

# Characterization of an Anomalous Behavior of a Practical Smoothing Technique

Pedro Borges\*

## Abstract

A practical smoothing method was analyzed and tested against state-of-the-art solvers for some non-smooth optimization problems in [BSS20a; BSS20b]. This method can be used to smooth the value functions and solution mappings of fully parameterized convex problems under mild conditions. In general, the smoothing of the value function lies from above the true value function and the smoothing of the solution mapping is an approximate selection over the true solution mapping. In this paper we say exactly when the smoothed value function is equal to the true value function and when the smoothed solution is an actual solution. Moreover, we describe a further application of this smoothing technique for building smooth approximations of optimal value reformulations of optimistic bilevel problems that converge epigraphically to the original bilevel problem. This new application motivates the characterization we prove.

**keywords** Smoothing Methods, Interior Penalty Methods, Bilevel Optimization, Non-smooth Optimization

## 1 Introduction

Non-smooth and non-convex optimization problems can arise in bilevel optimization [FJQ99], equilibrium programming [FP03], stochastic optimization [BSS20a] and as subproblems of methods for smooth optimization [BH13]. The non-smooth function can be explicitly defined [FJQ99] or can be implicitly defined [BSS20a; BSS20b]. To date, important classes of practical methods to deal with these problems are the cutting plane methods [Kel60], the bundle methods [OSL14], the subgradient methods [Kiw64], the smoothing methods [Nes05; BT12; BSS20a; BSS20b], the non-smooth Newton-type methods [IS14] and the derivative-free methods [RC-SNVnu].

A question central to most methods for non-smooth optimization is how to compute a subgradient [Kiw64]. This generalized derivative can be computed efficiently for some important special cases [Kel60; OSL14]. Practical smoothing methods can provide a general answer to this problem via the gradient consistency property [Che12]. Therefore, the importance of smoothing methods is justified. Nonetheless, such theory can be quite hard to develop [BHK13; BH13; BSS20a; BSS20b]. In particular, it involves understanding well at least the diverging eigenvalues of the Hessian of the smoothing [BSS20b, Proposition 2.2]. In the other hand, epigraphical convergence [RW09, Chapter 7] is a closely related concept of convergence and is easy to prove [BSS20b]. In short, gradient consistency is useful to the limiting analysis of stationary points of the smoothed problems and epigraphical convergence to the limiting analysis of local solutions of the smoothed problems as the smoothing parameter goes to zero.

Some algorithms already developed can be applied for the optimistic bilevel optimization problem [DDM07]. When problems are linear we can use the 0-1 reformulation based on mixed integer linear programming [Dem02; Bar98]. Methods focused on explicitly smoothing complementarity functions [FJQ99] and methods based on integral smoothing methods [XJY14; XWYJ14] have been considered. On the other hand, the optimal value reformulation of the bilevel problem remained under explored [DF14; DF16] because it is not clear how to compute subgradients for general value functions. Another efficient smoothing method for the optimistic bilevel problem with a convex subproblem is shown in [OA18]. Our method is close in spirit to [OA18].

---

\*Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil (pborges@impa.br)

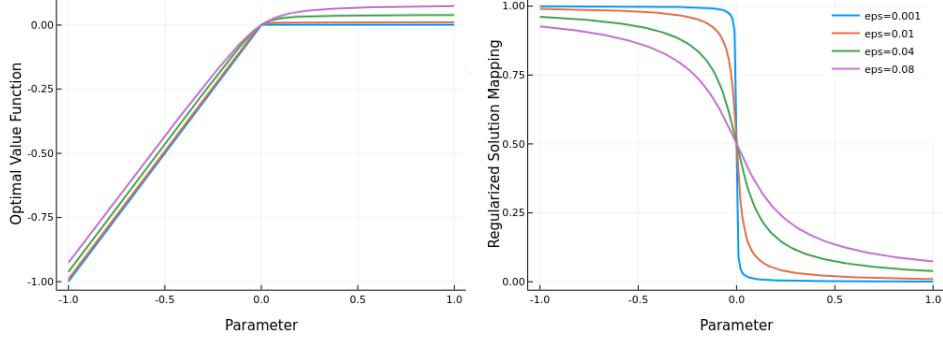


Figure 1: Let  $y$  the a parameter and consider the parameterized problem  $v(y) = \min_{x \in [0,1]} yx$ . This figure shows a situation where the anomalous behavior of the smoothing happens at  $y = 0$ . Note that  $v(y) = y$  if  $y \leq 0$  and  $v(y) = 0$  if  $y > 0$ . At  $y = 0$ , the smoothing on the left side agrees with  $v(y)$  for all  $\epsilon > 0$ . Moreover, the smoothed solution mapping denoted by  $x^\epsilon(y)$  at  $y = 0$  is  $x^\epsilon(y) = 0.5$  for all  $\epsilon > 0$ . At  $y = 0$ , the solution set associated with  $v(y)$  is  $S(y) = [0, 1]$ . The analytic center of  $S(0)$  is 0.5. For these figures, the smoothing is generated with  $\mu = 0$ .

As shown in [BSS20a; BSS20b], the smoothing presented here can be numerically effective if properly used. This includes selecting properly the sequence of regularizing parameters, taking care of numerical failures because we are potentially solving thousands of optimization problems, using previous iterates to restart new optimization problems and providing feasible iterates for the subproblems among others. All these conditions serve to increase the chances that in practice we really compute precisely enough the regularized solution mapping  $x^\epsilon(y)$  illustrated in Figure 1.

Our technique relies heavily on interior point solvers for solving the subproblems, which happens to be convenient since there are efficient implementations [WLB06]. As we shall explain, the cost of computing  $x^\epsilon(y)$  is the same of following the central path until  $\epsilon > 0$ . See the *mu-target* option of Ipopt [WLB06]. When  $x^\epsilon(y)$  is the regularized solution mapping of a fully parameterized linear problem (LP), one would be able to compute  $x^\epsilon(y)$  with the same efficiency of solving a LP with an interior point method. However, we are not aware that this option is available on the major LP solvers.

The paper is organized as follows. In Section 2 we present some preliminaries. In Section 3 we characterize the anomalous behavior of the smoothing. In Section 4 we present the application to optimistic bilevel optimization.

## 2 Preliminaries

Let us consider the parameterized problem

$$S(y) := \arg \min_x f(x, y) \quad \text{s.t.} \quad A(y)x = b(y), \quad g_i(x, y) \leq 0 \quad \forall i = 1, \dots, m \quad (1)$$

that satisfies the set of assumptions below and whose optimal value is denoted by  $v(y)$ .

**Assumption 2.1.** Problem (1) is such that (i) the Slater condition holds for all  $y$ , (ii) convexity holds for all  $y$ , (iii) its problem data is sufficiently smooth both in  $x$  and  $y$  and (iv) the lines of the matrix  $A(y)$  are linearly independent for all  $y$ .

**Remark 2.2.** Note that Assumption 2.1 does not imply that the optimal value of (1) is finite.

For a fixed  $\mu \geq 0$ , the parameterized interior penalty associated with (1) is given by

$$\phi(x, y) = - \sum_{i=1}^m \lg\{-g_i(x, y)\} + \frac{\mu}{2} \|x\|_2^2. \quad (2)$$

The restricted inf-compactness condition requires that

$$\limsup_{y' \in \text{dom } S, y' \rightarrow y} \left\{ \min_{x \in S(y')} \|x\|_2^2 \right\} < +\infty. \quad (3)$$

**Theorem 2.3.** Assume condition (3) and Assumption 2.1. Consider for all  $\epsilon > 0$  the functions

$$x^\epsilon(y) := \arg \min_x f(x, y) + \epsilon \phi(x, y) \quad \text{s.t.} \quad A(y)x = b(y) \quad (4)$$

and

$$v^\epsilon(y) := f(x^\epsilon(y), y). \quad (5)$$

If either  $\mu > 0$  or  $S(y)$  is non-empty and bounded for all  $y$  and the constraints  $x \geq 0$  are part of (1), the following items are true.

1.  $x^\epsilon(y)$  is well defined and smooth as a function of  $y$  and its derivatives satisfy a linear system.
2. For all  $y$  we have  $v^\epsilon(y) \searrow v(y)$  as  $\epsilon \searrow 0$  and for all  $y \in \text{dom } S$  it holds that

$$v(y) \leq v^\epsilon(y) \leq v(y) + m\epsilon + \epsilon \frac{\mu}{2} \min_{x \in S(y)} \|x\|_2^2. \quad (6)$$

3. Denote by  $\text{lsc } v(y)$  the lower semi-continuous closure of  $v(y)$ . It holds that

$$\liminf_{\epsilon \searrow 0, y' \rightarrow y} v^\epsilon(y') = \text{lsc } v(y) \quad \forall y. \quad (7)$$

4. It holds that

$$\lim_{\epsilon \searrow 0, y' \rightarrow y} v^\epsilon(y') = v(y) \quad \forall y \in \text{int dom } S. \quad (8)$$

*Proof.* See [BSS20b]. □

A sequence of functions  $\rho_k : \mathbb{R}^q \rightarrow \overline{\mathbb{R}}$  is said to converge epigraphically, see [RW09, Proposition 7.2], to  $\rho : \mathbb{R}^q \rightarrow \overline{\mathbb{R}}$  if the following two conditions hold for all  $y \in \mathbb{R}^q$ :

$$\liminf_k \rho_k(y_k) \geq \rho(y) \quad \text{for all } y_k \rightarrow y,$$

and

$$\limsup_k \rho_k(y_k) \leq \rho(y) \quad \text{for some } y_k \rightarrow y.$$

The notion of epigraphical convergence of functions is tightly related to the convergence of minimizers thanks to the following theorem, adapted from [RW09, Theorem 7.31] and [RW09, Theorem 7.4 (d)].

**Theorem 2.4.** For extended-valued functions  $\rho_k$  converging epigraphically to  $\rho$ , the following holds.

- (i)  $\inf \rho_k \rightarrow \inf \rho$  if and only if for all  $\epsilon > 0$  there is a compact set  $B \subset \mathbb{R}^q$  such that  $\inf_B \rho_k \leq \inf \rho + \epsilon$  for all  $k$  large enough,
- (ii)  $\limsup_k \{\epsilon_k - \arg \min \rho_k\} \subset \arg \min \rho$  for all  $\epsilon_k \rightarrow 0$ .

Moreover,

- (iii) If  $\rho_{k+1} \geq \rho_k$  for all  $k$ , then  $\rho_k$  converges epigraphically to the lsc. function  $\sup_k \{\text{lsc } \rho_k\}$ .

### 3 Anomalous Behavior

In this section we consider the non-parameterized convex smooth problem

$$\min_x f(x) \quad \text{s.t.} \quad Ax = b, \quad g_i(x) \leq 0 \quad \forall i = 1, \dots, m \quad (9)$$

satisfying the Slater condition and such that  $A$  has linearly independent lines. We also assume that the optimal value  $v$  of problem (9) is finite and denote by  $S \neq \emptyset$  the respective solution set. The interior of the feasible set of problem (9) is denoted by

$$D^0 = \{x \in \mathbb{R}^n : Ax = b, g_i(x) < 0 \quad \forall i = 1, \dots, m\}. \quad (10)$$

For a fixed  $\mu \geq 0$ , the interior penalty of problem (9) is given by

$$\phi(x) = -\sum_{i=1}^m \lg\{-g_i(x)\} + \frac{\mu}{2} \|x\|_2^2. \quad (11)$$

Besides, we also assume that either  $\mu > 0$  or the constraints  $x \geq 0$  are part of problem (9). In particular, this implies that the Hessian of  $\phi$  is positive definite for all  $x \in D^0$ . The analytic center of the feasible set  $D^0$  is the solution, whenever it exists, of the problem

$$\min_x \phi(x) \quad \text{s.t.} \quad x \in D^0. \quad (12)$$

Note that we are not talking about the usual analytic center of the solution set  $S$  to which interior penalty trajectories might converge [MZ98; DS99]. For this paper, the analytic center of the solution set is, whenever it exists, a solution of the problem

$$\min_x \phi(x) \quad \text{s.t.} \quad x \in S \cap D^0. \quad (13)$$

The interior penalty solution, whenever it exists, is given by

$$x(\epsilon) := \arg \min_x f(x) + \epsilon \phi(x) \quad \text{s.t.} \quad x \in D^0. \quad (14)$$

From [IS06; BSS20a], we know that the function  $x(\epsilon)$  satisfies

$$f(x(\epsilon)) \searrow v \quad \text{as} \quad \epsilon \searrow 0 \quad (15)$$

and

$$v \leq f(x(\epsilon)) \leq v + m\epsilon + \epsilon \frac{\mu}{2} \min_{x \in S} \|x\|_2^2. \quad (16)$$

The Hessian of  $\phi$  being positive definite does not guarantee that  $x(\epsilon)$  exists. For such, one has to assume that  $D^0$  is bounded or most generally that  $S$  is bounded [MZ98; DS99]. In general, it holds that  $f(x(\epsilon)) \geq v$  for all  $\epsilon > 0$ . The anomalous behavior we refer to is when  $f(x(\epsilon)) = v$  happens. In this section, we relate the possibility that  $f(x(\epsilon)) = v$  with the possibility that the analytic center of the feasible set intersects the solution set of the convex programming problem in consideration. We start with an example.

**Example 3.1.** Take  $f : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $f(x) = \max\{0, x\}^4 \in C^2$  and consider the problem

$$\min_x f(x) \quad \text{s.t.} \quad -1 \leq x \leq \eta$$

whose solution set is denoted by  $S(\eta)$ . The associated interior penalty for  $\mu = 0$  is

$$\phi(x) = -\ln\{x+1\} - \ln\{\eta-x\}.$$

Define the analytic center  $C(\eta)$  of the feasible set  $D^0$  as

$$\arg \min_x \phi(x) \quad \text{s.t.} \quad -1 < x < \eta.$$

Fix  $\eta > -1$ . It can be seen that  $C(\eta) = \{0.5(\eta-1)\}$  and that

$$S(\eta) = [-1, \min\{\eta, 0\}].$$

We can observe that  $f(x(\epsilon)) = v$  if and only if

$$S(\eta) \cap C(\eta) \neq \emptyset.$$

For instance, take  $\eta = 1$ . Then,

$$C(1) \cap S(1) = \{0\} \cap [-1, 0] \neq \emptyset$$

and  $x(\epsilon) = 0$  for all  $\epsilon > 0$ . In this case,  $0 \in \mathbb{R}$  is also the analytic center of  $S(1)$  because it solves

$$\min_x \phi(x) \quad \text{s.t.} \quad x \in S(1) \cap D^0.$$

For  $\eta > 1$  we have  $f(x(\epsilon)) > v = 0$ . The results that follow generalize these observations.

**Lemma 3.2.** Assume that  $x(\epsilon)$  exists and that there is  $\bar{\epsilon} > 0$  such that  $f(x(\bar{\epsilon})) = v$ . It follows that

$$\phi(x(\epsilon)) = \inf_{x \in S \cap D^0} \phi(x) \quad \forall \epsilon \in (0, \bar{\epsilon}]. \quad (17)$$

In other words, the point  $x(\epsilon)$  is the analytic center of the solution set  $S$  for all  $\epsilon \in (0, \bar{\epsilon}]$ .

*Proof.* Recall that

$$f(x(\bar{\epsilon})) \geq f(x(\epsilon)) \geq v \quad \forall \epsilon \in (0, \bar{\epsilon}]. \quad (18)$$

It follows from  $f(x(\bar{\epsilon})) = v$  and the last observation that

$$f(x(\epsilon)) = v \quad \forall \epsilon \in (0, \bar{\epsilon}]. \quad (19)$$

Take  $\epsilon \in (0, \bar{\epsilon}]$ . It holds that  $x(\epsilon) \in D^0 \cap S \neq \emptyset$ . Provided that  $S \cap D^0 \subset D^0$  we have

$$\begin{aligned} f(x(\epsilon)) + \epsilon \phi(x(\epsilon)) &= \inf \{f(x) + \epsilon \phi(x) : x \in D^0\} \\ &\leq \inf \{f(x) + \epsilon \phi(x) : x \in S \cap D^0\} \\ &= v + \epsilon \inf \{\phi(x) : x \in S \cap D^0\}. \end{aligned} \quad (20)$$

It follows that

$$\phi(x(\epsilon)) \leq \inf_{x \in S \cap D^0} \phi(x) \quad (21)$$

and therefore

$$\phi(x(\epsilon)) = \inf_{x \in S \cap D^0} \phi(x) \quad \forall \epsilon \in (0, \bar{\epsilon}]. \quad (22)$$

The last equality proves that  $x(\epsilon)$  is the analytic center of  $S$ .  $\square$

**Lemma 3.3.** Assume that  $x(\epsilon)$  exists. It holds that  $x(\epsilon) : (0, \infty) \rightarrow \mathbb{R}^n$  and  $\lambda(\epsilon) : (0, \infty) \rightarrow \mathbb{R}^l$  are  $C^1$  functions. Moreover, the following items are equivalent for a fixed  $\epsilon > 0$ :

1.  $x'(\epsilon) = 0$ .
2.  $[\nabla \phi](x(\epsilon)) - A^\top \lambda'(\epsilon) = 0$ .
3.  $x(\epsilon)$  solves  $\min_{x \in D^0} \phi(x)$ .

*Proof.* To prove that  $x(\epsilon), \lambda(\epsilon) \in C^1$  we just differentiate the KKT conditions for (14) with respect to  $\epsilon$  and apply the Implicit Function Theorem [Lan93]. The resulting equation is

$$\begin{bmatrix} \nabla^2 f(x(\epsilon)) + \epsilon \nabla^2 \phi(x(\epsilon)) & -A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x'(\epsilon) \\ \lambda'(\epsilon) \end{bmatrix} = \begin{bmatrix} -\nabla \phi(x(\epsilon)) \\ 0 \end{bmatrix}. \quad (23)$$

Define  $M = \nabla^2 f(x(\epsilon)) + \epsilon \nabla^2 \phi(x(\epsilon))$ . If  $\mu = 0$  the presence of  $x \geq 0$  makes  $X^{-2}(\epsilon) = \text{diag}^{-2}(x(\epsilon)) > 0$  appear, making  $M$  positive definite. If  $\mu > 0$  the positive definiteness of  $M$  is trivial. Using that  $M$  is non-singular, it follows that  $x'(\epsilon) = 0$  if and only if  $Mx'(\epsilon) = 0$  if and only if  $[\nabla \phi](x(\epsilon)) - A^\top \lambda'(\epsilon) = 0$ . Moreover, it follows that if  $[\nabla \phi](x(\epsilon)) - A^\top \lambda'(\epsilon) = 0$  then  $x(\epsilon)$  solves  $\min_{x \in D^0} \phi(x)$  because we can take the vector  $\lambda'(\epsilon)$  as a Lagrange multiplier. Assume now that  $x(\epsilon)$  solves  $\min_{x \in D^0} \phi(x)$ . We can find  $\eta \in \mathbb{R}^l$  such that  $[\nabla \phi](x(\epsilon)) - A^\top \eta = 0$ . Using (23) we have  $Ax'(\epsilon) = 0$  and  $Mx'(\epsilon) = A^\top(\lambda'(\epsilon) - \eta)$ . It follows that  $x'(\epsilon)^\top Mx'(\epsilon) = 0$  because  $Ax'(\epsilon) = 0$ , Therefore, we have  $x'(\epsilon) = 0$  because  $M$  is positive definite. The full row rank of  $A$  implies that  $\eta = \lambda'(\epsilon)$ . Therefore, item 3 implies item 2.  $\square$

**Proposition 3.4.** *Assume that  $x(\epsilon)$  exists. The following items are equivalent:*

1. *There is  $\bar{\epsilon} > 0$  such that  $f(x(\bar{\epsilon})) = v$ .*
2. *There is  $\bar{\epsilon} > 0$  such that  $x'(\epsilon) = 0$  for all  $\epsilon \in (0, \bar{\epsilon}]$ .*
3. *There is  $\bar{\epsilon} > 0$  such that*

$$x(\epsilon) = \arg \min_{x \in S \cap D^0} \phi(x) = \arg \min_{x \in D^0} \phi(x) \quad \forall \epsilon \in (0, \bar{\epsilon}]. \quad (24)$$

*Proof.* Assume item 1. It follows from Lemma 3.2 that  $x(\epsilon)$  is a constant function on some interval  $(0, \bar{\epsilon}]$ . Therefore, it has to hold that  $x'(\epsilon) = 0$  for all  $(0, \bar{\epsilon}]$ . Then, item 1 implies item 2. Recall now that  $x'(\epsilon) = 0$  is equivalent to  $x(\epsilon)$  solving  $\min_{x \in D^0} \phi(x)$ . Therefore, lemmas 3.2 and 3.3 show that item 1 also implies item 3. Assume item 2. It follows that there is  $\bar{\epsilon} > 0$  such that  $x(\epsilon)$  is constant for all  $\epsilon \in (0, \bar{\epsilon})$ . Then,  $f(x(\epsilon))$  is constant for  $\epsilon > 0$  small enough. As  $f(x(\epsilon)) \searrow v$  when  $\epsilon \searrow 0$  we have that  $f(x(\epsilon)) = v$  for all  $\epsilon \in (0, \bar{\epsilon})$ . Then, item 2 implies item 1. Also by lemmas 3.2 and 3.3 and the equivalence between items 1 and 2 we have that item 2 implies item 3. Item 3 implies 2 because  $x(\epsilon)$  is constant for  $\epsilon \in (0, \bar{\epsilon})$ .  $\square$

Note that the analytic center of the set  $D^0$  with respect to the penalty  $\phi$  may fail to exist. This is not a problem. In this case the inequality  $f(x(\epsilon)) > v$  for all  $\epsilon > 0$  has to hold. Moreover, if  $S \cap D^0 = \emptyset$  such inequality also holds. The last case corresponds to requiring that the solution set  $S$  is contained in the boundary of the feasible set in consideration. The next result characterizes the anomalous behavior of the trajectory  $x(\epsilon)$ .

**Theorem 3.5.** *Assume that  $x(\epsilon)$  exists. Define*

$$U = \arg \min\{\phi(x) : x \in S \cap D^0\} \quad \text{and} \quad V = \arg \min\{\phi(x) : x \in D^0\}. \quad (25)$$

*It holds that  $V \cap U = \emptyset$  if and only if  $f(x(\epsilon)) > v$  for all  $\epsilon > 0$ .*

*Proof.* Recall that  $U$  and  $V$  are both at most singletons because  $\phi$  is strictly convex. If  $U \cap V = \emptyset$  it is trivial that  $f(x(\epsilon)) = v$  for some  $\epsilon > 0$  cannot happen because of Proposition 3.4. It follows that  $f(x(\epsilon)) > v$  for all  $\epsilon > 0$ . Assume now that  $f(x(\epsilon)) > v$ . It follows from Proposition 3.4 that  $U \cap V \neq \{x(\epsilon)\}$ . Note that  $U \cap V$  is at most a singleton. If  $U \cap V = \emptyset$  we are done. If  $U \cap V \neq \emptyset$  we take  $\bar{x} \in U \cap V$ . It is trivial to see that  $\bar{x}$  solves  $\min_{x \in D^0} f(x) + \epsilon \phi(x)$ . Therefore,  $\bar{x} = x(\epsilon)$  for all  $\epsilon > 0$ . This is a contradiction. It follows that  $U \cap V = \emptyset$ .  $\square$

**Corollary 3.6.** *Assume that  $x(\epsilon)$  exists. The following items are equivalent:*

1. *There is  $\bar{\epsilon} > 0$  such that  $f(x(\bar{\epsilon})) = v$ .*
2. *For all  $\epsilon > 0$  we have  $f(x(\epsilon)) = v$ .*

*Proof.* Let  $U = \arg \min\{\phi(x) : x \in S \cap D^0\}$  and  $V = \arg \min\{\phi(x) : x \in D^0\}$ . Note that  $V \cap U$  do not depend on  $\epsilon > 0$ . If there is  $\bar{\epsilon} > 0$  such that  $f(x(\bar{\epsilon})) = v$  we know that  $V \cap U = \{x(\epsilon)\}$  for all  $\epsilon > 0$  and  $f(x(\epsilon)) = v$  because  $x(\epsilon) \in S$ . The converse is immediate.  $\square$

## 4 Bilevel Optimization

The optimistic bilevel optimization problem [DDM07] is given by

$$\min_{y,x} h(y,x) \quad \text{s.t.} \quad (y,x) \in Z, \quad x \in S(y). \quad (26)$$

Its interpretation is that there is a leader player that decides  $y$  and the follower responds with a decision  $x$ . The decisions of the leader and the follower together have to satisfy the clearing conditions  $(y,x) \in Z$ . The specific form of  $Z$  is not important for our presentation. However, we have in mind that  $Z$  is described by a system of smooth possibly non-convex constraints satisfying a suitable constraint qualification. It is optimistic because the leader assumes that the follower collaborates. In other words, that the follower selects a decision that is best for the leader. On the other hand, there is the pessimistic version of this bilevel problem [WTKR13].

The reason that bilevel optimization is considered separately from smooth nonconvex optimization is the constraint  $x \in S(y)$ . It can be reformulated depending on the assumptions on problem (1). The most common is that for all  $y$  problem (1) is convex, smooth and satisfies the Mangasarian-Fromovitz constraint qualification. Such assumption allows us to change  $x \in S(y)$  by the KKT conditions of problem (1). The most important detail is that the resulting formulation does not satisfy usual constraint qualifications due to the complementarity relations [FJQ99]. This motivated many techniques for optimization problems with equilibrium constraints [FJQ99]. Alternatively, the non-smooth optimization approach for solving (26) is to consider the optimal value reformulation given by problem

$$\min_{y,x} h(y,x) \quad \text{s.t.} \quad (y,x) \in Z, \quad A(y)x = b(y), \quad g_i(x,y) \leq 0 \quad \forall i, \quad f(x,y) \leq v(y). \quad (27)$$

At first, problem (27) is much harder than the KKT reformulation because  $v(y)$  is a non-smooth function and we do not know how to compute its directional derivatives in general. Some special cases where we have derivative information for  $v(y)$  involve second order sufficient conditions or convexity. However, such conditions can be strong. Recall now that the smoothing  $v^\epsilon(y)$  of  $v(y)$  is such that  $v^\epsilon(y) \geq v(y)$ . This gives the smooth relaxation of (27) given by

$$\min_{y,x} h(y,x) \quad \text{s.t.} \quad (y,x) \in Z, \quad A(y)x = b(y), \quad g_i(x,y) \leq 0 \quad \forall i, \quad f(x,y) \leq v^\epsilon(y) + \epsilon. \quad (28)$$

Note that we changed  $v(y)$  to  $v^\epsilon(y) + \epsilon$  on problem (28). This is the motivation for the developments of Section 3 because if  $v(y) = v^\epsilon(y)$ , then problem (28) does not have a Slater point. Then, to guarantee the existence of an interior point we consider  $v^\epsilon(y) + \epsilon$  instead of  $v^\epsilon(y)$ . Moreover, because  $v^\epsilon(y) \searrow v(y)$  as  $\epsilon \searrow 0$ , we know that problem (28) has a good chance of converging monotonically in terms of  $\epsilon$  to (27). Nonetheless, there is a technical question because we can have  $v(y) = -\infty$ . This is not bad because if  $\mu > 0$ , then  $v^\epsilon(y)$  is always finite-valued. In general, the function  $v^\epsilon(y)$  is much more well behaved than  $v(y)$ . For instance, it has smaller Lipschitz constants and computable derivatives. The next proposition summarizes the convergence properties of (28).

**Proposition 4.1.** *Assume that  $Z$  is closed. Assume condition (3) and Assumption 2.1. Assume either that  $\mu > 0$  or that  $S(y)$  is non-empty and bounded for all  $y$  and the constraints  $x \geq 0$  are part of (1). Then,  $v^\epsilon(y)$  is well defined and smooth. Moreover, (28) converges monotonically and epigraphically to (27).*

*Proof.* First recall that the intersection of arbitrarily many closed sets is closed. Then, the intersection of the feasible sets of (28) is closed. Now note that such intersection is the feasible set of problem (27), which has to be closed. Writing the feasible sets using “delta” functions, the monotone behavior of the feasible sets of (28) fits into item (iii) of Theorem 2.4. Then, the epigraphical convergence follows.  $\square$

**Acknowledgments** Research of the author is funded by CNPq.

## References

- [Bar98] J. F. Bard. *Practical Bilevel Optimization: Algorithms and Applications*. Springer, US, 1998 (cit. on p. 1).
- [BH13] J. Burke and T. Hoheisel. “Epi-convergent Smoothing with Applications to Convex Composite Functions”. *SIAM Journal on Optimization* 23.3 (2013), pp. 1457–1479 (cit. on p. 1).
- [BHK13] J. Burke, T. Hoheisel, and C. Kanzow. “Gradient Consistency for Integral-Convolution Smoothing Functions”. *Set-Valued and Variational Analysis* 21.2 (2013), pp. 359–376 (cit. on p. 1).
- [BSS20a] P. Borges, C. Sagastizábal, and M. Solodov. “A Regularized Smoothing Method for Fully Parameterized Convex Problems with Applications to Convex and Nonconvex Two-Stage Stochastic Programming”. [http://www.optimization-online.org/DB\\_FILE/2020/01/7556.pdf](http://www.optimization-online.org/DB_FILE/2020/01/7556.pdf) (2020) (cit. on pp. 1, 2, 4).
- [BSS20b] P. Borges, C. Sagastizábal, and M. Solodov. “Decomposition Algorithms for Two-Stage Stochastic Hierarchical Optimization”. [http://www.optimization-online.org/DB\\_HTML/2020/05/7781.html](http://www.optimization-online.org/DB_HTML/2020/05/7781.html) (2020) (cit. on pp. 1–3).
- [BT12] A. Beck and M. Teboulle. “Smoothing and First Order Methods: A Unified Framework”. *SIAM Journal on Optimization* 22.2 (2012), pp. 557–580 (cit. on p. 1).
- [Che12] X. Chen. “Smoothing Methods for Nonsmooth, Nonconvex Minimization”. *Mathematical Programming* 134.1 (2012), pp. 71–99 (cit. on p. 1).
- [DDM07] S. Dempe, J. Dutta, and B. S. Mordukhovich. “New necessary optimality conditions in optimistic bilevel programming”. *Optimization* 56.5-6 (2007), pp. 577–604 (cit. on pp. 1, 7).
- [Dem02] S. Dempe. *Foundations of Bilevel Programming*. Springer, US, 2002 (cit. on p. 1).
- [DF14] S. Dempe and S. Franke. “Solution Algorithm for an Optimistic Linear Stackelberg Problem”. *Computers & Operations Research* 41.1 (2014), pp. 277–281 (cit. on p. 1).
- [DF16] S. Dempe and S. Franke. “On the Solution of Convex Bilevel Optimization Problems”. *Computational Optimization and Applications* 63.3 (2016), pp. 685–703 (cit. on p. 1).
- [DS99] L. Drummond and B. Svaiter. “On well definedness of the Central Path”. *Journal of Optimization Theory and Applications* 102.2 (1999), pp. 223–237 (cit. on p. 4).
- [FJQ99] F. Facchinei, H. Jiang, and L. Qi. “A smoothing method for mathematical programs with equilibrium constraints”. *Mathematical Programming* 85.1 (1999), pp. 107–134 (cit. on pp. 1, 7).
- [FP03] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, volumes 1 and 2*. Springer Series in Operations Research and Financial Engineering, 2003 (cit. on p. 1).
- [IS06] A. Izmailov and M. Solodov. “A Note on Error Estimates for some Interior Penalty Methods”. *Recent Advances in Optimization. Lecture Notes in Economics and Mathematical Systems* 563 (2006) (cit. on p. 4).
- [IS14] A. Izmailov and M. Solodov. *Newton-type methods for optimization and variational problems*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014 (cit. on p. 1).
- [Kel60] J. Kelley. “The cutting-plane method for solving convex programs”. *Journal of the Society for Industrial and Applied Mathematics* 8.1 (1960), pp. 703–712 (cit. on p. 1).
- [Kiw64] K. C. Kiwiel. *Methods of Descent for Nondifferentiable Optimization*. Lecture Notes in Mathematics, Springer, 1964 (cit. on p. 1).
- [Lan93] S. Lang. *Real and Functional Analysis*. Graduate Texts in Mathematics, Springer, 1993 (cit. on p. 5).
- [MZ98] R. D. Monteiro and F. Zhou. “On the Existence and Convergence of the Central Path for Convex Programming and Some Duality Results”. *Computational Optimization and Applications* 10.1 (1998), pp. 51–77 (cit. on p. 4).

- [Nes05] Y. Nesterov. “Smooth Minimization of Non-Smooth Functions”. *Mathematical Programming* 103.1 (2005), pp. 127–152 (cit. on p. 1).
- [OA18] A. Ouattara and A. Aswani. “Duality Approach to Bilevel Programs with a Convex Lower Level”. *Proceedings of the American Control Conference* (2018) (cit. on p. 1).
- [OSL14] W. Oliveira, C. Sagastizábal, and C. Lemaréchal. “Convex proximal bundle methods in depth: a unified analysis for inexact oracles”. *Mathematical Programming* 148.1-2 (2014), pp. 241–277 (cit. on p. 1).
- [RCSNVnu] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia, January, 2009 (cit. on p. 1).
- [RW09] T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, 2009 (cit. on pp. 1, 3).
- [WLB06] A. Wachter and T. L. Biegler. “On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming”. *Mathematical Programming* 106.1 (2006), pp. 25–57 (cit. on p. 2).
- [WTKR13] W. Wiesemann, A. Tsoukalas, P.-M. Kleniati, and B. Rustem. “Pessimistic Bilevel Optimization”. *SIAM Journal on Optimization* 23.1 (2013), pp. 353–380 (cit. on p. 7).
- [XJY14] M. Xu and J. J. Ye. “A smoothing augmented Lagrangian method for solving simple bilevel programs”. *Computational Optimization and Applications* 59.1 (2014), pp. 353–377 (cit. on p. 1).
- [XWYJ14] M. Xu, S.-Y. Wu, and J. Ye J. “Solving semi-infinite programs by smoothing projected gradient method”. *Computational Optimization and Applications* 59.3 (2014), pp. 591–616 (cit. on p. 1).