FULL LENGTH PAPER

Series A

Lyapunov stability of the subgradient method with constant step size

Cédric Josz[1] · **Lexiao Lai[1]**

Abstract

We consider the subgradient method with constant step size for minimizing locally Lipschitz semi-algebraic functions. In order to analyze the behavior of its iterates in the vicinity of a local minimum, we introduce a notion of discrete Lyapunov stability and propose necessary and sufficient conditions for stability.

1 Introduction

The subgradient method with constant step size for minimizing a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ consists in choosing an initial point $x_0 \in \mathbb{R}^n$ and generating a sequence of iterates according to the update rule $x_{k+1} \in x_k - \alpha \partial f(x_k)$, $\forall k \in \mathbb{N} := \{0, 1, 2, \ldots\}$, where $\alpha > 0$ is the step size and ∂f is the Clarke subdifferential [14, Chapter 2]. While this method is often used in practice to solve nonconvex and nonsmooth problems, there is little theoretical understanding of the behavior of its iterates. To the best of our knowledge, the only known results are in the convex setting, as we next describe.

✉ Cédric Josz
cj2638@columbia.edu

Lexiao Lai
ll3352@columbia.edu

[1] Department of Industrial Engineering and Operations Research, Columbia University, S. W. Mudd Building, 500 W 120th St, New York, NY 10027-6623, USA

Published online: 16 February 2023

Springer

If f is convex and the Euclidean norm of its subgradients is bounded above by a constant c, then $\liminf f(x_k) - \inf f \leq c^2\alpha/2$ provided that the infimum is reached [6, Proposition 3.2.3]. In order to get within ϵ accuracy of that bound, $\lfloor d(x_0, X)^2/(\alpha\epsilon) \rfloor$ iterations suffice where $\lfloor \cdot \rfloor$ denotes the floor of a real number and $d(x_0, X)$ is the distance between the initial iterate x_0 and the set of minimizers $X \subset \mathbb{R}^n$ [6, Proposition 3.2.4]. If the objective function grows quadratically (at least as fast as $t \in \mathbb{R} \mapsto \beta t^2$ for some $\beta > 0$) around the set of minimizers, then the iterates asymptotically get within $c\sqrt{\alpha}/\sqrt{2\beta}$ distance to the set of minimizers if $\alpha \in (0, 1/(2\beta)]$ [6, Proposition 3.2.5]. If we relax the boundedness assumption on the subgradients to $\|s\| \leq c\sqrt{1 + d(x, X)^2}$ for all (x, s) in the graph of ∂f, then we get the slightly weaker bound $\liminf f(x_k) - \inf f \leq c^2\alpha/2(1 + d(x_0, X))$ [6, Exercise 3.6].

Given the absence of theoretical results in the nonconvex setting, in this note we take a first step by investigating the behavior of the subgradient method in the vicinity of a local minimum of f. In order to do so, we propose a notion of stability akin to Lyapunov stability in dynamical systems [21] [26, Equation (5.6)]. Informally, a point is stable if all of the iterates of the subgradient method remain in any neighborhood of it, provided that the initial point is close enough to it and that the step size is small enough.

Without any further assumptions on f, the notions of stability and local optimality are decorrelated. Indeed, the classical counterexample $f(x) = x^2 \sin(1/x)$ admits a stable point which is not a local minimum, while the Rockafellar function [25, p. 5] (see also [20, Proposition (1.9)]) admits a local minimum that is not stable. Assuming sharpness [16, Assumption A 2] and weak convexity [16, Assumption A 1], it can easily be shown that strict local minima are stable. These assumptions may not hold in practice however [15, p. 121] [14, 2.3.6 Proposition], and can be difficult to check [13, Conjecture 8.7]. We thus confine our investigation to locally Lipschitz semi-algebraic functions [9, 24]. The tame generalization [28] is immediate and captures seemingly all applications of interest nowadays.

When f is locally Lipschitz and semi-algebraic, the set of stable points and local minima coincide in two cases. The first is when f is continuously differentiable with a locally Lipschitz gradient. The fact that local minima are stable in this regime can be deduced using arguments from [2, Proposition 3.3]. The converse is a consequence of one of our results (Theorem 1). The second case is that of continuous-time subgradient dynamics, where instead of iterates one considers absolutely continuous solutions to the differential inclusion $x' \in -\partial f(x)$. This is a simple generalization of [1, Theorem 3] which holds for real analytic functions. Much of modern numerical optimization however falls outside the scope of these two cases, namely that of smooth objective functions and continuous-time dynamics. It is thus important to determine in the discrete case and when f is merely locally Lipschitz and semi-algebraic, whether local minima are stable, and conversely, whether stable points are local minima. In this note, we show that for a point to be stable, it is necessary for it to be a local minimum and it suffices for it to be a strict local minimum.

2 Characterizing stability

Let $\| \cdot \|$ be the induced norm of an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$. Let $B(a, r)$ and $\mathring{B}(a, r)$ respectively denote the closed ball and the open ball of center $a \in \mathbb{R}^n$ and radius $r > 0$. We next define the notion of discrete Lyapunov stability.

Definition 1 We say that $x^* \in \mathbb{R}^n$ is a stable point of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ if for all $\epsilon > 0$, there exist $\delta > 0$ and $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, the subgradient method with constant step size α initialized in $B(x^*, \delta)$ has all its iterates in $B(x^*, \epsilon)$.

In order to characterize stability, we rely on the theory of differential inclusions [4]. However, existing results cannot directly be applied, mainly because the locally Lipschitz assumption is too weak. We thus adapt them to our setting via Lemma 1. Note that the analysis of the stochastic subgradient method also involves differential inclusions [5, 7, 8, 11, 15].

Lemma 1 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be an upper semicontinuous mapping with nonempty, compact, and convex values. Let X_0 be a compact subset of $\mathbb{R}^n$ and let $T > 0$. Assume that there exist $\bar{\alpha}, r > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$ and for all sequence $(x_k)_{k \in \mathbb{N}}$ such that*

$$x_{k+1} \in x_k + \alpha F(x_k), \quad \forall k \in \mathbb{N}, \quad x_0 \in X_0, \tag{1}$$

we have that $x_0, \ldots, x_{\lfloor T/\alpha \rfloor + 1} \in B(0, r)$. For all $\epsilon > 0$, there exists $\hat{\alpha} \in (0, \bar{\alpha}]$ such that for all $\alpha \in (0, \hat{\alpha}]$ and for all sequence $(x_k)_{k \in \mathbb{N}}$ satisfying (1), there exists an absolutely continuous function $x : [0, T] \to \mathbb{R}^n$ such that

$$x'(t) \in F(x(t)), \quad \text{for a.e. } t \in (0, T), \quad x(0) \in X_0, \tag{2}$$

for which $\|\bar{x}(t) - x(t)\| \leq \epsilon$ for all $t \in [0, T]$, where $\bar{x} : [0, T] \to \mathbb{R}^n$ is defined by $\bar{x}(t) := x_k + (t - \alpha k)/\alpha(x_{k+1} - x_k)$ for all $t \in [\alpha k, \min\{\alpha(k + 1), T\}]$ and $k \in \{0, \ldots, \lfloor T/\alpha \rfloor\}$.

Proof Let $(\alpha_m)_{m \in \mathbb{N}}$ denote a sequence of positive numbers that converges to zero. Without loss of generality, we may assume that the sequence is bounded above by $\bar{\alpha}$. To each term in the sequence, we attribute a sequence $(x_k^m)_{k \in \mathbb{N}}$ generated by the Euler method with step size α_m and initialized in X_0, that is to say, which satisfies (1) with $\alpha := \alpha_m$. By assumption, $x_0^m, \ldots, x_{\lfloor T/\alpha_m \rfloor + 1}^m \in B(0, r)$. Consider the linear interpolation of those iterates, that is to say, the function $\bar{x}^m : [0, T] \to \mathbb{R}^n$ defined by $\bar{x}^m(t) := x_k^m + (t - \alpha_m k)(x_{k+1}^m - x_k^m)/\alpha_m$ for all $t \in [\alpha_m k, \min\{\alpha_m(k + 1), T\}]$ and $k \in \{0, \ldots, \lfloor T/\alpha_m \rfloor\}$. Since $B(0, r)$ is convex, it holds that $\|\bar{x}^m(t)\| \leq r$ for all $t \in [0, T]$. In addition, since F is upper semicontinuous and compact valued, by [4, Proposition 3 p. 42] there exists $r' > 0$ such that $F(B(0, r)) \subset B(0, r')$. Observe that $(\bar{x}^m)'(t) = (x_{k+1}^m - x_k^m)/\alpha_m \in F(x_k^m)$ for all $t \in (\alpha_m k, \min\{\alpha_m(k + 1), T\})$ and $k \in \{0, \ldots, \lfloor T/\alpha_m \rfloor\}$. Hence, we have that $\|(\bar{x}^m)'(t)\| \leq r'$ for almost every $t \in (0, T)$. By successively applying the Arzelà-Ascoli and the Banach-Alaoglu theorems (see

[4, Theorem 4 p. 13]), there exists a subsequence (again denoted $(\alpha_m)_{m \in \mathbb{N}}$) and an absolutely continuous function $x : [0, T] \rightarrow \mathbb{R}^n$ such that $\bar{x}^m(\cdot)$ converges uniformly to $x(\cdot)$ and $(\bar{x}^m)'(\cdot)$ converges weakly to $x'(\cdot)$ in $L^1([0, T], \mathbb{R}^n)$. Furthermore, for all $t \in (\alpha_m k, \min\{\alpha_m(k + 1), T\})$ and $k \in \{0, \dots, \lfloor T/\alpha_m \rfloor\}$, observe that

$$
\begin{aligned}
(\bar{x}^m(t), (\bar{x}^m)'(t)) &= \left(x_k^m + (t - \alpha_m k) \frac{x_{k+1}^m - x_k^m}{\alpha_m}, \frac{x_{k+1}^m - x_k^m}{\alpha_m} \right) \\
&\in \left(\{x_k^m\} + (t - \alpha_m k) F(x_k^m) \right) \times F(x_k^m) \\
&= \{x_k^m\} \times F(x_k^m) + (t - \alpha_m k) F(x_k^m) \times \{0\} \\
&\subset \operatorname{graph}(F) + B(0, r'\alpha_m) \times \{0\}.
\end{aligned}
$$

According to [4, Convergence Theorem p. 60], it follows that $x'(t) \in F(x(t))$ for almost every $t \in (0, T)$. The sequence of initial points $(x^m(0))_{m \in \mathbb{N}}$ lies in the closed set X_0, hence its limit $x(0)$ lies in X_0 as well. As a result, $x(\cdot)$ is a solution to the differential inclusion (2).

To sum up, we have shown that for every sequence $(\alpha_m)_{m \in \mathbb{N}}$ of positive numbers converging to zero, there exists a subsequence for which the corresponding linear interpolations uniformly converge towards a solution of the differential inclusion (2). The conclusion of the theorem now easily follows. To see why, one can reason by contradiction and assume that there exists $\epsilon > 0$ such that for all $\hat{\alpha} \in (0, \bar{\alpha}]$, there exist $\alpha \in (0, \hat{\alpha}]$ and a sequence $(x_k)_{k \in \mathbb{N}}$ generated by the Euler method with step size α and initialized in X_0 such that, for any solution $x(\cdot)$ to the differential inclusion (2), it holds that $\|\bar{x}(t) - x(t)\| > \epsilon$ for some $t \in [0, T]$. We can then generate a sequence $(\alpha_m)_{m \in \mathbb{N}}$ of positive numbers converging to zero such that, for any solution $x(\cdot)$ to the differential inclusion (2), it holds that $\|\bar{x}^m(t) - x(t)\| > \epsilon$ for some $t \in [0, T]$. Since there exists a subsequence $(\alpha_{\varphi(m)})_{m \in \mathbb{N}}$ such that $\bar{x}^{\varphi(m)}$ uniformly converges to a solution to the differential inclusion (2), we obtain a contradiction. $\qquad \square$

We will apply Lemma 1 to the case where the set-valued mapping F is the opposite of the Clarke subdifferential of a locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Recall that a point $x^* \in \mathbb{R}^n$ is a local minimum (respectively strict local minimum) of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ if there exists a positive constant ϵ such that $f(x^*) \leq f(x)$ for all $x \in B(x^*, \epsilon) \backslash \{x^*\}$ (respectively $f(x^*) < f(x)$). Using Lemma 1, we obtain the following necessary condition for stability.

Theorem 1 *Stable points of locally Lipschitz semi-algebraic functions are local minima.*

Proof Let $x^* \in \mathbb{R}^n$ denote a stable point of a locally Lipschitz semi-algebraic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. We reason by contradiction and assume that x^* is not a local minimum of f. Given $x \in \mathbb{R}^n$ and $S \subset \mathbb{R}^n$, let $d(x, S) := \inf\{\|x - y\| : y \in S\}$. According to the Kurdyka-Łojasiewicz inequality [10, Theorem 14] (see also [3, Theorem 4.1]), there exist $r, \rho > 0$ and a strictly increasing concave continuous semi-algebraic function $\psi : [0, \rho) \rightarrow [0, \infty)$ that is continuously differentiable on $(0, \rho)$ with $\psi(0) = 0$ such that $d(0, \partial f(x)) \geq 1/\psi'(|f(x) - f(x^*)|)$ for all $x \in B(x^*, r)$ whenever $0 <$

$|f(x) - f(x^*)| < \rho$. After possibly reducing r, the inequality holds for all $x \in B(x^*, r)$ such that $f(x) \neq f(x^*)$.

Let $\epsilon \in (0, r/2)$. By the definition of stability (Definition 1), there exist $\delta > 0$ and $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, the subgradient method with constant step size α initialized in $B(x^*, \delta)$ has all its iterates in $B(x^*, \epsilon)$. Since x^* is not a local minimum, we can take an initial iterate x_0 in $B(x^*, \delta)$ such that $f(x_0) < f(x^*)$. Let $L > 0$ denote a Lipschitz constant of f on $B(x^*, r)$ and let $T := (2\epsilon + \delta) L \psi'(f(x^*) - f(x_0))^2 > 0$. Consider the differential inclusion

$$x'(t) \in -\partial f(x(t)), \quad \text{for a.e. } t \in (0, T), \quad x(0) = x_0. \tag{4}$$

Since f is locally Lipschitz, the set-valued function $-\partial f$ is upper semicontinuous [14, 2.1.5 Proposition (d)] with nonempty, compact and convex values [14, 2.1.2 Proposition (a)]. By Lemma 1, there exists $\hat{\alpha} \in (0, \bar{\alpha}]$ such that, for all $\alpha \in (0, \hat{\alpha}]$ and for all sequence $(x_k)_{k \in \mathbb{N}}$ generated by the subgradient method with constant step size α and initialized at x_0, there exists a solution $x(\cdot)$ to the differential inclusion (4) for which $\|\bar{x}(t) - x(t)\| \leq \epsilon/2$ for all $t \in [0, T]$, where $\bar{x} : [0, T] \to \mathbb{R}^n$ is the piecewise linear function defined by $\bar{x}(t) := x_k + (t - \alpha k)/\alpha (x_{k+1} - x_k)$ for all $t \in [\alpha k, \min\{\alpha(k+1), T\}]$ and $k \in \{0, \ldots, \lfloor T/\alpha \rfloor\}$.

Let us fix some $\alpha \in (0, \min\{\hat{\alpha}, T/10\}]$ from now on. Consider a sequence $(x_k)_{k \in \mathbb{N}}$ generated by the subgradient method with constant step size α and initialized at x_0. Consider also the linear interpolation $\bar{x}(\cdot)$ of those iterates up to iteration $K+1$ where $K := \lfloor T/\alpha \rfloor$, as well as a solution $x(\cdot)$ to the differential inclusion (4) such that $\|\bar{x}(t) - x(t)\| \leq \epsilon/2$ for all $t \in [0, T]$. Since f is semi-algebraic, by [15, Lemma 5.2] (see also [17]) it holds that

$$f(x(t)) - f(x(0)) = -\int_0^t d(0, \partial f(x(\tau)))^2 d\tau, \quad \forall t \in [0, T]. \tag{5}$$

As a result, $f(x(t)) \leq f(x(0)) = f(x_0)$ for all $t \in [0, T]$. Also, $\|x(t) - x^*\| \leq \|x(t) - \bar{x}(t)\| + \|\bar{x}(t) - x^*\| \leq \epsilon/2 + \epsilon = 3\epsilon/2 \leq r$, where the inequality $\|\bar{x}(t) - x^*\| \leq \epsilon$ follows from the convexity of $B(x^*, \epsilon)$. Hence $0 < f(x^*) - f(x_0) \leq f(x^*) - f(x(t))$ and $d(0, \partial f(x(t))) \geq 1/\psi'(f(x^*) - f(x(t))) \geq 1/\psi'(f(x^*) - f(x_0))$ for all $t \in [0, T]$ by concavity of ψ. Together with (5), it follows that $f(x(K\alpha)) - f(x_0) \leq -K\alpha M^2$ where $K := \lfloor T/\alpha \rfloor$ and $M := 1/\psi'(f(x^*) - f(x_0))$. Recall that L is a Lipschitz constant of f on $B(x^*, r)$, so that we have $|f(x(K\alpha)) - f(x_0)| \leq L\|x(K\alpha) - x_0\|$. We thus obtain the lower bound $\|x(K\alpha) - x_0\| \geq K\alpha M^2/L = \lfloor T/\alpha \rfloor \alpha M^2/L \geq (T - \alpha) M^2/L \geq (T - T/10) M^2/L = 9/10(2\epsilon + \delta)$ (recall that $T = (2\epsilon + \delta) L/M^2$). Hence $\|x(K\alpha) - x^*\| \geq \|x(K\alpha) - x_0\| - \|x_0 - x^*\| \geq 9/10(2\epsilon + \delta) - \delta = 9\epsilon/5 - \delta/10 \geq 9\epsilon/5 - \epsilon/10 = 17\epsilon/10$ and finally $\|x_K - x^*\| \geq \|x(K\alpha) - x^*\| - \|x_K - x(K\alpha)\| \geq 17\epsilon/10 - \epsilon/2 = 6\epsilon/5 > \epsilon$. However, since $0 < \alpha \leq \hat{\alpha} \leq \bar{\alpha}$ and $x_0 \in B(x^*, \delta)$, by stability of x^* we have $x_K \in B(x^*, \epsilon)$. We have reached a contradiction. □

By appealing to Lemma 1 again, we obtain the following sufficient condition for stability.

Theorem 2 *Strict local minima of locally Lipschitz semi-algebraic functions are stable.*

Proof Let x^* denote a strict local minimum of locally Lipschitz semi-algebraic function $f : \mathbb{R}^n \to \mathbb{R}$. Since x^* is a strict local minimum, by the Łojasiewicz inequality [19, Theorem 0] (see also [22, §2], [23, §17], [18, (2.1)]) and the Kurdyka-Łojasiewicz inequality [10, Theorem 14], there exist $r, \rho > 0$ and strictly increasing continuous semi-algebraic functions $\sigma, \psi : [0, \rho) \to [0, \infty)$ that are continuously differentiable on $(0, \rho)$ with $\sigma(0) = \psi(0) = 0$ such that ψ is concave, $f(x) - f(x^*) \geq \sigma(\|x - x^*\|)$, and $d(0, \partial f(x)) \geq 1/\psi'(f(x) - f(x^*))$ for all $x \in B(x^*, r)$ whenever $0 < f(x) - f(x^*) < \rho$. After possibly reducing r, the two inequalities above hold for all $x \in B(x^*, r) \setminus \{x^*\}$. In order to prove stability, it suffices to prove the statement in Definition 1 for all $\epsilon > 0$ sufficiently small. We may thus restrict ourselves to the case where $0 < \epsilon < r$. Given such a fixed ϵ, we next describe a possible choice for δ.

Given $\Delta \geq f(x^*)$, let $L_{x^*}(f, \Delta)$ denote the connected component of the sublevel set $L(f, \Delta) := \{x \in \mathbb{R}^n : f(x) \leq \Delta\}$ containing x^*. By taking $\Delta_\epsilon := f(x^*) + \sigma(\epsilon/2)$, we find that $L_{x^*}(f, \Delta_\epsilon)$ is contained in $B(x^*, \epsilon/2)$. Indeed, one can reason by contradiction and assume that there exists $x \in L_{x^*}(f, \Delta_\epsilon) \setminus B(x^*, \epsilon/2)$. Then $x \notin B(x^*, r)$, otherwise $\sigma(\|x - x^*\|) \leq f(x) - f(x^*) \leq \sigma(\epsilon/2)$ and thus $\|x - x^*\| \leq \epsilon/2$. Therefore $L_{x^*}(f, \Delta_\epsilon)$ is the disjoint union of $L_{x^*}(f, \Delta_\epsilon) \cap \mathring{B}(x^*, r)$ and $L_{x^*}(f, \Delta_\epsilon) \setminus B(x^*, r)$, both of which are nonempty and open in $L_{x^*}(f, \Delta_\epsilon)$. This contradicts the connectedness of $L_{x^*}(f, \Delta_\epsilon)$, which yields that $L_{x^*}(f, \Delta_\epsilon) \subset B(x^*, \epsilon/2)$. By continuity of f we may choose $\delta > 0$ such that

$$B(x^*, \delta) \subset L_{x^*}(f, \Delta_\epsilon) \subset B(x^*, \epsilon/2). \tag{6}$$

We next describe a possible choice for $\bar{\alpha}$. Let $L > \sup\{\|s\| : s \in \partial f(x), x \in B(x^*, \epsilon)\}$ be a Lipschitz constant of f on $B(x^*, \epsilon)$ and let $T := \epsilon/(3L)$, where the supremum is finite due to [4, Proposition 3 p. 42]. Consider the differential inclusion

$$x'(t) \in -\partial f(x(t)), \quad \text{for a.e. } t \in (0, T), \quad x(0) \in L_{x^*}(f, \Delta_\epsilon). \tag{7}$$

Since f is continuous, the set of initial values $L_{x^*}(f, \Delta_\epsilon)$ is closed. By virtue of the second inclusion in (6), $L_{x^*}(f, \Delta_\epsilon)$ is in fact a compact set. Let $\alpha \in (0, T/2]$ and consider a sequence $(x_k)_{k \in \mathbb{N}}$ generated by the subgradient method with constant step size α and initialized in $L_{x^*}(f, \Delta_\epsilon)$. According to the second inclusion in (6), the initial iterate x_0 lies in $B(x^*, \epsilon/2)$. Hence $\|x_1 - x^*\| = \|x_0 - \alpha s_0 - x^*\| \leq \|x_0 - x^*\| + \|\alpha s_0\| \leq \epsilon/2 + L\alpha$ for some $s_0 \in \partial f(x_0)$. Repeating this process until iteration $K + 1$ where $K := \lfloor T/\alpha \rfloor$, we find that $\|x_k - x^*\| \leq \epsilon/2 + kL\alpha \leq \epsilon/2 + (K+1)L\alpha \leq \epsilon/2 + (T/\alpha + 1)L\alpha = \epsilon/2 + (\epsilon/(3L\alpha) + 1)L\alpha = 5\epsilon/6 + L\alpha \leq 5\epsilon/6 + LT/2 = 5\epsilon/6 + L\epsilon/(3L)/2 = \epsilon$. In other words, the iterates $x_0, \ldots, x_{K+1}$ lie in $B(x^*, \epsilon)$. Let $\epsilon' := \min\{\epsilon L, \sigma(\epsilon/2), \xi^2 T\}/(2L) > 0$ where $\xi := 1/\psi'(\sigma(\epsilon/2)/2) > 0$. All of the conditions of Lemma 1 are met, hence there exists $\bar{\alpha} \in (0, T/2]$ such that, for all $\alpha \in (0, \bar{\alpha}]$ and for all sequence $(x_k)_{k \in \mathbb{N}}$ generated by the subgradient method with constant step size α and initialized in $L_{x^*}(f, \Delta_\epsilon)$, there exists a solution to the differential inclusion (7) for which $\|\bar{x}(t) - x(t)\| \leq \epsilon'$ for all $t \in [0, T]$ where $\bar{x} : [0, T] \to \mathbb{R}^n$
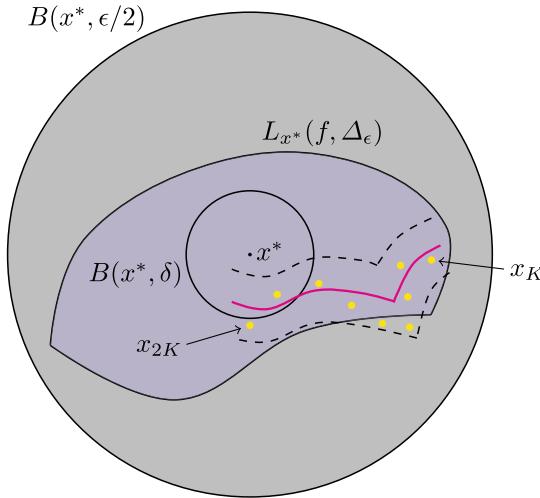
Fig. 1 Induction step

is the piecewise linear function defined by $\bar{x}(t) := x_k + (t - \alpha k)/\alpha(x_{k+1} - x_k)$ for all $t \in [\alpha k, \min\{\alpha(k+1), T\}]$ and $k \in \{0, \ldots, \lfloor T/\alpha \rfloor\}$. In particular, it holds that

$$\|x_k - x(k\alpha)\| \le \epsilon', \quad k = 0, \ldots, \lfloor T/\alpha \rfloor. \tag{8}$$

Having chosen δ and $\bar{\alpha}$, let us fix some $\alpha \in (0, \bar{\alpha}]$ from now on. Consider a sequence $(x_k)_{k\in\mathbb{N}}$ generated by the subgradient method with constant step size α and initialized in $B(x^*, \delta)$. Our goal is to show that all the iterates lie in $B(x^*, \epsilon)$. According to the first inclusion in (6), the initial iterate x_0 lies in $L_{x^*}(f, \Delta_\epsilon)$. A previous argument shows that the iterates $x_0, \ldots, x_K$ lie in $B(x^*, \epsilon)$ where $K := \lfloor T/\alpha \rfloor$. In order to show that the ensuing iterates also lie in $B(x^*, \epsilon)$, we will show that $x_K \in L_{x^*}(f, \Delta_\epsilon)$. The same argument used previously then yields that $x_{K+1}, \ldots, x_{2K} \in B(x^*, \epsilon)$. Since $K = \lfloor T/\alpha \rfloor \ge \lfloor T/\bar{\alpha} \rfloor \ge 2$, we may conclude by induction that all the iterates belong to $B(x^*, \epsilon)$. This is illustrated in Fig. 1.

For the remainder of the proof, we seek to show that $x_K \in L_{x^*}(f, \Delta_\epsilon)$. In order to do so, we prove that $B(x(K\alpha), \epsilon')$ is a connected subset of $L(f, \Delta_\epsilon)$ that has nonempty intersection with $L_{x^*}(f, \Delta_\epsilon)$. Since $L_{x^*}(f, \Delta_\epsilon)$ is a connected component of $L(f, \Delta_\epsilon)$, by maximality and (8) we then have $x_K \in B(x(K\alpha), \epsilon') \subset L_{x^*}(f, \Delta_\epsilon)$.

We begin by showing that $x(K\alpha) \in B(x(K\alpha), \epsilon') \cap L_{x^*}(f, \Delta_\epsilon)$. Since f is semi-algebraic, by [15, Lemma 5.2] (see also [17]) it holds that

$$f(x(t)) - f(x(0)) = -\int_0^t d(0, \partial f(x(\tau)))^2 d\tau, \quad \forall t \in [0, T]. \tag{9}$$
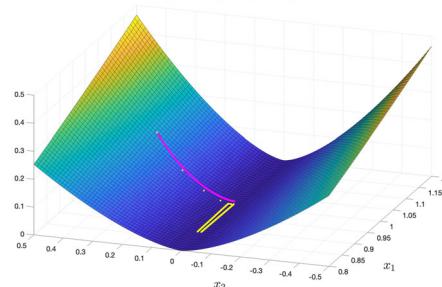
As a result, $f(x(t)) \le f(x(0)) \le \Delta_\epsilon$ for all $t \in [0, T]$. We thus have that $x(0) \in L_{x^*}(f, \Delta_\epsilon) \cap x([0, T])$, both of which are connected subsets of the sublevel

$f(x_1, x_2) = \max\{-18x_1^2 + 12|x_2|, 6x_1^2 + 3|x_2|\}$
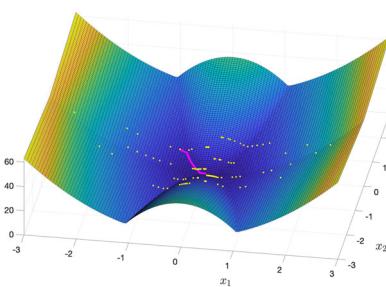
$f(x_1, x_2) = |x_1 x_2|^{3/2}$

(a) Stable local minimum.

(b) Unstable local minimum.

Fig. 2 Continuous and discrete subgradient trajectories in magenta and yellow respectively

set $L(f, \Delta_\epsilon)$. By maximality of $L_{x^*}(f, \Delta_\epsilon)$, it follows that $x(K\alpha) \in x([0, T]) \subset L_{x^*}(f, \Delta_\epsilon)$.

We next show that $B(x(K\alpha), \epsilon') \subset L(f, \Delta_\epsilon)$. For all $\tilde{x} \in B(x(K\alpha), \epsilon')$, we have

$$f(\tilde{x}) - f(x^*) = f(\tilde{x}) - f(x(K\alpha)) + f(x(K\alpha)) - f(x^*) \tag{10a}$$

$$\leq L\|\tilde{x} - x(K\alpha)\| + \max\{\sigma(\epsilon/2)/2, \sigma(\epsilon/2) - \xi^2 T/2\} \tag{10b}$$

$$\leq L\epsilon' + \sigma(\epsilon/2) - \min\{\sigma(\epsilon/2)/2, \xi^2 T/2\} \tag{10c}$$

$$\leq \sigma(\epsilon/2). \tag{10d}$$

Indeed, $\tilde{x}$ and $x(K\alpha)$ belong to $B(x^*, \epsilon)$ so we may invoke the Lipschitz constant L of f on $B(x^*, \epsilon)$ in order to bound the first term in (10a). Recall from the previous paragraph that $x(K\alpha) \in L_{x^*}(f, \Delta_\epsilon) \subset B(x^*, \epsilon/2)$ and, since $\epsilon' \leq \epsilon/2$, we have $\tilde{x} \in B(x(K\alpha), \epsilon') \subset B(x^*, \epsilon)$. As for the second term in (10a), if it is greater than or equal to $\sigma(\epsilon/2)/2$, then for all $t \in [0, K\alpha]$, we have $\sigma(\epsilon/2)/2 \leq f(x(K\alpha)) - f(x^*) \leq f(x(t)) - f(x^*)$ and thus $d(0, \partial f(x(t))) \geq 1/\psi'(f(x(t)) - f(x^*)) \geq 1/\psi'(\sigma(\epsilon/2)/2) = \xi$. By (9), it follows that $f(x(K\alpha)) - f(x^*) \leq f(x(0)) - f(x^*) - \int_0^{K\alpha} \xi^2 d\tau \leq f(x(0)) - f(x^*) - K\alpha\xi^2 \leq \sigma(\epsilon/2) - \xi^2 T/2$. The last inequality is due to the fact that $x(0) \in L_{x^*}(f, \Delta_\epsilon)$ and $K\alpha = \lfloor T/\alpha \rfloor \alpha \geq T - \alpha \geq T - \bar{\alpha} \geq T/2$. In (10c), we use the fact that $\tilde{x} \in B(x(K\alpha), \epsilon')$ and rewrite the maximum into a minimum. Finally, (10d) holds because $\epsilon' \leq \min\{\sigma(\epsilon/2), \xi^2 T\}/(2L)$. □

Observe that non-strict local minima need not be stable. While the strict local minimum in Fig. 2a is stable by Theorem 2, the non-strict local minimum in Fig. 2b is unstable. In the former, both continuous and discrete subgradient dynamics are stable, with the continuous trajectory converging to the local minimum, while the discrete trajectory hovers around it. In the latter, continuous and discrete trajectories become decoupled as they approach the local minimum; the continuous dynamics are stable but the discrete dynamics are not.

We conclude this note by proving that the local minimum in Fig. 2b is unstable with respect to the Euclidean inner product. We show that there exists $\epsilon > 0$ such that for all but finitely many constant step sizes $\alpha > 0$ and for almost every initial point

in $B(x^*, \epsilon)$, at least one of the iterates of the subgradient method does not belong to $B(x^*, \epsilon)$. Let $\epsilon \in (0, 1/2)$ and consider the set $S := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 0\}$. By the cell decomposition theorem [27, (2.11) p. 52] and [12, Claim 3], there exist $\alpha_1, \ldots, \alpha_m > 0$ such that for all constant step sizes $\alpha \in (0, \infty) \setminus \{\alpha_1, \ldots, \alpha_m\}$, there exists a null subset $I_\alpha \subset \mathbb{R}^2$ such that, for every initial point $(x_1^0, x_2^0) \in \mathbb{R}^2 \setminus I_\alpha$, none of the iterates $(x_1^k, x_2^k)_{k \in \mathbb{N}}$ of the subgradient method belong to the semi-algebraic null set S. In this case, the update rule of the subgradient method is given for all $k \in \mathbb{N}$ by

$$x_1^{k+1} = x_1^k - \frac{3}{2}\alpha |x_1^k|^{1/2} |x_2^k|^{3/2} \text{sign}(x_1^k),$$

$$x_2^{k+1} = x_2^k - \frac{3}{2}\alpha |x_1^k|^{3/2} |x_2^k|^{1/2} \text{sign}(x_2^k),$$

where $\text{sign}(x) := -1$ if $x < 0$ and $\text{sign}(x) := 1$ if $x > 0$. Assume that $(x_1^k, x_2^k) \in B((1, 0), \epsilon)$ for all $k \in \mathbb{N}$. Since $\epsilon \leq 1/2$, we have $x_1^k \geq 1/2$. If $0 < |x_2^k| \leq \alpha^2/32$ for some $k \in \mathbb{N}$, then $|x_2^{k+1}| = |x_2^k - 3\alpha/2|x_1^k|^{3/2}|x_2^k|^{1/2}\text{sign}(x_2^k)| \geq 3\alpha/2|x_1^k|^{3/2}|x_2^k|^{1/2} - |x_2^k| \geq (3\alpha/\sqrt{32|x_2^k|} - 1)|x_2^k| \geq 2|x_2^k|$. As a result, x_2^k does not converge to zero. This yields the following contradiction:

$$\frac{1}{2} \leq x_1^{k+1} = x_1^0 - \frac{3\alpha}{2} \sum_{i=0}^{k} |x_1^i|^{1/2} |x_2^i|^{3/2} \leq x_1^0 - \frac{3\alpha}{2\sqrt{2}} \sum_{i=0}^{k} |x_2^i|^{3/2} \to -\infty.$$

References

1. Absil, P.A., Kurdyka, K.: On the stable equilibrium points of gradient systems. Syst. Control Lett. **55**(7), 573–577 (2006)
2. Absil, P.A., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**(2), 531–547 (2005)
3. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. Math. Oper. Res. **35**, 438–457 (2010)
4. Aubin, J.P., Cellina, A.: Differential Inclusions: Set-Valued Maps and Viability Theory, vol. 264. Springer, Berlin (1984)
5. Benaïm, M., Hofbauer, J., Sorin, S.: Stochastic approximations and differential inclusions. SIAM J. Control Optim. **44**(1), 328–348 (2005)
6. Bertsekas, D.: Convex Optimization Algorithms. Athena Scientific, Nashua (2015)
7. Bianchi, P., Hachem, W., Salim, A.: Constant step stochastic approximations involving differential inclusions: stability, long-run convergence and applications. Stochastics **91**(2), 288–320 (2019)
8. Bianchi, P., Hachem, W., Schechtman, S.: Convergence of constant step stochastic gradient descent for non-smooth non-convex functions. Set-Valued Var. Anal. **30**(3), 1117–1147 (2022)
9. Bochnak, J., Coste, M., Roy, M.F.: Real Algebraic Geometry, vol. 36. Springer, Berlin (2013)
10. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
11. Bolte, J., Pauwels, E.: Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. Math. Program. **188**, 19–51 (2020)

12. Bolte, J., Pauwels, E.: A mathematical model for automatic differentiation in machine learning. NeurIPS **33**, 10809–10819 (2020)
13. Charisopoulos, V., Chen, Y., Davis, D., Díaz, M., Ding, L., Drusvyatskiy, D.: Low-rank matrix recovery with composite optimization: good conditioning and rapid convergence. Found. Comput. Math. **21**(6), 1505–1593 (2021)
14. Clarke, F.H.: Optimization and non smooth analysis. In: SIAM Classics in Applied Mathematics (1990)
15. Davis, D., Drusvyatskiy, D., Kakade, S., Lee, J.D.: Stochastic subgradient method converges on tame functions. Found. Comput. Math. **20**(1), 119–154 (2020)
16. Davis, D., Drusvyatskiy, D., MacPhee, K.J., Paquette, C.: Subgradient methods for sharp weakly convex functions. J. Optim. Theory Appl. **179**(3), 962–982 (2018)
17. Drusvyatskiy, D., Ioffe, A.D., Lewis, A.S.: Curves of descent. SIAM J. Control Optim. **53**(1), 114–138 (2015)
18. Hörmander, L.: On the division of distributions by polynomials. Arkiv för matematik **3**(6), 555–568 (1958)
19. Kurdyka, K.: On gradients of functions definable in o-minimal structures. In: Annales de l'institut Fourier, vol 48, pp 769–783 (1998)
20. Lebourg, G.: Generic differentiability of lipschitzian functions. Trans. Am. Math. Soc. **256**, 125–144 (1979)
21. Liapounoff, A.: Problème général de la stabilité du mouvement. In: Annales de la Faculté des sciences de Toulouse: Mathématiques, vol. 9, pp. 203–474 (1907)
22. Łojasiewicz, S.: Division d'une distribution par une fonction analytique de variables réelles. Comptes rendus hebdomadaires des séances de l'Académie des sciences. Paris, pp. 683–686 (1958)
23. Łojasiewicz, S.: Sur le problème de la division. Studia Mathematica, pp. 87–136 (1959)
24. Pham, T.S., Vui, H.H.: Genericity in Polynomial Optimization, vol. 3. World Scientific, Singapore (2016)
25. Rockafellar, R.T.: Favorable classes of Lipschitz continuous functions in subgradient optimization. IIASA Working Paper (1981)
26. Sastry, S.: Nonlinear Systems: Analysis, Stability, and Control, vol. 10. Springer, Berlin (2013)
27. Van den Dries, L.: Tame Topology and O-Minimal Structures, vol. 248. Cambridge University Press, Cambridge (1998)
28. Van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**(2), 497–540 (1996)