École polytechnique fédérale de Lausanne

Semester project

Minor in Computational Science and Engineering

On the continuous and discrete gradient conjecture

Author:
Florian Grün

Supervisor:
PhD Quentin Rebjock
Professor Nicolas Boumal

Contents

1 Introduction

We study geometrical convergence properties of gradient descent in the continuous and discrete settings of the function $f : \mathbb{R}^n \to \mathbb{R}$. In the former case, we will study the *continuous gradient descent*, i.e. the flow of the system of differential equations given by

$$x'(t) = -\nabla f(x(t)), \quad t \geq 0, \quad x(0) = x_0. \tag{1.1}$$

The *discrete gradient descent* version in turn is given by a sequence $\{x_k\}_{k=0}^{\infty}$ with

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \tag{1.2}$$

where α_k denotes the step size at step k. Discrete gradient descent (and its variants like BFGS) are widely used in numerical optimization, since it is fairly robust (i.e. usually convergence guaranteed), easy to implement and only requires information on the first derivative.

Naturally one is interested in the convergence of the solution (which of course depends on the initial point) as t (or k respectively) tends to ∞, however theoretical guarantee of convergence is a non-trivial problem. A major breakthrough (Theorem 1.11) for the continuous system was done in [Loj84] under the assumption that f is analytic, which was extended to the discrete case in [AMA05]. Both results make use of the *Lojasiewicz inequality* for analytic functions; in the non-analytic setting counterexamples are well known, e.g. "Mexican hat" functions, see [AMA05]. We note that the results mentioned above, use tools from real analytic geometry, in particular semialgebraic and subanalytic sets, stratifications and resolution of singularities. Lojasiewicz-Simon inequalities are an extension to infinite dimensional spaces and find applications in the study of convergence properties of PDEs, see [CM14]. Under additional assumptions on f, like nondegeneracy of the Hessian, convergence properties are easier to obtain (using tools from dynamical systems instead), see [HM94].

In this exposition we study in what manner trajectories approach their limit points. Do they go in a straight line or circle around the minimum? Here the major result (the gradient conjecture, Theorem 2.1) was obtained in [KMP99], showing that as long as f is analytic, the trajectories of the continuous system converge in a line to the critical point. The question whether that will also happen for the discrete trajectory remains open. First we recall the main results from [KMP99] and fill in some explanatory details. Consequently we compare different versions of the gradient conjecture and show that for a smooth, non-analytic function, it can happen that the trajectory converges in a spiral. Lastly we discuss difficulties and problem in our discretization attempt and provide a different strategy based on tools from dynamical systems. We show a stronger version of the gradient conjecture and a result in the discrete case, but have additional assumptions on the critical point.

1.1 Definitions and preliminary results

Definition 1.1. Given an open set $U \subseteq \mathbb{R}^n$, we say a function $f : U \to \mathbb{R}$ is *(real) analytic* if for each x_0 in U, there exists a neighborhood $V(x_0)$ and a sequence of homogeneous polynomials $P_k : \mathbb{R}^n \to \mathbb{R}$ with degree n such that

$$f(x) = \sum_{k=0}^{\infty} P_k(x - x_0) \qquad \forall x \in V(x_0).$$

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is analytic if each component function is analytic.

Definition 1.2. A set $A \subset \mathbb{R}^n$ is called *analytic* if it can be (locally) written as the finite union of zero sets of analytic functions $\{f_k\}_{k=1}^{K}$, i.e.

$$A = \bigcup_{1 \leq k \leq K} \{x \in \mathbb{R}^n : f_k(x) = 0\}.$$

Definition 1.3. A set $S \subset \mathbb{R}^n$ is called *semianalytic* if each $x_0 \in \mathbb{R}^n$ admits a neighborhood $V(x_0)$ such that

$$S \cap V(x_0) = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in V : f_{ij}(x) = 0, g_{ij}(x) > 0\},$$

where f_{ij}, g_{ij} are real analytic functions.

Definition 1.4. A set $S \subset \mathbb{R}^n$ is called *subanalytic* if each $x_0 \in \mathbb{R}^n$ admits a neighborhood $V(x_0)$ such that

$$S \cap V(x_0) = \{x \in \mathbb{R}^n : (x, y) \in B\},$$

with B being a bounded semianalytic subset of $\mathbb{R}^n \times \mathbb{R}^m$, i.e. S is the projection of a compact semianalytic set.

Definition 1.5. A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called *subanalytic* if its graph is a subanalytic set of $\mathbb{R}^n \times \mathbb{R}$.

We have the following elementary properties:

- Analytic functions are closed under addition, multiplication, composition and division.
- Subanalytic, semianalytic and semialgebraic sets are closed under locally finite union and intersection, and complement.
- Given a subanalytic set S, the distance $d_S(x) := \inf\{\|x - a\| : a \in S\}$ is a subanalytic function.
- Any semianalytic set is subanalytic, in $\mathbb{R}^2$ any subanalytic set is semianalytic.
- Any smooth manifold is a semianalytic and subanalytic set.

Lemma 1.6. (Curve selection lemma, [Kur98]) *Let S be a subanalytic set such that $\bar{S} \ni 0$. There exists a real analytic curve $\alpha : (-\varepsilon, 0] \to \mathbb{R}^n$ with $\alpha(0) = 0$ and $\alpha(t) \in S$.*

Theorem 1.7. (Lojasiewicz inequality [Loj65]) *Let $f : U \to \mathbb{R}$ be an analytic function with $Z := \{x : f(x) = 0\}$. Then for any compact set $K \subset U$, there exists $\alpha, \alpha' = 1/\alpha$ and C such that*

$$\operatorname{dist}(x, Z)^\alpha \le C|f(x)| \quad \forall x \in K \quad \text{or equivalently} \quad \operatorname{dist}(x, Z) \le C^{\alpha'}|f(x)|^{\alpha'} \quad \forall x \in K.$$

Theorem 1.8. (Lojasiewicz inequality for the gradient [Loj65]) *Let $f : U \to \mathbb{R}$ be an analytic function with a zero at the origin. Then there exists $p < 1$, $c > 0$ such that*

$$|\nabla f(x)| \ge c|f(x)|^\rho,$$

for all x in a neighborhood U_0 of the origin.

Remark 1.9. The Lojasiewicz inequalities are closely connected. The first is used to prove the second one. If the Lojasiewicz inequality holds for exponent ρ, it holds automatically for all $\rho' > \rho$, i.e. the "better" the function, the smaller the Lojasiewicz exponent. Since analytic functions are differentiable, it must hold that $\rho \ge \frac{1}{2}$. The Lojasiewicz inequality holds for more functions in general, in particular [Kur98] shows the Lojasiewicz gradient inequality for differentiable subanalytic functions. We could not find a reference whether the standard Lojasiewicz inequality holds for subanalytic functions.

Theorem 1.10. (Bochnak–Lojasiewicz inequality [KMP99, Lemma 4.3]) *Let f be an analytic function with a zero at the origin. Then there exists a neighborhood U_0 and a constant $c_f > 0$*

$$|x||\nabla f(x)| = r|\nabla f(x)| \ge c_f|f| \qquad \forall x \in U_0.$$

Theorem 1.11. (Lojasiewicz theorem [Loj84]) *If $x(t)$ has a limit point x^*, i.e. $x(t_v) \to x^*$ for some sequence $t_v \to \beta$, then the length of $x(t)$ is finite (to be exact $\int_0^\beta |x'(t)|dt \le c_1|f(x(0))|^{1-\rho}$) and $\beta = \infty$. Hence $x(t) \to x^*$ as $t \to \infty$.*

Remark 1.12. A function f is **Polyak-Lojasiewicz (PL)** if the Lojasiwiecz exponent is $\rho = 1/2$. Furthermore a non-degenerate critical point x^* implies PL and PL holds if and only if f is a Morse-Bott function.

Theorem 1.13. (Puiseux expansion [MC13, PW]) *Let $P : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ a polynomial in two variables, i.e. $P(x, y) = 0$. Then each solution $y^i(x) = (y_1^i(x), ..., y_m^i(x))$ can be written locally as a Puiseux series, i.e. for $k = 1, ..., m$,*

$$y_k^i(x) = a_{r_1}x^{r_1} + a_{r_2}x^{r_2} + \dots \qquad a_{r_1} \neq 0 \text{ and } r_i \in \mathbb{Q}$$

$$\implies P(x, y(x)) = b_{q_1}x^{q_1} + b_{q_1}x^{q_1} + b_{q_2}x^{q_2} + \dots \qquad a_{r_1} \neq 0 \text{ and } q_i \in \mathbb{Q}$$

Lemma 1.14. *For any $f \in C^2$, a trajectory of $x'(s) = \frac{\nabla f}{|\nabla f|}$ (a reparametrization by arclength), converging to x^* has a uniform bound on its curvature κ.*

Proof. Since s is the arc-length parameter, $\kappa = |x''(s)|$ and

$$\begin{aligned} x''(s) = \frac{\nabla f(x(s))}{|\nabla f(x(s))|} &= \frac{d}{ds}\frac{1}{|\nabla f(x(s))|}\nabla f(x(s)) + \frac{1}{|\nabla f(x(s))|}H_f(x(s)) \cdot \nabla f(x(s)) \\ &= \frac{-\nabla f(x(s)) \cdot (H_f(x(s)) \cdot \nabla f(x(s)))}{|\nabla f(x(s))|^3}\nabla f(x(s)) + \frac{1}{|\nabla f(x(s))|}H_f(x(s)) \cdot \nabla f(x(s)). \end{aligned}$$

Thus by Cauchy-Schwarz,

$$\kappa = |x''(s)| \leq 2\|H_f(x(s))\| \leq C,$$

since f as C^2 function has bounded second derivatives in a compact neighborhood of x^*. □

2 Proof of the gradient conjecture following [KMP99]

We review the result in [KMP99] and explain in detail their approach. The proof is based on a control function and a Lojasiewicz type argument (cf. [Loj84]). Note that [KMP99] deals with gradient ascent, i.e. for this section consider $\dot{x}(t) = \nabla f(x(t))$. Recall also the arclength reparametrization $x(s) = \frac{f(x(s))}{|\nabla f(x(s))|}$. Without loss of generality, assume that $f(x(t)) \leq 0$ for all $s < s_0$, since $\frac{df(x(t))}{dt} = |\nabla f(x(t))| \geq 0$, the critical point x^* is the origin and $f(x^*) = 0$. We prove:

Theorem 2.1. (The Gradient Conjecture) *Suppose that $x(t) \to x*$. Then $x(t)$ has a tangent at x^*, that is the limit of secants $\lim_{t\to\infty} \frac{x(t)-x^*}{|x(t)-x^*|}$ exists.*

2.1 Characteristic exponents

First split the gradient ∇f into its radial $\partial_r f \partial_r$ and spherical $\nabla' f(x) := \nabla f - \partial_r f \partial_r$ component. Define, for $\varepsilon > 0$, the set

$$W^\varepsilon = \{x : f(x) \neq 0, \quad \varepsilon|\nabla' f| \leq |\partial_r f|\},$$

that is the spherical change is controlled by the radial change. Clearly $W^\varepsilon \subset W^{\varepsilon'}$ whenever $\varepsilon' < \varepsilon$. In the following lemma, we provide a detailed explanation of a classical argument in subanalytic geometry using Lemma 1.6 similar to the proof of Theorem 1.10 in [MR09], which is not given in [KMP99].

Lemma 2.2. [KMP99, Lemma 4.1] *For each $\varepsilon > 0$, there exist a constant $c = c(\varepsilon)$ such that inside W^ε*

$$|f| \geq cr^{(1-\rho)^{-1}}.$$

Proof. Assume that the origin is in the closure of $S = W^\varepsilon \cap \{f < cr^{(1-\rho)^{-1}}\}$. Then as S is a semianalytic set, there exists an analytic curve $\gamma_1 : (-\varepsilon, 0]$, such that $\gamma_1(t) \in S$ for $t \in (-\varepsilon, 0)$ and $\gamma_1(0) = 0$. If we show that $|f|r^{-(1-\rho)^{-1}}$ is bounded away from zero for any analytic curve $\gamma(t) \in W^\varepsilon$ with $\gamma(t) \to 0$ as $t \to 0$, then γ_1 can not exist. Then $0 \notin \bar{S}$ and $\bar{S} \cap B_\delta(0) = \varnothing$ for δ sufficiently small, hence in all of $W^\varepsilon \cap B_\delta(0)$, we have $|f| \geq cr^{(1-\rho)^{-1}}$.

Reparametrize γ by its distance to the origin $|\gamma(t(r))| = r$, so $\gamma(r) = r\theta(r)$ with $\theta : \mathbb{R} \to \mathbb{S}^{n-1}$. We note that $\gamma'(r) = \theta(r) + r\theta'(r)$ with $r\theta'(r) \to 0$ as $r \to 0$. Now by Theorem 1.13 with $x = r$ and $y(x) = \theta(r)$, we have a Puiseux expansion of the form

$$f(\gamma(r)) = a_l r^l + ... \quad , \qquad a_l \neq 0, l \in \mathbb{Q}^+ .$$

$$\begin{aligned} l a_l r^{l-1} + ... = \frac{df}{dr}(\gamma(r)) &= \langle \nabla f(\gamma(r)), \gamma'(r)\rangle = \langle \partial_r f \partial_r + \nabla' f, \theta(r) + r\theta'(r)\rangle \\ &= \langle \partial_r f \partial_r, \theta(r)\rangle + \langle \nabla' f, \theta(r)\rangle + \langle \partial_r f \partial_r, r\theta'(r)\rangle + \langle \nabla' f, r\theta'(r)\rangle \\ &= 0 + |\nabla' f| + \langle \partial_r f \partial_r, \partial_r\rangle + \langle \nabla' f, r\theta'(r) \begin{cases} \leq |\partial_r f| + \frac{2}{\varepsilon}|\partial_r f| \\ \geq |\partial_r f| - \frac{2}{\varepsilon}|\partial_r f|. \end{cases} \end{aligned}$$

This step is not completely clear to us, why the need for a Puiseux expansion? Is not $f(\gamma(r))$ analytic in r? Also we could not find a clean statement of the Puiseux theorem. Continuing, it follows that

$$r|\nabla f| = r(|\nabla' f| + |\partial_r f|) \leq r\left(1 + \frac{1}{\varepsilon}\right)|\partial_r f| \implies r|\nabla f| \sim r|\partial_r f| \sim r\frac{df}{dr} \sim r^l,$$

so $|\nabla f| \sim r^{l-1}$ and $|f| \sim r^l$. Using Theorem 1.8 around the origin,

$$r^l \geq c_1 r|\nabla f| \geq c_1 r|f|^\rho \geq c_1 r c_2 r^{l\rho} \iff r^{l-1-l\rho} \geq c \iff \frac{l-1}{l} \leq \rho \iff (1-\rho)^{-1} \geq l.$$

Thus, for r small, the result $|f| \geq cr^l \geq cr^{(1-\rho)^{-1}}$ follows directly. $\square$

Now the set W^ε will be stratified, depending on how x approaches the origin, i.e. the limit of the function $\frac{r\partial_r f}{f}(x)$ as $x \to 0$. The next proposition consist of three statements, each characterizing the structure of the stratification of W^ε. We provide a detailed explanation of the second part.

Proposition 2.3. [KMP99, Proposition 4.2] *Let $\phi : W^\varepsilon \to \mathbb{R}$ be defined as above. Then*

(i) There exists a finite subset of positive rationals $L \subset \mathbb{Q}^+$, such that for any $\varepsilon > 0$

$$\frac{r\partial_r f}{f}(x) \to L \qquad \text{as } x \to 0 \quad (\text{i.e. for some } l_i \in L,\ \frac{r\partial_r f}{f}(x) \to\to l_i).$$

(ii) Define $W^\varepsilon_{l_i} = \{x \in W^\varepsilon : \left|\frac{r\partial_r f}{f}(x) - l_i\right| \leq r^\delta\}$ ($\delta > 0$ sufficiently small), then locally around the origin $W\varepsilon$ can be stratified as

$$W^\varepsilon = \bigsqcup_{l_i \in L} W^\varepsilon_{l_i}.$$

(iii) There exist positive constants $c_\varepsilon, C_\varepsilon$ such that

$$c_\varepsilon < \frac{|f|}{r^{l_i}} < C_\varepsilon \qquad \text{on } W^\varepsilon_{l_i}.$$

Proof. If $0 < \varepsilon' \leq \varepsilon$, we have $W^\varepsilon \subset W^{\varepsilon'}$, so also $L_\varepsilon \subset L_{\varepsilon'}$. Furthermore we have the stabilizing property $L_{\varepsilon_1} = L_{\varepsilon'}$ if $\varepsilon_1 > 0$ sufficiently small and $0 < \varepsilon' \leq \varepsilon_1$. This follows apparently from some Noetherian ring theorem. Then we first show that L_ε is subanalytic in $\mathbb{R}_\infty$. Using the curve selection lemma we can show that $L_\varepsilon \subset \mathbb{Q}^+$, for each curve γ, the limit, when approaching the origin, is l, with l the exponent from the Puisseux expansion. This then somehow implies that L_ε finite for each ε. Then just take L to be L_{ε_1} with ε_1 small enough, by stabilizing and inclusion property. This part is still unclear to us, it uses heavy results from analytic geometry, for details see [KMP99].

Since L is finite, let $\beta = \max_{i,j} |l_i - l_j|$ and let c such that $c > |x|$ implies

$$\left|\frac{r\partial_r f}{f}(x) - l_i\right| < \frac{\beta}{3} \qquad \text{for some } l_i \in L,$$

which is possible by the convergence in *(i)*. Hence we can already define the disjoint sets

$$\tilde{W}^\varepsilon_{l_i} := \{x \in W^\varepsilon : |x| < c\left|\frac{r\partial_r f}{f}(x) - l_i\right| \leq \frac{\beta}{3}\}.$$

Now it remains to improve the bound $\frac{\beta}{3}$. Pick one $\tilde{W}^\varepsilon_{l_i}$, then by [Loj65, Section 24, Corollary 1],

$$\left|\frac{r\partial_r f}{f}(x) - l_i\right| = \left|\frac{r\partial_r f}{f}(x) - l_i\right| = \lim_{\tilde{W}^\varepsilon_{l_i} \ni y \to 0}\left|\frac{r\partial_r f}{f}(x) - \frac{r\partial_r f}{f}(y)\right| \leq C \lim_{\tilde{W}^\varepsilon_{l_i} \ni y \to 0} |x-y|^{\delta_i} = C|x|^{\delta_i}.$$

The result is only stated for semianalytic functions, we have to assume it holds for the subanalytic function $\frac{r\partial_r f}{f}$, it is also not clear to us why this function is subanalytic. Maybe [KMP99] is using something else from [Loj65]. The exponents δ_i depend on $\tilde{W}^\varepsilon_{l_i}$, but since there are only finitely many, let δ be their minimum. For the third part, let $\gamma(r) \to 0$ be a curve in $W^\varepsilon_{l_i}$. The Puiseux expansion of $f(\gamma(r))$ has l_i as its first exponent, otherwise $\left|\frac{r\partial_r f}{f} - l_i\right| \geq c|l - l_i| \neq 0$, for $l \neq l_i$ and thus γ does not stay in $W^\varepsilon_{l_i}$ and the estimate holds along γ, the result follows by the curve selection lemma, similar to Lemma 2.2. $\square$

2.2 Asymptotic critical values

In this section we will introduce a generalization of the concept of criticality. We follow the outline in [KMP99], but skip one proof, moreover we are not entirely sure about the proof of Proposition 2.6, a strong version of Lojasiewicz's inequality for asymptotic critical values.

Definition 2.4. Let $F : U \to \mathbb{R}$ be a subanalytic C^1 function and $0 \in \bar{U}$. We call $a \in \mathbb{R}$ an *asymptotic critical value of F at the origin* if there exists a sequence $x_k \to 0, x_k \in U$ such that

(a) $|x_k||\nabla F(x_k)| \to 0$ or $|\nabla_\theta F(x_k)| = |x_k||\nabla' F(x_k)| \to 0$,

(b) $F(x_k) \to a$.

Clearly any usual critical value is also an asymptotic critical value, however e.g. 1 is a critical value for $F(x) = e^x$, despite $F'(x) = e^x \neq 0$ as $x \to 0$. The condition $\nabla F(x) = 0$ of a critical point is replaced by a weaker condition that $\nabla F(x) = o(1/|x|)$ along some sequence. The theory for asymptotic critical values is necessary, since later we will not directly at the critical value of f, but at $F = \frac{f}{r^l}$, which has only an asymptotic critical value at the origin.

Proposition 2.5. [KMP99, Proposition 5.1] *Let $F : U \to \mathbb{R}$ be a subanalytic C^1 function, then the set of asymptotic critical values is finite (and thereby isolated).*

Proof. [KMP99] uses the (w)-condition from Kuo-Verdier, for more see also [KOS00, Kur98]. □

The next proposition gives a Lojasiewicz estimate even for asymptotic critical values, provided that we are in a restricted domain. This is a big step towards the gradient conjecture. It is also important to note that there is an extra factor of r, like in Theorem 1.10, but an exponent $\rho_a < 1$. Estimates like this are known as **strong Lojasiewicz estimates**.

Proposition 2.6. [KMP99, Proposition 5.3] *Let F be C^1 and subanalytic on U (and $0 \in \bar{U}$) and a an asymptotic critical value at the origin. Then for any $\eta > 0$, there exists an exponent $\rho_a < 1$ and constants $c, c_a > 0$ such that in the set*

$$Z = Z_\eta := \{x \in U : |\partial_r F| \leq r^\eta |\nabla F|, |F(x) - a| \leq c_a\}$$

we have the Lojasiewicz type estimate

$$r|\nabla F| \geq C|F - a|^{\rho_a}.$$

Proof. Without loss of generality assume $a = 0$ (otherwise use $\tilde{F} := F - a$) and fix c_0 such that $\{t : |t| \leq c_0\}$ does not contain any other asymptotic critical values (asymptotic critical values are isolated). Then on Z_η

$$\frac{\langle \nabla F(x), x \rangle}{|\nabla F(x)||x|} = \frac{\partial_r F|x|}{|\nabla F(x)||x|} = \frac{\partial_r F}{|\nabla F(x)|} \leq r^\eta \to 0 \quad \text{as } Z \ni x \to 0.$$

Let now $\gamma(t)$ be a real analytic curve such that $\gamma(t) \to 0$ and $F(\gamma(t)) \to 0$ as $t \to 0$. The normalizations $\frac{d\gamma/dt}{|d\gamma/dt|}$ and $\frac{\gamma}{|\gamma|}$ have the same limit (this follows from our understanding by the analyticity of γ),

$$\frac{\langle \nabla F(\gamma(t)), d\gamma/dt \rangle}{|\nabla F(\gamma(t))||d\gamma/dt|} = \left\langle \nabla F(\gamma(t)), \frac{1}{|\nabla F(\gamma(t))||d\gamma/dt|} \right\rangle = \frac{\langle \nabla F(\gamma(t)), \gamma \rangle}{|\nabla F(\gamma(t))||\gamma|} \to 0 \quad \text{as } t \to 0.$$

(Would the equivalence of the limits for analytic curves not prove the strong gradient conjecture?) Hence $dF/dt = \langle \nabla F, d\gamma/dt \rangle = o(|\nabla F||d\gamma/dt|)$ and so $F(x) = o(|\nabla F||x|)$ along γ and

$$\frac{F(x)}{|\nabla F(x)||x|} \to 0 \quad \text{as } Z \ni x \to 0 \text{ and } F(x) \to 0.$$

Now let $\gamma(t) \to x_0 \neq 0$, then since the Lojasiewicz inequality holds for F being a C^1 subanalytic function,

$$\left| \frac{F(x)}{|\nabla F(x)||x|} \right| \leq \frac{|F|}{C|F|^\rho c} = C|F|^{1-\rho} \to 0 \quad \text{as } c < |x| \in Z \text{ and } F(x) \to 0.$$

In the case where $|x|$ is not bounded below, the bound by $|F|^\alpha$ holds nevertheless. Let y_k be such that $|y_k| \leq |F(y_k)|^\alpha \to 0$ (exist by Def. 2.4), from [Loj65, Section 24, Corollary 1] and Theorem 1.7,

$$\frac{|F(x)|}{|\nabla F(x)||x|} = \lim_{y_k \to 0} \left| \frac{F(x)}{|\nabla F(x)||x|} - \frac{|F(y)|}{|\nabla F(x)||y|} \right| \leq C \lim_{y_k \to 0} |x-y|^{\alpha_1} = C|x|^{\alpha_1} \leq C|F|^{\alpha_1\alpha_2} = C|F|^\alpha$$

$$\implies \frac{|F(x)|}{|\nabla F(x)||x|} \leq C|F|^\alpha \quad \text{on } Z \text{ for } \alpha > 0 \text{ small} \iff |F(x)|^{1-\alpha} \leq C|\nabla F(x)||x|.$$

We are not sure if the last estimate is right, as we use a theorem for semianalytic functions and the standard Lojasiewicz inequality on F, which was only subanalytic. Maybe [KMP99] does some different argument using results from [Loj65]. □

Proposition 2.7. [KMP99, Proposition 5.4] *Let $F = \frac{f}{r^l}$ with f analytic and $l \in L$. Then alternatively, $\mathbb{R} \ni a \neq 0$ is an asymptotic critical value if and only if there exists a sequence $x_k \to 0$, $x_k \neq 0$ such that*

(a') $\frac{|\nabla' f(x_k)|}{|\partial_r f(x_k)|} \to 0$,

(b) $F(x_k) \to a$.

Proof. We show the equivalence $(a) \equiv (a')$, with (a) from Definition 2.4. First, assume that $x_k \to 0$ is a sequence in W_ε^c, then

$$\frac{|\nabla' f(x_k)|}{|\partial_r f(x_k)|} \geq \varepsilon > 0$$

so (a') does not hold. As in $(W^\varepsilon)^c$, $(1+\varepsilon)|\nabla' f| \geq |\nabla f|$, using $r|\nabla f| \sim |f|$, we have

$$|x_k||\nabla F(x_k)| \geq r|\nabla' F| = \frac{r|\nabla' f|}{r^l} \geq \frac{1}{1+\varepsilon}\frac{r|\nabla f|}{r^l} \geq c\frac{|f|}{r^l} \geq \frac{1}{2}c|a| > 0,$$

so (a) is not satisfied. Hence assume $W^\varepsilon \ni x \to 0$. Let us show that $|x_k||\nabla F(x_k)| \to 0$ if and only if $|x_k||\nabla' F(x_k)| \to 0$, in other words, we want to show that $r|\partial_r F(x_k)| \to 0$. But if not, then $|\partial_r F(x_k)| \sim \frac{1}{r}$ as $x_k \to 0$, the derivative blows-up, hence it $F(x_k)$ can not approach the finite value a. In turn, as $|\partial_r f| \sim r^{l-1}$,

$$\frac{|\nabla' f(x_k)|}{|\partial_r f(x_k)|} \sim \frac{|\nabla' f(x_k)|}{r^{l-1}} = r|\nabla' F(x_k)| \to 0,$$

and thereby (a) if and only if (a'). □

Remark 2.8. Note that $F = \frac{f}{r^l}$ is differentiable outside the origin and as quotient of two analytic functions a subanalytic C^1 function in the sense of Definition 2.4 and Proposition 2.6 applies.

2.3 Estimates on trajectory

Using the tools, develloped in the previous two sections, we will now start to estimate the trajectory. Let us repeat the following definitions on $\mathbb{R}^n \setminus \{0\}$:

- $W^\varepsilon := \{x : \varepsilon|\nabla' f| \leq |\partial_r f|\} = \bigcup W_{l_i}^\varepsilon$.
- $W_{l_i}^\varepsilon := \{x \in W^\varepsilon : |\frac{r\partial_r f(x)}{f(x)} - l_i| \leq r^\delta\}$ with $\delta > 0$ sufficiently small.
- $W_{-w,l_i} := \{x \in W_{l_i}^\varepsilon : r^{-w}|\nabla' f| \leq |\partial_r f|\}$ with $\omega > 0$ which can be thought of as a refinement to $W_{l_i}^\varepsilon$; close to the origin the radial part is much bigger than the spherical one, the ratio $\frac{|\partial_r f|}{|\nabla' f|}$ will converge to infinity.
- $U_l := \{x : c < \frac{|f(x)|}{r^l} < C\}$ with $c, C > 0$.
- On W^ε (and so on W_l^ε) we have the equivalences $r|\nabla f| \sim r|\partial_r f| \sim r|df/dr| \sim r^l \sim |f|$.
- From now on $F(x) = \frac{f(x)}{r^l}$ with f analytic and l not necessarily in $L = L_f$.

Again, we follow the main outline in [KMP99], but provide a more detailed argument in the proof of Proposition 2.10 and split up the proof of Proposition 2.15 into several parts, where we prove the bounds from [KMP99, Remark 6.3] in detail.

Proposition 2.9. [KMP99, Proposition 6.1] *For each $l > 0$ there exist $\varepsilon, w > 0$ such that $F(x(s))$ is strictly increasing in either*

(a) the complement of $S_1 := \bigcup_{l_i \in L, l_i < l} W^\varepsilon_{l_i}$ *if* $l \notin L$

(b) the complement of $S_2 := W_{-w,l} \cup \bigcup_{l_i \in L, l_i < l} W^\varepsilon_{l_i}$ *if* $l \in L$

Also S_1 and S_2 are both subsets of W^ε.

Proof. Assume that $F(x(s))$ decreases and show that the trajectory is necessarily in S_1 or S_2.If

$$\frac{dF(x(s))}{ds} = \langle (x(s)), x'(s)\rangle = \left\langle \frac{\nabla' f}{r^l} + \left(\frac{\partial_r f}{r^l} - \frac{lf}{r^{l+1}}\right)\partial_r, \frac{\nabla f}{|\nabla f|}\right\rangle$$
$$= \frac{1}{|\nabla f| r^l}\left(|\nabla' f|^2 + |\partial_r f|^2\left(1 - \frac{lf}{r\partial_r f}\right)\right) \leq 0,$$

then in turn

$$0 \geq |\nabla' f|^2 + |\partial_r f|^2\left(1 - \frac{lf}{r\partial_r f}\right) = |\nabla' f|^2 + |\partial_r f|^2 - \frac{lf\partial_r f}{r} \implies r|\nabla f|^2 \leq lf\partial_r f.$$

Using the Bochnak–Lojasiewicz inequality (Theorem 1.10),

$$lf\partial_r f \geq r|\nabla f|^2 \geq c_f|f||\nabla f| \implies |\partial_r f| \geq \frac{c_f}{l}|\nabla f|,$$

hence we are in W^ε with $\varepsilon = c_f/l$. For the curve $x(s)$, $\frac{r\partial_r f}{f} \to l_i$ and so $x(s)$ lies in $W^\varepsilon_{l_i}$. We have the following:

1. If $l < l_i$, then $\left(1 - \frac{lf}{r\partial_r f}\right) \to (1 - l/l_i) > 0$, so $F(x(s))$ is actually strictly increasing.

2. If $l = l_i$, then $\left(1 - \frac{lf}{r\partial_r f}\right) \to 0$ and so by the definition of $W^\varepsilon_{l_i}$ there exists $\omega > 0$ such that $\left|1 - \frac{lf}{r\partial_r f}\right| \leq r^{2\omega}$. Thereby

   $$|\nabla' f|^2 - r^{2\omega}|\partial_r f|^2 \leq 0 \implies |\nabla' f| r^{-w} \leq |\partial_r f|,$$

   and the trajectory is in $W_{-\omega,l}$.

3. If $l > l_i$, then $F(x(s))$ can decrease in $W^\varepsilon_{l_i}$.

The case (a) follows directly from the first and third point. For case (b) (if $l \in L$), as $l > l_j$ for all $l_j < l_i$, we can decrease in $W_{-w,l} \cup \bigcup_{l_i \in L, l_i < l} W^\varepsilon_{l_i}$; note that this set is contained in $\bigcup_{l_i \in L, l_i \leq l} W^\varepsilon_{l_i}$ the estimate is "a bit better" than in the case $l \notin L$. □

From the next proposition, we have that around the origin the trajectory spends at least some time in the "good" set W^ε and a subset U_l, containing the trajectory completely.

Proposition 2.10. [KMP99, Proposition 6.2] *There exists a unique $l = l_i \in L$ and $0 < \varepsilon, c, C < \infty$ such that $x(s)$ passes through W^ε_l in any neighborhood of the origin. Also $x(s)$ is completely contained in U_l for s close to s_0.*

Proof. Let $\varepsilon > 0$ such that $\varepsilon < c_f(1 - \rho_f)$. Suppose that $x(s)$ does not pass through W^ε for any $\varepsilon > 0$. Since $(W^\varepsilon)^c$ is a subset of the sets in Proposition 2.9, $F(x(s)) = \frac{f(x(s))}{r^l}$ is increasing as long as $l > c_f/\varepsilon > (1 - \rho_f)^{-1}$. Now, as F is increasing and $f = -|f|$ along the trajectory, in a neighborhood of the origin

$$-C_l \leq F(x(s)) = \frac{-|f(x(s))|}{r^l} \implies |f(x(s))| \leq C_l r^l.$$

Remember that s is the arc length parameter, so $|s - s_0| = \int_s^{s_0} |x'(\tau)| d\tau$ and by Theorem 1.11, the arclength is bounded, i.e.

$$|s - s_0| \leq c_1 |f(x(s))|^{1-\rho_f}.$$

Yet also,

$$r = |x(s)| = |x(s) - x(s_0)| = \int_s^{s_0} x'(\tau)d\tau \leq \int_s^{s_0} |x'(\tau)|d\tau = |s - s_0|$$

thus from the previous three estimates

$$|s - s_0| \leq c_1|f(x(s))|^{1-\rho_f} \leq Cr^{l(1-\rho_f)} \leq C|s - s_0|^{l(1-\rho_f)} \qquad \forall s < s_0.$$

This leads to a contradiction as $s \to s_0$, since $l(1 - \rho_f) > 1$, thus $x(s)$ passes through W^ε in any neighborhood of the origin. Since there are only finitely many $W^\varepsilon_{l_i}$, $x(s)$ passes through $W^\varepsilon_{l_i}$ in any neighborhood of the origin.

To show $x(s) \in U_l$ for any s close to s_0, notice that each $W^\varepsilon_{l_i}$ is contained in U_{l_i} (at least close to the origin) by Proposition 2.3. Choose c and C to be the maximum and minimum over all l_i, the bounds in U_{l_i} are independent of l_i. Then the U_{l_i}'s do not intersect close to the origin, due to different exponents l_i. Let $U := \bigcup U_{l_i}$, we have $U \supset W^\varepsilon$. Hence $F(x(s))$ is increasing in $U^c \subset (W^\varepsilon)^c \subset S_2^c$ to the origin, by Proposition 2.9. Since the U_{l_i} are disjoint close to the origin $F(x(s))$ is increasing on $\bigcup U^c_{l_i}$ and in particular on $U^c_{l_i}$ and ∂U_{l_i}. Since f is negative, we define $\partial^+ U_{l_i} := \{x : F(x) = -C\}$ and $\partial^- U_{l_i} := \{x : F(x) = -c\}$. On $\partial^+ U_{l_i}$ the trajectory crosses into U_{l_i} and on $\partial^- U_{l_i}$ it will increase further, i.e. once the trajectory leaves U_{l_i}, there is no chance of coming back to U_{l_i} ($F(x)$ was increasing). Yet if it leaves it can not pass through $W^\varepsilon_{l_i} \subset U_{l_i}$ in any neighborhood of the origin, hence it has to stay in U_{l_i} indefinitely. This in turn then implies the uniqueness of W^ε_l. □

Corollary 2.11. *The trajectory $x(s)$ only intersects W^ε_l and never $W^\varepsilon_{l_i}$ for $l \neq l_i$ close to the origin, since $W^\varepsilon_{l_i} \subset U_{l_i}$ and the U_{l_i}'s were disjoint.*

From now on $F := \frac{f}{r^l}$ with l the unique l_i from Proposition 2.10 and fix $\varepsilon \leq \frac{1}{2}c_f(1 - \rho_f) \leq \frac{1}{2}c_f/l$.

Lemma 2.12. *On W^ε_l, we have $\partial_r f < 0$.*

Proof. This follows essentially from Corollary 2.11 and the proof of Proposition 2.9, as $f\partial_r f > 0$. □

Lemma 2.13. [KMP99, Remark 6.3] *There exist an exponent $\omega > 0$ such that for any $l \in L$ on W^ε_l, we have on W^ε_l*

$$|1 - \frac{lf}{r\partial_r f}| \leq r^{2\omega}.$$

Proof. Since ε is fixed, for any $l \in L$ there exist constants δ such that $|1 - \frac{lf}{r\partial_r f}| \leq r^{2\omega}$. Since L is finite the result follows. □

Lemma 2.14. [KMP99, Remark 6.3] *On $U_l \backslash W^\varepsilon_l$ we have*

$$\frac{dF(x(s))}{ds} \geq \frac{|\nabla f|}{2r^l} \geq \frac{c_f|f|}{2r^{l+1}} \geq \frac{C}{r}.$$

Proof. Since the trajectory $x(s)$ does not intersect any other $W^\varepsilon_{l_i}$ except when $l = l_i$, $\frac{dF(x(s))}{ds} \geq 0$ on $U_l \backslash W^\varepsilon_l$ by Proposition 2.10 and Proposition 2.9. Thus on $U_l \backslash W^\varepsilon_l \subset (W^\varepsilon)^c$,

$$\frac{c_f}{2l}|\nabla f| \geq \varepsilon|\nabla f| \geq \varepsilon|\nabla' f| \geq |\partial_r f| \iff |\nabla f| \geq \frac{2l}{c_f}|\partial_r f|,$$

and together with Theorem 1.10,

$$r|\nabla f|^2 \geq c_f|f||\nabla f| \geq c_f|f|\frac{2l}{c_f}|\partial_r f| = 2lf\partial_r f.$$

Hence once again Theorem 1.10 and the fact that on U_l $cr^l < |f|$ gives

$$\frac{dF(x(s))}{ds} = \frac{r|\nabla f|^2 - lf\partial_r f}{|\nabla f|r^{l+1}} \geq \frac{r|\nabla f|^2}{2|\nabla f|r^{l+1}} \geq \frac{|\nabla f|}{2r^l} \geq \frac{c_f|f|}{2r^{l+1}} \geq \frac{c_f cr^l}{2r^{l+1}} = \frac{C}{r}.$$

□

Finally we can show that not only f, but also F has a limiting value at the origin (0 for f) when approached along the trajectory.

Proposition 2.15. [KMP99, Proposition 6.4] *For $\alpha < 2\omega$, the function $g := F - r^\alpha$ is strictly increasing on $x(s)$. Moreover,*

$$F(x(s)) \to a_0 < 0, \qquad \text{as } s \to s_0,$$

and a_0 is an asymptotic critical value of F at the origin.

Proof. Again we consider three cases and follow the exact same steps as in [KMP99]:

1. $r^{-\omega}|\nabla' f| \le |\partial_r f|$: We are in $W_{-\omega,l} \subset W_l^\varepsilon$ close to the origin, so from Proposition 2.9, Lemma 2.12 and the equivalences ($|\partial_r f| \sim r^{l-1}$),

$$\left|\frac{dF(x(s))}{ds}\right| \le \frac{1}{|\nabla f| r^l}\left(|\nabla' f|^2 + |\partial_r f|^2 r^{2\omega}\right) \le \frac{C}{r^{l-1} r^l}\left(|\partial_r f|^2 r^{2\omega} + |\partial_r f|^2 r^{2\omega}\right) \le C r^{2\omega - 1},$$

$$\frac{d(-r^\alpha)}{ds} = -\alpha r^{\alpha-1}\frac{\partial_r f}{|\nabla f|} = -\alpha r^{\alpha-1}\frac{\partial_r f}{|\partial_r f| + |\nabla' f|} \ge -\alpha r^{\alpha-1}\frac{\partial_r f}{|\partial_r f| + r^\omega |\partial_r f|} \ge \frac{\alpha}{2} r^{\alpha-1}.$$

Since $\alpha < 2\omega$, $\frac{d(-r^\alpha)}{ds} \gg \frac{dF(x(s))}{ds}$ and so $g = F - r^\alpha$ is increasing on the trajectory.

2. $\varepsilon|\nabla' f| \le |\partial_r f| < r^{-\omega}|\nabla' f|$: Suppose $\frac{dF(x(s))}{ds} \le 0$, then

$$|\nabla' f|^2 - r^{2\omega}|\partial_r f|^2 \le 0 \implies |\nabla' f| r^{-w} \le |\partial_r f|,$$

which contradicts $|\partial_r f| < r^{-\omega}|\nabla' f|$. Also, since $\partial_r f(x(s)) < 0$ for $x(s) \in W_l^\varepsilon$ by Lemma 2.12,

$$\frac{d(-r^\alpha)}{ds} = -\alpha r^{\alpha-1}\frac{\partial_r f}{|\nabla f|} \ge 0,$$

so F and $-r^\alpha$ are increasing and so is g.

3. $|\partial_r f| < \varepsilon|\nabla' f|$: We are in $U_l \backslash W_l^\varepsilon$ and by Lemma 2.14 for r sufficiently small,

$$\frac{dg(x(s))}{ds} = \frac{dF}{ds} + \frac{d(-r^\alpha)}{ds} \ge Cr^{-1} - \left|\frac{dr^\alpha}{ds}\right| = Cr^{-1} - \left|\alpha r^{\alpha-1}\frac{|\partial_r f|}{|\nabla f|}\right| \ge C(r^{-1} - r^{\alpha-1}) \ge 0.$$

Since $x(s) \in U_l$, $F(x(s))$ and so g are bounded away from zero. Also $g(x(s))$ is negative and increasing and thus has the limit $a_0 < 0$ ($F(x(s))$ has same limit). Suppose now that a_0 is not an asymptotic critical value. Then by Proposition 2.7, there exists C such that for s close to s_0

$$|\nabla' f(x(s))| \ge C|\partial_r f(x(s))|.$$

On W_l^ε we have

$$\frac{dF(x(s))}{ds} = \frac{1}{r^l|\nabla f|}\left(|\nabla' f|^2 + |\partial_r f|^2\left(1 - \frac{lf}{r\partial_r f}\right)\right) = \frac{|\partial_r f|}{r^l|\nabla f|}\left(1 + C^2 - \frac{lf}{r\partial_r f}\right) \ge C'\frac{1}{r},$$

the same bound holds on $U_l \backslash W_l^\varepsilon$ by Lemma 2.14, thus $\frac{dF}{ds} \ge \frac{C}{r}$ holds on U_l and thereby for any s sufficiently close to s_0 (so that $x(s) \in U_l$). Note that since s is the arc-length parameter, $r \le s_0 - s$ and so

$$\frac{dF(x(s))}{ds} \ge \frac{C}{s_0 - s} \implies \int_s^{s_0} \frac{C}{s_0 - s} \le \int_s^{s_0} \frac{dF(x(s))}{ds} = \lim_{\tau \to s_0} F(x(\tau)) - F(x(s)) = a_0 - F(x(s)).$$

Yet the RHS is bounded (by existence of the limit) whereas the LHS is not, thus by contradiction, a_0 is an asymptotic critical value. □

The following Corollary will be studied in detail in a later chapter.

Corollary 2.16. ("length-distance convergence"), [KMP99, Corollary 6.5] *Let $\sigma(s)$ be the length of the trajectory from $x(s)$ to the origin, i.e. $\sigma(s) := \int_s^{s_0} |x'(s)| ds$. Then as $s \to s_0$*

$$\frac{\sigma(s)}{|x(s)|} \to 1.$$

2.4 Proof of main theorem

First we reparametrize again; consider the projection onto the unit sphere $\tilde{x}(s) = \frac{x(s)}{|x(s)|}$. This is a curve on $\mathbb{S}^{n-1}$, which has an arclength parameter $\tilde{s}$. We rewrite $x(s) = x(s(\tilde{s}))$, hence

$$\left|\frac{d}{d\tilde{s}}\tilde{x}(s(\tilde{s}))\right| = \left|\frac{d}{ds}\tilde{x}(s)\frac{ds}{d\tilde{s}}\right| = 1.$$

The reparametrization is possible, since $\tilde{s} \neq 0$. If $\tilde{s} = 0$, then the analytic trajectory is a straight line for some part and stays a line (as all derivatives vanish), the gradient conjecture holds trivially.

So far we have a unique asymptotic critical value $a < 0$ of $F = \frac{f}{r^l}$ for some unique $l \in L$ and thus $F(x(s)) \to a$ as $s \to s_0$. Hence we can assume that

$$x(s) \in U_{l,a} := \{x \in U_l : |F(x) - a| \leq c\} \qquad \text{for any } c > 0.$$

In particular, we can pick $c = c_a$ from Proposition 2.6, if we can somehow bound $|\partial_r F|$ by the gradient of F, we are in the setting of Proposition 2.6 and get a Lojasiewicz-type inequality. Conversely, intuitively if $|\partial_r F|$ is large, the trajectory approaches the origin in a straight manner.

For the exact proof, we follow a Lojasiewicz type argument; construct a *control function* g defined in a neighborhood around the trajectory, satisfying certain growth properties. Here, this neighborhood will be split into four parts where the control function will be estimated separately. We will follow the outline in [KMP99], but provide a detailed calculation of the Lojasiewicz argument (Proposition 2.18) and do one extra stratification in Proposition 2.20, to demonstrate all details of the estimates.

Definition 2.17. From Remark 2.13, there exists $\omega > 0$ such that $|1 - \frac{lf}{r\partial_r f}| \leq \frac{1}{2}r^{2w}$. Let now $\eta < \omega$ and $\rho(\eta) < 1$ from Proposition 2.6.

Proposition 2.18. (Lojasiewicz argument) *Let g be a suitable control function defined in a neighborhood of the trajectory (here in $U_{l,a}$), that is on the trajectory we have*

$$\frac{d}{d\tilde{s}}(g(x(\tilde{s}))) \geq c|g(x(\tilde{s}))|^{\xi},$$

with $0 < \xi < 1$ and $c > 0$. Then the length of $\tilde{x}(s)$ is finite.

Proof. Start by calculating (differentiate the absolute value away from zero)

$$\begin{aligned}
\frac{d}{d\tilde{s}}(|g(x)|^{1-\xi}) &= (1-\xi)|g(x(\tilde{s}))|^{-\xi-1}g(x(\tilde{s})) \cdot \frac{d}{d\tilde{s}}g(x(\tilde{s})) \\
&= (1-\xi)\frac{g(x(\tilde{s}))}{|g(x(\tilde{s}))|}|g(x(\tilde{s}))|^{-\xi} \cdot \frac{d}{d\tilde{s}}g(x(\tilde{s})) \\
&= -(1-\xi)|g(x(\tilde{s}))|^{-\xi} \cdot \frac{d}{d\tilde{s}}g(x(\tilde{s})) \\
&\leq -(1-\xi)|g(x(\tilde{s}))|^{-\xi} \cdot c|g(x(\tilde{s}))|^{\xi} \\
&= -(1-\xi)c < 0 \\
&\implies \frac{-1}{(1-\xi)c}\frac{d}{d\tilde{s}}(|g(x)|^{1-\xi}) \geq 1.
\end{aligned}$$

Let $s_1 < s_0$ be such that $x(s) \in U_{l,a}$ for any $s_1 \leq s < s_0$, which is known to exist by Proposition 2.10 and 2.15. We can now calculate

$$\begin{aligned}
&\ell = \text{length}_{s_1 \leq s < s_0}\{\tilde{x}(s)\} = \tilde{s}(s_0) - \tilde{s}(s_1) = \int_{s_1}^{s_0}\frac{d\tilde{s}}{ds}ds \leq \int_{s_1}^{s_0}\frac{-1}{(1-\xi)c}\frac{d|g|^{1-\xi}}{d\tilde{s}}\frac{d\tilde{s}}{ds}ds \\
&= \frac{1}{c(1-\xi)}\int_{s_0}^{s_1}\frac{d|g|^{1-\xi}}{ds}ds = \frac{1}{c(1-\xi)}(|g(x(s_1))|^{1-\xi} - |g(x(s_0))|^{1-\xi}) \leq \frac{1}{c(1-\xi)}|g(x(s_1))|^{1-\xi},
\end{aligned}$$

which is clearly finite. □

It remains to show existence of such a g, satisfying the necessary conditions of a control function. Here we will consider for $\alpha > 0$ sufficiently small (in fact $\alpha < \eta < \omega$),

$$g := F - a - r^{\alpha}.$$

Lemma 2.19. [KMP99, Lemma 7.2] *For any $\eta < 0$ sufficiently small, $F - a$ is a control function on the set $\{x \in U_{l,a} : |\partial_r f| \leq r^{-\eta}|\nabla' f|\}$, i.e. for some $\rho = \rho(\eta) < 1$ and $c > 0$,*

$$\frac{d(F(x(s)) - a)}{d\tilde{s}} \geq c|F(x(s)) - a|^{\rho}.$$

Proof. If $|\partial_r f| < \varepsilon|\nabla' f|$, then by Lemma 2.14

$$\frac{dF}{d\tilde{s}} \geq \frac{C}{r}\frac{ds}{d\tilde{s}} = \frac{C}{r}\frac{r|\nabla f|}{|\nabla' f|} \geq C > 0.$$

So let $\varepsilon|\nabla' f| \leq |\partial_r f| \leq r^{-\eta}|\nabla' f|$, that is we are inside W^{ε} and the equivalences $|\nabla f| \sim |\partial_r f| \sim r^{l-1}$ hold. Then

$$\begin{aligned}
r|\nabla' F| &= \frac{|\nabla' f|}{r^{l-1}} \geq \frac{|\nabla' f|}{C|\partial_r f|} \geq Cr^{\eta} \\
r^{1-\eta}|\partial_r F| &= r^{1-\eta}\left|\frac{\partial_r f r^l - l r^{l-1} f}{r^{2l}}\right| = r^{-\eta}\left|\frac{\partial_r f}{r^{l-1}}\left(1 - \frac{lf}{r\partial_r f}\right)\right| \leq c r^{2\omega - \eta} \\
\eta < \omega &\implies r|\nabla' F| \geq Cr^{\eta} \geq cr^{2\omega-\eta} \geq r^{1-\eta}|\partial_r F| \iff |\partial_r F| \leq r^{\eta}|\nabla F| \quad \text{for } r \text{ small.}
\end{aligned}$$

It then follows from Proposition 2.6 that $r|\nabla F| \geq c|F - a|^{\rho}$, so

$$\begin{aligned}
\frac{dF}{d\tilde{s}} &= \frac{|\nabla' f|}{r^{l-1}} + \frac{|\partial_r f|^2}{r^{l-1}|\nabla' f|}\left(1 - \frac{lf}{r\partial_r f}\right) \geq \frac{|\nabla' f|}{r^{l-1}} - \frac{|\partial_r f|^2}{r^{l-1}|\nabla' f|}\left|1 - \frac{lf}{r\partial_r f}\right| \geq r|\nabla' F| - Cr^{2\omega-\eta} \\
&\geq r|\nabla' F| - Cr^{2\omega-\eta} \geq r|\nabla' F| - \frac{1}{2}r|\nabla' F| = \frac{1}{4}r|\nabla' F| + \frac{1}{4}r|\nabla' F| \geq \frac{1}{4}r|\nabla' F| + \frac{1}{4}r^{1-\eta}|\partial_r F| \\
&\geq \frac{1}{4}r|\nabla' F| + \frac{1}{4}r|\partial_r F| = \frac{1}{4}r|\nabla F| \geq c|F - a|^{\rho},
\end{aligned}$$

where we use that $Cr^{2\omega-\eta} \leq C' r^{\omega} \leq \frac{1}{2}r|\nabla' F|$ since $0 < \eta < \omega$. □

Proposition 2.20. [KMP99, Lemma 7.3, Theorem 7.1] *For $\alpha > 0$ sufficiently small, $g = F - a - r^{\alpha}$ is a control function on $U_{l,a}$, i.e. for some $\xi < 1$ and $c > 0$,*

$$\frac{d}{d\tilde{s}}g(x(\tilde{s})) \geq c|g(x(\tilde{s}))|^{\xi}.$$

Proof. Note that from the arc-length parametrization $\tilde{s}$ we have

$$\begin{aligned}
\frac{d}{ds}\tilde{x}(s(\tilde{s})) &= \frac{d}{ds}\frac{x(s)}{|x(s)|} = \frac{1}{|x(s)|}\frac{\nabla f(x)}{|\nabla f(x)|} - \frac{x(s)\cdot\nabla f(x)}{|x(s)|^3|\nabla f(x)|}x(s) = \frac{r^2\nabla f(x) - r\partial_r f(x) x(s)}{r^3|\nabla f(x)|} \\
&= \frac{\nabla f(x) - \partial_r f(x)\frac{x(s)}{r}}{r|\nabla f(x)|} = \frac{\nabla f(x) - \partial_r f(x)\partial_r}{r|\nabla f(x)|} = \frac{\nabla' f(x)}{r|\nabla f(x)|}
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{ds}{d\tilde{s}} &= \left|\frac{d}{ds}\tilde{x}(s(\tilde{s}))\right|^{-1} = \frac{r|\nabla f|}{|\nabla' f|} \\
\frac{dF}{d\tilde{s}} &= \frac{dF}{ds}\frac{r|\nabla f|}{|\nabla' f|} = \frac{1}{r^l|\nabla f|}\left(|\nabla' f|^2 + |\partial_r f|^2\left(1 - \frac{lf}{r\partial_r f}\right)\right)\frac{r|\nabla f|}{|\nabla' f|} \\
&= \frac{1}{|\nabla' f| r^{l-1}}\left(|\nabla' f|^2 + |\partial_r f|^2\left(1 - \frac{lf}{r\partial_r f}\right)\right) \\
\frac{d}{d\tilde{s}}(-r^{\alpha}) &= -\alpha r^{\alpha-1}\frac{dr}{ds}\frac{ds}{d\tilde{s}} = -\alpha r^{\alpha-1}\left(\frac{x(s)\cdot x'(s)}{|x(s)|}\frac{r|\nabla f|}{|\nabla' f|}\right) = -\alpha r^{\alpha-1}\left(\frac{r\partial_r f}{r|\nabla f|}\frac{r|\nabla f|}{|\nabla' f|}\right) = -\alpha r^{\alpha}\frac{\partial_r f}{|\nabla' f|}
\end{aligned}$$

First let ω be the exponent given by Definition 2.17. Take now $\alpha, \delta, \eta > 0$ such that $\delta < \alpha < \eta < \omega$. Define $W_{-\eta,l} := \{x \in W_l^{\varepsilon} : r^{-\eta}|\nabla' f| \leq |\partial_r f|\}$. We have the following chain of inclusions,

$$W_{-\eta,l} \subset W_{-\delta,l} \subset W_l^{\varepsilon} \subset U_{l,a}.$$

The proof splits in four cases:

1. On $U_{l,a}\backslash W_l^\varepsilon$: Since on $(W_l^\varepsilon)^c$ we have $\frac{|\nabla' f|}{|\partial_r f|} \geq \frac{1}{\varepsilon}$. Thus using Lemma 2.14,

$$\frac{dg}{d\tilde{s}} = \frac{dF}{d\tilde{s}} + \frac{d(-r^\alpha)}{d\tilde{s}} = \frac{dF}{ds}\frac{r|\nabla f|}{|\nabla' f|} + \alpha r^\alpha \frac{|\partial_r f|}{|\nabla' f|} \geq \frac{C}{r}\frac{r|\nabla f|}{|\nabla' f|} \geq C \geq |g|^\xi,$$

because $g(x(\tilde{s})) \to 0$, g is a control function for $c = 1$ and any $\xi \geq 0$.

2. On $W_l^\varepsilon\backslash W_{-\delta,l}$: By Lemma 2.19, on $(W_{-\delta,l})^c$ we have some $\rho_1 = \rho_1(\delta)$ such that

$$\frac{d(F(x(s)) - a)}{d\tilde{s}} \geq c|F(x(s)) - a|^{\rho_1}.$$

Moreover from the proof of the same lemma,

$$\frac{dF}{d\tilde{s}} \geq \frac{1}{2}r|\nabla F| \geq \frac{1}{2}r|\nabla' F| = \frac{1}{2}\frac{|\nabla' f|}{r^{l-1}} \geq \frac{r^\delta|\partial_r f|}{r^{l-1}} \geq c' r^\delta.$$

Thus letting $\rho_2 = 1 - \frac{\delta}{\alpha}$, $\rho = \max\{\rho_1, \rho_2\}$,

$$\begin{aligned}\frac{dg}{d\tilde{s}} &= \frac{dF}{d\tilde{s}} + \frac{d(-r^\alpha)}{d\tilde{s}} = \frac{dF}{d\tilde{s}} + \alpha r^\alpha\frac{|\partial_r f|}{|\nabla' f|} \geq \frac{1}{2}\left(\frac{dF}{d\tilde{s}} + \frac{dF}{d\tilde{s}}\right) \geq c|F(x(s)) - a|^{\rho_1} + c' r^\delta \\ &\geq c|F(x(s)) - a|^\rho + c'(r^\alpha)^\rho \geq C\left(|F(x(s)) - a|^\rho + |r^\alpha|^\rho\right) \\ &\geq C\left(|F(x(s)) - a| + |r^\alpha|\right)^\rho \geq C|F(x(s)) - a - r^\alpha|^\rho \\ &= C|g|^\rho,\end{aligned}$$

so g is a control function with $\xi_2 := \rho < 1$.

3. On $W_{-\delta,l}\backslash W_{-\eta,l}$: By Lemma 2.19, on $(W_{-\eta,l})^c$ we have some $\rho_1 = \rho_1(\eta)$ such that

$$\frac{d(F(x(s)) - a)}{d\tilde{s}} \geq c|F(x(s)) - a|^{\rho_1}.$$

Moreover on $W_{-\delta,l}$ we have

$$\frac{d(-r^\alpha)}{d\tilde{s}} = -\alpha r^\alpha\frac{\partial_r f}{|\nabla' f|} \geq \alpha r^{\alpha-\delta} \geq \alpha(r^\alpha)^{\rho_2} \qquad \text{for } \rho_2 := 1 - \frac{\delta}{\alpha} < 1.$$

Let $\rho = \max\{\rho_1, \rho_2\}$, then on $W_{-\delta,l}\backslash W_{-\eta,l}$,

$$\begin{aligned}\frac{dg}{d\tilde{s}} &= \frac{dF}{d\tilde{s}} + \frac{d(-r^\alpha)}{\tilde{s}} \geq c|F(x(s)) - a|^{\rho_1} + \alpha(r^\alpha)^{\rho_2} \geq C\left(|F(x(s)) - a|^\rho + |r^\alpha|^\rho\right) \\ &\geq C\left(|F(x(s)) - a| + |r^\alpha|\right)^\rho \geq C|F(x(s)) - a - r^\alpha|^\rho = C|g|^\rho,\end{aligned}$$

g is a control function with $\xi_3 = \rho < 1$.

4. On $W_{-\eta,l} \subset W_l^\varepsilon$: Now we finally have control over the spherical part of the gradient. We estimate, using the bound on $W_{-\eta,l}$ and the fact that $|\partial_r f| \sim r^{l-1}$,

$$\frac{d(-r^\alpha)}{d\tilde{s}} = \alpha r^\alpha\frac{|\partial_r f|}{|\nabla' f|}(\geq \alpha r^{\alpha-\eta} \geq c > 0),$$

$$\begin{aligned}\frac{dF}{d\tilde{s}} &= \frac{|\nabla' f|}{r^{l-1}} + \frac{|\partial_r f|}{r^{l-1}}\frac{|\partial_r f|}{|\nabla' f|}\left(1 - \frac{lf}{r\partial_r f}\right) = \frac{|\nabla' f|}{r^{l-1}} + \frac{|\partial_r f|}{r^{l-1}}\frac{|\partial_r f|}{|\nabla' f|}O(r^{2\omega}) \\ &= \frac{|\nabla' f|}{r^{l-1}} + \frac{|\partial_r f|}{r^{l-1}}\frac{|\partial_r f|}{|\nabla' f|}O(r^{2\omega}) \geq \frac{|\nabla' f|}{C|\partial_r f|} + c\frac{|\partial_r f|}{|\nabla' f|}O(r^{2\omega}) \geq O(r^\eta) + \frac{|\partial_r f|}{|\nabla' f|}O(r^{2\omega}).\end{aligned}$$

Hence we conclude

$$\begin{aligned}\frac{dg}{d\tilde{s}} &= \frac{dF}{d\tilde{s}} + \frac{d(-r^\alpha)}{d\tilde{s}} \geq O(r^\eta) + \frac{|\partial_r f|}{|\nabla' f|}(\alpha r^\alpha + O(r^{2\omega})) \geq O(r^\eta) + r^{-\eta}(\alpha r^\alpha + O(r^{2\omega})) \\ &= O(r^\eta) + \alpha r^{\alpha-\eta} + O(r^{2\omega-\eta}) \geq c_0 > 0,\end{aligned}$$

since $\alpha \leq \eta$ and $\omega > \eta$ and surely $\frac{dg}{d\tilde{s}} \geq c|g|^\xi$ holds. Actually for $\alpha < \eta$, $\frac{dg}{d\tilde{s}}$ would even blow up, something way stronger than $\frac{dg}{d\tilde{s}} \geq c|g|^\xi$.

The exponents ξ_2 and ξ_3 respectively may be different in each region, yet taking the maximum gives and exponent ξ that holds everywhere on $U_{l,a}$. $\square$

Thereby we can conclude Propositions 2.18 and 2.20 into the following, which implies Theorem 2.1.

Theorem 2.21. [KMP99, Theorem 7.1] *Let $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$, converging to x^* as $s \to s_0$. Then $\tilde{x}(s) = \frac{x(s)-x^*}{|x(s)-x^*|}$ is of finite length.*

3 Gradient conjectures in comparison

Here we compare explicitly different limiting behaviours of the trajectory, only mentioned briefly in [KMP99] and similar papers. So far we have:

Theorem 3.1. [KMP99, Theorem 7.1] *Let $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$, converging to x^* as $s \to s_0$. Then $\tilde{x}(s) = \frac{x(s)-x^*}{|x(s)-x^*|}$ is of finite length.*

Corollary 3.2. (Gradient Conjecture) *Let $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$ and $x(s) \to x^*$ as $s \to s_0$. Then the limit of secants, $\lim_{s\to s_0} \frac{x(s)-x^*}{|x(s)-x^*|}$ exists and so $x(s)$ has a tangent at x^*.*

Proof. Assume the limit does not exist. Then for some $\delta > 0$, $\tilde{x}(s)$ does not stay inside a δ-neighborhood around any point on the unit sphere. In other words, for any $s_1 < s_0$, there exists $s_1 < s_2 < s_0$ such that $|\tilde{x}(s_1) - \tilde{x}(s_2)| \geq \delta$. Thus the length of the curve $\tilde{x}(s)$ can be bounded below,

$$\text{length}_{s_1 \leq s < s_0}\{\tilde{x}(s)\} \geq \delta + \text{length}_{s_2 \leq s < s_0}\{\tilde{x}(s)\} \geq \delta + \delta + \text{length}_{s_3 \leq s < s_0}\{\tilde{x}(s)\} = \sum_{i=1}^{\infty} \delta = \infty$$

since the same applies for s_2 and s_3. Thus its length is infinite, a contradiction to Theorem 1.11. $\square$

Conjecture 3.3. (Strong Gradient Conjecture) *Let f be analytic and $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$, converging to the origin as $s \to s_0$. Then $\lim_{s\to s_0} x'(s)$ (the unit tangents to the curve) exists.*

Proposition 3.4. *The strong gradient conjecture implies the gradient conjecture, i.e. existence of the limit of unit tangents implies existence of the limit of unit secants and the limits are the same.*

Proof. Let $v := \lim x'(s)$, then $x'(s) - v =: \eta(s)$ where $\eta(s) \to 0$ as $s \to s_0$. Since $|v| = 1$,

$$\begin{aligned}
\lim_{s\to s_0} \left| \frac{x(s)}{|x(s)|} - v \right| &= \lim_{s\to s_0} \left| \frac{\int_s^{s_0} x'(\tau)d\tau}{\left|\int_s^{s_0} x'(\tau)d\tau\right|} - v \right| = \lim_{s\to s_0} \left| \frac{\int_s^{s_0} v + \eta(\tau)d\tau}{\left|\int_s^{s_0} v + \eta(\tau)d\tau\right|} - v \right| \\
&= \lim_{s\to s_0} \left| \frac{v(s_0 - s) + \int_s^{s_0} \eta(\tau)d\tau}{\left|v(s_0 - s) + \int_s^{s_0} \eta(\tau)d\tau\right|} - v \right| \\
&= \left| \frac{\lim_{s\to s_0} v(s_0 - s) + \lim_{s\to s_0} \int_s^{s_0} \eta(\tau)d\tau}{\left|\lim_{s\to s_0} v(s_0 - s) + \lim_{s\to s_0} \int_s^{s_0} \eta(\tau)d\tau\right|} - v \right| \\
&= \left| \lim_{s\to s_0} \frac{v(s_0 - s)}{|v(s_0 - s)|} - v \right| = \left| \frac{v}{|v|} - v \right| = 0,
\end{aligned}$$

and the gradient conjecture follows. $\square$

The next corollary gives a relation between the distance to the origin and the remaining arc-length. It can be proven directly from 2.2. We will also provide a second "proof" from the strong gradient conjecture. It should also be possible to make a geometric proof from the Gradient Conjecture, using the convergence of secants and the bound on the curvature, Lemma 1.14. One might think the result would prevent the trajectory from spiraling, since intuitively the "additional length" that can make the trajectory behave wildly goes to zero. However this is not the case as shown in the next subsection.

Corollary 3.5. ("length-distance convergence"), [KMP99, Corollary 6.5] *Let $x(s)$ be a trajectory of $\frac{\nabla f}{|\nabla f|}$ and $x(s) \to 0$ as $s \to s_0$. Denote by $\sigma(s)$ its arc-length to the origin, then*

$$\frac{\sigma(s)}{|x(s)|} \to 1 \qquad \text{as } s \to s_0.$$

Proof. (Directly from 2.2 as done in [KMP99])
Define $U_{l,\tilde{c},\tilde{C}} := \{x : 0 < \tilde{c} < \frac{|f|}{r^l} < \tilde{C} < \infty\}$. Let $\gamma(r)$ be a real analytic curve, parameterized by its distance to the origin with $\gamma(r) \to 0$. Thus from the Puiseux expansion (see Lemma 2.2) of $f(\gamma(r))$,

$$f(\gamma(r)) = a_{l_1} r^{l_1} + a_{l_2} r^{l_2} + \dots,$$

$$|\nabla f| = |\partial_r f| + |\nabla' f| \geq |\partial_r f| = |l_1 a_{l_1} r^{l_1-1} + l_2 a_{l_2} r^{l_2-1} + \dots| = l_1 |a_{l_1}| r^{l_1-1} + O(r^{l_2-1}),$$

$$\frac{l_1|f|}{r} = |l_1 a_{l_1} r^{l_1-1} + l_1 a_{l_2} r^{l_2-1} + \dots| = |l_1| a_{l_1} r^{l_1-1} + O(r^{l_2-1}).$$

$$\implies |\nabla f| \geq \frac{l_1|f|}{r} - o(r^{l_1-1})$$

along the curve $\gamma(r)$, and in the same way as in Lemma 2.2, this estimate holds on $U_{l,\tilde{c},\tilde{C}}$. Now for the curve $x(s) = x_r(s(r))$ (reparametrized) we want to show $l_1 = l$. Since $x_r(r) \in U_{l,\tilde{c},\tilde{C}}$,

$$\tilde{c} < \frac{|f(x_r(r))|}{r^l} = \frac{b_{l_1} r^{l_1} + O(r^{l_2})}{r^l} < \tilde{C} \implies l_1 = l.$$

Thus along $x(s)$,

$$\frac{df}{ds} = \left\langle \nabla f(x(s)), \frac{\nabla f(x(s))}{|\nabla f(x(s))|} \right\rangle = |\nabla f| \geq \frac{l|f|}{r} - o(r^{l-1}) \geq l\tilde{c}^{1/l} |f|^{\frac{l-1}{l}} - o(r^{l-1}).$$

Now

$$\frac{d|f|^{1-\frac{l-1}{l}}}{ds} = \left(1 - \frac{l-1}{l}\right) |f|^{\left(-\frac{l-1}{l}-1\right)} f \frac{df}{ds} \leq \left(1 - \frac{l-1}{l}\right) |f|^{\left(-\frac{l-1}{l}-1\right)} f \left(l\tilde{c}^{1/l}|f|^{\frac{l-1}{l}} - o(r^{l-1})\right)$$

$$= -\left(1 - \frac{l-1}{l}\right) l\tilde{c}^{\frac{1}{l}} + o(r^{l-1})|f|^{-\frac{l-1}{l}} = -c_0 + o(r^{l-1})O(r^l) = -c_0 + o(1),$$

$$\implies 1 \leq \frac{1}{-c_0 + o(1)} \frac{d|f|^{1-\frac{l-1}{l}}}{ds},$$

where in the first inequality we use the fact that f is negative and define $c_0 := l\tilde{c}^{1/l}(1 - \frac{l-1}{l}) = \tilde{c}^{1/l}$. (The inequality flips since we divide by something negative.) Then, following Lojasiewicz argument,

$$\sigma(\bar{s}) = \int_{\bar{s}}^{s_0} ds \leq \int_{\bar{s}}^{s_0} \frac{-1}{c_0 + o(1)} \frac{d|f|^{1-\frac{l-1}{l}}}{ds} \leq \frac{1}{c_0 + o(1)} |f|^{1-\frac{l-1}{l}} = \frac{1}{\tilde{c}^{1/l} + o(1)} |f|^{1/l} \leq \frac{\tilde{C}^{1/l} r}{\tilde{c}^{1/l} + o(1)}.$$

Since $F(x(s)) = \frac{f(x(s))}{r^l} \to a_0$ as $s \to s_0$, we can choose $\tilde{c}, \tilde{C}$ arbitrarily close to a_0, the asymptotic critical value, e.g. $a_0 \pm \varepsilon$. Then

$$\frac{\sigma(s)}{|x(s)|} = \frac{\sigma(s)}{r} \leq \frac{\frac{\tilde{C}^{1/l} r}{\tilde{c}^{1/l} + o(1)}}{r} = \frac{\tilde{C}^{1/l}}{\tilde{c}^{1/l} + o(1)} \to 1 \qquad \text{as } s \to s_0 \text{ and } r = |x(s)| \to 0.$$

Note that there are two terms converging separately, once $F(x(s)) \to a_0$ and secondly $|x(s)| \to 0$, so in order to obtain a quantitative rate of convergence, both terms need to be taken into consideration. □

Proof. (from the Strong Gradient Conjecture)
Since $\lim_{s \to s_0} \frac{x'(s)}{|x'(s)|} =: v$, for all s with $|s - s_0| < \delta$, $\|v - \frac{x'(s)}{|x'(s)|}\| < \varepsilon$, thus close to the limit point, the unit tangent of the trajectory has to lie in a small disk on $\mathbb{S}^{n-1}$. Equivalently we can define $\alpha(s)$ as the angle between $\frac{x'(s)}{|x'(s)|}$ and v. Note that $\alpha(s) \to 0$ as $s \to s_0$.

Let now C_s be the closed one sided cone centered at $x(s)$ with axis v and opening angle $\alpha(s)$, containing all possible direction of the trajectory after $x(s)$. Firstly, we claim that $x(s')$ lies within C_s whenever $s < s' \leq s_0$, in particular $0 \in C_s$. Suppose that $x(s') \notin C_s$, then there exists a $\bar{s}$ where $x(\bar{s})$ crosses ∂C_s (i.e. $B_r(x(\bar{s})) \cap C_s^c \neq \varnothing$ and $B_r(x(\bar{s})) \cap C_s \neq \varnothing$ holds for any $r > 0$). But then $\frac{x'(\bar{s})}{|x'(\bar{s})|} \notin C_s$, i.e. pointing outwards of C_s, a contradiction.

Secondly, we want to bound $\sigma(s)$, by constructing a curve $\gamma(\tau)$, that has the maximal possible length between $x(s)$ and x^*, without violating the condition $\gamma'(\tau) \in C_s$. Let s be fixed and consider the disk $D := \{w \in \mathbb{S}^{n-1} : \|w - v\| < \varepsilon\}$. Define now

$$d_1 := \arg\min_{w \in \partial D} \left\| w - \frac{x'(s)}{|x'(s)|} \right\| \qquad \text{and} \qquad d_2 := \arg\max_{w \in \partial D} \left\| w - \frac{x'(s)}{|x'(s)|} \right\|,$$

and $\gamma(\tau)$ as the unique piecewise linear function following first d_2 up to some point $y \in \partial C_s$ and then d_1 connecting $x(s)$ and the origin. Note that any other piecewise linear function that travels first on the cone and then connects to the origin has the same length. To estimate the length of γ, note that $x(s), y, 0$ span a triangle, where the hypotenuse $|x(s)|$ is known and the angle $\angle(y, x(s)) =: \beta < \alpha(s)$. Hence we can estimate (isosceles triangle, set $\beta = \alpha(s)$),

$$\sigma(s) \leq |x(s) - y| + |y| \leq \frac{|x(s)|}{\cos(\alpha(s))},$$

which in turn gives

$$1 \leq \frac{\sigma(s)}{|x(s)|} \leq \frac{|x(s)|}{\cos(\alpha(s))|x(s)|} = \frac{1}{\cos(\alpha(s))} \to 1 \qquad \text{as } \alpha(s) \to 0.$$

Any other curve connecting $x(s)$ and the origin with tangents inside C_s has smaller length. $\square$

Let $x(t)$ be the trajectory of $x'(t) = -\nabla f(x(t))$ for an analytic function f. It is well known that the length of $x(t)$ is finite. Moreover, we have the following chain of implications:

$$\text{Strong Gradient Conjecture } (\textit{not proven}) \implies \text{Gradient Conjecture} \implies \lim_{s \to s_0} \frac{\sigma(s)}{|x(s)|} \to 1.$$

However for $f \in C^\infty$ functions there might be cases where:

- $x(t)$ does not even converge (see Mexican hat function in [AMA05]).
- $x(t)$ does converge but its arclength is infinite. We conjecture that counterexamples could be based on e.g. the curve $\gamma(t) = \left(\frac{1}{t}\cos t, \frac{1}{t}\sin t\right)$.
- $x(t)$ does converge, its arclength is finite and even $\lim_{s \to s_0} \frac{\sigma(s)}{|x(s)|} \to 1$ holds, yet the gradient conjecture (limit of secants) does not hold (see next section).
- The gradient conjecture holds (limit of secants), but not the strong gradient conjecture (limit of tangents). We conjecture counterexamples could be based on e.g. the curve $\gamma(t) = \left(t, t^2 \sin(1/t)\right)$.

3.1 Counterexample

Here we provide a counterexample to the length-distance corollary and the gradient conjecture in general, inspired by the Mexican hat function [AMA05]. That is we construct a smooth function $f : \mathbb{R}^2 \to \mathbb{R}$, such that the gradient flow of $x'(t) = -\nabla f(x(t))$ starting from some x_0 converges to the critical point at the origin as $t \to \infty$. The trajectory is finite and the stronger condition $\frac{\sigma(s)}{|x(s)|} \to 1$ holds as well, where $\sigma(s)$ denotes the arclength from $x(s)$ to the origin. (There exist many curves with finite length where this convergence of arclength does not hold.) Yet the secants $\frac{x(t)}{|x(t)|}$ do not converge, in fact they wander endlessly on the unit circle.

The main idea is to construct a spiral curve $\gamma : (1, \infty) \to \mathbb{R}^2$, written in polar coordinates as $\gamma(t) = (r(t), \theta(t))$ satisfying the desired properties and then construct a function $f(r, \theta)$ such that a gradient flow of f starting at e.g. $x_0 = \gamma(1.1)$ follows the trajectory of γ. Define the curve γ as

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

We obtain the following results:

- $\gamma(t) \to 0$ as $t \to \infty$, since $r(t) = \frac{1}{t} \to 0$.
- γ has finite length: Since $\gamma'(t) = (x(t), y(t))' = (r(t)\cos\theta(t), r(t)\sin\theta(t))'$,

$$\begin{aligned}
\|\gamma'(t)\| &= \left((r'\cos(\theta) - r\sin(\theta)\theta')^2 + (r'\sin(\theta) + r\cos(\theta)\theta')^2\right)^{1/2} \\
&= \sqrt{r'^2\cos^2\theta - 2r'r\cos\theta\sin\theta\theta' + r^2\sin^2\theta\theta'^2 + r'^2\sin^2\theta + 2rr'\sin\theta\cos\theta\theta' + r^2\cos^2\theta\theta'^2} \\
&= \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2} = \sqrt{\frac{1}{t^4} + \frac{1}{t^4}\frac{1}{\log(t)}}
\end{aligned}$$

$$\int_1^\infty \|\gamma'(t)\| dt = \int_1^\infty \sqrt{\frac{1}{t^4} + \frac{1}{t^4}\frac{1}{\log(t)}} dt \leq C \int_1^\infty \frac{1}{t^2} < \infty.$$

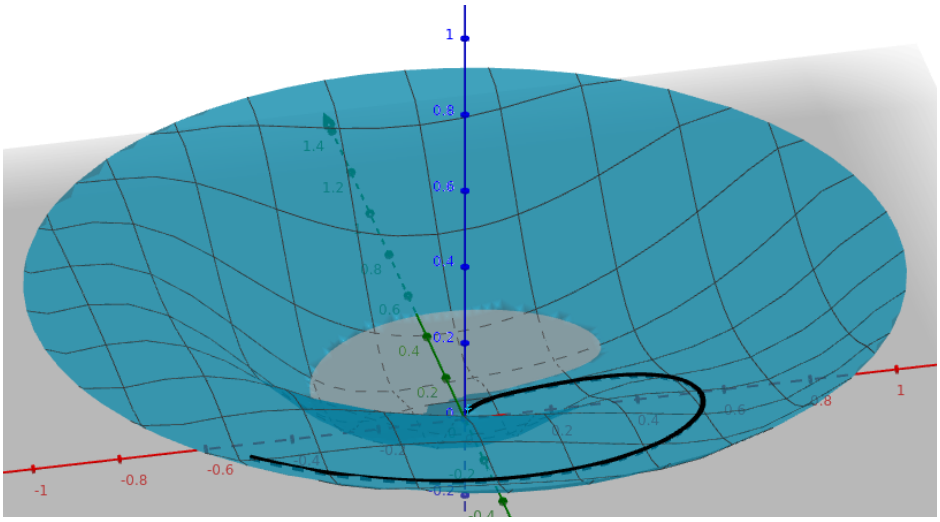

Figure 1: The function f and its gradient trajectory curve

- $\frac{\sigma(s)}{|\gamma(s)|} = \frac{\int_s^\infty \|\gamma'(t)\|dt}{|\gamma(s)|} \to 1$ as $s \to \infty$: For this, we estimate

$$\int_s^\infty \|\gamma'(t)\|dt = \int_s^\infty \sqrt{\frac{1}{t^4} + \frac{1}{t^4}\frac{1}{\log(t)}}dt \leq \int_s^\infty \frac{1}{t^2} + \int_s^\infty \frac{1}{t^2}\frac{1}{\log(t)}dt$$
$$\leq \frac{1}{s} + \frac{1}{\log(s)}\int_s^\infty \frac{1}{t^2}dt = \frac{1}{s} + \frac{1}{s}\frac{1}{\log(s)}$$
$$\implies \frac{\sigma(s)}{|\gamma(s)|} = \frac{\int_s^\infty \|\gamma'(t)\|dt}{|\gamma(s)|} \leq \frac{\frac{1}{s} + \frac{1}{s}\frac{1}{\log(s)}}{\frac{1}{s}} = 1 + \frac{1}{\log(s)} \to 1.$$

- However $\theta(t) = \log(\log(t)) \to \infty$ as $t \to \infty$.

Then we define the function $f : B_1 = \{x \in \mathbb{R}^n : |x| < 1\} \to \mathbb{R}$ as

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

Note that $f(r,\theta) > 0$ for any $0 < r < 1$, even on the curve γ. The exponential factor gives smoothness of the function and all its derivatives at the origin. By e.g. Whitney extension theorem the function could be extended to $\tilde{f}$ defined on $\mathbb{R}^2$; here it is sufficient to consider the behaviour close to the origin. We show that the gradient flow of f, starting at a point γ_0 on γ, stays always on γ, i.e. the polar gradient $(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \theta}\frac{1}{r})$ at points of the form $\{(r,\theta) : \theta - \log(\log(1/r)) = 2\pi k\}$ is parallel to $-\gamma'$.

First we note

$$\gamma'(t) = (r'(t), \theta'(t)) = \left(\frac{-1}{t^2}, \frac{1}{\log(t)}\frac{1}{t}\right) = \left(-r(t)^2, \frac{r(t)}{\log(1/r(t))}\right).$$

Thus it remains to check that for some scaling function $b(r)$ (in fact $b(r) = \frac{1}{r^2}e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{r + r^2\log^2(1/r)}\right)$) on the trajectory, that is whenever $\theta = \log(\log(1/r))$, we have

$$\frac{\partial f}{\partial r} = b(r)r^2 \qquad \text{and} \qquad \frac{\partial f}{\partial \theta}\frac{1}{r} = b(r)\frac{-r}{\log(1/r)}.$$

We compute that,

$$\begin{aligned}
\frac{\partial f}{\partial r}\bigg|_\gamma &= \frac{1}{r^2}e^{\frac{-1}{r}}\left(1-a(r)\sin(\theta-\log(\log(1/r)))\right) \\
&+ e^{\frac{-1}{r}}\left(-a'(r)\sin(\theta-\log(\log(1/r))-a(r)\cos(\theta-\log(\log(1/r))\frac{1}{r\log(1/r)}\right) \\
&= \frac{1}{r^2}e^{\frac{-1}{r}}\left(1-\frac{ra(r)}{\log(1/r)}\right) = \frac{1}{r^2}e^{\frac{-1}{r}}\left(1-\frac{r\log(1/r)}{r\log(1/r)+r^2\log^3(1/r)}\right) = e^{\frac{-1}{r}}\left(\frac{\log^2(1/r)}{r+r^2\log^2(1/r)}\right), \\
\frac{\partial f}{\partial \theta}\bigg|_\gamma &= -e^{\frac{-1}{r}}a(r)\cos(\theta-\log(\log(1/r)) = -e^{\frac{-1}{r}}a(r) = -e^{\frac{-1}{r}}\frac{\log(1/r)}{r+r^2\log^2(1/r)}.
\end{aligned}$$

Thus it follows that

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

Since $b(r)>0$ for all $0<r<1$,

$$-\nabla_{r,\theta}f|_\gamma = -\left(\frac{\partial f}{\partial r},\frac{\partial f}{\partial \theta}\frac{1}{r}\right)\bigg|_\gamma = b(r)\left(\frac{-1}{r^2},\frac{r}{\log(1/r)}\right)$$

is parallel to γ' for any point on γ, so the gradient flow stays on the curve (albeit with a different speed than its parametrization). Moreover since $\frac{\partial f}{\partial r}$ and $\frac{\partial f}{\partial \theta}$ converge to zero as $r\to 0$ ($a(r)\to 0$, which can be checked with e.g `Mathematica`), the origin is a critical point. To conclude, the trajectory of the continuous gradient descent flow converges to a point, has convergent arclength, yet the secants do not have any limit. It would be interesting to know relations to even stronger algebraic-geometric results like the *(analytic) finiteness conjecture for the gradient*.

4 Discretization attempts based on original paper

We now want to discretize the gradient conjecture to gradient descent, following a similar approach to [AMA05]. Then not only the arclength of discrete gradient descent is finite (shown in [AMA05]), but also the discrete trajectory would approach the origin in a straight line. In [AMA05] an inequality of the form

$$f(x_k)-f(x_{k+1})\geq\sigma|\nabla f(x_k)||x_{k+1}-x_k|\geq\sigma c|f(x_k)|^\rho|x_{k+1}-x_k|,\quad \rho<1,$$

is used to show convergence. The second inequality comes from the Lojasiewicz inequality, the first comes from the so called *primary descent condition* (satisfied for constant step size α or under Armijo condition),

$$f(x_k)-f(x_{k+1})\geq\sigma|\nabla f(x_k)||x_{k+1}-x_k|.$$

Having such an estimate leads then to

$$|x_{k+1}-x_k|\leq\frac{f(x_k)-f(x_{k+1})}{f(x_k)^\rho}\leq\frac{1}{c(1-\rho)}\left(f(x_k)^{1-\rho}-f(x_{k+1})^{1-\rho}\right), \tag{4.1}$$

which in turn gives a telescoping sum

$$\{\text{length up to } x_K\}=\sum_{k=0}^{K-1}|x_{k+1}-x_k|\leq C\left(f(x_0)^{1-\rho}-f(x_K)^{1-\rho}\right)\leq Cf(x_0)^{1-\rho}.$$

Now the upper bound is independent of K, thus passing to the limit as $K\to\infty$ gives that the length of the discrete trajectory is finite. For more details see [AMA05, Theorem 3.2].

To prove a discrete version of the gradient conjecture it is enough to find a function g (there is absolutely no need to use the f itself here) with a limit $g(x)\to L<\infty$ as $|x|\to 0$, that has some Lojasiewicz type estimate (i.e. some differential quantity is bounded below by the function raised to a power less than 1),

$$g(x_k)-g(x_{k+1})\geq\sigma|Dg(x_k)||\tilde{x}_{k+1}-\tilde{x}_k|\geq\sigma c|g(x_k)|^\rho|\tilde{x}_{k+1}-\tilde{x}_k|.$$

Unfortunately we could not succeed in showing it for general analytic functions. Note that even with a direct Lojasiewicz inequality (i.e. $Dg \geq |g|^\rho$, D denotes some differential quantity) it is non trivial to show the LHS inequality, since for $\tilde{x}$ (the projection onto $\mathbb{S}^{n-1}$) a factor of $\frac{1}{|x_{k+1}|}$ appears.

Assume that we can control $\frac{1}{|x_{k+1}|}$ by $\frac{1}{|x_k|}$ somehow, an extra factor of $\frac{1}{r}$. Morally speaking the problem increases in difficulty compared to the mere convergence in [AMA05], just because of the extra factor. Ideally we would like a strong Lojasiewicz inequality of the form $rDg \geq c|g|^\rho$ and $g(x_k) - g(x_{k+1}) \geq c|Dg||x_{k+1} - x_k|$ that holds around the origin, then it would follow

$$g(x_k) - g(x_{k+1}) \geq \frac{|x_k|}{|x_k|} Dg(x_k)|x_{k+1} - x_k| \geq c|g|^\rho \frac{1}{|x_k|}|x_{k+1} - x_k| = c|g|^\rho \frac{1}{|x_k|}|\tilde{x}_{k+1} - \tilde{x}_k|.$$

Straight from [KMP99] there is the estimate of Proposition 2.6, but for the discrete version it is not guaranteed that the iterates stay also in the strata $U_{l,a}$ where g is defined and well behaved. There is another estimate in [KMP99, Lemma 8.7], but with an exponent possibly greater than 1, the estimate of (4.1) does not work anymore.

5 Discretization based on dynamical systems

In this section we will aim for a discretization of the gradient conjecture using tools from dynamical systems. First we show a result in the continuous case and then discretize it. The idea is to linearize around the critical point (without loss of generality we assume it to be the origin) and use the "topological equivalence" of the trajectories to arrive at the gradient conjecture. However to make use of Hartman–Grobman type theorems, additional second order conditions on f (hyperbolic equilibrium point) are necessary.

5.1 Continuous version

Theorem 5.1. (Hartman, 1960 in [Per13, Section 2.8]) *Let $u : \mathbb{R}^n \to \mathbb{R}^n$ be of type C^2 around the origin, denote by Φ_t the flow of the nonlinear system $x'(t) = u(x(t))$. Suppose that $u(0) = 0$ and that all eigenvalues of $Du(0)$ have the same sign.*
Then there exists a C^1–diffeomorphism ψ of a neighborhood of the origin U onto an open neighborhood of the origin V such that for each $x \in U$ there is an open interval $I(x) \subset \mathbb{R}$ containing zero and

$$\psi \circ \Phi_t(x) = e^{(Du)t}\psi(x) \qquad \forall t \in I(x) \quad \forall x \in U.$$

Note that for gradient descent flow we set $u = -\nabla f$ and $Du = -H_f$. Here only the case for an attracting equilibrium point is of interest, in order to make use of the theorem we require that the eigenvalues of the Hessian at the origin are strictly positive. Without that condition i.e. only non-degeneracy of the equilibrium point, the conjugacy ψ may not be differentiable.

Proposition 5.2. *Let f be a three times differentiable cost function with an isolated, non-degenerate minimum at the origin and $H_f(0) > 0$. Then the trajectory of $x'(t) = -\nabla(x(t))$ (with initial point x_0 close enough to the origin) is differentiable at the origin, and the limit $\lim_{t\to\infty} \frac{x'(t)}{|x'(t)|}$ exists (strong gradient conjecture).*

Proof. At the origin the eigenvalues of $-H_f(0)$ are strictly negative. By the Theorem 5.1, the trajectory of x starting at $x_0 \in U$ is given by

$$x(t) = \psi^{-1}(e^{-H_f(0)t}\psi(x_0)), \tag{5.1}$$

where ψ is a diffeomorphism. Here U is a sufficiently small neighborhood of the origin, note that by the convergence of Theorem 1.11, $x(t) \in U$ for all t.

Actually from the theorem we only get topological equivalence on parts of the trajectory, but they can be pasted together. To be precise, let $I_0 = I(x_0) = (a_0, b_0)$ such that (5.1) holds. Then around $x(b_0) \in U$, construct another interval $I_1 = I(x(b_0)) = (a_1, b_1)$ where (5.1) holds. As $x((0, b_0)) \cap x((a_1, 0)) \neq \varnothing$, the intervals where (5.1) holds overlap and thus $x(t) = \psi^{-1}(e^{-H_f(0)t}\psi(x_0))$ holds for all t in $(a_0, b_0 + b_1)$. Iterating this procedure gives (5.1) on $(0, +\infty)$. If the series $b_0 + b_1 + b_2 + ...$ were to converge, there would be a minimal $b \in \mathbb{R}$ such that (5.1) holds in $(0, b)$ but not on $(b, b + \varepsilon)$ for any $\varepsilon > 0$. Nonetheless there exists an interval $I(x(b))$ such that (5.1) holds on $(0, b) \cup I(x(b)) + b$, contradicting the fact that (5.1) did not hold on $(b, b + \varepsilon)$, hence (5.1) holds for any $t > 0$.

In other words $x(t) = \Psi \circ y(t)$, for $\Psi = \psi^{-1}$ and $y(t) = e^{-H_f(0)t}\psi(x_0)$, the solution of the linear system $y'(t) = -H_f(0)y(t)$ with initial condition $\psi(x_0)$. Since the trajectory $y(t)$ will approach the origin along the eigenspace spanned by the largest eigenvalue(s) (compare to Lemma 5.8), the limit $\lim_{t\to\infty} \frac{y'(t)}{|y'(t)|} =: v$ exists. For the original trajectory $x(t)$,

$$x'(t) = \frac{d}{dt}(\Psi \circ y(t)) = D\Psi(y(t)) \cdot y'(t),$$

with $D\Psi : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ being continuous (here use the fact that ψ is diffeomorphism and thus differentiable in a neighborhood around the origin). Thus, for $M := D\Psi(0)$ being a nonzero matrix, the limit

$$\begin{aligned}\lim_{t\to\infty} \frac{x'(t)}{|x'(t)|} &= \lim_{t\to\infty} \frac{D\Psi(y(t)) \cdot y'(t)}{|D\Psi(y(t)) \cdot y'(t)|} = \lim_{t\to\infty} \frac{D\Psi(y(t)) \cdot \frac{y'(t)}{|y'(t)|}}{|D\Psi(y(t)) \cdot \frac{y'(t)}{|y'(t)|}|} = \frac{\lim_{t\to\infty} D\Psi(y(t)) \cdot \lim_{t\to\infty} \frac{y'(t)}{|y'(t)|}}{|\lim_{t\to\infty} D\Psi(y(t)) \cdot \lim_{t\to\infty} \frac{y'(t)}{|y'(t)|}|} \\ &= \frac{D\Psi(\lim_{t\to\infty} y(t)) \cdot v}{|D\Psi(\lim_{t\to\infty} y(t)) \cdot v|} = \frac{D\Psi(0) \cdot v}{|D\Psi(0) \cdot v|} = \frac{M \cdot v}{|M \cdot v|} \in \mathbb{R}^n,\end{aligned}$$

exists and so the strong gradient conjecture holds. ☐

Corollary 5.3. *Since the strong gradient conjecture implies the gradient conjecture (Theorem 3.4), the limit, i.e.* $\lim_{t\to\infty} \frac{x(t)}{|x(t)|}$ *exists, under the same conditions on* f*, i.e.* C^2 *and* $H_f(0) > 0$.

Remark 5.4. Actually $D\Psi$ is continuous in a neighborhood of the origin, but in the proof only the continuity at the origin is used to move the limit inside, this idea made the proof in the next section possible.

5.2 Discrete version

We will now apply a similar argument to show that the gradient conjecture holds not only for the continuous gradient flow, but also for the discrete gradient descent with fixed step size, at least for a non-degenerate local minimum.

Proposition 5.5. ([LSJR16, Proposition 4.5]) *Assume* f *is of type* C^k, ∇f *is* L*–Lipschitz and let* $\alpha < \frac{1}{L}$. *Then the map* $F : \mathbb{R}^n \to \mathbb{R}^n$ *given by* $F(x) = x - \alpha\nabla f(x)$ *is a* C^{k-1} *diffeomorphism in a neighborhood of the origin.*

Proof. Since $DF(0) = I - \alpha H_f(0)$ is invertible for $\alpha < 1/L$ (L is the Lipschitz constant of ∇f, the eigenvalues are all positive) and $F \in C^{k-1}$ the result follows from the inverse function theorem, see [Vas06]. ☐

Now a general case of Theorem 5.1, that can be applied to discrete dynamical system given by diffeomorphisms [Cic06, Section 4.3.1] is needed we propose the following:

Theorem 5.6. (van Strien, 1989 in [vS90]) *Let* $F : \mathbb{R}^n \to \mathbb{R}^n$ *be a* C^2 *diffeomorphism with hyperbolic fixpoint at the origin, i.e. no eigenvalue of* $DF(0)$ *lies on the unit circle. Then there exists a neighborhood* U *of the origin and a homeomorphism* $\psi : V \to \psi(V)$ *such that:*

(a) $F(x) = \psi^{-1}(DF(0)\psi(x))$ *on* $U \subset V$,

(b) ψ, ψ^{-1} *are differentiable at the origin,*

(c) ψ, ψ^{-1} *are Hölder continuous in* U.

Remark 5.7. The C^2 regularity can be weakened to $C^{1,\alpha}$ and extended to a Banach space as described in [ZLZ15] for a more complicated condition on the spectrum of $DF(0)$.

Lemma 5.8. *Define a discrete dynamical linear system by* $y_k = Ay_{k-1}$ *for* $k \geq 1$ *and* y_0 *fixed. If the eigenvalues of* $A \in \mathrm{Sym}_n$ *lie all inside the interval* $(0, 1)$ *, then the limit* $\lim_{k\to\infty} \frac{y_k}{|y_k|}$ *exists.*

Proof. Start with $y_k = Ay_{k-1} = A^k y_0$, then write $A^k = SD^kS^{-1}$, for a diagonal matrix D, by the spectral theorem. Without loss of generality assume that the diagonal entries of D are

$$|\lambda_1| = \cdots = |\lambda_j| > |\lambda_{j+1}| \geq \cdots \geq |\lambda_n| \qquad \text{for } j \geq 1. \tag{5.2}$$

It then follows that $y_k = SD^kS^{-1}y_0$ can be written as

$$y_k = S(\lambda_1^k c_1, \ldots, \lambda_j^k c_j, \lambda_{j+1}^k c_{j+1}, \ldots \lambda_n^k c_n) = \lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k}).$$

It then follows, that

$$\begin{aligned}
\frac{y_k}{|y_k|} &= \frac{\lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k})}{|\lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k})|} = \frac{\lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k})}{|\lambda_1^k c_1 e_1| + \cdots + |\lambda_j^k c_j e_j| + |o(|\lambda_1|^{-k})|} \\
&= \frac{\lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k})}{|\lambda_1|^k(c_1 + \cdots + c_j) + |o(|\lambda_1|^{-k})|} = \frac{\frac{1}{|\lambda_1|^k}}{\frac{1}{|\lambda_1|^k}} \cdot \frac{\lambda_1^k c_1 e_1 + \cdots + \lambda_j^k c_j e_j + o(|\lambda_1|^{-k})}{|\lambda_1|^k(c_1 + \cdots + c_j) + |o(|\lambda_1|^{-k})|} \\
&\longrightarrow \frac{c_1 e_1 + \cdots + c_j e_j}{c_1 + \cdots + c_j} \quad \text{as } k \to \infty.
\end{aligned}$$

That is the iterates will approach the eigenspace corresponding to the largest eigenvalue(s). □

Theorem 5.9. *Let f be a function of type C^3 with non-degenerate Hessian and isolated strict minimum ($\nabla f(0) = 0$) at the origin. Suppose the gradient descent with $x_{k+1} = F(x_k) = x_k - \alpha \nabla f(x_k) = F^k(x_0)$, where α sufficiently small, converges to the origin.*
Then the iterates x_k satisfy the gradient conjecture, i.e. $\lim_{k\to\infty} \frac{x_k}{|x_k|}$ exists.

Proof. The map $x \mapsto F(x) = x - \alpha \nabla f(x)$ is a C^2 diffeomorphism if α is sufficiently small (e.g. $\alpha < \frac{1}{L}$, with L the Lipschitz constant of f). Moreover, from the non-degeneracy of the Hessian the real eigenvalues are all strictly positive and so the eigenvalues of $DF(0)$ are all inside $(0, 1)$, if α small. F has thus a hyperbolic fixed point at the origin and by van Strien Theorem 5.9, there exists a homeomorphism ψ defined on a neighborhood U of the origin, such that

$$F(x) = \psi^{-1}(DF(0)\psi(x)) =: \psi^{-1}(A\psi(x)),$$

and $D\psi(0)$, $D(\psi^{-1})(0)$ exist. Since $x_k \to 0$, there exists some $K > 0$ s.t. $x_k \in U$ for all $k \geq K$, in particular (after relabeling $k := k - K$),

$$\begin{aligned}
x_1 &= F(x_0) = \psi^{-1}(A\psi(x_0)), \\
x_k &= F^k(x_0) = \psi^{-1}(A\psi(x_k)) = \psi^{-1}\left(A\psi\left(F^{k-1}(x_0)\right)\right) = \psi^{-1}\left(A\psi\left(\psi^{-1}(A\psi(x_{k-1}))\right)\right) \\
&= \psi^{-1}\left(A^2\psi(x_{k-1})\right) = \cdots = \psi^{-1}\left(A^k\psi(x_0)\right) =: \Psi(A^k\psi(x_0)),
\end{aligned}$$

with $\Psi := \psi^{-1}$ differentiable at the origin. Let $y_0 := \psi(x_0)$ and consider the linear discrete dynamical system

$$y_k = Ay_{k-1}, \quad k = 1, 2, ..., \qquad y_0 \text{ fixed.} \tag{5.3}$$

For the linear system the trajectory is well understood, it approaches the origin along the slowest eigenspace(s). In particular $\lim_{k\to\infty} \frac{y_k}{|y_k|} =: v$ exists by Lemma 5.8. Now, by Taylor expansion (for that only differentiablity at the origin is needed!),

$$\Psi(x) = \Psi(0) + D\Psi(0) \cdot x + o(|x|) = D\Psi(0) \cdot x + o(|x|),$$

since for the homeomorphism $\psi(0) = 0 = \Psi(0)$. Now

$$\begin{aligned}
\frac{x_{k+1}}{|x_{k+1}|} &= \frac{F(x_k)}{|F(x_k)|} = \frac{\Psi(A^k\psi(x_0))}{|\Psi(A^k\psi(x_0))|} = \frac{D\Psi(0) \cdot A^k\phi(x_0) + o(|A^k\phi(x_0)|)}{|D\Psi(0) \cdot A^k\phi(x_0) + o(|A^k\phi(x_0)|)|} \\
&= \frac{|A^k\phi(x_0)|}{|A^k\phi(x_0)|} \frac{D\Psi(0) \cdot A^k\phi(x_0) + o(|A^k\phi(x_0)|)}{|D\Psi(0) \cdot A^k\phi(x_0) + o(|A^k\phi(x_0)|)|} = \frac{D\Psi(0) \cdot \frac{A^k\phi(x_0)}{|A^k\phi(x_0)|} + \frac{o(|A^k\phi(x_0)|)}{|A^k\phi(x_0)|}}{|D\Psi(0) \cdot \frac{A^k\phi(x_0)}{|A^k\phi(x_0)|} + \frac{o(|A^k\phi(x_0)|)}{|A^k\phi(x_0)|}|} \\
&= \frac{D\Psi(0) \cdot \frac{y_{k+1}}{|y_{k+1}|} + \frac{o(|y_{k+1}|)}{|y_{k+1}|}}{|D\Psi(0) \cdot \frac{y_{k+1}}{|y_{k+1}|} + \frac{o(|y_{k+1}|)}{|y_{k+1}|}|},
\end{aligned}$$

thus when passing to the limit

$$\lim_{k\to\infty} \frac{x_{k+1}}{|x_{k+1}|} = \lim_{k\to\infty} \frac{D\Psi(0)\cdot \frac{y_{k+1}}{|y_{k+1}|} + \frac{o(|y_{k+1}|)}{|y_{k+1}|}}{|D\Psi(0)\cdot \frac{y_{k+1}}{|y_{k+1}|} + \frac{o(|y_{k+1}|)}{|y_{k+1}|}|} = \frac{D\Psi(0)\cdot v}{|D\Psi(0)\cdot v|} \in \mathbb{S}^{n-1}.$$

This finishes the proof. $\square$

Remark 5.10. The strong gradient conjecture in the discrete setting could be interpreted as existence of the limit $\lim_{k\to\infty} \frac{x_{k+1}-x_k}{|x_{k+1}-x_k|}$ which we have not shown to be true and would be a stronger result. Not only does the vectors x_k converges, but also the vectors $x_k - x_{k-1}$ connecting them. Another question could be the extension to non-degenerate critical points. The same argument with the Taylor expansion should work in the continuous case with $x'(t) = F(x) = -\nabla f(x(t))$, so $\lim_{t\to} \frac{x(t)}{|x(t)|}$ exists as long as the Hessian is non-degenerate.

It remains open if similar results could be shown if f is a Morse-Bott function, i.e. the Hessian is non-degenerate in the normal space, but we would maybe need to establish that the trajectory stays in the normal space to use similar Hartman–Grobman type arguments.

References

[AMA05] 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.

[Cic06] Carmen Cicone. *Ordinary Differential Equations with Applications*. Texts in Applied Mathematics. Springer, 2006.

[CM14] Tobias Holck Colding and William P. Minicozzi. Lojasiewicz inequalities and applications, 2014.

[HM94] U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. Springer London, 1994.

[KMP99] Krzysztof Kurdyka, Tadeusz Mostowski, and Adam Parusinski. Proof of the gradient conjecture of R. Thom. *Annals of Mathematics*, 152, 07 1999.

[KOS00] Krzysztof Kurdyka, Patrice Orro, and Stéphane Simon. Semialgebraic sard theorem for generalized critical values. *Journal of Differential Geometry*, 56:67–92, 2000.

[Kur98] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier*, 48(3):769–783, 1998.

[Loj65] Stanislaw Lojasiewicz. Ensembles semi-analytiques. 1965.

[Loj84] Stanislaw Lojasiewicz. Trajectoires du gradient d'une fonction analytique. *Seminari di Geometria 1982-1983*, pages 115–117, 1984.

[LSJR16] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient descent converges to minimizers. 2016.

[MC13] Bassel Mannaa and Thierry Coquand. Dynamic Newton–Puiseux theorem. *Journal of Logic and Analysis*, pages 1–22, 2013.

[MR09] Carlos Gustavo Moreira and Maria Aparecida Soares Ruas. The curve selection lemma and the Morse–Sard theorem. *manuscripta mathematica*, 129:401–408, 2009.

[Per13] Lawrence Perko. *Differential Equations and Dynamical Systems*, volume 7 of *Texts in Applied Mathematics*. Springer, New York, NY, third edition. edition, 2013.

[PW] PW. Puiseux series - wikipedia. Accessed Jan 5, 2023. `https://en.wikipedia.org/wiki/Puiseux_series`.

[Vas06] András Vasy. Lecture notes on the inverse function theorem, 2006. Accessed Jan 5, 2023 `http://virtualmath1.stanford.edu/~andras/174A-2.pdf`.

[vS90] Sebastian van Strien. Smooth linearization of hyperbolic fixed points without resonance conditions. *Journal of Differential Equations*, 85(1):66–90, 1990.

[ZLZ15] Wenmeng Zhang, Kening Lu, and Weinian Zhang. Differentiability of the conjugacy in the Hartman-Grobman theorem. *Transactions of the American Mathematical Society*, 369:1, 09 2015.