every $u \in U$. Since $Y(x, u)$ is non-empty and compact and $f(x, \cdot, u)$ is continuous, the minimum defining $Q(x, u)$ is attained. Therefore, for every $u \in U$, there exists $y \in Y(x, u)$ such that $f(x, y, u) \leq x_0$ holds. Hence, the additional condition of Corollary 3 is satisfied, and the result follows. $\square$

6.3. Residual-Based Separation Value Function. In Ayoub and Poss (2016), the authors introduce a separation strategy based on the Farkas lemma applied to a linear second-stage problem. Later, it was also extended to a more general convex setting by Lefebvre et al. (2025). Interestingly, this separation problem can be obtained by resorting to a residual maximization problem that fits the general theory detailed in this paper. To that end, let A6.3 hold and let $\rho \in \mathbb{R}_{>0}^{m_y}$ be given. We define the (weighted) residual-based inner objective and separation value function as

$$\Psi_\rho^{\text{res}}(x_0, x, y, u) := \left\| \begin{bmatrix} f(x, y, u) - x_0 \\ \text{diag}(\rho) \cdot g(x, y, u) \end{bmatrix}^+ \right\|,$$

and,

$$\Phi_\rho^{\text{res}}(x_0, x, u) := \inf_{y \in \mathbb{R}^{n_y}} \Psi_\rho^{\text{res}}(x_0, x, y, u),$$

respectively. The next example shows how the results from Lefebvre et al. (2025), and, thus, those from Ayoub and Poss (2016), can be derived.

Example 8. *We consider the setting presented in Lefebvre et al. (2025), i.e., we study the convex adjustable robust problem*

$$\begin{aligned} \min_{x_0, x} \quad & x_0 \\ \text{s.t.} \quad & x \in X, \\ & \forall u \in U, \exists y \in Y(x, u) : x_0 \geq f(x, y), \end{aligned}$$

where $Y(x, u) := \{y \in \mathbb{R}^{n_y} : T(x)u + \tilde{g}(x, y) \leq 0\}$ and, for fixed x , the functions $\tilde{g}_i(x, \cdot)$, $i = 1, \dots, m_y$, as well as $f(x, \cdot)$ are convex and real-valued. Let $(x_0, x) \in$

$\mathbb{R} \times X$ be given and let ρ be the vector of all ones. Using the ℓ_1 -norm, the separation problem reads

$$\begin{aligned} & \max_{u \in U} \min_{y \in \mathbb{R}^{n_y}} \left\| \begin{bmatrix} f(x, y) - x_0 \\ T(x)u + \tilde{g}(x, y) \end{bmatrix}^+ \right\|_1 \\ &= \max_{u \in U} \min_{y \in \mathbb{R}^{n_y}} [f(x, y) - x_0]^+ + \sum_{i=1}^{m_y} [T_i(x)u + \tilde{g}_i(x, y)]^+. \end{aligned}$$

Observe that the inner minimization problem is convex, feasible and bounded from below. Thus, Fenchel duality gives that the separation problem can be expressed as

$$\begin{aligned} & \max_{u, \lambda_0, \lambda, z^0, \dots, z^m} -\lambda_0(f|_x)^* \left(\frac{z^0}{\lambda_0} \right) - \sum_{i=1}^{m_y} \lambda_i(\tilde{g}_i|_x)^* \left(\frac{z^i}{\lambda_i} \right) + \lambda^\top T(x)u - \lambda_0 x_0 \\ & \text{s.t.} \quad \sum_{i=0}^{m_y} z^i = 0, \quad \|(\lambda_0, \lambda)\|_\infty \leq 1, \quad u \in U, \\ & \quad (\lambda_0, \lambda) \in \mathbb{R}_{\geq 0}^{m_y+1}, \\ & \quad z^i \in \mathbb{R}^{n_y}, \quad i = 0, \dots, m_y. \end{aligned}$$

Here, $f|_x := f(x, \cdot)$ and $(f|_x)^*$ denotes its convex conjugate defined as

$$(f|_x)^*(\pi) := \sup_{y \in \mathbb{R}^{n_y}} \pi^\top y - f|_x(y).$$

Moreover, $\lambda_0(f|_x)^*(z^0/\lambda_0)$ denotes the closed perspective of $(f|_x)^*$; see theorem 1 from Lefebvre et al. (2025) for more details. When $f(x, \cdot)$ and $\tilde{g}_i(x, \cdot)$ are affine, their convex conjugates reduce to indicator functions of affine equalities, and the formulation recovers the bilinear separation problem of Ayoub and Poss (2016). For more details, we refer to Example 1 of Lefebvre et al. (2025).

We first show that the residual separation value satisfies the requested violation tolerance properties.

Lemma 6. Let A6.3 hold and let $\rho \in \mathbb{R}_{>0}^{m_y}$ and a norm $\|\cdot\|$ on $\mathbb{R}^{m_y+1}$ be given. Suppose that the inner residual infimum defining Φ_ρ^{res} is attained for every $(x_0, x, u) \in \mathbb{R} \times X \times U$. For $i = 0, \dots, m_y$, let $\mu_i := \max \{|v_i| : v \in \mathbb{R}^{m_y+1}, \|v\| \leq 1\}$. Then, for every $\tau \geq 0$ and $\gamma \geq 0$, Φ_ρ^{res} satisfies the (τ, γ) -violation tolerance property with

$$\bar{\varepsilon} = \min \left\{ \frac{\tau}{\mu_0}, \min_{i=1, \dots, m_y} \frac{\gamma \rho_i}{\mu_i} \right\}.$$

Proof. The constants μ_i are positive and finite by equivalence of norms in finite dimension, so $\bar{\varepsilon} \geq 0$. Let $\tau \geq 0$, $\gamma \geq 0$, $0 \leq \varepsilon \leq \bar{\varepsilon}$, and let (x_0, x, u) satisfy $\Phi_\rho^{\text{res}}(x_0, x, u) \leq \varepsilon$. Choose a minimizer y . The residual vector has norm at most ε . By the definition of μ_i , every coordinate of a vector $v \in \mathbb{R}^{m_y+1}$ satisfies $|v_i| \leq \mu_i \|v\|$. Applied to the residual vector, this gives $[f(x, y, u) - x_0]^+ \leq \varepsilon \mu_0$, and hence $f(x, y, u) \leq x_0 + \varepsilon \mu_0$. For every $i = 1, \dots, m_y$, it also gives $\rho_i [g_i(x, y, u)]^+ \leq \varepsilon \mu_i$. Since $\rho_i > 0$, $g_i(x, y, u) \leq [g_i(x, y, u)]^+ \leq \varepsilon \mu_i / \rho_i \leq \gamma$. Similarly, $f(x, y, u) - x_0 \leq \varepsilon \mu_0 \leq \tau$. Therefore, $y \in Y_\gamma(x, u)$ holds. $\square$

Remark 2. Level-boundedness of the residual is easy to verify for common linear systems. If $Y(x, u) = \{y : A(x, u)y \leq b(x, u)\}$ and the second-stage objective is linear in y , the residual is level-bounded whenever the recourse inequalities together with the objective inequality have no nonzero recession direction. In particular, bounded polyhedral recourse is sufficient.

Combining this lemma with the inner-representation results developed in Section 5.1 yields some sufficient conditions for the finite termination of Algorithm 1 with a positive tolerance.

Theorem 8. *Let A6.3 hold and let $\tau \geq 0$, $\gamma \geq 0$, $\rho \in \mathbb{R}_{>0}^{m_y}$, and a norm $\|\cdot\|$ on $\mathbb{R}^{m_y+1}$ be given. Suppose that the following statements hold.*

- i. *f and g are jointly continuous.*
- ii. *For all $(x_0, x, u) \in \mathbb{R} \times X \times U$, $\Psi_\rho^{\text{res}}(x_0, x, \cdot, u)$ is level-bounded.*

Let $\bar{\varepsilon}$ be as in Lemma 6. Then, A3.1, A4.1, and A5.1 hold for $\Phi = \Phi_\rho^{\text{res}}$.

In particular, for any $\varepsilon \in (0, \bar{\varepsilon}]$ and $\Phi = \Phi_\rho^{\text{res}}$, Algorithm 1 is well-posed and one of the following two cases occurs.

- *Algorithm 1 correctly identifies Problem (1) as infeasible after a finite number of iterations.*
- *Algorithm 1 terminates after a finite number of iterations and returns (x_0^t, x^t) such that, for every $u \in U$, there exists $y \in Y_\gamma(x^t, u)$ satisfying $f(x^t, y, u) \leq x_0^t + \tau$.*

Proof. Continuity of g makes $\text{graph}(Y)$ closed and hence gives outer semicontinuity of both $Y(\cdot, u)$ and $Y(x, \cdot)$. Together with continuity of f , this gives A3.1 and A5.1.i–A5.1.ii. Compactness of $X \times \bar{Y} \times U$ also gives A5.1.iii. Ψ_ρ^{res} is nonnegative and jointly continuous. Thus, Proposition 5 gives A4.1 and attainment. Its residuals vanish whenever $y \in Y(x, u)$ and $f(x, y, u) \leq x_0$. Hence, Proposition 6 gives A5.1.iv and A5.1.v. Consequently, A3.1, A4.1, and A5.1 hold. Theorem 2 gives the two alternatives. If the algorithm returns (x_0^t, x^t) , then $\Phi_\rho^{\text{res}}(x_0^t, x^t, u) \leq \varepsilon$ for every $u \in U$. The claimed result follows from Lemma 6. $\square$

We next focus on the convergence result using $\varepsilon = 0$. The only additional requirement is locally uniform level-boundedness of the inner residual problem, i.e., for every $(\bar{x}_0, \bar{x}, \bar{u}) \in \mathbb{R} \times X \times U$ and $\alpha \in \mathbb{R}$, there exist a neighborhood V of $(\bar{x}_0, \bar{x}, \bar{u})$ and a bounded set $\Omega \subseteq \mathbb{R}^{n_y}$ such that $\{y: \Psi_\rho^{\text{res}}(x_0, x, y, u) \leq \alpha\} \subseteq \Omega$ holds for all $(x_0, x, u) \in V$.

Theorem 9. *Let A6.3 hold and let $\rho \in \mathbb{R}_{>0}^{m_y}$ and a norm $\|\cdot\|$ on $\mathbb{R}^{m_y+1}$ be given. Suppose that the following statements hold.*

- i. *f and g are jointly continuous.*
- ii. *$\Psi_\rho^{\text{res}}(x_0, x, \cdot, u)$ is level-bounded locally uniformly in (x_0, x, u) .*

Then A3.1, A4.1, A5.1, and A5.2 hold for $\Phi = \Phi_\rho^{\text{res}}$.

In particular, for $\varepsilon = 0$ and $\Phi = \Phi_\rho^{\text{res}}$, Algorithm 1 is well-posed and one of the following three alternatives occurs.

- *Algorithm 1 correctly identifies Problem (1) as infeasible after a finite number of iterations.*
- *Algorithm 1 terminates after a finite number of iterations and returns a solution to Problem (1).*
- *Algorithm 1 does not terminate and, for any accumulation point (x_0^*, x^*, u^*) of its iterates (x_0^t, x^t, u^t) , x^* is a solution to Problem (1).*

Proof. Continuity of g gives a closed graph and outer semicontinuity of $Y(\cdot, u)$ and $Y(x, \cdot)$. Continuity of f and compactness of $X \times \bar{Y} \times U$ therefore give A3.1 and A5.1.i–A5.1.iii. The inner objective is finite-valued, nonnegative and jointly continuous. Pointwise level-boundedness and Proposition 5 give A4.1 and attainment. Its residuals vanish at any feasible y with $f(x, y, u) \leq x_0$, so Proposition 6 gives A5.1.iv and A5.1.v. Proposition 7 gives A5.2. Consequently, A3.1, A4.1, A5.1, and A5.2 hold. Thus, Theorem 4 applies. If $\max_{u \in U} \Phi_\rho^{\text{res}}(x_0, x, u) \leq 0$, nonnegativity gives $\Phi_\rho^{\text{res}}(x_0, x, u) = 0$ for every $u \in U$. Attainment and positivity of ρ then give,

for every u , a point y with $g(x, y, u) \leq 0$ and $f(x, y, u) \leq x_0$. Thus, the additional condition of Corollary 3 is satisfied. $\square$

Thus, residual separation gives explicit violation bounds at positive tolerance and recovers exact feasibility at zero tolerance. To the best of our knowledge, the latter result is new even in the convex setting of Lefebvre et al. (2025).

6.4. Graph-Distance Separation Value Function. While the residual-based separation value function allows us to consider infeasible first-stage decisions, it requires the functional representation from A6.3. In contrast, the canonical separation value function uses a general set-valued map but excludes infeasible first-stage decisions. We introduce a separation value function that does not require a functional representation of the second-stage feasible set and measures feasibility geometrically through distance to its restricted graph. Let $\rho > 0$ be given and assume that A6.2 holds. We define the graph-distance inner objective and separation value function as

$$\Psi_\rho^{\text{gr}}(x_0, x, y, u) := [f(x, y, u) - x_0]^+ + \rho \text{dist}((x, u, y), \text{graph}_{X,U}(Y)),$$

and

$$\Phi_\rho^{\text{gr}}(x_0, x, u) := \inf_{y \in \bar{Y}} \Psi_\rho^{\text{gr}}(x_0, x, y, u), \quad (12)$$

respectively. We now give an example in a setting without complete recourse, i.e., A6.1 fails to be satisfied.

Example 9. Let us define $X = U = [0, 1]$ as well as

$$Y(x, u) := \begin{cases} \{0\} & \text{if } u \leq x, \\ \emptyset & \text{otherwise.} \end{cases}$$

We consider the two-stage robust problem defined by

$$\min_{x \in X} x \quad \text{s.t.} \quad \forall u \in U, \exists y \in Y(x, u).$$

It is easy to see that $x = 1$ is the only feasible point to this problem and that $\bar{Y} = \{0\}$ is a suitable choice so that $Y(x, u) \subseteq \bar{Y}$ holds for all $x \in X$ and all $u \in U$. The graph of Y can be expressed as

$$\text{graph}_{X,U}(Y) = \{(x, u, y) \in X \times U \times \bar{Y} : u \leq x\}.$$

Thus, using the ℓ_2 -norm, we have that $\text{dist}((x, u, y), \text{graph}_{X,U}(Y)) = [u - x]^+ / \sqrt{2}$ holds for any $(x, y, u) \in X \times \bar{Y} \times U$. Since $y = 0$ is the only point in $\bar{Y}$, the graph-distance separation value function reads

$$\Phi_\rho^{\text{gr}}(x_0, x, u) = [x - x_0]^+ + \frac{\rho}{\sqrt{2}}[u - x]^+.$$

Now, at any iteration of Algorithm 1, any iterate (x_0^t, x^t) trivially satisfies $x_0^t = x^t$. Hence, $\Phi_\rho^{\text{gr}}(x_0^t, x^t, u) = (\rho/\sqrt{2})[u - x^t]^+$ holds and we obtain

$$\max_{u \in U} \Phi_\rho^{\text{gr}}(x_0^t, x^t, u) = \frac{\rho}{\sqrt{2}}(1 - x^t).$$

Thus, in contrast to the canonical separation value function that assigns ∞ to all infeasible first-stage decisions, the graph-distance separation value function can distinguish slightly infeasible scenarios from strongly infeasible ones.

We next show that the graph-distance separation value function satisfies the requested graph tolerance properties.

Lemma 7. Let A6.2 hold and let a norm on $\mathbb{R}^{n_x + n_u + n_y}$ be given. Suppose that f is jointly continuous and that $\text{graph}_{X,U}(Y)$ is closed. Then, for every $\tau > 0$ and $\gamma > 0$ and every $\rho \geq \tau/\gamma$, Φ_ρ^{gr} satisfies the (τ, γ) -graph tolerance property with $\bar{\varepsilon} = \tau$.

Proof. By A6.2, $\text{graph}_{X,U}(Y)$ is non-empty. It is compact by closedness and compactness of $X \times U \times \bar{Y}$. Hence, distance to $\text{graph}_{X,U}(Y)$ is finite-valued and continuous. Since f is continuous and $\bar{Y}$ is compact, the inner infimum defining Φ_ρ^{gr} is attained.

Let $\tau > 0$, $\gamma > 0$, $\rho \geq \tau/\gamma$, and $0 \leq \varepsilon \leq \bar{\varepsilon} := \tau$. Suppose that $\Phi_\rho^{\text{gr}}(x_0, x, u) \leq \varepsilon$ and let $y_\rho \in \bar{Y}$ be a minimizer. By nonnegativity of the two terms in the inner objective,

$$[f(x, y_\rho, u) - x_0]^+ \leq \Phi_\rho^{\text{gr}}(x_0, x, u) \leq \varepsilon$$

and

$$\rho \text{dist}((x, u, y_\rho), \text{graph}_{X,U}(Y)) \leq \Phi_\rho^{\text{gr}}(x_0, x, u) \leq \varepsilon.$$

Therefore, $f(x, y_\rho, u) - x_0 \leq \varepsilon \leq \tau$ and

$$\text{dist}((x, u, y_\rho), \text{graph}_{X,U}(Y)) \leq \frac{\varepsilon}{\rho} \leq \frac{\tau}{\rho} \leq \gamma.$$

This proves the (τ, γ) -graph tolerance property with $\bar{\varepsilon} = \tau$. $\square$

We end this section with a finite termination result.

Theorem 10. *Let A6.2 hold and let a norm on $\mathbb{R}^{n_x+n_u+n_y}$ be given. Suppose that the following statements hold.*

- i. *f is jointly continuous.*
- ii. *Y is jointly outer semicontinuous.*

Let $\tau > 0$ and $\gamma > 0$ be given, choose any $\rho \geq \tau/\gamma$, and set $\bar{\varepsilon} = \tau$. Then A3.1, A4.1, A5.1, and A5.2 hold for $\Phi = \Phi_\rho^{\text{gr}}$, and Φ_ρ^{gr} satisfies the (τ, γ) -graph tolerance property.

In particular, for every $\varepsilon \in (0, \bar{\varepsilon}]$ and $\Phi = \Phi_\rho^{\text{gr}}$, Algorithm 1 is well-posed and one of the following two cases occurs.

- *Algorithm 1 correctly identifies Problem (1) as infeasible after a finite number of iterations.*
- *Algorithm 1 terminates after a finite number of iterations and returns (x_0^t, x^t) such that, for every $u \in U$, there exists $y \in \mathbb{R}^{n_y}$ satisfying $f(x^t, y, u) \leq x_0^t + \tau$ and $\text{dist}((x^t, u, y), \text{graph}_{X,U}(Y)) \leq \gamma$.*

Proof. Joint outer semicontinuity of Y gives both A3.1.iv and A5.1.ii. Together with joint continuity of f , this gives A3.1 and A5.1.i–A5.1.iii. Moreover, joint outer semicontinuity of Y implies that $\text{graph}(Y)$ is closed. Hence,

$$\text{graph}_{X,U}(Y) = \text{graph}(Y) \cap (X \times U \times \bar{Y})$$

is closed. By A6.2, it is also non-empty, and therefore compact as a closed subset of the compact set $X \times U \times \bar{Y}$. Hence, distance to $\text{graph}_{X,U}(Y)$ is finite-valued and continuous. Since $s \mapsto [s]^+$ is continuous, Ψ_ρ^{gr} is finite-valued, nonnegative, and jointly continuous on $\mathbb{R} \times X \times \bar{Y} \times U$. Since $\bar{Y}$ is non-empty and compact, Lemma 1 gives attainment of the inner infimum and joint continuity and finite-valuedness of Φ_ρ^{gr} follows from Berge theorem. Thus, A4.1 and A5.2 hold, and joint continuity gives A5.1.iv.

Finally, suppose that $Q(x, u) \leq x_0$. By A3.1, the minimum defining $Q(x, u)$ is attained at some $y \in Y(x, u)$. For this point, graph distance is zero and $[f(x, y, u) - x_0]^+ = 0$. Therefore, $\Phi_\rho^{\text{gr}}(x_0, x, u) = 0$, where equality follows from nonnegativity. Hence, A5.1.v holds. Consequently, A3.1, A4.1, A5.1, and A5.2 hold. Theorem 2 gives the two alternatives for every positive $\varepsilon \leq \bar{\varepsilon}$, and Lemma 7 gives the stated (τ, γ) -graph tolerance guarantee. $\square$

7. APPLICATION: ROBUST BILEVEL OPTIMIZATION WITH A CONVEX FOLLOWER

In this section, we apply the preceding results to robust bilevel optimization problems with a wait-and-see convex follower. We derive sufficient conditions for the existence of solutions and for the convergence of Algorithm 1 directly from the data of the follower problem.

7.1. Problem Statement. Bilevel optimization models a hierarchical decision process in which a first decision maker, called the leader, anticipates the optimal response of a second decision maker, called the follower; see, e.g., Beck et al. (2026) for an introduction. Here, some parameters of the follower problem are uncertain from the leader’s perspective. The leader first chooses x , the uncertainty realization u is then revealed, and the follower solves its problem knowing both x and u . The resulting decision timing is

leader chooses x $\curvearrowright$ uncertainty u is revealed $\curvearrowright$ follower chooses $y \in Y(x, u)$.

This timing is referred to as a “wait-and-see follower”; see, e.g., Beck et al. (2023). Following Lefebvre and Thuerlauf (2026), we adopt the optimistic convention: if the follower problem has multiple solutions, the one most favorable to the leader is selected.

We consider Problem (1) with $Y(x, u)$ given by the solution set of the follower problem. Let $F(x, u) \subseteq \bar{Y}$ be its feasible region and let $h : \mathbb{R}^{n_x+n_y+n_u} \rightarrow \mathbb{R}$ be its objective function. For every $x \in X$ and $u \in U$, the follower solves

$$\min_{y \in F(x, u)} h(x, y, u), \quad (13)$$

Its optimal-value function is

$$\varphi(x, u) := \min_{y \in F(x, u)} h(x, y, u).$$

and its solution map is

$$Y(x, u) := \arg \min_{y \in F(x, u)} h(x, y, u).$$

We impose the following baseline assumption throughout this section.

A7.1. *The following statements hold.*

- i. h is jointly continuous.
- ii. For all $(x, u) \in X \times U$, $F(x, u)$ is non-empty.

Note that the leader objective f need not coincide with h and is not assumed to be convex.

7.2. Regularity of the Follower Solution Map. We give two alternative structural conditions under which the follower solution map has the regularity required for the convergence theory developed in this paper. The first one uses convex inequality constraints and a pointwise Slater condition. The second one uses a right-hand-side parameterization and weak analyticity.

Following Bank et al. (1982), a function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is weakly analytic if, for every $y, d \in \mathbb{R}^n$, whenever the univariate function $\alpha \mapsto \phi(y + \alpha d)$ is constant on a nonempty open interval, it is constant on the entire line. Affine functions, convex polynomial functions, and every real-analytic function are weakly analytic. A real-analytic function is C^∞ and locally coincides with its convergent Taylor series; the converse does not hold in general. Weak analyticity is less restrictive than real analyticity and requires neither differentiability nor a local power-series representation. It is also not a constraint qualification and, more specifically, it does not require strict feasibility. This is useful in the present robust setting, where

worst-case uncertainty realizations may naturally place the follower feasible region on its boundary and thus make Slater-type assumptions unattractive. We now state the two assumptions.

A7.2. *The following statements hold.*

- i. *There exist jointly continuous functions g_i , $i = 1, \dots, m_y$, such that, for all $(x, u) \in X \times U$,*

$$F(x, u) = \{y \in \bar{Y} : g_i(x, y, u) \leq 0, i = 1, \dots, m_y\}.$$
- ii. *The set $\bar{Y}$ is convex and, for all $(x, u) \in X \times U$, the functions $g_i(x, \cdot, u)$, $i = 1, \dots, m_y$, are convex on $\bar{Y}$.*
- iii. *For all $(x, u) \in X \times U$, there exists $y \in \bar{Y}$ satisfying $g_i(x, y, u) < 0$, $i = 1, \dots, m_y$.*
- iv. *For all $(x, u) \in X \times U$, $h(x, \cdot, u)$ is strictly convex on $\bar{Y}$.*

A7.3. *The following statements hold.*

- i. *There exist continuous maps $T : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{m_y \times n_u}$ and $g^X : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{m_y}$, and a map $g^Y : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{m_y}$ such that, for all $(x, u) \in X \times U$,*

$$F(x, u) = \{y \in \mathbb{R}^{n_y} : T(x)u + g^X(x) + g^Y(y) \leq 0\}.$$
- ii. *The components of g^Y are convex and weakly analytic.*
- iii. *There exist a continuous function $h^{XU} : X \times U \rightarrow \mathbb{R}$ and a function $h^Y : \mathbb{R}^{n_y} \rightarrow \mathbb{R}$ such that, for all $(x, y, u) \in X \times \mathbb{R}^{n_y} \times U$, $h(x, y, u) = h^Y(y) + h^{XU}(x, u)$.*
- iv. *The function h^Y is convex and weakly analytic.*

Assumptions A7.2 and A7.3 provide two alternative routes to the regularity of the follower solution map. Under A7.2, the feasible region is described by convex inequalities and satisfies a pointwise Slater condition. The latter is imposed separately for each $(x, u) \in X \times U$. Together with the strict convexity condition in A7.2.iv, this route yields a unique follower response.

Assumption A7.3 instead exploits a right-hand-side parameterization of the follower feasible region. In this case, the convex weakly analytic structure in A7.3.ii allows one to apply Bank et al. (1982, Theorem 3.3.3) without a Slater-type condition. The objective decomposition in A7.3.iii together with A7.3.iv then permits the use of Bank et al. (1982, Theorem 4.3.5(2)) for the follower solution map. In contrast to A7.2, this second route does not require strict convexity and therefore allows multiple follower-optimal solutions.

The next proposition derives the required regularity of Y under either condition.

Proposition 8. *Let A7.1 hold and assume that either A7.2 or A7.3 holds. Then, the following statements are true.*

- i. φ is finite-valued and continuous on $X \times U$.
- ii. For all $(x, u) \in X \times U$, $Y(x, u)$ is non-empty and compact.
- iii. $\text{graph}(Y)$ is closed.
- iv. Y is jointly outer semicontinuous on $X \times U$.
- v. Y is jointly inner semicontinuous relative to $X \times U$.

Proof. We first establish the required regularity of F . Suppose that A7.2 holds. By A7.2.i, $\text{graph}(F)$ is closed. Moreover, A7.2.i–A7.2.iii allow us to apply Bank et al. (1982, Theorem 3.1.6) to the constant set $\bar{Y}$ and the convex inequality system, with (x, u) as parameter. Hence, F is jointly inner semicontinuous relative to $X \times U$.

Suppose now that A7.3 holds. Let

$$\lambda(x, u) := -T(x)u - g^X(x), \text{ and } M(t) := \{y \in \mathbb{R}^{n_y} : g_i^Y(y) \leq t_i, i = 1, \dots, m_y\},$$

and set $\Lambda := \{\lambda(x, u) : (x, u) \in X \times U\}$. Then $F(x, u) = M(\lambda(x, u))$ holds. By A7.3.ii, Bank et al. (1982, Theorem 3.3.3) gives inner semicontinuity of M relative to Λ . Continuity of λ gives joint inner semicontinuity of F relative to $X \times U$. Furthermore, the finite-valued convex functions g_i^Y are continuous, and therefore $\text{graph}(F)$ is closed.

Thus, under either assumption, F has a closed graph and is jointly inner semicontinuous relative to $X \times U$. Since $F(x, u) \subseteq \bar{Y}$ and $\bar{Y}$ is compact, F has compact values and is jointly outer semicontinuous on $X \times U$. By A7.1.i and Lemma 1, φ is finite-valued and continuous on $X \times U$. The follower minimum is attained for every $(x, u) \in X \times U$, and hence $Y(x, u)$ is non-empty and compact. Moreover,

$$Y(x, u) = \{y \in F(x, u) : h(x, y, u) = \varphi(x, u)\}.$$

Continuity of h and φ and closedness of $\text{graph}(F)$ imply that $\text{graph}(Y)$ is closed. Since $Y(x, u) \subseteq \bar{Y}$, the map Y is jointly outer semicontinuous on $X \times U$.

It remains to prove joint inner semicontinuity. If A7.2 holds, then A7.2.ii makes $F(x, u)$ convex, while A7.2.iv makes $h(x, \cdot, u)$ strictly convex. Thus, $Y(x, u)$ is a singleton for every $(x, u) \in X \times U$, and joint outer semicontinuity implies joint inner semicontinuity.

If A7.3 holds, then A7.3.iii gives

$$Y(x, u) = \arg \min_{y \in M(\lambda(x, u))} h^Y(y).$$

By A7.3.iv and Bank et al. (1982, Theorem 4.3.5(2)), the map $t \mapsto \arg \min_{y \in M(t)} h^Y(y)$ is inner semicontinuous relative to Λ . Composition with λ gives joint inner semicontinuity of Y relative to $X \times U$. $\square$

The existence result for Problem (1) now follows immediately.

Corollary 4. *Let A7.1 hold and suppose that either A7.2 or A7.3 holds. If f is jointly continuous, then A3.1 holds and Problem (1) attains a finite optimal value.*

7.3. Convergence Properties of Algorithm 1. In this section, we study the convergence property of Algorithm 1 when using the canonical separation value function Φ^{can} . We start with finite termination.

Theorem 11. *Let A7.1 hold and suppose that either A7.2 or A7.3 holds. Assume that f is jointly continuous. Then, A3.1, A4.1, and A5.1 hold for $\Phi = \Phi^{\text{can}}$. Moreover, for every $\varepsilon > 0$, Algorithm 1 terminates after a finite number of iterations and returns (x_0^t, x^t) such that, for every $u \in U$, there exists $y \in Y(x^t, u)$ satisfying $f(x^t, y, u) \leq x_0^t + \varepsilon$. In addition, x^t is ε -optimal for Problem (1).*

Proof. Proposition 8 gives non-empty compact values and joint inner and outer semicontinuity of Y on $X \times U$, as well as closedness of $\text{graph}(Y)$. Together with continuity of f , these are the hypotheses of the canonical finite-termination theorem in Section 6. Hence, A3.1, A4.1, and A5.1 hold, and that theorem gives finite termination and the stated guarantee. In particular, $\xi(x^t) \leq x_0^t + \varepsilon$. Since the restricted master problem is a relaxation of Problem (1), $x_0^t \leq v^*$, where v^* is its optimal value. Therefore, $v^* \leq \xi(x^t) \leq v^* + \varepsilon$. $\square$

The same regularity gives continuity of the canonical value and, therefore, leads to a convergence result for $\varepsilon = 0$.

Theorem 12. *Let A7.1 hold and suppose that either A7.2 or A7.3 holds. Assume that f is jointly continuous. Then, A3.1, A4.1, A5.1, and A5.2 hold for $\Phi = \Phi^{\text{can}}$. Moreover, Algorithm 1 with $\varepsilon = 0$ and $\Phi = \Phi^{\text{can}}$ satisfies one of the following two alternatives.*

- Algorithm 1 terminates after a finite number of iterations and returns a solution to Problem (1).
- Algorithm 1 does not terminate and, for any accumulation point (x_0^*, x^*, u^*) of its iterates (x_0^t, x^t, u^t) , x^* is a solution to Problem (1).

Proof. Proposition 8 and continuity of f verify the hypotheses of the canonical zero-tolerance theorem in Section 6. Thus, A3.1, A4.1, A5.1, and A5.2 hold. Since $Y(x, u)$ is non-empty for all $(x, u) \in X \times U$, its two feasible alternatives give the result. $\square$

Remark 3. In Lefebvre and Thuerauf (2026), the authors consider linearly constrained robust bilevel problems with a wait-and-see follower and a polyhedral uncertainty set. Their framework does not require the follower problem to be feasible for every leader decision and every uncertainty realization, and they establish finite termination of a variant of Algorithm 1 in this setting.

The analysis developed here instead allows for general compact uncertainty sets, at the price of assuming complete recourse. For polyhedral uncertainty sets, this requirement could in principle be relaxed by augmenting the algorithm with an exact feasibility check for the follower problem, thereby separating feasibility from optimality as in Lefebvre and Thuerauf (2026). However, developing such an extension is beyond the scope of this paper. We also note that Lefebvre and Thuerauf (2026) goes beyond the continuous-follower setting considered here and additionally studies mixed-integer follower problems.

8. CONCLUSION

In this paper, we developed a general convergence framework for column-and-constraint generation algorithms for two-stage robust optimization. Rather than relying on the particular algebraic structure of a given problem class, the analysis is formulated in terms of continuity and semicontinuity properties of the recourse map, the objective function, and the separation value function. This yields sufficient conditions for well-posedness, finite termination with a positive tolerance, and convergence of accumulation points when zero tolerance is used.

We also studied several concrete separation value functions and showed how their stopping conditions translate into different notions of approximate second-stage feasibility. Finally, we applied the general theory to robust bilevel optimization with a wait-and-see convex follower and obtained convergence results from regularity properties of the follower solution map.

Future works include an extension of the analysis to nonconvex recourse problems, decision-dependent uncertainty sets, and further investigations on error bounds linking constraint violation to distances from the recourse set.

REFERENCES

- Aliprantis, C. D. and K. C. Border (2006). *Infinite dimensional analysis: a hitchhiker's guide*. Springer. DOI: [10.1007/3-540-29587-9](https://doi.org/10.1007/3-540-29587-9).
- Ayoub, J. and M. Poss (2016). “Decomposition for adjustable robust linear optimization subject to uncertainty polytope.” In: *Computational Management Science* 13.2, pp. 219–239. DOI: [10.1007/s10287-016-0249-2](https://doi.org/10.1007/s10287-016-0249-2).
- Bank, B., J. Guddat, D. Klatte, B. Kummer, and K. Tammer (1982). *Non-Linear Parametric Optimization*. Birkhäuser Basel. DOI: [10.1007/978-3-0348-6328-5](https://doi.org/10.1007/978-3-0348-6328-5).
- Beck, Y., I. Ljubić, and M. Schmidt (2023). “A survey on bilevel optimization under uncertainty.” In: *European Journal of Operational Research* 311.2, pp. 401–426. DOI: [10.1016/j.ejor.2023.01.008](https://doi.org/10.1016/j.ejor.2023.01.008).

- Beck, Y., I. Ljubić, and M. Schmidt (2026). *Linear and Mixed-Integer Bilevel Optimization: Theory and Algorithms*. Pre-publication version available online. URL: <https://yasminebeck.github.io/files/bilevel-optimization-cup.pdf>.
- Hu, H., J. Li, J. Tang, and B. Zeng (2026). *Two-stage robust bilevel optimization model for facility location considering operational service level under disruption risk*. arXiv: 2603.21225 [math.OC]. URL: <https://arxiv.org/abs/2603.21225>.
- Lefebvre, H., E. Malaguti, and M. Monaci (2025). “Exact approaches for convex adjustable robust optimization.” In: *Mathematical Programming*. DOI: [10.1007/s10107-025-02311-3](https://doi.org/10.1007/s10107-025-02311-3).
- Lefebvre, H., M. Schmidt, and J. Thürauf (2026). “Column generation in column-and-constraint generation for adjustable robust optimization with Interdiction-type linking Constraints.” In: *Mathematical Programming Computation* 18:2, pp. 259–296. DOI: [10.1007/s12532-025-00300-3](https://doi.org/10.1007/s12532-025-00300-3).
- Lefebvre, H. and A. Subramanyam (2025). “Remarks on: A Lagrangian dual method for two-stage robust optimization with binary uncertainties.” In: *Optimization and Engineering* 27:1, pp. 531–553. DOI: [10.1007/s11081-025-10015-y](https://doi.org/10.1007/s11081-025-10015-y).
- Lefebvre, H. and J. Thürauf (2026). *Robust Bilevel Optimization with a Wait-and-See Follower: A Column-and-Constraint Generation Approach*. URL: <https://optimization-online.org/?p=34583>.
- Mutapcic, A. and S. Boyd (2009). “Cutting-set methods for robust convex optimization with pessimizing oracles.” In: *Optimization Methods and Software* 24:3, pp. 381–406. DOI: [10.1080/10556780802712889](https://doi.org/10.1080/10556780802712889).
- Rockafellar, R. T. and R. J. B. Wets (1998). *Variational Analysis*. Springer Berlin Heidelberg. DOI: [10.1007/978-3-642-02431-3](https://doi.org/10.1007/978-3-642-02431-3).
- Subramanyam, A. (2022). “A Lagrangian dual method for two-stage robust optimization with binary uncertainties.” In: *Optimization and Engineering* 23:4, pp. 1831–1871. DOI: [10.1007/s11081-022-09710-x](https://doi.org/10.1007/s11081-022-09710-x).
- Vershynin, R. (2026). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.
- Xu, L. (2023). “A Study on Multilevel Optimization Subject to Uncertainty.” Doctoral dissertation. University of Pittsburgh. URL: <https://d-scholarship.pitt.edu/concern/etds/52182e1e-5550-4916-8066-d856f315251f>.
- Zeng, B. and W. Wang (2022). *Two-Stage Robust Optimization with Decision Dependent Uncertainty*. arXiv: 2203.16484 [math.OC]. URL: <https://arxiv.org/abs/2203.16484>.
- Zeng, B. and L. Zhao (2013). “Solving two-stage robust optimization problems using a column-and-constraint generation method.” In: *Operations Research Letters* 41:5, pp. 457–461. DOI: [10.1016/j.orl.2013.05.003](https://doi.org/10.1016/j.orl.2013.05.003).
- Zhao, L. and B. Zeng (2012). *An Exact Algorithm for Two-stage Robust Optimization with Mixed Integer Recourse Problems*. URL: <https://optimization-online.org/2012/01/3310/>.

APPENDIX A. RELATIONS BETWEEN THE DIFFERENT TOLERANCE PROPERTIES

Throughout this appendix, $(\tau, \gamma) \in \mathbb{R}_{\geq 0}^2$ denotes a pair of actual target tolerances, and $\bar{\varepsilon} \geq 0$ denotes a corresponding tolerance property. The norm on the product space is compatible with the norm on $\mathbb{R}^{n_y}$ in the sense that $\|(0, 0, y)\| = \|y\|$ and $\|y\| \leq \|(x, u, y)\|$. Fix also a monotone norm on $\mathbb{R}^{m_y}$ and let $c_e := \|e\|$. We first record the direct self-improvement of pointwise tolerance and then follow the implications displayed in Figure 3.

Proposition 9. *Suppose that there exists $L_f \geq 0$ such that*

$$f(x, y', u) - f(x, y, u) \leq L_f \|y' - y\|$$

holds for all $(x, u) \in X \times U$ and all $y, y' \in \mathbb{R}^{n_y}$. If Φ satisfies the (τ, γ) -pointwise tolerance property, then it also satisfies the $(\tau + L_f \gamma, 0)$ -pointwise tolerance property.

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. By pointwise tolerance, there exist $y \in \mathbb{R}^{n_y}$ and $\hat{y} \in Y(x, u)$ such that

$$f(x, y, u) - x_0 \leq \tau, \quad \|y - \hat{y}\| \leq \gamma.$$

Hence,

$$f(x, \hat{y}, u) - x_0 \leq f(x, y, u) - x_0 + L_f \|\hat{y} - y\| \leq \tau + L_f \gamma.$$

Moreover, $\hat{y} \in Y(x, u) = Y(x, u) + B_0$. Thus, the same threshold $\bar{\varepsilon}$ proves the result. $\square$

We now compare pointwise and graph tolerance.

Proposition 10. *If Φ satisfies the (τ, γ) -pointwise tolerance property, then it also satisfies the (τ, γ) -graph tolerance property.*

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. By pointwise tolerance, there exist y and $\hat{y} \in Y(x, u)$ such that

$$f(x, y, u) - x_0 \leq \tau, \quad \|y - \hat{y}\| \leq \gamma.$$

Since $(x, u, \hat{y}) \in \text{graph}_{X,U}(Y)$, the compatibility of the norms gives the (τ, γ) -graph tolerance property with the same threshold. $\square$

Proposition 11. *Let A6.1 hold and assume that Y is closed-valued on $X \times U$. Suppose that there exists $L_Y \geq 0$ such that $\text{dist}(\hat{y}, Y(x, u)) \leq L_Y \|(x, u) - (x', u')\|$ holds for all $(x, u), (x', u') \in X \times U$ and all $\hat{y} \in Y(x', u')$. If Φ satisfies the (τ, γ) -graph tolerance property, then it satisfies the $(\tau, \gamma(1 + L_Y))$ -pointwise tolerance property.*

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. Graph tolerance gives y and $(x', u', y') \in \text{graph}_{X,U}(Y)$ such that

$$f(x, y, u) - x_0 \leq \tau, \quad \|(x, u, y) - (x', u', y')\| \leq \gamma.$$

The global Aubin-type estimate gives

$$\text{dist}(y', Y(x, u)) \leq L_Y \gamma.$$

Since $Y(x, u)$ is non-empty and closed, this distance is attained at some $\hat{y} \in Y(x, u)$. The triangle inequality and compatibility of the norms yield

$$\|y - \hat{y}\| \leq \gamma(1 + L_Y),$$

proving the result. $\square$

We next relate pointwise tolerance to the functional violation measure.

Proposition 12. *Let A6.1 and A6.3 hold, and assume that Y is closed-valued on $X \times U$. Suppose that there exists $H_Y^y > 0$ such that*

$$\text{dist}(y, Y(x, u)) \leq H_Y^y \| [g(x, y, u)]^+ \|$$

holds for all $(x, y, u) \in X \times \mathbb{R}^{n_y} \times U$. If Φ satisfies the (τ, γ) -violation tolerance property, then it satisfies the $(\tau, H_Y^y \gamma c_e)$ -pointwise tolerance property.

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. Violation tolerance gives y satisfying $f(x, y, u) - x_0 \leq \tau$ and $g(x, y, u) \leq \gamma e$, and therefore

$$\| [g(x, y, u)]^+ \| \leq \gamma \| e \| = \gamma c_e.$$

The error bound gives $\text{dist}(y, Y(x, u)) \leq H_Y^y \gamma c_e$. Since $Y(x, u)$ is non-empty and closed, this distance is attained, which proves the $(\tau, H_Y^y \gamma c_e)$ -pointwise tolerance property with the same threshold. $\square$

In affine or polyhedral settings, the error bound in the preceding proposition may follow from Hoffman's lemma.

Proposition 13. *Let A6.3 hold and suppose that there exists $L_g^y \geq 0$ such that $g(x, y, u) - g(x, y', u) \leq L_g^y \| y - y' \| e$ holds for all $(x, u) \in X \times U$ and all $y, y' \in \mathbb{R}^{n_y}$. If Φ satisfies the (τ, γ) -pointwise tolerance property, then it satisfies the $(\tau, L_g^y \gamma)$ -violation tolerance property.*

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. Pointwise tolerance gives y and $\hat{y} \in Y(x, u)$ such that

$$f(x, y, u) - x_0 \leq \tau, \quad \| y - \hat{y} \| \leq \gamma.$$

Since $g(x, \hat{y}, u) \leq 0$, the stated estimate gives $g(x, y, u) \leq L_g^y \gamma e$, proving the result. $\square$

It remains to compare violation and graph tolerance.

Proposition 14. *Let A6.3 hold and assume that $\text{graph}_{X,U}(Y)$ is non-empty and closed. Suppose that there exists $H_Y > 0$ such that $\text{dist}((x, u, y), \text{graph}_{X,U}(Y)) \leq H_Y \| [g(x, y, u)]^+ \|$ holds for all $(x, y, u) \in X \times \mathbb{R}^{n_y} \times U$. If Φ satisfies the (τ, γ) -violation tolerance property, then it satisfies the $(\tau, H_Y \gamma c_e)$ -graph tolerance property.*

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. Violation tolerance gives y satisfying $f(x, y, u) - x_0 \leq \tau$ and

$$\| [g(x, y, u)]^+ \| \leq \gamma c_e.$$

The joint error bound gives

$$\text{dist}((x, u, y), \text{graph}_{X,U}(Y)) \leq H_Y \gamma c_e.$$

Since $\text{graph}_{X,U}(Y)$ is non-empty and closed, the distance is attained. Hence, $(x, u, y) \in \text{graph}_{X,U}(Y) + B_{H_Y \gamma c_e}$, which proves the $(\tau, H_Y \gamma c_e)$ -graph tolerance property. $\square$

In affine or polyhedral settings, the joint error bound above may follow from Hoffman's lemma.

Proposition 15. *Let A6.3 hold and suppose that there exists $L_g \geq 0$ such that*

$$g(x, y, u) - g(x', y', u') \leq L_g \| (x, u, y) - (x', u', y') \| e$$

holds for all $(x, u, y), (x', u', y') \in X \times U \times \mathbb{R}^{n_y}$. If Φ satisfies the (τ, γ) -graph tolerance property, then it satisfies the $(\tau, L_g \gamma)$ -violation tolerance property.

Proof. Let $\bar{\varepsilon} > 0$ be a threshold for the assumed property, and let $0 \leq \varepsilon \leq \bar{\varepsilon}$. Suppose that $\Phi(x_0, x, u) \leq \varepsilon$. Graph tolerance gives y and $(x', u', y') \in \text{graph}_{X,U}(Y)$ such that

$$f(x, y, u) - x_0 \leq \tau, \quad \|(x, u, y) - (x', u', y')\| \leq \gamma.$$

Since $g(x', y', u') \leq 0$, the stated estimate gives $g(x, y, u) \leq L_g \gamma e$, proving the result. $\square$

(Henri Lefebvre) LIRMM, UNIVERSITY OF MONTPELLIER, CNRS, MONTPELLIER, FRANCE
Email address: `henri.lefebvre@cnrs.fr`