On the continuous and discrete Gradient Conjecture

Semester project in CSE

Florian Grün
École Polytechnique Fédérale de Lausanne, Switzerland

OPTIM Professor Nicolas Boumal
Supervisor PhD Quentin Rebjock

11.01.2023

Outline

Introduction and known results

- continuous gradient descent flow:

$$x'(t) = -\nabla f(x(t)), \quad x(0) = x_0$$

- discrete gradient descent:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad x_0 \text{ given}$$

Introduction and known results

- Convergence? Yes, thanks to Lojasiewicz [Loj84] (1984, *Trajectoires du gradient d'une fonction analytique*)
  Thm: **if** f real-analytic, convergence to critical point x^*
  or $|x(t)| \to \infty$.
  $\hookrightarrow$ Lojasiewicz inequality: $|\nabla f| \geq c|f|^\rho, \quad \rho < 1$

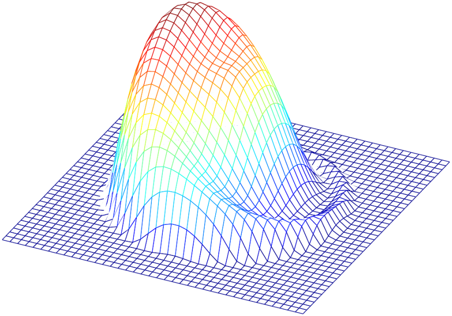

Figure: Mexican hat from [AMA05]

Introduction and known results

- Convergence also for discrete case
  (2005, P.-A. Absil, R. Mahony and B. Andrews: *Convergence of the iterates of descent methods for analytic cost functions*) [AMA05]

- infinite dimensional Lojasiewicz inequalities for PDEs

Gradient conjectures in comparison

- (1) Gradient Conjecture (GC): $x'(t) = -\nabla f(x(t))$, $x(t) \to x^*$

$$\text{Then} \qquad \lim_{t\to\infty} \frac{x(t) - x^*}{|x(t) - x^*|} \qquad \text{exists.}$$

$\hookrightarrow$ trajectory "does not wiggle"

- proven by K. Kurdyka, T. Mostowski, and A. Parusinski in *Proof of the gradient conjecture of R. Thom* [KMP99]

- (2) strong GC: $\lim_{t\to\infty} \frac{x'(t)}{|x'(t)|}$ exists ? (implies GC)

Gradient conjectures in comparison

- (1) Gradient Conjecture (GC): $x'(t) = -\nabla f(x(t))$, $x(t) \to x^*$

$$\text{Then} \qquad \lim_{t\to\infty} \frac{x(t) - x^*}{|x(t) - x^*|} \qquad \text{exists.}$$

$\hookrightarrow$ trajectory "does not wiggle"

- proven by K. Kurdyka, T. Mostowski, and A. Parusinski in *Proof of the gradient conjecture of R. Thom* [KMP99]

- (2) strong GC: $\lim_{t\to\infty} \frac{x'(t)}{|x'(t)|}$ exists ? (implies GC)

Gradient conjectures in comparison

- (3) length-distance convergence:

$$\lim_{t\to\infty} \frac{\sigma(t)}{|x(t)|} = \lim_{t\to\infty} \frac{\int_t^\infty |x'(t)|}{|x(t)|} \to 1$$

- holds for f analytic (Corollary in [KMP99] proof)
- strong GC $\implies$ length-distance convergence
- GC $\implies$ length-distance convergence ?

Gradient conjectures in comparison

- Counterexample: $f \in C^\infty$ and $x(t)$ solution to $x' = -\nabla f(x)$
  length-distance convergence
  but converging in spiral

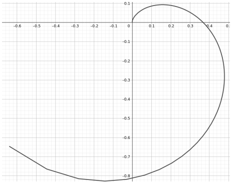

Figure: The curve γ

- $\gamma(t) = (r(t), \theta(t)) = \left(\frac{1}{t}, \log(\log(t))\right)$

- $f(r, \theta) = e^{\frac{-1}{r}} \left(1 - \frac{\log(1/r)}{r + r^2 \log(1/r)^2} \sin(\theta - \log(\log(1/r)))\right)$

- $\implies -\nabla f\big|_{\gamma(t)} \parallel \gamma'(t)$, thus gradient flow follows γ.

Gradient conjectures in comparison

- Counterexample: $f \in C^\infty$ and $x(t)$ solution to $x' = -\nabla f(x)$
  length-distance convergence
  but converging in spiral

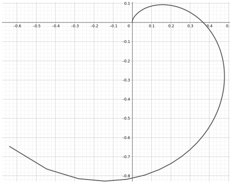

Figure: The curve γ

- $\gamma(t) = (r(t), \theta(t)) = \left(\frac{1}{t}, \log(\log(t))\right)$
- $f(r,\theta) = e^{\frac{-1}{r}}\left(1 - \frac{\log(1/r)}{r + r^2\log(1/r)^2}\sin(\theta - \log(\log(1/r))\right)$
- $\implies -\nabla f\big|_{\gamma(t)} \parallel \gamma'(t)$, thus gradient flow follows γ.

Gradient conjectures in comparison

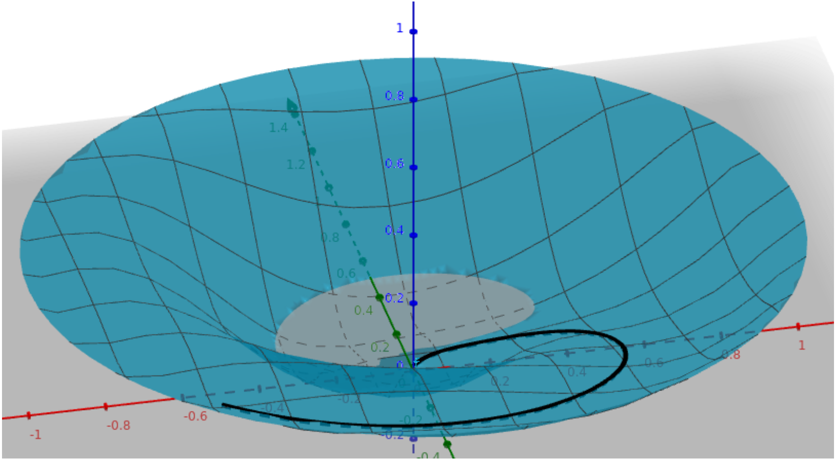

Figure: f and a gradient descent trajectory

The proof of [KMP99]

- split up into ∇f into $\partial_r f$ and $\nabla' f$ and stratify
- $F(x(s)) = \frac{f}{r^\ell} = \frac{f(x(s))}{|x(s)|^\ell}$ has limit a as $s \to s_0$ $(x(s) \to 0)$
  $\hookrightarrow$ ℓ characteristic to f
- strong Lojasiewicz inequality: $r|\nabla F| \geq |F - a|^\rho$ on some strata

The proof of [KMP99]

- split up into ∇f into $\partial_r f$ and $\nabla' f$ and stratify
- $F(x(s)) = \frac{f}{r^\ell} = \frac{f(x(s))}{|x(s)|^\ell}$ has limit a as $s \to s_0$ $(x(s) \to 0)$
  $\hookrightarrow \ell$ characteristic to f
- strong Lojasiewicz inequality: $r|\nabla F| \geq |F - a|^\rho$ on some strata

The proof of [KMP99]

- split up into ∇f into $\partial_r f$ and $\nabla' f$ and stratify
- $F(x(s)) = \frac{f}{r^\ell} = \frac{f(x(s))}{|x(s)|^\ell}$ has limit a as $s \to s_0$ $(x(s) \to 0)$
  $\hookrightarrow \ell$ characteristic to f
- strong Lojasiewicz inequality: $r|\nabla F| \geq |F - a|^\rho$ on some strata

The proof of [KMP99]

- $g(x(s)) := F - a - r^\alpha$ is *control function*

  $\hookrightarrow$ bounded and $\frac{dg}{d\tilde{s}} \geq |g|^\xi, \xi < 1$ ($\tilde{s}(s)$ is arclength of $\frac{x(s)}{|x(s)|}$)

- Lojasiewicz argument: $\frac{dg(x(\tilde{s}))}{d\tilde{s}} \geq c|g(x(\tilde{s}))|^\xi \implies \tilde{x}(s)$ finite

  $\hookrightarrow$ in original [Loj84] theorem: $\frac{df}{ds} \geq |\nabla f| \geq c|f|^\rho \implies x(s)$ finite

- finite length of projection $\tilde{x}(s)$ on $\mathbb{S}^{n-1} \implies$ GC

The proof of [KMP99]

- $g(x(s)) := F - a - r^\alpha$ is *control function*

  $\hookrightarrow$ bounded and $\frac{dg}{d\tilde{s}} \geq |g|^\xi, \xi < 1$ $\quad$ ($\tilde{s}(s)$ is arclength of $\frac{x(s)}{|x(s)|}$)

- Lojasiewicz argument: $\frac{dg(x(\tilde{s}))}{d\tilde{s}} \geq c|g(x(\tilde{s}))|^\xi \implies \tilde{x}(s)$ finite

  $\hookrightarrow$ in original [Loj84] theorem: $\frac{df}{ds} \geq |\nabla f| \geq c|f|^\rho \implies x(s)$ finite

- finite length of projection $\tilde{x}(s)$ on $\mathbb{S}^{n-1} \implies$ GC

Problems with a direct discretization

- [AMA05]: discrete Lojasiewicz argument with telescoping sum
  $\hookrightarrow$ crucial that $\boldsymbol{\rho < 1}$!

- works for various optimization schemes

Problems with a direct discretization

- [AMA05]: discrete Lojasiewicz argument with telescoping sum
  $\hookrightarrow$ crucial that $\boldsymbol{\rho < 1}$!
- works for various optimization schemes

Problems with a direct discretization

- now $\frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$, extra factor of $\frac{1}{|x_k|}$, need strong Lojasiewicz

* for f , Bochnak-Lojasiewicz: $r|\nabla f| \geq c|f|$ not enough

* for F, strong Lojasiewicz, but only on strata $(|\partial_r F| \ll |\nabla' F|)$

* $r\frac{dg}{ds} \geq |g|^{\mu}$, but maybe $\mu \geq 1$

* $\frac{dg}{d\tilde{s}} \geq c|g|^{\xi}$ only along the trajectory

Problems with a direct discretization

- now $\frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$, extra factor of $\frac{1}{|x_k|}$, need strong Lojasiewicz

* for f , Bochnak-Lojasiewicz: $r|\nabla f| \geq c|f|$ not enough

* for F, strong Lojasiewicz, but only on strata $(|\partial_r F| \ll |\nabla' F|)$

* $r\frac{dg}{ds} \geq |g|^\mu$, but maybe $\mu \geq 1$

* $\frac{dg}{d\tilde{s}} \geq c|g|^\xi$ only along the trajectory

Problems with a direct discretization

- now $\frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$, extra factor of $\frac{1}{|x_k|}$, need strong Lojasiewicz

* for f , Bochnak-Lojasiewicz: $r|\nabla f| \geq c|f|$ not enough

* for F, strong Lojasiewicz, but only on strata $(|\partial_r F| \ll |\nabla' F|)$

* $r\frac{dg}{ds} \geq |g|^{\mu}$, but maybe $\mu \geq 1$

* $\frac{dg}{d\tilde{s}} \geq c|g|^{\xi}$ only along the trajectory

Problems with a direct discretization

- now $\frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$, extra factor of $\frac{1}{|x_k|}$, need strong Lojasiewicz

∗ for f , Bochnak-Lojasiewicz: $r|\nabla f| \geq c|f|$ not enough

∗ for F, strong Lojasiewicz, but only on strata $(|\partial_r F| \ll |\nabla' F|)$

∗ $r\frac{dg}{ds} \geq |g|^{\mu}$, but maybe $\mu \geq 1$

∗ $\frac{dg}{d\tilde{s}} \geq c|g|^{\xi}$ only along the trajectory

Problems with a direct discretization

- now $\frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$, extra factor of $\frac{1}{|x_k|}$, need strong Lojasiewicz

* for f , Bochnak-Lojasiewicz: $r|\nabla f| \geq c|f|$ not enough

* for F, strong Lojasiewicz, but only on strata $(|\partial_r F| \ll |\nabla' F|)$

* $r\frac{dg}{ds} \geq |g|^{\mu}$, but maybe $\mu \geq 1$

* $\frac{dg}{d\tilde{s}} \geq c|g|^{\xi}$ only along the trajectory

Dynamical Systems: Continuous case

- Hartman-Grobman: $x'(t) = u(x(t))$. Around hyperbolic fixpoint, flows of linear and non-linear system are homeomorphic.
- Hartman: If $Du(0) < 0$, then trajectories of linear and non-linear system are diffeomorphic.
- for gradient descent: $H_f(0) > 0$
- trajectory for linear system well-understood
- implies **strong** gradient conjecture $x'(t) = -\nabla f(x(t))$

Dynamical Systems: Continuous case

- Hartman-Grobman: $x'(t) = u(x(t))$. Around hyperbolic fixpoint, flows of linear and non-linear system are homeomorphic.
- Hartman: If $Du(0) < 0$, then trajectories of linear and non-linear system are diffeomorphic.
- for gradient descent: $H_f(0) > 0$
- trajectory for linear system well-understood
- implies **strong** gradient conjecture $x'(t) = -\nabla f(x(t))$

Dynamical Systems: Continuous case

- Hartman-Grobman: $x'(t) = u(x(t))$. Around hyperbolic fixpoint, flows of linear and non-linear system are homeomorphic.
- Hartman: If $Du(0) < 0$, then trajectories of linear and non-linear system are diffeomorphic.
- for gradient descent: $H_f(0) > 0$
- trajectory for linear system well-understood
- implies **strong** gradient conjecture $x'(t) = -\nabla f(x(t))$

Dynamical Systems: Discrete case

- Gradient map $F(x) = x - \alpha \nabla f(x) : x_k \to x_{k+1}$ is diffeomorphism around strict local minimum, if α small.

- Sebastian van Strien, 1990 in *Smooth linearization of hyperbolic fixed points without resonance conditions* [vS90]: If F is C^2 diffeo,

$$F(x) = \psi^{-1} \circ DF(0) \circ \psi(x), \quad \text{with } \psi, \psi^{-1} \text{ differentiable at 0.}$$

- $x_k = F^k(x_0) = F \circ F \circ \cdots \circ F(x_0)$, use same ψ for F^k

$$F^k(x) = \psi^{-1} \circ DF(0)^k \circ \psi(x)$$

Dynamical Systems: Discrete case

- Gradient map $F(x) = x - \alpha \nabla f(x) : x_k \to x_{k+1}$ is diffeomorphism around strict local minimum, if α small.

- Sebastian van Strien, 1990 in *Smooth linearization of hyperbolic fixed points without resonance conditions* [vS90]: If F is C^2 diffeo,

$$F(x) = \psi^{-1} \circ DF(0) \circ \psi(x), \quad \text{with } \psi, \psi^{-1} \text{ differentiable at } 0.$$

- $x_k = F^k(x_0) = F \circ F \circ \cdots \circ F(x_0)$, use same ψ for F^k

$$F^k(x) = \psi^{-1} \circ DF(0)^k \circ \psi(x)$$

Dynamical Systems: Discrete case

- Gradient map $F(x) = x - \alpha\nabla f(x) : x_k \to x_{k+1}$ is diffeomorphism around strict local minimum, if α small.
- Sebastian van Strien, 1990 in *Smooth linearization of hyperbolic fixed points without resonance conditions* [vS90]: If F is C^2 diffeo,

$$F(x) = \psi^{-1} \circ DF(0) \circ \psi(x), \quad \text{with } \psi, \psi^{-1} \text{ differentiable at 0.}$$

- $x_k = F^k(x_0) = F \circ F \circ \cdots \circ F(x_0)$, use same ψ for F^k

$$F^k(x) = \psi^{-1} \circ DF(0)^k \circ \psi(x)$$

Dynamical Systems: Discrete case

- for linear system $\lim_k \frac{y_k}{|y_k|}$ exists.
- $\implies \lim_k \frac{x_k}{|x_k|}$ exists .
- discrete Gradient Conjecture: if $f \in C^3$ and x^* strict local minimum
- What if f is Morse-Bott ?

Dynamical Systems: Discrete case

- for linear system $\lim_k \frac{y_k}{|y_k|}$ exists.
- $\implies \lim_k \frac{x_k}{|x_k|}$ exists .
- discrete Gradient Conjecture: if $f \in C^3$ and x^* strict local minimum
- What if f is Morse-Bott ?

Dynamical Systems: Discrete case

- for linear system $\lim_k \frac{y_k}{|y_k|}$ exists.
- $\implies \lim_k \frac{x_k}{|x_k|}$ exists .
- discrete Gradient Conjecture: if $f \in C^3$ and x^* strict local minimum
- What if f is Morse-Bott ?

Thank you for listening!

Pierre-Antoine Absil, Robert E. Mahony, and B. Andrews.
Convergence of the iterates of descent methods for analytic cost functions.
SIAM J. Optim., 16:531–547, 2005.

Krzysztof Kurdyka, Tadeusz Mostowski, and Adam Parusinski.
Proof of the gradient conjecture of R. Thom.
Annals of Mathematics, 152, 07 1999.

Stanislaw Lojasiewicz.
Trajectoires du gradient d'une fonction analytique.
Seminari di Geometria 1982-1983, pages 115–117, 1984.

Sebastian van Strien.
Smooth linearization of hyperbolic fixed points without resonance conditions.
Journal of Differential Equations, 85(1):66–90, 1990.