Global convergence of the gradient method for functions definable in o-minimal structures

Cédric Josz

IEOR Columbia University

Dynamical systems and Semi-algebraic geometry: interactions with Optimization and Deep Learning

University of Dalat, Vietnam

July 20$^{\text{th}}$, 2023

O-minimal structures on the real field

van den Dries, 1996

An *o-minimal structure* on the real field is a sequence $S = (S_k)_{k \in \mathbb{N}}$ such that for all $k \in \mathbb{N}$:

1. S_k is a boolean algebra of subsets of $\mathbb{R}^k$, with $\mathbb{R}^k \in S_k$;
2. S_k contains the diagonal $\{(x_1, \dots, x_k) \in \mathbb{R}^k : x_i = x_j\}$ for $1 \leqslant i < j \leqslant k$;
3. If $A \in S_k$, then $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to S_{k+1};
4. If $A \in S_{k+1}$ and $\pi : \mathbb{R}^{k+1} \to \mathbb{R}^k$ is the projection onto the first k coordinates, then $\pi(A) \in S_k$;
5. S_3 contains the graphs of addition and multiplication;
6. S_1 consists exactly of the finite unions of open intervals and singletons.

$\rightarrow$ for e.g., semi-algebraic sets (Tarski 1951, Seidenberg 1954), globally subanalytic sets (Gabrielov, 1968), $(\mathbb{R}, \exp)$ (Wilkie, 1996), ...

Examples of functions definable in o-minimal structures

1. Semi-algebraic functions (Tarski 1951, Seidenberg 1954):

$$f(x,y) = \sqrt{x^4 + y^4}$$

2. Globally subanalytic functions (Gabrielov 1968);

$$f(x,y) = \frac{y}{\sin x}, \quad x \in (0,\pi)$$

3. $(\mathbb{R}, \exp)$ – definable functions (Wilkie 1996):

$$f(W_1,\ldots,W_L) = (W_L\sigma(W_{L-1}\cdots\sigma(W_1 x)) - y)^2, \quad \sigma(t) = \frac{1}{1+e^{-t}}$$

4. $(\mathbb{R}_{\mathrm{an}}, \exp)$ – definable functions (van den Dries *et al.* 1994):

$$f(x,y) = |x|^{\sqrt{2}} \ln(\sin y), \quad y \in (0,\pi).$$

Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier, tome 48, no 3 (1998), p. 769-783.

Gradient method

$$\inf_{x \in \mathbb{R}^n} f(x) \quad \text{where} \quad f : \mathbb{R}^n \to \mathbb{R} \quad \text{is differentiable}$$

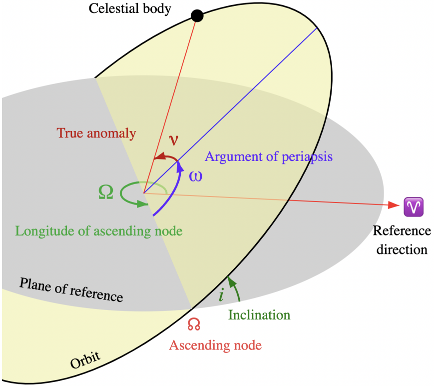
$$\begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \alpha_k \nabla f(x_k), \;\; k = 0, 1, 2, \ldots \end{cases}$$

$\alpha_k > 0$ are the step sizes, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is the gradient wrt $\langle \cdot, \cdot \rangle$

- Proposed by Cauchy in 1847 for solving $f_1(x) = \cdots = f_m(x) = 0$ by setting $f(x) = f_1(x)^2 + \cdots + f_m(x)^2$

- Application to orbits of astronomical bodies (six unknowns)

- Lemaréchal 2012: *"some form of uniformity is required from the objective's continuity"*

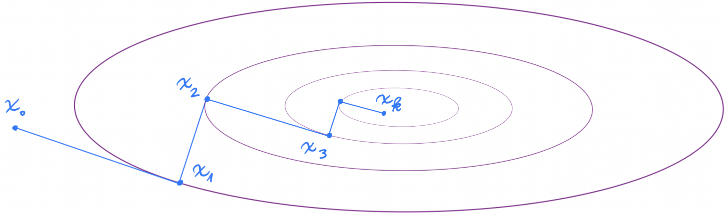


Literature review: first results

1. If f is a strongly convex quadratic, then exact line search yields $\|\nabla f(x_k)\| \to 0$ (Temple 1939).
2. If f is continuously differentiable and coercive, then the Curry step yields $\|\nabla f(x_k)\| \to 0$ (1944).
3. If f is a convex quadratic, $\inf f = 0$, and exact line search is used, then by the Kantorovich inequality (1948)

$$f(x_{k+1}) \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 f(x_k)$$

4. Forsythe and Motkzin (1951), Akaike (1959) argue that the rate is close to its worst possible value for ill-conditioned quadratics.

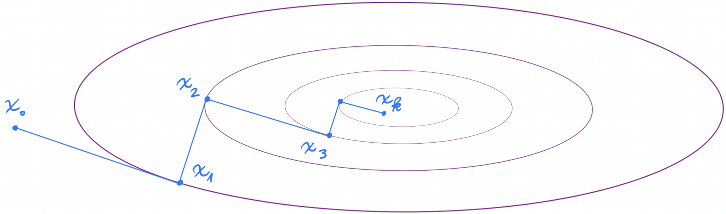


Literature review: first results

1. If f is a strongly convex quadratic, then exact line search yields $\|\nabla f(x_k)\| \to 0$ (Temple 1939).
2. If f is continuously differentiable and coercive, then the Curry step yields $\|\nabla f(x_k)\| \to 0$ (1944).
3. If f is a convex quadratic, $\inf f = 0$, and exact line search is used, then by the Kantorovich inequality (1948)

$$f(x_{k+1}) \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 f(x_k)$$

4. Forsythe and Motkzin (1951), Akaike (1959) argue that the rate is close to its worst possible value for ill-conditioned quadratics.

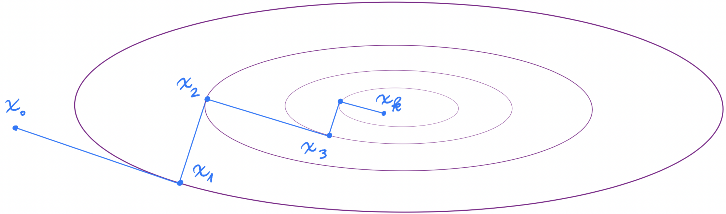


Literature review: first results

1. If f is a strongly convex quadratic, then exact line search yields $\|\nabla f(x_k)\| \to 0$ (Temple 1939).
2. If f is continuously differentiable and coercive, then the Curry step yields $\|\nabla f(x_k)\| \to 0$ (1944).
3. If f is a convex quadratic, $\inf f = 0$, and exact line search is used, then by the Kantorovich inequality (1948)

$$f(x_{k+1}) \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 f(x_k)$$

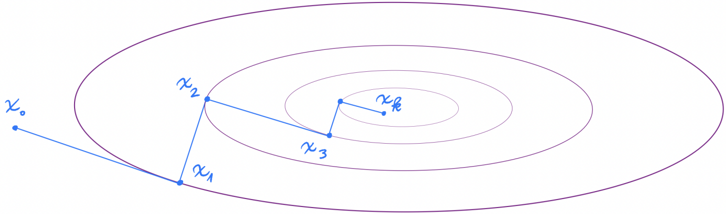
4. Forsythe and Motkzin (1951), Akaike (1959) argue that the rate is close to its worst possible value for ill-conditioned quadratics.



1. If f is a strongly convex quadratic, then exact line search yields $\|\nabla f(x_k)\| \to 0$ (Temple 1939).
2. If f is continuously differentiable and coercive, then the Curry step yields $\|\nabla f(x_k)\| \to 0$ (1944).
3. If f is a convex quadratic, $\inf f = 0$, and exact line search is used, then by the Kantorovich inequality (1948)

$$f(x_{k+1}) \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 f(x_k)$$

4. Forsythe and Motkzin (1951), Akaike (1959) argue that the rate is close to its worst possible value for ill-conditioned quadratics.



Literature review: first results

1. If f is a strongly convex quadratic, then exact line search yields $\|\nabla f(x_k)\| \to 0$ (Temple 1939).
2. If f is continuously differentiable and coercive, then the Curry step yields $\|\nabla f(x_k)\| \to 0$ (1944).
3. If f is a convex quadratic, $\inf f = 0$, and exact line search is used, then by the Kantorovich inequality (1948)

$$f(x_{k+1}) \leqslant \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 f(x_k)$$

4. Forsythe and Motkzin (1951), Akaike (1959) argue that the rate is close to its worst possible value for ill-conditioned quadratics.
5. Goldstein (1965), Armijo (1966), and Wolfe (1969) propose inexact line search methods.

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:

$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$

2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally

$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$

3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally

$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$

4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small

$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1} t^{1-\theta}$$

5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:
$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$
2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally
$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$
3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally
$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$
4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small
$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1} t^{1-\theta}$$
5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise
$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:
$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$
2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally
$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$
3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally
$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$
4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small
$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1} t^{1-\theta}$$
5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise
$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:
$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$
2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally
$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$
3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally
$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$
4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small
$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1}t^{1-\theta}$$
5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise
$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:
$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$
2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally
$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$
3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally
$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$
4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small
$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1} t^{1-\theta}$$
5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise
$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: fundamental inequalities

1. If f is polynomial (Hörmander 1958) or analytic (Łojasiewicz 1958) and $f(0) = 0$, then in any bounded set containing zero $\exists c, \theta > 0$:
$$|f(x)| \geqslant c\, d(x, f^{-1}(0))^{\theta}$$
2. Polyak (1962) proves linear convergence if $\inf f = 0$ and $\exists c > 0$: globally
$$\|\nabla f(x)\| \geqslant c f(x)^{1/2}$$
3. If f is analytic (Łojasiewicz 1963) and $f(0) = 0$, then $\exists c > 0$, $\exists \theta \in (0,1)$: locally
$$\|\nabla f(x)\| \geqslant c|f(x)|^{\theta}$$
4. If f is definable and differentiable (Kurdyka 1998), then in any bounded set $\exists \psi : \mathbb{R}_+ \to \mathbb{R}_+$: for $|f|$ small
$$\|\nabla(\psi \circ |f|)(x)\| \geqslant 1 \quad \psi(t) = c^{-1}(1-\theta)^{-1} t^{1-\theta}$$
5. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then likewise
$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

Literature review: state-of-the-art convergence results

1. If f is $C^{1,1}_{\text{loc}}$ definable, bounded iterates converge to critical point x^* if
   - Wolfe's line search is used (Absil *et al.* 2005) or
   - f is $C^{1,1}_L$ and $\alpha_k \in (0, 2/L - \epsilon]$ non-summable (Attouch *et al.* '13).
2. If $f \in C^{1,1}_L \cap C^2$, $x_k \to x^*$, $\alpha_k := \alpha \in (0, 2/L)$, and $\lambda_{\min}(\nabla^2 f(x)) < 0$

   $$\text{for all } x \in \mathbb{R}^n : \ \nabla f(x) = 0 \ \ \& \ \ x \text{ is not a local minimum,}$$

   then x^* is a local minimum almost surely

   (Shub 1986, Pemantle 1990, Lee *et al.* 2016, Panageas and Piliouras 2017).
3. If f is $C^{1,1}_L$ definable in a polynomially bounded o-minimal structure, $x_k \to x^*$, $\alpha_k \in [\underline{\alpha}, \overline{\alpha}] \subset (0, 2/L)$, then $\exists p > 0$, $\exists q \in (0,1)$:

   $$\|x_k - x^*\| \leqslant \begin{cases} pq^k & \text{if } \theta \in (0, 1/2] \\ pk^{-(1-\theta)(2\theta-1)} & \text{if } \theta \in (1/2, 1). \end{cases}$$

Literature review: state-of-the-art convergence results

1. If f is $C^{1,1}_{\text{loc}}$ definable, bounded iterates converge to critical point x^* if
   - Wolfe's line search is used (Absil *et al.* 2005) or
   - f is $C^{1,1}_L$ and $\alpha_k \in (0, 2/L - \epsilon]$ non-summable (Attouch *et al.* '13).
2. If $f \in C^{1,1}_L \cap C^2$, $x_k \to x^*$, $\alpha_k := \alpha \in (0, 2/L)$, and $\lambda_{\min}(\nabla^2 f(x)) < 0$

$$\text{for all } x \in \mathbb{R}^n : \; \nabla f(x) = 0 \;\;\&\;\; x \text{ is not a local minimum,}$$

$$\text{then } x^* \text{ is a local minimum almost surely}$$

(Shub 1986, Pemantle 1990, Lee *et al.* 2016, Panageas and Piliouras 2017).

3. If f is $C^{1,1}_L$ definable in a polynomially bounded o-minimal structure, $x_k \to x^*$, $\alpha_k \in [\underline{\alpha}, \overline{\alpha}] \subset (0, 2/L)$, then $\exists p > 0$, $\exists q \in (0,1)$:

$$\|x_k - x^*\| \leqslant \begin{cases} pq^k & \text{if } \theta \in (0, 1/2] \\ pk^{-(1-\theta)(2\theta-1)} & \text{if } \theta \in (1/2, 1). \end{cases}$$

Literature review: state-of-the-art convergence results

1. If f is $C^{1,1}_{\text{loc}}$ definable, bounded iterates converge to critical point x^* if
   - Wolfe's line search is used (Absil *et al.* 2005) or
   - f is $C^{1,1}_L$ and $\alpha_k \in (0, 2/L - \epsilon]$ non-summable (Attouch *et al.* '13).
2. If $f \in C^{1,1}_L \cap C^2$, $x_k \to x^*$, $\alpha_k := \alpha \in (0, 2/L)$, and $\lambda_{\min}(\nabla^2 f(x)) < 0$

   $$\text{for all } x \in \mathbb{R}^n : \ \nabla f(x) = 0 \ \ \& \ \ x \text{ is not a local minimum,}$$

   $$\text{then } x^* \text{ is a local minimum almost surely}$$

(Shub 1986, Pemantle 1990, Lee *et al.* 2016, Panageas and Piliouras 2017).

3. If f is $C^{1,1}_L$ definable in a polynomially bounded o-minimal structure, $x_k \to x^*$, $\alpha_k \in [\underline{\alpha}, \overline{\alpha}] \subset (0, 2/L)$, then $\exists p > 0$, $\exists q \in (0,1)$:

$$\|x_k - x^*\| \leqslant \begin{cases} p q^k & \text{if } \theta \in (0, 1/2] \\ p k^{-(1-\theta)(2\theta-1)} & \text{if } \theta \in (1/2, 1). \end{cases}$$

Literature review: state-of-the-art convergence results

1. If f is $C^{1,1}_{\text{loc}}$ definable, bounded iterates converge to critical point x^* if
   - Wolfe's line search is used (Absil *et al.* 2005) or
   - f is $C^{1,1}_L$ and $\alpha_k \in (0, 2/L - \epsilon]$ non-summable (Attouch *et al.* '13).
2. If $f \in C^{1,1}_L \cap C^2$, $x_k \to x^*$, $\alpha_k := \alpha \in (0, 2/L)$, and $\lambda_{\min}(\nabla^2 f(x)) < 0$

   $$\text{for all } x \in \mathbb{R}^n : \ \nabla f(x) = 0 \ \ \& \ \ x \text{ is not a local minimum,}$$

   $$\text{then } x^* \text{ is a local minimum almost surely}$$

(Shub 1986, Pemantle 1990, Lee *et al.* 2016, Panageas and Piliouras 2017).

3. If f is $C^{1,1}_L$ definable in a polynomially bounded o-minimal structure, $x_k \to x^*$, $\alpha_k \in [\underline{\alpha}, \overline{\alpha}] \subset (0, 2/L)$, then $\exists p > 0$, $\exists q \in (0,1)$:

$$\|x_k - x^*\| \leqslant \begin{cases} pq^k & \text{if } \theta \in (0, 1/2] \\ pk^{-(1-\theta)(2\theta-1)} & \text{if } \theta \in (1/2, 1). \end{cases}$$

Global convergence

Theorem (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is $C^{1,1}_{\text{loc}}$ definable and any $x : [0, T) \to \mathbb{R}^n$ such that

$$x' = -\nabla f(x) \quad \text{on} \quad (0, T)$$

is bounded, then $\forall x_0 \in \mathbb{R}^n, \ \exists \bar{\alpha} > 0 : \ \forall \alpha \in (0, \bar{\alpha}]$, the sequence

$$\begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \alpha \nabla f(x_k), \ k = 0, 1, 2, \ldots \end{cases}$$

converges to a critical point of f and

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\| \ = \ o(1/k).$$

N.B.: convex functions have bounded flows (Bruck 1975) and admit the same convergence rate if $C^{1,1}_{\text{loc}}$ (Nesterov 2012, Lee and Wright 2019)

Global convergence

Theorem (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is $C^{1,1}_{\text{loc}}$ definable and any $x : [0, T) \to \mathbb{R}^n$ such that

$$x' = -\nabla f(x) \quad \text{on} \quad (0, T)$$

is bounded, then $\forall x_0 \in \mathbb{R}^n, \; \exists \bar{\alpha} > 0 : \; \forall \alpha \in (0, \bar{\alpha}]$, the sequence

$$\begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k - \alpha \nabla f(x_k), \; k = 0, 1, 2, \ldots \end{cases}$$

converges to a critical point of f and

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\| \; = \; o(1/k).$$

N.B.: convex functions have bounded flows (Bruck 1975) and admit the same convergence rate if $C^{1,1}_{\text{loc}}$ (Böhm and Daniilidis 2022)

Global convergence (general case)

Theorem (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is $C^{1,1}_{\text{loc}}$ definable and any $x : [0, T) \to \mathbb{R}^n$ such that

$$x' = -\nabla f(x) \quad \text{on} \quad (0, T)$$

is bounded, then $\forall X_0 \subset \mathbb{R}^n$ bounded, $\exists \bar{\alpha}, c > 0 : \ \forall \alpha_0, \alpha_1, \ldots \in (0, \bar{\alpha}]$, the sequences

$$\begin{cases} x_0 \in X_0 \\ x_{k+1} = x_k - \alpha_k \nabla f(x_k), \ k = 0, 1, 2, \ldots \end{cases}$$

obey $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| \leqslant c$. If $\underline{\alpha} := \inf \alpha_k > 0$, then

$$\min_{i=0,\ldots,k} \|\nabla f(x_i)\| \ \leqslant \ \frac{2\underline{\alpha}^{-1}}{k+2} \sum_{i=\lfloor k/2 \rfloor}^{\infty} \|x_{i+1} - x_i\|.$$

Global convergence to a local minimum

Corollary (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 definable with strict saddles and any $x : [0, T) \to \mathbb{R}^n$ such that

$$x' = -\nabla f(x) \quad \text{on} \quad (0, T)$$

is bounded, then $\forall X_0 \subset \mathbb{R}^n$ bounded, $\exists \bar{\alpha} > 0 : \; \forall \alpha \in (0, \bar{\alpha}]$, the sequences

$$\begin{cases} x_0 \in X_0 \\ x_{k+1} = x_k - \alpha \nabla f(x_k), \; k = 0, 1, 2, \ldots \end{cases}$$

converge to a local minimum of f almost surely.

N.B.: strict saddles are saddles whose Hessian has a negative eigenvalue

Applications

The following problems

1. principal component analysis (Baldi *et al.* 1989)

$$\inf_{X,Y} \quad \|XY^\top - M\|_F^2$$

2. matrix sensing (Li *et al.* 2020)

$$\inf_{X,Y} \quad \sum_{i=1}^{m} (\langle A_i, XY^\top \rangle - y_i)^2$$

3. linear neural networks (Kawaguchi 2016, Laurent and von Brecht 2018, Zhang 2019)

$$\inf_{W_1,\ldots,W_L} \quad \|W_L \cdots W_1 X - Y\|_F^2$$

have bounded continuous gradient trajectories by

J. and Li, Certifying the absence of spurious local minima at infinity, 2023

Applications

The following problems

1. principal component analysis (Baldi *et al.* 1989) → global convergence to global minimum almost surely

$$\inf_{X,Y} \quad \|XY^\top - M\|_F^2$$

2. matrix sensing (Li *et al.* 2020) → global convergence to global minimum almost surely

$$\inf_{X,Y} \quad \sum_{i=1}^{m} (\langle A_i, XY^\top \rangle - y_i)^2$$

3. linear neural networks (Kawaguchi 2016, Laurent and von Brecht 2018, Zhang 2019) → global convergence to critical point

$$\inf_{W_1,\ldots,W_L} \quad \|W_L \cdots W_1 X - Y\|_F^2$$

Application to principal component analysis

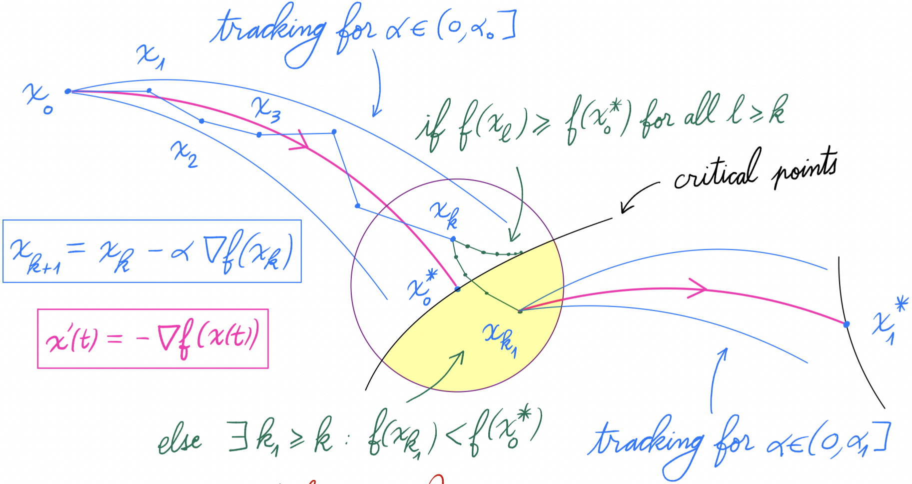
$$\boxed{\begin{array}{c}\inf_{X,Y} \|XY^\top - M\|_F^2 \\ \begin{cases} \dot{X} = -2(XY^\top - M)Y \\ \dot{Y} = -2(XY^\top - M)^\top X \end{cases}\end{array}}$$

X and Y are bounded because

$$\overline{X^\top X - Y^\top Y}^{\,\cdot} = \dot{X}^\top X + X^\top \dot{X} - \dot{Y}^\top Y - Y^\top \dot{Y} = 0$$

and

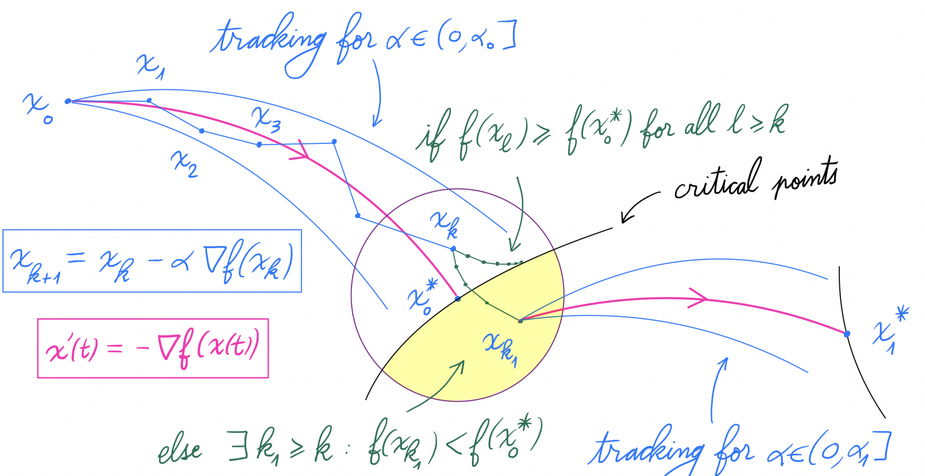
$$\begin{aligned} \|X\|_2^4 + \|Y\|_2^4 &\leqslant \|X^\top X\|^2 + \|Y^\top Y\|^2 \\ &= \|X^\top X - Y^\top Y\|^2 + 2\langle X^\top X, Y^\top Y\rangle \\ &= \|X^\top X - Y^\top Y\|^2 + 2\|XY^\top\|^2 \\ &\leqslant \|X^\top X - Y^\top Y\|^2 + 2(\|XY^\top - M\| + \|M\|)^2 \\ &\leqslant \|X(0)^\top X(0) - Y(0)^\top Y(0)\|^2 + \\ &\quad\; 2(\|X(0)Y(0)^\top - M\| + \|M\|)^2. \end{aligned}$$



1) What if $\alpha_1 < \alpha_0$?

2) How to get close to a critical point ?

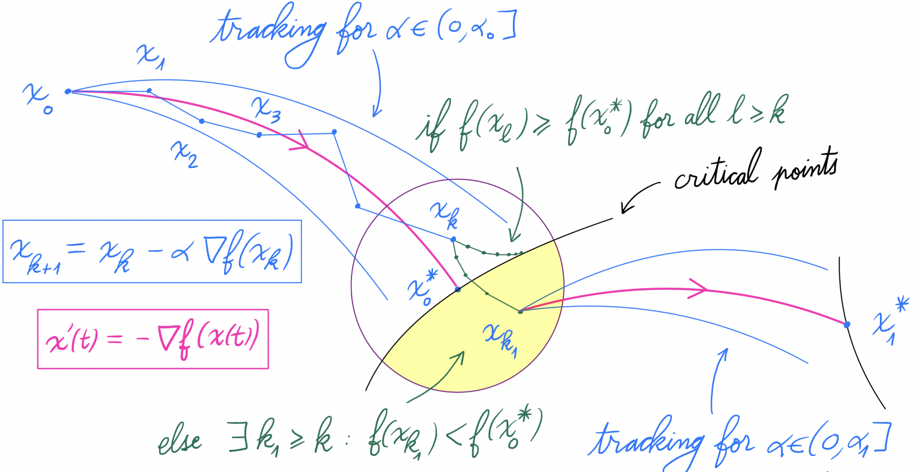
3) How to guarantee $f(x_0^*) > f(x_1^*)$?



1) What if $\alpha_1 < \alpha_0$? INITIALIZE IN A BOUNDED SET

2) How to get close to a critical point ?

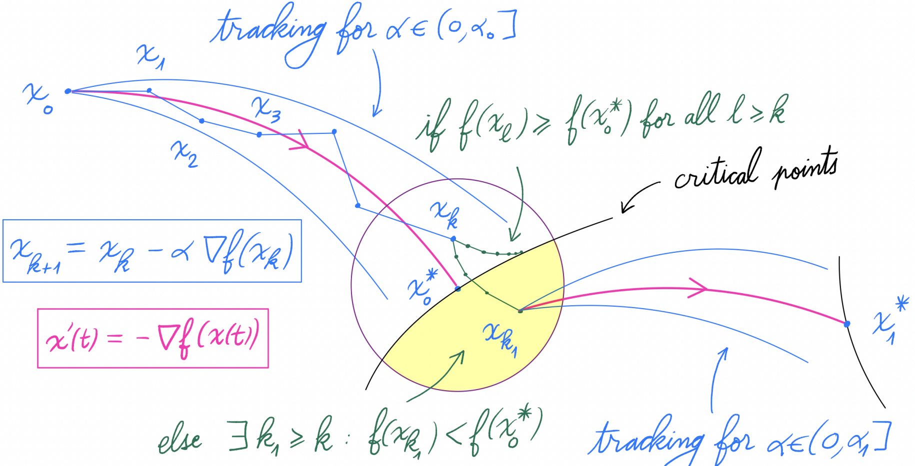
3) How to guarantee $f(x_0^*) > f(x_1^*)$?



1) What if $\alpha_1 < \alpha_0$? INITIALIZE IN A BOUNDED SET

2) How to get close to a critical point ? LENGTH

3) How to guarantee $f(x_0^*) > f(x_1^*)$?



1) What if $\alpha_1 < \alpha_0$? INITIALIZE IN A BOUNDED SET

2) How to get close to a critical point ? LENGTH

3) How to guarantee $f(x_0^*) > f(x_1^*)$? DESINGULARIZATION

Key proof idea

$$\sigma(X_0) = \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0 \end{cases}$$

$$\sigma(X_0, \bar{\alpha}) = \sup_{\substack{x \in (\mathbb{R}^n)^{\mathbb{N}} \\ \alpha \in (0, \bar{\alpha}]^{\mathbb{N}}}} \sum_{k=0}^{\infty} \|x_{k+1} - x_k\|$$

$$\text{subject to} \quad \begin{cases} x_{k+1} = x_k - \alpha_k \nabla f(x_k), \ \forall k \in \mathbb{N}, \\ x_0 \in X_0 \end{cases}$$

$$\boxed{\sigma(X_0) < \infty \quad \Longrightarrow \quad \sigma(X_0, \bar{\alpha}) < \infty}$$

Reminder

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0 \end{cases}$$

$$\sigma(X_0, \bar{\alpha}) := \sup_{\substack{x \in (\mathbb{R}^n)^{\mathbb{N}} \\ \alpha \in (0, \bar{\alpha}]^{\mathbb{N}}}} \sum_{k=0}^{\infty} \|x_{k+1} - x_k\|$$

$$\text{subject to} \quad \begin{cases} x_{k+1} = x_k - \alpha_k \nabla f(x_k), \ \forall k \in \mathbb{N}, \\ x_0 \in X_0 \end{cases}$$

Bellman's principle of optimality (1957): *"An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."*

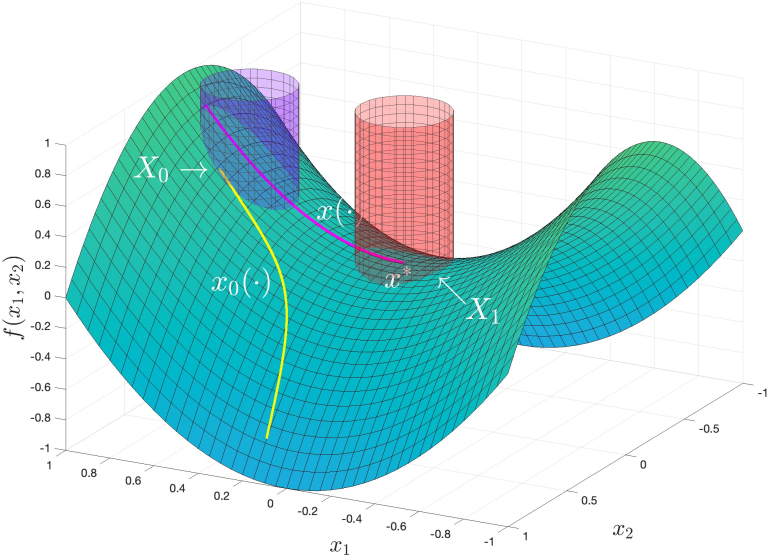
Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

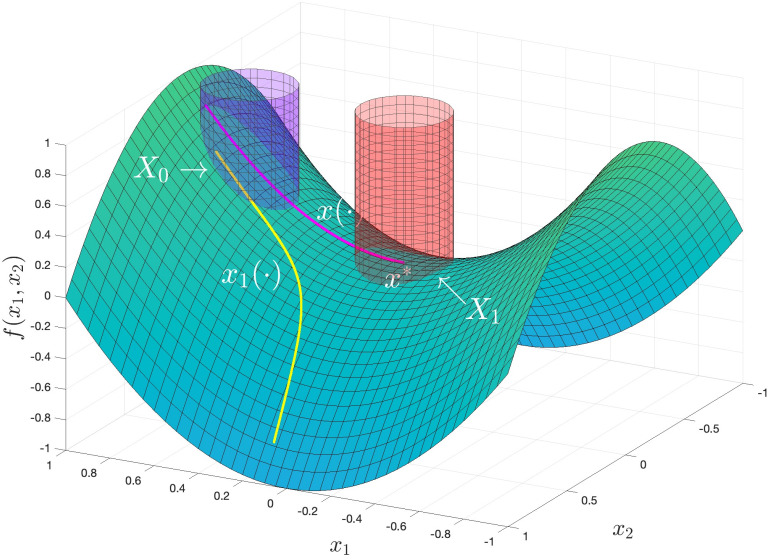
$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

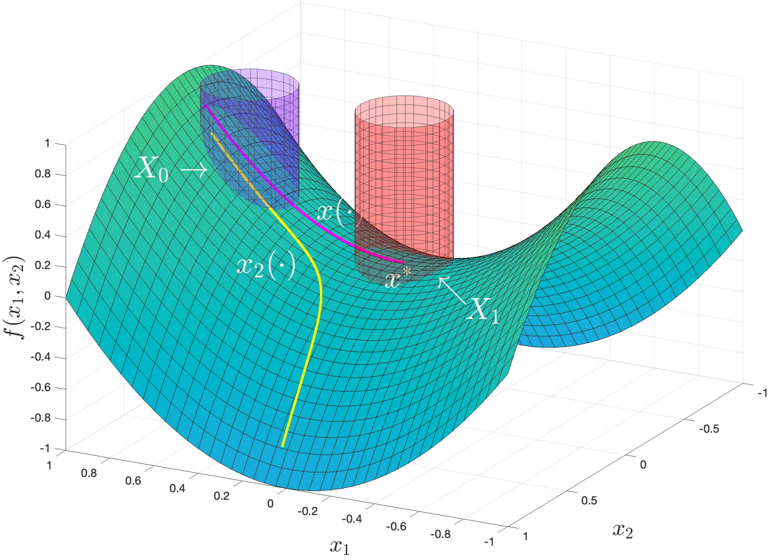
Bellman's principle of optimality



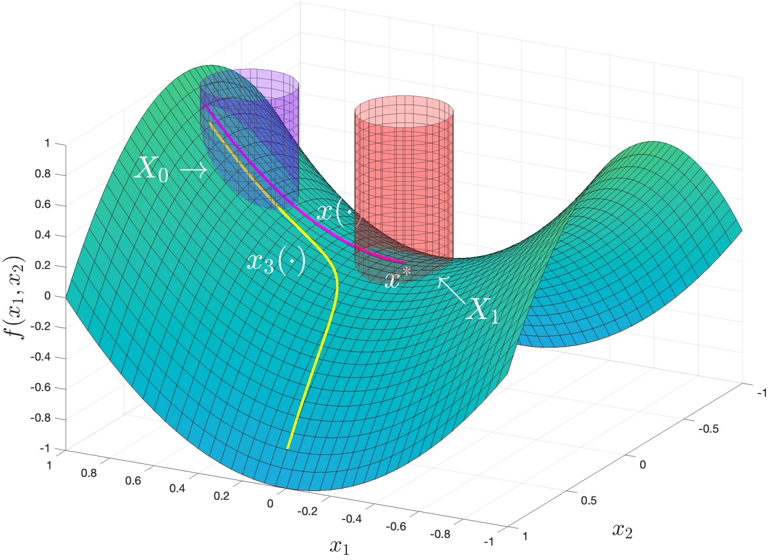
Bellman's principle of optimality



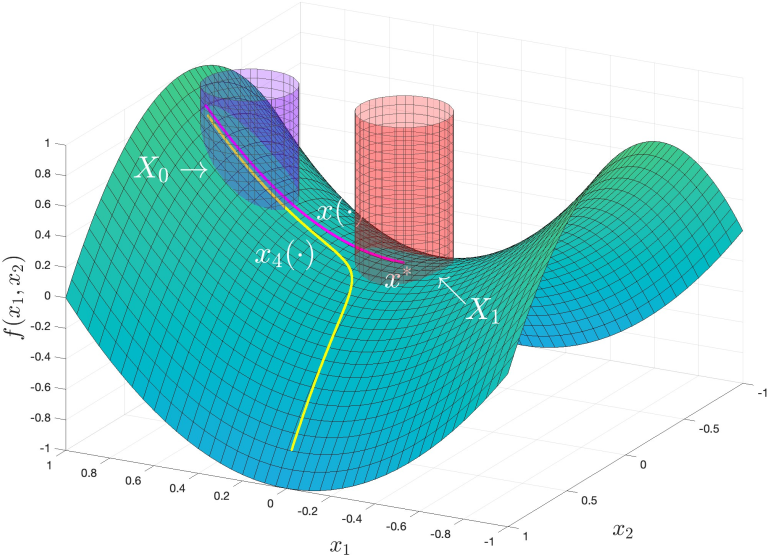
Bellman's principle of optimality



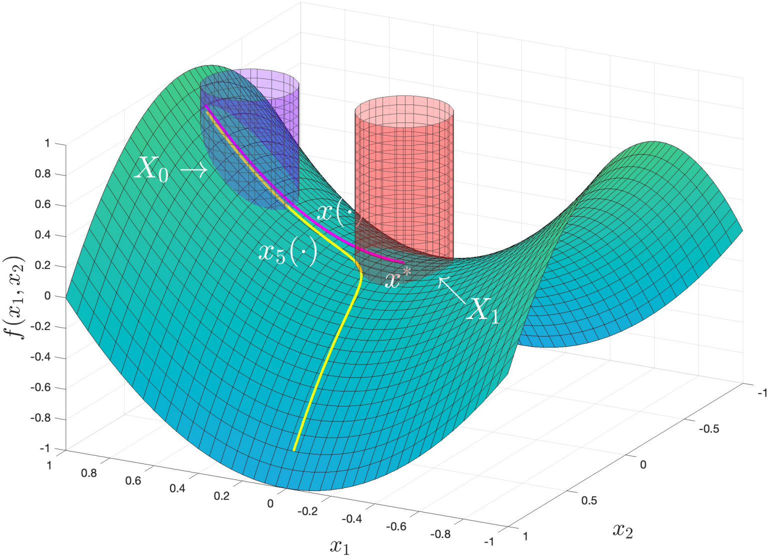
Bellman's principle of optimality



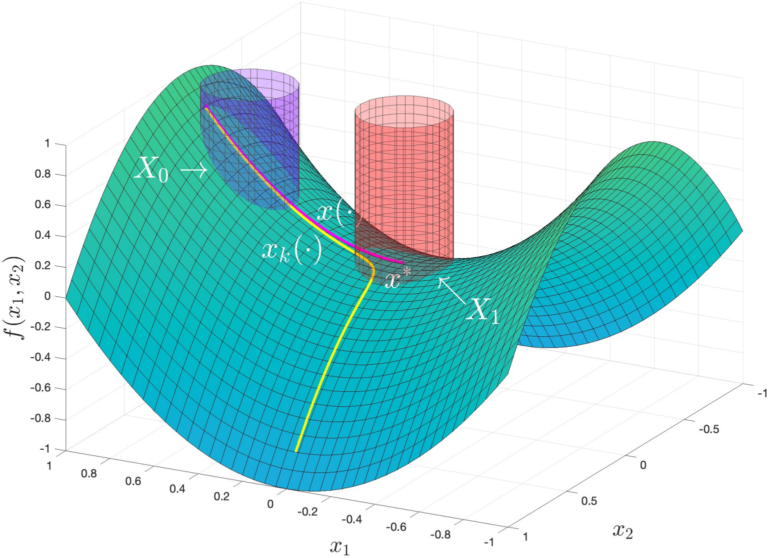
Bellman's principle of optimality



Bellman's principle of optimality



Bellman's principle of optimality



Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2)
$$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$
$$\leqslant \psi \left(\sup_{X_0} f - \min_{B(x^*, \epsilon)} f \right) + \sigma(X_1)$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \ \text{for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2) $$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$
$$\leqslant \psi\left(\sup_{X_0} f - \min_{B(x^*,\epsilon)} f\right) + \sigma(X_1)$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \ \text{for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2)
$$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$
$$\leqslant \psi\left(\sup_{X_0} f - \min_{B(x^*,\epsilon)} f\right) + \sigma(X_1)$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2) $$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$

$$\leqslant \psi\left(\sup_{X_0} f - \min_{B(x^*,\epsilon)} f\right) + \sigma(X_1)$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \ \text{for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2)

$$\begin{aligned} \int_0^\infty \|x_k'(t)\| dt &= \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt \\ &\leqslant \psi\left(\sup_{X_0} f - \min_{B(x^*,\epsilon)} f\right) + \sigma(X_1) \end{aligned}$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \text{ for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2)
$$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$

$$\sigma(X_0) \leqslant \psi \left(\sup_{X_0} f - \min_{B(x^*, \epsilon)} f \right) + \sigma(X_1)$$

where $X_1 := B(x^*, \epsilon)$.

Lemma (J. 2023)

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz definable with bounded continuous subgradient trajectories and $X_0 \subset \mathbb{R}^n$ is bounded, then $\sigma(X_0) < \infty$ where

$$\sigma(X_0) := \sup_{x \in \mathcal{A}(\mathbb{R}_+, \mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt$$

$$\text{subject to} \quad \begin{cases} x'(t) \in -\partial f(x(t)), \ \text{for a.e. } t > 0, \\ x(0) \in X_0. \end{cases}$$

Proof sketch: Let $x_k(\cdot)$ be a maximizing sequence.

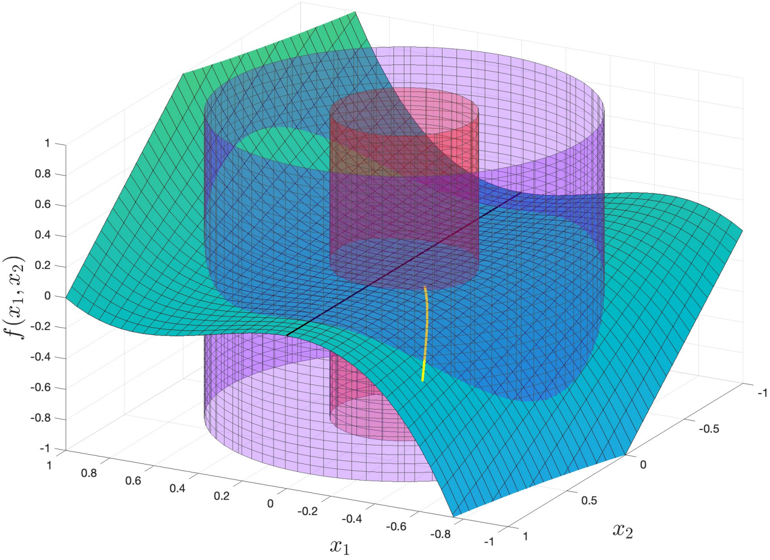
1) $x_k(\cdot) \to x(\cdot)$ uniformly over compact intervals & $x(t) \to x^*$.

2)
$$\int_0^\infty \|x_k'(t)\| dt = \int_0^{T_k} \|x_k'(t)\| dt + \int_{T_k}^\infty \|x_k'(t)\| dt$$

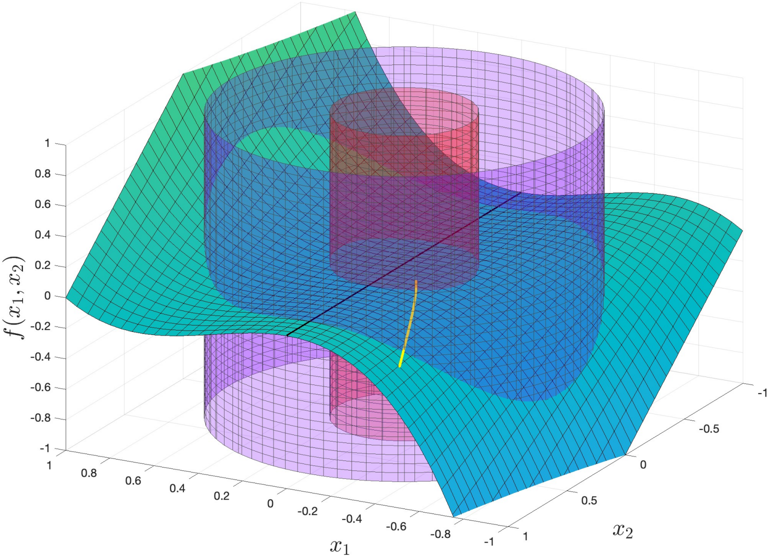
$$\sigma(X_0) \leqslant \psi \left(\sup_{X_0} f - \min_{B(x^*,\epsilon)} f \right) + \sigma(X_1)$$

3) $X_1 := B(x^*, \epsilon) \cap [f \leqslant f(x^*) - \epsilon']$.

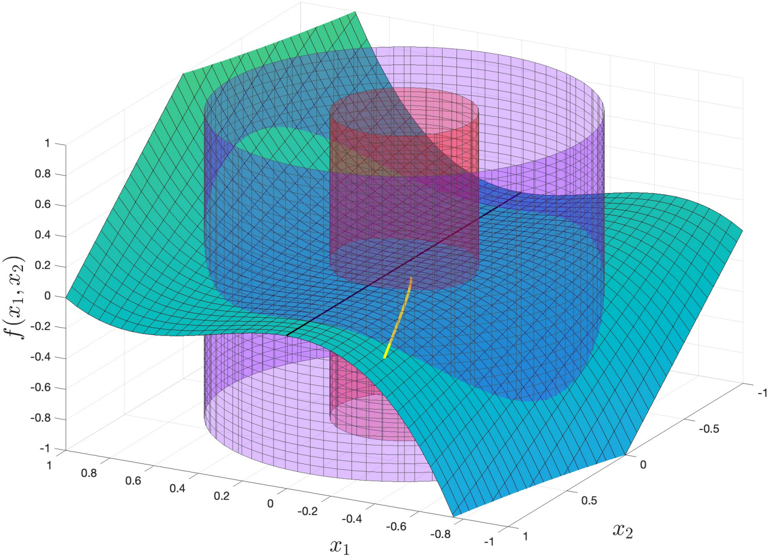
How to guarantee a sufficient decrease?



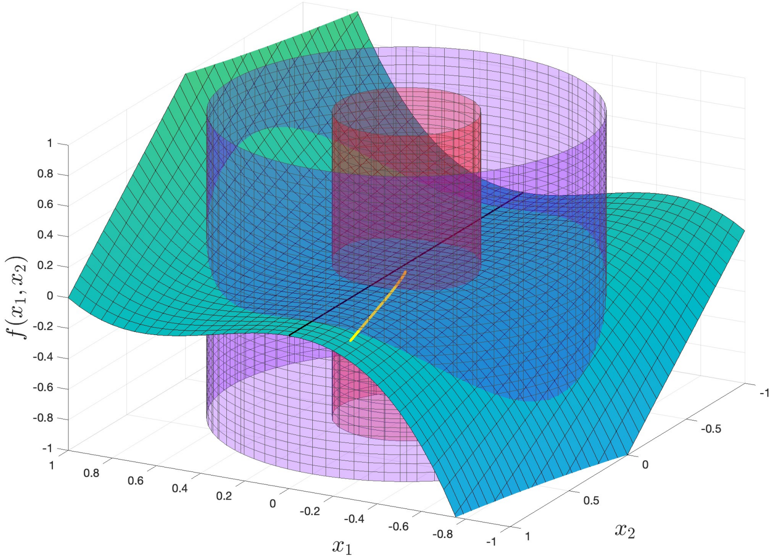
How to guarantee a sufficient decrease?



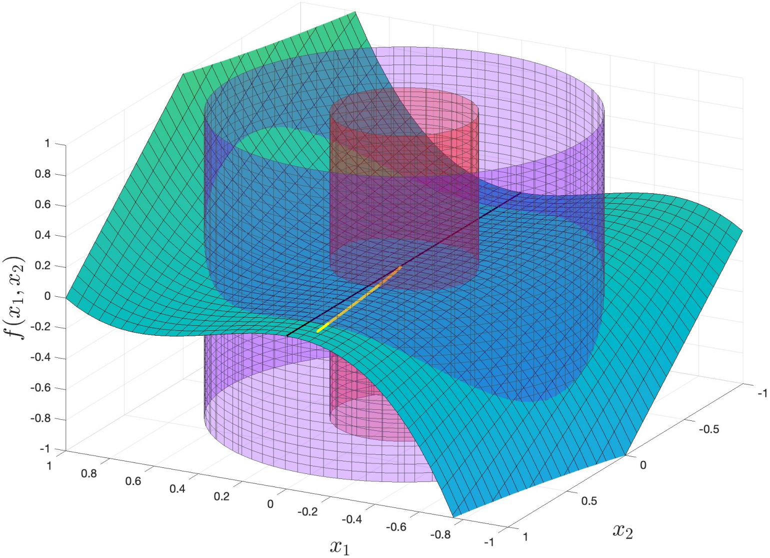
How to guarantee a sufficient decrease?



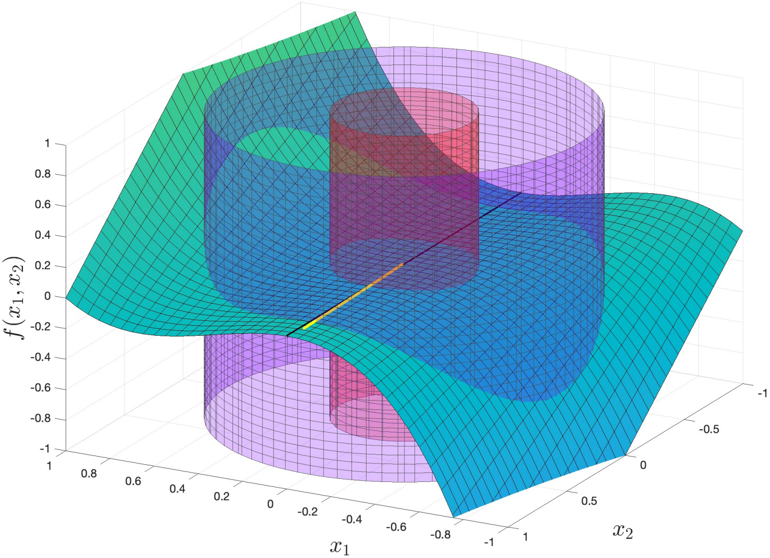
How to guarantee a sufficient decrease?



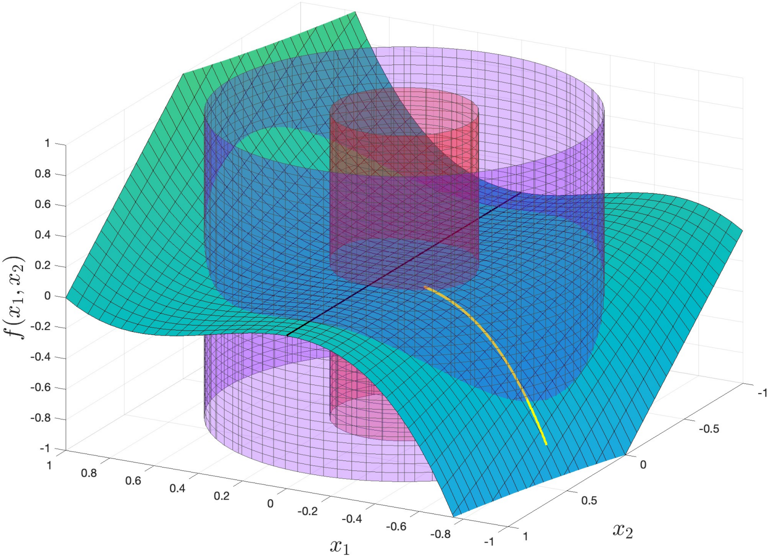
How to guarantee a sufficient decrease?



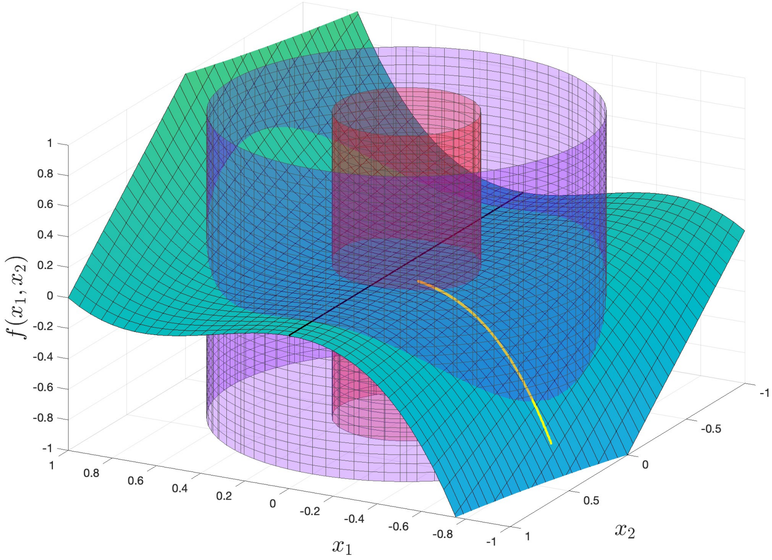
How to guarantee a sufficient decrease?



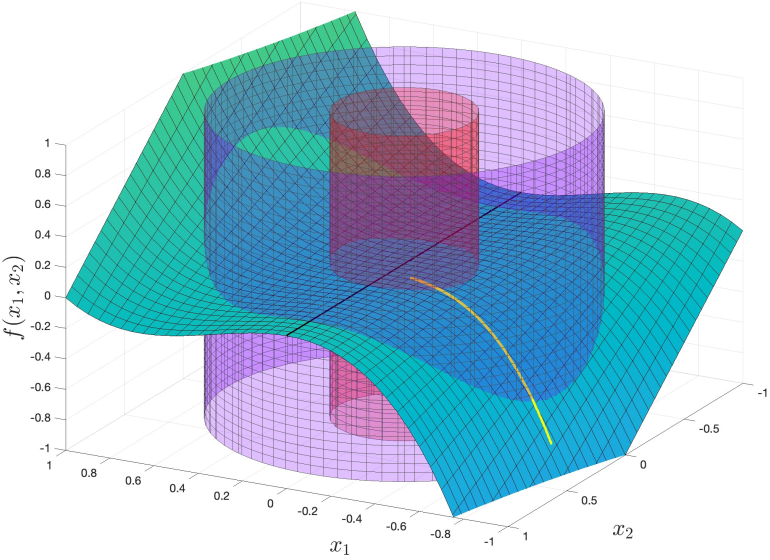
How to guarantee a sufficient decrease?



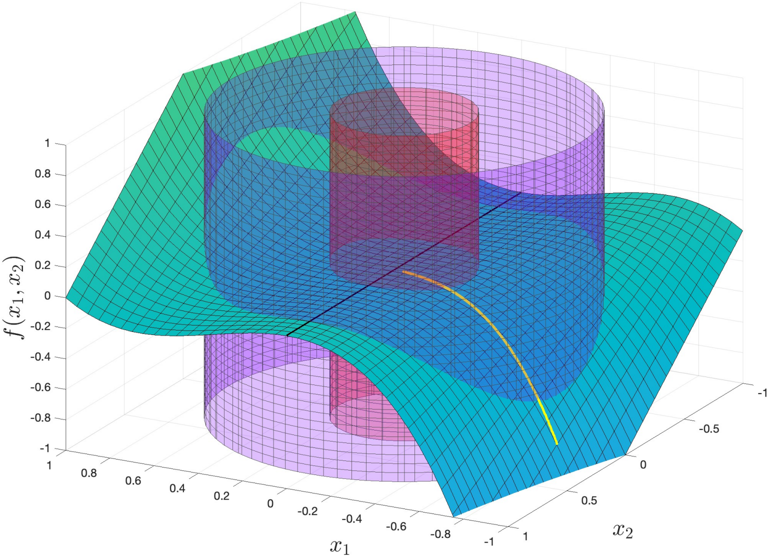
How to guarantee a sufficient decrease?



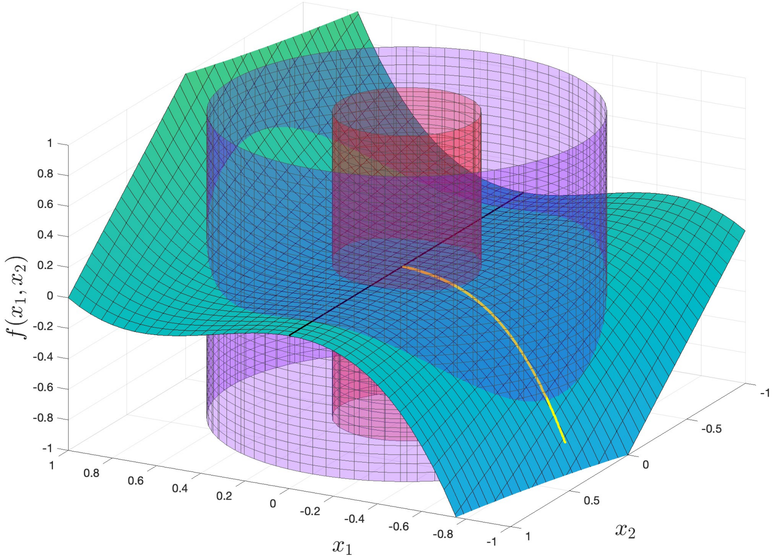
How to guarantee a sufficient decrease?



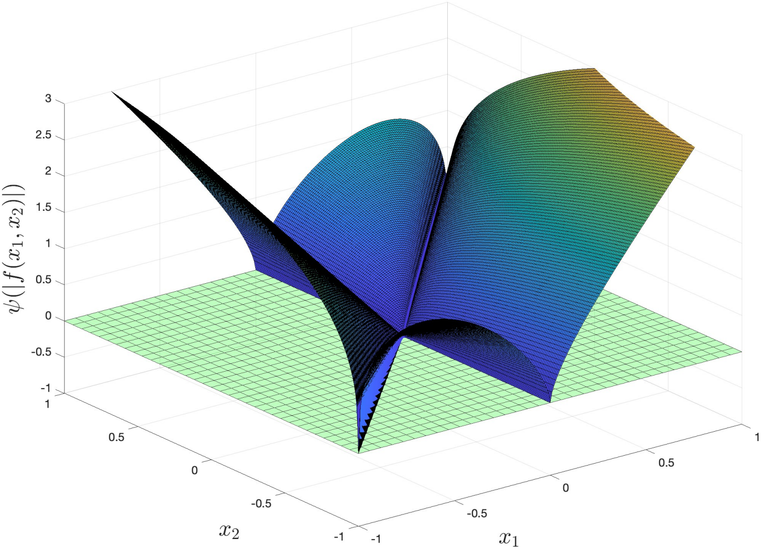
How to guarantee a sufficient decrease?



How to guarantee a sufficient decrease?



... by desingularizing $\psi \circ |f|$



Uniform Kurdyka-Łojasiewicz inequality

1. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then in any bounded set and with $|f|$ small

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

2. If f is definable and locally Lipschitz (J. 2023), in any bounded set

$$\boxed{d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1} \quad \tilde{f}(x) := d(f(x), V)$$

where V are the critical values of f in the closure of the bounded set.

3. Continuous length formula (J. 2023): in any bounded set

$$\frac{1}{2m} \int_0^T \|x'(t)\| dt \leqslant \psi\left(\frac{f(x(0)) - f(x(T))}{2m}\right)$$

4. Discrete length formula (J. 2023): in any bounded set

$$\frac{1}{(2+\epsilon)m} \sum_{k=0}^{K} \|x_{k+1} - x_k\| \leqslant \psi\left(\frac{f(x_0) - f(x_K)}{2m}\right) + \frac{2L}{2+\epsilon} \max_{k=0,\ldots,K} \alpha_k$$

Uniform Kurdyka-Łojasiewicz inequality

1. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then in any bounded set and with $|f|$ small

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

2. If f is definable and locally Lipschitz (J. 2023), in any bounded set

$$\boxed{d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1} \quad \tilde{f}(x) := d(f(x), V)$$

where V are the critical values of f in the closure of the bounded set.

3. Continuous length formula (J. 2023): in any bounded set

$$\frac{1}{2m} \int_0^T \|x'(t)\| dt \leqslant \psi\left(\frac{f(x(0)) - f(x(T))}{2m}\right)$$

4. Discrete length formula (J. 2023): in any bounded set

$$\frac{1}{(2+\epsilon)m} \sum_{k=0}^{K} \|x_{k+1} - x_k\| \leqslant \psi\left(\frac{f(x_0) - f(x_K)}{2m}\right) + \frac{2L}{2+\epsilon} \max_{k=0,\ldots,K} \alpha_k$$

Uniform Kurdyka-Łojasiewicz inequality

1. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then in any bounded set and with $|f|$ small

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

2. If f is definable and locally Lipschitz (J. 2023), in any bounded set

$$\boxed{d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1} \quad \tilde{f}(x) := d(f(x), V)$$

where V are the critical values of f in the closure of the bounded set.

3. Continuous length formula (J. 2023): in any bounded set

$$\frac{1}{2m} \int_0^T \|x'(t)\| dt \ \leqslant \ \psi \left(\frac{f(x(0)) - f(x(T))}{2m} \right)$$

4. Discrete length formula (J. 2023): in any bounded set

$$\frac{1}{(2+\epsilon)m} \sum_{k=0}^{K} \|x_{k+1} - x_k\| \ \leqslant \ \psi \left(\frac{f(x_0) - f(x_K)}{2m} \right) + \frac{2L}{2+\epsilon} \max_{k=0,\ldots,K} \alpha_k$$

Uniform Kurdyka-Łojasiewicz inequality

1. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then in any bounded set and with $|f|$ small

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

2. If f is definable and locally Lipschitz (J. 2023), in any bounded set

$$\boxed{d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1} \quad \tilde{f}(x) := d(f(x), V)$$

where V are the critical values of f in the closure of the bounded set.

3. Continuous length formula (J. 2023): in any bounded set

$$\frac{1}{2m} \int_0^T \|x'(t)\| dt \;\leqslant\; \psi\left(\frac{f(x(0)) - f(x(T))}{2m}\right)$$

4. Discrete length formula (J. 2023): in any bounded set

$$\frac{1}{(2+\epsilon)m} \sum_{k=0}^{K} \|x_{k+1} - x_k\| \;\leqslant\; \psi\left(\frac{f(x_0) - f(x_K)}{2m}\right) + \frac{2L}{2+\epsilon} \max_{k=0,\ldots,K} \alpha_k$$

Uniform Kurdyka-Łojasiewicz inequality

1. If f is definable and locally Lipschitz (Bolte *et al.* 2007), then in any bounded set and with $|f|$ small

$$d(0, \partial(\psi \circ |f|)(x)) \geqslant 1$$

2. If f is definable and locally Lipschitz (J. 2023), in any bounded set

$$\boxed{d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1} \quad \tilde{f}(x) := d(f(x), V)$$

where V are the critical values of f in the closure of the bounded set.

3. Continuous length formula (J. 2023): in any bounded set

$$\psi^{-1}\left(\frac{1}{2m}\int_0^T \|x'(t)\|dt\right) \leqslant \frac{f(x(0)) - f(x(T))}{2m}$$

4. Discrete length formula (J. 2023): in any bounded set

$$\psi^{-1}\left(\frac{1}{(2+\epsilon)m}\sum_{k=0}^{K}\|x_{k+1} - x_k\| - \frac{2L}{2+\epsilon}\max_{k=0,\dots,K}\alpha_k\right) \leqslant \frac{f(x_0) - f(x_K)}{2m}$$

On gradients of functions definable in o-minimal structures (Kurdyka, 1998)

Theorem 2. — *Let $f : U \longrightarrow \mathbb{R}$ be a function of class C^1, where U is an open and bounded subset of $\mathbb{R}^n$. Suppose that f is an $\mathcal{M}$-function, for some o-minimal structure $\mathcal{M}$.*

a) *Then there exists $A > 0$ such that all trajectories of $-\mathbf{grad}\, f$ have length bounded by A.*

b) *More precisely, there exists $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ a continuous strictly increasing $\mathcal{M}$-function, with $\lim_{t\to 0} \sigma(t) = 0$, such that if γ is a trajectory of $-\mathbf{grad}\, f$ and $a, b \in \gamma$, then*

$$|\gamma(a, b)| \leq \sigma(|f(b) - f(a)|).$$

On trajectories of the gradient of an analytic function 1983

SUR LES TRAJECTOIRES DU GRADIENT D'UNE FONCTION ANALYTIQUE

par S. Łojasiewicz

On va démontrer le suivant

Théorème (1). Soit f une fonction analytique et $\geqslant 0$ au voisinage de zéro de $\mathbb{R}^n$, telle que $f(0) = 0$. Alors il existe un voisinage U de zéro tel que chaque trajectoire y_x, vérifiant $y_x(0) = x$, du système

$$y' = -\operatorname{grad} f(y)$$

est définie dans $[0,\infty)$, est de longueur finie, et tend vers un point de $Z = \{f(x) = 0\}$ lorsque $t \to \infty$, la convergence étant uniforme.

Alternative proof using Kurdyka's length formula

Let $f : \mathbb{R}^n \to \mathbb{R}$ nonnegative analytic with $f(0) = 0$ and $\epsilon > 0$.

Since f is continuous, $\exists \delta \in (0, \epsilon) : \sup_{B(0,\delta)} f \leqslant \sigma^{-1}(\epsilon - \delta)$.

Consider a gradient trajectory

$$\begin{cases} x'(t) = -\nabla f(x(t)), \ \forall t > 0, \\ x(0) \in \mathring{B}(0, \delta). \end{cases}$$

If $T = \inf\{t \geqslant 0 : x(t) \notin B(0, \epsilon)\} < \infty$, then

$$\epsilon - \delta < \|x(T) - x(0)\| \leqslant \int_0^T \|x'(t)\| dt \leqslant \sigma\left(f(x(0)) - f(x(T))\right)$$

$$\leqslant \sigma(f(x(0))) \leqslant \epsilon - \delta.$$

Kurdyka-Łojasiewicz inequality (Kurdyka 1998, Bolte *et al.* 2007)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let X be a bounded subset of $\mathbb{R}^n$ and $v \in \mathbb{R}$ be a critical value of f in $\overline{X}$. There exists $\rho > 0$ and a strictly increasing continuous definable function $\psi : [0, \rho) \to [0, \infty)$ which belongs to $C^1((0, \rho))$ with $\psi(0) = 0$ such that

$$\forall x \in X, \quad |f(x) - v| \in (0, \rho) \quad \Longrightarrow \quad d(0, \partial(\psi \circ |f - v|)(x)) \geqslant 1.$$

Uniform Kurdyka-Łojasiewicz inequality (J. 2023)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let X be a bounded subset of $\mathbb{R}^n$ and V be the set of critical values of f in $\overline{X}$ if it is non-empty, otherwise $V := \{0\}$. There exists a concave definable diffeomorphism $\psi : [0, \infty) \to [0, \infty)$ such that

$$\forall x \in X \setminus (\partial \tilde{f})^{-1}(0), \quad d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1,$$

where $\tilde{f}(x) := d(f(x), V)$ for all $x \in \mathbb{R}^n$.

Kurdyka-Łojasiewicz inequality (Kurdyka 1998, Bolte *et al.* 2007)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let X be a bounded subset of $\mathbb{R}^n$ and $v \in \mathbb{R}$ be a critical value of f in $\overline{X}$. There exists $\rho > 0$ and a strictly increasing continuous definable function $\psi : [0, \rho) \to [0, \infty)$ which belongs to $C^1((0, \rho))$ with $\psi(0) = 0$ such that

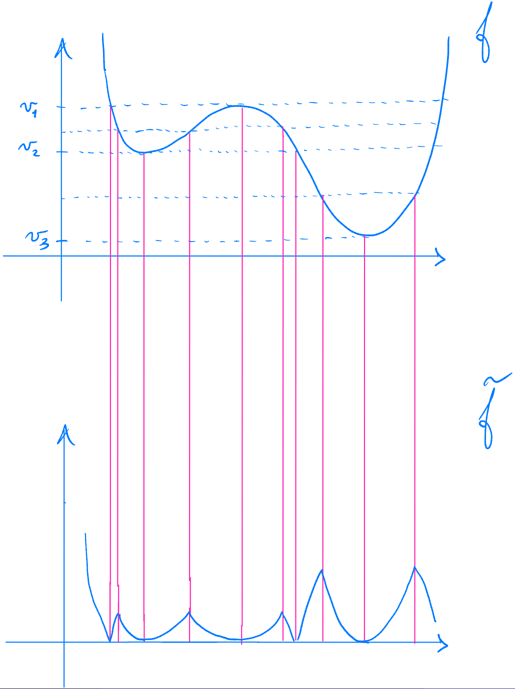
$$\forall x \in X, \quad |f(x) - v| \in (0, \rho) \quad \Longrightarrow \quad d(0, \partial(\psi \circ |f - v|)(x)) \geqslant 1.$$

Uniform Kurdyka-Łojasiewicz inequality (J. 2023)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and definable. Let X be a bounded subset of $\mathbb{R}^n$ and V be the set of critical values of f in $\overline{X}$ if it is non-empty, otherwise $V := \{0\}$. There exists a concave definable diffeomorphism $\psi : [0, \infty) \to [0, \infty)$ such that

$$\forall x \in X \setminus (\partial \tilde{f})^{-1}(0), \quad d(0, \partial(\psi \circ \tilde{f})(x)) \geqslant 1,$$

where $\tilde{f}(x) := d(f(x), V)$ for all $x \in \mathbb{R}^n$.


f
v_1
v_2
v_3
$\tilde{f}$

Continuous-time length formula

Proposition (J. 2023)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz definable function and let X be a bounded subset of $\mathbb{R}^n$. Assume that f has at most $m \in \mathbb{N}^*$ critical values in $\overline{X}$. Let ψ be a desingularizing function of f on X. If $x : [0, T] \to X$ with $T \geqslant 0$ is an absolutely continuous function such that $x'(t) \in -\partial f(x(t))$ for almost every $t \in (0, T)$, then

$$\frac{1}{2m} \int_0^T \|x'(t)\| dt \ \leqslant \ \psi\left(\frac{f(x(0)) - f(x(T))}{2m}\right).$$

- Refinement of Kurdyka's length formula
- New proof integrates the Kurdyka-Łojasiewicz inequality
- It also uses Jensen's inequality
- Discrete version appears to be new

Special case where $0 \notin \partial\tilde{f}(x(t))$ for all $t \in (0, T)$

$$\begin{aligned}
\overline{T} &\leqslant \int_0^{\overline{T}} d(0, \partial(\psi \circ \tilde{f})(\bar{x}(\bar{t})))\, d\bar{t} \\
&= \int_0^{\overline{T}} \|\bar{x}'(\bar{t})\| d\bar{t} \\
&= \int_0^{T} \|x'(t)\| dt \\
&\leqslant \sqrt{\overline{T}} \sqrt{\int_0^{\overline{T}} \|\bar{x}'(\bar{t})\|^2 d\bar{t}} \\
&= \sqrt{\overline{T}} \sqrt{|(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})|}.
\end{aligned}$$

Hence $\overline{T} \leqslant |(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})|$ and thus

$$\begin{aligned}
\int_0^{T} \|x'(t)\| dt &\leqslant |(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})| \\
&\leqslant \psi(|(\tilde{f} \circ \bar{x})(0) - (\tilde{f} \circ \bar{x})(\overline{T})|) \\
&= \psi(f(x(0)) - f(x(T)))
\end{aligned}$$

Special case where $0 \notin \partial\tilde{f}(x(t))$ for all $t \in (0, T)$

$$
\begin{aligned}
\overline{T} &\leqslant \int_0^{\overline{T}} d(0, \partial(\psi \circ \tilde{f})(\bar{x}(\bar{t})))\, d\bar{t} \\
&= \int_0^{\overline{T}} \|\bar{x}'(\bar{t})\| d\bar{t} \\
&= \int_0^{T} \|x'(t)\| dt \\
&\leqslant \sqrt{\overline{T}} \sqrt{\int_0^{\overline{T}} \|\bar{x}'(\bar{t})\|^2 d\bar{t}} \\
&= \sqrt{\overline{T}} \sqrt{|(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})|}.
\end{aligned}
$$

Hence $\overline{T} \leqslant |(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})|$ and thus

$$
\begin{aligned}
\int_0^T \|x'(t)\| dt &\leqslant |(\psi \circ \tilde{f} \circ \bar{x})(0) - (\psi \circ \tilde{f} \circ \bar{x})(\overline{T})| \\
&\leqslant \psi(|(\tilde{f} \circ \bar{x})(0) - (\tilde{f} \circ \bar{x})(\overline{T})|) \\
&= \psi(f(x(0)) - f(x(T)))
\end{aligned}
$$

General case where $0 \in \partial \tilde{f}(x(t_i))$ for $i = 1, \dots, k$

$$\int_0^T \|x'(t)\| dt = \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} \|x'(t)\| dt$$

$$\leqslant \sum_{i=0}^{k} \psi(f(x(t_i)) - f(x(t_{i+1})))$$

$$\leqslant (k+1)\,\psi\left(\frac{1}{k+1}\sum_{i=0}^{k} f(x(t_i)) - f(x(t_{i+1}))\right)$$

$$= (k+1)\,\psi\left(\frac{f(x(0)) - f(x(T))}{k+1}\right)$$

$$\leqslant 2m\,\psi\left(\frac{f(x(0)) - f(x(T))}{2m}\right)$$

From continuous to discrete time

$$\sup_{x \in \mathcal{A}(\mathbb{R}_+,\mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt \quad \text{s.t.} \quad \begin{cases} x'(t) = -\nabla f(x(t)), \ \forall t > 0, \\ x(0) \in X_0 \end{cases}$$

$$\sup_{\substack{x \in (\mathbb{R}^n)^{\mathbb{N}} \\ \alpha \in (0,\bar{\alpha}]^{\mathbb{N}}}} \sum_{k=0}^{\infty} \|x_{k+1} - x_k\| \quad \text{s.t.} \quad \begin{cases} x_{k+1} = x_k - \alpha_k \nabla f(x_k), \ \forall k \in \mathbb{N}, \\ x_0 \in X_0 \end{cases}$$

Tracking lemma (J. 2023)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be $C^{1,1}_{\text{loc}}$ and lower bounded.

$$\forall \epsilon, \ \forall T > 0, \ \exists \bar{\alpha} > 0 : \ \forall (x,\alpha) \in \Sigma(X_0, \bar{\alpha}), \ \exists x(\cdot) \in \Sigma(X_0) :$$

$$\forall k \in \mathbb{N}^*, \quad \alpha_0 + \cdots + \alpha_{k-1} \leqslant T \quad \Longrightarrow \quad \|x_k - x(\alpha_0 + \cdots + \alpha_{k-1})\| \leqslant \epsilon.$$

Getting close to critical points

$$\sup_{x\in\mathcal{A}(\mathbb{R}_+,\mathbb{R}^n)} \int_0^\infty \|x'(t)\|dt \quad \text{s.t.} \quad \begin{cases} x'(t) = -\nabla f(x(t)), \ \forall t > 0, \\ x(0) \in X_0 \end{cases}$$

- $\Phi : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is the gradient flow
- C is the set of critical points of f in $\overline{\Phi(\mathbb{R}_+, X_0)}$
- $\nu := \inf_{\overline{\Phi(\mathbb{R}_+,X_0)}\setminus\mathring{B}(C,\delta)} \|\nabla f\| > 0$
- $T := 2\sigma(X_0)/\nu < \infty$

If $\|x'(t)\| \geqslant 2\sigma(X_0)/T$ for all $t \in (0,T)$, then

$$\sigma(X_0) < T\frac{2\sigma(X_0)}{T} \leqslant \int_0^T \|x'(t)\|dt \leqslant \int_0^\infty \|x'(t)\|dt \leqslant \sigma(X_0).$$

Thus $\exists t^* \in (0,T) :$ $\|x'(t^*)\| < 2\sigma(X_0)/T = \nu$ and $x(t^*) \in \mathring{B}(C,\delta)$.

$$\sup_{x \in \mathcal{A}(\mathbb{R}_+,\mathbb{R}^n)} \int_0^\infty \|x'(t)\| dt \quad \text{s.t.} \quad \left\{ \begin{array}{l} x'(t) = -\nabla f(x(t)), \ \forall t > 0, \\ x(0) \in X_0 \end{array} \right.$$

- $\Phi : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is the gradient flow
- C is the set of critical points of f in $\overline{\Phi(\mathbb{R}_+, X_0)}$
- $\nu := \inf_{\overline{\Phi(\mathbb{R}_+,X_0)} \setminus \mathring{B}(C,\delta)} \|\nabla f\| > 0$
- $T := 2\sigma(X_0)/\nu < \infty$

If $\|x'(t)\| \geqslant 2\sigma(X_0)/T$ for all $t \in (0, T)$, then

$$\sigma(X_0) < T\frac{2\sigma(X_0)}{T} \leqslant \int_0^T \|x'(t)\| dt \leqslant \int_0^\infty \|x'(t)\| dt \leqslant \sigma(X_0).$$

Thus $\exists t^* \in (0, T) :$ $\|x'(t^*)\| < 2\sigma(X_0)/T = \nu$ and $x(t^*) \in \mathring{B}(C, \delta)$.

Getting close to critical points

$$\sup_{x\in\mathcal{A}(\mathbb{R}_+,\mathbb{R}^n)} \int_0^\infty \|x'(t)\|dt \quad \text{s.t.} \quad \begin{cases} x'(t) = -\nabla f(x(t)), \ \forall t > 0, \\ x(0) \in X_0 \end{cases}$$

- $\Phi : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is the gradient flow
- C is the set of critical points of f in $\overline{\Phi(\mathbb{R}_+, X_0)}$
- $\nu := \inf_{\overline{\Phi(\mathbb{R}_+,X_0)}\setminus \mathring{B}(C,\delta)} \|\nabla f\| > 0$
- $T := 2\sigma(X_0)/\nu < \infty$

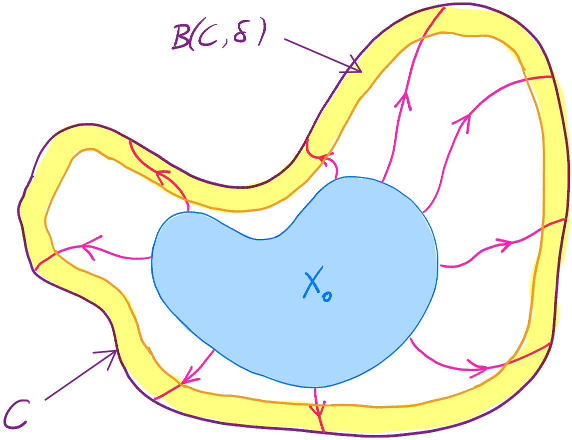
If $\|x'(t)\| \geqslant 2\sigma(X_0)/T$ for all $t \in (0, T)$, then

$$\sigma(X_0) < T\frac{2\sigma(X_0)}{T} \leqslant \int_0^T \|x'(t)\|dt \leqslant \int_0^\infty \|x'(t)\|dt \leqslant \sigma(X_0).$$

Thus $\exists t^* \in (0, T):$ $\|x'(t^*)\| < 2\sigma(X_0)/T = \nu$ and $x(t^*) \in \mathring{B}(C, \delta)$.

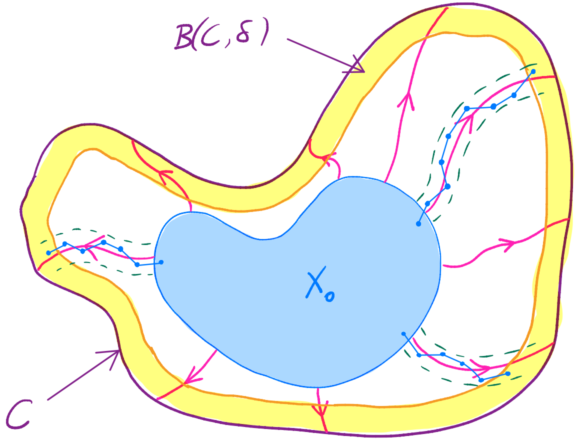
Getting close to critical points

$$\exists\, t^* \in (0,T): \quad x(t^*) \in B(C,\delta)$$



Getting close to critical points

$$\exists\, t^* \in (0,T): \quad x(t^*) \in B(C,\delta)$$



Conclusion

1. Continuous/discrete time gradient dynamics intricately linked

2. Length is no longer a byproduct, but a crucial tool

3. KL $\Longrightarrow$ length formula $\Longrightarrow$ convergence analysis

Thank you for your attention!

Feel free to contact me at

cj2638@columbia.edu

for questions or suggestions.