ON CONVERGENCE TO ESSENTIAL SINGULARITIES*

NATHANIEL J. MCCLATCHEY†

Abstract. An iterative optimization method applied to a function \( f \) on \( \mathbb{R}^n \) will produce a sequence of arguments \( \{x_k\}_{k \in \mathbb{N}} \); this sequence is often constrained such that \( \{f(x_k)\}_{k \in \mathbb{N}} \) is monotonic. As part of the analysis of an iterative method, one may ask under what conditions the sequence \( \{x_k\}_{k \in \mathbb{N}} \) converges. In 2005, Absil et al. employed the Lojasiewicz gradient inequality in a proof of convergence; this requires that the objective function exist at a cluster point of the sequence. Here we provide a convergence result that does not require \( f \) to be defined at the limit \( \lim_{k \to \infty} x_k \), should the limit exist. We show that a variant of the Lojasiewicz gradient inequality holds on sets adjacent to singularities of bounded multivariate rational functions. We extend the results of Absil et al. to prove how this may be employed to analyze divergent sequences by mapping them to projective space, and consider the implications this has for the study of low-rank tensor approximations.

Key words. Lojasiewicz gradient inequality, convergence

AMS subject classifications. 40A05, 90C26

1. Introduction. Historically, a variety of approaches have been employed to examine the convergence of sequences produced by iterative optimization methods. Some examine local convergence – whether small perturbations of a solution are corrected, and at what rate. Such results are typically proved directly – as in the case of the Newton-Raphson method – or by comparison to a method with known behavior [4, 13]. Others attempt to show global convergence – that a particular method will produce a convergent sequence. Though global convergence can be proved directly in some cases [5, 12], more general approaches exist. Since 1971, the Wolfe conditions [16] have been employed to ensure that the gradient of the objective function – evaluated at the points of a sequence – tends to zero. More recently, seminal work by Absil et al. [1] employed the Lojasiewicz gradient inequality to provide a stronger result.

For convenience, we state the gradient form of Stanislaw Lojasiewicz’s theorem [8, 9] below, incorporating an improvement from [11].

**Theorem 1 (Lojasiewicz gradient inequality).** Let \( f \) be a real-analytic function on a neighborhood of \( x_* \) in \( \mathbb{R}^n \). Then there are constants \( c > 0 \) and \( \theta \in (0, 1/2] \) such that

\[
|f(x) - f(x_*)|^{1-\theta} \leq c\|\nabla f(x)\|
\]

for any \( x \) in some neighborhood of \( x_* \).

Though the constant \( \theta \) can be determined in some cases [3] and estimated in others [6, 10], it is in general a priori unknown.

The result of Absil et al. [1] states that a sequence \( \{x_k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) converges if it has a cluster point and if there is some real-analytic function \( f : \mathbb{R}^n \to \mathbb{R} \) such that the “descent conditions” (A1–A2) hold. The first descent condition, that

\[
(A1) \quad f(x_k) - f(x_{k+1}) \geq \sigma \|\nabla f(x_k)\| \|x_{k+1} - x_k\|
\]

*January 4, 2018

**Funding:** Supported by the National Science Foundation under Grant 1418787.

†Ohio University, Athens, OH (nml60111@ohio.edu)
for some $\sigma > 0$ and all sufficiently large $k$, ensures that the sequence of $f$-values decreases sufficiently quickly. The second descent condition prevents cycles by requiring that

$$f(x_k) = f(x_{k+1}) \implies x_k = x_{k+1}.$$ (A2)

Moreover, Uschmajew, et al. proved that if also there exists some $\kappa > 0$ such that for all sufficiently large $k$,

$$\|x_{k+1} - x_k\| \geq \kappa \|\nabla f(x_k)\|,$$ (A3)

then the rate of convergence of the sequence may be bounded [11, 14].

In recent years, the results of Absil, Uschmajew, et al. [1, 11, 14] have been of interest in the field of tensor approximation. Given a target tensor $T$, one may attempt to find the tensor of a fixed rank $r$ which best approximates the target. This problem is often described as minimization of the “error” $\|T - \tau(x)\|$, where $x \in \mathbb{R}^n$ is a tuple of parameters, and $\tau$ is a multilinear map from parameters to tensors of rank $r$. If $r = 1$, there exist parameters $x$ that minimize error, and several methods, including Alternating Least Squares (ALS), are known to produce convergent sequences [14, 7]. If $r \geq 2$, then the approximation $\tau(x)$ is a sum of two or more separable tensors $X^1 + \cdots + X^r$. Because no orthogonality constraint is imposed on $X^1, \ldots, X^r$, there exist problems for which error cannot be minimized [2, 15]. For these ill-posed problems, if $\|T - \tau(x_k)\|$ is to approach its infimum, the sequence of parameters $\{x_k\}_{k \in \mathbb{N}}$ must diverge. As yet, little intuition exists for the behavior of these sequences; in particular, it is not known whether the summands $X^1, \ldots, X^r$ maintain some steady configuration, changing little except in scale, or whether they cycle among different configurations.

To gain intuition, the divergent case may be converted to a projective space by rescaling; one may then ask whether the normalized sequence $\{x_k/\|x_k\|\}_{k \in \mathbb{N}}$ converges. When this re-scaling is performed, however, one must alter the objective function such that it is optimized by a unit vector, as in $x \mapsto \min_{\lambda \in \mathbb{R}} f(\lambda x)$. In the tensor approximation problem, this results in a bounded multivariate rational function. Targets for which the best approximation problem is ill-posed correspond to singularities of this function; some of these singularities are known to be essential — that is, neither removable nor unbounded.

**Fig. 1.** Line Search along the gradient maximizes $f(x, y) = \frac{-xy}{(x^2+y^2)(1+x^2+y^2)}$ from an initial estimate of $(x_0, y_0) = (2, -0.1)$. 
Wherever a function fails to be continuous, the Lojasiewicz gradient inequality cannot apply. This does not prevent optimization methods from producing convergent sequences, as is illustrated in Figure 1, but it does preclude the use of the theorems of Absil, Uschmajew, et al. Fortunately, those theorems may be strengthened to require only that the Lojasiewicz gradient inequality hold on points in the sequence.

The conditions required by the following theorems will follow from the main results of this paper. Specifically, Theorems 10 and 13 will establish (L*) for sequences such as that illustrated in Figure 1.

**Theorem 2.** Let \( U \subset \mathbb{R}^n \) be an open set, let \( f : U \to \mathbb{R} \) be a differentiable function, and let \( \{x_k\}_{k=1}^\infty \subset U \) be a sequence of vectors satisfying Assumptions (A1) and (A2). If a cluster point \( x^* \) of the sequence \( \{x_k\}_{k=1}^\infty \) admits an open neighborhood \( V \subset \mathbb{R}^n \) and constants \( \theta \in (0,1/2] \) and \( c \in \mathbb{R} \) such that for all \( k \),

\[
(L_*) \quad x_k \in V \implies |f(x_k) - \lim_{j \to \infty} f(x_j)|^{1-\theta} \leq c\|\nabla f(x_k)\|
\]

then \( x^* \) must be the limit of the sequence \( \{x_k\}_{k=1}^\infty \).

**Theorem 3.** Under the conditions of Theorem 2, if also Assumption (A3) holds, then \( \nabla f(x_k) \to 0 \) and

\[
\|x^* - x_k\| = \begin{cases} O(q_\theta^k) & \text{if } \theta = \frac{1}{2} \quad (\text{for some } 0 < q < 1), \\ O(k^{-\frac{1}{q-1}q}) & \text{if } 0 < \theta < \frac{1}{2} \end{cases}
\]

where \( \theta \) is such that (L*) holds.

The following short proof lists only those changes required to extend Uschmajew’s result. For a full proof of Theorems 2 and 3, see Appendix A.

**Proof of Theorems 2 and 3.** Though \( f(x^*) \) is not assumed to exist, (L*) implies that \( \lim_{k \to \infty} f(x_k) \) exists and is finite. With this established, only three changes are needed to extend the proof in [11, p. 644] to a proof of Theorems 2 and 3:

1. Replacing all occurrences of \( f(x^*) \) with \( \lim_{k \to \infty} f(x_k) \) we remove the requirement that \( x^* \) be in the domain of \( f \).
2. Inequality (A.1) in [11] need only hold when \( x = x_k \). Thus, it follows from (L*).
3. The statement that \( n \) may be selected so large that \( \|x_n - x^*\| < \frac{\epsilon}{q} \) and \( \frac{d}{d_n} f_n^\theta < \frac{\epsilon}{2} \) does not require that \( f \) be continuous at \( x^* \). Instead, it follows from the existence of \( \lim_{k \to \infty} f(x_k) \) and the assumption that \( x^* \) is a cluster point.

**Remark 4.** Theorems 2 and 3 are strictly stronger than the results of Absil, Uschmajew, et al. in that weaker hypotheses allow the same conclusions. Specifically, if \( f \) is analytic on a neighborhood of \( x^* \), (L*) follows from Theorem 1.

Our Theorems 2 and 3 do not require continuity of the objective function, but do require a weaker form (L*) of the Lojasiewicz gradient inequality. Beyond Theorems 2 and 3, the contributions of this paper are four-fold.

- In subsection 2.1, we contribute a version of the Lojasiewicz gradient inequality that may be employed at essential singularities of bounded multivariate rational functions. This inequality holds not on a neighborhood of a singularity, but instead on open sets which have boundaries containing the singularity.
- Subsection 2.2 establishes that our Lojasiewicz gradient inequality holds on the tail of the sequence, under a technical condition. Specifically, if the lim-
iting behavior of \( f \) near the singularity does not depend continuously on the direction of approach, then the sequence must avoid the “unsafe” directions at which discontinuities occur. We show that the set of “unsafe” directions of approach is closed and has Lebesgue measure zero.

- Subsection 2.3 combines these results to prove that if this and (A1–A3) hold, then the sequence \( \{x_{k+1}\}_{k \in \mathbb{N}} \) converges.
- Finally, section 3 provides additional tools for examination of convergence in direction and briefly discusses the implications for tensor approximation.

This partially closes a gap in the theory of tensor approximation, and provides a general tool for analysis of sequences.

2. On the assumption (L*)

Before we begin our analysis of (L*), we note two obstacles and the means by which we circumvent them.

First, we must assume that \( f \) has some structure strict enough that its behavior may be analyzed near a singularity \( x^* \), yet not so strict as to require that \( f \) be continuous at \( x^* \). Throughout the paper, we will consider multivariate rational functions on \( \mathbb{R}^n \), with the assumption that the domain of such functions is defined implicitly to be all points in \( \mathbb{R}^n \) at which division by zero does not occur. More formally, we define a multivariate rational function as follows:

**Definition 5.** A function \( r \) is a multivariate rational function on \( \mathbb{R}^n \) if and only if there exist multivariate polynomials \( p, q : \mathbb{R}^n \rightarrow \mathbb{R} \) such that \( r = \frac{p}{q} \).

Second, we cannot expect the Lojasiewicz gradient inequality to hold on open neighborhoods of \( x^* \). For an essential singularity, however, (2) can fail to hold on open neighborhoods of \( x^* \). Rather than examine open neighborhoods of singularities, we establish that (2) holds on certain open sets adjacent to singularities of bounded multivariate rational functions. We then provide conditions under which sequences \( \{x_k\}_{k \in \mathbb{N}} \) remain within those open sets.

2.1. A Lojasiewicz-like inequality holds on cones

Bounded multivariate rational functions may admit essential singularities, such as that illustrated in Figure 1. Bounded univariate rational functions instead exhibit only removable singularities. In this subsection, we parameterize multivariate functions in terms of directions and distance; that is, as images of lines under a multivariate function. Lemmas 6 and 8 establish that these parameterizations are themselves analytic, if not rational. The parameterization is employed in Theorem 10 to establish a Lojasiewicz inequality on cones near singularities of the original function.

**Lemma 6.** Suppose \( f \) is a multivariate rational function on \( \mathbb{R}^m \times \mathbb{R} \). If

\[
f_x(t) = \lim_{s \to t} f(x, s)
\]
is defined for some $x \in \mathbb{R}^m$ and all $t \in \mathbb{R}$, then there exists an open set $O \subset \mathbb{R}^m$ of full measure and multivariate rational functions $c_n : O \to \mathbb{R}$ defined everywhere on $O$ such that, for every $x \in O$ there exists $\rho_x > 0$ such that

$$|t| < \rho_x \implies f_x(t) = \sum_{n=0}^{\infty} c_n(x)t^n.$$ 

**Proof.** Suppose $f_x(t) = \lim_{s \to t} f(x, s)$ is defined for all $t \in \mathbb{R}$. By fixing $x \in \mathbb{R}^m$, we reduce $f(x, t)$ to a univariate rational function with respect to $t$, so its continuous extension $f_x(t)$ is a univariate rational function. Because $f_x(t)$ is defined everywhere, it is everywhere real-analytic.

Let $f(x, t) = \frac{f_{\text{numer}}(x, t)}{f_{\text{denom}}(x, t)}$, where $f_{\text{numer}}$ and $f_{\text{denom}}$ are multivariate polynomials. $f_x(t)$ is analytic, so the coefficients $c_n$ of its Taylor expansion around 0 exist, and for $n = 0, \ldots, \infty$,

$$c_n(x) = \frac{1}{n!} \lim_{t \to 0} \frac{d^n f_{\text{numer}}(x, t)}{dt^n f_{\text{denom}}(x, t)}.$$ 

Because the function $f_x$ is analytic for each $x$, there exist $\rho_x$ such that $|t| < \rho_x$ implies $f_x(t) = \sum_{n=0}^{\infty} c_n(x)t^n$. Note that the neighborhood of convergence depends on the direction parameter $x$. We now turn our attention to properties of the coefficients $c_n(x)$.

By repeated application of the quotient rule, and omitting the numerators because they are not required for the proof, we obtain an expression of the form

$$c_n(x) = \frac{1}{n!} \lim_{t \to 0} \frac{\cdots}{(f_{\text{denom}}(x, t))^2},$$ 

which is a limit of a rational function. These limits exist by analyticity of $f_x(t)$, so they may be evaluated by repeated application of L’Hospital’s rule.

Let $f_n(x)$ denote the coefficients of the Taylor expansion, with respect to $t$, of $f_{\text{denom}}(x, t)$. Note that these are multivariate polynomials with respect to $x$. Let

$$n_{\text{min}} = \min\{n \in \mathbb{Z} : n \geq 0 \text{ and } \exists x \text{ st. } f_n(x) \neq 0\},$$

and define

$$O = \{x : f_{n_{\text{min}}}(x) \neq 0\}.$$ 

It must be shown that $n_{\text{min}}$ is well-defined. The function $f_x$ is assumed to be defined for some $x \in \mathbb{R}^m$, so $f_{\text{denom}}$ is nonzero at some $(x, t)$. From this and analyticity of $f_{\text{denom}}$, it follows that $f_n(x)$ is non-zero for some $n$. This establishes that $n_{\text{min}}$ is well-defined.

A multivariate real or complex polynomial is either identically zero or non-zero almost-everywhere with respect to Lebesgue measure. Further, the set on which such a polynomial is equal to a given constant is closed. Thus $O$ is an open set of full measure.

For every $x \in O$, the coefficients $c_n(x)$ may be evaluated by applying L’Hospital’s rule exactly $2^n n_{\text{min}}$ times. This implies that on $O$, every $c_n$ is a multivariate rational function with respect to $x$, and is defined everywhere on $O$. $\square$
It is well known that the trailing coefficients of the Taylor series expansion of a univariate rational function satisfy a linear recurrence relation. For completeness, we provide a proof of this here:

**Proposition 7.** Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, let $f, g : \mathbb{K} \to \mathbb{K}$ be polynomials of degree $d_f$ and $d_g$ given as $f(x) = \sum_{i=0}^{d_f} f_i x^i$ and $g(x) = \sum_{j=0}^{d_g} g_j x^j$, and let $\{c_n\}_{n=0}^\infty \subset \mathbb{K}$ such that $(\frac{g}{f})(x) = \sum_{n=0}^\infty c_n x^n$ on some open $U \subset \mathbb{K}$. Then for all $n > \max\{d_f, d_g\}$, the coefficients $c_n$ are given by $c_n = -\sum_{j=1}^{d_g} c_{n-j} \frac{g_j}{f_0}$.

**Proof.** For all $x \in U$, we have $\frac{f(x)}{g(x)} = \sum_{n=0}^\infty c_n x^n$. Then, multiplying by $g(x)$,

$$
\sum_{i=0}^{d_f} f_i x^i = \left(\sum_{j=0}^{d_g} g_j x^j\right) \left(\sum_{n=0}^\infty c_n x^n\right)
$$

Matching terms gives the equation $0 = \sum_{j=0}^{d_g} c_{n-j} g_j$ for all $n > \max\{d_f, d_g\}$, and solving for $c_n$ completes the proof. \( \square \)

We have shown that the Taylor series expansion in Lemma 6 converges within some radius of convergence $\rho_x$, but this radius of convergence might depend on $x$. Before we can use Lemma 6 for its intended purpose, we must show that $\rho_x$ may be bounded away from 0. We will do this by showing that $\rho_x$ depends continuously on $x$ so that we may later apply a compactness argument.

**Lemma 8.** If the maps $c_n$ are defined as in Lemma 6, then there exists a continuous map $\rho : \mathcal{O} \to (0, \infty)$ and a corresponding set

$$
U_\rho = \{ (x, t) \mid x \in \mathcal{O}, t \in (-\rho(x), \rho(x)) \}
$$

on which $\sum_{n=0}^\infty |c_n(x)t^n|$ converges uniformly.

**Proof.** By Proposition 7, there exist $N, r \in \mathbb{N}$ and matrices $C_x \in \mathbb{R}^{r \times r}$ such that for all $n > N$,

$$
\begin{bmatrix}
  c_{n+1}(x) \\
  \vdots \\
  c_{n+r+1}(x)
\end{bmatrix} = C_x \begin{bmatrix}
  c_n(x) \\
  \vdots \\
  c_{n+r}(x)
\end{bmatrix}.
$$

Further, the recurrence matrices $C_x$ and its $\infty$-norm $\|C_x\|_\infty$ depend continuously on the choice of $x \in \mathcal{O}$. We may thus bound the coefficients $c_n$ by a geometric sequence and control the convergence of the terms $c_n t^n$ by choice of $t$. Select $N \in \mathbb{N}$ such that recurrence relation holds for all $n \geq N - r$. Note that $N$ may be chosen identically for all $x \in \mathcal{O}$. For any $0 < k < 1$, selecting $t$ such that

$$
|t| \leq \frac{k}{\max\{1, \|C_x\|_\infty\}}
$$

guarantees that for $n \geq N$, we may bound $|c_n t^n|$ by

$$
|c_n(x)t^n| \leq \left(\max_{j=0,\ldots,r} |c_{N-j}(x)|\right) \|C_x\|_\infty^{n-N} \left(\frac{k}{\max\{1, \|C\|_\infty\}}\right)^n \\
\leq \left(\max_{j=0,\ldots,r} |c_{N-j}(x)|\right) k^n.
$$

(4)
Then
\[ \sum_{n=N}^{\infty} |c_n(x)t^n| \leq k^N \left( \frac{1}{1-k} \right) \max_{j=0,\ldots,r} |c_{N-j}(x)|. \]
This may be bounded arbitrarily by choice of \( k \). In particular, for a given bound \( \epsilon > 0 \), the above bound shows that \( \sum_{n=N}^{\infty} |c_n(x)t^n| \leq \epsilon \) if
\[ \left( \max_{j=0,\ldots,r} |c_{N-j}(x)| \right) k^N + \epsilon k - \epsilon = 0. \]
This equation has a solution \( k = k_{x,\epsilon} \in (0,1) \) which depends continuously on \( x \) and \( \epsilon \).

Fix \( \epsilon > 0 \), and define \( \rho \) by
\[ x \mapsto \frac{k_{x,\epsilon}}{2 \max\{1,\|C_x\|_{\infty}\}}. \]
Note that this map is continuous with respect to \( x \), and that
\[ |t| \leq 2\rho(x) \implies \sum_{n=N}^{\infty} |c_n(x)t^n| \leq \epsilon. \]
Then, for \( |t| \leq \rho(x) \) and \( n \geq N \), the terms \( |c_n(x)t^n| \) are bounded by
\[ |c_n(x)t^n| \leq \frac{\epsilon}{2^n}. \]
The series \( \sum_{n=N}^{\infty} \frac{\epsilon}{2^n} \) converges, so \( \sum_{n=N}^{\infty} |c_n(x)t^n| \) converges uniformly for \((x,t) \in U_{\rho}\).

We now define \( O_{x^*} \), the largest set of directions from which \( x^* \) can be approached while ensuring that the parameterized function \((x,t) \mapsto f(tx + x^*)\) behaves consistently.

**Definition 9.** Given a bounded multivariate rational function \( f \) on \( \mathbb{R}^n \) and a point \( x^* \in \mathbb{R}^n \), and denoting by \( f_n(x) \) the \( n \)-th Maclaurin series coefficient, with respect to \( t \), of the denominator of the rational function defined by \((x,t) \mapsto f(tx + x^*)\), we define
\[ O_{x^*} = \{ x : f_{n_{\min}}(x) \neq 0 \}, \]
where \( n_{\min} = \min\{ n \in \mathbb{Z}_{\geq 0} \mid \exists x : f_n(x) \neq 0 \} \).

**Definition 9** will be used to describe those directions of approach near which the function \( f \) is sufficiently well-behaved. Later sections will demonstrate that if a sequence admits a subsequence that approaches \( x^* \) along a direction in \( O_{x^*} \), then \( x^* \) is the limit of that sequence. That is, approaching \( x^* \) from almost any direction will trap the sequence at that cluster point.

Note that \( O_{x^*} \) satisfies the conditions of the open set specified by **Lemma 6** for the function \( f_s(t) = \lim_{s \to t} f(sx + x^*) \), as it is identical to the set constructed in the proof of **Lemma 6**. Note also that for any \( t \in \mathbb{R} \setminus \{0\} \) and \( x \in O_{x^*} \), we have \( tx \in O_{x^*} \).

We are now prepared to show that a generalized Lojasiewicz gradient inequality (2) holds near the singularities of bounded multivariate rational functions.

**Theorem 10.** Suppose \( E \) is a bounded multivariate rational function on \( \mathbb{R}^m \). For any point \( p \in \mathbb{R}^m \) and any direction \( d \in O_p \setminus \{0\} \), where \( O_p \) is as in **Definition 9**, then...
there exist an open set $U_\infty \subset \mathbb{R}^m$ and constants $\theta \in (0, \frac{1}{2}]$ and $k > 0$ such that for all $x \in \text{domain}(E) \cap U_\infty$,

$$\left| E(x) - \lim_{s \to 0} E(p + sd) \right|^{1-\theta} \leq k \|\nabla E(x)\|.$$  

Further, $U_\infty$ may be chosen so that

1. there exists some $s \in (0, 1]$ for which $p + sd \in U_\infty$, and
2. for all $y \in U_\infty$ and $t \in [-1, 0) \cup (0, 1]$, the set $U_\infty$ contains $p + t(y - p)$.

Proof. If $E$ is nowhere defined, the theorem holds vacuously. We assume that $E$ is somewhere defined.

Consider lines in parameter space, beginning at $p$ and with direction $d \in \mathbb{R}^m$: $p + td, \ t \in \mathbb{R}$.

$E(p + td)$ is a univariate bounded real rational function, with respect to $t$. As such, it either exists nowhere or the limit $\lim_{s \to t} E(p + sd)$ exists for all $t \in \mathbb{R}$. If the limit exists, it is a univariate real rational function defined everywhere on $\mathbb{R}$, and thus it is everywhere real-analytic.

By Lemma 6, there exists some open subset $O \subset \mathbb{R}^m$ of full measure on which there are defined rational functions $c_n : O \to \mathbb{R}$ such that for all $y \in O$,

$$\lim_{s \to t} E(p + sy) = \sum_{n=0}^{\infty} c_n(y)t^n$$

if $|t| < \rho_y$. This may be selected such that $O = O_p$, as in Definition 9. By Lemma 8, select $\rho_y$ to depend continuously on $y$.

Select a reference direction $d \in O$. We may assume, by rescaling, that $\|d\| = 1$.

The parameter space is a finite-dimensional Banach space and $O$ is open, so there exist a compact set $K$ and an open neighborhood $U$ of $d$ such that

$$d \in U \subset K \subset O.$$  

$\rho_x$ is continuous on $K$ and on $U$. Let $M = \min_{x \in K} \rho_x$. Because $K$ is compact, $M$ is well-defined. Further, $\rho_x > 0$ by definition, so $M > 0$. Define

$$(6) \quad U_d = \left\{ \begin{bmatrix} y \\ t \end{bmatrix} : y \in U \text{ and } |t| < M \right\}.$$  

Define the extended function $F$ by

$$(7) \quad F(y, t) = \lim_{s \to t} E(p + sy) = \sum_{n=0}^{\infty} c_n(y)t^n.$$  

Within $U_d$, (7) is a convergent sum of analytic functions, and thus (7) is itself a real-analytic function. By Theorem 1 there exist an open ball $V \subset U_d$ centered at $\begin{bmatrix} d \\ 0 \end{bmatrix}$ and constants $\theta \in (0, \frac{1}{2}]$ and $k > 0$, such that for all $y \in V$,

$$(8) \quad |F(y, t) - F(d, 0)|^{1-\theta} \leq \frac{k}{2} \|\nabla F(y, t)\|.$$
Define the open set $V_* \subset V$ by

$$V_* = \left\{ \begin{bmatrix} y \\ t \end{bmatrix} \in V : \|y - d\| < \frac{1}{2} \quad \text{and} \quad 0 < |t| < \frac{1}{2} \right\}.$$  

From the definition of $V$ as an open ball centered at $\begin{bmatrix} d \\ 0 \end{bmatrix}$, we conclude that if $\begin{bmatrix} y \\ t \end{bmatrix} \in V_*$ then $\begin{bmatrix} y - t \\ 0 \end{bmatrix} \in V_*$. For any $\begin{bmatrix} y \\ t \end{bmatrix} \in V_*$, if $E(p + ty)$ exists then its gradient exists, and the gradient of $F$ is characterized as

$$\nabla F(y, t) = \begin{bmatrix} t \nabla E(p + ty) \\ (y, \nabla E(p + ty)) \end{bmatrix},$$  

where $\nabla E(p + ty)$ is the gradient of $E$ at the point $p + ty$. Decompose $\nabla F(y, t) = a + b$, where

$$a = \begin{bmatrix} 0 \\ (y, \nabla E(p + ty)) \end{bmatrix}, \quad \text{and} \quad b = t \begin{bmatrix} \nabla E(p + ty) \\ 0 \end{bmatrix}.$$  

By the Cauchy-Schwartz inequality and definition of $\| \cdot \|$,  

$$\|a\| \leq \|y\| \|\nabla E(p + ty)\|, \quad \text{and} \quad \|b\| = |t|\|\nabla E(p + ty)\|.$$  

Using the triangle inequality, relate $\|\nabla F(y, t)\|$ and $\|\nabla E(p + ty)\|$ by

$$\|\nabla F(y, t)\| \leq \|a\| + \|b\|$$

$$\leq \|y\| \|\nabla E(p + ty)\| + |t|\|\nabla E(p + ty)\|$$

$$= (\|y\| + |t|)\|\nabla E(p + ty)\|.$$  

Note that by definition of $V_*$, $|t| < \frac{1}{2}$ and $\|y\| < \|d\| + \frac{1}{2} = \frac{3}{2}$, so

$$\|\nabla F(y, t)\| \leq 2\|\nabla E(p + ty)\|.$$  

We now establish (5) for points $p + ty$ using (8):

$$\left| E(p + ty) - \lim_{s \to 0} E(p + sd) \right|^{1-\theta} = |F(y, t) - F(d, 0)|^{1-\theta}$$

$$\leq \frac{k}{2} \|\nabla F(y, t)\|$$

$$\leq k \|\nabla E(p + ty)\|. \quad (9)$$  

Because this holds, with identical constants, for all $\begin{bmatrix} y \\ t \end{bmatrix} \in V_*$, we may establish (5) on a subset of $\mathbb{R}^m$. Define the cone $U_{\beta_1}$ by

$$U_{\beta_1} = \left\{ p + ty \mid \begin{bmatrix} y \\ t \end{bmatrix} \in V_* \right\} \subset \mathbb{R}^m.$$
Note that $U_\infty$ is open and satisfies the requirements of the theorem. By (9), for any $x \in U_\infty$,
\[
\left| E(x) - \lim_{s \to 0} E(p + sd) \right|^{1-\theta} \leq k\|\nabla E(x)\|.
\]
This completes the proof.

Though we have shown that a generalized Lojasiewicz gradient inequality holds in cones near the discontinuities of bounded rational functions, it is important to note that these cones are not neighborhoods of the discontinuities. Before this theorem may be used to show convergence, it must be shown that a sequence approaching a discontinuity remains within a single cone, or within a finite union of cones.

2.2. A Lojasiewicz inequality for sequences approaching singularities.

This section proves conditions under which sequences remain within sets such as those produced by Theorem 10. This is broken into Lemmas 11 and 12 and Theorem 13.

Lemma 11 shows that a continuous function on a product space may be used as a funnel, guiding a sequence into an open subset of one of its factor spaces, such as a set provided by Theorem 10. If one set provided by Theorem 10 is insufficient to capture the full behavior of a sequence, Lemma 11 allows one to form the union of multiple such sets. Finally, Theorem 13 constructs a set on which the generalized Lojasiewicz gradient inequality holds, and provides conditions under which a sequence will be fummeled into the constructed set.

**Lemma 11.** Let $U_1, \ldots, U_m$ be subsets of $\mathbb{R}^n$, and let the function $f : \bigcup_{i=1}^m U_i \to \mathbb{R}$ be bounded and differentiable. If there exist constants $L \in \mathbb{R}$, $\theta_1, \ldots, \theta_m \in (0, \frac{1}{2}]$, and $c_1, \ldots, c_m \in \mathbb{R}$ such that, for all $x \in U_i$,
\[
|f(x) - L|^{1-\theta_i} \leq c_i\|\nabla f(x)\|,
\]
then there exist $\theta_* \in (0, \frac{1}{2})$ and $c_* \in \mathbb{R}$ such that, for all $x \in \bigcup_{i=1}^m U_i$,
\[
|f(x) - L|^{1-\theta_*} \leq c_*\|\nabla f(x)\|.
\]

**Proof.** Let $M = \sup_{x \in \bigcup_{i=1}^m U_i} |f(x) - L|$. Because $f$ is bounded, $M < \infty$. Let $\theta_* = \min_i \theta_i$. Then, for any $x \in U_i$,
\[
|f(x) - L|^{1-\theta_i} \leq M^{\theta_i-\theta_*} |f(x) - L|^{1-\theta_i} \leq M^{\theta_i-\theta_*} c_i\|\nabla f(x)\|.
\]
Selecting $c_* = \max_i (M^{\theta_i-\theta_*} c_i)$ completes the proof.

In the proof of Theorem 13, we will parameterize elements of a sequence $\{x_k\}_{k=1}^\infty$ as pairs of directions $x_k/\|x_k\|$ and distances $\|x_k\|$. Using this parameterization, we must show that if $\|x_k\|$ is sufficiently small, then the direction $x_k/\|x_k\|$ must be within a specified set. That argument is greatly simplified by Lemma 12, which states formally an intuitive property of continuous functions and direct products of compact sets illustrated by Figure 2.

**Lemma 12.** Given metric spaces $X_1, X_2$, compact sets $K_1 \subset X_1$ and $K_2 \subset X_2$, and a continuous function $f : X_1 \times X_2 \to \mathbb{R}$, if $f|_{K_1 \times K_2} > 0$, then there exists $\epsilon > 0$ such that, for each $x_1 \in X_1$ and $x_2 \in X_2$, at least one of the statements
1. $f(x_1, x_2) < \epsilon$,
2. $d(x_1, K_1) < \epsilon$,
3. $d(x_2, K_2) < \epsilon$
Lemma can be applied only to finitely many cones, and these cones for all sufficiently large $K$ theorem, such that $f$ union is also compact. The closures of the sets in this subcover are compact, by the Heine-Borel property, so their union is also compact.

The set $N^C = (X_1 \times X_2 \setminus \left(\bigcup_{i=1}^n N_{q_i,r_i}^{x_i,y_i}\right)$ is closed, so $\epsilon_1 = \frac{1}{2} d_\infty(N^C,K_1 \times K_2)$ is positive. Define $K_{\epsilon_1} = \{(x_1,x_2) \in X_1 \times X_2 | d_\infty((x_1,x_2),K_1 \times K_2) \leq \epsilon_1\}$, and note that $K_{\epsilon_1} \subseteq \bigcup_{i=1}^n N_{q_i,r_i}^{x_i,y_i}$. As a closed subset of the union of the closures of $\{N_{q_i,r_i}^{x_i,y_i}\}_{i=1}^n$, the set $K_{\epsilon_1}$ is compact. Thus $f$ attains a lower bound $\epsilon_2 > 0$ on $K_{\epsilon_1}$. Define $\epsilon = \min\{\epsilon_1, \epsilon_2\}$.

Given a point $(x_1,x_2) \in X_1 \times X_2$, the statements $d(x_1,K_1) < \epsilon$ and $d(x_2,K_2) < \epsilon$ together imply $d_\infty((x_1,x_2),K_1 \times K_2) < \epsilon_1$. This in turn implies $(x_1,x_2) \in K_{\epsilon_1}$, so $f(x_1,x_2) \geq \epsilon_2 \geq \epsilon$.

With the lemmas above, we can formalize a statement that certain sequences are funneled into cones on which the Łojasiewicz inequality holds. Though we have a statement that the Łojasiewicz inequality holds on cones adjacent to essential singularities, infinitely many such cones may be required to cover a deleted neighborhood of a given point. Lemma 11 can be applied only to finitely many cones, and these cones must have an identical limiting value $L$. We can ensure a finite union by imposing a condition which induces a compact set covered by cones, such as condition (10) below.

**Theorem 13.** Let $f$ be a bounded multivariate rational function on $\mathbb{R}^n$, and let $\{x_k\}_{k=1}^\infty \subset \mathbb{R}^n$ be a sequence such that $\{f(x_k)\}_{k=1}^\infty$ is monotonic. Suppose that $x^*$ is a cluster point of the sequence $\{x_k\}_{k=1}^\infty$ and there exists a closed set $V \subset \mathcal{O}_x$ such that for all sufficiently large $k$,

$$\frac{x_k - x^*}{\|x_k - x^*\|} \in V.$$

Then there exist an open neighborhood $U$ of $x^*$ and constants $\theta \in (0,1/2]$ and $c \in \mathbb{R}$ such that if $k$ is sufficiently large and if $x_k \in U$, then

$$\left|f(x_k) - \lim_{i \to \infty} f(x_i)\right|^{1-\theta} \leq c\|\nabla f(x_k)\|.$$
Proof. If $f(x^*)$ is defined, then $f$ is real-analytic in an open neighborhood of $x^*$, and the conclusion of the theorem follows directly from the Lojasiewicz gradient inequality.

Assume instead that $f(x^*)$ is undefined. Then $x_k \neq x^*$ for all $k \in \mathbb{N}$.

The sequence $\{f(x_k)\}_{k=1}^{\infty}$ is monotonic and bounded, so it admits a limit

$$ L = \lim_{k \to \infty} f(x_k). $$

Without loss of generality, we may assume that $x^* = 0$ and that every $u \in V$ has $\|u\| = 1$. Define $F(x, t) = \lim_{s \to t} f(sx)$ and note that it is analytic on $O_{x^*} \times \mathbb{R}$.

Let $K \subset O_{x^*}$ be defined by

$$ K = V \cap \left\{ x \in \mathbb{R}^n \mid \lim_{z \to x} \inf_{y} |F(z, 0) - L| = 0 \right\}. $$

As an intersection of a compact set and a closed set, $K$ is compact. To show that $K$ is non-empty, take a subsequence $\{x_k\}_{k=1}^{\infty}$ such that $x_k \to x^*$, and note that $\{(x_k, \|x_k\|, \|x_k\|)\}_{k=1}^{\infty}$ admits a cluster point $(u^*, 0)$, for some $u^* \in V$. Because $F$ is continuous, $F(u^*, 0) = \lim_{i \to \infty} F(x_k, \|x_k\|, \|x_k\|) = L$, and thus $u^* \in K$. Further, we may use continuity of $F$ to calculate the image $F[K \times \{0\}] = \{L\}$.

Using Theorem 10, create collections

$$ U_u \subset \mathbb{R}^n, \quad \Theta_u \in (0, \frac{1}{2}], \quad \text{and} \quad C_u \in \mathbb{R} $$

such that, for every $u \in K$,

$$ y \in U_u \implies |f(y) - L|^{1-\Theta_u} \leq C_u \|\nabla f(y)\|. $$

By their definition, each $U_u$ admits an open subset $V_u \subset U_u$ and a real number $\epsilon_{1,u} \in (0, 1]$ such that

$$ 0 < |t| \leq \epsilon_{1,u} \wedge y \in V_u \implies t \frac{y}{\|y\|} \in V_u, $$

and such that there exists $t > 0$ such that $tu \in V_u$. For each $u \in K$, let $V'_u = \left\{ \frac{t}{\epsilon_{1,u}} y \mid y \in V_u \right\}$. Note that $\|u\| = 1$ by assumption on $V$, and thus that $\frac{t}{\epsilon_{1,u}} \in V_u$; it follows that $u \in V'_u$. The collection $\{V'_u\}_{u \in K}$ is then an open cover of $K$, and thus admits a finite subcover $\{V'_{u_i}\}_{i=1}^{m}$, where $\{u_i\}_{i=1}^{m} \subset K$. Let

$$ \epsilon_1 = \min_{i=1, \ldots, m} \epsilon_{1,u_i}, $$

$$ U = \bigcup_{i=1}^{m} U_{u_i}. $$

By Lemma 11, there exist $\theta \in (0, 1/2]$ and $c \in \mathbb{R}$ such that

$$ y \in U \implies |f(y) - L|^{1-\theta} \leq c \|\nabla f(y)\|. $$

Note that $\epsilon_1 K = \{ \epsilon_1 u \mid u \in K \} \subset \bigcup_{i=1}^{m} V_{u_i} \subset U$, and that the distance between a compact set and a closed set is positive if the sets are disjoint. Thus, we may define

$$ \delta = \frac{d(\epsilon_1 K, \mathbb{R}^n \setminus U)}{\epsilon_1} > 0. $$
Recalling that \( u \in K \Rightarrow \|u\| = 1 \), and applying \((11)\) and \((12)\) gives that if
\[
\|x_k\| < \epsilon_1 \text{ and } d(x_k/\|x_k\|, K) < \delta,
\]
then \( x_k \in U \).

Let \( V_\delta = \{ u \in V | \|u\| = 1 \wedge d(u, K) \geq \delta \} \). By \((10)\),
\[
\frac{x_k}{\|x_k\|} \notin V_\delta \iff d\left(\frac{x_k}{\|x_k\|}, K\right) < \delta.
\]

\( V_\delta \) is compact, and \(|F - L| > 0\) on \( V_\delta \times \{0\} \). By Lemma \(12\), there exists \( \epsilon_2 > 0 \) such that \(\|x_k\| < \epsilon_2\) and \(|F\left(\frac{x_k}{\|x_k\|}, \|x_k\|\right) - L| < \epsilon_2\) together imply \(\frac{x_k}{\|x_k\|} \notin V_\delta\). Note that \( F\left(\frac{x_k}{\|x_k\|}, \|x_k\|\right) = f(x_k)\).

If \( k \) is large enough that \(|f(x_k) - L| < \epsilon_2\) and if \(\|x_k\| < \min\{\epsilon_1, \epsilon_2\}\), then
\[
d\left(\frac{x_k}{\|x_k\|}, K\right) < \delta, \text{ and } x_k \in U.
\]
Selecting \( U = \{ v \in \mathbb{R}^n | \|v - x^*\| < \min\{\epsilon_1, \epsilon_2\}\} \) completes the proof.

With the possible exception of \((10)\), the conditions of Theorem \(13\) impose no great burden. Any sequence produced by a hill-climbing algorithm will necessarily have monotonic \(\{f(x_k)\}_{k=1}^\infty\).

**Proposition 14.** The set \( \mathcal{O}_{x^*} \) defined in Definition \(9\) contains every cluster point of the sequence \(\left\{\frac{x_k - x^*}{\|x_k - x^*\|}\right\}_{k=1}^\infty\) if and only if there exists a closed set \( V \subset \mathcal{O}_{x^*}\) such that \(\frac{x_k - x^*}{\|x_k - x^*\|} \in V \) for all sufficiently large \( k \).

**Proof.** If the set \( V \) exists, then the set \(\{ u \in V | \|u\| = 1 \} \) is compact, so the “if” direction holds.

If no such set \( V \) exists, then some subsequence \(\{u_i\}_{i=1}^\infty = \left\{\frac{x_{i_k} - x^*}{\|x_{i_k} - x^*\|}\right\}_{i=1}^\infty\) must approach the compact set \( K = \{ u \in \mathbb{R}^n | \|u\| = 1 \} \setminus \mathcal{O}_{x^*}\). That is \( d(u_i, K) \to 0\). This may be shown by contradiction; if \( d(u_i, K) \neq 0\), a set of points \( c\)-distant from \( K \) would fulfill the requirements of \( V \), which would contradict the assumption that no such set exists. Define a sequence of closest points in the complement of \( \mathcal{O}_{x^*}\) by selecting, for each \( i \), a point
\[
k_i \in \{ k \in K | d(u_i, k_i) = d(u_i, K) \}
\]
The sequence \(\{k_i\}_{i=1}^\infty\) admits a cluster point, and \( d(u_i, k_i) \to 0\), so the sequence \(\{u_i\}_{i=1}^\infty\) also admits a cluster point \( u^* \in K\). By definition of \( K\), \( u^* \notin \mathcal{O}_{x^*}\), so the “only if” direction holds.

### 2.3. New convergence theorems

In the previous sections, we have established that a generalized Lojasiewicz gradient inequality holds on cones and established conditions under which sequences are funneled into those cones. We now combine this with Theorems \(2\) and \(3\) to establish conditions under which sequences converge to essential singularities of bounded multivariate rational functions.

We say that a sequence \(\{x_k\}_{k=1}^\infty \subset \mathbb{R}^n\) satisfies Assumption \((A4)\) if
- the set \( \mathcal{O}_{x^*} \) in Definition \(9\) contains every cluster point of the sequence

\[
\left\{\frac{x_k - x^*}{\|x_k - x^*\|}\right\}_{k \in \mathbb{N}}.
\]

**Theorem 15.** Let \( f \) be a bounded multivariate rational function on \( \mathbb{R}^n \), and let \(\{x_k\}_{k=1}^\infty \subset \mathbb{R}^n\) be a sequence of vectors with a cluster point \( x^* \). If Assumptions \((A1)\), \((A2)\), and \((A4)\) hold on the tail of the sequence, then \( x^* \) is the limit of \(\{x_k\}_{k=1}^\infty\).
Proof. Assumption (A1) guarantees that the sequence \( \{ f(x_k) \}_{k \in \mathbb{N}} \) is decreasing. Assumption (A4) and Proposition 14 then fulfill the conditions of Theorem 13. This fulfills condition (L*) of Theorem 2.

The conditions of Theorem 2 are thus satisfied, so the conclusions of Theorem 15 hold.

The proof of Theorem 15 merely uses Theorem 13 to show that the conditions of Theorem 2 hold, so the same argument provides the following specialization of Theorem 3.

**Theorem 16.** Under the conditions of Theorem 15, if Assumption (A3) holds, then \( \nabla f(x_k) \to 0 \) and the convergence rate may be estimated as

\[
\| x^* - x_k \| = \left\{ \begin{array}{ll}
O(k^{\frac{q}{q}}) & \text{if } \theta = \frac{1}{2} \\
O(k^{-\frac{q}{q}}) & \text{if } 0 < \theta < \frac{1}{2}
\end{array} \right.
\]

for some \( 0 < q < 1 \).

where \( \theta \) is such that \( (L_*) \) holds.

**Proof.** Follows immediately from the proof of Theorem 15 and from the conditions of Theorem 3.

These theorems can be used to show convergence of sequences to cluster points not in a function’s domain, but both theorems rely on assumption (A4), which may be difficult to verify a priori.

3. Algorithms, examples, and implications. Uchmajew et al. have shown that (A1), (A2), and (A3) hold for the sequences produced by various optimization algorithms [14, 11, 7]. The reader should note, however, that these results may rely on additional restrictions on the sequence or on the objective function, which may preclude application to sequences approaching singularities of an objective function.

Notably, in the analysis of the Gradient-Related Projection Method with Line-Search (GRPMLS) from [11], the assumption (A0) required by Corollary 2.9 in [11] need only hold on the sequence itself, and thus (A1) and (A2) hold even if the sequence diverges or has a cluster point not in the domain of the objective function. Unfortunately, the additional conditions that Uchmajew used to prove that (A3) holds, in particular that \( \nabla f \) must be Lipschitz continuous on a neighborhood of \( x^* \), do not typically hold if \( x^* \) is an essential singularity.

If the gradient-related projection method with line-search is applied to a bounded multivariate rational function then the sequence produced satisfies (A1) and (A2). One must still guarantee the existence of cluster points, and show that (A4) holds.

3.1. Example: **Figure 1.** Figure 1 illustrates a sequence that maximizes the rational function defined by \( (x, y) \mapsto \frac{-xy}{(x^2 + y^2)(1 + x^2 + y^2)} \). We will employ the results of this paper to establish its convergence. More precisely, we will examine the equivalent problem of minimizing \( f(x, y) = \frac{xy}{(x^2 + y^2)(1 + x^2 + y^2)} \).

The sequence is produced by the method of [11], and thus satisfies (A1) and (A2).

Any cluster point of the sequence will occur either in the domain of \( f \) or at its singularity — that is, at \( 0 \). To guarantee convergence in the latter case, we establish that \( \mathcal{O}_{x^*} = \mathbb{R}^n \setminus \{0\} \). As established in Definition 9, we may examine the Maclaurin series coefficients of the denominator of \( (x, y, t) \mapsto f(tx, ty) \). The zeroth and first Maclaurin series coefficients are identically \( 0 \). The second Maclaurin series coefficient is \( f_2(x, y) = \frac{d^2}{dt^2} \big|_{t=0} (x^2 + y^2)(1 + t^2(x^2 + y^2)) \big| = 2(x^2 + y^2) \), which is nonzero on \( \mathbb{R}^n \setminus \{0\} \). Thus \( \mathcal{O}_{x^*} = \mathbb{R}^n \setminus \{0\} \), and (A4) holds regardless of the direction from which the sequence approaches.
Noting that the degree of the denominator \((x^2 + y^2)(1 + x^2 + y^2)\) exceeds that of the numerator \(xy\) and that the sequence of images \(f(x_k) = f(x_k, y_k)\) is negative and decreasing ensures that \(\|x_k\|\) remains bounded. Thus a cluster point exists. This cluster point must be the limit of the sequence. If it is in the domain of \(f\), convergence would be guaranteed by the results of [11]. Otherwise, convergence would be guaranteed by Theorem 15.

In this example, it is also trivial to establish that the sequence converges to \(0\). If the sequence were to converge to some \(x^*\) in the domain of \(f\), then [11, Corrollary 2.11] and continuity would provide \(\nabla f(x^*) = 0\). Calculation gives \(\nabla f(x, y) \neq 0\), so this case does not occur.

Though this trivial example could be solved using alternate methods, the results of this paper provide more interesting results when a function admits multiple singularities. As an example of this, the function \(f(x, y) = \sum_{i=1}^j w_i f(x - u_i, y - v_i)\), for constants \(w_i, u_i, v_i \in \mathbb{R}\) and \(i = 1, \ldots, j\), admits multiple singularities \(x_i^* = (u_i, v_i)\), each of which has \(O_{x_i^*} = \mathbb{R}^n \setminus \{0\}\). Any sequence produced by GRPMLS would be precluded from cycling among the singularities by the results of this paper.

3.2. Convergence in direction. If a sequence admits a cluster point, one may ask whether this cluster point is the limit of the sequence. If a sequence does not admit a cluster point, one may instead ask whether there exists a related convergent sequence that approximates a solution of a corresponding problem. In this subsection we examine a class of problems for which such related convergent sequences may be produced.

If a sequence fails to converge because it is unbounded, one may optimize instead a homogeneous function of degree 0, such as that defined by

\[
\tilde{f}(x) = \min_{\alpha \in \mathbb{R}} f(\alpha x).
\]

A sequence that optimizes \(\tilde{f}\) may be paired with a sequence of scalar constants to describe a sequence that optimizes the original objective function \(f\). In many cases, this is significantly more difficult to analyze than the original objective function \(f\), in part because \(\tilde{f}\) may introduce singularities. For some notable problems, such as low-rank approximation of tensors, \(\tilde{f}\) is a bounded multivariate rational function, and is thus within the scope of this paper.

Optimizing \(\tilde{f}\) may still produce a divergent sequence \(\{x_i\}_{i \in \mathbb{N}}\). To address this, consider instead the normalized sequence \(\{x_i/\|x_i\|\}_{i \in \mathbb{N}}\), which also optimizes \(\tilde{f}\). This normalized sequence must admit a cluster point, by compactness of the unit ball in a finite-dimensional Banach space. We will show that if Assumption (A1) holds on \(\{x_i\}_{i \in \mathbb{N}}\), it holds also on the normalized sequence \(\{x_i/\|x_i\|\}_{i \in \mathbb{N}}\).

**Lemma 17.** Let \(V\) be a Hilbert space, and let \(u, v \in V\) be vectors such that \(\|u\| = 1\). Then

\[
\left\| u - \frac{v}{\|v\|} \right\| \leq 2\|u - v\|.
\]

**Proof.** Let \(k = \|v\|\), let \(w = v/k\), and let \(x = \text{Re}(u, w)\).

\[
\begin{align*}
\frac{1}{2} & (4\|u - kw\|^2 - \|u - w\|^2) \geq 0 \iff 2k^2 - 4kx + 1 + x \geq 0 \iff 2(k - x)^2 + 1 + x - 2x^2 \geq 0
\end{align*}
\]
If $\frac{1}{2} x \leq x \leq 1$, then $1 + x - 2x^2 \geq 0$, and (15) holds. If $x \leq 0$, then $(k - x)^2 \geq x^2$ by the choice of $k$, and (15) holds if $1 + x \geq 0$. By the Cauchy-Schwarz inequality, $x \in [-1, 1]$, so (15) holds.

**Proposition 18.** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ has the property that $f(x) = f(cx)$ for all $c \in \mathbb{R}$. Suppose also that \{x_i\}$_{i \in \mathbb{N}}$ is a sequence such that Assumption (A1) holds. Then Assumption (A1) holds also for the sequence \{x_i/\|x_i\|\}$_{i \in \mathbb{N}}$.

**Proof.** For clarity, let $u_i = x_i/\|x_i\|$. By Assumption (A1), there is some $\sigma > 0$ such that

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

It is easily verified, by definitions of the derivative and $f$, that

$$\nabla f(u_i) = \|x_i\|\nabla f(x_i).$$

Further, by Lemma 17,

$$\|u_i - u_{i+1}\| \leq 2 \left\| u_i - \frac{x_i + 1}{\|x_i\|} \right\| = 2 \frac{\|x_i - x_{i+1}\|}{\|x_i\|}. $$

Thus,

$$\frac{\sigma}{2} \|\nabla f(u_i)\| \|u_i - u_{i+1}\| = \frac{\sigma}{2} \|\nabla f(x_i)\| \|x_i\| \|u_i - u_{i+1}\| \leq \sigma \|\nabla f(x_i)\| \|x_i - x_{i+1}\| \leq f(x_i) - f(x_{i+1}) = f(u_i) - f(u_{i+1}),$$

So Assumption (A1) holds for the sequence \{u_i\}$_{i \in \mathbb{N}}$.

**3.3. Implications for tensor approximation.** Creation of low-rank approximations of tensors, specifically when approximating a tensor using two or more separable tensors, is prone to produce divergent sequences of parameters. We make no effort to establish convergence of such a sequence. Instead, we consider the convergence of the sequence of normalized parameters, as in **subsection 3.2**.

Let $\tau(x)$ be a multilinear map from parameters $x \in \mathbb{R}^n$ to tensors, and let $\mathcal{T}$ be the tensor one wishes to approximate. If one selects the usual objective function $f(x) = \|\tau(x) - \mathcal{T}\|^2$, one obtains a polynomial function on $\mathbb{R}^n$. If instead one selects $\hat{f}(x) = \|\frac{\tau(x)}{\|\tau(x)\|} - \mathcal{T}\|^2$, one obtains a bounded multivariate rational function with denominator $\|\tau(x)\|^4$. Moreover, $\hat{f}$ is a homogeneous function of degree 0, and $\hat{f}(x) = \min_{a \in \mathbb{R}} f(ax)$ wherever $\hat{f}$ is defined.

A full analysis of the Maclaurin series coefficients $\frac{d^m}{dx^m} \|\tau(x + t \mathbf{d})\|^4$ would be beyond the scope of this paper. We can, however, draw conclusions from established properties of the set $\mathcal{O}_{\mathbb{R}^n}$. In particular, $\mathbb{R}^n \setminus \mathcal{O}_{\mathbb{R}^n}$ is of Lebesgue measure 0, so it covers almost every direction of approach.

If one uses the GRPML method from [11] to optimize $\hat{f}$, and normalizes the resulting parameters, then it is likely that the resulting sequence of parameters will converge. That is, it is likely that the approximation’s summands will form a stable configuration, changing little except in scale.
4. Concluding remarks. We have shown that a generalized Lojasiewicz inequality holds on sets adjacent to essential singularities of bounded multivariate rational functions. With this result, we have provided sufficient conditions under which a sequence will converge to a singularity of the objective function. Finally, we have shown that these results may be employed in the study of unbounded sequences by converting to a projective space.

The results employed in this paper could potentially be extended to sequences for which Assumption (A4) does not hold. In particular, we expect that if a sequence is well-described by some smooth curve, that is, \( x_i = \gamma(1/i) \) for some smooth \( \gamma \) with \( \gamma(0) = x^* \), then a generalized Lojasiewicz inequality should hold on that sequence. This would, however, be of little practical utility due to the difficulty of showing that an algorithm produces sequences well-described by smooth curves.

Acknowledgement. This paper was completed with the guidance of Martin J. Mohlenkamp.

REFERENCES

[1] P.-A. Absil, R. Mahony, and B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, SIAM J. Optim., 16 (2005), pp. 531–547, https://doi.org/10.1137/040605266.

[2] G. Beylkin and M. J. Mohlenkamp, Algorithms for numerical analysis in high dimensions, SIAM Journal on Scientific Computing, 26 (2005), pp. 2133–2159, https://doi.org/10.1137/040604959, http://dx.doi.org/10.1137/040604959.

[3] S. Brzostowski, The Lojasiewicz exponent of semiquasihomogeneous singularities, Bulletin of the London Mathematical Society, 47 (2015), pp. 848–852, https://doi.org/10.1112/blms/bdv058.

[4] A. I. Cohen, Rate of convergence of several conjugate gradient algorithms, SIAM J. Numer. Anal., 9 (1972), pp. 248–259.

[5] M. Espig, W. Hackbusch, and A. Khachatryan, On the convergence of alternating least squares optimisation in tensor format representations, preprint on webpage at https://arxiv.org/abs/1506.00062.

[6] J. Gwoździewicz, The lojasiewicz exponent of an analytic function at an isolated zero, Commentarii Mathematici Helvetici, 74 (1999), pp. 364–375, https://doi.org/10.1007/s000140050094.

[7] Z. Li, A. Uschmajew, and S. Zhang, On convergence of the maximum block improvement method, SIAM J. Optim., 25 (2015), pp. 210–233, https://doi.org/10.1137/130939110.

[8] S. Lojasiewicz, Sur le problème de la division, Studia Mathematica, 18 (1959), pp. 87–136, http://eudml.org/doc/216949.

[9] S. Lojasiewicz, Sur la géométrie semi- et sous- analytique, Annales de l’institut Fourier, 43 (1993), pp. 1575–1595, http://eudml.org/doc/75048.

[10] G. Oleśkiewicz, The Lojasiewicz exponent of nondegenerate surface singularities, Acta Mathematica Hungarica, 138 (2013), pp. 179–199, https://doi.org/10.1007/s10474-012-0285-5.

[11] R. Schneider and A. Uschmajew, Convergence results for projected line-search methods on varieties of low-rank matrices via lojasiewicz inequality, SIAM J. Optim., 25 (2015), pp. 622–646, https://doi.org/10.1137/140957822.

[12] Z. Shi, G. Yang, and Y. Xiao, A limited memory BFGS algorithm for non-convex minimization with applications in matrix largest eigenvalue problem, Math. Methods Oper. Res., 83 (2016), pp. 243–264, https://doi.org/10.1007/s00186-015-0527-7, http://dx.doi.org/10.1007/s00186-015-0527-7.

[13] A. Uschmajew, Local convergence of the alternating least squares algorithm for canonical tensor approximation, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 639–652, https://doi.org/10.1137/110843587.

[14] A. Uschmajew, A new convergence proof for the higher-order power method and generalizations, Pac. J. Optim., 11 (2015), pp. 309–321, http://www.ybook.co.jp/
Appendix A. Full proof of Theorems 2 and 3.

Proof. Assumption (L*) implies that $\lim_{k \to \infty} f(x_k)$ exists. The gradient of $f$ is finite where defined, so assumption (L*) requires that $\lim_{k \to \infty} f(x_k)$ be finite.

A proof of the original theorems of Absil, Uschmajew, et al. is provided in [11, p. 644]. Because only three changes must be made to the original proof, I quote Uschmajew’s proof here as indented block quotes. Between block quotes, I supply the alterations which extend Uschmajew’s proof to a proof of Theorems 2 and 3. Let $f_k = f(x_k)$, and let $g_k^- = \|\nabla f(x_k)\|$. We can assume that $g_k^- > 0$ for all $k$ since otherwise the sequence becomes stationary and there is nothing to prove.

There will also be no loss of generality to assume that (A1) and (A2) hold for all $k$ and that $\lim_{k \to \infty} f_k = 0$. Then $0 \leq f_k$ for all $k$ and the Lojasiewicz gradient inequality at $x^*$ reads as

(A.1)  \[ f_k^{1-\theta} \leq \Lambda g_k^- \]

whenever $x_k \in U$. The set $U$ contains an open ball $B_{\delta}(x^*)$ for some $\delta > 0$.

Let $\epsilon \in (0, \delta]$, and assume $\|x_k - x^*\| < \delta$. Then, by (A.1) and (A1),

\[ \|x_k - x_{k+1}\| \leq \frac{\Lambda}{\sigma} f_k^{\theta-1} (f_k - f_{k+1}) . \]

Using the fact that for $\phi \in [f_{k+1}, f_k]$ there holds $f_k^{\theta-1} \leq \phi^{\theta-1} \leq f_{k+1}^{\theta-1}$, we can estimate

\[ f_k^{\theta-1} (f_k - f_{k+1}) \leq \int_{f_{k+1}}^{f_k} \phi^{\theta-1} \, d\phi = \frac{1}{\theta} (f_k^{\theta} - f_{k+1}^{\theta}) \]

and thus obtain

\[ \|x_k - x_{k+1}\| \leq \frac{\Lambda}{\sigma\theta} (f_k^{\theta} - f_{k+1}^{\theta}) . \]

More generally, let $\|x_j - x^*\| < \epsilon \leq \delta$ for $k \leq j < m$; we get by this argument that

(A.2)  \[ \|x_m - x_k\| \leq \sum_{j=k}^{m} \|x_{j+1} - x_j\| \leq \sum_{j=k}^{m} \frac{\Lambda}{\sigma\theta} (f_j^{\theta} - f_{j+1}^{\theta}) \leq \frac{\Lambda}{\sigma\theta} f_k^{\theta} . \]

Since $x^*$ is an accumulation point, and because $f_k \to 0$, we can pick $n$ so large that

\[ \|x_k - x^*\| < \frac{\epsilon}{2} \quad \text{and} \quad \frac{\Lambda}{\sigma\theta} f_k^{\theta} < \frac{\epsilon}{2} . \]

Then (A.2) inductively implies $\|x_m - x^*\| < \epsilon$ for all $m > k$. This proves that $x^*$ is the limit point of the sequence, and, by (A3), $g_k^- \to 0$.

\[1\] To simplify reading, I have changed Uschmajew’s notation to match mine.
To estimate the convergence rate, let

$$r_k = \sum_{j=k}^{\infty} \|x_{j+1} - x_j\|.$$

Then $$\|x_k - x^*\| \leq r_k$$, so it suffices to estimate the latter. By (A.2), (A.1), and (A3), there exists $$k_0 \geq 1$$ such that for $$k \geq k_0$$ it holds that

$$\frac{1}{r_k} \leq \left(\frac{\Lambda}{\sigma\theta}\right)^{\frac{1-\theta}{\theta}} \frac{\Lambda}{\kappa} \|x_{k+1} - x_k\| = \left(\frac{\Lambda}{\sigma\theta}\right)^{\frac{1-\theta}{\theta}} \frac{\Lambda}{\kappa} (r_k - r_{k+1}),$$

that is,

(A.3) $$r_{k+1} \leq r_k - \nu r_k^{\frac{1-\theta}{\theta}}$$

with $$\nu = \left(\frac{\Lambda}{\sigma\theta}\right)^{\frac{\theta-1}{\theta}} \frac{\kappa}{\Lambda}$$. Now, if $$\theta = 1/2$$, we get from (A.3) that $$\nu \in (0,1)$$, and

$$r_k \leq r_{k_0} (1 - \nu)^{k-k_0}$$

for $$k \geq k_0$$. The case $$0 \leq \theta < 1/2$$ is more delicate. We follow Levitt: put $$p = \frac{\theta}{\theta-2\theta}, C \geq \max\{\frac{\kappa}{p}, r_{k_0} k_0^{-p}\}$$, and $$s_k = C k^{-p}$$; then $$s_{k_0} \geq r_{k_0}$$, and

$$s_{k+1} = s_k (1+k^{-1})^{-p} \geq s_k (1-pk^{-1}) = s_k - \frac{p}{C^{1/p}} s_k^{\frac{p+1}{p}} \geq s_k - \nu s_k^{\frac{p}{p}}$$

(the first inequality holding by convexity of $$x^{-p}$$). Using induction, it now follows from (A.3) that $$r_k \leq s_k$$ for all $$k \geq k_0$$, which finishes the proof.