HAL
open science

Softening bilevel problems via two-scale Gibbs measures

Guillaume Carlier, Lina Mallozzi

To cite this version:

Guillaume Carlier, Lina Mallozzi. Softening bilevel problems via two-scale Gibbs measures. Set-Valued and Variational Analysis, 2022, 30 (2), pp.573-595. 10.1007/s11228-021-00605-0 . hal-02305909

HAL Id: hal-02305909

https://hal.science/hal-02305909v1

Submitted on 4 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Softening bilevel problems via two-scale Gibbs measures

Guillaume Carlier* Lina Mallozzi †

September 30, 2019

Abstract

We introduce a new, and elementary, approximation method for bilevel optimization problems motivated by Stackelberg leader-follower games. Our technique is based on the notion of two-scale Gibbs measures. The first scale corresponds to the cost function of the follower and the second scale to that of the leader. We explain how to choose the weights corresponding to these two scales under very general assumptions and establish rigorous Γ-convergence results. An advantage of our method is that it is applicable both to optimistic and to pessimistic bilevel problems.

Keywords: bilevel optimization, Stackelberg games, Gibbs measures, Γ-convergence, Laplace method.

MS Classification: 49K35, 90C26, 60B05.

1 Introduction

Bilevel optimization is defined as mathematical programming where an optimization problem contains another one as a constraint. It consists of decision making problems with hierarchical leader-follower structure and has a natural interpretation in game theory. Bilevel problems have a long history that dates back to von Stackelberg [12] and have been intensively studied from

*Université Paris Dauphine, PSL, CNRS, CEREMADE, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, France and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

†University of Naples Federico II, Department of Mathematics and Applications, Via Claudio 21, 80125 Naples, Italy mallozzi@unina.it

a theoretical point of view as well as in applications to various domains including traffic planning, security, supply chain management, principal-agent models, production planning, market deregulation, optimal taxation, parameter estimation, see the recent surveys [6, 8] and the references therein. In the context of Cournot duopolies, von Stackelberg investigated the leader-follower model where the leader firm maximizes profit under the constraint that the follower firm reacts with an optimal choice of the quantity that is supposed to be unique. Later on, Leitmann [10] discussed the case where the optimal solutions for the follower's problem form a set that the leader has to take into account in order to solve her own optimization problem.

In case the follower's program has several solutions, we see that there is some ambiguity even in the definition of the leader's program. In the literature, several concepts have been considered. The optimistic (or strong) Stackelberg solution assumes a cooperative like behavior between the agents: the leader expects the follower to choose solutions leading to the best outcome for her. On the contrary, the pessimistic (or weak) Stackelberg solution assumes that the follower always breaks ties by choosing the worst actions for the leader which corresponds to a security strategy for her, see [2, 5]. Some intermediate cases can also be considered. In [1], a cooperation degree is assumed leading to the optimization of a convex combination of the best and the worst payoff value for the leader, while, in [11], a probabilistic information about the follower's behavior is assumed resulting in the optimization of an average payoff.

Both optimistic and pessimistic bilevel programs are challenging and often difficult to solve in practice. In the present paper, we present a new and quite simple (unconstrained) approximation scheme for such problems based on the notion of two-scale Gibbs measures. Our method is directly inspired by the classical Laplace method: Gibbs probability measures which have a density proportional to $e^{-\lambda u}$ with respect to a reference measure with full support concentrate on the set where u is minimal as $\lambda \to \infty$. We refer to Hwang [9] for a fine study of the method and precise statements in smooth finite-dimensional situations. In the context of bilevel optimization, we have to take into account the objective of both the leader and the follower and a single parameter λ is not enough capture the nested structure of the leader-follower problem. This is why we introduce two-scale Gibbs measures where the first scale (with weight λ) takes into account the follower's objective and the secondary one (with a smaller weight to be chosen properly) takes into account the leader's objective. We investigate in details convergence issues (both in the pointwise and Γ-convergence sense) and the choice of the secondary scale only in terms of the reference measure and the modulus

of continuity of the leader and follower objective functions. Our method is flexible enough to cope with both the optimistic and the pessimistic case.

The paper is organized as follows: the setting of our analysis is introduced in section 2. In section 3, we recall some basic facts on Gibbs measures and analyze the convergence of two-scale Gibbs measures in a simple case. In section 4, we give a construction for the weights which guarantees convergence of the corresponding two-scale Gibbs measures under general assumptions. We then establish Γ-convergence results, first for the optimistic case in section 5 and then in the pessimistic case in section 6. Finally, section 7 concludes with some remarks and examples.

2 Setting

Throughout this paper, X (strategy set for the leader) and Y (strategy set for the follower) will be two compact metric spaces. We will also assume that the cost functions of both the leader and the follower are continuous, $\varphi \in C(X \times Y)$, $\psi \in C(X \times Y)$ will denote the cost function of the leader (who chooses $x \in X$) and the follower (choosing $y \in Y$) respectively. The Stackelberg problem is the program of the leader which reads

$$\inf_{(x,y)\in X\times Y} \Big\{ \varphi(x,y) \ : \ y \in \operatorname{argmin} \psi(x,.) \Big\}. \tag{2.1}$$

Under our assumptions, it is obvious that (2.1) admits at least one solution but finding such solutions in practice is a challenging task due to the constraint $y \in \operatorname{argmin} \psi(x,.)$. Of course, one can rewrite (2.1) as a minimization problem with respect to x only:

$$\inf_{x\in X} \varphi_*(x) \tag{2.2}$$

where

$$\varphi_*(x) := \min \Big\{ \varphi(x,y) \ : \ y \in \operatorname{argmin} \psi(x,.) \Big\}. \tag{2.3}$$

Problem (2.1) is usually refered to as the *optimistic* problem since it assumes that in case the follower has several optimal strategies she will break ties by chosing one which is optimal for the leader. The *pessimistic* problem consists, on the contrary, in assuming that the follower actually breaks ties by chosing strategies which are the worst for the leader. The corresponding bilevel pessimistic program therefore consists in

$$\inf_{x\in X} \varphi^*(x) \tag{2.4}$$

where

$$\varphi^*(x) := \max\Big\{\varphi(x,y) \ : \ y \in \operatorname{argmin} \psi(x,.)\Big\}. \tag{2.5}$$

In general, the pessimistic value φ^* is not lower-semicontinuous (lsc) so (2.4) does not necessarily admit solutions which makes the pessimistic problem more involved than the optimistic one and requires some suitable relaxation of φ^*.

Our goal is to approximate the bilevel problem (2.1) by a family of unconstrained ones (we will also address the approximation of the pessimistic bilevel problems (2.4) in section 6). We shall indeed prove that the somehow rough function φ_* can be approximated by a family of more regular ones defined by an integral depending on a parameter. By approximated we mean both in the pointwise sense and in the sense of Γ-convergence we recall below (see [4] or [7] for an overview of Γ-convergence and its applications):

Definition 2.1. *Let $F : X \to \mathbb{R}$ and let for every $\lambda > 0$, $F_\lambda : X \to \mathbb{R}$, then F_λ is said to Γ-converge to F as $\lambda \to +\infty$ if the following two conditions are satisfied:*

- *for every $x \in X$ and every family $(x_\lambda)_{\lambda>0}$ converging to x as $\lambda \to +\infty$, one has the Γ-liminf inequality:*

$$F(x) \leq \liminf_{\lambda\to+\infty} F_\lambda(x_\lambda),$$

- *for every $x \in X$, there exists a (so-called recovery) family $(x_\lambda)_{\lambda>0}$ converging to x as $\lambda \to +\infty$ such that the following Γ-limsup inequality holds*

$$F(x) \geq \limsup_{\lambda\to+\infty} F_\lambda(x_\lambda).$$

Our approximation is a variant of the celebrated Laplace method, which as far as we know, has not been investigated in the bilevel framework. First of all, we give ourselves a Borel probability measure ν on Y (which we will denote $\nu \in \mathcal{P}(Y)$) and assume that it has full support

$$\operatorname{spt}(\nu) = Y. \tag{2.6}$$

For $r > 0$ and $y \in Y$, we denote by $B_r(y)$ the closed ball of radius r centered at y and set for every $r \geq 0$,

$$\alpha_\nu(r) := \inf_{y\in Y} \nu(B_r(y)). \tag{2.7}$$

Note that the full support assumption (2.6) and the compactness of Y ensure that $\alpha_\nu(r) > 0$ for every $r > 0$.

Given $x \in X$, $\lambda > 0$ and $\delta > 0$, we consider the probability measure on Y

$$\mu_{\lambda,\delta}(\mathrm{d}y|x) := Z_{\lambda,\delta}(x) e^{-\lambda(\psi(x,y)+\delta\varphi(x,y))} \nu(\mathrm{d}y) \tag{2.8}$$

where $Z_{\lambda,\delta}(x)$ is the normalizing constant which makes $\mu_{\lambda,\delta}(.|x)$ a probability measure i.e.

$$Z_{\lambda,\delta}(x) := \frac{1}{\int_Y e^{-\lambda(\psi(x,y)+\delta\varphi(x,y))}\nu(\mathrm{d}y)}.$$

Our main result is that one can choose the secondary scale $\delta = \delta_\lambda$ with

$$\lim_{\lambda\to+\infty} \delta_\lambda = 0, \ \lim_{\lambda\to+\infty} \lambda\delta_\lambda = +\infty \tag{2.9}$$

in such a way that the family

$$x \mapsto \varphi_\lambda(x) := \int_Y \varphi(x,y)\mu_{\lambda,\delta_\lambda}(\mathrm{d}y|x)$$

Γ-converges and converges pointwise to φ_* as $\lambda \to \infty$. Our construction of δ_λ only depends on the function α_ν defined in (2.8) and a modulus of continuity of φ and ψ, it will be detailed in section 4.

3 On Gibbs measures

3.1 On standard Gibbs measures

In this section, we temporarily leave the approximation of Stackelberg problems and focus on the asymptotic behavior of Gibbs measures. Given $\nu \in \mathcal{P}(Y)$ with full support as in (2.6), $\lambda > 0$ and $w \in C(Y)$, we define the Gibbs measure

$$\nu_{\lambda,w} := Z_{\lambda,w} e^{-\lambda w}\nu, \ Z_{\lambda,w} := \frac{1}{\int_Y e^{-\lambda w(y)}\nu(\mathrm{d}y)}. \tag{3.1}$$

Of course, $\nu_{\lambda,w}$ is unchanged if one adds a constant to w so there is no loss of generality in normalizing w in some way, and the most natural way is to assume that its minimum is 0. The following elementary result will be used intensively in the sequel:

Lemma 3.1. *Let $w \in C(Y)$ be such that $\min_Y w = 0$ and let $\nu_{\lambda,w}$ be defined by (3.1) then for every $\varepsilon > 0$ there holds*

$$\nu_{\lambda,w}(\{w \geq \varepsilon\}) \leq \frac{e^{-\frac{\lambda\varepsilon}{2}}}{\nu(\{w \leq \frac{\varepsilon}{2}\})}. \tag{3.2}$$

Proof. Let us write

$$\nu_{\lambda,w}(\{w \geq \varepsilon\}) = \frac{\int_{\{w\geq\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y)}{\int_{\{w\geq\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y) + \int_{\{w<\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y)}$$

so that

$$\frac{1}{\nu_{\lambda,w}(\{w \geq \varepsilon\})} = 1 + \frac{\int_{\{w<\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y)}{\int_{\{w\geq\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y)}$$

then observe that

$$\int_{\{w\geq\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y) \leq e^{-\lambda\varepsilon}$$

and

$$\int_{\{w<\varepsilon\}} e^{-\lambda w(y)}\nu(\mathrm{d}y) \geq \int_{\{w\leq\frac{\varepsilon}{2}\}} e^{-\lambda w(y)}\nu(\mathrm{d}y) \geq e^{-\frac{\lambda\varepsilon}{2}}\nu(\{w \leq \frac{\varepsilon}{2}\})$$

so putting eveything together yields (3.2).

□

Since, for every λ, $\nu_{\lambda,w}$ is a probabilty measure and Y is compact, there is a sequence $\lambda_n \to \infty$ and $\overline{\nu} \in \mathcal{P}(Y)$ such that, $\nu_{\lambda_n,w}$ weakly[1] star converges to $\overline{\nu}$ as $n \to \infty$. Since the set $\{w > \varepsilon\}$ is open, it follows from Portmanteau's theorem (see [3]) that

$$\overline{\nu}(\{w > \varepsilon\}) \leq \liminf_n \nu_{\lambda_n,w}(\{w > \varepsilon\}).$$

Hence, for every $\varepsilon > 0$ thanks to (3.2) and the fact that $\nu(\{w \leq \frac{\varepsilon}{2}\}) > 0$ (because ν has full support and the minimum of w is 0), we get

$$\overline{\nu}(\{w > \varepsilon\}) = 0$$

letting $\varepsilon \to 0^+$ and using the fact that w is continuous, we conclude that $\overline{\nu}$ concentrates on the set where $w = 0$. We thus recover the well-known fact that Gibbs measures concentrate on the set where the potential is minimal:

Corollary 3.2. *Let $w \in C(Y)$ and $\nu_{\lambda,w}$ be defined by (3.1), then any weak star cluster point of $\nu_{\lambda,w}$ as $\lambda \to \infty$ has its support in* $\operatorname{argmin}_Y w$.

[1]Recall that $(\nu_n)_n \in \mathcal{P}(Y)^{\mathbb{N}}$ weakly star converges to $\overline{\nu}$ if $\int_Y u(y)\nu_n(\mathrm{d}y) \to \int_Y u(y)\overline{\nu}(\mathrm{d}y)$ for every $u \in C(Y)$.

3.2 Convergence of two-scale Gibbs measures in a simple case

To understand how to approximate bilevel problems with Gibbs measures, we first have to understand the following question. Given two functions u and v in $C(Y)$, we want to find a weight δ_λ such that $\delta_\lambda \to 0$, $\lambda\delta_\lambda \to +\infty$ as $\lambda \to +\infty$ in such a way that the two-scale Gibbs measure $\nu_{\lambda,u+\delta_\lambda v}$ concentrates when $\lambda \to \infty$ on the double argmin set:

$$\operatorname{argmin}_{\operatorname{argmin} u} v := \{y \in \operatorname{argmin} u \,:\, v(y) = \min\{v(z) \,:\, z \in \operatorname{argmin} u\}\}. \tag{3.3}$$

We will give a general constructive answer in paragraph 4 (depending on the modulus of continuity of u and v and the function α_ν in (2.7)). Yet, for now, we prefer to focus on a rather simple case where the explicit choice $\delta_\lambda := \frac{1}{\sqrt{\lambda}}$ works (as well as many other simple ones, see Remark 3.4 below). This simple case corresponds to the extra assumptions that both u and v are Hölder continuous and the function α_ν is bounded from below by a power function. Denoting by dist the distance on Y and $\operatorname{diam}(Y)$ its diameter, these assumptions mean that there exist $C > 0$, $\alpha \in (0, 1]$ and $\beta \in \mathbb{R}_+^*$ such that

$$\max(|u(y) - u(z)|, |v(y) - v(z)|) \leq C \operatorname{dist}(y, z)^\alpha, \ \forall (y, z) \in Y^2, \tag{3.4}$$

and

$$\alpha_\nu(r) \geq \frac{r^\beta}{C}, \ \forall r \in [0, \operatorname{diam}(Y)]. \tag{3.5}$$

To shorten notations, let us set

$$\widetilde{\nu}_\lambda := \widetilde{Z}_\lambda e^{-\lambda u - \sqrt{\lambda} v} \nu, \ \widetilde{Z}_\lambda := \frac{1}{\int_Y e^{-\lambda u - \sqrt{\lambda} v} \nu} \tag{3.6}$$

which corresponds to the the two-scale Gibbs measure $\nu_{\lambda, u + \frac{1}{\sqrt{\lambda}} v}$.

Proposition 3.3. *Assume that u and v satisfy (3.4), that ν satisfies (3.5) and define $\widetilde{\nu}_\lambda$ by (3.6) then any weak star cluster point of $\widetilde{\nu}_\lambda$ as $\lambda \to \infty$ has its support in the double argmin set* $\operatorname{argmin}_{\operatorname{argmin} u} v$.

Proof. To ease notations, let us normalize u and v in such a way that

$$\min_Y u = 0, \ \min\{v(y) \,:\, y \in Y, \ u(y) = 0\} = 0. \tag{3.7}$$

Also define $w_\lambda := u + \frac{v}{\sqrt{\lambda}}$ and observe that (3.7) implies that $\min_Y w_\lambda \leq 0$. Let then $\lambda_n \to \infty$, $\overline{\nu} \in \mathcal{P}(Y)$ such that $\widetilde{\nu}_{\lambda_n}$ weakly star converges to $\overline{\nu}$. Let

$\varepsilon > 0$, for λ large enough $\{u > \varepsilon\} \subset \{w_\lambda > \frac{\varepsilon}{2}\} \subset \{w_\lambda > \min_Y w_\lambda + \frac{\varepsilon}{2}\}$ and $\{u \leq \frac{\varepsilon}{8}\} \subset \{w_\lambda \leq \min_Y w_\lambda + \frac{\varepsilon}{4})$. It then follows from the inequality (3.2) of Lemma 3.1 that

$$\begin{aligned}\widetilde{\nu}_\lambda(\{u > \varepsilon\}) &\leq \widetilde{\nu}_\lambda(\{w_\lambda > \min_Y w_\lambda + \frac{\varepsilon}{2}\}) \\ &\leq \frac{e^{-\frac{\lambda\varepsilon}{4}}}{\nu(\{w_\lambda \leq \min_Y w_\lambda + \frac{\varepsilon}{4}\})} \leq \frac{e^{-\frac{\lambda\varepsilon}{4}}}{\nu(\{u \leq \frac{\varepsilon}{8}\})}\end{aligned}$$

so that, again thanks to Portmanteau's Theorem, $\overline{\nu}(\{u > \varepsilon\}) = 0$ and $\overline{\nu}$ is supported by $\operatorname{argmin} u = \{u = 0\}$. In particular, with (3.7), $v \geq 0$ on $\operatorname{spt}(\overline{\nu})$. To conclude, we thus have to show that for every $\varepsilon > 0$, $\overline{\nu}(\{v > \varepsilon\}) = 0$. Since $u \geq 0$ and $\min_Y w_\lambda \leq 0$, we have $\{v > \varepsilon\} \subset \{w_\lambda > \min_Y w_\lambda + \frac{\varepsilon}{\sqrt{\lambda}}\}$ so using Lemma 3.1 again we get

$$\widetilde{\nu}_\lambda(\{v > \varepsilon\}) \leq \frac{e^{-\frac{\sqrt{\lambda}\varepsilon}{2}}}{\nu(\{w_\lambda \leq \min_Y w_\lambda + \frac{\varepsilon}{2\sqrt{\lambda}}\})}. \tag{3.8}$$

Let y_λ be a point where w_λ achieves its minimum, then it follows from (3.4) that for $\lambda \geq 1$, $w_\lambda \leq \min_Y w_\lambda + \frac{\varepsilon}{2\sqrt{\lambda}}$ in the ball of center y_λ and radius $(\frac{\varepsilon}{4C\sqrt{\lambda}})^{\frac{1}{\alpha}}$, hence

$$\nu(\{w_\lambda \leq \min_Y w_\lambda + \frac{\varepsilon}{2\sqrt{\lambda}}\}) \geq \alpha_\nu\Big(\Big(\frac{\varepsilon}{4C\sqrt{\lambda}}\Big)^{\frac{1}{\alpha}}\Big) \geq \frac{1}{C}\Big(\frac{\varepsilon}{4C\sqrt{\lambda}}\Big)^{\frac{\beta}{\alpha}}$$

where the last inequality follows from (3.5). With (3.8), this yields

$$\widetilde{\nu}_\lambda(\{v > \varepsilon\}) \leq \varepsilon^{-\frac{\beta}{\alpha}} 4^{\frac{\beta}{\alpha}} C^{1+\frac{\beta}{\alpha}} \sqrt{\lambda}^{\frac{\beta}{\alpha}} e^{-\frac{\sqrt{\lambda}\varepsilon}{2}}$$

since for every $\varepsilon > 0$ the right hand side tends to 0 as $\lambda \to \infty$, Portmanteau's Theorem again allows us to conclude that $\overline{\nu}(\{v > \varepsilon\}) = 0$. □

Remark 3.4. Our choice $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ above is just for illustrative purpose and by no means the only possible one or optimal in any sense. It is indeed straightforward to check, with the same proof as above, that under the assumptions of Proposition 3.3, any choice of δ_λ such that

$$\lim_{\lambda\to+\infty} \delta_\lambda = 0, \lim_{\lambda\to+\infty} \lambda\delta_\lambda = +\infty, \lim_{\lambda\to+\infty} \frac{\log(\delta_\lambda)}{\lambda\delta_\lambda} = 0 \tag{3.9}$$

guarantees that the corresponding two-scale measures $\nu_{\lambda,u+\delta_\lambda v}$ tend to concentrate on the double argmin set (3.3) as $\lambda \to +\infty$. In particular any power

choice for δ_λ i.e. $\delta_\lambda = \lambda^{-\gamma}$ with $\gamma \in (0,1)$ (or much larger weights such as $\delta_\lambda = \frac{1}{\log(\lambda)}$, $\delta_\lambda = \frac{1}{\log\log(\lambda)}$) ensures convergence to the double argmin set. Note that smaller weights such as $\delta_\lambda = \frac{\log(\lambda)}{\lambda}$ violate condition (3.9). Assumptions (3.4) and (3.5) are essential if one wishes to use power like weights. In section 7, we will consider examples where α_ν is much smaller than a power function. In such cases, the choice $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ may rule out the desired convergence property (see Example 7.1). Even worse, it may be the case that no power-like weight converges to the double argmin set (see Example 7.2).

An immediate consequence of Proposition 3.3 is

Corollary 3.5. *Under the same assumptions as in Proposition 3.3, we have*

$$\lim_{\lambda\to\infty} \int_Y v(y)\widetilde{\nu}_\lambda(dy) = \min\{v(y) \ : \ y \in \operatorname{argmin} u\}.$$

4 Chosing the weights under general assumptions

Now we consider the general case where u and v are continuous and ν has full support. We wish to find secondary weights δ_λ satisfying (2.9) in such a way that defining the two-scale Gibbs measures,

$$\widetilde{\nu}_\lambda := \widetilde{Z}_\lambda e^{-\lambda(u+\delta_\lambda v)}\nu, \ \widetilde{Z}_\lambda := \frac{1}{\int_Y e^{-\lambda(u+\delta_\lambda v)}\nu} \tag{4.1}$$

every weak star cluster point of $\widetilde{\nu}_\lambda$ as $\lambda \to \infty$ is supported by the double argmin set (3.3).

First of all, we fix a common modulus of continuity ω for both u and v in the sense that:

$$|u(y) - u(z)| + |v(y) - v(z)| \leq \omega(\operatorname{dist}(y,z)), \ \forall (y,z) \in Y^2 \tag{4.2}$$

without loss of generality, we may assume that ω is strictly increasing and concave (hence continuous) on the whole of $\mathbb{R}_+$ and of course $\omega(t) \to 0^+$ as $t \to 0$. We the denote by Γ, $\mathbb{R}_+^* \to \mathbb{R}_+^*$ the inverse of ω, $\Gamma := \omega^{-1}$, by construction we thus have

$$z \in B_{\Gamma(\varepsilon)}(y) \Rightarrow |u(y) - u(z)| + |v(y) - v(z)| \leq \varepsilon. \tag{4.3}$$

Recalling that α_ν is defined by (2.7), we define for every $t > 0$

$$\theta(t) := \lim_{\delta\to 0^+} \log(\alpha_\nu(\Gamma(t-\delta))). \tag{4.4}$$

Note then that θ is nondecreasing, lsc (this is why we define it as a left limit) and $\theta(t) \to -\infty$ as $t \to 0^+$.

Lemma 4.1. *Let θ be given by (4.4), for every $\lambda > 0$ define*

$$t_\lambda := \inf\{t > 0 \ : \ \theta(t) + \frac{\lambda t}{2} \geq 0\} \tag{4.5}$$

and

$$\delta_\lambda := 2\sqrt{t_\lambda} \tag{4.6}$$

then $\lambda \mapsto \delta_\lambda$ is decreasing, satisfies (2.9) and

$$\lim_{\lambda \to \infty} \theta\Big(\frac{\delta_\lambda^2}{2}\Big) + \frac{\lambda \delta_\lambda^2}{2} = +\infty. \tag{4.7}$$

Proof. First of all, it is clear that t_λ is well defined since $t \mapsto \theta(t) + \frac{\lambda t}{2}$ is an increasing function which tends to $-\infty$ as $t \to 0^+$ and $+\infty$ as $t \to +\infty$. Moreover since we chose θ lsc we have

$$\theta(t_\lambda) + \frac{\lambda t_\lambda}{2} \leq 0 \tag{4.8}$$

and since $2t_\lambda > t_\lambda$ we also have

$$\theta(2t_\lambda) + \lambda t_\lambda \geq 0. \tag{4.9}$$

It is obvious by monotonicity that $\lambda \mapsto t_\lambda$ is nonincreasing and dividing (4.8) by λ and letting $\lambda \to +\infty$ one easily sees that t_λ converges to 0 as $\lambda \to +\infty$, this gives $\theta(2t_\lambda) \to -\infty$ as $\lambda \to +\infty$. Hence, thanks to (4.9), we have

$$\lim_{\lambda \to +\infty} \lambda t_\lambda = +\infty. \tag{4.10}$$

The fact that δ_λ decreases to 0 and $\lambda \delta_\lambda$ tends to $+\infty$ immediately follow. Finally, by definition of δ_λ and using (4.9), we have

$$\theta\Big(\frac{\delta_\lambda^2}{2}\Big) + \frac{\lambda \delta_\lambda^2}{2} = \theta(2t_\lambda) + 2\lambda t_\lambda \geq \lambda t_\lambda$$

so (4.7) directly follows from (4.10).

□

Proposition 4.2. *Let δ_λ be defined as in Lemma 4.1 and $\widetilde{\nu}_\lambda$ be defined by (4.1). Then any weak star cluster point of $\widetilde{\nu}_\lambda$ as $\lambda \to \infty$ has its support included in the double argmin set (3.3).*

Proof. Again we can normalize u and v so that (3.7) holds and set $w_\lambda := u + \delta_\lambda v$, since $\delta_\lambda \to 0$, one can proceed as in the proof of Proposition 3.3 to show that for every $\varepsilon > 0$, $\widetilde{\nu}_\lambda(\{u > \varepsilon\})$ tends to 0 as $\lambda \to +\infty$ and thus deduce that any weak star cluster point of $\widetilde{\nu}_\lambda$ as $\lambda \to \infty$ has its support included in $\{u = 0\} = \operatorname{argmin} u$. To conclude that such weak star cluster points are in fact supported by the double argmin set (3.3), it is enough to show that for every $\varepsilon > 0$, $\widetilde{\nu}_\lambda(\{v > \varepsilon\})$ tends to 0 as $\lambda \to +\infty$. To prove this, we observe that since $u \geq 0$, $\{v > \varepsilon\} \subset \{w_\lambda \geq \delta_\lambda \varepsilon\}$ and since $\min_Y w_\lambda \leq 0$ (because of (3.7)), we have $\{v > \varepsilon\} \subset \{w_\lambda \geq \min_Y w_\lambda + \delta_\lambda \varepsilon\}$. Since $\widetilde{\nu}_\lambda = \nu_{\lambda, w_\lambda}$, our basic inequality (3.2) in Lemma 3.1 gives:

$$\widetilde{\nu}_\lambda(\{v > \varepsilon\}) \leq \widetilde{\nu}_\lambda(\{w_\lambda \geq \min_Y w_\lambda + \delta_\lambda \varepsilon\}) \leq \frac{e^{-\frac{\lambda \delta_\lambda \varepsilon}{2}}}{\nu(\{w_\lambda \leq \min_Y w_\lambda + \frac{\delta_\lambda \varepsilon}{2}\})}. \tag{4.11}$$

Choose now λ large enough so that $\delta_\lambda \leq 1$, doing so ω is a modulus of continuity of w_λ. Hence if y_λ is a minimum point of w_λ, the ball $B_{\Gamma(\frac{\delta_\lambda \varepsilon}{2})}(y_\lambda)$ is contained in $\{w_\lambda \leq \min_Y w_\lambda + \frac{\delta_\lambda \varepsilon}{2}\}$ hence by definition of Γ, α_ν and θ we get

$$\nu(\{w_\lambda \geq \min_Y w_\lambda + \frac{\delta_\lambda \varepsilon}{2}\}) \geq \alpha_\nu\Big(\Gamma(\frac{\delta_\lambda \varepsilon}{2})\Big) \geq e^{\theta(\frac{\delta_\lambda \varepsilon}{2})}$$

replacing in (4.11) gives

$$\widetilde{\nu}_\lambda(\{v > \varepsilon\}) \leq e^{-\frac{\lambda \delta_\lambda \varepsilon}{2} - \theta(\frac{\delta_\lambda \varepsilon}{2})} \tag{4.12}$$

for λ large enough, $\delta_\lambda \leq \varepsilon$ so that, using the monotonicity of θ:

$$\frac{\lambda \delta_\lambda \varepsilon}{2} + \theta(\frac{\delta_\lambda \varepsilon}{2}) \geq \frac{\lambda \delta_\lambda^2}{2} + \theta\Big(\frac{\delta_\lambda^2}{2}\Big)$$

and since the right hand side tends to $+\infty$ as $\lambda \to +\infty$ thanks to (4.7), we obtain that for every $\varepsilon > 0$

$$\lim_{\lambda \to +\infty} \widetilde{\nu}_\lambda(\{v > \varepsilon\}) = \lim_{\lambda \to +\infty} e^{-\frac{\lambda \delta_\lambda \varepsilon}{2} - \theta(\frac{\delta_\lambda \varepsilon}{2})} = 0,$$

which gives the desired conclusion.

□

Remark 4.3. In fact the proof of Proposition 4.2 gives a more precise information than the fact that weak cluster points of $\widetilde{\nu}_\lambda$ concentrate on the double argmin set (3.3). Thanks to (4.12), we in fact have

$$\widetilde{\nu}_\lambda(\{v > \min_{\operatorname{argmin} u} v + \varepsilon\}) \leq e^{-\frac{\lambda \delta_\lambda \varepsilon}{2} - \theta(\frac{\delta_\lambda \varepsilon}{2})} \tag{4.13}$$

and by construction for every $\varepsilon > 0$:

$$\frac{\lambda\delta_\lambda\varepsilon}{2} + \theta(\frac{\delta_\lambda\varepsilon}{2}) \to +\infty \text{ as } \lambda \to \infty. \tag{4.14}$$

Now observe that the quantity in (4.14) only depends on u and v through the function θ i.e. through the modulus ω. So the speed at which $\widetilde{\nu}_\lambda(\{v > \min_{\text{argmin}\, u} v + \varepsilon\})$ converges to 0 is uniform with respect to v and u admitting ω as modulus of continuity. This uniform behavior will be useful in section 6 devoted to the pessimistic case.

Remark 4.4. Let us also emphasize that what really matters for convergence is the fact that the left-hand side of (4.13) tends to 0 as $\lambda \to +\infty$ for every $\varepsilon > 0$ and not really the precise construction of δ_λ. Our construction based on Lemma 4.1 is just a cooking recipe which guarantees this property, and we are not claiming that it is optimal in any sense. In fact, the choice $\delta_\lambda = 2\sqrt{t_\lambda}$ in Lemma 4.1 is a bit arbitrary, taking $\delta_\lambda = t_\lambda^\gamma$ with $\gamma \in (0,1)$ or $\delta_\lambda = -t_\lambda \log(t_\lambda)$ would have worked just as well.

5 Γ-convergence to the bilevel problem

We are now ready to go back to the bilevel problem (2.2) of paragraph 2. Since both φ and ψ are continuous on $X \times Y$ which is compact, we can find ω, a (strictly increasing and concave) modulus of continuity of $\varphi(x,.)$, $\psi(x,.)$ which is uniform with respect to $x \in X$ i.e.:

$$|\varphi(x,y) - \varphi(x,z)| + |\psi(x,y) - \psi(x,z)| \leq \omega(\text{dist}(y,z)),\ \forall (x,y,z) \in X \times Y^2 \tag{5.1}$$

we then define $\Gamma := \omega^{-1}$ as in the previous paragraph, θ as in (4.4) and then the weights δ_λ by (4.5)-(4.6) as in Lemma 4.1. Recalling (2.8), let us set for every $x \in X$,

$$\widetilde{\mu}_\lambda(\mathrm{d}y|x) := \mu_{\lambda,\delta_\lambda}(\mathrm{d}y|x)$$

i.e.

$$\widetilde{\mu}_\lambda(\mathrm{d}y|x) := \widetilde{Z}_\lambda(x) e^{-\lambda(\psi(x,y)+\delta_\lambda\varphi(x,y))}\nu(\mathrm{d}y)$$

with

$$\widetilde{Z}_\lambda(x) := Z_{\lambda,\delta_\lambda}(x) = \frac{1}{\int_Y e^{-\lambda(\psi(x,y)+\delta_\lambda\varphi(x,y))}\nu(\mathrm{d}y)}.$$

Finally, let us define

$$\varphi_\lambda(x) := \int_Y \varphi(x,y)\widetilde{\mu}_\lambda(\mathrm{d}y|x),\ \forall x \in X. \tag{5.2}$$

Our main result is then the following

Theorem 5.1. *Let φ_* be defined by (2.3) and φ_λ be as above. Then φ_λ Γ-converges and converges pointwise to φ_* as $\lambda \to +\infty$.*

Proof. The pointwise convergence follows directly from Proposition 4.2. The Γ-limsup inequality straighforwardly follows (taking the constant $x_\lambda = x$ as recovery sequence). It remains to show the Γ-liminf inequality. Let then $x \in X$, $x_\lambda \to x$ as $\lambda \to +\infty$, since $\varphi(x_\lambda, .)$ converges uniformly to $\varphi(x, .)$, we first have

$$\liminf_{\lambda\to\infty} \varphi_\lambda(x_\lambda) = \liminf_{\lambda\to\infty} \int_Y \varphi(x,y)\widetilde{\mu}_\lambda(\mathrm{d}y|x_\lambda) \tag{5.3}$$

we may then find a sequence $\lambda_n \to +\infty$ in such a way that $\mu_n := \widetilde{\mu}_{\lambda_n}(.|x_{\lambda_n})$ weakly star converges to some $\mu \in \mathcal{P}(Y)$ and

$$\liminf_{\lambda\to\infty} \int_Y \varphi(x,y)\widetilde{\mu}_\lambda(\mathrm{d}y|x_\lambda) = \lim_n \int_Y \varphi(x,y)\widetilde{\mu}_{\lambda_n}(\mathrm{d}y|x_{\lambda_n}) = \int_Y \varphi(x,y)\mu(\mathrm{d}y)$$

so if we show that

$$\psi(x,.) = \min_Y \psi(x,.) \ \mu\text{-a.e.} \tag{5.4}$$

this will show that $\int_Y \varphi(x,y)\mu(\mathrm{d}y) \geq \varphi_*(x)$ by definition of φ_* yielding the desired Γ-liminf inequality. To prove (5.4), we take $\varepsilon > 0$ and observe that for large enough n, the set $\{y \in Y \ : \ \psi(x,y) > \min_Y \psi(x,.) + \varepsilon\}$ is included in

$$\{y \in Y \ : \ \psi(x_{\lambda_n}, y) + \delta_{\lambda_n}\varphi(x_{\lambda_n}, y) > \min_Y\{\psi(x_{\lambda_n}, .) + \delta_{\lambda_n}\varphi(x_{\lambda_n}, .)\} + \frac{\varepsilon}{2}\}.$$

Let us set $f_n := \psi(x_{\lambda_n}, .) + \delta_{\lambda_n}\varphi(x_{\lambda_n}, .)$ to shorten notations. Thanks to (3.2) again, we have

$$\mu_n\Big(\{\ \psi(x,.) > \min_Y \psi(x,.) + \varepsilon\}\Big) \leq \frac{e^{-\frac{\lambda_n\varepsilon}{4}}}{\nu(\{f_n \leq \min_Y f_n + \frac{\varepsilon}{4}\})}$$

and since f_n converges uniformly to $\psi(x,.)$, for large enough n, $\{f_n \leq \min_Y f_n + \frac{\varepsilon}{4}\}$ contains a neighbourhood of a minimizer of $\psi(x,.)$, we obtain

$$\lim_n \mu_n\Big(\{\ \psi(x,.) > \min_Y \psi(x,.) + \varepsilon\}\Big) = 0.$$

Portmanteau's theorem thus implies that $\mu\Big(\{\ \psi(x,.) > \min_Y \psi(x,.) + \varepsilon\}\Big) = 0$ which ends the proof.

□

As a consequence of the previous Γ-convergence, we immediately have:

Corollary 5.2. *Let φ_* be defined by (2.3) and φ_λ be as above, then*

$$\lim_{\lambda\to+\infty} \min_X \varphi_\lambda = \min_{x\in X} \varphi_* = \min\{\varphi(x,y),\ (x,y)\in X\times Y,\ y\in \operatorname{argmin}\psi(x,.)\}$$

and if x_λ is a minimizer of φ_λ and x is a cluster point of x_λ as $\lambda\to+\infty$ then x is a minimizer of φ_*.

6 The pessimistic case

Let us now address the pessimistic case and the approximation via Gibbs-measures of the pessimistic value function φ^* defined in (2.5). Since φ^* is not necessarily lsc, we have to relax it and consider its lsc envelope (that is the largest lsc function lying below φ^* on X) which we denote by $\underline{\varphi^*}$ and is given by

$$\underline{\varphi^*}(x) := \liminf_{x'\to x} \varphi^*(x') = \lim_{r\to 0^+} \inf_{B_r(x)} \varphi^*. \tag{6.1}$$

We construct δ_λ exactly as in paragraph 5, and for every $x\in X$, we consider the Gibbs measure:

$$\widetilde{\mu}^+_\lambda(\mathrm{d}y|x) := \widetilde{Z}^+_\lambda(x) e^{-\lambda(\psi(x,y)-\delta_\lambda\varphi(x,y))}\nu(\mathrm{d}y)$$

with

$$\widetilde{Z}^+_\lambda(x) := \frac{1}{\int_Y e^{-\lambda(\psi(x,y)-\delta_\lambda\varphi(x,y))}\nu(\mathrm{d}y)}.$$

We consider then the approximations

$$\varphi^+_\lambda(x) := \int_Y \varphi(x,y)\widetilde{\mu}^+_\lambda(\mathrm{d}y|x),\ \forall x\in X. \tag{6.2}$$

Our convergence result concerning the pessimistic value is the following:

Theorem 6.1. *Let φ^* be defined by (2.5) and φ^+_λ be as above. Then, as $\lambda\to+\infty$, φ^+_λ converges pointwise to φ^* and Γ-converges to its lsc envelope, $\underline{\varphi^*}$, defined in (6.1).*

Proof. The pointwise convergence of φ^+_λ to φ^* follows directly from Proposition 4.2.

Let us then introduce the Γ-liminf and Γ-limsup:

$$\Gamma-\liminf_\lambda \varphi^+_\lambda(x) := \inf\left\{\liminf_\lambda \varphi^+_\lambda(x_\lambda)\ :\ x_\lambda\to x\right\}$$

and

$$\Gamma - \limsup_{\lambda} \varphi_\lambda^+(x) := \inf \Big\{ \limsup_{\lambda} \varphi_\lambda^+(x_\lambda) \ : \ x_\lambda \to x \Big\}.$$

Since φ_λ^+ converges pointwise to φ^* we obviously have

$$\Gamma - \limsup_{\lambda} \varphi_\lambda^+ \leq \varphi^*$$

but since $\Gamma - \limsup_\lambda \varphi_\lambda^+$ is lsc (see Proposition 1.28 in [4]) this also implies

$$\Gamma - \limsup_{\lambda} \varphi_\lambda^+ \leq \underline{\varphi^*}. \tag{6.3}$$

We now aim to show that

$$\Gamma - \liminf_{\lambda} \varphi_\lambda^+ \geq \underline{\varphi^*}. \tag{6.4}$$

Let us argue by contradiction assuming that (6.4) is false, then there exists $x \in X$, $r > 0$, $\varepsilon > 0$, x_λ converging to x as $\lambda \to +\infty$ such that

$$\liminf_{\lambda} \varphi_\lambda^+(x_\lambda) \leq \inf_{B_r(x)} \varphi^* - 4\varepsilon,$$

To ease notations, choose a sequence $x_n = x_{\lambda_n}$ with $\lambda_n \to +\infty$ in such a way that

$$\liminf_{\lambda} \varphi_\lambda^+(x_\lambda) = \lim_n \int_Y \varphi(x_n, y) \widetilde{\mu}_{\lambda_n}^+(\mathrm{d}y|x_n).$$

and set $\varphi_n := \varphi(x_n, .)$, $\psi_n := \psi(x_n, .)$ and $\mu_n^+ := \widetilde{\mu}_{\lambda_n}^+(.|x_n)$. Since for n large enough $x_n \in B_r(x)$, we have for large enough n

$$\int_Y \varphi_n(y) \mu_n^+(\mathrm{d}y) \leq \varphi^*(x_n) - 3\varepsilon = \max_{\operatorname{argmin} \psi_n} \varphi_n - 3\varepsilon. \tag{6.5}$$

But it follows from (4.13)-(4.14) in remark 4.3 (actually applied to $u = \psi_n$ and $v = -\varphi_n$) that

$$\lim_n \mu_n^+(\{\varphi_n < \varphi^*(x_n) - \varepsilon\}) = 0$$

and since

$$\begin{aligned} \int_Y \varphi_n(y) \mu_n^+(\mathrm{d}y) \geq & -\|\varphi\|_\infty \mu_n^+(\{\varphi_n < \varphi^*(x_n) - \varepsilon\}) \\ & + (\varphi^*(x_n) - \varepsilon) \mu_n^+(\{\varphi_n \geq \varphi^*(x_n) - \varepsilon\}) \end{aligned}$$

for large enough n, we have

$$\int_Y \varphi_n(y) \mu_n^+(\mathrm{d}y) \geq \varphi^*(x_n) - 2\varepsilon$$

which contradicts (6.5) and therefore establishes (6.4). Finally, it is well-known (see Theorem 1.17 in [4]) that (6.3)-(6.4) imply that φ_λ Γ-converges to $\underline{\varphi^*}$ as $\lambda \to +\infty$.

□

7 Examples and remarks

If one wishes to implement our approximation on concrete problems (which, at the moment, we leave for future works), a certain number of issues have to be seriously addressed. Among them one can think of the choice of the reference measure ν as well as of the numerical method to efficiently compute the integrals which appear in φ_λ. But the first question that naturally comes to mind is the choice of the weight δ_λ for the secondary scale. Since δ_λ captures the tradeoff between the upper and lower level, a small δ_λ will result in a good accuracy for the follower's optimizing behavior but might too slowly take into account the leader's objective. We already saw that δ_λ cannot be chosen too small for the convergence to be guaranteed but choosing it too large may affect the speed of convergence to the value function of the leader. A universallly good choice for δ_λ is certainly impossible and the aim of this final section is precisely to illustrate on some particular examples the behavior of our approximations.

7.1 A case where δ_λ cannot be too small

We first consider an example where the assumption (3.5) of a power-like lower bound on α_ν is relaxed. This example shows that one cannot hope for a universal choice of δ_λ, in particular $\delta_\lambda = \lambda^{-\frac{1}{2}}$ does not guarantee convergence to the double argmin set if α_ν happens to be too small near 0.

Example 7.1. Consider the case where $Y = [-1, 1]$, and let u and v Lipschitz, nonnegative with

- $u = 0$ exactly on $K := [-3/4, -1/2]$ and at 0, $u(y) = |y|$ in a neighbourhood of 0 (for instance $u(y) = \min(|y|, d_K(y))$ where d_K is the distance to K),
- v vanishes at 0 only.

We thus have $\operatorname{argmin} u = K \cup \{0\}$ and the double argmin set $\operatorname{argmin}_{\operatorname{argmin} u} v$ is reduced to $\{0\}$. Now choose ν absolutely continuous with respect to the 1-dimensional Lebesgue measure $\mathrm{d}\nu = f \mathrm{d}y$ with f bounded, $f > 0$ Lebesgue

almost everywhere so that ν has full support but f extremely small in a neighbourhood of 0, more precisely:

$$f(y) = \frac{3}{y^4} e^{-\frac{1}{|y|^3}}, \text{ for } |y| > 0 \text{ small and } , f(0) = 0.$$

so that for small $\delta > 0$

$$\nu([-\delta, \delta]) = 2e^{-\frac{1}{\delta^3}}.$$

Consider now the two-scale measure with a square root weight for the secondary function v:

$$\widetilde{\nu}_\lambda := \widetilde{Z}_\lambda e^{-\lambda u - \sqrt{\lambda} v} \nu \text{ with } \widetilde{Z}_\lambda := \frac{1}{\int_Y e^{-\lambda u - \sqrt{\lambda} v} \nu}.$$

First observe that

$$\int_Y e^{-\lambda u - \sqrt{\lambda} v} \nu \geq \int_K f e^{-\sqrt{\lambda} v} \geq \Big(\int_K f \Big) e^{-\max_K v \sqrt{\lambda}} =: C e^{-M\sqrt{\lambda}}$$

then for $\delta > 0$ small we have

$$\widetilde{\nu}_\lambda([-\delta, \delta]) \leq \frac{2e^{M\sqrt{\lambda}}}{C} \int_0^\delta e^{-\lambda y} f$$

since

$$\begin{aligned} \int_0^\delta e^{-\lambda y} f &= \int_0^{\frac{\delta}{\lambda^{1/4}}} e^{-\lambda y} f + \int_{\frac{\delta}{\lambda^{1/4}}}^\delta e^{-\lambda y} f \\ &\leq e^{-\frac{\lambda^{3/4}}{\delta^3}} + e^{-\lambda^{3/4}\delta} \|f\|_\infty \end{aligned}$$

we finally get

$$\widetilde{\nu}_\lambda([-\delta, \delta]) \leq 2C^{-1} e^{\sqrt{\lambda} M} (e^{-\frac{\lambda^{3/4}}{\delta^3}} + e^{-\lambda^{3/4}\delta} \|f\|_\infty)$$

since the right hand-side goes to 0 as $\lambda \to +\infty$, we deduce that there exists a neighbourhood of 0 which has zero measure for any weak cluster point of $\widetilde{\nu}_\lambda$, in particular none of these cluster points can concentrate on $\{0\}$. In an example like this one, one typically has $\theta(t) \sim \frac{-1}{t^3}$, so applying the cooking recipe of Lemma 4.1 one finds $t_\lambda \sim \lambda^{-\frac{1}{4}}$ and weights δ_λ which guarantee convergence are $\delta_\lambda = \lambda^{-\gamma}$ with $\gamma \in (0, \frac{1}{4})$ or $\delta_\lambda = \lambda^{-\frac{1}{4}} \log(\lambda)$.

7.2 A case where δ_λ cannot be power-like

We now consider a variant of the previous example where any power choice for δ_λ fails to guarantee convergence of the two-scale measure to the double argmin set.

Example 7.2. Consider exactly the same u and v as in example 7.1, but with an even worse reference measure whose density f is of the form

$$f(y) = \frac{1}{y^2} \exp\Big(\frac{1}{|y|} - \exp(\frac{1}{|y|})\Big)$$

in a neighbourhood of 0 (and bounded and bounded away from zero elsewhere, say) so that for small $\delta > 0$, $\nu([-\delta,\delta]) = 2\exp(-\exp(\delta^{-1}))$ and $\alpha_\nu(r) \sim \exp(-\exp(\frac{1}{r}))$ for small r. Let us try to take a power function λ^γ (i.e. $\delta_\lambda = \lambda^{\gamma-1}$) for some $\gamma \in (0,1)$ as a weight on the function v i.e. consider

$$\widetilde{\nu}_\lambda := \widetilde{Z}_\lambda e^{-\lambda u - \lambda^\gamma v}\nu \text{ with } \widetilde{Z}_\lambda := \frac{1}{\int_Y e^{-\lambda u - \lambda^\gamma v}\nu}.$$

As in example 7.1, we have

$$\int_Y e^{-\lambda u - \lambda^\gamma v}\nu \geq Ce^{-M\lambda^\gamma}$$

hence

$$\widetilde{\nu}_\lambda([-\delta,\delta]) \leq \frac{2}{C}e^{M\lambda^\gamma}\int_0^\delta e^{-\lambda y}f(y)dy. \tag{7.1}$$

For any $\sigma > 1$ we have

$$\begin{aligned}\int_0^\delta e^{-\lambda y}f &= \int_0^{\frac{\delta}{\sigma}} e^{-\lambda y}f + \int_{\frac{\delta}{\sigma}}^\delta e^{-\lambda y}f \\ &\leq \exp(-\exp(\frac{\sigma}{\delta})) + \|f\|_\infty e^{-\frac{\lambda\delta}{\sigma}}\end{aligned}$$

choosing $\sigma = \lambda^{\frac{1-\gamma}{2}}$ and replacing in (7.1), we arrive at

$$\widetilde{\nu}_\lambda([-\delta,\delta]) \leq \frac{2}{C}e^{M\lambda^\gamma}\Big(\exp(-\exp(\delta^{-1}\lambda^{\frac{1-\gamma}{2}})) + \|f\|_\infty \exp(-\delta\lambda^{\frac{1+\gamma}{2}})\Big)$$

and since $\gamma < \frac{1+\gamma}{2}$, we reach the conclusion that $\widetilde{\nu}_\lambda([-\delta,\delta])$ tends to 0 as $\lambda \to \infty$, ruling out the convergence of $\widetilde{\nu}_\lambda$ to the Dirac mass at 0 (recall that $\{0\}$ is the double argmin set in this example). In other words, no power like

secondary weight gives convergence. But using Lemma 4.1 and Remark 4.4 the (very slowly decaying!) weight

$$\delta_\lambda = \frac{\log(\log(\log(\lambda)))}{\log(\log(\lambda))}$$

ensures the desired convergence. Of course, one may think that it is crazy to use such a pathological reference measure, but a rough behavior of u, v or the boundary of the set Y in higher dimensions may generate similar pathologies as well.

7.3 A case where the pessimistic value is not lsc

The following explicit example illustrates the convergence of the approximations in a case where the pessimistic value is not lsc.

Example 7.3. Let us consider the case $X = Y = [0,1]$, take ν to be the Lebesgue measure on Y and

$$\varphi(x,y) = x + y, \ \psi(x,y) = xy,$$

then

$$R(x) = \operatorname{argmin}_{y\in Y} xy = \begin{cases} 0 & \text{if } x \in (0,1] \\ [0,1] & \text{if } x = 0. \end{cases}$$

Since φ and ψ are Lipschitz and ν is the Lebesgue measure any choice of δ_λ of the form $\delta_\lambda = \lambda^{-\gamma}$ with $\gamma \in (0,1)$ ensures the validity of our convergence result.

In the optimistic case, the leader minimizes the optimistic value

$$\varphi_*(x) = \min_{y\in R(x)} x + y \ = x$$

and the solution is 0.

Now, the approximation scheme proposed in the paper can be explicitly computed. For a given $\lambda > 0$, consider $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$,

$$\widetilde{\mu}_\lambda = \widetilde{Z}_\lambda(x) e^{-\lambda xy - \sqrt{\lambda}(x+y)} dy,$$

where

$$\widetilde{Z}_\lambda(x) = \frac{1}{\int_0^1 e^{-\lambda xy - \sqrt{\lambda}(x+y)} dy} = e^{\sqrt{\lambda}x} \frac{\lambda x + \sqrt{\lambda}}{[1 - e^{-(\lambda x + \sqrt{\lambda})}]}$$

then

$$\varphi_\lambda(x) = \tilde{Z}_\lambda(x) \int_0^1 (x+y) e^{-\lambda xy - \sqrt{\lambda}(x+y)} dy = x + \frac{e^{-\lambda x - \sqrt{\lambda}}}{e^{-\lambda x - \sqrt{\lambda}} - 1} + \frac{1}{\lambda x + \sqrt{\lambda}}$$

converges to $\varphi_*(x) = x$ as $\lambda \to +\infty$.
In the pessimistic case, the leader minimizes the pessimistic value

$$\varphi^*(x) = \max_{y \in R(x)} x + y \ = \begin{cases} 1 & \text{if } x = 0 \\ x & \text{if } x > 0 \end{cases}$$

which is not lsc at 0 and the infimum of φ^* which is 0 is not achieved. Note that the lsc envelope of φ^* coincides with the optimistic value φ_*.

In this case, for a given $\lambda > 0$, consider $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$,

$$\widetilde{\mu}_\lambda^+ = \widetilde{Z}_\lambda^+(x) e^{-\lambda xy + \sqrt{\lambda}(x+y)} dy$$

where

$$\widetilde{Z}_\lambda^+(x) = \frac{1}{\int_0^1 e^{-\lambda xy + \sqrt{\lambda}(x+y)} dy} = \begin{cases} e^{\sqrt{\lambda} x \frac{[e^{(-\lambda x + \sqrt{\lambda})} - 1]}{-\lambda x + \sqrt{\lambda}}} & \text{if } x \neq \frac{1}{\sqrt{\lambda}} \\ \frac{1}{e} & \text{if } x = \frac{1}{\sqrt{\lambda}} \end{cases}$$

then

$$\varphi_\lambda^+(x) = \tilde{Z}_\lambda(x) \int_0^1 (x+y) e^{-\lambda xy + \sqrt{\lambda}(x+y)} dy = \begin{cases} x + \frac{e^{-\lambda x + \sqrt{\lambda}}}{e^{-\lambda x + \sqrt{\lambda}} - 1} - \frac{1}{-\lambda x + \sqrt{\lambda}} & \text{if } x \neq \frac{1}{\sqrt{\lambda}} \\ \frac{1}{\sqrt{\lambda}} + \frac{1}{2} & \text{if } x = \frac{1}{\sqrt{\lambda}}. \end{cases}$$

We know that φ_λ^+ pointwise converges to φ^* and Γ-converges to $\underline{\varphi}^*$ as $\lambda \to +\infty$. The pointwise convergence is of course slower near 0 and we have tested various exponents for δ_λ (the square root as above but also $\gamma = \frac{1}{4}$, for which the convergence is even slower and $\gamma = \frac{3}{4}$ which seems to give more accurate approximations).

7.4 The choice of δ_λ is critical in practice

Even in the case where argmin $\psi(x, .)$ is a singleton for any $x \in X$, so that optimistic and pessimistic solutions coincide, the optimistic and pessimistic λ-approximations converge, with a different convergence speeed, to the common value $\varphi_* = \varphi^*$, as shown in the next example in which X is two-dimensional and for which the choice of δ_λ seems to be crucial for practical convergence.

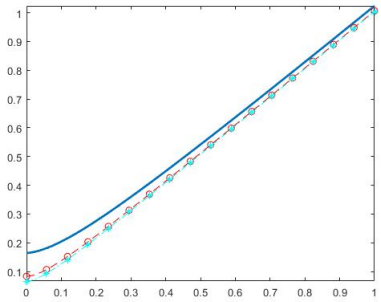


Figure 1: Graph of φ_λ for $\lambda = 36$, $\lambda = 144$ (circle), $\lambda = 225$ (square) with $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ (Example 7.3).

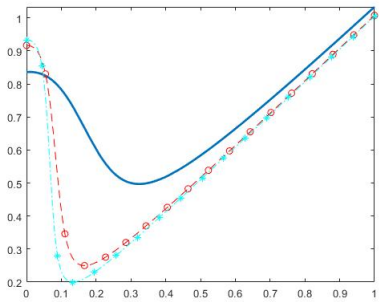


Figure 2: Graph of φ_λ^+ for $\lambda = 36$, $\lambda = 144$ (circle), $\lambda = 225$ (square) with $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ (Example 7.3).

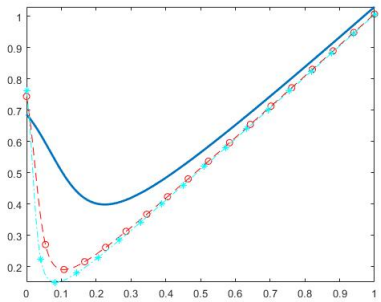


Figure 3: Graph of φ_λ^+ for $\lambda = 36$, $\lambda = 144$ (circle), $\lambda = 225$ (square) with $\delta_\lambda = \frac{1}{\sqrt[4]{\lambda^3}}$ (Example 7.3).

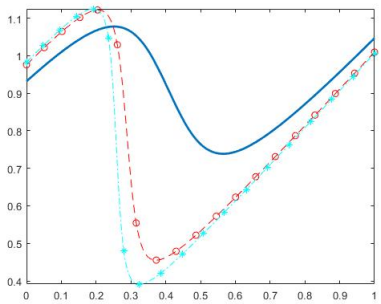


Figure 4: Graph of φ_λ^+ for $\lambda = 36$, $\lambda = 144$ (circle), $\lambda = 225$ (square) with $\delta_\lambda = \frac{1}{\sqrt[4]{\lambda}}$ (Example 7.3).

Example 7.4. Consider the case where

$$X = [0,1]^2, \ Y = [0,1],$$

and equip Y with the Lebesgue measure as reference measure. Consider the objectives and $\varphi, \psi : X \times Y \to \mathbb{R}$ defined by

$$\varphi(x,y) = x_1^2 + 4x_2 y - x_1 x_2$$

$$\psi(x,y) = \frac{y^2}{2} - y\frac{x_1 + x_2}{4} + x_1 - x_2.$$

For a given $x = (x_1, x_2)$, the follower solves the problem

$$\min_{y \in [0,1]} \frac{y^2}{2} - y\frac{x_1 + x_2}{4} + x_1 - x_2$$

whose unique solution is

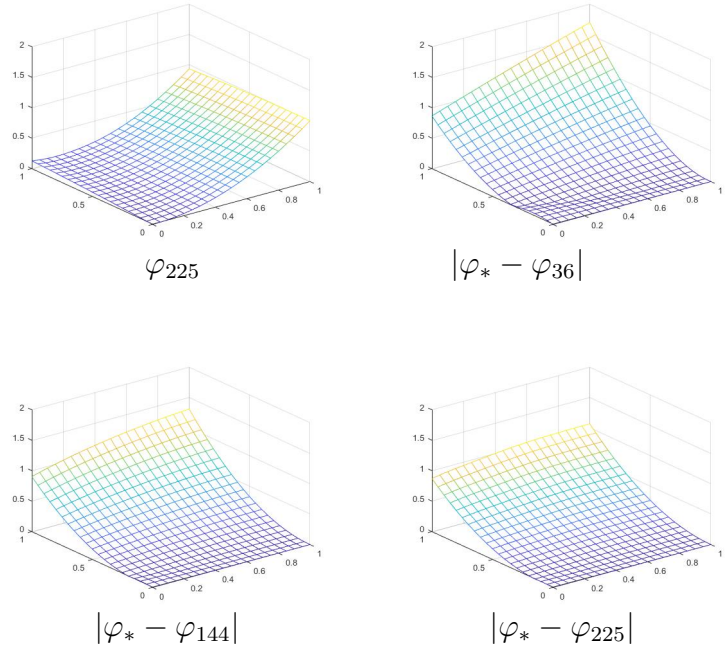
$$R(x) = \frac{x_1 + x_2}{4}$$

and the leader solves the problem $\min_{x \in [0,1]^2} \varphi(x, R(x))$ where

$$\varphi_*(x) = \varphi^*(x) = \varphi(x, R(x)) = x_1^2 + x_2^2.$$

The Stackelberg solution is $\overline{x} = (0,0)$ and $\overline{y} = 0$. Since φ and ψ are Lipschitz we can chose any power function for δ_λ, $\delta_\lambda = \lambda^{-\gamma}$. Our approximation is given by:

$$\varphi_\lambda(x) = \frac{\int_Y \varphi(x,y) e^{-\lambda\psi(x,y) - \lambda^{1-\gamma}\varphi(x,y)} dy}{\int_Y e^{-\lambda\psi(x,y) - \lambda^{1-\gamma}\varphi(x,y)} dy}$$

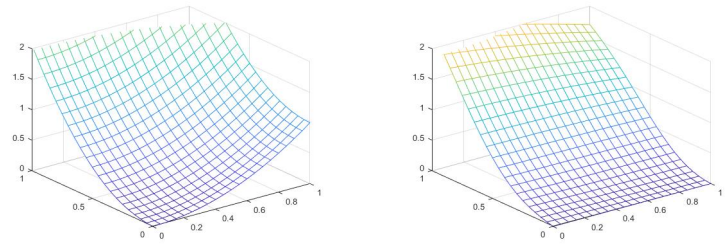


φ_{225} $\quad$ $|\varphi_* - \varphi_{36}|$

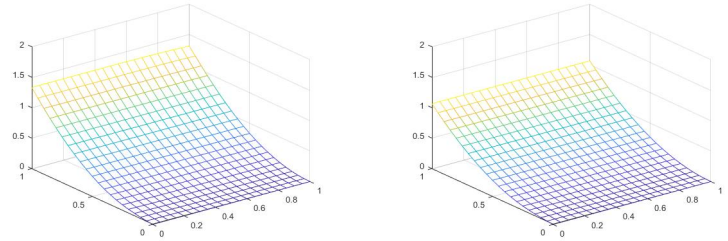
$|\varphi_* - \varphi_{144}|$ $\quad$ $|\varphi_* - \varphi_{225}|$

Figure 5: Optimistic case for $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ (Example 7.4).

(and the pessimistic approximation φ_λ^+ is given by a smilar formula, just by changing the sign of the term involving φ) which converges as $\lambda \to +\infty$ to φ^* which achieves its minimum at $(0, 0)$. We illustrate the convergence with various exponents and values of λ. The convergence turns out to be very bad for $\gamma = \frac{1}{2}$ but very good for γ close to 1 as in the case $\gamma = \frac{9}{10}$.

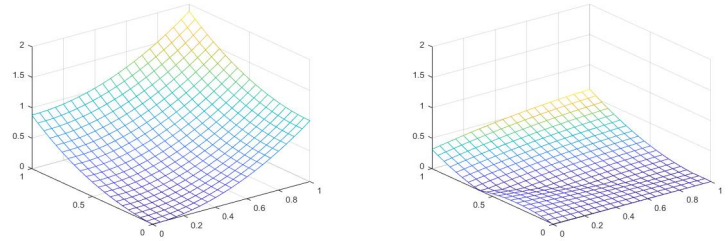


φ_{225}^+ $\quad$ $|\varphi^* - \varphi_{36}^+|$



$|\varphi^* - \varphi^+_{144}|$ $\qquad$ $|\varphi^* - \varphi^+_{225}|$

Figure 6: Pessimistic case for $\delta_\lambda = \frac{1}{\sqrt{\lambda}}$ (Example 7.4).



φ_{225} $\qquad$ $|\varphi_* - \varphi_{36}|$

References

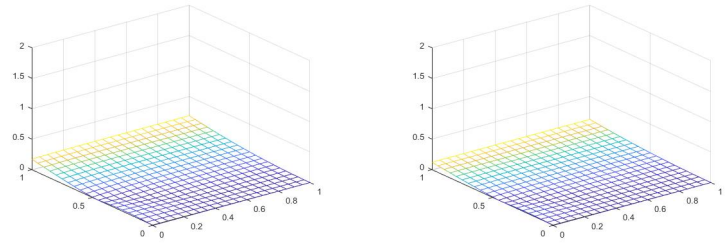
[1] A. Aboussoror and P. Loridan. Strong-weak Stackelberg problems in finite-dimensional spaces. *Serdica Math. J.*, 21(2):151–170, 1995.

[2] T. Başar and G. J. Olsder. *Dynamic noncooperative game theory.* Classics in Applied Mathematics, vol. 23. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. Reprint of the second edition (1995).

[3] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

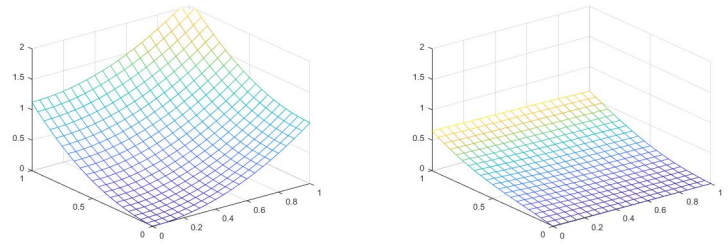
[4] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, Oxford, 2002.

[5] M. Breton, A. Alj, and A. Haurie. Sequential Stackelberg equilibria in two-person games. *J. Optim. Theory Appl.*, 59(1):71–97, 1988.

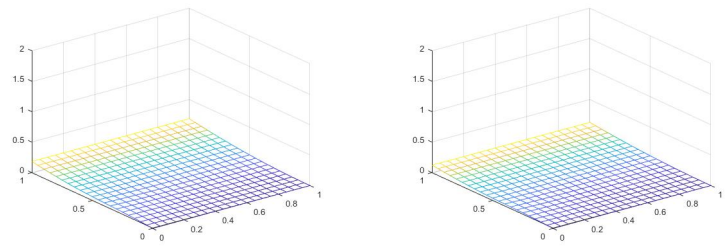


$|\varphi_* - \varphi_{144}|$ $\qquad$ $|\varphi_* - \varphi_{225}|$

Figure 7: Optimistic case for $\delta_\lambda = \frac{1}{\sqrt[10]{\lambda^9}}$ (Example 7.4).

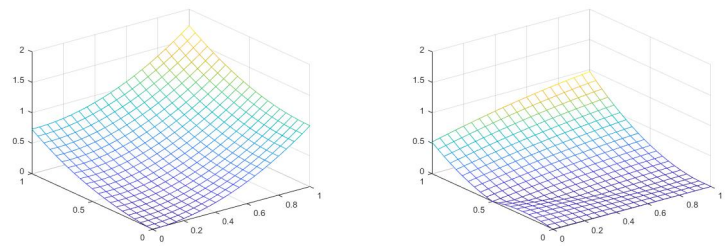


φ^+_{225} $\qquad$ $|\varphi^* - \varphi^+_{36}|$

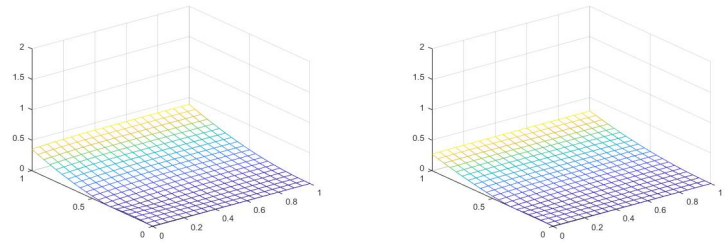


$|\varphi^* - \varphi^+_{144}|$ $\qquad$ $|\varphi^* - \varphi^+_{225}|$

Figure 8: Pessimistic case for $\delta_\lambda = \frac{1}{\sqrt[10]{\lambda^9}}$ (Example 7.4).

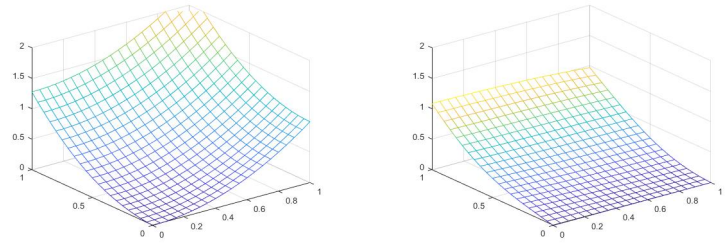


φ_{225} $\qquad$ $|\varphi_* - \varphi_{36}|$

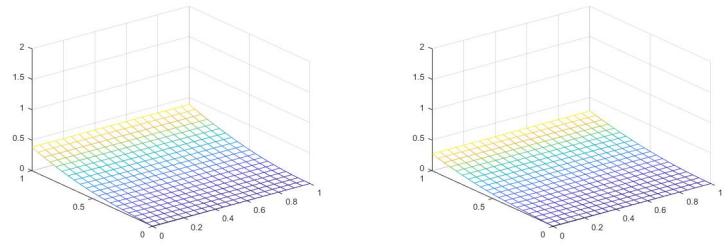


$|\varphi_* - \varphi_{144}|$ $|\varphi_* - \varphi_{225}|$

Figure 9: Optimistic case for $\delta_\lambda = \frac{1}{\sqrt[4]{\lambda^3}}$ (Example 7.4).



φ^+_{225} $|\varphi^* - \varphi^+_{36}|$



$|\varphi^* - \varphi^+_{144}|$ $|\varphi^* - \varphi^+_{225}|$

Figure 10: Pessimistic case for $\delta_\lambda = \frac{1}{\sqrt[4]{\lambda^3}}$ (Example 7.4).

[6] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Ann. Oper. Res.*, 153:235–256, 2007.

[7] Gianni Dal Maso. *An introduction to* Γ*-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1993.

[8] S. Dempe. Bilevel optimization: theory, algorithms and applications. Preprint 2018-11, 2018.

[9] C.-R. Hwang. Laplace's method revisited: weak convergence of probability measures. *Ann. Probab.*, 8(6):1177–1182, 1980.

[10] G. Leitmann. On generalized Stackelberg strategies. *J. Optim. Theory Appl.*, 26(4):637–643, 1978.

[11] L. Mallozzi and J. Morgan. Oligopolistic markets with leadership and demand functions possibly discontinuous. *J. Optim. Theory Appl.*, 125(2):393–407, 2005.

[12] H. von Stackelberg. *Marktform und Gleichgewicht*. J. Springer, 1934.