Structure Preserving Approximation of Semiconcave Functions

Karl Kunisch[1,2]
karl.kunisch@uni-graz.at

Donato Vásquez-Varas[2,3,4]
dvasquez@dim.uchile.cl, donato.vasquez-varas@ricam.oeaw.ac.at

[1]Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria
[2]Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria
[3]Department of Mathematical Engineering, University of Chile, Beauchef 851, 8370458 Santiago, RM, Chile
[4]Center for Mathematical Modelling, Beauchef 851, 8370456 Santiago, RM, Chile

February 8, 2026

Abstract

This article addresses structure-preserving smooth approximation of semiconcave functions. semiconcave functions are of particular interest because they naturally arise in a variety of variational problems, including optimal feedback control, game theory, and optimal transport. We leverage the fact that any semiconcave function can be represented as the infimum of a countable family of C^2 functions. This infimum is expressed in a form that allows approximation by finitely many functions, combined with smoothing operations, such that each element of the approximating sequence remains semiconcave. The active sets of indices contributing to the representation of the semiconcave function and its approximations are analyzed in detail. Moreover, we show that the gradients of the elements in the expansion of the approximating functions form a probability distribution, a property of particular interest for the value function in optimal control. Approximation results are established in $C(\bar{\Omega})$ and in $W^{1,p}(\Omega)$ for $p \in [1, \infty)$ and $p = \infty$. Finally, numerical results are presented to illustrate the approach on a test example.

Keywords: semiconcavity, approximation of semiconcave functions, exponential distance function, feedback synthesis.

AMS classification: 26B25, 49J52, 41A30, 49L25.

1 Introduction

This work is concerned with the approximation of semiconcave functions. Here we say that for a convex, open, and bounded set $\Omega \subset \mathbb{R}^d$ and $C > 0$ a function $v \in C(\overline{\Omega})$ is $C-$semiconcave if the mapping

$$x \mapsto v(x) - \frac{C}{2}|x|^2$$

is concave. We are particularly interested in approximations schemes which preserve the semiconcavity structure and in their asymptotic limits. Our main motivation for studying these type of functions lies in their connection with control theory. Indeed, under appropriate hypotheses, the value function of an optimal control problem is semiconcave (see [6] and [3]). In the same direction, the semiconcavity is a natural regularity property for certain Hamilton-Jacobi-Bellman equations (see [9, 11, 18]), since it provides uniqueness and it is connected with the developed of maxplus algebra methods (see [1, 10, 13]). Further, Lyapunov semiconcave functions can be used to design discontinuous feedback laws (see [25, 26]) for problems that do not admit smooth ones. In [20] the existence of a sequence of smooth consistent feedback laws is proved provided that the value functions of the underlying control problem is semiconcave, whereas in [19] this result is used to prove the convergence of a machine learning scheme for the synthesis of optimal feedback-laws. Besides optimal control, semiconcave functions also arise in optimal transport [27], where they are crucial for the regularity and stability of transport maps, and in the analysis of free boundary value problems, where they help to characterize the structure of the free boundary [14]. Let us also mention that in the literature also the terminology 'weakly-convex' is used, which would, in analogy to the present setting, be referred to as semiconvex. We mention, for instance, [16] where weakly-convex functions are learned by means of neural networks and used as sparsity promoting regularizers in image reconstruction problems.

In this paper we develop a method for approximating semiconcave functions with the property that all elements of the approximating family preserve semiconcavity. In order to achieve this, we design a parametrization which corresponds to a regularized version of the minimum of an array of real numbers applied to a family of smooth parametrized functions. Additionally, by means of analyzing the gradients of the regularized functions, we identify a family of sets in which the convergence of the gradients of the approximation is uniform. Remarkably, this family covers part of the discontinuities of the approximated function. As we will explain, this is particularly relevant for the approximation of solutions to Hamilton Jacobi Bellman equations.

To the best of our knowledge, semiconcave preserving parametrizations have not been addressed yet. However, there are many contributions to the approximation of convex functions. We point out that many of them use as starting point the fact that convex functions are the maximum of a family of affine functions. In [4], they are parametrized by the maximum of a finite number of affine functions, which are learned by an alternating optimization method. The convergence of this approach together with an initilization method can be found in [15] under further assumptions on the initial guess of the parameters. Concerning neural network approaches, Input Convex Neural Networks (ICNN) are introduced in [2]. They consist in a fully connected neural network with non-negative weights and increasing convex activation function. In the case of ReLU activation these types of neural networks are universal approximators of globally Lipschitz continuous and convex functions, see[17]. This approach is extended in [7] by Input Convex Recurrent Neural Network (ICRNN), where a recurrent neural network is considered. The authors used this to estimate the dynamics of a control process in a manner that the resulting approximated control problem is convex. In [28] the GroupMax Neural Network architecture is proposed and its universal approximation property with respect to convex functions is proved. Similarly to ICNN, GroupMax Neural Networks utilize increasing convex activation functions together with positive weights. The main difference between them is that GroupMax Neural Networks consider group maximization layers. In the previous cases the approximation is convex and Lipschitz, but it is not smooth. In contrast, in [5] Log-sum-exp neural networks are presented which are based on using the Log-Sum-Exp function to approximate the maximum of a family of affine functions. This architecture is smooth and it enjoys the universal approximation property for convex functions. A closer connection between semiconcavity and neural networks can be found in [8], where a neural network architecture based on the min-plus algebras is proposed. The approximation is constructed as the minimum of a family of functions which are quadratic in the space variables and also time dependent. This corresponds to a semiconcave approximation of the value function of an optimal control problem.

The organization of the article is as follow. In Section 2 the semiconcavity preserving parametrization is introduced. In Section 3 a probabilistic interpretation connected with the parametrization is introduced which helps to analyze the behavior of the gradient of the approximation. Section (4) is devoted to investigate the convergence of the approximation. Additionally, in Section 5 a semi-concave test function is proposed for which the results of the proposed parametrization are illustrated and compared towards their numerical performance. Finally, in Section 6 the main results and conclusions of this work are summarized.

Before continuing, we give some notation. For a vector $x \in \mathbb{R}^d$ we denote its Euclidean norm by $|x|$, its $p-$norm by $|x|_p = \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$, for $p \in [1, \infty)$, and the supremum norm by $|x|_\infty = \max_{i\in\{1,\dots,d\}} |x_i|$. For $A \subset \mathbb{R}^d$ and $B \subset X$, with X a normed vector space equipped with a norm $\|\cdot\|_X$, we denote the space of continuous functions from A to B by $C(A; B)$ equipped with the norm

$$\|v\|_{C(A;B)} = \sup_{x\in A} \|v(x)\|_X ,$$

where $v \in C(A; B)$. If A is open or it is the closure of an open set which Lipschitz boundary, for $m \in \mathbb{N}$ we denote the space of functions form A to B which are m times continuously differentiable by $C^m(A; B)$.

For $A \subset \mathbb{R}^d$ measurable, B a measurable space, and $p \in [1, \infty)$ we denote the space of $p-$integrable functions from A to B by $L^p(A; B)$ and the space of essentially bounded functions from A to B by $L^\infty(A; B)$ equipped with their usual norms. For $n, m \in \mathbb{N}$ we denote the space of matrices of dimension $n \times m$ by $\mathbb{R}^{n\times m}$ equipped with the operator norm which we denote by $|A| = \sup_{x\in\mathbb{R}^m} \frac{|Ax|}{|x|}$ for $A \in \mathbb{R}^{n\times m}$.

2 Semiconcavity preserving parametrization

In this section we describe the architecture that we propose to approximate semiconcave functions while preserving this semiconcavity property. The approximation is based on the fact that a function is semiconcave if and only if it is the poinwise minimum over a family of C^2 functions with uniformly bounded second derivatives. Hence, to approximate semiconcave functions we consider a smooth approximation of the minimum together with a parameterized family of $C^2(\overline{\Omega})$ functions. To describe the parametrization in more detail we need some concepts and results that we describe in the following. Throughout we consider Ω to be an open, bounded, and convex set in $\mathbb{R}^d$.

We start by recalling the definition of a semiconcave function.

Definition 2.1. *Let $v : \Omega \mapsto \mathbb{R}$ be a continuous function and $C > 0$. We say that v is $C-$semiconcave if $x \in \Omega \mapsto v(x) - C|x|^2$ is concave.*

According to Proposition 1.1.3 in [6], a semiconcave function can be characterized as the minimum over a family of C^2 function as is stated in the following result. Here we present a slightly modified version, assuming that v is Lipschitz continuous on $\bar{\Omega}$, rather than on Ω, which is more convenient for our needs.

Theorem 2.1. *Let $v \in C(\overline{\Omega})$. Then v is Lipschitz on $\overline{\Omega}$ and semiconcave with constant $C > 0$ if and only if there exists a family of functions $\{\phi_i\}_{i\in\mathcal{I}} \subset C^2(\overline{\Omega})$ uniformly bounded in $C^2(\overline{\Omega})$ such that*

$$\sup_{i\in\mathcal{I}} \left\|\nabla^2\phi_i\right\|_{C(\overline{\Omega};\mathbb{R}^{d\times d})} \leqslant C,$$

and

$$v(x) = \inf_{i\in\mathcal{I}} \phi_i(x) \text{ for all } x \in \overline{\Omega}, \tag{2.1}$$

where $\mathcal{I}$ is an uncountable index set.

Proof of Theorem 2.1. Let us first suppose that v satisfies the specified properties and prove the existence of a family of functions $\{\phi_i\}_{i\in\mathcal{I}} \subset C^2(\overline{\Omega})$ satisfying the conclusions of the theorem. In the following we denote by L the Lipschitz constant of v.

Due to the semiconcavity of v and the convexity of Ω, we have by Proposition 3.3.1 in [6] that for all $x, y \in \Omega$,

$$p \in D^+v(y) \text{ if and only if } v(x) \leqslant v(y) + p(x-y) + \frac{C}{2}|x-y|^2, \tag{2.2}$$

where $D^+v(y)$ is the upper-differential of v at y, which is defined by

$$D^+v(y) := \left\{ p \in \mathbb{R}^d : \limsup_{x\to y} \frac{v(x) - v(y) - p\cdot(x-y)}{|y-x|} \leqslant 0 \right\}.$$

according to Definition 3.1.1 in [6]. We note that $D^+v(x)$ is not empty for all $x \in \Omega$ since v is semiconcave (see Theorem 3.3.4 (b) in [6, Chapter 3]).

By continuity this inequality also holds for $x \in \bar{\Omega}$. We will construct the family $\{\phi_i\}_{i\in\mathcal{I}}$ by using (2.2). For this purpose we verify that for every $y \in \Omega$ the norm of the elements in $D^+v(y)$ is bounded by L.

Let us consider $y \in \Omega$ and $p \in D^+v(y)$. Without loss of generality we assume $p \neq 0$. Let us set $x = y - \frac{\varepsilon}{|p|}p$. Since Ω is open, for ε small enough we have that $x \in \Omega$. Then for such ϵ we have by means of (2.2) that

$$v\left(y - \frac{\varepsilon}{|p|}p\right) \leqslant v(y) - \varepsilon|p| + \frac{C}{2}\varepsilon^2.$$

Rearranging the terms in the previous inequality and using the Lipschitz continuity of v we obtain

$$\varepsilon|p| \leqslant L\varepsilon + \frac{C}{2}\varepsilon^2.$$

Dividing both sides of the inequality by ε and letting $\varepsilon \to 0^+$ we obtain that $|p| \leqslant L$, which proves the claim.

For $y \in \partial\Omega$ it is not true that all the elements $D^+v(y)$ are bounded. However, arguing by continuity and using (2.2) it is possible to prove that there exists at least one element satisfying (2.2) which is bounded by L. This observation proves that for every $y \in \overline{\Omega}$ there exists $p_y \in \mathbb{R}^d$ with $|p_y| \leqslant L$ such that

$$v(x) \leqslant v(y) + p_y(x-y) + \frac{C}{2}|y-x|^2.$$

Then, observing that the previous inequality is attained for $x = y$, setting $\mathcal{I} = \overline{\Omega}$, and

$$\phi_y(x) = v(y) + p_y(x-y) + \frac{C}{2}|y-x|^2 \text{ for y} \in \overline{\Omega},$$

we obtain (2.1), and the asserted uniform bounds on $\{\phi_y\}_{y\in\bar{\Omega}}$.

The reciprocal implication is given by Corollary 2.1.6 in [6] and the fact that the infimum over a family of Lipschitz functions is Lipschitz as well.

□

We next assert that the index set $\mathcal{I}$ in the previous theorem can be replaced by a countable one.

Proposition 2.1. *Let $v \in C(\overline{\Omega})$ be a semiconcave function with constant $C > 0$ and Lipschitz continuous in $\overline{\Omega}$ with constant $L > 0$. Let $\{\phi_i\}_{i\in\mathcal{I}}$ denote the family appearing in Theorem 2.1. Then there exists $\{\tilde{\phi}_i\}_{i=1}^{\infty} \subset \{\phi_i\}_{i\in\mathcal{I}}$ such that*

$$v_n(x) = \min_{i=1,\dots,n} \tilde{\phi}_i(x)$$

satisfies

$$\lim_{n\to\infty} \|v_n - v\|_{C(\overline{\Omega})} + \|\nabla v_n - \nabla v\|_{W^{1,p}(\Omega)} = 0, \tag{2.3}$$

for all $p \in [1, \infty)$, and

$$\lim_{n\to\infty} \nabla v_n(x) = \nabla v(x) \text{ a.e. in } \Omega. \tag{2.4}$$

Further, for each compact subset K of Ω, such that $D^+v(x)$ is a singleton for all $x \in K$, we have

$$\lim_{n\to\infty} v_n = v \text{ in } C^1(K). \tag{2.5}$$

Proof of Proposition 2.1. For $n \in \mathbb{N}$ let us consider a family of rectangles $\{\tau_{i,h}\}_{i=1}^{N_n}$ such that $diam(\tau_{i,n}) \leqslant \frac{1}{n}$, where $diam(\cdot)$ denotes the diameter of $\tau_{i,n}$, and

$$\overline{\Omega} \subset \bigcup_{i=1}^{N_n} \tau_{i,n} \text{ and } \overline{\Omega} \cap \tau_{i,n} \neq \emptyset \text{ for all } i = 1, \dots, N_n.$$

For each $i = 1, \dots, N_n$ we choose $x_{i,n} \in \tau_{i,n} \cap \overline{\Omega}$. By the definition of infimum, for each $x_{i,n}$ there exists $j(i,n)$ such that $\phi_{j(i,n)}(x_{i,n}) \leqslant v(x_{i,n}) + \frac{1}{n}$. This combined with the Lipschitz continuity of v on $\overline{\Omega}$ and the uniform boundedness of $\{\phi_i\}_{i\in\mathcal{I}}$ in $C^1(\overline{\Omega})$ we have for each $i = 1, \dots, N_n$

$$|\phi_{j(i,n)}(x) - v(x)| \leqslant \frac{(2L+1)}{n}, \text{ for all } x \in \tau_{i,n}. \tag{2.6}$$

Let us denote by $\tilde{\mathcal{I}}_n$ the finite set composed by the indices $j(i,n)$ defined above. Further set $\mathcal{I}_n = \bigcup_{j\leqslant n} \tilde{\mathcal{I}}_j$ and $\mathcal{I}_\infty = \cup_{n\in\mathbb{N}} \mathcal{I}_n$, and define $v_n \in C(\overline{\Omega})$ as follows

$$v_n(x) = \min_{i\in\mathcal{I}_n} \phi_i(x).$$

Observe that v_n is semiconcave. Further from (2.6) we deduce that

$$|v_n(x) - v(x)| \leqslant \frac{(2L+1)}{n} \text{ for all } x \in \overline{\Omega},$$

from which the uniform convergence of v_n to v follows. Almost everywhere convergence of the gradients as claimed in (2.4) follows from [6, Theorem 3.3.3].

The convergence of v_n in $W^{1,p}(\Omega)$ for all $p \in [1, \infty)$ is a simple consequence of the a.e. convergence of ∇v_n and the dominated convergence theorem. Consequently (2.3) is verified.

We now prove the last part of the result and assume that $Dv^+(x)$ is a singleton for all $x \in \bar{K}$. Consequently v is differentiable for all $v \in \bar{\Omega}$. Let us denote by $B \subset \Omega$ the set where the gradients v_n exists for all n, and observe that $|\Omega \setminus B| = 0$, where $|\cdot|$ denotes the measure of a set. Proceeding by contradiction let us assume that v_n does not converge to v in $W^{1,\infty}(K)$. In this case, there exists $\varepsilon > 0$, a sequence $x_k \in B$, and a sub-sequence of v_n denoted by $v_{n(k)}$ such that

$$|\nabla v_{n(k)}(x_k) - \nabla v(x_k)| \geqslant \varepsilon. \tag{2.7}$$

Since K is compact and $\{v_n\}_n$ is uniformly bounded in $W^{1,\infty}(\Omega)$, there exists $\bar{x}$ in $\bar{K}$ and $p \in \mathbb{R}^d$ such that x_k converges to $\bar{x}$ and $\nabla v_{n(k)}$ converges to p. By the semiconcavity of v_n is easy to verify that

$$v(y) - v(\bar{x}) - p \cdot (y - \bar{x}) \leqslant C|\bar{x} - y|^2, \text{ for all } y \in \Omega.$$

This implies that $p \in D^+v(\bar{x})$ and by assumption $p = \nabla v(\bar{x})$. However, by (2.7) we have $|p - \nabla v(\bar{x})| \geqslant \epsilon$. This gives the desired contradiction, and $\lim_{n\to\infty} v_n = v$ in $C^1(K)$ follows. $\square$

Our parametrization will build on this result. It consists in replacing the infinite family $\{\tilde{\phi}_i\}_{i=1}^{\infty}$ by a family of n parameterized functions and it utilizes a smooth approximation of the minimum operation ψ_n of n real numbers. We first describe the construction of the smooth approximation $\psi_{n,\varepsilon}$, $\varepsilon > 0$, and its properties. Then, relying on these properties, we introduce the family of parameterized functions together with the usage of $\psi_{n,\varepsilon}$, see (2.27) below.

For $n \leqslant m \in \mathbb{N} \setminus \{0\}$ and $a \in \mathbb{R}^m$ let us denote by $\psi_n(a)$ the minimum over a, that is,

$$\psi_n(a) = \min_{i \in \{1,\dots,n\}} a_i. \tag{2.8}$$

Thus, if $n < m$ then $\psi_n(a)$ only considers the first n elements of a. We also observe that ψ_n is a $1-$Lipschitz continuous function on $\mathbb{R}^m$ endowed with the maximum norm, but is not C^1.

For obtaining a C^1 approximation of ψ_n, we note that it can be written in a recursive manner, namely, for $i \in \{1, \dots, n-1\}$ we have the following relation

$$\psi_{i+1}(a) = \min\{a_{i+1}, \psi_i(a)\}, \text{ with } \psi_1 = a_1. \tag{2.9}$$

Further, for the case of the minimum of only two elements $x, y \in \mathbb{R}$ we have

$$\min(x, y) = x - (x - y)_+. \tag{2.10}$$

where $(\cdot)_+$ stands for the positive part function. Combining (2.9) and (2.10), we can evaluate ψ_n by the following recursive formula

$$\psi_{i+1}(a) = a_{i+1} - (a_{i+1} - \psi_i(a))_+, \ \psi_1 = a_1, \ i \in \{1, \dots, n-1\}. \tag{2.11}$$

It is noteworthy that an equivalent formula holds for the case of the maximum of a vector and this allowed the authors of [17] and [7] to prove the universal approximation property of ICNN and ICRNN. In our case we will use this to provide a smooth approximation of the minimum.

We shall employ a smooth regularization $g_\varepsilon \in C^{1,1}(\mathbb{R})$, $\varepsilon > 0$, of the positive part function $(\cdot)_+$, which satisfies the following properties

$$g_\varepsilon(x) \geqslant 0 \text{ for all } x \in \mathbb{R}, \tag{2.12a}$$

$$g'_\varepsilon(x) \in [0, 1] \text{ for all } x \in \mathbb{R}, \tag{2.12b}$$

$$g''_\varepsilon(x) \geqslant 0 \text{ for all almost all } x \in \mathbb{R}, \tag{2.12c}$$

$$\|(\cdot)_+ - g_\varepsilon\|_{C(\mathbb{R})} \leqslant \varepsilon, \tag{2.12d}$$

$$\lim_{\varepsilon\to 0^+} g'_\varepsilon = \chi_{[0,\infty)} \text{ in } C_{loc}(\mathbb{R} \setminus \{0\}), \tag{2.12e}$$

$$g''_\varepsilon \in L^\infty(\mathbb{R}) \text{ for each } \varepsilon > 0. \tag{2.12f}$$

Remark 2.2. *We provide examples of functions g_ε which* (2.12)*. We shall return to them in the following section. First, we consider the Moreau envelope of the positive part function, namely,*

$$g_{\varepsilon,M}(s) = \min_{t\in\mathbb{R}} (t)_+ + \frac{1}{2\varepsilon}|t - s|^2 = \begin{cases} 0 & \text{if } s < 0 \\ \frac{1}{2\varepsilon}s^2 & \text{if } s \in [0, \varepsilon) \\ s - \frac{\varepsilon}{2} & \text{if } s \geqslant \varepsilon \end{cases}$$

It is not hard to see that $g_{\varepsilon,M}$ satisfies (2.12)*. Further, it satisfies that $g'_{\varepsilon,M}(0) = 0$ for all $\varepsilon > 0$ and $g_{\varepsilon,M} \in C^{1,1}(\mathbb{R})$, but it is not in $C^2(\mathbb{R})$.*

Another example is $g_{\varepsilon,A}$ defined as:

$$g_{\varepsilon,A}(s) = \frac{1}{2}\left(s + \sqrt{s^2+\varepsilon^2} - \varepsilon\right).$$

To see that this function satisfies (2.12) *we note that $(s)_+ = \frac{1}{2}(|s|+s)$ and that $\sqrt{s^2+\varepsilon^2}-\varepsilon$ is a smooth approximation of the absolute value. Using these properties it is easy to see that $g_{\varepsilon,A}$ satisfies* (2.12)*.*

By replacing the positive part in (2.11) by g_ε we obtain a smooth approximation of ψ_n given by the following process:

$$\psi_{i+1,\varepsilon}(a) = a_{i+1} - g_\varepsilon(a_{i+1} - \psi_{i,\varepsilon}(a)),\ i \in \{1,\dots,n-1\},\ \psi_{1,\varepsilon} = a_1. \tag{2.13}$$

In Proposition 2.2 below, some useful properties of $\psi_{n,\varepsilon}$ are shown. These properties will permit us to construct a semiconcavity preserving parametrization.

Proposition 2.2. *Let $n \in \mathbb{N}\cap[2,\infty)$ and $\varepsilon > 0$ be arbitrarily fixed, and assume that $g_\varepsilon \in C^{1,1}(\mathbb{R})$ satisfies* (2.12)*. Then the function $\psi_{n,\varepsilon}$ defined in* (2.13) *is of class $C^{1,1}_{loc}(\mathbb{R}^n)$ and satisfies for all $i,j \in \{1,\dots,n\}$ that*

$$\frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(a) = \delta_{i,n} - g'_\varepsilon(a_n - \psi_{n-1,\varepsilon}(a))(\delta_{i,n} - \frac{\partial \psi_{n-1,\varepsilon}}{\partial a_i}(a)),\ \textit{for all } a \in \mathbb{R}^n, \tag{2.14}$$

$$\begin{aligned}\frac{\partial^2 \psi_{n,\varepsilon}}{\partial a_i, \partial a_j}(a) &= -g''_\varepsilon(a_n - \psi_{n-1,\varepsilon}(a))\left(\delta_{i,n} - \frac{\partial \psi_{n-1,\varepsilon}}{\partial a_i}(a)\right)\left(\delta_{j,n} - \frac{\partial \psi_{n-1,\varepsilon}}{\partial a_j}(a)\right)\\ &+g'_\varepsilon(a_n - \psi_{n-1,\varepsilon}(a))\frac{\partial^2 \psi_{n-1,\varepsilon}}{\partial a_i, \partial a_j}(a),\ \textit{for almost all } a \in \mathbb{R}^n,\end{aligned} \tag{2.15}$$

where $\delta_{i,j}$ stand for the Kronecker delta. Further, we have that

$$\|\psi_{n,\varepsilon} - \psi_n\|_{L^\infty(\mathbb{R}^n)} \leqslant (n-1)\varepsilon, \tag{2.16}$$

$$\frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(a) \geqslant 0, \quad \sum_{j=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_j}(a) = 1,\ \textit{for all } a \in \mathbb{R}^n \tag{2.17}$$

$$-2(n-1)\left\|g''_\varepsilon\right\|_{L^\infty(\mathbb{R})} \leqslant b^\top \nabla^2 \psi_{n,\varepsilon}(a) b \leqslant 0,\ \textit{for almost all } a \in \mathbb{R}^n, \tag{2.18}$$

for all $b \in \mathbb{R}^n$ with $|b| = 1$, and

$$\left|\nabla^2 \psi_{n,\varepsilon}(a)\right| \leqslant 2(n-1)\left\|g''_\varepsilon\right\|_{L^\infty(\mathbb{R})}\ \textit{for almost all } a \in \mathbb{R}^n. \tag{2.19}$$

Proof. By the chain rule it is immediate to obtain (2.14) and (2.15). In (2.17), the positivity of the partial derivatives is a direct consequence of (2.12b). For the second part of (2.17) we proceed by induction on n. The base case is $\psi_{1,\epsilon}(a) = a_1$, which clearly holds true. Let us assume as inductive hypothesis that the claim holds for $n-1$ with $n \geqslant 2$. By (2.14) and again (2.12b), we see that $\nabla\psi_{n.\varepsilon}(a)$ is a convex combination between e_n (the $n-th$ vector of the canonical basis of $\mathbb{R}^n$) and $\nabla\psi_{n-1}(a)$, where the gradient is with respect to a. Since the sum of the e_n is equal to 1 and by the inductive hypothesis the same holds true for $\nabla\psi_{n-1}(a)$, we get that the sum of the elements of $\nabla\psi_n(a)$ is 1 as well, which proves the second part of (2.17).

Let $b \in \mathbb{R}^n$ be such that $|b| = 1$. By (2.15) we have that

$$
\begin{aligned}
\sum_{i,j=1}^{n} \frac{\partial^2 \psi_{n,\varepsilon}(a)}{\partial a_i \partial a_j} b_i b_j &= -g_\varepsilon''(a_n - \psi_{n-1,\varepsilon}(a))((e_n - \nabla\psi_{n-1,\varepsilon}(a)) \cdot b)^2 \\
&+ g_\varepsilon'(a_n - \psi_{n-1,\varepsilon}(a)) \sum_{i,j=1}^{n} \frac{\partial^2 \psi_{n-1,\varepsilon}(a)}{\partial a_i \partial a_j} b_i b_j.
\end{aligned} \tag{2.20}
$$

We note that the first term in the right hand-side of the above expression is bounded from below. To see this, we first point out that

$$
((e_n - \nabla\psi_{n-1,\varepsilon}(a)) \cdot b)^2 = \left(b_n - \sum_{i=1}^{n-1} \frac{\partial \psi_{n-1,\varepsilon}}{\partial a_i}(a) b_i \right)^2 \leqslant 2 \left(b_n^2 + \sum_{i=1}^{n-1} \frac{\partial \psi_{n-1,\varepsilon}}{\partial a_i}(a) b_i^2 \right) \leqslant 2,
$$

where we have used (2.17), Jensen inequality, $|b| = 1$, and the fact that $(x-y)^2 \leqslant 2(x^2 + y^2)$ for all $x, y \in \mathbb{R}$. Using this and (2.12c) in (2.20) we obtain

$$
\begin{aligned}
&-2g_\varepsilon''(a_n - \psi_{n-1,\varepsilon}(a)) + g_\varepsilon'(a_n - \psi_{n-1,\varepsilon}(a)) \sum_{i,j=1}^{n} \frac{\partial^2 \psi_{n-1,\varepsilon}(a)}{\partial a_i \partial a_j} b_i b_j \\
&\leqslant \sum_{i,j=1}^{n} \frac{\partial^2 \psi_{n,\varepsilon}(a)}{\partial a_i \partial a_j} b_i b_j \leqslant \sum_{i,j=1}^{n} \frac{\partial^2 \psi_{n-1,\varepsilon}(a)}{\partial a_i \partial a_j} b_i b_j.
\end{aligned}
$$

Using induction in both of the above inequalities we obtain (2.18).

This implies that for almost all $a \in \mathbb{R}^n$, the eigenvalues of $\nabla^2 \psi_{n,\varepsilon}(a)$ are contained in $[-2(n-1) \left\| g_\varepsilon'' \right\|_{L^\infty(\mathbb{R})}, 0]$, and hence (2.19) holds.

We turn now our attention to the proof of (2.16). By subtracting (2.11) from (2.13) we have that for all $a \in \mathbb{R}^n$ and $i = 1, \ldots, n-1$

$$
\psi_{i+1,\varepsilon}(a) - \psi_{i+1}(a) = (a_{i+1} - \psi_i(a))_+ - g_\varepsilon\left(a_{i+1} - \psi_{i,\varepsilon}(a)\right).
$$

Subtracting and adding the term $(a_{i+1} - \psi_{i,\varepsilon}(a))_+$ on the right hand-side of the previous inequality we obtain

$$
\begin{aligned}
&\psi_{i+1,\varepsilon}(a) - \psi_{i+1}(a) = (a_{i+1} - \psi_i(a))_+ - (a_{i+1} - \psi_{i,\varepsilon}(a))_+ + \\
&(a_{i+1} - \psi_{i,\varepsilon}(a))_+ - g_\varepsilon\left(a_{i+1} - \psi_{i,\varepsilon}(a)\right).
\end{aligned}
$$

Using the 1-Lipschitz continuity of the positive part and (2.12d) in the above expression we obtain that

$$
|\psi_{i+1,\varepsilon}(a) - \psi_{i+1}(a)| \leqslant \varepsilon + |\psi_{i,\varepsilon}(a) - \psi_i(a)|.
$$

Summing from $i = 1$ to $i = n-1$ we obtain the desired estimate for all $a \in \mathbb{R}^n$

$$
|\psi_{n,\varepsilon}(a) - \psi_n(a)| \leqslant (n-1)\varepsilon.
$$

□

Remark 2.3. *It is interesting to observe that* (2.17) *implies that the partial derivative of* $\psi_{n,\varepsilon}$ *at a point in* $\mathbb{R}^n$ *form a probability distribution. We will delve into this in Section 3, where via* (2.17)*, we give a probabilistic interpretation of the partial derivatives of* $\psi_{n,\varepsilon}$ *and connect it with the discontinuities of the gradient of a semiconcave function.*

To construct an approximation of the type (2.1) we choose a finite family of functions $\{\phi_i\}_{i=1}^n \subset C^2(\overline{\Omega})$ and set

$$\Phi_n = (\phi_1, \ldots, \phi_n), \quad v_n = \psi_n \circ \Phi_n, \quad \text{and } v_{n,\varepsilon} = \psi_{n,\varepsilon} \circ \Phi_n. \tag{2.21}$$

In Proposition 2.3 below, we prove that $v_{n,\varepsilon}$ is a $C^{1,1}(\overline{\Omega})$ approximation of v_n and that it is semiconcave.

Proposition 2.3. *Let $\{\phi_i\}_{i=1}^n \subset C^2(\overline{\Omega})$. Then $v_{n,\varepsilon}$ is $C-$semiconcave and $L-$Lipschitz continuous, with*

$$C = \max_{i\in\{1,\ldots,n\}} \left\|\nabla^2 \phi_i\right\|_{C(\overline{\Omega};\mathbb{R}^{d\times d})} \text{ and } L = \max_{i\in\{1,\ldots,n\}} \|\nabla \phi_i\|_{C(\overline{\Omega};\mathbb{R}^d)}.$$

Further,

$$\|v_{n,\varepsilon} - v_n\|_{C(\overline{\Omega})} \leqslant (n-1)\varepsilon. \tag{2.22}$$

and $v_{n,\varepsilon}$ is of class $C^{1,1}$ with

$$\left\|\nabla^2 v_{n,\varepsilon}\right\|_{L^\infty(\Omega;\mathbb{R}^{d\times d})} \leqslant C + 2n(n-1)L^2 \left\|g_\varepsilon''\right\|_{L^\infty(\mathbb{R})}. \tag{2.23}$$

Proof. We note that $v_{n,\varepsilon}$ is in $C^{1,1}(\overline{\Omega})$ since it is the composition of $\psi_{n,\varepsilon} \in C^{1,1}(\mathbb{R}^n)$ and $\Phi_n \in C^2(\overline{\Omega};\mathbb{R}^n)$. For $x \in \overline{\Omega}$, by the chain rule we have that

$$\nabla v_{n,\varepsilon} = \sum_{i=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(\Phi_n(x))\nabla\phi_i(x),$$

and by (2.17) with $a = \Phi_{n,\epsilon}(x)$ we find that

$$|\nabla v_{n,\varepsilon}(x)| = \left|\sum_{i=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(\Phi_n(x))\nabla\phi_i(x)\right| \tag{2.24}$$

$$\leqslant \sum_{i=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(\Phi_n(x)) \|\nabla\phi_i\|_{C(\overline{\Omega};\mathbb{R}^d)} \leqslant \max_{i\in\{1,\ldots,n\}} \|\nabla \phi_i\|_{C(\overline{\Omega};\mathbb{R}^d)}. \tag{2.25}$$

The asserted Lipschitz continuity of $v_{n,\varepsilon}$ follows from this estimate. We next argue the semiconcavity of $v_{n,\varepsilon}$ and the bound (2.23). For this purpose, let us consider $y \in \mathbb{R}^d$. By the chain rule we have for almost all $x \in \Omega$ and all $i, j \in \{1, \ldots, d\}$ that

$$\frac{\partial v_{n,\varepsilon}}{\partial x_i \partial x_j}(x) = \sum_{k,r=1}^n \frac{\partial^2 \psi_{n,\varepsilon}}{\partial a_r \partial a_k}(\Phi_n(x))\frac{\partial \phi_r}{\partial x_i}(x)\frac{\partial \phi_k}{\partial x_j}(x) + \sum_{k=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_k}(\Phi_n(x))\frac{\partial^2 \phi_k}{\partial x_i \partial x_j}(x). \tag{2.26}$$

We note that by (2.18), we have for almost all $x \in \Omega$ and all $y \in \mathbb{R}^d$, with $|y| = 1$, that

$$\begin{aligned} &-2n(n-1)\left\|g''\right\|_{L^\infty(\mathbb{R})} \max_{i\in\{1,\ldots,n\}} \|\nabla\phi_i\|^2_{C(\overline{\Omega};\mathbb{R}^d)} \\ &\leqslant \sum_{i,j=1}^d \sum_{k,r=1}^n \frac{\partial^2 \psi_{n,\varepsilon}}{\partial a_r \partial a_k}(\Phi_n(x))\frac{\partial \phi_r}{\partial x_i}(x)\frac{\partial \phi_k}{\partial x_j}(x) y_i y_j = \sum_{k,r=1}^n z_r(x)\frac{\partial^2 \psi_{n,\varepsilon}}{\partial a_r \partial a_k}(\Phi_n(x)) z_k(x) \leqslant 0, \end{aligned}$$

where $z_r(x) = \sum_{i=1}^d \frac{\partial \phi_r}{\partial x_i}(x) y_i$, $z_k(x) = \sum_{j=1}^d \frac{\partial \phi_k}{\partial x_j}(x) y_j$. Combining this with (2.26), (2.17), and (2.19) , we obtain

$$-C - 2n(n-1)\left\|g''\right\|_{L^\infty(\mathbb{R})} L^2 \leqslant y^\top \cdot \nabla^2 v_{n,\varepsilon}(x) \cdot y \leqslant C, \text{ for all } y \in \mathbb{R}^d, \text{ with } |y| = 1.$$

This implies that the eigenvalues of $\nabla^2 v_{n,\varepsilon}(x)$ are in $[-C - 2n(n-1)\left\|g''\right\|_{L^\infty(\mathbb{R})} L^2, C]$ from which we deduce (2.23), and that the semiconcavity constant of $v_{n,\varepsilon}$ is bounded by C, see Proposition 1.1.3 in [6].

To conclude the proof, we note that (2.22) is a direct consequence of (2.16). □

We are now in position to introduce our semiconcavity preserving parametrization. With Proposition 2.1 and (2.21) in mind, for the purpose of numerical realization, it remains to choose an approximation of the family of functions ϕ_i in order achieve a finite dimensional semiconcave parametrization for functions v. For this purpose, let us consider a finite dimensional Banach space Θ equipped with a norm $\|\cdot\|_\Theta$ and a continuous function $\xi : \Theta \mapsto C^2(\overline{\Omega})$. For $n \in \mathbb{N}$ and $\theta = (\theta^1, \dots, \theta^n) \in \Theta^n$, we set $\Xi_n(\theta) = (\xi(\theta^1), \dots, \xi(\theta^n))$. Thus, for $\theta \in \Theta^n$ we have that $\Xi_n(\theta)$ acts as a family of smooth functions which parameterizes Φ. We equip Θ^n with the supremum norm, namely, for $\theta \in \Theta^n$ we use $\|\theta\|_{\Theta^n} = \sup_{i=1,\dots,n} \|\theta_i\|_\Theta$. We call the tuple (Θ, ξ) a setting.

We propose the following semiconcavity preserving parametrization

$$v_{n,\varepsilon}(\theta) = \psi_{n,\varepsilon} \circ \Xi_n(\theta), \tag{2.27}$$

with $\varepsilon \geqslant 0$. If $\varepsilon = 0$ we drop the ε from the sub-index and write

$$v_n(\theta) = \psi_n \circ \Xi_n(\theta).$$

By Proposition 2.3 it is clear that for $\theta \in \Theta^n$ fixed the semiconcavity constant of $v_{n,\varepsilon}(\theta)$ will be bounded by

$$\max_{i \in \{1,\dots,n\}} \left\| \nabla^2 \xi(\theta^i) \right\|_{C(\overline{\Omega};\mathbb{R}^{d\times d})}. \tag{2.28}$$

Therefore, controlling this quantity is crucial for preserving semiconcavity.

In section 4 the convergence properties of $v_{n,\varepsilon}$ are studied with respect to a sequence of settings $\{(\Theta^m, \xi^m)\}_{m\in\mathbb{N}}$ as $\varepsilon \to 0^+$ and $n \to \infty$.

3 Probabilistic interpretation and active sets

In this section we give a probabilistic interpretation of the parametrization introduced in Section 2. To this end, we consider a family of functions $\{\phi_i\}_{i=1}^n \subset C^2(\overline{\Omega})$ and the functions v_n and $v_{n,\varepsilon}$ as in (2.21). The interpretation consists in setting for each $a \in \mathbb{R}^n$ a probability $\{p_{n,i,\varepsilon}(a)\}_{i=1}^n$ over the indices $\{1, \dots, n\}$ by using the partial derivatives of ψ_n. We will see that in the limit as $\varepsilon \to 0^+$, the probability $\{p_{n,i,\varepsilon}(a)\}_{i=1}^n$ has support on the active set of indices $\{i \in \mathbb{N} :\ 1 \leqslant i \leqslant n,\ \psi_n(a) = a_i\}$ and we will provide hypotheses under which an explicit formula for the limit probabilities exists. This will help to characterize the *active sets* for $v_{n,\varepsilon}$ and v_n which are the sets where ϕ_i is equal to v_n, and where $\nabla\phi_i$ contributes the most to the gradient of v_n.

The relevance of this probabilitic interpretation and the active sets described above lies in the fact that the discontinuities of ∇v_n occur in the boundary of the active sets, thus understanding the behavior of the probabilities and the actives sets is important regarding the characterization of the discontinuities of the gradient of v_n. Further, a correct representation of the active sets by the approximation $v_{n,\varepsilon}$ is crucial for capturing the correct behavior of v_n and its gradient around the discontinuities.

The behavior of the gradients is particularly importance if v_n represents the value function of an optimal control problem. In this case, the optimal control in feedback form can be expressed by means of the gradient of the value function. We explain this in more details in Remark 3.1.

We commence by observing that

$$p_{n,j,\varepsilon}(a) = \frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(a) \text{ for } a \in \mathbb{R}^n \text{ and } j \in \{1, \dots, n\}. \tag{3.1}$$

By(2.17) the family $\{p_{n,j,\varepsilon}(a)\}_{j=1}^n$ can be interpreted as a probability distribution over the indices. Furthermore, by the chain rule we have for $x \in \Omega$ that

$$\nabla v_{n,\varepsilon}(x) = \sum_{i=1}^n \frac{\partial \psi_{n,\varepsilon}}{\partial a_i}(\Phi_n(x)) \nabla\phi_i(x) = \sum_{i=1}^n p_{n,i,\varepsilon}(\Phi_n(x)) \nabla\phi_i(x), \tag{3.2}$$

from where we see that $p_{n,i,\varepsilon}(\Phi_n(x))$ measure the contribution of $\nabla\phi_i(x)$ in the summation and also that $\nabla v_{n,\varepsilon}(x)$ correspond to the average of $\{\nabla\phi_i(x)\}_{i=1}^n$ weighted by $p_{n,i,\varepsilon}(\Phi_n(x))$.

It is noteworthy that by (2.13) the following recursive relations hold

$$p_{m,j,\varepsilon}(a) = \delta_{m,j} - g'_\varepsilon\left(a_m - \psi_{m-1,\varepsilon}(a)\right)\left(\delta_{m,j} - p_{m-1,j,\varepsilon}(a)\right) \text{ for } j \in \{1,\dots,n\} \text{ and } m \in \{2,\dots,n\} \tag{3.3}$$

and

$$p_{1,j,\varepsilon}(a) = \delta_{1,j} \text{ for } j \in \{1,\dots,n\}, \tag{3.4}$$

for $a \in \mathbb{R}^n$, where it is understood that $p_{i,j,\varepsilon} = 0$ for $i \in \{1,\dots,n-1\}$ and $j \in \{i+1,\dots,n\}$.

We can iterate these relations to obtain

$$p_{m,j,\varepsilon}(a) = \left(\prod_{i=j+1}^{m} g'_\varepsilon\left(a_i - \psi_{i-1,\varepsilon}(a)\right)\right)\left(1 - g'_\varepsilon(a_j - \psi_{j-1,\varepsilon}(a))\right) \tag{3.5}$$

for all $m \in \{2,\dots,n\}$ and $j \in \{1,\dots,m-1\}$, and

$$p_{j,j,\varepsilon}(a) = \left(1 - g'_\varepsilon(a_j - \psi_{j-1,\varepsilon}(a))\right). \tag{3.6}$$

Next we address the limiting behavior of the probability distributions $p_{n,j,\varepsilon}(a)$ as ε tends to 0. For this will make use of the following assuption on the family of functions $\{g_\varepsilon\}_{\varepsilon>0}$ which approximates the positive part:

$$g_\varepsilon(r) \leqslant (r)_+ \text{ for all } r \in \mathbb{R}, \tag{3.7}$$

and

$$\text{there exits } s_0 \text{ satisfying } \lim_{\varepsilon\to 0^+} \sup_{s\in[-s_0,0]} |g'_\varepsilon(s)| = 0. \tag{3.8}$$

For this we will need the following technical lemma:

Lemma 3.1. *Let $\{g_\varepsilon\}_{\varepsilon>0}$ be a family of $C^{1,1}(\mathbb{R})$ functions satisfying* (2.12) *and* (3.7). *Then for all $m \in \{1,\dots,n\}$ we have*

$$\psi_m(a) \leqslant \psi_{m,\varepsilon}(a) \textit{ for } a \in \mathbb{R}^n \textit{ and all } \varepsilon \in (0,\infty). \tag{3.9}$$

Proof. We proceed by induction. Let us consider $a \in \mathbb{R}^n$ fixed. The base case $m = 1$ is trivial since $\psi_{1,\varepsilon}(a) = a_1 = \psi_1(a)$. Let us assume that the claim holds for $m \in \{2,\dots,n-1\}$, that is, $\psi_m(a) \leqslant \psi_{m,\varepsilon}(a)$. By (3.7) we have that

$$\psi_{m+1,\varepsilon}(a) \geqslant a_{m+1} - (a_{m+1} - \psi_{m,\varepsilon}(a))_+.$$

By the monotonicity of the positive part we have that $(a_{m+1} - \psi_{m,\varepsilon})_+ \leqslant (a_{m+1} - \psi_m)_+$, due to the fact that $\psi_m(a) \leqslant \psi_{m,\varepsilon}(a)$. From this we deduce that $\psi_{m+1,\varepsilon}(a) \geqslant \psi_{m+1}(a)$, which proves the result.

□

In order to simplify the notation in what follows, for $a \in \mathbb{R}^n$, we define the set of active indices $I_n(a) \subset \{1,\dots,n\}$ as the argmin over the family $\{\phi_i(a)\}_{i=1}^n$, and $\hat{i}_n(a) \in \{1,\dots,n\}$ as the largest active index, that is,

$$I_n(a) = \{i \in \{1,\dots,n\} : \psi_n(a) = a_i\}, \text{ and } \hat{i}_n(a) = \max I_n(a). \tag{3.10}$$

Proposition 3.1. *Let $g_\varepsilon \in C^{1,1}(\mathbb{R})$ with $\varepsilon > 0$ be a family of functions satisfying* (2.12)*. Then for $a \in \mathbb{R}^n$ and all $j \notin I_n(a)$ we have*

$$\lim_{\varepsilon\to 0^+} p_{n,j,\varepsilon}(a) = 0. \tag{3.11}$$

If $I_n(a)$ is a singleton, i.e. $I_n(a) = \{\hat{i}_n(a)\}$, then

$$\lim_{\varepsilon\to 0^+} p_{n,\hat{i}_n(a),\varepsilon}(a) = 1. \tag{3.12}$$

Further, if g_ε satisfies (3.7) *and* (3.8)*, then for all $j \in \{1,\dots,n\}$ we have*

$$\lim_{\varepsilon\to 0^+} p_{n,j,\varepsilon}(a) = \begin{cases} 1 & \textit{if } j = \hat{i}_n(a) \\ 0 & \textit{otherwise.} \end{cases} \tag{3.13}$$

Proof. Let us start by proving (3.11) and choose some $j \in \{1,\dots,n\} \setminus I_n(x)$. We consider separately the cases $a_j \leqslant \psi_{j-1}(a)$ and $a_j > \psi_{j-1}(a)$. In the first case, since $j \notin I_n(a)$, there exist $i \in \{j+1,\dots,n\}$ such that $a_i < \psi_{i-1}(a)$, otherwise we have that $a_j = \psi_n(a)$ which is a contradiction. Then by (2.16) and (2.12e), we have

$$\lim_{\varepsilon\to 0^+} g'_\varepsilon(a_i - \psi_{i-1,\varepsilon}(a)) = 0,$$

which by means of (3.5) implies that $\lim_{\varepsilon\to 0^+} p_{n,j,\varepsilon}(a) = 0$. On the other hand, if $a_j > \psi_{j-1}(a)$, there exists $\delta > 0$ such that

$$a_j - \psi_{j-1}(a) > \delta. \tag{3.14}$$

By (2.16) we have that for all $\varepsilon \in \left(0, \frac{\delta}{2(j-2)}\right)$ the estimate

$$|\psi_{j-1}(a) - \psi_{j-1,\varepsilon}(a)| \leqslant \frac{\delta}{2}. \tag{3.15}$$

Combining (3.14) and (3.15), we arrive at

$$\frac{\delta}{2} \leqslant a_j - \psi_{j-1,\varepsilon}(a) < a_j - \psi_{j-1}(a) + \frac{\delta}{2}, \text{ for all } \varepsilon \in \left(0, \frac{\delta}{2(j-2)}\right). \tag{3.16}$$

This implies that $\{a_j - \psi_{j-1,\varepsilon}(a)\}_{\varepsilon\in J}$, with $J = \left(0, \frac{\delta}{2(j-2)}\right)$ lies in a compact subset of $\mathbb{R} \setminus \{0\}$. Using this, (2.16), and (2.12e), we get that

$$\lim_{\varepsilon\to 0^+} g'_\varepsilon(a_j - \psi_{j-1,\varepsilon}(a)) = \chi_{[0,\infty)}(a_j - \psi_{j-1}(a)) = 1.$$

Combining this with (3.5), we again obtain (3.11).

We turn to the verification of (3.12) and assume that assume that $I_n(a) = \{\hat{i}_n(a)\}$. Then by (3.11) we know that

$$\lim_{\varepsilon\to 0^+} \sum_{j\in\{1,\dots,n\}\setminus I_n(a)} p_{n,j,\varepsilon}(a) = 0.$$

Since $\{p_{n,j,\varepsilon}(a)\}_{i=1}^n$ sum up one, $\lim_{\varepsilon\to 0^+} p_{n,\hat{i}(a),\varepsilon}(a) = 1 - \lim_{\varepsilon\to 0^+} \sum_{j\in\{1,\dots,n\}\setminus I_n(a)} p_{n,j,\varepsilon}(a) = 1$ follows. This proves (3.12).

We turn our attention now to the proof of (3.13). Let assume that (3.7) and (3.8) hold. We note that given the fact $\{p_{n,j,\varepsilon}(a)\}_{j=1}^n$ is a probability distribution, we only need to prove that

$$\lim_{\varepsilon\to 0^+} p_{n,\hat{i}_n(a),\varepsilon}(a) = 1, \tag{3.17}$$

to demonstrate (3.13). By Lemma 3.1 we know that

$$a_{\hat{i}_n(a)} = \psi_{\hat{i}_n(a)}(a) \leqslant \psi_{\hat{i}_n(a)-1}(a) \leqslant \psi_{\hat{i}_n(a)-1,\varepsilon}(a).$$

Then $a_{\hat{i}_n(a)} - \psi_{\hat{i}_n(a)-1,\varepsilon}(a) \leqslant 0$ for $\varepsilon > 0$. Combining this with (3.8), (2.12c), and (2.12e) (in the case that $a_{\hat{i}_n(a)} - \psi_{\hat{i}_n(a)-1}(a) < s_0$) we obtain

$$\lim_{\varepsilon\to 0^+} g'_\varepsilon(a_{\hat{i}_n(a)} - \psi_{\hat{i}_n(a)-1,\varepsilon})(a) = 0. \tag{3.18}$$

If $\hat{i}_n(a) = n$, this together with (3.6) proves (3.17). On the other hand, if $\hat{i}_n(a) < n$, we can make use of (3.5) instead. To this end, it is important to observe that for each $j \in \{1, \dots, n\}$ strictly larger than $\hat{i}_n(a)$, we have, by the definition of $\hat{i}_n(a)$, that $j \notin I_n(a)$ and therefore arguing as in the proof of (3.12) we have that

$$\lim_{\varepsilon\to 0^+} \prod_{j=\hat{i}_n(a)+1}^{n} g'_\varepsilon(a_j - \psi_{j-1,\varepsilon}(a)) = 1.$$

Combining this with (3.18), and (3.6) we obtain that (3.17) holds. □

Corollary 3.1. *If* (3.7) *and* (3.8) *hold, then for each $x \in \Omega$ all the probability is assigned to $\hat{i}_n(\Phi_n(x))$ and consequently $\nabla v_{n,\varepsilon}(x)$ converges to $\nabla\phi_{\hat{i}_n(\Phi_n(x))}(x)$ as $\varepsilon \to 0^+$.*

Remark 3.1. *The property described in the previous corollary is particularly convenient for the value function in optimal control as we proceed to explain. Let v_n be a viscosity solution of an equation of the form*

$$F(x, \nabla v(x)) = 0 \text{ in } \Omega \tag{3.19}$$

with F continuous, and convex in the second argument. For $x \in \Omega$ let

$$\begin{aligned} D^* v_n(x) := \; & \{p \in \mathbb{R}^d : p = \lim_{k\to\infty} \nabla v_n(x_k),\ v_n \text{ differentiable at } x_k \text{ for all } k \in \mathbb{N}, \\ & \text{and } x = \lim_{k\to\infty} x_k\}. \end{aligned}$$

We recall that v_n is semiconcave and hence $D^* v_n(x)$ is nonempty for each $x \in \Omega$, see [6, Proposition 3.3.4(c), Theorem 3.3.6]. Then , v_n satisfies (3.19) *a.e. and we have that $F(x,p) = 0$ for all $x \in \Omega$ and all $p \in D^* v_n(x)$. Here we use the continuity of F and [6, Remark 5.4.1]. In particular, if for all $x \in \Omega$ and $i \in \{j \in \mathbb{N} : 1 \leqslant j \leqslant n,\ \phi_j(x) = v_n(x)\}$ it holds*

$$\nabla\phi_i(x) \in D^* v_n(x), \tag{3.20}$$

then

$$F(x, \nabla\phi_i(x)) = 0 \text{ for all } i \in \{1, \dots, n\} \text{ such that } v_n(x) = \phi_i(x). \tag{3.21}$$

In particular, assume that (3.20) *holds with $i = \hat{i}_n(\Phi_n(x))$, i.e.*

$$\nabla\phi_{\hat{i}_n(\Phi_n(x))}(x) \in D^* v_n(x), \tag{3.22}$$

for all $x \in \Omega$. In this situation assigning all the probability to only one of the active functions is advantageous for the asymptotic behavior of the $\varepsilon-$approximation. Indeed, by (3.13) *and* (3.2) *we have*

$$\lim_{\varepsilon\to 0^+} \nabla v_{n,\varepsilon}(x) = \nabla\phi_{\hat{i}_n(\Phi_n(x))}(x), \text{ for all } x \in \Omega,$$

which together with the continuity of F and (3.22) *implies*

$$\lim_{\varepsilon\to 0^+} F(x, \nabla v_{n,\varepsilon}(x)) = F(x, \nabla\phi_{\hat{i}_n(\Phi_n(x))}(x)) = 0, \text{ for all } x \in \Omega. \tag{3.23}$$

If (3.7) *and* (3.8) *do not hold, the limit of* $\nabla v_{n,\varepsilon}(x)$ *may lie in* $D^+v_n(x) \setminus D^*v_n(x)$ *and by Remark 5.4.1 in [6][Chapter 5] we can only ensure*

$$\lim_{\varepsilon \to 0^+} F(x, \nabla v_{n,\varepsilon}(x)) \leqslant 0.$$

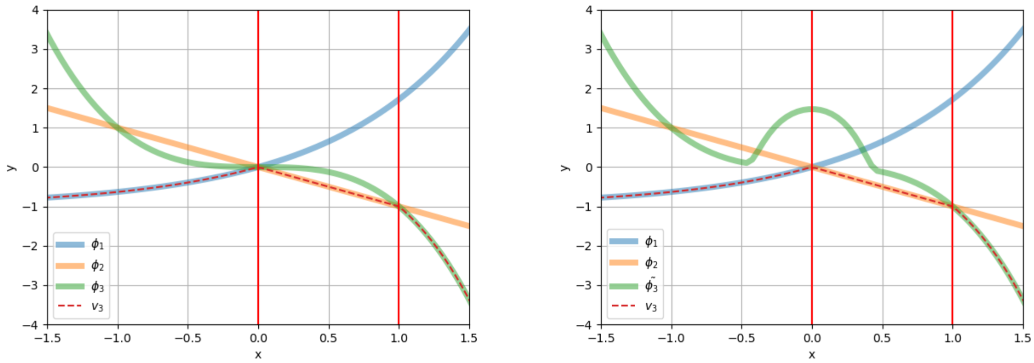
We shall return to this case in Remark 3.3 below.

Remark 3.2. *It is important to observe that condition* (3.20) *does not hold in general. If* $n = 2$*, then it holds trivially for arbitrary* d*, however if* $n \geqslant 3$ *it is possible to construct a counter example as follows. Let* $\phi_1(x) = -x$*,* $\phi_2(x) = \exp(x) - 1$*,* $\phi_3(x) = -x^3$ *and* $v_3(x) = \min_{i \in \{1,2,3\}} \phi_i(x)$ *for* $x \in \mathbb{R}$*.*

We have that

$$v_3(x) = \begin{cases} \exp(x) - 1 & \text{if } x \leqslant 0 \\ -x & \text{if } x \in (0, 1] \\ -x^3 & \text{if } x > 1 \end{cases}$$

as is depicted in Figure 1a



(a) Unmodified functions.

(b) Modified functions.

Figure 1: Illustrative example for (3.20).

At $x = 0$ *the three functions* $\{\phi_i\}_{i=1}^3$ *are equal to* $v_3(0)$*. Additionally, we have that*

$$D^*v_3(0) = \{-1, 1\}. \tag{3.24}$$

To prove this, we note that, for $x < 0$*,* v_3 *is differentiable and* $v_3'(x) = \exp(x)$*. Then clearly we have* $\exp(0) = 1$ *is an element of* $D^*v_3(0)$*. For proving that* -1 *is an element of* $D^*v_3(0)$*, we note that for* $x \in (0, 1)$ *we have that* $v_n = -x$ *and consequently* $v_n'(x) = -1$ *which implies that* $-1 \in D^*v_3(0)$*. It is important to observe, that* $D^*v_3(0)$ *does not include any other element, since any sequence* $p_n = v_3'(x_n)$ *with* x_n *converging to* 0 *and* p_n *converging, is such that* $p_n = \exp(x_n)$ *for all* n *large enough or* $p_n = -1$ *for* n *large enough.*

In addition, we have that $\phi_3'(0) = 0$ *and therefore by* (3.24)*, it is clear that* $\phi_3'(0)$ *is not an element of* $D^*(0)$*.*

We can remedy this issue by modifying ϕ_3 *in a neighborhood of* 0 *in a way that it will not change* v_3*. For instance, let us consider* $\tilde{\phi}_3(x) = \phi_3\,(x) + f(2x)$ *with*

$$f(x) = \begin{cases} \exp(-1/(1 - |x|^2)) & \text{if } x \in (-1, 1) \\ 0 & \text{if } |x| \geqslant 1 \end{cases}$$

As is depicted in Figure 1b, we have that $v_3(x) = \min\{\phi_1(x), \phi_2(x), \tilde{\phi}_3(x)\}$*, but now* $v_3(0) \neq \tilde{\phi}_3(0)$ *and consequently the new family* $\{\phi_1, \phi_2, \tilde{\phi}_3\}$ *satisfies* (3.20)*.*

Using the same idea, that is, modifying the problematic ϕ_i's around the point where (3.20) *does not hold, we show in Remark 3.4 a general way to modify the family of functions* $\{\phi_i\}_{i=1}^n$ *in such a way that* (3.20) *is fulfilled everywhere.*

In view of Proposition 3.1 with (3.7) and (3.8) holding, the expression

$$p_{n,i}(a) := \lim_{\varepsilon\to 0^+} p_{n,i,\varepsilon}(a) \in \{0,1\}, \tag{3.25}$$

is welldefined for each $a \in \mathbb{R}^n$ and $i \in \{1,\dots,n\}$. Now for each $i = 1,\dots,n$ we define the Ω^i and their corresponding approximations by $\Omega^{i,\varepsilon}$ are defined by

$$\Omega^i = \{x \in \overline{\Omega} : p_{n,i}(\Phi(x)) = \max_{j=1,\dots,n} p_{n,j}(\Phi(x))\}, \tag{3.26}$$

$$\Omega^{i,\varepsilon} = \{x \in \overline{\Omega} : p_{n,i,\varepsilon}(\Phi(x)) = \max_{j=1,\dots,n} p_{n,j,\varepsilon}(\Phi(x))\}. \tag{3.27}$$

We refer to the sets $\{\Omega^i\}_{i=1}^n$ respectively $\{\Omega^{i,\varepsilon}\}_{i=1}^n$ as gradient active sets, since if $x \in \Omega^{i,\varepsilon}$, then ϕ_i is the function with the largest contribution to $\nabla v_{n,\varepsilon}(x)$ for ε sufficiently small. We also observe that $\overline{\Omega} = \cup_{i=1}^n \Omega^i$. The following convergence result with respect to the gradient active sets can be obtained.

Proposition 3.2. *Let* $\{g_\varepsilon\}_{\varepsilon>0} \subset C^{1,1}(\mathbb{R})$ *be a family of functions satisfying* (2.12), (3.7), *and* (3.8). *Then for every* $x \in \overline{\Omega}$ *we have that*

$$\lim_{\varepsilon\to 0^+} \chi_{\Omega_{i,\varepsilon}}(x) = \chi_{\Omega_i}(x),$$

where χ_ω *denotes the characteristic function of* $\omega \subset \mathbb{R}^d$.

Proof of Proposition 3.2. Let $x \in \overline{\Omega}$ fixed. We start by noting that by (3.13) we have for all $i \in \{1,\dots,n\}$ that

$$p_{n,j}(\Phi(x)) = \begin{cases} 1, & i = \hat{i} \\ 0, & i \neq \hat{i}, \end{cases}$$

where $\hat{i}$ is the largest index for which $\phi_i(x) = v_n(x)$. From this we deduce that there exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$ we have

$$p_{n,i,\varepsilon}(\Phi(x)) < \frac{1}{2} \text{ for } i \in \{1,\dots,n\} \setminus \{\hat{i}\}, \text{ and } p_{n,\hat{i},\varepsilon}(\Phi(x)) > \frac{1}{2}.$$

By the definitions of $\Omega_{i,\varepsilon}$ and Ω_i this implies that for all $\varepsilon \in (0,\varepsilon_0)$ it holds that

$$\chi_{\Omega^{i,\varepsilon}}(x) = \chi_{\Omega^i}(x),$$

which implies that the pointwise convergence of $\chi_{\Omega_{i,\varepsilon}}$ as $\varepsilon \to 0^+$. □

Remark 3.3. *In the following we demonstrate the consequences of the limit behavior of* g'_ε *in a neighborhood of* 0 *on the limiting behavior of the probability distribution given by* $p_{n,j,\varepsilon}(a)$ *as* $\varepsilon \to 0^+$.

Returning to $g_{\varepsilon,M}$ *and* $g_{\varepsilon,A}$ *introduced in Remark 2.2, both of them satisfy* (3.7) *and* (3.8). *Therefore, by means of Proposition 3.1 we obtain for* $a \in \mathbb{R}^n$ *and* $j \in \{1,\dots,n\}$ *that*

$$p_{n,j}(a) = \begin{cases} 1 & \text{if } j = \hat{i}_n(a) \\ 0 & \text{otherwise}, \end{cases}$$

where we recall the notation introduced in (3.25).

Another choice of smooth approximation for the minimum of a vector of real numbers is the Log-Sum-Exp approximation which is given by

$$\tilde{\psi}_{n,\varepsilon}(a) = -\varepsilon \log \left(\frac{1}{n} \sum_{i=1}^{n} \exp\left(-\frac{a_i}{\varepsilon}\right) \right).$$

for $a \in \mathbb{R}^n$ and $\varepsilon > 0$, which can be considered in place of $\psi_{n,\varepsilon}(a)$ defined in (2.13)*.*

The gradient of $\tilde{\psi}_n$ is given by the soft-min function:

$$\nabla\tilde{\psi}_{n,\varepsilon}(a) = \frac{1}{\sum_{i=1}^{n} \exp\left(-\frac{a_i}{\varepsilon}\right)} \left(\exp\left(-\frac{a_1}{\varepsilon}\right), \ldots, \exp\left(-\frac{a_n}{\varepsilon}\right) \right).$$

See [12] for a collection of results regarding this functions and its application to machine learning. We also point out that this approximation was used in the context of convexity preserving neural networks in [5].

Although both $\psi_{n,\varepsilon}$ and $\tilde{\psi}_{n,\varepsilon}$ converge to ψ_n uniformly as $\varepsilon \to 0^+$, the limiting behaviors as $\varepsilon \to 0^+$ of the derivatives of these two approximation are different. This has an impact on the limiting properties of the approximation of ∇v_n provided by $\tilde{v}_{n,\epsilon} = \tilde{\psi}_{n,\varepsilon} \circ \Phi_n$ in comparison to $v_{n,\varepsilon}$. To see this, we note that setting $I(a) = \arg\min\{a_i : i \in \{1, \ldots, n\}\}$ for $a \in \mathbb{R}^n$ we have for $j \in \{1, \ldots, n\}$ that

$$\lim_{\varepsilon \to 0^+} \frac{\partial \tilde{\psi}_{n,\varepsilon}}{\partial a_j}(a) = \begin{cases} \frac{1}{|I(a)|} & \text{if } j \in I(a), \\ 0 & \text{if } j \notin I(a). \end{cases}$$

From this observation we see that $\nabla\tilde{v}_{n,\varepsilon}$ satisfies for each $x \in \Omega$ that

$$\lim_{\varepsilon \to 0^+} \nabla\tilde{v}_{n,\varepsilon}(x) = \frac{1}{|I(\Phi_n(x))|} \sum_{i \in I(\Phi_n(x))} \nabla\phi_i(x),$$

that is, $\nabla\tilde{v}_{n,\varepsilon}$ converges point-wise to the average of the gradients of the active functions. In line with Remark 3.1, if v_n is a viscosity solution of an equation of the form of (3.19)*, we can only ensure that*

$$\lim_{\varepsilon \to 0^+} F(x, \nabla\tilde{v}_{n,\varepsilon}(x)) \leqslant 0, \tag{3.28}$$

compared to

$$\lim_{\varepsilon \to 0^+} F(x, \nabla v_{n,\varepsilon}(x)) = 0, \tag{3.29}$$

if (3.7) *and* (3.8) *hold. Consequently in the context of viscosity solutions, $\psi_{n,\varepsilon}$ provides a better representation of v_n at points of discontinuity than $\tilde{\psi}_{n,\varepsilon}$.*

Remark 3.4. *Regarding Remark 3.1, if condition* (3.20) *does not hold, we now explain how to modify the functions $\{\phi_i\}_{i=1}^n$ in a manner such that v_n does not change and the new family satisfies* (3.20)*. For this purpose let us first fix some $i \in \{1, \ldots, n\}$ and define*

$$A_i = \{x \in \overline{\Omega} : \nabla\phi_i(x) \notin D^* v_n(x)\}.$$

We argue that A_i is open in the relative topology of $\overline{\Omega}$. Indeed, let $\bar{x} \in \overline{\Omega}$ be such that $\nabla\phi_i(\bar{x}) \notin D^* v_n(\bar{x})$, then there exists $\delta > 0$ such that for all $x \in B(\bar{x}, \delta) \cap \overline{\Omega}$ we have $\nabla\phi_i(x) \notin D^* v_n(x)$. To prove this claim, let us proceed by contradiction. If the claim does not hold, then there exists $x_m \in \overline{\Omega}$ approaching $\bar{x}$ as $m \to \infty$ such that $\nabla\phi_i(x_m) \in D^* v_n(x_m)$. By the definition of $D^* v(x_m)$, for each $m \in \mathbb{N}$ there exists a sequence $x_{k,m}$ such that $x_{k,m} \to x_m$ as $k \to \infty$, v_n is differentiable at $x_{k,m}$ and $\nabla v_n(x_{k,m}) \to \nabla\phi_i(x_m)$ as $k \to \infty$. From this we deduce that for each $m \in \mathbb{N}$ there exists $k(m)$ such that

$$|\nabla v_n(x_{k(m),m}) - \nabla\phi_i(x_m)| + |x_{k,m} - x_m| < \frac{1}{m}.$$

Then by setting $\tilde{x}_m = x_{k(m),m}$ we have that

$$|\nabla v_n(\tilde{x}_m) - \nabla\phi_i(\bar{x})| + |x_{k,m} - \bar{x}| < \frac{1}{m} + |\nabla\phi_i(\bar{x}) - \nabla\phi_i(x_m)| + |x_m - \bar{x}|.$$

Thus, by the continuity of $\nabla\phi_i$ and the convergence of x_m to $\bar{x}$ as $m \to \infty$, we deduce that

$$\lim_{m\to\infty} \nabla v_n(\tilde{x}_m) = \nabla\phi_i(\bar{x}) \text{ and } \lim_{m\to\infty} \tilde{x}_m = \bar{x}.$$

This implies that $\nabla\phi_i(\bar{x}) \in D^*v_n(\bar{x})$, which is a contradiction and consequently A_i is open in the relative topolgy of $\overline{\Omega}$.

In particular, this implies that for each $x \in A_i$ there exists $j \in \{1,\dots,n\} \setminus \{i\}$ with $v_n(x) = \phi_j(x)$. Indeed, if the aforementioned assertion does not hold, then there exists $x \in A_i$ such that $v_n(x) = \phi_i(x)$ and $v_n(x) < \phi_j(x)$ for all $j \in \{1,\dots,n\} \setminus \{i\}$. By the continuity of ϕ_i, this implies that there exists $\delta > 0$ such that for all $y \in B(x,\delta) \cap \Omega$ we have that $\phi_i(y) < \phi_j(y)$ for $j \in \{1,\dots,n\} \setminus \{i\}$ and consequently $v_n = \phi_i$ in $B(x,\delta) \cap \Omega$. From this we deduce that v_n is C^1 at x and that $\nabla v_n(x) = \nabla\phi_i(x)$. However, this contradict the fact that $x \in A_i$.

Using that for all $x \in A_i$ there exists $j \in \{1,\dots,n\} \setminus \{i\}$ satisfying $v_n(x) = \phi_j(x)$, we have that adding a positive number to $\phi_i(x)$ in a neighborhood of x will not change $v_n(x)$ since $\phi_j(x)$ remains unchanged.

Since A_i is open and due to the Lindelof property of $\mathbb{R}^d$, there exists an open sub-covering $\{B(x_j,\delta_j) \cap \overline{\Omega}\}_{j=1}^{\infty}$ of A_i such that

$$A_i = \bigcup_{j=1}^{\infty} B(x_j,\delta_j) \cap \overline{\Omega}.$$

Since A_i is bounded we can assume that $\{\delta_j\}_{j=1}^{\infty}$ is bounded. Let us consider a function $\nu : \mathbb{R}^d \mapsto \mathbb{R}$ such that $\nu \in C_c^\infty(\mathbb{R}^d)$ and $\nu(x) > 0$ in $B(0,1)$ and $\nu(x) = 0$ in $\mathbb{R}^d \setminus B(0,\delta)$. We define $\tilde{\phi}_i$ by

$$\tilde{\phi}_i(x) = \sum_{j=1}^{\infty} \frac{\delta_j^2}{2^j} \nu\left(\frac{x - x_j}{\delta_j}\right) + \phi_i(x) \text{ for } x \in \overline{\Omega}.$$

We clearly have that $\tilde{\phi}_i \in C^2(\overline{\Omega})$, $\tilde{\phi}_i(x) > \phi_i(x)$ for all $x \in A_i$ and $\tilde{\phi}_i(x) = \phi_i(x)$ for all $x \in \overline{\Omega} \setminus A_i$. Then replacing $\tilde{\phi}_i$ by ϕ_i does not affect the definition of v_n and we have that $\nabla\tilde{\phi}_i(x) \in D^*v_n(x)$ for every $x \in \overline{\Omega}$ where $\tilde{\phi}_i(x) = v_n(x)$. Now we iterate from $i = 1$ to $i = n$. If A_i is empty we do not need to do anything, on the other hand, if $A_i \neq \emptyset$, we modify ϕ_i as above. In this manner we obtain that at the end of the i-th iteration we have

$$v_n(x) = \min\left\{\min_{k\in\{1,\dots,i\}} \tilde{\phi}_k(x), \min_{k\in\{i+1,\dots,n\}} \phi_k(x)\right\} \text{ for all } x \in \overline{\Omega}$$

$$\text{and } \{x \in \overline{\Omega} : \nabla\tilde{\phi}_j(x) \notin D^*v_n(x) \text{ and } \tilde{\phi}_j(x) = v_n(x)\} = \emptyset$$

for $j \in \{1,\dots,i\}$. In this manner we end up with a modified family $\{\tilde{\phi}_i\}_{i=1}^n$ such that

$$v_n(x) = \min_{i\in\{1,\dots,n\}} \tilde{\phi}_i(x), \text{ and } \{x \in \overline{\Omega} : \nabla\tilde{\phi}_i(x) \notin D^*v_n(x) \text{ and } \tilde{\phi}_i(x) = v_n(x)\} = \emptyset.$$

4 Approximation

Our next goal is to analyze the universality of the parametrization $v_{n,\varepsilon}$, which was introduced at the end of Section 2, for semiconcave functions. We also address approximation properties involving the active sets of semiconcave functions. For this purposes we introduce a sequence of settings $\{(\Theta_m, \|\cdot\|_m), \xi_m\}_{m\in\mathbb{N}}$, which will be required to satisfy the following property:

Hypothesis 4.1. *For all $\phi \in C^2(\overline{\Omega})$ and $m \in \mathbb{N}$ there exist $\theta_m \in \Theta_m$, such that*

$$\lim_{m\mapsto\infty} \xi_m(\theta_m) = \phi \text{ in } C^2(\overline{\Omega}). \tag{4.1}$$

In analogy to the notation introduced in Section 2 we define for the sequence of settings $\{(\Theta_m, \|\cdot\|_m), \xi_m\}_{m\in\mathbb{N}}$ and the mappings

$$\Xi_m : (\Theta_m)^n \mapsto (C^2(\Omega))^n \text{ by } \Xi_m(\theta_1, \dots, \theta_m) = (\xi_m(\theta_1), \dots, \xi_m(\theta_n)) \text{ and}$$

$$v_{m,n,\varepsilon} : \Theta_m^n \mapsto C^2(\overline{\Omega}) \text{ by } v_{m,n,\varepsilon}(\theta) = \psi_{n,\varepsilon}(\Xi_m(\theta_1, \dots, \theta_m))$$

for $\theta = (\theta_1, \dots, \theta_n) \in (\Theta_m)^n$. We start with a technical result.

Proposition 4.1. *For all $v \in Lip(\bar{\Omega})$ and $p \in [1, \infty)$ the following inequality holds:*

$$\|\nabla v\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant (2Kd^2 + d^{\frac{p}{2}}|\Omega|)^{\frac{1}{p}} \Big(\|v\|_{C(\overline{\Omega})}^{\frac{1}{p}} \|\nabla v\|_{L^\infty(\Omega;\mathbb{R}^d)}^{\frac{p-1}{p}} + \|v\|_{C(\overline{\Omega})}^{\frac{1}{1+p}} \|\nabla v\|_{L^\infty(\Omega;\mathbb{R}^d)}^{\frac{p}{1+p}} \Big), \tag{4.2}$$

where $K = \max_{i\in\{1,\dots,n\}} |\Omega_i|$ with

$$\Omega_i = \{y \in \mathbb{R}^{d-1} : \exists a \in \mathbb{R} \text{ such that } (y_1, \dots, y_{i-1}, a, y_{i+1}, \dots, y_d) \in \Omega\}. \tag{4.3}$$

Proof. We first consider the case that $v \in C^1(\bar{\Omega})$. Let $\delta > 0$ be arbitrary. The L^p norm of the gradient can be split in the following manner

$$\int_\Omega |\nabla v|^p dx = \int_{|\nabla v|_\infty > \delta} |\nabla v|^p dx + \int_{|\nabla v|_\infty \leqslant \delta} |\nabla v|^p dx,$$

where for $x \in \mathbb{R}^d$, $|x|_\infty = \max_{i=1,\dots,d} |x_i|$ and $|x|$ is the Euclidean norm of x. Clearly the last term of the above inequality can be bounded by $d^{\frac{p}{2}}|\Omega|\delta^p$. Further, thanks to the Coarea formula, see [21, Theorem 3.11, Chapter 3]), we can express the first integral on the right hand side of the above inequality in terms of the level sets of v to obtain

$$\int_\Omega |\nabla v|^p dx \leqslant \int_{-\|v\|_{C(\Omega)}}^{\|v\|_{C(\Omega)}} \int_{v=s, |\nabla v|_\infty > \delta} |\nabla v|^{(p-1)} dH_{d-1}(x) ds + d^{\frac{p}{2}} \delta^p |\Omega|.$$

Here dH_{d-1} denotes the $(d-1)$-dimensional Hausdorff measure. Noticing that the set $\{x \in \Omega : |\nabla v(x)|_\infty > \delta\}$ is a subset of $\bigcup_{i=1}^d \{x : \left|\frac{dv}{dx_i}(x)\right| > \delta\}$ we obtain

$$\int_\Omega |\nabla v|^p dx \leqslant \sum_{i=1}^d \int_{-\|v\|_{C(\overline{\Omega})}}^{\|v\|_{C(\overline{\Omega})}} \int_{v=s, \left|\frac{dv}{dx_i}\right| > \delta} |\nabla v|^{(p-1)} dH_{d-1}(x) ds + d^{\frac{p}{2}} \delta^p |\Omega|.$$

In order to estimate the right hand-side of the above inequality, we will bound the terms inside the summation for each $i \in \{1, \dots, d\}$. For this purpose, we note that by the implicit function theorem the set $\{v = s, \left|\frac{dv}{dx_i}\right| > \delta\}$ is an open C^1 sub-manifold of $\{v = s\}$ which is short for $\{x \in \Omega : v(x) = s\}$. Thus for each $s \in [-\|v\|_{C(\overline{\Omega})}, \|v\|_{C(\overline{\Omega})}]$ and each i there exists an open set $V_i(s) \subset \mathbb{R}^{d-1}$ and a bijective function $x : V_i(s) \mapsto \{v = s : \left|\frac{dv}{dx_i}\right| > \delta\}$. Further $V_i(s)$ is bounded by some constant depending on Ω, d, and $\|v\|_{C(\bar{\Omega})}$. This allows us to use the area formula (see Theorem 3.9 in [21, Chapter 3, Section 3.3.2]) to obtain

$$\begin{aligned} &\textstyle\int_{\{v=s, \left|\frac{dv}{dx_i}\right| > \delta\}} |\nabla v|^{(p-1)} dH_{d-1}(x) \\ &= \textstyle\int_{V_i(s)} |\nabla v(x(y))|^{p-1} \left(1 + \left(\frac{dv}{dx_i}(x(y))\right)^{-2} \sum_{j\neq i} \left(\frac{dv}{dx_j}(x(y))\right)^2\right)^{\frac{1}{2}} dy. \end{aligned}$$

Here, following the notation of [21, Chapter 3, Section 3.3.4.B], we used that the Jacobian can be expressed as $J(x(y)) = \left(1 + \left(\frac{dv}{dx_i}(x(y))\right)^{-2} \sum_{j \neq i} \left(\frac{dv}{dx_j}(x(y))\right)^2\right)^{\frac{1}{2}}$.

Since v is Lipschitz and $|\frac{dv}{dx_i}(x(y))| > \delta$ for $y \in V_i$ we can bound the right hand-side of the above inequality to get

$$\int_{\{v=s, \left|\frac{dv}{dx_i}\right|>\delta\}} |\nabla v|^{(p-1)} dH_{d-1}(x) \leqslant d \left(\|\nabla v\|^{p-1}_{L^\infty(\Omega;\mathbb{R}^d)} + \frac{1}{\delta} \|\nabla v\|^{p}_{L^\infty(\Omega;\mathbb{R}^d)} \right) |V_i(s)|.$$

Consequently we get

$$\int_\Omega |\nabla v|^p dx \leqslant 2d^2 C \|v\|_{C(\overline{\Omega})} \left(\|\nabla v\|^{p-1}_{L^\infty(\Omega;\mathbb{R}^d)} + \frac{1}{\delta} \|\nabla v\|^{p}_{L^\infty(\Omega;\mathbb{R}^d)} \right) + d^{\frac{p}{2}} \delta^p |\Omega|,$$

where we have used that $V_i(s) \subset \Omega_i$ for all $i = 1, \ldots, d$ and $s \in [-\|v\|_{C(\overline{\Omega})}, \|v\|_{C(\overline{\Omega})}]$. Choosing $\delta = \|v\|^{\frac{1}{1+p}}_{C(\overline{\Omega})} \|\nabla v\|^{\frac{p}{1+p}}_{L^\infty(\Omega;\mathbb{R}^d)}$ in the above inequality we obtain

$$\int_\Omega |\nabla v|^p dx \leqslant 2d^2 K \|v\|_{C(\overline{\Omega})} \|\nabla v\|^{p-1}_{L^\infty(\Omega;\mathbb{R}^d)} + (2d^2 C + d^{\frac{p}{2}} |\Omega|) \|v\|^{\frac{p}{1+p}}_{C(\overline{\Omega})} \|\nabla v\|^{\frac{p^2}{1+p}}_{L^\infty(\Omega;\mathbb{R}^d)}. \tag{4.4}$$

This estimate implies (4.2) for the case $v \in C^1(\overline{\Omega})$.

For $v \in Lip(\overline{\Omega}) = W^{1,\infty}(\Omega)$, let us consider $\Omega_\varepsilon = \{x \in \Omega : \ dist(x, \partial\Omega) > \varepsilon\}$ for $\varepsilon > 0$. Since $v \in Lip(\overline{\Omega})$ there exists a sequence $v_n \in C^\infty(\Omega)$ such that $\lim_{n\to\infty} \|v_n - v\|_{C(\overline{\Omega_\varepsilon})} = 0$, $\lim_{n\to\infty} \|v_n - v\|_{W^{1,p}(\Omega_\varepsilon)} = 0$, for each $p \in [1, \infty)$, and $\|\nabla v_n\|_{L^\infty(\Omega)} \to \|\nabla v\|_{L^\infty(\Omega)}$ see e.g. [22, Theorem 11.24, Exercise 11.31]. In particular $v_n \in C^1(\overline{\Omega}_\varepsilon)$ for each n, and we can apply (4.2) for v_n with the domain Ω replaced by Ω_ε for all of the function spaces. Passing to the limit as $n \to \infty$ we obtain that (4.2) holds with Ω_ε instead of Ω for $v \in Lip(\overline{\Omega})$. Then, passing to the limit as $\varepsilon \to 0^+$, the desired estimate is obtained. □

Remark 4.1. *We observe that Proposition 4.1 allows to transfer information from a function to its gradient. For instance, for $v = f - f_n$, with $\|f\|_{W^{1,\infty}(\overline{\Omega})} \leqslant L$ and $\|f_n\|_{W^{1,\infty}(\overline{\Omega})} \leqslant L$ for all $n \in \mathbb{N}$ the previous proposition allows us to measure the convergence of ∇f_n to ∇f in terms of the $C(\overline{\Omega})$ distance between f_n and f.*

Remark 4.2. *In the case $\Omega = (a_1, b_1) \times \ldots \times (a_d, b_d)$ with $a, b \in \mathbb{R}^d$ and $a < b$ coordinate-wise, the constant K appearing in Proposition 4.1 is given by*

$$K = |\Omega| \sup_{i \in \{1,\ldots,d\}} \frac{1}{b_i - a_i}. \tag{4.5}$$

Making use of Proposition 4.1, we obtain Theorem 4.1 below. This is an approximation result for semiconcave functions for the architecture proposed in this work. The main emphasis lies on the fact that an approximation of the functions $\{\phi_i\}_{i=1}^n$ in the $C(\overline{\Omega})$ norm, induces information on the approximation of the gradient of a semiconcave function.

Theorem 4.1. *For $i = 1, \ldots, n$ let $\{\phi_{i,m}\}_{m=1}^\infty \subset C^2(\overline{\Omega})$ be sequences of functions converging to ϕ_i in $C(\overline{\Omega})$. Additionally, assume that there exist $L > 0$ and $C > 0$ satisfying*

$$\limsup_{m\to\infty} L_m \leqslant L \tag{4.6}$$

and

$$\limsup_{m\to\infty} C_m \leqslant C, \tag{4.7}$$

where

$$L_m = \sup_{i\in\{1,\dots,n\}} \|\nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)} \ \ \textit{and}\ C_m = \sup_{i\in\{1,\dots,n\}} \left\|\nabla^2\phi_{i,m}\right\|_{C(\overline{\Omega};\mathbb{R}^{d\times d})}.$$

For $m\in\mathbb{N}$ and $\varepsilon\geqslant 0$ let us set

$$v_n = \min_{i=1,\dots,n} \phi_i,\ \ v_{n,m,\varepsilon} = \psi_{n,\varepsilon}\circ\Phi_{n,m}$$

where $\Phi_{n,m}\in C^2(\overline{\Omega};\mathbb{R}^n)$ is defined by $x\in\overline{\Omega}\mapsto\Phi_{n,m}(x) = (\phi_{1,m}(x),\dots,\phi_{n,m}(x))$. Then

$$\|v_n - v_{n,m,\varepsilon}\|_{C(\overline{\Omega})} \leqslant \sup_{i=1,\dots,n} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})} + (n-1)\varepsilon, \tag{4.8}$$

$$\begin{aligned} &\|\nabla v_{n,m,\varepsilon} - \nabla v_n\|_{W^{1,p}(\Omega)} \leqslant (2Kd^2 + d^{\frac{p}{2}}|\Omega|)^{\frac{1}{p}} \\ &\Big\{ ((n-1)\varepsilon + \max_{i\in\{1,\dots,n\}} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})})^{\frac{1}{p}} (L_n + L)^{\frac{p-1}{p}} + \\ &((n-1)\varepsilon + \max_{i\in\{1,\dots,n\}} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})})^{\frac{1}{1+p}} (L_n + L)^{\frac{p}{1+p}} \Big\}, \end{aligned} \tag{4.9}$$

where K is the constant from the previous proposition. Further v_n is $C-$semiconcave and $L-$ Lipschitz.

Proof. To prove this result we recall that for all $\varepsilon>0$, the function $\psi_{n,\varepsilon}$ is a 1-Lipschitz for the topology induced by the norm $\|\cdot\|_\infty$. Therefore, we have that

$$\begin{aligned} &|v_n(x) - v_{n,m,\varepsilon}(x)| = |\psi_n(\Phi(x)) - \psi_{n,\varepsilon}(\Phi_m(x))(x)| \leqslant \\ &|\psi_n(\Phi(x)) - \psi_n(\Phi_m(x))(x)| + |\psi_{n,\varepsilon}(\Phi_m(x)) - \psi_n(\Phi_m(x))(x)| \\ &\leqslant \sup_{i\in\{1,\dots,n\}} |\phi_i(x) - \phi_{i,m}(x)| + (n-1)\varepsilon, \end{aligned}$$

where we have used (2.17). This proves (4.8).

To prove (4.9) we note that $v_{n,m,\varepsilon}$ is L_m-Lipschitz, due to (2.17), (2.12b), and (3.2). Therefore we have

$$|v_{n,m,\varepsilon}(x) - v_{n,m,\varepsilon}(y)| \leqslant L_m|x-y| \text{ for all } x,y\in\overline{\Omega}.$$

Using this, (4.8), (4.6), and the assumption that $\phi_{i,m}\to\phi_i$ in $C(\bar{\Omega})$ we obtain

$$|v_n(x) - v_n(y)| \leqslant L|x-y| \text{ for all } x,y\in\overline{\Omega},$$

i.e., v is $L-$ *Lipschitz*. This allows us to apply Proposition 4.1 to $v_{n,m,\varepsilon} - v_n$ and obtain

$$\begin{aligned} &\|\nabla v_{n,m,\varepsilon} - \nabla v_n\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant (2Kd^2 + d^{\frac{p}{2}}|\Omega|)^{\frac{1}{p}} \\ &\left(\|v_{n,m,,\varepsilon} - v_n\|_{C(\overline{\Omega})}^{\frac{1}{p}} (L_m + L)^{\frac{p-1}{p}} + \|v_{n,m,,\varepsilon} - v_n\|_{C(\overline{\Omega})}^{\frac{1}{1+p}} (L + L_m)^{\frac{p}{1+p}} \right). \end{aligned}$$

From this estimate and (4.8) we obtain (4.9).

Finally, since $v_{n,m}$ is C_m-semiconcave, we have, for all $x,y\in\Omega$ and $t\in[0,1]$,

$$tv_{n,m}(x) + (1-t)v_{n,m}(y) - v_{n,m}(tx + (1-t)y) \leqslant \frac{C_m}{2}t(1-t)|x-y|^2,$$

see [6, Proposition 1.1.3]. Applying the uniform convergence of $v_{n,m}$ to v_n and (4.7) we deduce, for all $x,y\in\Omega$ and $t\in[0,1]$, that

$$tv_n(x) + (1-t)v_n(y) - v_n(tx + (1-t)y) \leqslant \frac{C}{2}t(1-t)|x-y|^2.$$

which implies that $x\mapsto v(x) - \frac{C}{2}x^2$ is concave and consequently v_n is $C-$semiconcave.

□

Remark 4.3. *The fact that this result allows to measure the L^p-error between ∇v_n by $\nabla v_{n,m}$ in terms of the $C(\overline{\Omega})$-error between ϕ_i and $\phi_{i,m}$, is of importance for feedback control, since the feedback operator can be expressed in terms of the gradient of the value function. Semiconcavity of the value function is a well-studied property, see for instance [3, 6]. The established estimate can also be relevant for obtaining error bounds which do not increase exponentially with the dimension.*

We are now in position to prove the main theorem regarding the approximation properties of the proposed architecture.

Theorem 4.2. *Let v be a C-semiconcave function, which is Lipschitz continuous in $\overline{\Omega}$ with constant $L > 0$, and suppose that Hypothesis 4.1 holds. Then for each $p \in [1, \infty)$ and each $\delta > 0$, there exist $\varepsilon > 0$, $m \in \mathbb{N}$, $n \in \mathbb{N}$, and $\theta_m = (\theta_{m,1}, \dots, \theta_{n,m}) \in \Theta_m^n$ such that we have*

$$\|v_{n,m,\varepsilon}(\theta_m) - v\|_{C(\overline{\Omega})} + \|\nabla v_{n,m,\varepsilon}(\theta_m) - \nabla v\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant \delta,$$

$v_{n,m,\varepsilon}$ is $(L+\delta)$ Lipschitz and $(C+\delta)$-semiconcave.

Proof. Let $p \in [1, \infty)$, $\delta > 0$, and $\{\phi\}_{i=1}^{\infty}$ be the family of function established in Proposition 2.1. For each $n \in \mathbb{N}$, let us denote

$$v_n = \min_{i=1,\dots,n} \phi_i(x).$$

By construction each v_n is semiconcave with constant C and Lipschitz continuous with constant L. By Proposition 2.1 we can find $n \in \mathbb{N}$ such that

$$\|v_n - v\|_{C(\overline{\Omega})} + \|\nabla v_n - \nabla v\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant \frac{\delta}{2}. \tag{4.10}$$

Further $\|v_n\|_{W^{1,\infty}(\Omega)} \leqslant L$, and v_n is $C-$semiconcave.

In addition, by Hypothesis 4.1, applied for $\xi = \phi_i$ with $i = \{1, \dots, n\}$, and Theorem 4.1 we have that there exists $\theta_m = (\theta_{m,1}, \dots, \theta_{n,m}) \in \Theta_m^n$ and $\varepsilon > 0$ satisfying

$$\|v_{n,m,\varepsilon}(\theta_m) - v_n\|_{C(\overline{\Omega})} + \|\nabla v_{n,m,\varepsilon}(\theta) - \nabla v_n\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant \frac{\delta}{2}, \tag{4.11}$$

$\|v_{n,m,\varepsilon}\|_{W^{1,\infty}(\Omega)} \leqslant L + \delta$, and $v_{n,m,\varepsilon}$ is $(C+\delta)-$semiconcave. Combining (4.10), and (4.11)+ we arrive at

$$\|v_{n,m,\varepsilon}(\theta_m) - v\|_{C(\overline{\Omega})} + \|\nabla v_{n,m,\varepsilon}(\theta) - \nabla v\|_{L^p(\Omega;\mathbb{R}^d)} \leqslant \delta,$$

$\|\nabla v_{n,m,\varepsilon}(\theta_m)\|_{C^1(\overline{\Omega})} \leqslant L + \delta$, and $v_{n,m,\varepsilon}(\theta_m)$ is $(C+\delta)$-semiconcave. $\square$

We return to Theorem 4.1 and observe that it only connects the errors in the norms of $C(\overline{\Omega})$ and $W^{1,p}(\Omega)$, with $p \in [1, \infty)$. However, for some applications it can be important to control the $W^{1,\infty}(\Omega)$ error. Since the function v_n is only Lipschitz continuous it is not possible to obtain convergence in $W^{1,\infty}(\Omega)$ of $v_{n,m}$, since such a convergence would imply that v_n is $C^1(\Omega)$. Nevertheless, we can still identify regions where uniform convergence of the gradients holds. This involves the set of active indices. For a given $x \in \Omega$, these are the indices for which $\min_{i=1,\dots,n} \phi_i(x)$, respectively $\min_{i=1,\dots,n} \phi_{i,m}(x)$, are achieved.

Theorem 4.3. *Let $\{\phi_{i,m}\}_m \subset C^1(\overline{\Omega})$ be sequences of functions converging to ϕ_i in $C^1(\overline{\Omega})$ for $i = 1, \dots, n$. Further set $v_n = \min_{i=1,\dots,n} \phi_i$, $v_{n,m} = \min_{i=1,\dots,n} \phi_{i,m}$, and introduce the sets of active indices for each $x \in \overline{\Omega}$*

$$I_n(x) = \{i : \phi_i(x) = v_n(x)\},\ I_{n,m}(x) = \{i : \phi_{i,m}(x) = v_{n,m}(x)\},$$

and for $\delta > 0$ the sets

$$\Omega_\delta = \{x \in \overline{\Omega} : v_n(x) \leqslant \phi_j(x) - \delta \text{ for all } j \notin I_n(x)\}.$$

Then, for all $m \in \mathbb{N}$ such that $\|v_n - v_{n,m}\|_{C(\overline{\Omega})} < \frac{\delta}{2}$ we have that $I_{n,m}(x) \subset I_n(x)$, for $x \in \Omega_\delta$, and

$$\|\nabla v_n - \nabla v_{n,m}\|_{L^\infty(\Omega_\delta;\mathbb{R}^d)} \leqslant \sup_{i=1,\dots,n} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)} . \tag{4.12}$$

Additionally, defining $v_{n,\varepsilon} = \psi_{n,\varepsilon}(\phi_1, \dots, \phi_n)$ and $v_{n,m,\varepsilon} = \psi_{n,\varepsilon}(\phi_{1,m}, \dots, \phi_{n,m})$, if (3.7) holds, then we have for $\varepsilon \in \left(0, \frac{\delta}{2(n-1)}\right)$ that

$$\|\nabla v_n - \nabla v_{n,\varepsilon}\|_{L^\infty(\Omega_\delta;\mathbb{R}^d)} \leqslant 2L(g'_\varepsilon(0) + (1 - (g'_\varepsilon(\delta)^{(n-1)}))) \tag{4.13}$$

and

$$\begin{aligned} &\|\nabla v_n - \nabla v_{n.m,\varepsilon}\|_{L^\infty(\Omega_\delta;\mathbb{R}^d)} \leqslant 2L(g'_\varepsilon(0) + (1 - (g'_\varepsilon(\delta)^{(n-1)})))+ \\ &2n(n-1)L \left\|g''_\varepsilon\right\|_{L^\infty(\mathbb{R})} \sup_{i=1,\dots,n} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})} + \sup_{i=1,\dots,n} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)} . \end{aligned} \tag{4.14}$$

where

$$L = \sup_{i\in\{1,\dots,n\}} \|\nabla\phi_i\|_{C(\overline{\Omega};\mathbb{R}^d)} .$$

Proof. We first verify that $I_{n,m}(x) \subset I(x)$ for $x \in \Omega_\delta$. For $m \in \mathbb{N}$ let us set

$$\delta_m := \sup_{i=1,\dots,n} \|\phi_{i,m} - \phi_i\|_{C(\overline{\Omega})} .$$

Let $m \in \mathbb{N}$ be such that $\delta_m \leqslant \frac{\delta}{2}$ and consider $x \in \Omega_\delta$ and $i \in I_{n,m}(x)$. Then we have that

$$v_{n,m}(x) = \phi_{i,m}(x) \geqslant \phi_i(x) - \delta_m.$$

Proceeding by contradiction, if $i \notin I_n(x)$, then $v_n(x) \leqslant \phi_i(x) - \delta$. Combining these two inequalities we obtain

$$v_n(x) \leqslant v_{n,m}(x) - \delta + \delta_m.$$

This implies that $v_{n,m}(x) - v_n(x) \geqslant \delta - \delta_m$ and hence

$$|v_{n,m}(x) - v_n(x)| \geqslant \delta - \delta_m \geqslant \frac{\delta}{2}.$$

On the other hand, we also know that $|v_n(x) - v_{n,m}(x)| < \frac{\delta}{2}$, which leads to a contradiction. Thus $I_{n,m}(x) \subset I_n(x)$ for $x \in \Omega_\delta$ follows.

Turning to (4.12), let us first observe that Ω_δ is cloased and there measurable. Further, both functions v_n and $v_{n,m}$ are a.e. differentiable in Ω_δ. Let x be an arbitrary element of Ω_δ where both functions are differentiable. Choose $i \in I_{n,m}(x)$. Then by the previous step we have $i \in I_n(x) \cap I_{n,m}(x)$. This implies that $v_n(x) = \phi_i(x)$ and $v_{n,m}(x) = \phi_{i,m}(x)$. By Theorems 3.2.13 and 3.2.2 in [23, Part I, Section 3.1, pg.37,47], we have $\nabla v_n(x) = \nabla\phi_i(x)$ and $\nabla v_{n,m}(x) = \nabla\phi_{i,m}(x)$. Furthermore, for $j \in I_{n,m}(x)$

$$\nabla v_{n,m}(x) = \nabla\phi_{j,m}(x) \tag{4.15}$$

holds. Then it is clear that

$$|\nabla v_n(x) - \nabla v_{n,m}(x)| \leqslant \sup_{j=1,\dots,n} \|\nabla\phi_j - \nabla\phi_{j,m}\|_{L^\infty(\Omega;\mathbb{R}^d)} .$$

Since v_m and v are differentiable a.e. the above equality holds a.e. in Ω_δ and consequently

$$\|\nabla v_n - \nabla v_{n,m}\|_{L^\infty(\Omega_\delta;\mathbb{R}^d)} \leqslant \sup_{i=1,\dots,n} \|\nabla \phi_i - \nabla \phi_{i,m}\|_{L^\infty(\Omega;\mathbb{R}^d)},$$

and thus (4.12) is satisfied.

In order to prove (4.13) we will first prove the following:

$$|1 - p_{n,\hat{i},\varepsilon}(\Phi_n(x))| \leqslant g'_\varepsilon(0) + \left(1 - g'_\varepsilon\left(\frac{\delta}{2}\right)^{n-1}\right), \tag{4.16}$$

for $x \in \Omega_\delta$ and $\hat{i} = \hat{i}_n(\Phi_n(x))$. For this purpose, from (3.5) we deduce

$$\begin{aligned} |1 - p_{n,\hat{i},\varepsilon}(\Phi_n(x))| &= 1 - \prod_{j=\hat{i}+1}^{n} g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)) \left(1 - g'_\varepsilon(\phi_{\hat{i}}(x) - v_{\hat{i}-1,\varepsilon}(x))\right) \\ = 1 - \prod_{j=\hat{i}+1}^{n} g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)) &+ g'_\varepsilon(\phi_{\hat{i}}(x) - v_{\hat{i}-1,\varepsilon}(x)) \prod_{j=\hat{i}+1}^{n} g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)). \end{aligned} \tag{4.17}$$

We analyze separately each of the terms in rightmost expression in the previous equality. Since $\hat{i} \in I_n(\Phi_n(x))$ we know that

$$\phi_{\hat{i}}(x) \leqslant v_{\hat{i}-1}(x).$$

Additionally, since (3.7) holds, we can use Lemma 3.1 in the previous inequality to deduce that

$$\phi_{\hat{i}}(x) \leqslant v_{\hat{i}-1,\varepsilon}(x).$$

Since g'_ε is monotonically increasing, we have

$$g'_\varepsilon(\phi_{\hat{i}}(x) - v_{\hat{i}-1,\varepsilon}(x)) \leqslant g'_\varepsilon(0). \tag{4.18}$$

From this, and using (2.12b) we can bound the last term in (4.17) and obtain

$$g'_\varepsilon(\phi_{\hat{i}}(x) - v_{\hat{i}-1,\varepsilon}(x)) \prod_{j=\hat{i}+1}^{n} g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)) \leqslant g'_\varepsilon(0). \tag{4.19}$$

For the remaining term in (4.17), we note that since $\hat{i} = \max\{i \in I_n(\Phi_n(x))\}$ and $x \in \Omega_\delta$, we have that if $j \in \{\hat{i}+1, \dots, n\}$, then

$$\phi_j(x) - v_{j-1}(x) \geqslant \delta.$$

Using (2.16) in the previous inequality and the fact that $\varepsilon \in \left(0, \frac{\delta}{2(n-1)}\right)$ we infer that

$$\phi_j(x) - v_{j-1,\varepsilon}(x) > \frac{\delta}{2}.$$

Combining this with the monotonicity of g'_ε we deduce that

$$g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)) \geqslant g'_\varepsilon\left(\frac{\delta}{2}\right).$$

This implies that

$$1 - \prod_{j=\hat{i}+1}^{n} g'_\varepsilon(\phi_j(x) - v_{j-1,\varepsilon}(x)) \leqslant 1 - g'_\varepsilon\left(\frac{\delta}{2}\right)^{n-\hat{i}} \leqslant 1 - g'_\varepsilon\left(\frac{\delta}{2}\right)^{n-1}. \tag{4.20}$$

Using (4.19) and (4.20) in (4.17), we conclude that (4.16) holds.

Assuming that v_n is differentiable at x, we have $\nabla v_n(x) = \nabla \phi_{\hat{i}}(x)$. Combining this with (3.2) and (4.16) we obtain

$$\begin{aligned}
|\nabla v_{n,\varepsilon}(x) - \nabla v_n(x)| &= \Big|\nabla \phi_{\hat{i}}(x)(1 - p_{n,\hat{i},\varepsilon}(\Phi_n(x))) + \sum_{\hat{i}\neq j} \nabla \phi_j(x) p_{n,j,\varepsilon}(\Phi_n(x))\Big| \\
&\leqslant L(1 - p_{n,\hat{i},\varepsilon}(\Phi_n(x))) + L\sum_{j\neq \hat{i}} p_{n,j,\varepsilon}(\Phi_n(x)) \leqslant 2L(1 - p_{n,\hat{i},\varepsilon}(\Phi_n(x))) \\
&\leqslant 2L\left(g'_\varepsilon(0) + \left(1 - g'_\varepsilon\left(\frac{\delta}{2}\right)^{n-1}\right)\right),
\end{aligned}$$

where in the next to last inequality we used (2.17). Using the almost everywhere differentiability of v_n we obtain (4.13).

To prove (4.14) we will first derive an upper bound of

$$\|\nabla v_{n,m,\varepsilon} - \nabla v_{n,m}\|_{L^\infty(\Omega_\delta)}.$$

For this purpose, we note that by (2.19) we have for all $x \in \overline{\Omega}$

$$\begin{aligned}
&(\nabla v_{n,m,\varepsilon}(x) - \nabla v_{n,\varepsilon}(x))^\top = D\psi_{n,\varepsilon}(\Phi_{n,m}(x))D\Phi_{n,m}(x) - D\psi_{n,\varepsilon}(\Phi_n(x))D\Phi_n(x) \\
&= (D\psi_{n,\varepsilon}(\Phi_{n,m}(x)) - D\psi_{n,\varepsilon}(\Phi_n(x)))D\Phi_n(x) + D\psi_{n,\varepsilon}(\Phi_{n,m}(x))(D\Phi_{n,m}(x) - D\Phi_n(x)).
\end{aligned} \tag{4.21}$$

where D denotes the Jacobian matrix. To bound the first term on the right hand side of (4.21) we have

$$\begin{aligned}
&|D\psi_{n,\varepsilon}(\Phi_{n,m}(x)) - D\psi_{n,\varepsilon}(\Phi_n(x)))D\Phi_n(x)| \leqslant |\nabla\psi_{n,\varepsilon}(\Phi_{n,m}(x)) - \nabla\psi_{n,\varepsilon}(\Phi_n(x))| \cdot |D\Phi_n^\top(x)| \\
&\leqslant n^{\frac{1}{2}} L \left\|\nabla^2 \psi_{n,\varepsilon}\right\|_{L^\infty(\mathbb{R}^n;\mathbb{R}^{n\times n})} |\Phi_{n,m}(x) - \Phi_n(x)| \\
&\leqslant nL \left\|\nabla^2 \psi_{n,\varepsilon}\right\|_{L^\infty(\mathbb{R}^n;\mathbb{R}^{n\times n})} \max_{i\in\{1,\dots,n\}} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})},
\end{aligned} \tag{4.22}$$

as for the second one we have

$$\begin{aligned}
&|D\psi_{n,\varepsilon}(\Phi_{n,m}(x))(D\Phi_{n,m}(x) - D\Phi_n(x))| = \left(\sum_{i=1}^d \left(\sum_{j=1}^n \frac{\partial\psi_{n,\varepsilon}}{\partial a_j}(\Phi_{n,m}(x))\left(\frac{\partial\phi_{j,m}}{\partial x_i}(x) - \frac{\partial\phi_j}{\partial x_i}(x)\right)\right)^2\right)^{\frac{1}{2}} \\
&\leqslant \left(\sum_{i=1}^d \sum_{j=1}^n \frac{\partial\psi_{n,\varepsilon}}{\partial a_j}(\Phi_{n,m}(x))\left(\frac{\partial\phi_{j,m}}{\partial x_i}(x) - \frac{\partial\phi_j}{\partial x_i}(x)\right)^2\right)^{\frac{1}{2}} = \left(\sum_{j=1}^n \frac{\partial\psi_{n,\varepsilon}}{\partial a_j}(\Phi_{n,m}(x))|\nabla\phi_{j,m}(x) - \phi_j(x)|\right)^{\frac{1}{2}} \\
&\leqslant \max_{i\in\{1,\dots,n\}} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)},
\end{aligned} \tag{4.23}$$

where we have used Jensen's inequality. Plugging (4.22) and (4.23) in the right hand side of (4.21) we arrive at

$$\begin{aligned}
&|(\nabla v_{n,m,\varepsilon}(x) - \nabla v_{n,\varepsilon}(x))| \leqslant nL\left\|\nabla^2\psi_{n,\varepsilon}\right\|_{L^\infty(\mathbb{R}^n;\mathbb{R}^{n\times n})} \max_{i\in\{1,\dots,n\}} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})} + \\
&\max_{i\in\{1,\dots,n\}} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)}.
\end{aligned} \tag{4.24}$$

Using (2.19) on the first term in the right hand side of the above inequality we get

$$
\begin{aligned}
&|(\nabla v_{n,m,\varepsilon}(x) - \nabla v_{n,\varepsilon}(x))| \leqslant 2n(n-1)L \left\|g_\varepsilon''\right\|_{L^\infty(\mathbb{R})} \max_{i\in\{1,\dots,n\}} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})} + \\
&\max_{i\in\{1,\dots,n\}} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)} .
\end{aligned}
\tag{4.25}
$$

Finally, using (4.25) and (4.13), we obtain (4.14) by the triangle inequality. □

Remark 4.4. *Let us discuss some consequences of the previous theorem. For this purpose we observe that by* (2.12e) *and* (3.8) *the term* $(g_\varepsilon'(0) + (1 - (g_\varepsilon'(\delta)^{(n-1)})))$ *tends to* 0 *as* $\varepsilon \to 0$. *Moreover let us choose* $\varepsilon_m \to 0^+$ *such that in addition to* (2.12f) *the following holds:*

$$
\lim_{m\to\infty} \left(\left\|g_{\varepsilon_m}''\right\|_{L^\infty(\mathbb{R})} \max_{i\in\{1,\dots,n\}} \|\phi_{i,m} - \phi_i\|_{C(\overline{\Omega})} \right) = 0.
$$

Then as a consequence of Theorem 4.3 we obtain that v_{n,m,ε_m} *converges to* ∇v_n *in* $W^{1,\infty}(\Omega_\delta)$ *for each* $\delta > 0$, *as* $m \to \infty$.

As a second consequence, if we choose $g_\varepsilon = g_{\varepsilon,M}$ *defined in Remark 2.2, then we have that*

$$
g_{\varepsilon,M}'(0) = 0,\ g_\varepsilon'(\delta) = 1 \text{ and } \left\|g''\right\|_{L^\infty(\mathbb{R})} = \frac{1}{\varepsilon},
$$

if $\varepsilon < \delta$. *From this we deduce that under this choice the right hand side of* (4.14) *is bounded by*

$$
2n(n-1)\frac{L}{\varepsilon} \sup_{i\in\{1,\dots,n\}} \|\phi_{i,m} - \phi_i\|_{C(\overline{\Omega})} + \sup_{i\in\{1,\dots,n\}} \|\nabla\phi_{i,m} - \nabla\phi_i\|_{C(\overline{\Omega};\mathbb{R}^d)} . \tag{4.26}
$$

Consequently, for $\varepsilon \in \left(0, \frac{1}{2\delta(n-1)}\right)$ *fixed as specified in Theorem 4.3,* $v_{n,m,\varepsilon}$ *converges to* v_n *in* $W^{1,\infty}(\Omega_\delta)$ *as* $m \to \infty$.

Remark 4.5. *It should be noted that* (4.14) *does not imply convergence in* $C(\Omega_\delta;\mathbb{R}^d)$ *of* $\nabla v_{n,m,\varepsilon}$. *However, noticing that the function* $G(x) = \nabla\phi_{\hat{\imath}_n(x)}(x)$ *is almost everywhere in* Ω *equal to* ∇v_n *and that* G *is continuous in* Ω_δ, *we can infer, by* (4.14), *that*

$$
\begin{aligned}
&\|G - \nabla v_{n.m,\varepsilon}\|_{C(\Omega_\delta;\mathbb{R}^d)} \leqslant 2L(g_\varepsilon'(0) + (1 - (g_\varepsilon'(\delta)^{(n-1)}))) \\
&+2n(n-1)L \left\|g_\varepsilon''\right\|_{L^\infty(\mathbb{R})} \sup_{i=1,\dots,n} \|\phi_i - \phi_{i,m}\|_{C(\overline{\Omega})} + \sup_{i=1,\dots,n} \|\nabla\phi_i - \nabla\phi_{i,m}\|_{C(\overline{\Omega};\mathbb{R}^d)} .
\end{aligned}
\tag{4.27}
$$

The continuity of G *in* Ω_δ *is a consequence of the fact that the active set of incides* I_n *does not change locally, that is, for* $x \in \Omega_\delta$ *we have* $I_n(y) = I_n(x)$ *for all* $y \in B\left(x, \frac{\delta}{2L}\right) \cap \Omega$. *To prove this assertion, we note that if* $|y - x| < \frac{\delta}{2L}$, *then for all* $i \in \{1,\dots,n\}$ *we have* $|\phi_i(x) - \phi_i(y)| < \frac{\delta}{2}$. *By symmetry, it is enough to prove that* $I_n(y) \subset I_n(x)$ *for all* $y \in B(x, \frac{\delta}{2L}) \cap \Omega$. *If* $I_n(y) \not\subseteq I_n(x)$ *for some* $y \in B(x, \frac{\delta}{2L}) \cap \Omega$, *then there exists* $i \in I_n(y) \setminus I_n(x)$ *and thus we obtain:*

$$
v(y) = \phi_i(y) \geqslant \phi_i(x) - \frac{\delta}{2} \geqslant v(x) + \frac{\delta}{2} \geqslant v(y) + \frac{\delta}{2},
$$

which is a contradiction.

5 Example

In this section we introduce an example of a semiconcave function v_d for $d \in \mathbb{N}$, which can be explicitly represented by a family of $2d$ functions of class C^2, satisfies a Hamilton Jacobi Bellman equation, and which is not C^1. Additionally, we present a family of setting $S_m = (\Theta_m, \xi_m)$ for $m \in \mathbb{N}$ which satisfies (4.1). Utilizing this family of settings and the approximation of the positive part $g_\varepsilon = g_{\varepsilon,M}$ for $\varepsilon > 0$, which was introduced in Remark 2.2, we approximate v_d by means of the the parametrization introduced in (2.27).

In subsection 5.2 we verify that the hypotheses of Theorem 4.1 are met and we use it to prove the convergence of the approximation. Furthermore, by means 4.1 we will also prove that the the approximation is converging in $W^{1,\infty}(\Omega_\delta)$ for all $\delta > 0$ and that the approximation satisfies the same Hamilton Jacobi Bellman equation approximately.

In order to highlight the properties of the proposed approximation, we compere it with the approximation resulting from using the Log-Sum-Exp function (see Remark 3.3) which is commonly utilized as a smooth approximation of the minimum. In contrast to the proposed approximation, the Log-Sum-Exp approximation is not able to deal with the discontinuities of the gradient and it struggles to solve the Hamilton Jacobi Bellman equation. To support this numerically, in subsection 5.3 we implement both approximations and measure theirs convergences. Remarkably, these experiments suggest that the proposed approximation solves the Hamilton Jacobi Bellman equation approximately in a uniform sense.

5.1 Exponential Distance Function

In order to illustrate the properties of the proposed parametrization we introduce the Exponential Distance Function:

$$x \in \overline{\Omega} \mapsto v_d(x) := \min_{i \in \{1,\dots,2d\}} \phi_i(x) \tag{5.1}$$

with $\Omega = (-1,1)^d$ and

$$\phi_i(x) = \exp\left(-\frac{1}{2}|x - e_i|^2\right) \text{ and } \phi_{d+i}(x) = \exp\left(-\frac{1}{2}|x + e_i|^2\right) \tag{5.2}$$

for $i \in \{1,\dots,d\}$ where e_i is the i-th canocial vector of $\mathbb{R}^d$.

Clearly, the function v_d is semiconcave thanks to (2.1) and it is a viscosity solution of the following Hamilton Jacobi Bellman equation:

$$H(\nabla\phi(x), \phi(x)) = 0 \text{ for all } x \in \mathbb{R}^d. \tag{5.3}$$

for the Hamiltonian $H : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ given by

$$H(p,a) = \begin{cases} |p|^2 + 2\log(|a|)a^2 & \text{if } a \neq 0 \\ |p|^2 & \text{if } a = 0 \end{cases}, \text{ for } (p,a) \in \mathbb{R}^d \times \mathbb{R}.$$

Below we provide the proof.

Lemma 5.1. *For each $i \in \{1,\dots,2d\}$ the function ϕ_i satisfies*

$$H(\nabla\phi_i(x), \phi_i(x)) = 0 \textit{ for all } x \in \mathbb{R}^d, \tag{5.4}$$

and v_d is a viscosity solution of (5.3).

Proof. We start by noting that an easy calculation proves that for each $i \in \{1,\dots,n\}$, the function ϕ_i satisfies (5.4). Additionally, fixing $x \in \mathbb{R}^d$ with v_d differentiable at x, we have that $v_d(x) = \phi_i(x)$ and $\nabla v_d(x) = \nabla\phi_i(x)$ for some $i \in \{1,\dots,n\}$ and hence we have that v_d satsifies (5.3) at x. Then, due to the a.e. differentiability of v_d, we have that v_d satisfies (5.3) a.e. and thus by Proposition 5.3.1 in [6] and the semiconcavity of v_d, we have that v_d is a viscosity solution of (5.3). □

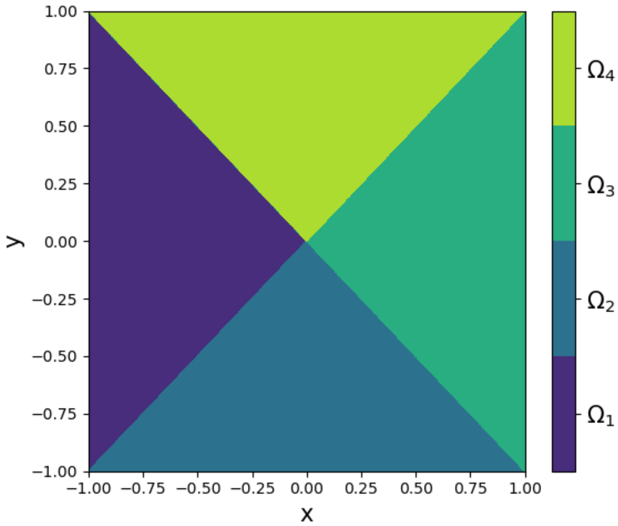
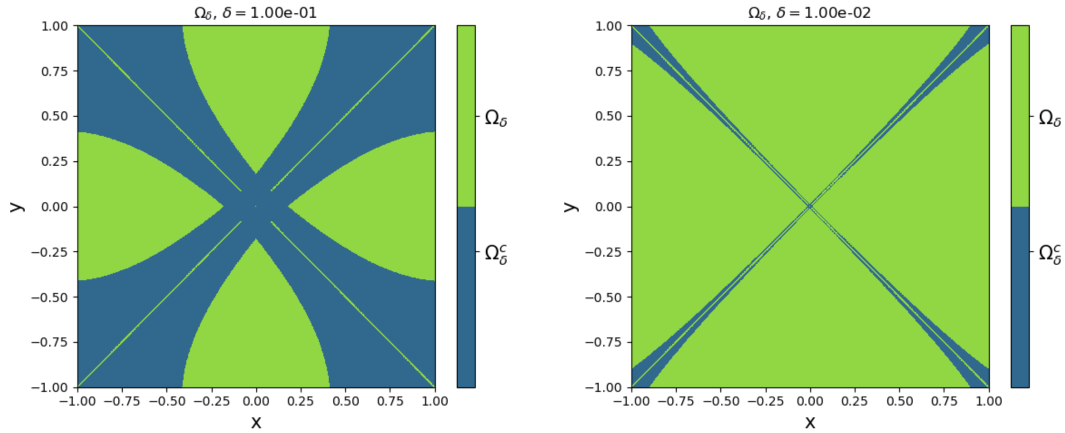


Figure 2: Active sets



(a) $\delta = 10^{-1}$

(b) $\delta = 10^{-2}$

Figure 3: Ω_δ sets of v_d.

In the remainder of this section, we use $d = 2$ for the ease of the exposition, since it allows to depict the active sets and the set Ω_δ. In Figure 2 each active set is colored differently. From this we see that the discontinuities occurs along the diagonals of the square $[-1, 1]^2$. Additionally, in Figure 3, the set Ω_δ is shown for $\delta \in \{10^{-2}, 10^{-1}\}$. As we see in Figure 3, the sets Ω_δ include part of the diagonals of the square $[-1, 1]^2$, which is where the discontinuities of ∇v_d occurs. To see this, we note that along the diagonals of $[-1, 1]^2$ there is more than one active indices and $\nabla\phi_i \neq \nabla\phi_j$ for $i \neq j$. For example, if $x_1 = x_2$ and $x_1 \in [0, 1]$, then $\phi_3(x_1, x_2) = \phi_4(x_1, x_2)$ and it can be verified that $\nabla v_3(x_1, x_2) \neq \nabla v_4(x_1, x_2)$.

5.2 Setting and approximation

For $m \in \mathbb{N}$ we consider Chebyshev polynomials as setting, that is:

$$\Theta_m = \mathbb{R}^{(m+1)\times(m+1)} \text{ and } \xi_m(\theta)(x) = \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} \theta_{i,j} T_{i-1}(x_1) T_{j-1}(x_2) \text{ for } x \in [-1, 1]^2 \text{ and } \theta \in \Theta_m,$$

where T_k is the $k-$th Chebyshev polynomial for $k \in \mathbb{N}_0$. By the results in [24, Section 4.5.1 in Chapter 4] we see that the setting $S_m = (\Theta_m, \xi_m)$ satisfies Hypothesis 4.1. In particular, for $\phi \in C^\infty([-1,1]^2)$ we have that for coefficients $\theta_m \in \Theta_m$ obtained by interpolating ϕ at the Chebyshev points of degree m (see [24, Section 4.5.1 in Chapter 4]) we have

$$\lim_{m\to\infty} \xi_m(\theta_m) = \phi \text{ in } C^2([-1,1]^2).$$

Concerning the approximation of the positive part, we choose

$$s \in \mathbb{R} \mapsto g_{\varepsilon,M}(s) = \begin{cases} 0 & \text{if } s < 0 \\ \frac{s^2}{2\varepsilon} & \text{if } s \in [0,\varepsilon) \\ s - \frac{\varepsilon}{2} & \text{if } s > \varepsilon. \end{cases}$$

According to Remarks 2.2 and 3.3, $g_{\varepsilon,M}$ satisfies (2.12), (3.7) and (3.8). We name the resulting approximation of the minimum $\psi_{n,\varepsilon}$ the *MoreauRegMin* since it comes from the Moreau regularization of the positive part.

Defining $\theta_m = (\theta_{1,m}, \dots, \theta_{2d,m})$ with $\theta_{i,m} \in \Theta_m$ as the coefficients obtained by interpolating ϕ_i by Chebychev polynomials of total degree less or equal than m, for $i \in \{1, \dots, 2d\}$ we propose the following approximation for v_d:

$$v_{d,m,\varepsilon} = \mathrm{v}_{2d,m,\varepsilon}(\theta_m),$$

We will refer to this approximation of v_d as the *MoreauRegMin approximation.*

By (4.1) we have

$$\lim_{\varepsilon\to 0^+, m\to\infty} \|v_{d,m,\varepsilon} - v_d\|_{C(\overline{\Omega})} + \|\nabla v_{d,m,\varepsilon} - \nabla v_d\|_{L^1(\Omega);\mathbb{R}^d} = 0.$$

Furthermore, for $\delta > 0$ and $\varepsilon \in (0, \frac{\delta}{2(2d-1)})$ fixed, we have that

$$\lim_{m\to\infty} \|v_{d,m,\varepsilon} - v_d\|_{W^{1,\infty}(\Omega_\delta)} = 0.$$

In particular, due the continuity of H and the fact that $\mathrm{v}_{2d,m,\varepsilon}(\theta_m)$ is of class C^1, we have that

$$\lim_{\varepsilon\to 0^+, m\to\infty} \|H(\nabla v_{d,m,\varepsilon}, v_{d,m,\varepsilon})\|_{C(\overline{\Omega_\delta})} = 0.$$

which in turns implies that

$$\lim_{\varepsilon\to 0^+, m\to\infty} H(\nabla v_{d,m,\varepsilon}(x), v_{d,m,\varepsilon}(x)) = 0$$

for all $x \in \overline{\Omega}$, since clearly

$$\overline{\Omega} = \lim_{\tilde{\delta}\to 0^+} \Omega_{\tilde{\delta}} = \bigcup_{\tilde{\delta}>0} \Omega_{\tilde{\delta}}.$$

It is noteworthy that the last three properties described above are not necessarily satisfied by a general approximation of v_d. For example, using the Log-Sum-Exp functions defined in Definition 3.3, the approximation of v_d given by

$$\tilde{v}_{d,m,\varepsilon} = \tilde{\psi}_{n,\varepsilon}(\xi_{\theta_{1,m}}, \dots, \xi_{\theta_{2d,m}}) \tag{5.5}$$

converges uniformly to v_d as $\varepsilon \to 0^+$ and $m \to \infty$, it is Lipschitz and semiconcave uniformly with respect to $m \in \mathbb{N}$, but its gradient does not converges pointwise as is explained in Definition 3.3 and furthermore $H(\nabla \tilde{v}_{d,m,\varepsilon}, \tilde{v}_{d,m,\varepsilon})$ does not vanishes as $\varepsilon \to 0^+$ and $m \to \infty$ pointwise nor uniformly in Ω_δ for any $\delta > 0$. In particular, if we consider $x = \left(-\frac{1}{2}, -\frac{1}{2}\right)$ we have that $v_d(x) = \phi_1(x) = \phi_2(x) = \exp\left(-\frac{5}{4}\right)$ and

$$\nabla\phi_1(x) = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \frac{\exp\left(-\frac{5}{4}\right)}{2} \neq \begin{pmatrix} 1 \\ 3 \end{pmatrix} \frac{\exp\left(-\frac{5}{4}\right)}{2} = \nabla\phi_2(x),$$

choosing $\varepsilon_m = |\phi_{1,m}(x) - \phi_{2,m}(x)|^{\frac{1}{2}}$ we have

$$\lim_{m\to\infty} \nabla\tilde{v}_{d,m,\varepsilon_m}(x) = \frac{1}{2}(\nabla\phi_1(x) + \nabla\phi_2(x)) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \exp\left(-\frac{5}{4}\right),$$

which by the continuity of H implies that

$$\lim_{m\to\infty} H(\nabla\tilde{v}_{d,m,\varepsilon_m}(x), \tilde{v}_{d,m,\varepsilon_m}(x)) = -\frac{1}{2}\exp(-5) < 0.$$

In the next subsection, we compare numerically the convergence of $v_{d,m,\varepsilon}$ and $\tilde{v}_{d,m,\varepsilon}$ as $\varepsilon \to 0^+$ and $m \to \infty$.

5.3 Numerical experiments

To carry out the numerical experiments we consider different degrees of the Chebyshev interpolation and regularization parameters $\varepsilon > 0$:

$$m \in \{2, 4, 6, 8, 10\} \text{ and } \varepsilon \in \{10^{-4}, 10^{-2}, 10^{-1}\}.$$

We will gauge the convergence of the different approximations of v_d in Ω_δ for $\delta \in \{0, 10^{-1}, 10^{-3}\}$, where in the case $\delta = 0$ we take that $\Omega_\delta = \overline{\Omega}$. For this, we consider a uniform grid $\mathcal{X} = \{(x_i, y_j)\}_{i,j=0}^{1000}$ of $[-1,1]^2$, where $\{x_i\}_{i=0}^{1000}$ is a uniform division of $[-1,1]$. We measure the convergence using the following metrics:

$$D_C(u,\delta) = \frac{1}{|\mathcal{X} \cap \Omega_\delta|} \max_{x\in\mathcal{X}\cap\Omega_\delta} |v_d(x) - u(x)|,$$

$$D_{W1}(u,\delta) = \frac{1}{|\mathcal{X} \cap \Omega_\delta|} \sum_{x\in\mathcal{X}\cap\Omega_\delta} |\nabla v_d(x) - \nabla u(x)|,$$

$$D_{W\infty}(u,\delta) = \max_{x\in\mathcal{X}\cap\Omega_\delta} |\nabla v_d(x) - \nabla u(x)|,$$

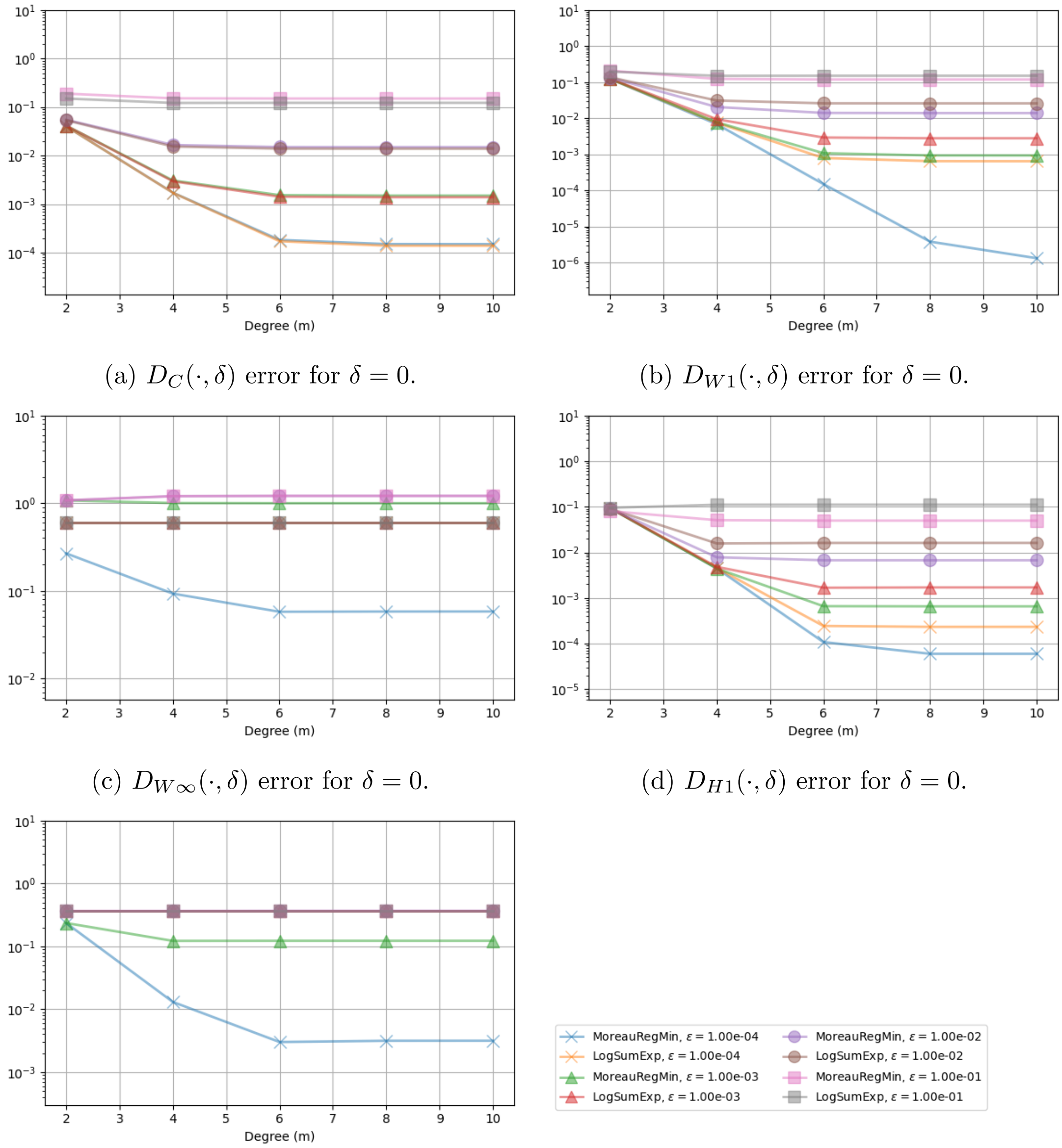
where $|\cdot|$ denotes the cardinality of a set as well as the $\mathbb{R}^n-$norm as appropriate. Additionally, to verify the convergence of the Hamiltonian of the approxmiation we use:

$$D_{H1}(u,\delta) = \frac{1}{|\mathcal{X} \cap \Omega_\delta|} \sum_{x\in\mathcal{X}\cap\Omega_\delta} |H(\nabla u(x), u(x))|,$$

$$D_{H\infty}(u,\delta) = \max_{x\in\mathcal{X}\cap\Omega_\delta} |H(\nabla u(x_i, y_j), u(x))|.$$

δ	10^{-4}	10^{-3}	10^{-2}	10^{-1}
$\frac{\lvert\mathcal{X}\cap\Omega_\delta\rvert}{\lvert\mathcal{X}\rvert}$	1	0.993	0.918	0.416

Table 1: Proportion of $\frac{|\mathcal{X}\cap\Omega_\delta|}{|\mathcal{X}|}$.



(a) $D_C(\cdot, \delta)$ error for $\delta = 0$.

(b) $D_{W1}(\cdot, \delta)$ error for $\delta = 0$.

(c) $D_{W\infty}(\cdot, \delta)$ error for $\delta = 0$.

(d) $D_{H1}(\cdot, \delta)$ error for $\delta = 0$.

(e) $D_{H\infty}(\cdot, \delta)$ error for $\delta = 0$.

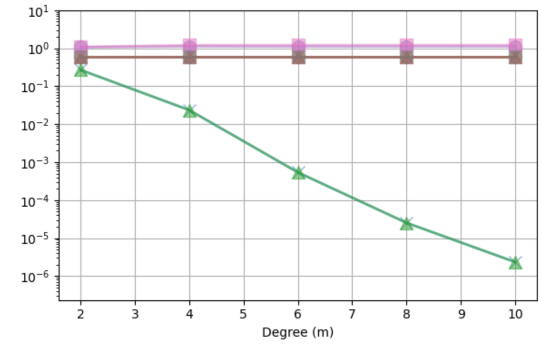
(f) Legend.

Figure 4: Error for $\delta = 0$.

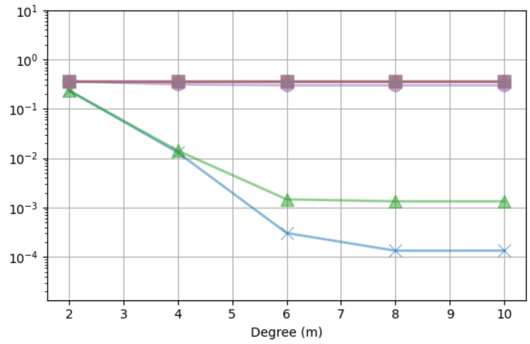
The results are summarized in Table 1, where the proportion of the grid that intersects Ω_δ, for each considered δ is recorded, and in Figures 4-6, where the behavior of the errors for $\delta \in \{0, 10^{-3}, 10^{-2}, 10^{-1}\}$ are depicted. In the following, we discuss these results.

We observe in Figure 3 that Ω_δ covers part of the diagonals of the square. As $\delta \to 0^+$ the proportion of the diagonals covered by Ω_δ increases and for $\delta = 10^{-2}$, at least visually, it seems that this proportion is close to 1. It is worth recalling that the discontinuities of the gradient of v_d occur precisely along the diagonals of the square. In this example, even if δ is not extremely close to 0, we can obtain a good representation of these discontinuities.

Regarding the behavior of the errors in the case of $\delta = 0$, and thus $\Omega_\delta = \overline{\Omega}$, from Figure 4 we see that the behavior of the D_C, $D_{W\infty}$, $D_{H\infty}$ is the same for the MoreuRegMin and the Log-Sum-Exp approximation. The main differences arise from the choice of ε. In contrast, in the case of D_{H1} and D_{W1}, independently of ε and the degree, MoreuRegMin achieves a better performance in terms of D_{W1}. This is consistent with the fact that MoreuRegMin is able to

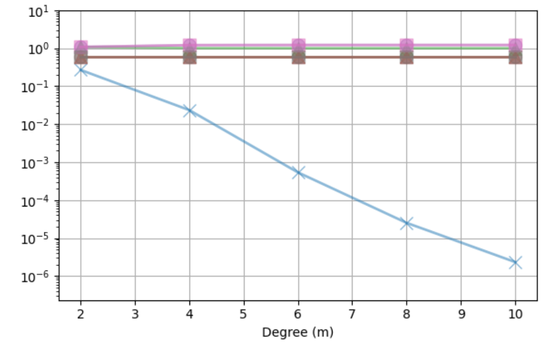


(a) $D_{W\infty}(\cdot,\delta)$ error for $\delta = 10^{-2}$.

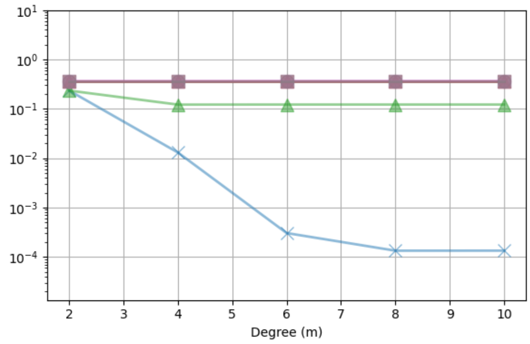


(b) $D_{H\infty}(\cdot,\delta)$ error for $\delta = 10^{-2}$.

Figure 5: Error for $\delta = 10^{-2}$



(a) $D_{W\infty}(\cdot,\delta)$ error for $\delta = 10^{-3}$.



(b) $D_{H\infty}(\cdot,\delta)$ error for $\delta = 10^{-3}$.

Figure 6: Error for $\delta = 10^{-3}$

represent in a better manner the gradient of v_d and viscosity solutions of HJB equations.

For the cases $\delta = 10^{-2}$ and $\delta = 10^{-3}$, D_C, D_{W1} and D_{H1} do not differ from what was already shown in Figure 4. Thus we decided in Figures 5 and 6 to focus on the behavior of $D_{H\infty}$ and $D_{W\infty}$. It is noteworthy that for all ε and degrees, both $D_{W\infty}$ and $D_{H\infty}$ are above 10^{-1} for the Log-Sum-Exp approximation, this indicates that this approximation is not able to approximate the gradient of v_d uniformly. In contrast, we see in Figure 5 that for $\delta = 10^{-2}$, the error $D_{W\infty}$ of the MoreuRegMin approximation decreases with the degree. The cases $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-4}$ behave identically to each other. As explained at the end of Section 4, this is due to the fact that $g'_{\varepsilon,M}(0) = g'_{\varepsilon,M}(\frac{\delta}{2}) = 0$ if $\varepsilon < \delta$. In the case of $D_{H\infty}$, the behavior for $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-4}$ is not identical because the Hamiltonian also receives as input the function itself for which the approximation error depends on ε. In agreement with these observations, in Figure 6 for $\delta = 10^{-3}$ we see that the only case in which the errors $D_{W\infty}$ and $D_{H\infty}$ decrease and achives values below 10^{-1} is when $\varepsilon = 10^{-4}$ for the MoreuRegMin approximation. As in the case of $\delta = 10^{-2}$ this is explained by the fact that $D_{W\infty}$ does not depend on ε if $\varepsilon < \delta$.

6 Conclusions

In this work a smooth, semiconcavity preserving, approximation was introduced. The proposed semiconcavity preserving approximation was devised with two components: a smooth approximation $\psi_{n,\varepsilon}$ of the minimum of $n-$ real-valued functions and a universal approximating parametrization of C^2 functions. The universality of the semiconcavity preserving approximation for semiconcave functions was proved in the $C(\overline{\Omega})$ and $W^{1,p}(\Omega)$ norms for $p \in [1, \infty)$. Further uniform error bounds were presented for the gradient of the proposed approximation in a family of sets converging to Ω, which includes part of the dicontinuities of the gradient. Such a result is not evident since the gradient of a semiconcave function is not necessarily continuous.

By analyzing the limiting behavior of the gradient of $\psi_{n,\varepsilon}$ as $\varepsilon \to 0^+$, the pertinence of this approximation for viscosity solutions of HJB equations was shown in Remark 3.1. It should be noted that although there exist other smooth approximations of the minimum function such as the Log-Sum-Exp, they do not necessarily exhibit the same behavior.

To illustrate the results of this article, in Section 5 the exponential distance function was introduced. It is semiconcave and satisfies a HJB equation. By means of Chebyshev polynomials, a setting satisfying Hypothesis 4.1 was presented in Subsection 5.2. The uniform convergence of the Hamiltonians in Ω_δ towards 0 was proved for this approximation, and for ε small enough, the numerical results using the MoreauRegMin framework confirm the behavior. On the other hand, the Log-Sum-Exp approximation is not able to achieve the same results.

To conclude, we expect to utilize the proposed parametrization for devising new machine learning based methodologies for the resolution of HJB equations and the design of optimal feedback laws in future work. For the aforementioned task, it has been proved in [20] that the semiconcavity of the approximation plays an important role for ensuring its performance. In addition, we believe that the analysis of the parametrization developed in this article can help to understand under which conditions one can provide a parametrization which mitigates the so called curse of dimensionality.

References

[1] M. Akian, S. Gaubert, and A. Lakhoua. The max-plus finite element method for solving deterministic optimal control problems: Basic properties and convergence analysis. *SIAM J. Control Optim.*, 47(2):817–848, 2008.

[2] B. Amos, L. Xu, and J. Z. Kolter. Input convex neural networks. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 146–155, New York, 06–11 Aug 2017. PMLR. URL `https://proceedings.mlr.press/v70/amos17b.html`.

[3] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhäuser, Boston, 1997.

[4] S. Boyd and A. Magnani. Convex piecewise-linear fitting. *Optim. Eng.*, 10:1–17, 2009. doi: 10.1007/s11081-008-9045-3.

[5] G. C. Calafiore, S. Gaubert, and C. Possieri. Log-sum-exp neural networks and posynomial models for convex and log-log-convex data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(3):827–838, 2020. doi: 10.1109/TNNLS.2019.2910417.

[6] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. Progress in Nonlinear Differential Equations and Their Appli. Birkhauser, Boston, Feb. 2006.

[7] Y. Chen, Y. Shi, and B. Zhang. Optimal control via neural networks: A convex approach. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=H1MW72AcK7`.

[8] J. Darbon, P. M. Dower, and T. Meng. Neural network architectures using min-plus algebra for solving certain high-dimensional optimal control problems and Hamilton–Jacobi PDEs. *Math. Control Signals Systems*, 35(1):1–44, Mar. 2023.

[9] A. Douglis. Solutions in the large for multi-dimensional non linear partial differential equations of first order. *Ann. I. Fourier*, 15(2):1–35, 1965. URL http://eudml.org/doc/73873.

[10] P. M. Dower, W. M. McEneaney, and H. Zhang. Max-plus fundamental solution semigroups for optimal control problems. In *2015 Proceedings of the Conference on Control and its Applications*, pages 368—-375, 2015.

[11] W. H. Fleming. The cauchy problem for a nonlinear first order partial differential equation. *J. Differ. Equ.*, 5(3):515–530, 1969. ISSN 0022-0396. doi: https://doi.org/10.1016/0022-0396(69)90091-6. URL https://www.sciencedirect.com/science/article/pii/0022039669900916.

[12] B. Gao and L. Pavel. On the properties of the softmax function with application in game theory and reinforcement learning, 2018. URL https://arxiv.org/abs/1704.00805.

[13] S. Gaubert, W. McEneaney, and Z. Qu. Curse of dimensionality reduction in max-plus based approximation methods: Theoretical estimates and improved pruning algorithms. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 1054–1061, 2011.

[14] F. Générau, É. Oudet, and B. Velichkov. Cut locus on compact manifolds and uniform semiconcavity estimates for a variational inequality. *Archive for Rational Mechanics and Analysis*, 246(2–3):561–602, 2022. doi: 10.1007/s00205-022-01821-0.

[15] A. Ghosh, A. Pananjady, A. Guntuboyina, and K. Ramchandran. Max-affine regression with universal parameter estimation for small-ball designs. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2706–2710, 2020. doi: 10.1109/ISIT44484.2020.9174116.

[16] A. Goujon, S. Neumayer, and M. Unser. Learning weakly convex regularizers for convergent image-reconstruction algorithms. *SIAM J. Imaging Sci.*, 17(1):91–115, 2024. doi: 10.1137/23M1565243. URL https://doi.org/10.1137/23M1565243.

[17] B. Hanin. Universal function approximation by deep neural nets with bounded width and relu activations. *Mathematics*, 7(10), 2019. ISSN 2227-7390. doi: 10.3390/math7100992. URL https://www.mdpi.com/2227-7390/7/10/992.

[18] S. N. Kružkov. Generalized solutions of the hamilton-jacobi equations of eikonal type. i. formulation of the problems; existence, uniqueness and stability theorems; some properties of the solutions. *Math. USSR-Sb.*, 27(3):406, apr 1975. doi: 10.1070/SM1975v027n03ABEH002522. URL https://dx.doi.org/10.1070/SM1975v027n03ABEH002522.

[19] K. Kunisch and D. Vásquez-Varas. Convergence of machine learning methods for feedback control laws: averaged feedback learning scheme and data driven methods, 2025. URL https://arxiv.org/abs/2407.18403.

[20] K. Kunisch and D. Vásquez-Varas. Consistent smooth approximation of feedback laws for infinite horizon control problems with non-smooth value functions. *J. Differ. Equ.*, 411:438–477, 2024. ISSN 0022–0396. doi: https://doi.org/10.1016/j.jde.2024.08.010. URL https://www.sciencedirect.com/science/article/pii/S0022039624004923.

[21] E. L.C. and R. Gariepy. *Measure Theory and Fine Properties of Functions*. Chapman and Hall/CRC, New York, revised edition (1st ed.) edition, 2015. doi: 10.1201/b18333.

[22] G. Leoni. *A First Course in Sobolev Spaces*. Graduate studies in mathematics. American Mathematical Soc., Providence, Rhode Island, 2009. ISBN 9780821884157. URL `https://books.google.at/books?id=W3RLWwnY0RkC`.

[23] M. Mäkelä and P. Neittaanmäki. *Nonsmooth Optimization*. World Scientific, New Jersey, 1992. doi: 10.1142/1493. URL `https://www.worldscientific.com/doi/abs/10.1142/1493`.

[24] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*. Springer series in computational mathematics. Springer, Berlin, Germany, 1 edition, Sept. 2008.

[25] L. Rifford. Existence of Lipschitz and semiconcave control-Lyapunov functions. *SIAM J. Control Optim.*, 39(4):1043–1064, 2000. doi: 10.1137/S0363012999356039.

[26] L. Rifford. Semiconcave control-Lyapunov functions and stabilizing feedbacks. *SIAM J. Control Optim.*, 41(3):659–681, 2002. doi: 10.1137/S0363012900375342.

[27] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Birkhäuser, Cham, 2015. ISBN 978-3-319-20828-2. doi: 10.1007/978-3-319-20828-2.

[28] X. Warin. The groupmax neural network approximation of convex functions. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):11608–11612, 2024. doi: 10.1109/TNNLS.2023.3240183.