

Lower- and upper- C^2 integrands: a structured class of functions for stochastic nonconvex optimization

Felipe Atenas¹ Alejandro Jofré^{1,2,*} Pedro Pérez-Aros^{1,2} David Torregrosa-Belén³

¹Centro de Modelamiento Matemático (CNRS IRL 2807), Universidad de Chile, Santiago, Chile

²Departamento de Ingeniería Matemática, Universidad de Chile, Santiago, Chile

³Department of Mathematics, University of Alicante, Alicante, Spain

fatenas@cmm.uchile.cl; ajofre@uchile.cl; pperez@dim.uchile.cl; david.torregrosa@ua.es

*Corresponding author

Abstract

We study nonconvex stochastic functions that admit a structured second-order representation, a generalization of lower- and upper- C^2 functions to normal integrands. Similarly to the deterministic case, lower- C^2 normal integrands are defined as the pointwise maxima over a compact random index set of functions that are twice-continuously differentiable in the decision variable with an integrable bounded Hessian. Our main result is a local value function characterization of this class of functions, namely, that they satisfy a local sufficient ascent property almost surely. Equivalently, they differ from a convex normal integrand by a random quadratic. The upper- C^2 theory follows by a change of sign, and delivers the corresponding lower envelopes, the form in which these integrands appear as second-stage value functions of stochastic programs with compact recourse. We also show that for this class integrability of the function values, of the local Lipschitz moduli, and of the minimum-norm Clarke subgradients are equivalent. We then study smooth integrands, finite lower envelopes, difference-of-convex integrands, Moreau and forward-backward envelopes, compositions with a smooth map, monotone convex compositions, and compact-recourse stochastic programs within these classes of nonconvex stochastic functions. As an application, we show that the optimistic loss of partial-label regression and the variational formulation of the conditional value-at-risk fit within this framework.

Keywords: Normal integrands, Upper- C^2 functions, Lower- C^2 functions, Clarke subdifferential, Stochastic programming, Value functions, Measurable multifunctions

MSC Classification: 49J52, 49J53, 90C15, 90C30

1 Introduction

Many stochastic optimization models involve expected-value functionals of the form

$$x \mapsto \mathbf{E}_\varphi(x) := \int_{\Xi} \varphi(\xi, x) \mathbb{P}(d\xi), \quad (1)$$

where $(\Xi, \mathcal{F}, \mathbb{P})$ is a probability space and φ is a normal integrand. Such functionals play a central role in stochastic programming, variational analysis, and sample-average approximation; see, e.g., [23–25]. A persistent difficulty is that the sample functions $\varphi_\xi := \varphi(\xi, \cdot)$ may fail to be smooth even when the underlying model data are smooth. This phenomenon arises naturally in two-stage and recourse models: the first-stage objective is defined through the optimal value of the second-stage problem, and such value functions are typically nondifferentiable in the first-stage variable.

The purpose of this work is to formulate and study structured classes of nonconvex functions at the level of normal integrands. In particular, we focus on stochastic versions of lower- and upper- C^2 functions (see, e.g., [23, Definition 10.29]) which enjoy a favorable second-order structure. More

precisely, a locally Lipschitz function is lower- C^2 if, locally, it can be represented as a maximum of C^2 functions. The upper- C^2 property is its natural counterpart, obtained by changing sign. The classes of lower- and upper- C^2 functions play a prominent role in nonsmooth and variational analysis: convex functions provide basic examples of lower- C^2 functions, whereas lower envelopes and value functions obtained through minimization typically lead to upper- C^2 behavior; we refer the reader to [10, 11, 23, 29, 30] for more details. Beyond their theoretical significance, these classes are also relevant to practical optimization problems and to the design and convergence analysis of numerical algorithms. The benign form of nonconvexity of the family of lower- C^2 functions, also known as locally *weakly convex* functions, has been exploited to design suitable proximal algorithms (see, e.g., [2, 7, 16, 26, 31]). In contrast, the upper- C^2 structure has received comparatively less attention in the development of new algorithms (see, e.g., [3–5]). Related developments also include Frank–Wolfe methods for the broader class of upper- $C^{1,\alpha}$ functions [13].

We define the lower- and upper- C^2 classes of normal integrands as the functions that can be obtained by optimizing, over a compact random index set, an objective that is twice-continuously differentiable with respect to the decision variable and whose Hessian is dominated by an integrable random modulus (see Definition 2.4). In the lower- C^2 case, this holds via maximization, and, after the change of sign, via minimization in the upper- C^2 case. This formulation is useful in applications because it reduces the verification of lower- or upper- C^2 regularity to standard hypotheses: compactness of the index (recourse) set, Carathéodory regularity, and an integrable second-order bound. In this way, the result separates the geometric content of the lower- and upper- C^2 property from the measurability issues handled by the theory of normal integrands and measurable multifunctions.

Our main contribution is a local “ascent lemma” (resp. “descent lemma”) for lower- C^2 (resp. upper- C^2) normal integrands. Moreover, we also prove that a normal integrand φ_ξ is lower- C^2 around a reference point with integrable curvature if and only if it is locally characterized by a stochastic upper envelope of quadratic-affine functions with an integrable quadratic coefficient. The mirror result states that lower envelopes of stochastic quadratic-affine functions are the upper- C^2 normal integrands. In particular, the latter have the form of the integrands that appear as second-stage value functions in stochastic programming.

In the stochastic setting, the quadratic local curvature modulus of lower- and upper- C^2 functions may naturally depend on the random parameter ξ , an essential dependence in applications. We therefore require it to belong to an appropriate space $L^p_{\mathbb{P}}$. We show that this integrability requirement is mild enough to cover random value function constructions, but strong enough to transfer samplewise estimates to the expected functional (1). In particular, it provides a convenient bridge between the local geometry of the sample functions and the analytic properties of the expected objective.

The paper also records how the integrability hypotheses interact with one another. For lower- and upper- C^2 normal integrands we show that p -integrability of the values on a neighborhood of a point, p -integrability of the local Lipschitz modulus, and p -integrability of the minimum-norm Clarke subgradient are equivalent, and that each of them needs to be checked at a single point only. We further relate these conditions to integrability assumptions commonly used in smooth stochastic optimization, such as bounded-variance and multiplicative-noise estimates.

The paper is organized as follows. Section 2 fixes notation and recalls the required tools from normal-integrand theory and variational analysis. Section 3 presents equivalent characterizations of lower- and upper- C^2 normal integrands. Section 4 collects important general classes of such integrands, including difference-of-convex functions, Moreau and forward–backward envelopes, and composite functions. Section 5 presents the relationship between integrability of the values, of the Lipschitz moduli, and of the minimum-norm Clarke subgradients of lower- and upper- C^2 normal integrands. In Section 6, we describe two example applications that belong to the families of lower- and upper- C^2 normal integrands.

2 Notation and preliminaries

Throughout the paper, $(\Xi, \mathcal{F}, \mathbb{P})$ denotes a complete probability space. All assertions involving the parameter ξ are understood to hold for $\mathbb{P}$ -almost every $\xi \in \Xi$. After modifying the corresponding objects on a null set if necessary, we shall simply write “for all $\xi \in \Xi$ ”. The space $\mathbb{R}^d$ is endowed with the Euclidean inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$. We denote by $\mathbb{B}_r(x)$ and $\mathbb{B}_r[x]$ the open and closed balls of radius $r > 0$ centered at x , respectively, and set $\mathbb{B} := \mathbb{B}_1[0]$. We denote the extended real line by $\overline{\mathbb{R}} := [-\infty, +\infty]$.

We say that a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is C^1 if it is *continuously differentiable*, that is, differentiable with a continuous gradient, and that it is C^2 if it is twice-continuously differentiable.

For $A \subset \mathbb{R}^d$, the symbols $\text{cl } A$, $\text{int } A$, $\text{bd } A$, $\text{co } A$, and $\text{dist}(x, A)$ denote, respectively, the closure, interior, boundary, convex hull, and distance from x to A . For an extended-real-valued function $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, we denote the domain of f by $\text{dom } f := \{x \in \mathbb{R}^d : f(x) < +\infty\}$ and the epigraph of f by $\text{epi } f := \{(x, \alpha) \in \mathbb{R}^d \times \mathbb{R} : f(x) \leq \alpha\}$. For a set-valued mapping $F : X \rightrightarrows Y$, its graph is denoted by $\text{gph } F := \{(x, y) \in X \times Y : y \in F(x)\}$. For a set $K \subseteq \mathbb{R}^d$, the *indicator function* $\iota_K : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ is defined by $\iota_K(x) := 0$ if $x \in K$ and $\iota_K(x) := +\infty$ otherwise.

For a proper function $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ and $\gamma > 0$, the *Moreau envelope* of f with parameter γ is

$$e_\gamma f(x) := \inf_{z \in \mathbb{R}^d} \left\{ f(z) + \frac{1}{2\gamma} \|z - x\|^2 \right\},$$

and f is said to be *prox-bounded with threshold γ_f* if $e_\gamma f > -\infty$ for every $\gamma \in (0, \gamma_f)$.

2.1 Normal integrands and measurable multifunctions

For $p \in [1, \infty]$, the notation $L^p_{\mathbb{P}}$ refers to the usual Lebesgue space with respect to $\mathbb{P}$. Let X be a metric space. A set-valued mapping $C : \Xi \rightrightarrows X$ is called *measurable* if, for every open set $U \subset X$, $\{\xi \in \Xi : C(\xi) \cap U \neq \emptyset\} \in \mathcal{F}$. When X is a Polish space and C has nonempty compact values, this is the standard measurability notion used for random closed sets and measurable selection arguments.

A mapping $a : \Xi \times X \rightarrow \mathbb{R}^m$ is called a *Carathéodory mapping* if $a(\cdot, t)$ is measurable for every $t \in X$ and $a(\xi, \cdot)$ is continuous for every $\xi \in \Xi$. Similarly, a mapping $g : \Xi \times X \times \mathbb{R}^d \rightarrow \mathbb{R}$ is Carathéodory if $g(\cdot, t, x)$ is measurable for every $(t, x) \in X \times \mathbb{R}^d$ and $g(\xi, \cdot, \cdot)$ is continuous for every $\xi \in \Xi$.

Given an integrand $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, we use the shorthand

$$\varphi_\xi(x) := \varphi(\xi, x), \quad \mathbb{E}_\varphi(x) := \int_{\Xi} \varphi_\xi(x) \mathbb{P}(d\xi),$$

whenever the integral is well-defined. An integrand $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ is called a *normal integrand* if the epigraphical mapping $\xi \mapsto \text{epi } \varphi_\xi$ is measurable and has closed values. All normal integrands considered here are proper, in the sense that $\varphi_\xi > -\infty$ on $\mathbb{R}^d$ and $\text{dom } \varphi_\xi \neq \emptyset$ for every $\xi \in \Xi$. Moreover, φ is a *convex (concave) normal integrand* if, in addition, φ_ξ is convex (concave) for $\mathbb{P}$ -almost every $\xi \in \Xi$.

We shall use the following consequence of measurable maximum theorems. If X is Polish, $C : \Xi \rightrightarrows X$ is measurable with nonempty compact values, and $g : \Xi \times X \times \mathbb{R}^d \rightarrow \mathbb{R}$ is Carathéodory, then the value function

$$\varphi_\xi(x) := \min_{t \in C(\xi)} g(\xi, t, x)$$

is well-defined, the minimum is attained, and the resulting mapping $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a normal integrand (see, e.g., [23, Chapter 14]).

2.2 Clarke subdifferential for locally Lipschitz functions

We adopt generalized differentiation for locally Lipschitz functions; see [10]. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be locally Lipschitz continuous around a given $x \in \mathbb{R}^d$. The *Clarke generalized directional derivative* of f at x in the direction $u \in \mathbb{R}^d$ is

$$f^\circ(x; u) := \limsup_{\substack{y \rightarrow x \\ t \downarrow 0}} \frac{f(y + tu) - f(y)}{t},$$

and the *Clarke subdifferential* of f at x is

$$\partial f(x) := \left\{ v \in \mathbb{R}^d : \langle v, u \rangle \leq f^\circ(x; u) \text{ for all } u \in \mathbb{R}^d \right\}.$$

Throughout the paper, whenever f is locally Lipschitz, the notation $\partial f(x)$ refers to the Clarke subdifferential. If f is locally Lipschitz around x , then $\partial f(x)$ is a nonempty, compact, and convex subset of $\mathbb{R}^d$. Moreover, if f is continuously differentiable around x , then $\partial f(x) = \{\nabla f(x)\}$, while if f is convex finite-valued (and thus locally Lipschitz continuous), the Clarke subdifferential coincides with the usual convex subdifferential.

Moreover, a locally Lipschitz continuous function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be *Clarke regular at x* if, for every direction $u \in \mathbb{R}^d$, the one-sided directional derivative

$$f'(x; u) := \lim_{t \downarrow 0} \frac{f(x + tu) - f(x)}{t}$$

exists and satisfies

$$f'(x; u) = f^\circ(x; u).$$

The function f is Clarke regular on a set D if it is Clarke regular at every point of D . Continuously differentiable functions and finite convex functions are Clarke regular. Notice, however, that Clarke regularity is not invariant under multiplication by -1 . For example, the concave function $x \mapsto -|x|$ is not Clarke regular at the origin. The importance of Clarke regularity in our work is twofold: on one hand, it establishes the equivalence among the Fréchet, limiting (Mordukhovich), and Clarke subdifferentials. On the other hand, it yields tight calculus rules that hold with equality.

The following result is well-known.

Lemma 2.1. Let $V \subset \mathbb{R}^d$ be convex, let $h : V \rightarrow \mathbb{R}$ be continuously differentiable, and let $\kappa \geq 0$ be such that

$$\|\nabla h(z) - \nabla h(w)\| \leq \kappa \|z - w\| \quad \text{for all } z, w \in V. \quad (2)$$

Then, for all $x, y \in V$,

$$|h(y) - h(x) - \langle \nabla h(x), y - x \rangle| \leq \frac{\kappa}{2} \|y - x\|^2. \quad (3)$$

In particular, (2) holds whenever h is twice-continuously differentiable on V with $\|\nabla^2 h(z)\| \leq \kappa$ for all $z \in V$.

2.3 Local Lipschitz moduli and lower- and upper- C^2 integrands

Definition 2.2 (Locally Lipschitz integrand with integrable modulus). Let $p \in [1, \infty]$. A normal integrand $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ is *locally Lipschitz around $\bar{x}$ with p -integrable modulus* if there exist a neighborhood W of $\bar{x}$ and a nonnegative $\kappa \in L^p_{\mathbb{P}}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$, the function φ_ξ is finite on W and

$$|\varphi_\xi(y) - \varphi_\xi(z)| \leq \kappa(\xi) \|y - z\| \quad \text{for all } y, z \in W. \quad (4)$$

When (4) holds for some measurable κ , not necessarily integrable, we simply say that φ is *locally Lipschitz around $\bar{x}$ (with measurable modulus)*. We shall say that these definitions hold on a set $U \subseteq \mathbb{R}^d$ if they hold around every point $\bar{x} \in U$.

Next, we recall that the local Lipschitz modulus of a convex function is controlled by a bound on its values. The following result is the corresponding statement for convex normal integrands, with the bound and the modulus required to be p -integrable.

Proposition 2.3. Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be a convex normal integrand. Let $p \in [1, \infty]$, let $\bar{x} \in \mathbb{R}^d$, and suppose that there exists a convex neighborhood V of $\bar{x}$ with $\varphi(\cdot, x) \in L_{\mathbb{P}}^p$ for all $x \in V$. Then φ is locally Lipschitz around $\bar{x}$ with p -integrable modulus.

Proof. Since $\bar{x} \in \text{int } V$, we may pick affinely independent points $q_0, \dots, q_d \in V$ and $r > 0$ such that $\mathbb{B}_{2r}[\bar{x}] \subseteq \text{co}\{q_i : i = 0, \dots, d\}$, and set

$$\mathbf{m}(\xi) := \max_{i=0, \dots, d} |\varphi_{\xi}(q_i)| + |\varphi_{\xi}(\bar{x})|,$$

which is p -integrable because $\varphi(\cdot, q_i) \in L_{\mathbb{P}}^p$ for $i = 0, \dots, d$ and $\varphi(\cdot, \bar{x}) \in L_{\mathbb{P}}^p$. Let $N \in \mathcal{F}$ be the $\mathbb{P}$ -null set of those ξ for which $\mathbf{m}(\xi) = +\infty$, and fix $\xi \in \Xi \setminus N$. Since φ_{ξ} is convex, its supremum over the simplex $\text{co}\{q_i : i = 0, \dots, d\}$ is attained at a vertex, and thus

$$\varphi_{\xi}(x) \leq \max_{i=0, \dots, d} \varphi_{\xi}(q_i) \leq \mathbf{m}(\xi) + \varphi_{\xi}(\bar{x}) \quad \text{for all } x \in \mathbb{B}_{2r}[\bar{x}]; \quad (5)$$

that is, φ_{ξ} is bounded above on $\mathbb{B}_{2r}[\bar{x}]$ by $\varphi_{\xi}(\bar{x}) + \mathbf{m}(\xi)$. Applying [33, Corollary 2.2.12] on the ball $\mathbb{B}_{2r}[\bar{x}]$, with excess $\mathbf{m}(\xi)$ of this upper bound over the value $\varphi_{\xi}(\bar{x})$ at the center, we obtain

$$|\varphi_{\xi}(x) - \varphi_{\xi}(y)| \leq \frac{3}{2r} \mathbf{m}(\xi) \|x - y\| \quad \text{for all } x, y \in \mathbb{B}_r(\bar{x}),$$

which ends the proof. $\square$

For an integrand $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ that is locally Lipschitz around x , we write

$$\partial\varphi_{\xi}(x) := \partial(\varphi_{\xi})(x).$$

We next define the classes of normal integrands of interest in this work.

Definition 2.4 (Lower- and upper- C^2 normal integrands). Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be a normal integrand, let $\bar{x} \in \mathbb{R}^d$, and let $p \in [1, \infty]$. We say that φ is a *lower- C^2 normal integrand around $\bar{x}$ with p -integrable (curvature) modulus* if there exist a compact convex neighborhood V of $\bar{x}$, a nonnegative function $\varrho \in L_{\mathbb{P}}^p$, a Polish space X , a measurable set-valued mapping $C : \Xi \rightrightarrows X$ with nonempty compact values, and a Carathéodory mapping $g : \Xi \times X \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$,

$$\varphi_{\xi}(x) = \max_{t \in C(\xi)} g(\xi, t, x) \quad \text{for all } x \in V, \quad (6)$$

where, for each $\xi \in \Xi$ and $t \in X$, the function $x \mapsto g(\xi, t, x)$ is twice-continuously differentiable on an open set containing V , the mappings g , $\nabla_x g$, and $\nabla_x^2 g$ are continuous in (t, x) , and

$$\|\nabla_x^2 g(\xi, t, x)\| \leq \varrho(\xi) \quad \text{for all } \xi \in \Xi, t \in C(\xi), x \in V. \quad (7)$$

If (6)–(7) hold for some nonnegative measurable ϱ , not necessarily integrable, we simply say that φ is a *lower- C^2 normal integrand around $\bar{x}$* .

Similarly, φ is an *upper- C^2 normal integrand around $\bar{x}$* (with p -integrable modulus) if $-\varphi$ is a lower- C^2 normal integrand around $\bar{x}$ (with p -integrable modulus). More precisely, this amounts to a representation

$$\varphi_{\xi}(x) = \min_{t \in C(\xi)} g(\xi, t, x) \quad \text{for all } x \in V, \quad (8)$$

with C , g and ϱ as above.

3 Characterizations of lower- and upper- C^2 normal integrands

We next establish a characterization of lower- C^2 normal integrands as local stochastic value functions, from which the upper- C^2 case follows by a change of sign. The first representation uses quadratic-affine functions, while the second uses smooth functions with an integrable Hessian bound.

Theorem 3.1 (Equivalent characterizations of lower- C^2 normal integrands). *Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, let $\bar{x} \in \mathbb{R}^d$, and let $p \in [1, \infty]$. Then the following assertions are equivalent:*

- (a) φ is a lower- C^2 normal integrand around $\bar{x}$ with $L_{\mathbb{P}}^p$ curvature modulus, i.e., Definition 2.4 holds.
- (b) There exist a compact convex neighborhood V of $\bar{x}$ and a nonnegative function $\varrho \in L_{\mathbb{P}}^p$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$, the function φ_{ξ} is locally Lipschitz on a neighborhood of V and

$$\varphi_{\xi}(y) \geq \varphi_{\xi}(x) + \langle v, y - x \rangle - \frac{\varrho(\xi)}{2} \|y - x\|^2 \quad \text{for all } x, y \in V \text{ and all } v \in \partial\varphi_{\xi}(x). \quad (9)$$

- (c) There exist a compact convex neighborhood V of $\bar{x}$, a nonnegative function $\varrho \in L_{\mathbb{P}}^p$, a Polish space X , a measurable set-valued mapping $C : \Xi \rightrightarrows X$ with nonempty compact values, and Carathéodory mappings $a : \Xi \times X \rightarrow \mathbb{R}^d$ and $b : \Xi \times X \rightarrow \mathbb{R}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$ and all $x \in V$,

$$\varphi_{\xi}(x) = \max_{t \in C(\xi)} \left\{ -\frac{\varrho(\xi)}{2} \|x\|^2 + \langle a(\xi, t), x \rangle + b(\xi, t) \right\}. \quad (10)$$

- (d) There exist a compact convex neighborhood V of $\bar{x}$, a nonnegative function $\varrho \in L_{\mathbb{P}}^p$, and a convex normal integrand $f : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$,

$$\varphi_{\xi}(x) = f_{\xi}(x) - \frac{\varrho(\xi)}{2} \|x\|^2 \quad \text{for all } x \in V. \quad (11)$$

Moreover, the modulus ϱ can be taken to be the same in all items (a)–(d).

Proof. We first show that (a) implies (b). Let V , ϱ , X , C and g be as in Definition 2.4. By (6), the compactness of $C(\xi) \times V$, and the continuity of $\nabla_x g(\xi, \cdot, \cdot)$, each φ_{ξ} is Lipschitz on V ; hence, shrinking V if necessary, we may assume that each φ_{ξ} is Lipschitz on a neighborhood of V . Let $\xi \in \Xi$ and $x, y \in V$, and define the set of active indices $C_x^*(\xi) := \operatorname{argmax}_{t \in C(\xi)} g(\xi, t, x)$, which is nonempty by compactness of $C(\xi)$ and continuity of $g(\xi, \cdot, x)$. In view of (6) and the smoothness of $x \mapsto g(\xi, t, x)$ (see, e.g., [23, Theorem 10.31]), for any $v \in \partial\varphi_{\xi}(x)$ there exist $n \geq 1$, $t_i \in C_x^*(\xi)$ and $\lambda_i \geq 0$, $i = 1, \dots, n$, such that $\sum_{i=1}^n \lambda_i = 1$ and

$$v = \sum_{i=1}^n \lambda_i \nabla_x g(\xi, t_i, x). \quad (12)$$

Furthermore, by (7) and Lemma 2.1, for all $i = 1, \dots, n$,

$$\varphi_{\xi}(y) \geq g(\xi, t_i, y) \geq g(\xi, t_i, x) + \langle \nabla_x g(\xi, t_i, x), y - x \rangle - \frac{\varrho(\xi)}{2} \|y - x\|^2.$$

Multiplying by λ_i and summing over $i = 1, \dots, n$ yields

$$\begin{aligned} \varphi_{\xi}(y) &\geq \sum_{i=1}^n \lambda_i g(\xi, t_i, x) + \left\langle \sum_{i=1}^n \lambda_i \nabla_x g(\xi, t_i, x), y - x \right\rangle - \frac{\varrho(\xi)}{2} \|y - x\|^2 \\ &= \varphi_{\xi}(x) + \langle v, y - x \rangle - \frac{\varrho(\xi)}{2} \|y - x\|^2, \end{aligned}$$

where in the last line we used that $t_i \in C_x^*(\xi)$, so that $g(\xi, t_i, x) = \varphi_\xi(x)$, together with (12). This is (9).

Next, we prove (b) implies (c). Let V and $\varrho \in L_{\mathbb{P}}^p$ be as in (b). Set $X := V \times \mathbb{R}^d$. For each fixed ξ , the set

$$C(\xi) := \text{gph } \partial\varphi_\xi \cap X \quad (13)$$

is nonempty and compact. Indeed, $C(\xi)$ is closed due to closedness of V and Lipschitz continuity of φ_ξ on a neighborhood of V , which guarantees closedness of the Clarke subdifferential at the points of V [10, Proposition 2.1.5(b)]. Furthermore, $C(\xi)$ is a subset of a compact set. Let $U \supseteq V$ be an open set on which φ_ξ is locally Lipschitz. For each $x \in V$, there exists an open ball $B_x \subseteq U$ on which φ_ξ is Lipschitz with some constant K_x , and by compactness $V \subseteq B_{x_1} \cup \dots \cup B_{x_m}$ for some $m \in \mathbb{N}$. In view of [10, Proposition 2.1.2], we have

$$\text{gph } \partial\varphi_\xi \cap (V \times \mathbb{R}^d) \subseteq V \times K_\xi \mathbb{B}, \quad K_\xi := \max_{i=1, \dots, m} K_{x_i} < +\infty,$$

which proves the claim. Furthermore, C is a measurable set-valued mapping by standard measurability properties (see, e.g., [23, Theorem 14.56]) with nonempty compact values. For $t = (z, v) \in X$, define

$$a(\xi, t) := v + \varrho(\xi)z, \quad b(\xi, t) := \varphi_\xi(z) - \frac{\varrho(\xi)}{2}\|z\|^2 - \langle v, z \rangle. \quad (14)$$

The mappings a and b are Carathéodory: a is affine in t , and since φ_ξ is continuous on V and φ is a normal integrand, b is continuous in $t = (z, v)$ for each ξ and measurable in ξ for each fixed t .

Fix $\xi \in \Xi$ and $y \in V$. From (9), for every $t = (z, v) \in C(\xi)$,

$$\begin{aligned} \varphi_\xi(y) &\geq \varphi_\xi(z) + \langle v, y - z \rangle - \frac{\varrho(\xi)}{2}\|y - z\|^2 \\ &= -\frac{\varrho(\xi)}{2}\|y\|^2 + \langle v + \varrho(\xi)z, y \rangle + \varphi_\xi(z) - \frac{\varrho(\xi)}{2}\|z\|^2 - \langle v, z \rangle \\ &= -\frac{\varrho(\xi)}{2}\|y\|^2 + \langle a(\xi, t), y \rangle + b(\xi, t). \end{aligned}$$

Taking the maximum over $t \in C(\xi)$ gives one inequality in (10). Conversely, since $\partial\varphi_\xi(y) \neq \emptyset$, pick $v_y \in \partial\varphi_\xi(y)$. Then $t = (y, v_y) \in C(\xi)$ and the corresponding quadratic-affine function equals $\varphi_\xi(y)$ at y . Hence (10) holds.

Now we show that (c) implies (a). Assume that (10) holds and define

$$g(\xi, t, x) := -\frac{\varrho(\xi)}{2}\|x\|^2 + \langle a(\xi, t), x \rangle + b(\xi, t).$$

Then g is Carathéodory, $x \mapsto g(\xi, t, x)$ is twice-continuously differentiable, the mappings g , $\nabla_x g = -\varrho(\xi)x + a(\xi, t)$ and $\nabla_x^2 g = -\varrho(\xi)I$ are continuous in (t, x) , and $\|\nabla_x^2 g(\xi, t, x)\| = \varrho(\xi)$. Hence (6)–(7) hold with the same V , ϱ , X and C , i.e., φ is a lower- C^2 normal integrand around $\bar{x}$ with $L_{\mathbb{P}}^p$ curvature modulus.

To show that (c) implies (d), observe that due to (10), it is enough to consider $f(\xi, x) := \max_{t \in C(\xi)} \{\langle a(\xi, t), x \rangle + b(\xi, t)\}$, which is clearly a finite convex normal integrand.

To conclude the proof, we show (d) implies (b). If (11) holds, then $\varphi_\xi + \frac{\varrho(\xi)}{2}\|\cdot\|^2 = f_\xi$ is convex on V . Shrinking V if necessary, φ_ξ is locally Lipschitz on a neighborhood of V , since f_ξ is convex and finite on a neighborhood of V . For $x, y \in V$ and $v \in \partial\varphi_\xi(x)$ we have $v + \varrho(\xi)x \in \partial f_\xi(x)$ by the smooth sum rule for the Clarke subdifferential, and the convex subgradient inequality gives

$$\varphi_\xi(y) + \frac{\varrho(\xi)}{2}\|y\|^2 \geq \varphi_\xi(x) + \frac{\varrho(\xi)}{2}\|x\|^2 + \langle v + \varrho(\xi)x, y - x \rangle,$$

which upon rearranging is exactly (9). □

We next record the corresponding characterization of upper- C^2 normal integrands, which follows by applying Theorem 3.1 to $-\varphi$. Observe that the characterization states that upper- C^2 normal integrands satisfy a generalization of the so-called *descent lemma* (cf. Lemma 2.1), pivotal in the convergence analysis of optimization algorithms.

Corollary 3.2 (Equivalent characterizations of upper- C^2 normal integrands). Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, let $\bar{x} \in \mathbb{R}^d$, and let $p \in [1, \infty]$. Then the following assertions are equivalent:

- (a) φ is an upper- C^2 normal integrand around $\bar{x}$ with $L_{\mathbb{P}}^p$ curvature modulus, i.e., there exist a compact convex neighborhood V of $\bar{x}$, a nonnegative $\varrho \in L_{\mathbb{P}}^p$, a Polish space X , a measurable set-valued mapping $C : \Xi \rightrightarrows X$ with nonempty compact values, and a Carathéodory mapping $g : \Xi \times X \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying (7) and the smoothness requirements of Definition 2.4, such that

$$\varphi_{\xi}(x) = \min_{t \in C(\xi)} g(\xi, t, x) \quad \text{for all } x \in V. \quad (15)$$

- (b) There exist a compact convex neighborhood V of $\bar{x}$ and a nonnegative function $\varrho \in L_{\mathbb{P}}^p$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$, the function φ_{ξ} is locally Lipschitz on a neighborhood of V and

$$\varphi_{\xi}(y) \leq \varphi_{\xi}(x) + \langle v, y - x \rangle + \frac{\varrho(\xi)}{2} \|y - x\|^2 \quad \text{for all } x, y \in V \text{ and all } v \in \partial \varphi_{\xi}(x). \quad (16)$$

- (c) There exist a compact convex neighborhood V of $\bar{x}$, a nonnegative function $\varrho \in L_{\mathbb{P}}^p$, a Polish space X , a measurable set-valued mapping $C : \Xi \rightrightarrows X$ with nonempty compact values, and Carathéodory mappings $a : \Xi \times X \rightarrow \mathbb{R}^d$ and $b : \Xi \times X \rightarrow \mathbb{R}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$ and all $x \in V$,

$$\varphi_{\xi}(x) = \min_{t \in C(\xi)} \left\{ \frac{\varrho(\xi)}{2} \|x\|^2 + \langle a(\xi, t), x \rangle + b(\xi, t) \right\}. \quad (17)$$

- (d) There exist a compact convex neighborhood V of $\bar{x}$, a nonnegative function $\varrho \in L_{\mathbb{P}}^p$, and a concave normal integrand $f : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$,

$$\varphi_{\xi}(x) = f_{\xi}(x) + \frac{\varrho(\xi)}{2} \|x\|^2 \quad \text{for all } x \in V. \quad (18)$$

Moreover, the modulus ϱ can be taken to be the same in all items (a)–(d).

In the following, we collect several direct consequences of Corollary 3.2, illustrating that the assumptions are easy to check in common stochastic programming constructions.

Example 3.3 (Smooth normal integrands). Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a Carathéodory integrand. Fix $\bar{x} \in \mathbb{R}^d$ and suppose that there exist a compact convex neighborhood V of $\bar{x}$ and nonnegative functions $\kappa \in L_{\mathbb{P}}^2$ and $\varrho \in L_{\mathbb{P}}^1$ such that, for every $\xi \in \Xi$, the function φ_{ξ} is twice-continuously differentiable on (an open set containing) V and

$$\|\nabla_x \varphi_{\xi}(y) - \nabla_x \varphi_{\xi}(z)\| \leq \varrho(\xi) \|y - z\|, \quad |\varphi_{\xi}(y) - \varphi_{\xi}(z)| \leq \kappa(\xi) \|y - z\|$$

for all $y, z \in V$. Then φ is locally Lipschitz around $\bar{x}$ with square-integrable modulus, and both lower- and upper- C^2 around $\bar{x}$ with integrable curvature modulus ϱ , in view of Lemma 2.1 and the identity $\partial \varphi_{\xi}(x) = \{\nabla_x \varphi_{\xi}(x)\}$.

Example 3.4 (Finite lower envelopes of smooth integrands). Let $m \in \mathbb{N}$ and let $f_i : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$, $i = 1, \dots, m$, be Carathéodory integrands. Define

$$\varphi_\xi(x) := \min_{1 \leq i \leq m} f_i(\xi, x).$$

Assume that, around a point $\bar{x}$, each f_i satisfies the hypotheses of Example 3.3 on the same compact convex neighborhood V , with nonnegative curvature moduli $\varrho_i \in L^1_{\mathbb{P}}$ and Lipschitz moduli $\kappa_i \in L^2_{\mathbb{P}}$. Then φ is locally Lipschitz around $\bar{x}$ with square-integrable modulus $\kappa(\xi) := \max_{1 \leq i \leq m} \kappa_i(\xi)$ and upper- C^2 around $\bar{x}$ with curvature modulus

$$\varrho(\xi) := \max_{1 \leq i \leq m} \varrho_i(\xi).$$

Indeed, the conclusion follows from Corollary 3.2(a) by taking $X = \{1, \dots, m\}$ and $C(\xi) = X$. By replacing the minimum with a maximum in the definition of φ_ξ , we obtain a lower- C^2 normal integrand.

Example 3.5 (A nonsmooth DC sample function). Let $A : \Xi \rightarrow \mathbb{R}^{m \times d}$ and $c : \Xi \rightarrow \mathbb{R}^m$ be measurable, and define

$$\varphi_\xi(x) := \frac{1}{2} \|A(\xi)x - c(\xi)\|^2 - \lambda(\xi) \|x\|_1,$$

where $\lambda : \Xi \rightarrow [0, \infty)$ is measurable. On any compact neighborhood V of $\bar{x}$, this has the form $g_\xi - h_\xi$ with

$$g_\xi(x) = \frac{1}{2} \|A(\xi)x - c(\xi)\|^2, \quad h_\xi(x) = \lambda(\xi) \|x\|_1 \quad \text{for } x \in V.$$

If $\|A(\xi)\|^2 \in L^1_{\mathbb{P}}$ and the corresponding local Lipschitz moduli of g and h on V belong to $L^2_{\mathbb{P}}$, then Corollary 4.1 below shows that φ is a locally Lipschitz upper- C^2 normal integrand.

Example 3.6 (Two-stage stochastic programs with compact recourse). Consider a two-stage stochastic program of the form

$$\min_{x \in \mathbb{R}^d} \psi(x) + \int_{\Xi} \varphi_\xi(x) \mathbb{P}(d\xi),$$

with second-stage value function

$$\varphi_\xi(x) := \min_{y \in C(\xi)} G(\xi, y, x).$$

Assume that $C : \Xi \rightrightarrows \mathbb{R}^m$ is measurable with nonempty compact values and that $G : \Xi \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a Carathéodory integrand. Fix $\bar{x} \in \mathbb{R}^d$ and suppose that there exist a compact convex neighborhood V of $\bar{x}$ and a nonnegative $\varrho \in L^1_{\mathbb{P}}$ such that, for every $\xi \in \Xi$ and every $y \in C(\xi)$, the mapping $x \mapsto G(\xi, y, x)$ is twice-continuously differentiable on (an open set containing) V , the mappings $\nabla_x G(\xi, \cdot, \cdot)$ and $\nabla_x^2 G(\xi, \cdot, \cdot)$ are continuous in (y, x) , and

$$\|\nabla_x^2 G(\xi, y, x)\| \leq \varrho(\xi) \quad \text{for all } x \in V.$$

Then Corollary 3.2(a) implies that φ is an upper- C^2 normal integrand around $\bar{x}$ with integrable curvature modulus. If, in addition, φ is locally Lipschitz around $\bar{x}$ with a square-integrable modulus, the reduced expected-value objective

$$x \mapsto \psi(x) + \mathbb{E}_\varphi(x)$$

falls within the upper- C^2 normal-integrand framework developed above, once the deterministic term ψ is handled separately.

4 Archetypal classes of lower- and upper- C^2 normal integrands

In this section we examine how some operations that are standard in optimization, including difference-of-convex decompositions, Moreau envelopes, and compositions, produce normal integrands that are lower- or upper- C^2 in the sense of Definition 2.4. At the level of a single sample function, each of these facts is classical, and the arguments below follow the deterministic ones closely. What is specific to the present setting is a quantitative treatment of the random curvature and Lipschitz moduli that these constructions generate.

In most statements we exhibit the moduli explicitly in terms of the data, and we track how their integrability exponents are produced from those of the data, typically through Hölder's inequality. This allows the samplewise estimates to be transferred to the expected functional $\mathbf{E}_\varphi$, which will be useful in Section 5. It is also where new hypotheses appear, as a construction may be harmless samplewise, but fail to produce an integrable modulus.

We start by showing that the difference of a Lipschitz smooth function and a convex function is an upper- C^2 normal integrand.

Corollary 4.1. Let $p, r \in [1, \infty]$ and let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a normal integrand. Fix $\bar{x} \in \mathbb{R}^d$ and suppose that there are normal integrands $g : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $h : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$, a compact convex neighborhood V of $\bar{x}$, and nonnegative functions $\varrho \in L^p_{\mathbb{P}}$, $\kappa \in L^r_{\mathbb{P}}$ and $\ell \in L^r_{\mathbb{P}}$ such that:

$$\varphi(\xi, x) = g(\xi, x) - h(\xi, x) \quad \text{for all } x \in V \text{ and all } \xi \in \Xi,$$

and

- (i) for every $\xi \in \Xi$, the function $g_\xi := g(\xi, \cdot)$ is continuously differentiable on V and

$$\|\nabla g_\xi(y) - \nabla g_\xi(z)\| \leq \varrho(\xi)\|y - z\| \quad \text{for all } y, z \in V;$$

- (ii) for every $\xi \in \Xi$, the function $h_\xi := h(\xi, \cdot)$ is finite and convex on V ;

- (iii) for all $y, z \in V$ and all $\xi \in \Xi$,

$$|g_\xi(y) - g_\xi(z)| \leq \kappa(\xi)\|y - z\|, \quad |h_\xi(y) - h_\xi(z)| \leq \ell(\xi)\|y - z\|.$$

Then φ is locally Lipschitz around $\bar{x}$ with r -integrable modulus $\kappa + \ell$ and upper- C^2 around $\bar{x}$ with p -integrable curvature modulus ϱ .

Proof. The local Lipschitz estimate follows immediately from (iii). Let $x, y \in V$ and $v \in \partial\varphi_\xi(x)$. Since g_ξ is continuously differentiable and h_ξ is finite convex, the Clarke calculus rules give

$$\partial\varphi_\xi(x) = \nabla g_\xi(x) - \partial h_\xi(x),$$

where $\partial h_\xi(x)$ is the convex subdifferential. Hence there exists $\eta \in \partial h_\xi(x)$ such that $v = \nabla g_\xi(x) - \eta$. By Lemma 2.1,

$$g_\xi(y) \leq g_\xi(x) + \langle \nabla g_\xi(x), y - x \rangle + \frac{\varrho(\xi)}{2}\|y - x\|^2,$$

while convexity of h_ξ gives

$$h_\xi(y) \geq h_\xi(x) + \langle \eta, y - x \rangle.$$

Subtracting the second inequality from the first yields

$$\varphi_\xi(y) \leq \varphi_\xi(x) + \langle v, y - x \rangle + \frac{\varrho(\xi)}{2}\|y - x\|^2,$$

which is (16). Therefore, by Corollary 3.2, we conclude the result. $\square$

Next, we show that under appropriate integrability assumptions, the Moreau envelope of a normal integrand is an upper- C^2 normal integrand.

Corollary 4.2 (Moreau envelope). Let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be a normal integrand and let $\bar{x} \in \mathbb{R}^d$. Let $p \in [2, \infty]$ and $r, s \in [1, \infty]$, and suppose that there exist $\gamma^\varphi > 0$, $\alpha : \Xi \rightarrow \mathbb{R}^d$ p -integrable and $\beta : \Xi \rightarrow \mathbb{R}$ r -integrable such that

$$\varphi_\xi(x) \geq -\frac{1}{2\gamma^\varphi}\|x\|^2 + \langle \alpha_\xi, x \rangle + \beta_\xi \quad \text{for all } x \in \mathbb{R}^d, \xi \in \Xi, \quad (19)$$

and that $\varphi(\cdot, \bar{x}) \in L^s_{\mathbb{P}}$. Then, for every $\gamma \in (0, \gamma^\varphi)$, the function $e_\gamma \varphi$ is a finite-valued upper- C^2 normal integrand around $\bar{x}$ with p' -integrable Lipschitz modulus and q' -integrable curvature modulus, where

$$p' = \min \left\{ \frac{p}{2}, r, s \right\} \quad \text{and} \quad q' = \infty.$$

Proof. Fix $\gamma \in (0, \gamma^\varphi)$ and set $c_\gamma := \frac{1}{\gamma} - \frac{1}{\gamma^\varphi} > 0$. By (19), φ_ξ is prox-bounded with threshold at least γ^φ , so $e_\gamma \varphi_\xi$ is finite-valued and, by [4, Proposition 2.7], for all $x \in \mathbb{R}^d$,

$$e_\gamma \varphi_\xi(x) = \frac{1}{2\gamma}\|x\|^2 - A_\gamma \varphi_\xi(x), \quad A_\gamma \varphi_\xi(x) := \sup_{z \in \mathbb{R}^d} \left\{ \frac{1}{\gamma} \langle z, x \rangle - \varphi_\xi(z) - \frac{1}{2\gamma}\|z\|^2 \right\}. \quad (20)$$

Being a pointwise supremum of affine functions of x , the function $A_\gamma \varphi_\xi$ is convex and lower semi-continuous. Next, using (19), we have that

$$A_\gamma \varphi_\xi(x) \leq \sup_{z \in \mathbb{R}^d} \left\{ \left\langle \frac{x}{\gamma} - \alpha_\xi, z \right\rangle - \frac{c_\gamma}{2}\|z\|^2 \right\} - \beta_\xi = \frac{1}{2c_\gamma} \left\| \frac{x}{\gamma} - \alpha_\xi \right\|^2 - \beta_\xi. \quad (21)$$

In particular, $A_\gamma \varphi_\xi$ is finite, hence continuous. For fixed x , $(\xi, z) \mapsto \varphi_\xi(z) + \frac{1}{2\gamma}\|z - x\|^2$ is a normal integrand [23, Proposition 14.44(d)], so $\xi \mapsto e_\gamma \varphi_\xi(x)$ is measurable [23, Theorem 14.37], and so is $\xi \mapsto A_\gamma \varphi_\xi(x)$ by (20). Hence, $A_\gamma \varphi$ is a Carathéodory integrand, and thus a normal integrand [23, Example 14.29]. Moreover, evaluating the supremum in (20) at $z = \bar{x}$,

$$A_\gamma \varphi_\xi(x) \geq \frac{1}{\gamma} \langle \bar{x}, x \rangle - \varphi_\xi(\bar{x}) - \frac{1}{2\gamma}\|\bar{x}\|^2. \quad (22)$$

Combining (21) and (22), $A_\gamma \varphi_\xi$ is finite on $\mathbb{R}^d$ and $A_\gamma \varphi(\cdot, x) \in L^{p'}_{\mathbb{P}}$ for every $x \in \mathbb{R}^d$. Hence, by Proposition 2.3, $A_\gamma \varphi$ is locally Lipschitz around any point with p' -integrable modulus. Finally, Corollary 4.1 gives the conclusion. $\square$

We next extend the composition rule in [12, Proposition 11 (ii)] to the case of normal integrands. We include it for completeness and to account for the integrability of the moduli.

Proposition 4.3 (Composition rule). Let $r \in [1, \infty]$ and let $p, q \in [r, \infty]$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Let $G : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ be a Carathéodory mapping and let $f : \Xi \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ be a convex normal integrand. Let $\bar{x} \in \mathbb{R}^d$ and let V be a neighborhood of $\bar{x}$, and assume that there exist $M \in L^p_{\mathbb{P}}$ and $L \in L^q_{\mathbb{P}}$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$,

- (i) G_ξ is differentiable on an open convex set containing V , and ∇G_ξ is Lipschitz continuous on V with modulus $L_\xi \geq 0$. Moreover, $\xi \mapsto \|\nabla G_\xi(\bar{x})\|$ belongs to $L^q_{\mathbb{P}}$.
- (ii) f_ξ is finite and Lipschitz continuous with modulus $M_\xi \geq 0$ on an open convex set containing $G_\xi(V)$.

Then $\varphi := f \circ G$, that is $\varphi_\xi := f_\xi \circ G_\xi$, is a lower- C^2 normal integrand around $\bar{x}$ with r -integrable curvature modulus and it is locally Lipschitz around $\bar{x}$ with r -integrable modulus.

Proof. Shrinking V if necessary, we may assume that $V = \mathbb{B}_\delta[\bar{x}]$ for some $\delta > 0$. By (ii), f is finite and continuous in the second variable on a neighborhood of $G_\xi(V)$, hence Carathéodory there. Since G is Carathéodory, so is φ on V , and therefore φ is a finite-valued normal integrand around $\bar{x}$.

Fix $\xi \in \Xi$. By (i), $\|\nabla G_\xi(x)\| \leq \|\nabla G_\xi(\bar{x})\| + L_\xi \delta$ for all $x \in V$, so (ii) gives, for $x, y \in V$,

$$|\varphi_\xi(y) - \varphi_\xi(x)| \leq M_\xi \|G_\xi(y) - G_\xi(x)\| \leq \kappa(\xi) \|y - x\|, \quad \kappa(\xi) := M_\xi (\|\nabla G_\xi(\bar{x})\| + L_\xi \delta).$$

Since $\|\nabla G(\cdot, \bar{x})\| + \delta L(\cdot) \in L^q_{\mathbb{P}}$ and $M \in L^p_{\mathbb{P}}$, the generalized Hölder inequality yields $\kappa \in L^r_{\mathbb{P}}$ with $\|\kappa\|_r \leq \|M\|_p \|\|\nabla G(\cdot, \bar{x})\| + \delta L(\cdot)\|_q$. In particular φ is locally Lipschitz around $\bar{x}$ with r -integrable modulus.

Let $x, y \in V$ and $v \in \partial\varphi_\xi(x)$. The function f_ξ is convex and finite on a neighborhood of $G_\xi(x)$, hence Clarke regular there, and G_ξ is continuously differentiable at x by (i). Clarke's chain rule [10, Theorem 2.3.10] therefore yields

$$\partial\varphi_\xi(x) = \nabla G_\xi(x)^\top \partial f_\xi(G_\xi(x)),$$

so there exists $u \in \partial f_\xi(G_\xi(x))$ with $v = \nabla G_\xi(x)^\top u$. Moreover $\|u\| \leq M_\xi$, because f_ξ is M_ξ -Lipschitz on a neighborhood of $G_\xi(x)$. Convexity of f_ξ then gives

$$\varphi_\xi(y) - \varphi_\xi(x) = f_\xi(G_\xi(y)) - f_\xi(G_\xi(x)) \geq \langle u, G_\xi(y) - G_\xi(x) \rangle.$$

Since V is convex, the fundamental theorem of calculus applied to $t \mapsto G_\xi(x + t(y - x))$ yields

$$G_\xi(y) - G_\xi(x) = \int_0^1 \nabla G_\xi(x + t(y - x))(y - x) dt.$$

Hence, using $v = \nabla G_\xi(x)^\top u$, the Cauchy–Schwarz inequality, (i) and $\|u\| \leq M_\xi$,

$$\begin{aligned} \langle u, G_\xi(y) - G_\xi(x) \rangle &= \langle v, y - x \rangle + \int_0^1 \langle (\nabla G_\xi(x + t(y - x)) - \nabla G_\xi(x))^\top u, y - x \rangle dt \\ &\geq \langle v, y - x \rangle - \int_0^1 L_\xi t \|y - x\| M_\xi \|y - x\| dt \\ &= \langle v, y - x \rangle - \frac{L_\xi M_\xi}{2} \|y - x\|^2. \end{aligned}$$

Consequently, combining the last two estimates, (9) holds on V with curvature modulus $\varrho = LM$, which belongs to $L^r_{\mathbb{P}}$ by the generalized Hölder inequality, with $\|\varrho\|_r \leq \|L\|_q \|M\|_p$. Theorem 3.1 gives the conclusion. $\square$

Given a composite objective $f + g$ with f smooth and g proper, lower semicontinuous and prox-bounded, the *forward-backward envelope* with parameter $\gamma > 0$ is the value function associated with the forward-backward minimization subproblem; see [27, 28]. This makes it a natural merit function for proximal-gradient-type algorithms, and an extension of the Moreau envelope. In the stochastic setting the same construction is performed samplewise: for a Carathéodory integrand f and a normal integrand g we define

$$\varphi_{\gamma, \xi}^{\text{fb}}(x) := \min_{z \in \mathbb{R}^d} \left\{ f_\xi(x) + \langle \nabla f_\xi(x), z - x \rangle + g_\xi(z) + \frac{1}{2\gamma} \|z - x\|^2 \right\}. \quad (23)$$

This function is real-valued and, for γ small enough, shares the minimizers and the optimal value of $f + g$.

The next result shows that, under a uniform prox-boundedness condition on g and a Lipschitz Hessian for f , the family (23) is an upper- C^2 normal integrand with integrable moduli. The second-order regularity of f cannot be weakened to a Lipschitz gradient: for $d = 1$, $g \equiv 0$ and $f'(x) = 1 - |x|$ near the origin, (23) reduces to $f - \frac{\gamma}{2}(f')^2$, which has a convex kink at 0 and is therefore not upper- C^2 there.

Corollary 4.4 (Forward–backward envelope). Let $s \in [1, \infty]$ and let $p, q \in [s, \infty]$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{s}$. Let $\bar{x} \in \mathbb{R}^d$, let $\delta > 0$, set $V := \mathbb{B}_\delta[\bar{x}]$ and $R := \|\bar{x}\| + \delta$. Let $f : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a Carathéodory integrand and let $g : \Xi \times \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ be a normal integrand such that:

- (i) for every $\xi \in \Xi$, f_ξ is twice-continuously differentiable on an open convex set containing V , and

$$\|\nabla^2 f_\xi(x)\| \leq L_\xi, \quad \|\nabla^2 f_\xi(x) - \nabla^2 f_\xi(y)\| \leq \Gamma_\xi \|x - y\| \quad \text{for all } x, y \in V,$$

with $L, \Gamma \in L^q_{\mathbb{P}}$;

- (ii) there exist $\gamma^g > 0$ and measurable $\alpha : \Xi \rightarrow \mathbb{R}^d$, $\beta : \Xi \rightarrow \mathbb{R}$ such that

$$g_\xi(z) \geq -\frac{1}{2\gamma^g} \|z\|^2 + \langle \alpha_\xi, z \rangle + \beta_\xi \quad \text{for all } z \in \mathbb{R}^d, \xi \in \Xi.$$

Fix $\gamma \in (0, \gamma^g)$, set $T_\xi := \text{id} - \gamma \nabla f_\xi$, and assume that the Asplund function $A_\gamma g_\xi$ of (20) is Lipschitz continuous with modulus M_ξ on an open convex set containing $T_\xi(V)$, with $M \in L^p_{\mathbb{P}}$. Then $\varphi_\gamma^{\text{fb}}$ given by (23) is a finite-valued normal integrand which is upper- C^2 around $\bar{x}$ with s -integrable curvature modulus

$$\varrho(\xi) := \frac{1}{\gamma} + L_\xi + R\Gamma_\xi + \gamma\Gamma_\xi M_\xi, \quad (24)$$

and locally Lipschitz around $\bar{x}$ with s -integrable modulus

$$\kappa(\xi) := R\left(\frac{1}{\gamma} + L_\xi\right) + M_\xi(1 + \gamma L_\xi). \quad (25)$$

Proof. Completing the square in (23) gives, for all $x \in \mathbb{R}^d$ (see [20]),

$$\varphi_{\gamma,\xi}^{\text{fb}}(x) = f_\xi(x) - \frac{\gamma}{2} \|\nabla f_\xi(x)\|^2 + e_\gamma g_\xi(T_\xi(x)). \quad (26)$$

By (ii) and the argument in the proof of Corollary 4.2, g_ξ is prox-bounded with threshold at least γ^g , the envelope $e_\gamma g_\xi$ is finite-valued, and (20) holds for g , with $A_\gamma g$ a finite convex Carathéodory integrand. Since ∇f is Carathéodory, (26) shows that $\varphi_\gamma^{\text{fb}}$ is Carathéodory, hence a finite-valued normal integrand.

Substituting (20) into (26) and expanding $\|T_\xi(x)\|^2$, the terms $\pm \frac{\gamma}{2} \|\nabla f_\xi(x)\|^2$ cancel and we obtain

$$\varphi_{\gamma,\xi}^{\text{fb}} = \Phi_\xi - \Psi_\xi \quad \text{on } V, \quad \begin{cases} \Phi_\xi(x) := f_\xi(x) + \frac{1}{2\gamma} \|x\|^2 - \langle x, \nabla f_\xi(x) \rangle, \\ \Psi_\xi := A_\gamma g_\xi \circ T_\xi. \end{cases} \quad (27)$$

On the one hand, by (i), Φ_ξ is continuously differentiable on V with

$$\nabla \Phi_\xi(x) = \frac{1}{\gamma} x - \nabla^2 f_\xi(x) x,$$

so that, for $x, y \in V$,

$$\|\nabla \Phi_\xi(x) - \nabla \Phi_\xi(y)\| \leq \frac{1}{\gamma} \|x - y\| + \|\nabla^2 f_\xi(x) - \nabla^2 f_\xi(y)\| \|x\| + \|\nabla^2 f_\xi(y)\| \|x - y\| \leq \varrho^1(\xi) \|x - y\|,$$

with $\varrho^1(\xi) := \frac{1}{\gamma} + L_\xi + R\Gamma_\xi$. Moreover $\|\nabla \Phi_\xi(x)\| \leq R(\frac{1}{\gamma} + L_\xi)$ on V .

On the other hand, by (i), T_ξ is Lipschitz on V with modulus $1 + \gamma L_\xi$, and continuously differentiable on a neighborhood of V with $\nabla T_\xi = I - \gamma \nabla^2 f_\xi$. Hence $\|\nabla T_\xi\| \leq 1 + \gamma L_\xi$ on V and ∇T_ξ is Lipschitz on V with modulus $\gamma \Gamma_\xi$. Since $1 + \gamma L \in L^q_{\mathbb{P}}$, $\gamma \Gamma \in L^q_{\mathbb{P}}$ and $M \in L^p_{\mathbb{P}}$, Proposition 4.3 applies with inner mapping $G := T$ and outer convex integrand $A_\gamma g$, and yields that Ψ_ξ is Lipschitz on V with modulus $M_\xi(1 + \gamma L_\xi)$ and lower- C^2 on V with s -integrable modulus $\sigma(\xi) := \gamma \Gamma_\xi M_\xi$; that is, $\Psi_\xi + \frac{\sigma(\xi)}{2} \|\cdot\|^2$ is convex on V for every $\xi \in \Xi$.

Next, rewrite (27) as

$$\varphi_{\gamma,\xi}^{\text{fb}} = \left[\Phi_\xi + \frac{\sigma(\xi)}{2} \|\cdot\|^2 \right] - \left[\Psi_\xi + \frac{\sigma(\xi)}{2} \|\cdot\|^2 \right].$$

The first bracket is continuously differentiable on V with gradient Lipschitz of modulus $\rho^1(\xi) + \sigma(\xi) = \varrho(\xi)$, the second is finite and convex on V , and both are Lipschitz on V . Corollary 4.1 then gives the upper- C^2 property with curvature modulus (24). The local Lipschitz modulus (25) follows directly from (27), since Φ_ξ is Lipschitz on V with modulus $R(\frac{1}{\gamma} + L_\xi)$ and Ψ_ξ with modulus $M_\xi(1 + \gamma L_\xi)$. Finally, $\mathbb{P}$ is a probability measure and $q \geq s$, so $L, \Gamma \in L_{\mathbb{P}}^s$, while ΓM and LM belong to $L_{\mathbb{P}}^s$ by the generalized Hölder inequality. Thus, $\varrho, \kappa \in L_{\mathbb{P}}^s$. $\square$

Remark 4.5 (On the Lipschitz assumption on $A_\gamma g$ in Corollary 4.4). Let (ii) of Corollary 4.4 hold, and let $\gamma \in (0, \gamma^g)$. Then, for every set $S \subseteq \mathbb{R}^d$, the function $A_\gamma g_\xi$ is Lipschitz on S with modulus

$$\frac{1}{\gamma} \sup \{ \|z\| : z \in P_\gamma g_\xi(w), w \in S \}, \quad (28)$$

where $P_\gamma g_\xi(w) := \operatorname{argmin}_z \{ g_\xi(z) + \frac{1}{2\gamma} \|z - w\|^2 \}$, $w \in \mathbb{R}^d$. First, observe that $P_\gamma g_\xi(w) \neq \emptyset$ for every $w \in \mathbb{R}^d$. Indeed, setting $c_\gamma := \frac{1}{\gamma} - \frac{1}{\gamma^g} > 0$, for all $z \in \mathbb{R}^d$ we have

$$g_\xi(z) + \frac{1}{2\gamma} \|z - w\|^2 \geq \frac{c_\gamma}{2} \|z\|^2 + \left\langle \alpha_\xi - \frac{w}{\gamma}, z \right\rangle + \beta_\xi + \frac{1}{2\gamma} \|w\|^2. \quad (29)$$

Since $c_\gamma > 0$, the lower semicontinuous function on the left-hand side is level-bounded in z , and thus its set of minimizers $P_\gamma g_\xi(w)$ is nonempty. Next, the supremum defining $A_\gamma g_\xi(w')$ is attained at any $z' \in P_\gamma g_\xi(w')$. Hence

$$A_\gamma g_\xi(w') - A_\gamma g_\xi(w) \leq \frac{1}{\gamma} \langle z', w' - w \rangle \leq \frac{1}{\gamma} \|z'\| \|w' - w\|,$$

and the claim follows by symmetry.

In particular, we have the following two important examples.

- (1) Let $g_\xi = \iota_{K(\xi)}$, the indicator function of $K(\xi)$, where $K : \Xi \rightrightarrows \mathbb{R}^d$ is measurable with nonempty closed values and $K(\xi) \subseteq k_\xi \mathbb{B}$ with $k \in L_{\mathbb{P}}^p$. Then Corollary 4.4 (ii) holds with $\alpha \equiv 0$, $\beta \equiv 0$ and any $\gamma^g > 0$. Since $P_\gamma g_\xi(w)$ is the set of projections of w onto $K(\xi)$, which lie in $K(\xi)$, it follows from (28) that $A_\gamma g_\xi$ is Lipschitz on all of $\mathbb{R}^d$ with modulus $M_\xi = k_\xi/\gamma$.
- (2) Let g_ξ be Lipschitz on $\mathbb{R}^d$ with modulus ℓ_ξ , where $\ell \in L_{\mathbb{P}}^p$ (e.g., $g_\xi = \lambda_\xi \|\cdot\|_1$ and minimax concave penalties [15, 34]). Since $|g_\xi(z) - g_\xi(0)| \leq \ell_\xi \|z\| \leq \frac{1}{2\gamma^g} \|z\|^2 + \frac{\gamma^g}{2} \ell_\xi^2$, Corollary 4.4 (ii) holds with $\alpha \equiv 0$, $\beta_\xi = g_\xi(0) - \frac{\gamma^g}{2} \ell_\xi^2$ and any $\gamma^g > 0$. For $w \in \mathbb{R}^d$ and $z \in P_\gamma g_\xi(w)$,

$$\frac{1}{2\gamma} \|z - w\|^2 \leq g_\xi(w) - g_\xi(z) \leq \ell_\xi \|z - w\|,$$

so

$$\|z\| \leq \|w\| + 2\gamma \ell_\xi \quad \text{for all } w \in \mathbb{R}^d, z \in P_\gamma g_\xi(w). \quad (30)$$

Let $\delta > 0$ and set $K_\xi := (1 + \gamma L_\xi)\delta + \|\bar{x}\| + \gamma \|\nabla f_\xi(\bar{x})\|$. Since T_ξ is Lipschitz on $\mathbb{B}_\delta[\bar{x}]$ with modulus $1 + \gamma L_\xi$ by assumption (i) of Corollary 4.4, for $y = T_\xi(w')$ with $w' \in \mathbb{B}_\delta[\bar{x}]$ we have

$$\|y\| \leq \|T_\xi(w') - T_\xi(\bar{x})\| + \|T_\xi(\bar{x})\| \leq (1 + \gamma L_\xi) \|w' - \bar{x}\| + \|\bar{x}\| + \gamma \|\nabla f_\xi(\bar{x})\| \leq K_\xi,$$

and thus $T_\xi(\mathbb{B}_\delta[\bar{x}]) \subseteq \mathbb{B}_{K_\xi}[0] \subseteq \mathbb{B}_{K_\xi+1}(0)$. Furthermore, (28) and (30) imply that $A_\gamma g_\xi$ is Lipschitz on the open convex set $\mathbb{B}_{K_\xi+1}(0)$ with modulus $M_\xi = (K_\xi + 1)/\gamma + 2\ell_\xi$, which lies in $L_{\mathbb{P}}^p$ provided that $\|\nabla f(\cdot, \bar{x})\|, L \in L_{\mathbb{P}}^p$.

The last result of this section extends Proposition 4.3 to compositions whose inner mapping is no longer smooth, but has lower- C^2 components. Smoothness of the inner mapping is replaced by a monotonicity requirement on the outer convex function, which guarantees that the ascent inequalities of the components are preserved under composition. This is the (stochastic) lower- C^2 counterpart of the classical rule that a nondecreasing convex function of convex functions is convex. It is also the structure of the conditional value-at-risk studied in Section 6, where a nondecreasing convex function is composed with a nonsmooth loss.

Proposition 4.6 (Monotone convex composition). Let $r \in [1, \infty]$ and let $p, q \in [r, \infty]$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Let $g = (g_1, \dots, g_m) : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}^m$, where each g_j is a finite-valued normal integrand, and let $f : \Xi \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a finite-valued convex normal integrand such that, for every $\xi \in \Xi$, f_ξ is nondecreasing with respect to the componentwise order of $\mathbb{R}^m$. Let $\bar{x} \in \mathbb{R}^d$ and assume that there exist a compact convex neighborhood V of $\bar{x}$, nonnegative functions $\varrho_1, \dots, \varrho_m \in L_{\mathbb{P}}^q$ and a nonnegative $M \in L_{\mathbb{P}}^p$ such that, for $\mathbb{P}$ -almost every $\xi \in \Xi$,

- (i) for each $j = 1, \dots, m$, g_j is locally Lipschitz on a neighborhood of V and lower- C^2 around $\bar{x}$, with the representation (6) holding on V with curvature modulus ϱ_j ;
- (ii) f_ξ is Lipschitz continuous with modulus M_ξ on an open convex set containing $g_\xi(V)$.

Then $\varphi := f \circ g$ is a lower- C^2 normal integrand around $\bar{x}$ with r -integrable curvature modulus $\sigma(\xi) := M_\xi \|\varrho(\xi)\|$, where $\varrho := (\varrho_1, \dots, \varrho_m)$. If, in addition, each g_j is Lipschitz on V with modulus $\kappa_j \in L_{\mathbb{P}}^q$, then φ is locally Lipschitz around $\bar{x}$ with r -integrable modulus $M\|\kappa\|$, $\kappa := (\kappa_1, \dots, \kappa_m)$.

Proof. Due to [23, Proposition 14.45(b)], φ is a normal integrand. Moreover, it is locally Lipschitz continuous on a neighborhood of V as the composition of locally Lipschitz maps.

Next, fix $\xi \in \Xi$, $x, y \in V$ and $v \in \partial\varphi_\xi(x)$. By Clarke's chain rule [10, Theorem 2.3.9],

$$v \in \text{co} \left\{ \sum_{j=1}^m \alpha_j \zeta_j : \alpha \in \partial f_\xi(g_\xi(x)), \zeta_j \in \partial(g_j)_\xi(x) \right\},$$

so $v = \sum_{k=1}^N \lambda_k w_k$ for some $N \in \mathbb{N}$, $\lambda_k \geq 0$, $\sum_{k=1}^N \lambda_k = 1$, and $w_k = \sum_{j=1}^m \alpha_j^k \zeta_j^k$ for some $\alpha^k \in \partial f_\xi(g_\xi(x))$ and $\zeta_j^k \in \partial(g_j)_\xi(x)$. By assumption (ii), $\|\alpha^k\| \leq M_\xi$. Using monotonicity and the convex subgradient inequality at $z := g_\xi(x)$, for $t > 0$ and e_j the j -th canonical vector in $\mathbb{R}^m$,

$$f_\xi(z) \geq f_\xi(z - te_j) \geq f_\xi(z) - t\alpha_j^k,$$

so that $\alpha_j^k \geq 0$ for every $j = 1, \dots, m$. Furthermore, by Theorem 3.1, assumption (i) implies that (9) holds on V with modulus $\varrho_j(\xi)$, and thus

$$\begin{aligned} \varphi_\xi(y) - \varphi_\xi(x) &\geq \langle \alpha^k, g_\xi(y) - g_\xi(x) \rangle \geq \sum_{j=1}^m \alpha_j^k \left(\langle \zeta_j^k, y - x \rangle - \frac{\varrho_j(\xi)}{2} \|y - x\|^2 \right) \\ &= \langle w_k, y - x \rangle - \sum_{j=1}^m \frac{\alpha_j^k \varrho_j(\xi)}{2} \|y - x\|^2 \geq \langle w_k, y - x \rangle - \frac{M_\xi \|\varrho(\xi)\|}{2} \|y - x\|^2. \end{aligned}$$

Multiplying by λ_k and summing over $k = 1, \dots, N$ gives (9) on V with modulus $\sigma = M\|\varrho\|$. Furthermore, by the generalized Hölder inequality, it follows that $\|\sigma\|_r \leq \|M\|_p \|\varrho\|_q$, so that $\sigma \in L_{\mathbb{P}}^r$. The conclusion follows from Theorem 3.1. Finally, for $x, y \in V$, $|\varphi_\xi(y) - \varphi_\xi(x)| \leq M_\xi \|g_\xi(y) - g_\xi(x)\| \leq M_\xi \|\kappa(\xi)\| \|y - x\|$, and again $M\|\kappa\| \in L_{\mathbb{P}}^r$ by the Hölder inequality. $\square$

5 Integrability of Lipschitz moduli, values and subgradients

In this section we show that, for lower- and upper- C^2 normal integrands, the integrability of the values, of the Lipschitz moduli, and of the Clarke subgradients are all related. In particular, we

prove some equivalences that make the samplewise estimates of Section 4 usable for the expected functional (1).

The first result extends Proposition 2.3 beyond the convex case. For lower- and upper- C^2 normal integrands, p -integrability of the values on a neighborhood of the reference point is both necessary and sufficient for local Lipschitz continuity with a p -integrable modulus.

Corollary 5.1. Let $p \in [1, \infty]$, let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, and let $\bar{x} \in \mathbb{R}^d$. Assume that φ is either lower- C^2 or upper- C^2 around $\bar{x}$ with p -integrable curvature modulus ϱ . Then the following assertions are equivalent:

- (a) there exists a neighborhood V of $\bar{x}$ such that $\varphi(\cdot, x) \in L_{\mathbb{P}}^p$ for every $x \in V$;
- (b) $\varphi(\cdot, \bar{x}) \in L_{\mathbb{P}}^p$ and φ is locally Lipschitz around $\bar{x}$ with p -integrable Lipschitz modulus.

Proof. Both (a) and (b) are invariant under the change of sign $\varphi \mapsto -\varphi$, which interchanges the lower- and upper- C^2 properties while preserving the curvature modulus. We may therefore assume that φ is lower- C^2 around $\bar{x}$.

Suppose that (a) holds. Shrinking V if necessary, we may assume that it is a compact convex neighborhood of $\bar{x}$ on which, by Theorem 3.1(d), the representation

$$\varphi_{\xi}(x) = f_{\xi}(x) - \frac{\varrho(\xi)}{2} \|x\|^2 \quad \text{for all } x \in V$$

holds with a convex normal integrand f . Since $\varrho \in L_{\mathbb{P}}^p$ and V is bounded, the function $\xi \mapsto \frac{\varrho(\xi)}{2} \|x\|^2$ belongs to $L_{\mathbb{P}}^p$ for every $x \in V$. Hence, $f(\cdot, x) \in L_{\mathbb{P}}^p$ for every $x \in V$. Proposition 2.3 then shows that f is locally Lipschitz around $\bar{x}$ with p -integrable Lipschitz modulus. As $x \mapsto \frac{\varrho(\xi)}{2} \|x\|^2$ is Lipschitz on V with modulus $\varrho(\xi) \sup_{u \in V} \|u\|$, which is again p -integrable, the same is true of φ . Finally, $\varphi(\cdot, \bar{x}) \in L_{\mathbb{P}}^p$ because $\bar{x} \in V$, so (b) holds.

Conversely, suppose that (b) holds, and let V be a neighborhood of $\bar{x}$ on which φ_{ξ} is Lipschitz with a p -integrable modulus κ . Then

$$|\varphi_{\xi}(x)| \leq |\varphi_{\xi}(\bar{x})| + \kappa(\xi) \|x - \bar{x}\| \quad \text{for all } x \in V,$$

and the right-hand side is p -integrable for each fixed $x \in V$, since $\varphi(\cdot, \bar{x}) \in L_{\mathbb{P}}^p$ and $\kappa \in L_{\mathbb{P}}^p$. Hence $\varphi(\cdot, x) \in L_{\mathbb{P}}^p$ for every $x \in V$, which is (a). $\square$

Combining Corollary 5.1 with the ascent inequality (9) yields a criterion in which the p -integrability of the values on a whole neighborhood is replaced by a one-sided bound together with a condition at the reference point alone, namely, the minimum-norm Clarke subgradient at $\bar{x}$ must be p -integrable.

Corollary 5.2. Let $p \in [1, \infty]$, let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, and let $\bar{x} \in \mathbb{R}^d$ be such that $\varphi(\cdot, \bar{x}) \in L_{\mathbb{P}}^p$. Assume that, for some $\varepsilon > 0$ and with $V := \mathbb{B}_{\varepsilon}[\bar{x}]$, one of the following conditions holds:

- (i) φ is a lower- C^2 normal integrand around $\bar{x}$ with p -integrable curvature modulus ϱ , the ascent inequality (9) holds on V , and $\xi \mapsto \sup_{u \in V} \varphi_{\xi}(u)$ belongs to $L_{\mathbb{P}}^p$;
- (ii) φ is an upper- C^2 normal integrand around $\bar{x}$ with p -integrable curvature modulus ϱ , the descent inequality (16) holds on V , and $\xi \mapsto \inf_{u \in V} \varphi_{\xi}(u)$ belongs to $L_{\mathbb{P}}^p$.

Then φ is locally Lipschitz around $\bar{x}$ with p -integrable Lipschitz modulus if and only if

$$\xi \mapsto \text{dist}(0, \partial \varphi_{\xi}(\bar{x})) \quad \text{belongs to } L_{\mathbb{P}}^p. \quad (31)$$

Proof. Necessity is immediate: if φ is locally Lipschitz around $\bar{x}$ with p -integrable Lipschitz modulus κ , then $\|v\| \leq \kappa(\xi)$ for every $\xi \in \Xi$ and every $v \in \partial\varphi_\xi(\bar{x})$, so that $\text{dist}(0, \partial\varphi_\xi(\bar{x})) \leq \kappa(\xi)$ and (31) follows.

For the converse, observe that (i) and (ii) may be exchanged by the change of sign $\varphi \mapsto -\varphi$, which affects neither (31) nor the conclusion, since $\text{dist}(0, \partial(-\varphi_\xi)(\bar{x})) = \text{dist}(0, \partial\varphi_\xi(\bar{x}))$. We may therefore assume without loss of generality that (i) holds.

Suppose that (31) holds, fix $\xi \in \Xi$, and let $\Pi(\bar{x}, \xi)$ be the projection of 0 onto the nonempty compact convex set $\partial\varphi_\xi(\bar{x})$, so that $\|\Pi(\bar{x}, \xi)\| = \text{dist}(0, \partial\varphi_\xi(\bar{x}))$. Applying (9) at $x = \bar{x}$ with $y \in V$ and $v = \Pi(\bar{x}, \xi)$, we obtain

$$\varphi_\xi(\bar{x}) - \text{dist}(0, \partial\varphi_\xi(\bar{x}))\varepsilon - \frac{\varrho(\xi)}{2}\varepsilon^2 \leq \varphi_\xi(y) \leq \sup_{u \in V} \varphi_\xi(u) \quad \text{for all } y \in V.$$

Both outer bounds are independent of y and p -integrable, so there exists $r \in L^p_{\mathbb{P}}$ with $|\varphi_\xi(y)| \leq r(\xi)$ for all $y \in V$. In particular $\varphi(\cdot, y) \in L^p_{\mathbb{P}}$ for every $y \in V$, and Corollary 5.1 gives the conclusion. $\square$

The previous result suggests studying the relation between integrability of the subgradients and Lipschitz continuity of the integrand over open convex sets. For lower- and upper- C^2 normal integrands, the next result states that a pointwise p -integrability condition on the minimum-norm Clarke subgradient already forces a locally uniform p -integrable bound on the entire subdifferential. Note also that the condition is imposed at each point separately, whereas the conclusion is local Lipschitz continuity with a single modulus valid on a whole neighborhood, and the p -integrability of the values is needed at one point only.

Corollary 5.3. Let $p \in [1, \infty]$, let $\varphi : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, and let $U \subseteq \mathbb{R}^d$ be an open convex set containing a point $\bar{x}$ with $\varphi(\cdot, \bar{x}) \in L^p_{\mathbb{P}}$. Assume that one of the following conditions holds:

- (i) φ is lower- C^2 around every point of U with p -integrable curvature modulus;
- (ii) φ is upper- C^2 around every point of U with p -integrable curvature modulus.

Then the following assertions are equivalent:

- (a) for every $x \in U$,

$$\mathbb{E} \left[\text{dist}(0, \partial\varphi_\xi(x))^p \right] < +\infty; \tag{32}$$

- (b) φ is locally Lipschitz continuous around every point of U with p -integrable Lipschitz modulus.

Proof. As before, both (a) and (b) are invariant under the change of sign $\varphi \mapsto -\varphi$, which interchanges (i) and (ii). We may therefore assume without loss of generality that φ is lower- C^2 around every point of U .

Suppose that (a) holds and consider the set

$$\tilde{U} := \{x \in U : \varphi(\cdot, x) \in L^p_{\mathbb{P}}\}, \tag{33}$$

which is nonempty because $\bar{x} \in \tilde{U}$. For $w \in U$, let $\Pi(w, \xi)$ denote, as in the proof of Corollary 5.2, the projection of 0 onto the nonempty compact convex set $\partial\varphi_\xi(w)$, so that $\|\Pi(w, \xi)\| = \text{dist}(0, \partial\varphi_\xi(w))$. By (32), the map $\xi \mapsto \|\Pi(w, \xi)\|$ belongs to $L^p_{\mathbb{P}}$ for every fixed $w \in U$.

First, we show that φ is locally Lipschitz around every point of the open set $\tilde{U}$ with p -integrable Lipschitz modulus. Fix $\hat{x} \in \tilde{U}$ and let $W \subseteq U$ be a compact convex neighborhood of $\hat{x}$ on which (9) and (11) hold with a nonnegative modulus $\varrho \in L^p_{\mathbb{P}}$ and a convex normal integrand f . Let $e_0, \dots, e_d \in \mathbb{R}^d$ be affinely independent points such that $\Delta := \text{co}\{e_0, \dots, e_d\} \subseteq W$ and $\hat{x} \in \text{int } \Delta$.

Applying (9) first at $x = \hat{x}$, $y = e_i$, $v = \Pi(\hat{x}, \xi)$, and then at $x = e_i$, $y = \hat{x}$, $v = \Pi(e_i, \xi)$, we obtain, for each $i = 0, 1, \dots, d$,

$$\varphi_\xi(\hat{x}) + \langle \Pi(\hat{x}, \xi), e_i - \hat{x} \rangle - \frac{\varrho(\xi)}{2} \|e_i - \hat{x}\|^2 \leq \varphi_\xi(e_i) \leq \varphi_\xi(\hat{x}) + \langle \Pi(e_i, \xi), e_i - \hat{x} \rangle + \frac{\varrho(\xi)}{2} \|e_i - \hat{x}\|^2.$$

Since $\varphi(\cdot, \hat{x}) \in L_{\mathbb{P}}^p$, it follows that $\varphi(\cdot, e_i) \in L_{\mathbb{P}}^p$ for every $i = 0, 1, \dots, d$.

Let now $w \in \Delta$, written as a convex combination $w = \sum_{i=0}^d \lambda_i e_i$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. By (11) and the convexity of f_ξ ,

$$\varphi_\xi(w) = f_\xi(w) - \frac{\varrho(\xi)}{2} \|w\|^2 \leq \sum_{i=0}^d \lambda_i f_\xi(e_i) = \sum_{i=0}^d \lambda_i \left(\varphi_\xi(e_i) + \frac{\varrho(\xi)}{2} \|e_i\|^2 \right),$$

while (9) at $x = \hat{x}$, $y = w$ and $v = \Pi(\hat{x}, \xi)$ yields

$$\varphi_\xi(w) \geq \varphi_\xi(\hat{x}) + \langle \Pi(\hat{x}, \xi), w - \hat{x} \rangle - \frac{\varrho(\xi)}{2} \|w - \hat{x}\|^2.$$

Both bounds of $\varphi_\xi(w)$ are dominated by p -integrable functions independent of w , so $\varphi(\cdot, w) \in L_{\mathbb{P}}^p$ for every $w \in \Delta$. Hence $\Delta \subseteq \tilde{U}$, and since $\hat{x} \in \text{int } \Delta$, the set $\tilde{U}$ is a neighborhood of $\hat{x}$. Moreover, $\xi \mapsto \sup_{u \in \Delta} \varphi_\xi(u)$ is p -integrable, so in view of (a), Corollary 5.2 shows that φ is locally Lipschitz around $\hat{x}$ with p -integrable Lipschitz modulus.

We next show that $\tilde{U} = U$. Suppose, on the contrary, that $U \setminus \tilde{U} \neq \emptyset$. Since U is nonempty, open, and convex, hence connected, the set $\tilde{U}$ cannot be relatively closed in U . Therefore, there exists $y \in U \cap \text{bd } \tilde{U}$, and $y \notin \tilde{U}$ because $\tilde{U}$ is open. As φ is lower- C^2 around y , there exist a neighborhood $V \subseteq U$ of y and a nonnegative $\varrho \in L_{\mathbb{P}}^p$ such that (9) holds on V . Since $y \in \text{bd } \tilde{U}$, we may pick $\hat{x} \in \tilde{U} \cap V$. Applying (9) at $x = \hat{x}$, y and then at $x = y$, $\hat{x}$ gives

$$\varphi_\xi(\hat{x}) + \langle \Pi(\hat{x}, \xi), y - \hat{x} \rangle - \frac{\varrho(\xi)}{2} \|y - \hat{x}\|^2 \leq \varphi_\xi(y) \leq \varphi_\xi(\hat{x}) + \langle \Pi(y, \xi), y - \hat{x} \rangle + \frac{\varrho(\xi)}{2} \|y - \hat{x}\|^2,$$

thus $\varphi(\cdot, y) \in L_{\mathbb{P}}^p$, that is, $y \in \tilde{U}$, which yields a contradiction. In this manner, $\tilde{U} = U$, and (b) follows.

Conversely, assume that (b) holds and fix $x \in U$. Let $\kappa \in L_{\mathbb{P}}^p$ be a Lipschitz modulus of φ on a neighborhood of x . Then $\|v\| \leq \kappa(\xi)$ for every $\xi \in \Xi$ and every $v \in \partial \varphi_\xi(x)$, so $\text{dist}(0, \partial \varphi_\xi(x)) \leq \kappa(\xi)$ and (32) follows. $\square$

Remark 5.4 (On bounded-variance and multiplicative-noise assumptions). In smooth stochastic optimization, it is common to impose square-integrability of the stochastic gradient noise. For instance, one assumes that

$$\sigma(x) := \left(\mathbb{E}[\|\nabla_x \varphi(\xi, x) - \nabla \mathbf{E}_\varphi(x)\|^2] \right)^{1/2} < +\infty,$$

or that a *multiplicative-noise condition* of the form

$$\sigma(x) \leq \sigma(\hat{x}) + \sigma_L \|x - \hat{x}\|, \quad x \in U, \quad (34)$$

holds for some $\hat{x} \in U$ and some constant $\sigma_L > 0$. It is straightforward to show that either condition implies

$$\mathbb{E}[\|\nabla_x \varphi(\xi, x)\|^2] < +\infty \quad \text{for all } x \in U,$$

and thus (32) holds for $p = 2$. Hence, under the assumptions of Corollary 5.3 with $p = 2$, φ is locally Lipschitz continuous around every point of U with square-integrable Lipschitz modulus.

6 Applications in structured optimization

In this section, we describe two structured applications that fit our framework of nonconvex normal integrands.

6.1 Partial-label regression

In partial-label regression, each observation is attached to a set of (real-valued) candidate labels that is known to contain the true one [8, 17]. A way of fitting a model to such data is to be optimistic with respect to the ambiguity, namely, the model chooses the most favorable among all candidate label assignments. We show that the resulting optimistic distributionally robust formulation reduces to the expectation of an upper- C^2 normal integrand.

Let $m \in \mathbb{N}$. The data consist of a measurable set-valued mapping $S : \Xi \rightrightarrows \mathbb{R}^m$ with nonempty closed values, representing the candidates, and a Carathéodory mapping $G : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}^m$, the prediction model. A typical instance is $G_\xi(\cdot) = h(\cdot; a(\xi))$, where $a(\xi)$ is a random feature vector and h is a parametric model. For the squared residual, the optimistic loss at $x \in \mathbb{R}^d$ of a sample ξ is

$$\varphi_\xi(x) := \min_{y \in S(\xi)} \frac{1}{2} \|G_\xi(x) - y\|^2 = \frac{1}{2} \text{dist}(G_\xi(x), S(\xi))^2. \quad (35)$$

Let $\mathcal{Q}$ be the set of probability measures on $\Xi \times \mathbb{R}^m$, endowed with the product σ -algebra $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^m)$, consisting of all joint laws of observations and labels that reproduce the observed data and assign to each observation a (possibly randomized) label among its candidates. In other words, $Q \in \mathcal{Q}$ if

$$Q(D \times \mathbb{R}^m) = \mathbb{P}(D) \quad \text{for all } D \in \mathcal{F} \quad \text{and} \quad Q(\text{gph } S) = 1. \quad (36)$$

Notice that $\text{gph } S \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^m)$ because S is measurable with closed values [23, Theorem 14.8].

For a linear map A , $\|A\|$ denotes the operator norm, and for $x \in \mathbb{R}^d$ we regard $\nabla^2 G_\xi(x)$ as a symmetric bilinear map $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ with norm

$$\|\nabla^2 G_\xi(x)\| := \sup \{ \|\nabla^2 G_\xi(x)[u, w]\| : \|u\| \leq 1, \|w\| \leq 1 \}.$$

Proposition 6.1. Let S , G and φ be as above. Fix $\bar{x} \in \mathbb{R}^d$ and $\delta > 0$, set $V := \mathbb{B}_\delta[\bar{x}]$, and assume that, for every $\xi \in \Xi$, the mapping G_ξ is twice-continuously differentiable on an open set containing V , and that there exist measurable functions $K_1, K_2 : \Xi \rightarrow [0, +\infty)$ such that

$$\|\nabla G_\xi(x)\| \leq K_1(\xi), \quad \|\nabla^2 G_\xi(x)\| \leq K_2(\xi) \quad \text{for all } x \in V, \xi \in \Xi. \quad (37)$$

Set $r_0(\xi) := \text{dist}(G_\xi(\bar{x}), S(\xi))$. Then the following assertions hold.

(a) For every $x \in \mathbb{R}^d$,

$$\inf_{Q \in \mathcal{Q}} \int_{\Xi \times \mathbb{R}^m} \frac{1}{2} \|G_\xi(x) - y\|^2 Q(d(\xi, y)) = \mathbb{E}_\varphi(x), \quad (38)$$

where both sides belong to $[0, +\infty]$ and the infimum is attained.

(b) The function φ is a finite-valued normal integrand which is upper- C^2 around $\bar{x}$ in the sense of Definition 2.4, with Polish space $X := \mathbb{R}^m$, compact-valued measurable mapping

$$C(\xi) := S(\xi) \cap \mathbb{B}_{\tau(\xi)}[G_\xi(\bar{x})], \quad \text{where } \tau(\xi) := r_0(\xi) + 2\delta K_1(\xi), \quad (39)$$

Carathéodory mapping $g(\xi, y, x) := \frac{1}{2} \|G_\xi(x) - y\|^2$, and curvature modulus

$$\varrho(\xi) := K_1(\xi)^2 + K_2(\xi)(r_0(\xi) + 3\delta K_1(\xi)). \quad (40)$$

(c) For every $\xi \in \Xi$, the function φ_ξ is Lipschitz continuous on V with modulus

$$\kappa(\xi) := K_1(\xi)(r_0(\xi) + \delta K_1(\xi)). \quad (41)$$

- (d) Let $p \in [1, \infty]$. If $K_1, r_0 \in L_{\mathbb{P}}^{2p}$ and $K_2 r_0, K_2 K_1 \in L_{\mathbb{P}}^p$, then $\varrho, \kappa \in L_{\mathbb{P}}^p$ and $\varphi(\cdot, x) \in L_{\mathbb{P}}^p$ for every $x \in V$.

Proof. For every $w \in \mathbb{R}^m$, the function $\xi \mapsto \text{dist}(w, S(\xi))$ is measurable since S is measurable with closed values [23, Theorem 14.3]. In particular, r_0 is measurable. As $w \mapsto \text{dist}(w, S(\xi))$ is 1-Lipschitz, $(\xi, w) \mapsto \text{dist}(w, S(\xi))$ is Carathéodory, and so is its composition with the Carathéodory mapping G . Consequently, φ is a finite-valued Carathéodory integrand, and thus a normal integrand [23, Example 14.29].

We first show (a). Fix $x \in \mathbb{R}^d$ and set $f_x(\xi, y) := \frac{1}{2} \|G_\xi(x) - y\|^2$, which is Carathéodory and hence $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^m)$ -measurable. Let $Q \in \mathcal{Q}$. Since $Q(\text{gph } S) = 1$, we have $f_x(\xi, y) \geq \varphi_\xi(x)$ for Q -almost every (ξ, y) , and since the Ξ -marginal of Q is $\mathbb{P}$, it follows that

$$\int f_x dQ \geq \int \varphi_\xi(x) Q(d(\xi, y)) = \mathbb{E}_\varphi(x).$$

Conversely, the mapping $\xi \mapsto \text{proj}_{S(\xi)}(G_\xi(x))$ is the argmin mapping of the normal integrand $(\xi, y) \mapsto f_x(\xi, y) + \iota_{S(\xi)}(y)$, which is level-bounded in y . Hence it is measurable with nonempty closed values [23, Theorem 14.37] and admits a measurable selection $\hat{y}_x(\xi) \in \text{proj}_{S(\xi)}(G_\xi(x))$ [23, Corollary 14.6]. The image Q_x of $\mathbb{P}$ under $\xi \mapsto (\xi, \hat{y}_x(\xi))$ belongs to $\mathcal{Q}$, and

$$\int f_x dQ_x = \int_{\Xi} \frac{1}{2} \|G_\xi(x) - \hat{y}_x(\xi)\|^2 \mathbb{P}(d\xi) = \mathbb{E}_\varphi(x),$$

which proves (38) together with the attainment of the infimum.

Next, to show (b), we first check that C in (39) is a measurable set-valued mapping with nonempty compact values. The mapping $\xi \mapsto \mathbb{B}_{\tau(\xi)}[G_\xi(\bar{x})]$ is measurable with closed values, because it is a level-set mapping of a normal integrand [23, Proposition 14.33], and thus so is the intersection in (39) [23, Proposition 14.11]. Its values are closed and bounded, hence compact. Now fix $\xi \in \Xi$ and $x \in V$. Since V is convex, (37) and the mean value inequality give

$$\|G_\xi(x) - G_\xi(\bar{x})\| \leq K_1(\xi) \|x - \bar{x}\| \leq \delta K_1(\xi). \quad (42)$$

Let $y \in \text{proj}_{S(\xi)}(G_\xi(x))$. As $\text{dist}(\cdot, S(\xi))$ is 1-Lipschitz, (42) yields

$$\|y - G_\xi(\bar{x})\| \leq \text{dist}(G_\xi(x), S(\xi)) + \|G_\xi(x) - G_\xi(\bar{x})\| \leq r_0(\xi) + 2\|G_\xi(x) - G_\xi(\bar{x})\| \leq \tau(\xi),$$

that is, $\text{proj}_{S(\xi)}(G_\xi(x)) \subseteq C(\xi)$. In particular, by taking $x = \bar{x}$, $C(\xi) \neq \emptyset$. Since $C(\xi) \subseteq S(\xi)$, we conclude that

$$\varphi_\xi(x) = \min_{y \in C(\xi)} g(\xi, y, x) \quad \text{for all } x \in V, \quad (43)$$

which is (8). Moreover, the mapping g is Carathéodory. Indeed, for each $\xi \in \Xi$ and $y \in \mathbb{R}^m$, the function $x \mapsto g(\xi, y, x)$ is twice-continuously differentiable on an open set containing V , such that

$$\nabla_x g(\xi, y, x) = \nabla G_\xi(x)^\top (G_\xi(x) - y)$$

and

$$\nabla_x^2 g(\xi, y, x)[u, u] = \|\nabla G_\xi(x)u\|^2 + \langle G_\xi(x) - y, \nabla^2 G_\xi(x)[u, u] \rangle,$$

and $g, \nabla_x g$ and $\nabla_x^2 g$ are continuous in (y, x) . Moreover, for $y \in C(\xi)$ and $x \in V$, (42) gives

$$\|G_\xi(x) - y\| \leq \|G_\xi(x) - G_\xi(\bar{x})\| + \|G_\xi(\bar{x}) - y\| \leq \delta K_1(\xi) + \tau(\xi) = r_0(\xi) + 3\delta K_1(\xi).$$

Combining these estimates with (37), for $\|u\| \leq 1$, we obtain

$$\|\nabla_x^2 g(\xi, y, x)\| \leq K_1(\xi)^2 + K_2(\xi)(r_0(\xi) + 3\delta K_1(\xi)) = \varrho(\xi) \quad \text{for all } y \in C(\xi), x \in V,$$

which yields (7). Hence all the requirements of Definition 2.4 are fulfilled, and φ is an upper- C^2 normal integrand around $\bar{x}$ with curvature modulus (40).

To show (c), set $d_\xi(x) := \text{dist}(G_\xi(x), S(\xi))$. By (42), $d_\xi(x) \leq r_0(\xi) + \delta K_1(\xi)$ and $|d_\xi(x) - d_\xi(x')| \leq K_1(\xi)\|x - x'\|$ for all $x, x' \in V$. Therefore

$$|\varphi_\xi(x) - \varphi_\xi(x')| = \frac{1}{2} |d_\xi(x) - d_\xi(x')| (d_\xi(x) + d_\xi(x')) \leq \kappa(\xi)\|x - x'\|.$$

We conclude by proving (d). Since $K_1 \in L_{\mathbb{P}}^{2p}$, we have $K_1^2 \in L_{\mathbb{P}}^p$, and by Hölder's inequality $\|K_1 r_0\|_p \leq \|K_1\|_{2p} \|r_0\|_{2p}$. Hence $\varrho = K_1^2 + K_2 r_0 + 3\delta K_2 K_1$ and $\kappa = K_1 r_0 + \delta K_1^2$ belong to $L_{\mathbb{P}}^p$. Finally, $0 \leq \varphi_\xi(x) \leq \frac{1}{2}(r_0(\xi) + \delta K_1(\xi))^2$ for all $x \in V$, and the right-hand side belongs to $L_{\mathbb{P}}^p$. $\square$

Remark 6.2 (Optimism versus pessimism). If S has compact values, replacing the infimum in (38) by a supremum leads, by the same argument, to the integrand $\max_{y \in S(\xi)} \frac{1}{2} \|G_\xi(x) - y\|^2$. This integrand is lower- C^2 by Definition 2.4, so pessimism and optimism over candidate labels correspond, respectively, to the lower- and upper- C^2 classes. When $S(\xi)$ is convex, as for interval-valued labels, the function $\frac{1}{2} \text{dist}(\cdot, S(\xi))^2$ is convex with a 1-Lipschitz gradient. Then φ_ξ is convex for linear models, and smooth whenever G_ξ is.

Remark 6.3 (Descent property on active candidate labels). By Corollary 3.2(b), Proposition 6.1 yields the descent inequality

$$\varphi_\xi(x') \leq \varphi_\xi(x) + \langle v, x' - x \rangle + \frac{\varrho(\xi)}{2} \|x' - x\|^2 \quad \text{for all } x, x' \in V \text{ and all } v \in \partial\varphi_\xi(x),$$

where, by the localized representation (43) and [23, Theorem 10.31] applied to $-\varphi_\xi$,

$$\partial\varphi_\xi(x) = \text{co} \left\{ \nabla G_\xi(x)^\top (G_\xi(x) - y) : y \in \text{proj}_{S(\xi)}(G_\xi(x)) \right\}. \quad (44)$$

In particular, the descent inequality holds for every choice of active candidate label. This is related to identification methods for partial-label learning [8, 35], in which the updates first identify a candidate (or model) and then perform the optimization step.

Example 6.4 (Amplitude-based phase retrieval). Let $a : \Xi \rightarrow \mathbb{R}^d$ and $b : \Xi \rightarrow [0, +\infty)$ be measurable. The amplitude-based loss underlying the reshaped Wirtinger flow [35] and the truncated amplitude flow [32] is

$$\varphi_\xi(x) := \frac{1}{2} (|\langle a(\xi), x \rangle| - b(\xi))^2. \quad (45)$$

This is the particular case of (35) with $m = 1$, the linear model $G_\xi(x) = \langle a(\xi), x \rangle$, and the two-point candidate set $S(\xi) := \{-b(\xi), b(\xi)\}$. Indeed, S is measurable with nonempty compact values, and since $b(\xi) \geq 0$,

$$\text{dist}(u, \{-b(\xi), b(\xi)\}) = \min \{|u - b(\xi)|, |u + b(\xi)|\} = ||u| - b(\xi)| \quad \text{for all } u \in \mathbb{R}.$$

By Proposition 6.1, φ is an upper- C^2 normal integrand around every point of $\mathbb{R}^d$, with $r_0(\xi) = ||\langle a(\xi), \bar{x} \rangle| - b(\xi)|$, $\kappa(\xi) = \|a(\xi)\|(r_0(\xi) + \delta\|a(\xi)\|)$ and $\varrho(\xi) = \|a(\xi)\|^2$. By (44), the Clarke subdifferential is

$$\partial\varphi_\xi(x) = \begin{cases} \{(\langle a, x \rangle - b \text{sgn}\langle a, x \rangle)a\}, & \langle a, x \rangle \neq 0, \\ [-b, b]a, & \langle a, x \rangle = 0, \end{cases}$$

where the dependence on ξ is omitted. Consequently, the descent inequality

$$\varphi_\xi(x') \leq \varphi_\xi(x) + \langle v, x' - x \rangle + \frac{\|a(\xi)\|^2}{2} \|x' - x\|^2$$

holds for all $x, x' \in \mathbb{R}^d$ and every $v \in \partial\varphi_\xi(x)$. In particular, it holds for the update direction of the reshaped Wirtinger flow [35, Algorithm 1]. The update rule in [32, Algorithm 1] of the truncated amplitude flow acts on these subgradients, by discarding some samples at each iteration.

6.2 Conditional value-at-risk

The conditional value-at-risk (CVaR), introduced in [21] (see also [1, 22]), is a risk measure that quantifies the expected loss in the tail of a distribution. Let $\alpha \in (0, 1)$ be a confidence level and let $L : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, where $L_\xi(x)$ represents the loss incurred by the decision $x \in \mathbb{R}^d$ under the scenario $\xi \in \Xi$, with $L(\cdot, x) \in L_{\mathbb{P}}^1$. Roughly speaking, the CVaR of $L_\xi(x)$ is the average of the worst $(1 - \alpha)$ -fraction of the losses. When the distribution of $L_\xi(x)$ is continuous, it coincides with the conditional expectation

$$\text{CVaR}_\alpha(L_\xi(x)) = \mathbb{E}[L_\xi(x) \mid L_\xi(x) \geq \text{VaR}_\alpha(L_\xi(x))],$$

where VaR_α denotes the value-at-risk, that is, the α -quantile of the loss. Unlike the VaR, the CVaR is a coherent risk measure in the sense of [6], and it interpolates between the expected loss ($\alpha \rightarrow 0$) and the worst-case loss ($\alpha \rightarrow 1$).

The CVaR has found applications in a wide range of fields. In finance, where it originated, it is used for portfolio selection and hedging [21], and it has replaced the VaR as the reference measure of market risk in the Basel regulatory framework; see [18] for a comprehensive account. In stochastic programming it is the prototypical coherent risk measure for risk-averse formulations [25], and CVaR constraints provide a convex conservative approximation of chance constraints, the tightest one within a natural class [19]. More recently, the CVaR has been adopted in machine learning as a training criterion that focuses on the hardest examples or subpopulations rather than on the average loss, owing to its dual representation as a worst-case expectation over distributions with bounded likelihood ratio [14], and in risk-sensitive reinforcement learning and control [9]. In most of these applications the CVaR enters either as an objective or as a constraint, so that its tractability for optimization is essential.

A key feature of the CVaR is that it admits a variational formulation [21], which makes it amenable to optimization. It can be equivalently described by

$$\text{CVaR}_\alpha(L_\xi(x)) = \min_{t \in \mathbb{R}} \mathbf{E}_\psi(x, t),$$

with

$$\psi_\xi(x, t) := t + \frac{1}{1 - \alpha} (L_\xi(x) - t)_+, \quad (x, t) \in \mathbb{R}^{d+1}, \quad (46)$$

where $(s)_+ := \max\{s, 0\}$. Moreover, the value-at-risk $\text{VaR}_\alpha(L_\xi(x))$ is a minimizer of the inner problem in t (the left endpoint of the set of minimizers). Consequently, minimizing the CVaR over x reduces to the joint stochastic problem

$$\min_{(x, t) \in \mathbb{R}^{d+1}} \mathbf{E}_\psi(x, t). \quad (47)$$

The next proposition shows that the integrand ψ in (47) inherits the lower- C^2 structure and the Lipschitz continuity of the loss L . It is a direct consequence of Proposition 4.6.

Proposition 6.5. Let $p \in [1, \infty]$, let $L : \Xi \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a finite-valued normal integrand, and let $\bar{x} \in \mathbb{R}^d$ be such that L is lower- C^2 around $\bar{x}$ with p -integrable curvature modulus. Then the integrand ψ defined in (46) is a finite-valued normal integrand on $\Xi \times \mathbb{R}^{d+1}$ which is lower- C^2 around $(\bar{x}, \bar{t})$ with p -integrable curvature modulus, for every $\bar{t} \in \mathbb{R}$. If, in addition, L is locally Lipschitz around $\bar{x}$ with p -integrable modulus, then ψ is locally Lipschitz around $(\bar{x}, \bar{t})$ with p -integrable modulus, for every $\bar{t} \in \mathbb{R}$.

7 Concluding comments

The preceding results show that lower- C^2 normal integrands admit a simple local representation as upper envelopes of random quadratic-affine functions with integrable curvature, and, by a change of

sign, that upper- C^2 normal integrands are the corresponding lower envelopes. This representation is suitable for stochastic programming, since it is stable under compact-recourse minimization and provides a direct route from samplewise second-order estimates to expected-value estimates. A natural next step is to combine this structure with algorithmic frameworks for nonsmooth stochastic optimization, where the integrable curvature modulus and square-integrable Lipschitz moduli can be used to control model errors and sample-average approximations.

Acknowledgements

Research of the authors was supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence. The second author was also supported by ANID-Chile grants Fondecyt Regular 1240120, Fondecyt Regular 1240335 and Fondecyt Regular 1261728. The fourth author was partially supported by grant PID2025-168246NB-I00 funded by ERDF/EU and by MICIU/AEI/10.13039/501100011033. The last two authors were partially supported by grant CIGE/2026/80 by Conselleria d'Educació, Cultura i Universitats, Generalitat Valenciana.

References

- [1] C. ACERBI AND D. TASCHÉ, *On the coherence of expected shortfall*, J. Bank. Finance, 26 (2002), pp. 1487–1503.
- [2] P. APKARIAN, D. NOLL, AND O. PROT, *A proximity control algorithm to minimize nonsmooth and nonconvex semi-infinite maximum eigenvalue functions*, J. Convex Anal., 16 (2009), pp. 641–666.
- [3] F. J. ARAGÓN-ARTACHO, R. CAMPOY, P. PÉREZ-AROS, AND D. TORREGROSA-BELÉN, *Nonmonotone subgradient methods based on a local descent lemma*, Computational Optimization and Applications, (2026). To appear.
- [4] F. J. ARAGÓN-ARTACHO, B. S. MORDUKHOVICH, AND P. PÉREZ-AROS, *Coderivative-based semi-Newton method in nonsmooth difference programming*, Math. Program., 213 (2025), pp. 385–432.
- [5] F. J. ARAGÓN-ARTACHO, P. PÉREZ-AROS, AND D. TORREGROSA-BELÉN, *The boosted double-proximal subgradient algorithm for nonconvex optimization*, Math. Program., 214 (2025), pp. 491–537.
- [6] P. ARTZNER, F. DELBAEN, J.-M. EBER, AND D. HEATH, *Coherent measures of risk*, Math. Finance, 9 (1999), pp. 203–228.
- [7] F. ATENAS, C. SAGASTIZÁBAL, P. J. S. SILVA, AND M. V. SOLODOV, *A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods*, SIAM J. Optim., 33 (2023), pp. 89–115.
- [8] X. CHENG, D.-B. WANG, L. FENG, M.-L. ZHANG, AND B. AN, *Partial-label regression*, in Proc. AAAI Conf. Artif. Intell., vol. 37, 2023.
- [9] Y. CHOW, M. GHAVAMZADEH, L. JANSON, AND M. PAVONE, *Risk-constrained reinforcement learning with percentile risk criteria*, J. Mach. Learn. Res., 18 (2017), pp. 1–51.
- [10] F. H. CLARKE, *Optimization and nonsmooth analysis*, vol. 5 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990.

- [11] F. H. CLARKE, R. J. STERN, AND P. R. WOLENSKI, *Proximal smoothness and the lower- C^2 property*, J. Convex Anal., 2 (1995), pp. 117–144.
- [12] A. DANIILIDIS AND J. MALICK, *Filling the gap between lower- C^1 and lower- C^2 functions*, J. Convex Anal., 12 (2005), pp. 315–329.
- [13] W. DE OLIVEIRA, *A note on the Frank–Wolfe algorithm for a class of nonconvex and nonsmooth optimization problems*, Open J. Math. Optim., 4 (2023), pp. 1–10.
- [14] J. C. DUCHI AND H. NAMKOONG, *Learning models with uniform performance via distributionally robust optimization*, Ann. Stat., 49 (2021), pp. 1378–1406.
- [15] J. FAN AND R. LI, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc., 96 (2001), pp. 1348–1360.
- [16] W. HARE AND C. SAGASTIZÁBAL, *A redistributed proximal bundle method for nonconvex optimization*, SIAM J. Optim., 20 (2010), pp. 2442–2473.
- [17] E. HÜLLERMEIER, *Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization*, Int. J. Approx. Reason., 55 (2014), pp. 1519–1534.
- [18] A. J. MCNEIL, R. FREY, AND P. EMBRECHTS, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, revised ed., 2015.
- [19] A. NEMIROVSKI AND A. SHAPIRO, *Convex approximations of chance constrained programs*, SIAM J. Optim., 17 (2006), pp. 969–996.
- [20] P. PATRINOS AND A. BEMPORAD, *Proximal Newton methods for convex composite optimization*, in Proc. IEEE Conf. Decis. Control, IEEE, 2013, pp. 2358–2363.
- [21] R. T. ROCKAFELLAR AND S. URYASEV, *Optimization of conditional value-at-risk*, J. Risk, 2 (2000), pp. 21–41.
- [22] —, *Conditional value-at-risk for general loss distributions*, J. Bank. Finance, 26 (2002), pp. 1443–1471.
- [23] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational analysis*, vol. 317 of Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1998.
- [24] J. O. ROYSET AND R. J.-B. WETS, *An optimization primer*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2021.
- [25] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Lectures on stochastic programming—modeling and theory*, vol. 28 of MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, third ed., 2021.
- [26] J. E. SPINGARN, *Submonotone mappings and the proximal point algorithm*, Numer. Funct. Anal. Optim., 4 (1981/82), pp. 123–150.
- [27] L. STELLA, A. THEMELIS, AND P. PATRINOS, *Forward-backward quasi-Newton methods for nonsmooth optimization problems*, Comput. Optim. Appl., 67 (2017), pp. 443–487.
- [28] A. THEMELIS, L. STELLA, AND P. PATRINOS, *Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms*, SIAM J. Optim., 28 (2018), pp. 2274–2303.

- [29] L. THIBAUT, *Unilateral variational analysis in Banach spaces. Part I—general theory*, World Scientific Publishing Co. Pte. Ltd., Singapore, [2023] ©2023.
- [30] —, *Unilateral variational analysis in Banach spaces. Part II—special classes of functions and sets*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, [2023] ©2023.
- [31] W. VAN ACKOOIJ AND W. DE OLIVEIRA, *Nonsmooth and nonconvex optimization via approximate difference-of-convex decompositions*, J. Optim. Theory Appl., 182 (2019), pp. 49–80.
- [32] G. WANG, G. B. GIANNAKIS, AND Y. C. ELDAR, *Solving systems of random quadratic equations via truncated amplitude flow*, IEEE Trans. Inf. Theory, 64 (2018), pp. 773–794.
- [33] C. ZĂLINESCU, *Convex analysis in general vector spaces*, Singapore: World Scientific, 2002.
- [34] C.-H. ZHANG, *Nearly unbiased variable selection under minimax concave penalty*, Ann. Statist., 38 (2010), pp. 894–942.
- [35] H. ZHANG, Y. ZHOU, Y. LIANG, AND Y. CHI, *A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms*, J. Mach. Learn. Res., 18 (2017).