Highlights

The Łojasiewicz Gradient Inequality on the Complex Stiefel Manifold

Matthew M. Lin

- Using the realification isomorphism Ψ_1, we show that $\mathrm{St}(m,n)$ admits the structure of a real-analytic submanifold of $\mathbb{R}^{2mn}$ with dimension $2mn - n^2$.
- The classical Łojasiewicz gradient inequality is extended to $\mathrm{St}(mr,n)$ with the intrinsic Riemannian (Frobenius) gradient by pulling back through a real-analytic coordinate chart.
- The key transfer step uses the uniform positive-definiteness of the Gram matrix $G(u) = D\varphi(u)^\top D\varphi(u)$ on a compactly contained ball, ensuring $\lambda_{\max} < \infty$ and yielding $\|\nabla_u \tilde{f}\| \le \lambda_{\max}^{1/2} \|\operatorname{grad}_{\mathcal{M}} f\|_F$.
- The proof flow is summarized in the following diagram.

Goal

$|f(K) - f(K^\star)|^\theta \le c \,\|\operatorname{grad}_{\mathrm{St}} f(K)\|_F$

↓

Lemma 1 — Real-analytic submanifold

Realification via Ψ_1 realizes $\mathrm{St}(m,n)$ as a real-analytic submanifold of $\mathbb{R}^{2mn}$ of dimension $d := 2mn - n^2$

↓

Lemma 2 — Real-analytic coordinate map

The analytic implicit function theorem yields $\varphi(u) := (u, \phi(u)) : U \subset \mathbb{R}^d \to \Psi_1(\mathrm{St}(m,n))$, $\quad \mathrm{Im}(D\varphi(u^\star)) = T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))$

↓

Lemma 3 — Gradient correspondence

$\mathrm{Proj}_{T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))} \nabla(f \circ \Psi_1^{-1})(\Psi_1(K^\star)) = \Psi_1\big(\operatorname{grad}_{\mathrm{St}} f(K^\star)\big)$

↓

(a) Proof of Theorem 1 — Pulled-back function

$\tilde{f} := f \circ \Psi_1^{-1} \circ \varphi : U \to \mathbb{R}$ is real-analytic, $\quad \nabla_u \tilde{f}(u^\star) = 0$

↓

(b) Proof of Theorem 1 — Classical Łojasiewicz on $\mathbb{R}^d$

$|\tilde{f}(u) - f^\star|^\theta \le \tilde{c}\,\|\nabla_u \tilde{f}(u)\|$ for all $u \in \overline{B_r(u^\star)}$

↓

(c) Proof of Theorem 1 — Transfer via Gram matrix

$G(u) := D\varphi(u)^\top D\varphi(u) \preceq \lambda_{\max} I_d$ on $\overline{B_{r'}(u^\star)} \subsetneq U$, $\lambda_{\max} < \infty$ by compactness

$\Longrightarrow \quad \|\nabla_u \tilde{f}(u)\| \le \lambda_{\max}^{1/2} \|\operatorname{grad}_{\mathrm{St}} f(K(u))\|_F$

↓

Conclusion

$|f(K) - f(K^\star)|^\theta \le c\,\|\operatorname{grad}_{\mathrm{St}} f(K)\|_F$ for all $K \in \mathcal{U} = \Psi_1^{-1}(\varphi(B_\rho(u^\star)))$, $\quad c := \tilde{c}\,\lambda_{\max}^{1/2}$, $\quad \rho := \min(r, r'/2)$

The Łojasiewicz Gradient Inequality on the Complex Stiefel Manifold

Matthew M. Lin [a,*,1]

[a]*Department of Mathematics, National Cheng Kung University, Tainan, 701, Taiwan*

ARTICLE INFO

Keywords:
Łojasiewicz gradient inequality
Stiefel manifold
real-analytic submanifold
Riemannian gradient
Gram matrix

ABSTRACT

We establish the Łojasiewicz gradient inequality for the restriction of an arbitrary real-analytic function to the complex Stiefel manifold $\mathrm{St}(m, n)$, equipped with the intrinsic Frobenius Riemannian metric. Unlike many convergence results on manifolds, our inequality applies at every stationary point and requires no assumptions beyond the real analyticity of the objective function. To prove this result, we first show that $\mathrm{St}(m, n)$ has the structure of a real-analytic submanifold of $\mathbb{R}^{2mn}$. We then connect the Riemannian gradient to the Euclidean gradient through local real-analytic coordinates. This connection allows us to transfer the classical Łojasiewicz gradient inequality from Euclidean space to the manifold setting. The resulting inequality provides a valuable tool for analyzing convergence in optimization problems on complex Stiefel manifolds.

1. Introduction, motivation and presentation of the main result

Optimization over matrix manifolds has become a central theme in modern computational mathematics. Many problems arising in signal processing, machine learning, and quantum information involve optimizing objective functions whose feasible set is a Riemannian manifold, rather than an open subset of Euclidean space. Among the most commonly encountered constraint sets is the complex Stiefel manifold

$$\mathrm{St}(m, n) = \left\{ K \in \mathbb{C}^{m\times n} \;\middle|\; K^*K = I_n \right\},$$

which occurs when the decision variable must consist of orthonormal columns. This work addresses a fundamental analytical question for functions defined on Riemannian manifolds: whether the Łojasiewicz gradient inequality holds, and, if so, whether it can be expressed using the intrinsic Riemannian gradient.

The Stiefel manifold arises naturally as the feasible set in a diverse range of applications. In quantum information theory, stacking the Kraus operators of a trace-preserving quantum channel $\mathcal{E}(\rho) = \sum_{i=1}^{r} K_i \rho K_i^*$ into a single block matrix $K = [K_1^\top, \dots, K_r^\top]^\top \in \mathbb{C}^{mr\times n}$ yields the constraint $K^*K = I_n$, placing K on $\mathrm{St}(mr, n)$; recovering K from measurement data then constitutes quantum process tomography (Chuang and Nielsen, 1997; Nielsen and Chuang, 2010). In grant-free radio access for the Internet of Things, collision resolution via blind source separation imposes a signal orthogonality constraint that places the demixer matrix directly on the Stiefel manifold (Feres and Ding, 2022). Real Stiefel manifold optimization is likewise central to dimensionality reduction and subspace learning, including principal component analysis, linear discriminant analysis (Cunningham and Ghahramani, 2015; Wang, Zhang and Li, 2023; Jiang, Xiao and Liu, 2026), and hyperspectral unmixing (Bioucas-Dias, Plaza, Dobigeon, Parente, Du, Gader and Chanussot, 2012).

In all of the above settings, a major theoretical question is whether a gradient flow or iterative descent method converges to a single stationary point instead of oscillating along the manifold of stationary points. A fundamental tool for addressing this question is the Łojasiewicz gradient inequality. Let $\|\cdot\|_F$ represent the Frobenious norm. In the landmark work (Łojasiewicz, 1963), Łojasiewicz showed that for a real-analytic function $f : \mathbb{R}^N \to \mathbb{R}$, and for any stationary point $x^\star$, there exist $c > 0$ and $\theta \in (0, 1)$ such that

$$|f(x) - f(x^\star)|^\theta \;\le\; c\, \|\nabla f(x)\|_F.$$

*Corresponding author.
mhlin@mail.ncku.edu.tw (M.M. Lin)
ORCID(s): 0000-0002-8144-947X (M.M. Lin)

[1]The author received support from the National Cen- ter for Theoretical Sciences of Taiwan and the National Science and Technology Council through grant 112-2628-M-006-009-MY4.

for all x sufficiently close to $x^\star$. The extension to manifolds is, however, not automatic. The inequality in Łojasiewicz (1963) is formulated in Euclidean space. On a manifold $\mathcal{M}$, however, one seeks an inequality involving the intrinsic gradient $\|\operatorname{grad}_{\mathcal{M}} f\|F$, which is the orthogonal projection of the Euclidean gradient $\|\nabla f\|_F$ onto the tangent space. Since one always has

$$\|\operatorname{grad}_{\mathcal{M}} f\|_F \leq \|\nabla f\|_F.$$

An inequality involving the larger quantity $\|\nabla f\|_F$ does not necessarily imply one that pertains only to the smaller quantity $\|\operatorname{grad}_{\mathcal{M}} f\|_F$.

For the general case, this intrinsic gradient has been used in the convergence analysis of line-search methods (Schneider and Uschmajew, 2015, Definition 2.1) and Jacobi-G algorithms on the orthogonal group or unitary group Li, Usevich and Comon (2018); Usevich, Li and Comon (2020). In the real case, the original proof was established for the real-algebraic variety $\mathcal{M}_{\leq k}$ of matrices with rank at most k (Schneider and Uschmajew, 2015, Theorem 3.8). In the complex case, the proof was obtained at stationary points satisfying the Morse-Bott property is satisfied (Usevich et al., 2020, Proposition 6.8), yielding a quadratic form of the Łojasiewicz inequality with $\theta = \frac{1}{2}$.

Without additional assumption, the present paper provides a self-contained proof that the Łojasiewicz gradient inequality holds on $\mathrm{St}(m, n)$ with the intrinsic Frobenius (equivalently Riemannian) gradient. Our main result is sated as follows.

Theorem 1. *Let $f : \mathrm{St}(m, n) \to \mathbb{R}$ be real-analytic and let $K^\star \in \mathrm{St}(m, n)$ be a stationary point of $f|_{\mathrm{St}(m,n)}$. Then there exist $c > 0$, $\theta \in (0, 1)$, and a neighborhood $\mathcal{U}$ of $K^\star$ in $\mathrm{St}(m, n)$ such that*

$$\left|f(K) - f(K^\star)\right|^\theta \leq c \left\|\operatorname{grad}_{\mathrm{St}(mr,n)} f(K)\right\|_F \quad \textit{for all } K \in \mathcal{U}.$$

The proof proceeds in four steps: (1) we claim the existence of a real-analytic coordinate map to pull f back to a Euclidean domain; (2) we verify that the pulled-back function has a genuine critical point; (3) we apply the classical Łojasiewicz theorem to the pulled-back function in $\mathbb{R}^d$ (where $d = 2mrn - n^2$); and (4) we transfer the resulting bound from the coordinate gradient to the Riemannian gradient via the uniform positive-definiteness of the Gram matrix.

The paper is organized as follows. Section 2 proves the existence of local real-analytic maps and demonstrates that $\mathrm{St}(m, n)$ has the structure of a real-analytic submanifold of $\mathbb{R}^{2mn}$, with dimension $2mn - n^2$. Section 3 contains the main result and its complete proof.

2. Real analytic parameterization

The key tool for generalizing the classical Łojasiewicz theorem is to transfer the resulting bound from the Euclidean gradient to the Riemannian gradient. Let $\mathcal{H}_n$ denote the real vector space of $n \times n$ Hermitian matrices with $\dim_{\mathbb{R}} \mathcal{H}_n = n^2$, and $\operatorname{vec}(\cdot)$ stack the columns of a matrix into a vector. Define $\Psi_1 : \mathbb{C}^{m\times n} \to \mathbb{R}^{2mn}$ and $\Psi_2 : \mathcal{H}_n \to \mathbb{R}^{n^2}$ by

$$\begin{cases} \Psi_1(K) = \operatorname{vec}\left(\begin{bmatrix} K_{\mathcal{R}} \\ K_{\mathcal{I}} \end{bmatrix}\right), & \text{(1a)} \\ \Psi_2(S) = \left[s_{11}, \ldots, s_{nn}, \operatorname{Re}(s_{12}), \operatorname{Im}(s_{12}), \ldots, \operatorname{Re}(s_{n-1,n}), \operatorname{Im}(s_{n-1,n})\right]^\top, & \text{(1b)} \end{cases}$$

where $K = K_{\mathcal{R}} + \imath K_{\mathcal{I}} \in \mathbb{C}^{m\times n}$, with $K_{\mathcal{R}}, K_{\mathcal{I}} \in \mathbb{R}^{m\times n}$ denoting the real and imaginary parts of K, respectively, and $S = [s_{ij}] \in \mathcal{H}_n$ with $s_{ij} = \overline{s_{ji}}$. Consider the complex constraint map

$$\Phi : \mathbb{C}^{m\times n} \longrightarrow \mathcal{H}_n, \qquad \Phi(K) := K^* K - I_n. \tag{2}$$

Then Φ is real-analytic (polynomial in real and imaginary parts of K), and $\mathrm{St}(m, n) = \Phi^{-1}(\{0\})$. Correspondingly, define the real-analytic map

$$\widetilde{\Phi} : \mathbb{R}^{2mn} \longrightarrow \mathbb{R}^{n^2}, \quad \widetilde{\Phi}\big(\Psi_1(K)\big) := \left[\begin{array}{c} \left\{\sum_\ell [(K_{\mathcal{R}})_{\ell i}(K_{\mathcal{R}})_{\ell j} + (K_{\mathcal{I}})_{\ell i}(K_{\mathcal{I}})_{\ell j}] - \delta_{ij}\right\}_{i\leq j} \\ \left\{\sum_\ell [(K_{\mathcal{R}})_{\ell i}(K_{\mathcal{I}})_{\ell j} - (K_{\mathcal{I}})_{\ell i}(K_{\mathcal{R}})_{\ell j}]\right\}_{i<j} \end{array}\right], \tag{3}$$

satisfying the commutative diagram

$$\begin{tikzcd} \mathbb{C}^{m\times n} \arrow[r, "\Phi"] \arrow[d, "\Psi_1"'] & \mathcal{H}_n \arrow[d, "\Psi_2"] \\ \mathbb{R}^{2mn} \arrow[r, "\widetilde{\Phi}"] & \mathbb{R}^{n^2} \end{tikzcd} \tag{4}$$

i.e., $\widetilde{\Phi} = \Psi_2 \circ \Phi \circ \Psi_1^{-1}$ and $\mathrm{St}(m, n) = \Psi_1^{-1}(\{\widetilde{\Phi} = 0\})$.

Lemma 2 below shows that at every point $\Psi_1(K^\star)$, the analytic implicit function theorem applied to the polynomial constraint map Φ produces a real-analytic chart $\varphi(u) = (u,\ \phi(u))$, one real-analytic submanifold chart for $\Psi_1(\mathrm{St}(m, n))$.

Lemma 2. *For every $K^\star \in \mathrm{St}(m, n)$, there exist an open set $U \subset \mathbb{R}^d$, where $d = 2mn - n^2$, and a real-analytic map $\phi : U \to \mathbb{R}^{n^2}$ such that $\varphi(u) := (u, \phi(u))$ defines a real-analytic bijection from U onto a neighborhood of $\Psi_1(K^\star)$ in $\Psi_1\big(\mathrm{St}(m, n)\big)$, with inverse $\varphi^{-1}(u, \phi(u)) = u$. In particular, $\Psi_1\big(\mathrm{St}(m, n)\big)$ is a real-analytic submanifold of $\mathbb{R}^{2mn}$ of dimension d.*

Proof. Let $K \in \mathrm{St}(m, n)$. For any $\Delta K \in \mathbb{C}^{m\times n}$, the directional derivative of Φ at K is

$$D\Phi(K)[\Delta K] = \frac{d}{dt}\Phi(K + t\Delta K)\Big|_{t=0} = \Delta K^* K + K^* \Delta K \in \mathcal{H}_n.$$

Observe that for all $S \in \mathcal{H}_n$, we have

$$D\Phi(K)\big[\tfrac{1}{2} KS\big] = \tfrac{1}{2} S^*(K^*K) + \tfrac{1}{2}(K^*K)S = S$$

while choosing $\Delta K = \frac{1}{2} KS$. Hence $D\Phi(K)$ is surjective for every $K \in \mathrm{St}(m, n)$). From the commutative diagram (4) and the chain rule, it follows that

$$\begin{aligned} D\widetilde{\Phi} &= \Psi_2 \circ D\Phi(K) \circ \Psi_1^{-1}, \\ \widetilde{\Phi}(\Psi_1(K)) &= (\Psi_2 \circ \Phi \circ \Psi_1^{-1})\big(\Psi_1(K)\big) = 0. \end{aligned}$$

Since Ψ_1 and Ψ_2 are real-linear isomorphisms, the surjectivity of $D\Phi(K)$ is equivalent to the surjectivity of $D\widetilde{\Phi}\big(\Psi_1(K)\big)$. Consequently, $D\widetilde{\Phi}$ has full row rank on $\Psi_1\big(\mathrm{St}(m, n)\big)$. By the implicit function theorem, for every $\Psi_1(K^\star) \in \Psi_1\big(\mathrm{St}(m, n)\big)$ there exist an open set $U \subset \mathbb{R}^d$ and a unique real-analytic map $\phi : U \to \mathbb{R}^{n^2}$ such that

$$\varphi : U \to \Psi_1\big(\mathrm{St}(m, n)\big), \qquad \varphi(u) = (u, \phi(u)),$$

is a real-analytic bijection in a neighborhood of $\Psi_1(K^\star)$ in $\Psi_1\big(\mathrm{St}(m, n)\big)$ with the inverse given by $\varphi^{-1}(u, \phi(u)) = u$. Furthermore, since $\Psi_1(\mathrm{St}(m, n)) = \widetilde{\Phi}^{-1}(0)$, the submersion theorem (see Proposition 3.3.3 in Absil, Mahony and Sepulchre (2008)) implies that $\Psi_1\big(\mathrm{St}(m, n)\big)$ is a real-analytic submanifold of $\mathbb{R}^{2mn}$ of dimension d. ☐

By Lemma 2, for every $K^\star \in \mathrm{St}(m, n)$, a neighborhood of $\Psi_1(K^\star)$ in $\Psi_1(\mathrm{St}(m, n))$ admits a real-analytic parameterization $u \mapsto (u, \phi(u))$. The following lemma characterizes the image of $D\phi(u^\star)$ geometrically.

Lemma 3. *Let $K^\star \in \mathrm{St}(m, n)$, and let $\varphi(u) := (u, \phi(u)) : U \subset \mathbb{R}^d \to \Psi_1(\mathrm{St}(m, n))$ be the local real-analytic parametrization provided by Lemma 2 with $d = 2mn - n^2$ and $\varphi(u^\star) = \Psi_1(K^\star)$. Then*

$$\mathrm{Im}\big(D\varphi(u^\star)\big) = T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m, n)). \tag{5}$$

Proof. Let $w = D\varphi(u^\star)\dot{u}$ for some $\dot{u} \in \mathbb{R}^d$ and define $\gamma(t) := \varphi(u^\star + t\dot{u})$. Then $\dot{\gamma}(0) = w$ and $\widetilde{\Phi}(\gamma(t)) = 0$ for all t small enough. Differentiating at $t = 0$ gives $D\widetilde{\Phi}\big(\Psi_1(K^\star)\big)w = 0$. Hence $w \in \ker D\widetilde{\Phi}\big(\Psi_1(K^\star)\big) = T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m, n))$, showing that $\mathrm{Im}(D\varphi(u^\star)) \subseteq T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m, n))$.

Since $\varphi(u) = (u, \phi(u))$, differentiating each component at $u = u^\star$ yields

$$D\varphi(u^\star) = \begin{bmatrix} I_d \\ D\phi(u^\star) \end{bmatrix} \in \mathbb{R}^{2mn\times d}, \tag{6}$$

where $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix and $D\phi(u^\star) \in \mathbb{R}^{n^2\times d}$ is the Jacobian of $\phi : U \to V$ at $u = u^\star$. Now suppose $D\varphi(u^\star)\,\dot{u} = 0$ for some $\dot{u} \in \mathbb{R}^d$. From (6), we see that

$$D\varphi(u^\star)\,\dot{u} = \begin{bmatrix} \dot{u} \\ D\phi(u^\star)\,\dot{u} \end{bmatrix} = 0 \in \mathbb{R}^N.$$

From the first d rows, we immediately obtain $\dot{u} = 0$, and therefore $\ker D\varphi(u^\star) = \{0\}$. By the rank-nullity theorem $\dim(\mathrm{Im}(D\varphi(u^\star))) = d - \dim(\ker D\varphi(u^\star)) = d$. Since $\dim\big(T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))\big) = d$, it follows that $\mathrm{Im}(D\varphi(u^\star)) = T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))$. □

Also, from the proof of Lemma 2, we can conclude that $\mathrm{St}(m,n)$ is an embedded submanifold of the space $\mathbb{C}^{m\times n}$. For each $K^\star \in \mathrm{St}(m,n)$, we equip the tangent space $T_{K^\star}\,\mathrm{St}(m,n)$ with the embedded Euclidean Riemannian metric and its induced norm (Edelman, Arias and Smith, 1998):

$$\langle X, Y\rangle_{K^\star} := \mathrm{Re}\left(\mathrm{tr}(X^*Y)\right), \qquad \|X\|_{K^\star} := \sqrt{\langle X, X\rangle_{K^\star}} = \|X\|_F, \qquad X, Y \in T_{K^\star}\,\mathrm{St}(m,n). \tag{7}$$

Furthermore, identifying complex matrices with their real representations through the map Ψ_1, the metric in (7) can be expressed as

$$\langle X, Y\rangle_{K^\star} = \sum_{i,j}\left[(X_{\mathcal{R}})_{ij}(Y_{\mathcal{R}})_{ij} + (X_{\mathcal{I}})_{ij}(Y_{\mathcal{I}})_{ij}\right] = \langle \Psi_1(X), \Psi_1(Y)\rangle_F, \tag{8}$$

where $\langle A, B\rangle_F := \mathrm{tr}(A^\top B)$ denotes the Frobenius inner product in the associated real matrix space.

The next lemma shows that the gradient on $\mathrm{St}(m,n)$ is naturally induced by the projected gradient on $\Psi_1(\mathrm{St}(m,n))$ through the $\mathbb{R}$-linear map Ψ_1.

Lemma 4. *Let $f : \mathrm{St}(m,n) \to \mathbb{R}$ be differentiable, and let $K^\star \in \mathrm{St}(m,n)$. Then*

$$\mathrm{Proj}_{T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))}\,\nabla(f\circ\Psi_1^{-1})(\Psi_1(K^\star)) = \Psi_1\left(\mathrm{grad}_{\mathrm{St}}\, f(K^\star)\right). \tag{9}$$

Proof. By (1a), $\Psi_1 : \mathbb{C}^{m\times n} \to \mathbb{R}^N$ is $\mathbb{R}$-linear, hence $D\Psi_1 = \Psi_1$. For any smooth curve $\gamma : (-\varepsilon, \varepsilon) \to \mathrm{St}(m,n)$ with $\gamma(0) = K^\star$, letting $\sigma = \Psi_1\circ\gamma$ yields $\sigma(0) = \Psi_1(K^\star)$ and $\dot{\sigma}(0) = \frac{d}{dt}\Psi_1(\gamma(t))\Big|_{t=0} = \Psi_1(\dot{\gamma}(0))$. Since $\gamma \mapsto \Psi_1\circ\gamma$ is bijective, it follows that $\Psi_1\left(T_{K^\star}\,\mathrm{St}(m,n)\right) = T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))$.

For any $\eta \in T_{K^\star}\,\mathrm{St}(m,n)$, the $\mathbb{R}$-linearity of Ψ_1 gives $\Psi_1(K^\star + t\eta) = \Psi_1(K^\star) + t\Psi_1(\eta)$, so:

$$Df(K^\star)[\eta] = \frac{d}{dt}f(K^\star + t\eta)\Big|_{t=0} = \frac{d}{dt}(f\circ\Psi_1^{-1})\big(\Psi_1(K^\star + t\eta)\big)\Big|_{t=0} = \nabla(f\circ\Psi_1^{-1})(\Psi_1(K^\star))[\Psi_1(\eta)].$$

Also, for any $\eta \in T_{K^\star}\,\mathrm{St}(m,n)$, we have

$$\left\langle \nabla(f\circ\Psi_1^{-1})(\Psi_1(K^\star)),\, \Psi_1(\eta)\right\rangle_F = Df(K^\star)[\eta] = \left\langle \mathrm{grad}_{\mathrm{St}}\, f(K^\star),\, \eta\right\rangle_{K^\star} = \left\langle \Psi_1(\mathrm{grad}_{\mathrm{St}}\, f(K^\star)),\, \Psi_1(\eta)\right\rangle_F.$$

Since $\Psi_1(\eta)$ ranges over all of $T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))$ as η ranges over $T_{K^\star}\,\mathrm{St}(m,n)$, this shows that

$$\left[\nabla(f\circ\Psi_1^{-1})(\Psi_1(K^\star)) - \Psi_1(\mathrm{grad}_{\mathrm{St}}\, f(K^\star))\right] \perp T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n)),$$

which completes the proof by the uniqueness of orthogonal projection. □

3. The Łojasiewicz Inequality on $\mathrm{St}(mr, n)$

Having established the relationship between $\mathrm{St}(m,n)$ and $\mathbb{R}^{2mn}$ (Lemma 2), we now establish the Łojasiewicz gradient inequality for the restriction of f to $\mathrm{St}(m,n)$, with the Riemannian gradient on the right-hand side. The argument proceeds by pulling back f to a Euclidean coordinate domain through a real-analytic parameterization, applying the classical Łojasiewicz theorem there and then transferring the resulting inequality back to the manifold using the uniform positive definiteness of the induced Gram matrix.

Proof of Theorem 1. The proof consists of four steps. First, retaining the notation of Lemmas 2 and 3, suppose that $\varphi(u^\star) = \Psi_1(K^\star)$. Let $\tilde{f}$ denote the pullback of f under the parametrization $K(u) = \Psi_1^{-1}(\varphi(u))$, namely,

$$\tilde{f} : U \longrightarrow \mathbb{R}, \qquad \tilde{f}(u) := f\left(\Psi_1^{-1}(\varphi(u))\right) = f(K(u)),$$

Since f is real-analytic by assumption, φ is real-analytic, and Ψ_1^{-1} is real-analytic, the composition $\tilde{f} = f \circ \Psi_1^{-1} \circ \varphi$ is real-analytic on $U \subset \mathbb{R}^d$.

Second, we verify that $\nabla_u \tilde{f}(u^\star) = 0$. By the chain rule, for any direction $v \in \mathbb{R}^d$,

$$\nabla_u \tilde{f}(u^\star)^\top v = \frac{d}{dt}\tilde{f}(u^\star + tv)\Big|_{t=0} = \frac{d}{dt}(f \circ \Psi_1^{-1})\big(\varphi(u^\star + tv)\big)\Big|_{t=0} = \left\langle \nabla(f \circ \Psi_1^{-1})(\Psi_1(K^\star)), D\varphi(u^\star)v \right\rangle_F.$$

By Lemma 3, the vector $D\varphi(u^\star)\, v$ lies in $T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))$ for every $v \in \mathbb{R}^d$. Thus,

$$\nabla_u \tilde{f}(u^\star)^\top v = \left\langle \mathrm{Proj}_{T_{\Psi_1(K^\star)}\Psi_1(\mathrm{St}(m,n))} \nabla(f \circ \Psi_1^{-1})(\Psi_1(K^\star)), D\varphi(u^\star)v \right\rangle_F = \left\langle \Psi_1\left(\mathrm{grad}_{\mathrm{St}} f(K^\star)\right), D\varphi(u^\star)v \right\rangle_F = 0,$$

where the last equality follows from Lemma 4 together with the fact $\mathrm{grad}_{\mathrm{St}} f(K^\star) = 0$. Hence, $\nabla_u \tilde{f}(u^\star) = 0$, i.e., $u^\star$ is a stationary point of $\tilde{f}$ in the Euclidean sense.

Third, it follows from the classical Łojasiewicz gradient inequality Łojasiewicz (1963) that there exist $\tilde{c} > 0$, $\theta \in (0,1)$, and $r > 0$ with $\overline{B_r(u^\star)} \subset U$ such that

$$\left|\tilde{f}(u) - \tilde{f}(u^\star)\right|^\theta \;\leq\; \tilde{c}\,\|\nabla_u \tilde{f}(u)\| \qquad \text{for all } u \in \overline{B_r(u^\star)}. \tag{10}$$

Fourth, for any $K(u) = \Psi_1^{-1}(\varphi(u)) \in \mathrm{St}(m,n)$, the Riemannian gradient $\mathrm{grad}_{\mathrm{St}} f(K(u)) \in T_{K(u)}\,\mathrm{St}(m,n)$ is defined as the unique tangent vector satisfying

$$\left\langle \mathrm{grad}_{\mathrm{St}} f(K(u)),\, X \right\rangle_F = Df(K(u))[X] \quad \text{for all } X \in T_{K(u)}\,\mathrm{St}(m,n), \tag{11}$$

where $Df(K(u))[X] = \nabla(f \circ \Psi_1^{-1})(\varphi(u))[\Psi_1(X)]$ denotes the directional derivative of $f \circ \Psi_1^{-1}$ at $\varphi(u)$ along the direction $\Psi_1(X) \in \mathbb{R}^{2mn}$. By Lemma 3, there exist vectors $\dot{g}, \dot{x} \in \mathbb{R}^d$ such that

$$\mathrm{grad}_{\mathrm{St}} f(K(u)) = \Psi_1^{-1}\big(D\varphi(u)\dot{g}\big), \qquad X = \Psi_1^{-1}\big(D\varphi(u)\dot{x}\big).$$

Substituting these expressions into the left-hand side of (11) yields

$$\left\langle \mathrm{grad}_{\mathrm{St}} f(K(u)),\, X \right\rangle_F = \left\langle D\varphi(u)\dot{g},\, D\varphi(u)\dot{x} \right\rangle_F = \dot{g}^\top G(u)\dot{x}, \tag{12}$$

where $G(u) := D\varphi(u)^\top D\varphi(u) \succ 0$ is a symmetric positive definite matrix for all $u \in U$. Meanwhile, the right-hand side of (11) can be written as

$$Df(K(u))[X] = \nabla(f \circ \Psi_1^{-1})(\varphi(u))\big[D\varphi(u)\dot{x}\big] = \left\langle \nabla(f \circ \Psi_1^{-1})(\varphi(u)), D\varphi(u)\dot{x} \right\rangle_F = \nabla_u \tilde{f}(u)^\top \dot{x}. \tag{13}$$

Combining (13) with (12) and using (11), we obtain

$$\dot{g} = G(u)^{-1}\nabla_u \tilde{f}(u). \tag{14}$$

Substituting this expression (14) back into the representation of the Riemannian gradient gives

$$\mathrm{grad}_{\mathrm{St}} f(K(u)) = \Psi_1^{-1}\left(D\varphi(u)\, G(u)^{-1}\, \nabla_u \tilde{f}(u)\right). \tag{15}$$

By (8), the squared Riemannian norm of the gradient (15) is given by

$$\left\|\mathrm{grad}_{\mathrm{St}} f(K(u))\right\|_{K(u)}^2 = \nabla_u \tilde{f}(u)^\top G(u)^{-1}\nabla_u \tilde{f}(u). \tag{16}$$

Last, we have to bound $\|\nabla_u \tilde{f}(u)\|_F$ in terms of $\left\|\mathrm{grad}_{\mathrm{St}} f(K(u))\right\|_F$. For each fixed u, the matrix $G(u)$ is a symmetric positive definite. Hence, its largest eigenvalue satisfies $\lambda_{max}(u) = \|G(u)\|_2$. Since φ is real-analytic, the Gram matrix

$G(u) = D\varphi(u)^\top D\varphi(u)$ is continuous on U. Choose $r' > 0$ such that $\overline{B_{r'}(u^\star)} \subsetneq U$. Since $\overline{B_{r'}(u^\star)}$ is compact and the map $u \mapsto \lambda_{\max}(u) := \|G(u)\|_2$ is continuous, the extreme value theorem ensures a maximal finite constant

$$\lambda_{\max} := \max_{u \in \overline{B_{r'}(u^\star)}} \lambda_{\max}(u) < \infty, \tag{17}$$

so $G(u)^{-1} \geq \lambda_{\max}^{-1} I_d$ for all $u \in B_{r'}(u^\star)$. From (16), we have

$$\left\|\mathrm{grad}_{\mathrm{St}}\, f(K(u))\right\|_{K(u)} = \left(\nabla_u \tilde{f}(u)^\top G(u)^{-1} \nabla_u \tilde{f}(u)\right)^{1/2} \geq \lambda_{\max}^{-1/2} \left\|\nabla_u \tilde{f}(u)\right\|_F. \tag{18}$$

Let $\rho := \min(r, r'/2)$. For all $u \in B_\rho(u^\star)$, combining (10) and (18):

$$\left|f(K(u)) - f(K^\star)\right|^\theta = \left|\tilde{f}(u) - \tilde{f}(u^\star)\right|^\theta \leq \tilde{c}\left\|\nabla_u \tilde{f}(u)\right\|_F \leq \tilde{c}\,\lambda_{\max}^{1/2}\left\|\mathrm{grad}_{\mathrm{St}}\, f(K(u))\right\|_F.$$

Setting $c := \tilde{c}\,\lambda_{\max}^{1/2}$, it remains to identify the open neighborhood $\mathcal{U}$ of $K^\star$ in $\mathrm{St}(m,n)$. Since $B_\rho(u^\star) \subset U$ is open and φ is a homeomorphism onto its image, $\varphi(B_\rho(u^\star))$ is open in $\Psi_1(\mathrm{St}(m,n))$. Given that Ψ_1 is a homeomorphism, we can define

$$\mathcal{U} := \Psi_1^{-1}\left(\varphi(B_\rho(u^\star))\right) \subset \mathrm{St}(m,n)$$

as an open neighborhood of $K^\star = \Psi_1^{-1}(\varphi(u^\star))$ in $\mathrm{St}(m,n)$. Then, the inequality holds for all $K \in \mathcal{U}$, which completes the proof. □

References

Absil, P.A., Mahony, R., Sepulchre, R., 2008. Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ. URL: https://doi.org/10.1515/9781400830244, doi:10.1515/9781400830244. with a foreword by Paul Van Dooren.

Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., Chanussot, J., 2012. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5, 354–379. doi:10.1109/JSTARS.2012.2194696.

Chuang, I.L., Nielsen, M.A., 1997. Prescription for experimental determination of the dynamics of a quantum black box. Journal of Modern Optics 44, 2455–2467. URL: https://www.tandfonline.com/doi/abs/10.1080/09500349708231894, doi:10.1080/09500349708231894, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/09500349708231894.

Cunningham, J.P., Ghahramani, Z., 2015. Linear dimensionality reduction: Survey, insights, and generalizations. Journal of Machine Learning Research 16, 2859–2900. URL: http://jmlr.org/papers/v16/cunningham15a.html.

Edelman, A., Arias, T.A., Smith, S.T., 1998. The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications 20, 303–353. URL: https://doi.org/10.1137/S0895479895290954, doi:10.1137/S0895479895290954, arXiv:https://doi.org/10.1137/S0895479895290954.

Feres, C., Ding, Z., 2022. A Riemannian geometric approach to blind signal recovery for grant-free radio network access. IEEE Trans. Signal Process. 70, 1734–1748. URL: https://doi.org/10.1109/tsp.2022.3159988, doi:10.1109/tsp.2022.3159988.

Jiang, L., Xiao, N., Liu, X., 2026. A smooth locally exact penalty method for optimization problems over generalized stiefel manifolds. arXiv preprint arXiv:2602.05631 .

Li, J., Usevich, K., Comon, P., 2018. Globally convergent Jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization. SIAM J. Matrix Anal. Appl. 39, 1–22. URL: https://doi.org/10.1137/17M1116295, doi:10.1137/17M1116295.

Łojasiewicz, S., 1963. Une propriété topologique des sous-ensembles analytiques réels, in: Les Équations aux Dérivées Partielles (Paris, 1962). Éditions du Centre National de la Recherche Scientifique, Paris, pp. 87–89.

Nielsen, M.A., Chuang, I.L., 2010. Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press.

Schneider, R., Uschmajew, A., 2015. Convergence results for projected line-search methods on varieties of low-rank matrices via łojasiewicz inequality. SIAM J. Optim. 25, 622–646. URL: https://doi.org/10.1137/140957822, doi:10.1137/140957822.

Usevich, K., Li, J., Comon, P., 2020. Approximate matrix and tensor diagonalization by unitary transformations: convergence of Jacobi-type algorithms. SIAM J. Optim. 30, 2998–3028. URL: https://doi.org/10.1137/19M125950X, doi:10.1137/19M125950X.

Wang, L., Zhang, L.H., Li, R.C., 2023. Trace ratio optimization with an application to multi-view learning. Math. Program. 201, 97–131. URL: https://doi.org/10.1007/s10107-022-01900-w, doi:10.1007/s10107-022-01900-w.