

On Error Bounds for Rank-Constrained Affine Matrix Sets

Ruoning Chen^{a,b}, Defeng Sun^{b,1}, Liping Zhang^{a,2,*}

^a*Department of Mathematical Sciences, Tsinghua University, Beijing, China*

^b*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

Abstract

Rank-constrained matrix problems appear frequently across science and engineering. The convergence analysis of iterative algorithms developed for these problems often hinges on local error bounds, which correlate the distance to the feasible set with a measure of how much the constraints are violated. Foundational results in semi-algebraic geometry guarantee that such bounds exist, yet the associated exponents are generally not explicitly determined. This paper establishes a local Hölderian error bound with an explicit exponent for the canonical rank-constrained affine feasibility set. This paper proves that, on any compact set, the distance to the feasible set is bounded by a power of a natural residual function capturing violations in both the rank and affine constraints. The exponent in this bound is given explicitly in terms of the problem's dimensions. This provides a fundamental quantitative result on the geometry of the solution set, paving the way for the convergence analysis of a broad class of numerical methods.

Keywords: Error bound, Low-rank constraints, Łojasiewicz inequality

2020 MSC: 14P10, 49J53, 90C26

*Corresponding Author. *Email address:* lipingzhang@tsinghua.edu.cn

¹Defeng Sun is partially supported by the Research Center for Intelligent Operations Research and RGC Senior Research Fellow Scheme No. SRF52223-5S02.

²Liping Zhang is funded by the National Natural Sciences Foundation of China (Grant No. 12571323).

1. Introduction

The set of rank-constrained matrices satisfying a system of linear equations,

$$\mathcal{S} = \{X \in \mathbb{R}^{m \times n} \mid \mathcal{A}(X) = b, \text{rank}(X) \leq r\}, \quad (1)$$

forms the feasibility region for a vast number of problems in modern data analysis and engineering. Applications include signal and image processing [3], system identification and control [5], and collaborative filtering, where it appears in the well-known matrix completion problem [1]. Despite its importance, the set \mathcal{S} is notoriously challenging from a computational standpoint. The rank constraint introduces a non-convex, combinatorial structure, rendering the general feasibility problem NP-hard [13].

A central question in the analysis of optimization algorithms is the rate of convergence to a solution set. The theory of error bounds provides a powerful framework for addressing this question by relating the distance from an arbitrary point to the solution set (i.e., $\text{dist}(X, \mathcal{S})$) to the value of a residual function that measures constraint violation [8, 10]. For an iterative algorithm that generates a sequence $\{X_k\}$, an error bound of the form $\text{dist}(X_k, \mathcal{S}) \leq \kappa \cdot \text{residual}(X_k)^\tau$ can provide a direct path to establishing convergence rates, showing that driving the residual to zero forces the iterates to approach the feasible set \mathcal{S} .

The study of error bounds for problems involving low-rank matrices has primarily proceeded along algorithmic lines. A significant body of work has analyzed convex relaxations, where the rank constraint is replaced by a constraint on the nuclear norm [9]. For solutions derived from such methods, error bounds have been established, often under statistical assumptions on the data or structural properties of the linear map \mathcal{A} , such as the Restricted Isometry Property (RIP) [9]. These results, however, characterize the properties of solutions to an auxiliary problem rather than providing a direct geometric property of the original non-convex set (1).

For non-convex methods that operate directly on low-rank factorizations (i.e., $X = UV^T$), local error bounds and related geometric properties like the quadratic growth condition or the Polyak-Łojasiewicz (PL) inequality have been established. These results typically demonstrate that, near a solution, the objective landscape is well-behaved, which guarantees linear convergence for methods like gradient descent [6, 11]. This line of inquiry provides crucial insight into algorithmic performance but couches the analysis in terms of

a surrogate optimization landscape, not the geometry of the fundamental constraint set (1).

Our work departs from this algorithmic focus to study the intrinsic geometry of the set (1) itself. Since \mathcal{S} is a semi-algebraic set (defined by polynomial equalities and inequalities), the foundational work of Łojasiewicz guarantees the existence of a Hölderian error bound, often called a Łojasiewicz inequality [7]. This classic result ensures that for a point $X_0 \in \mathcal{S}$ and a residual function f , there exists a neighborhood of X_0 , a constant c , and an exponent $\tau \in (0, 1]$ such that $\text{dist}(X, \mathcal{S}) \leq c \cdot f(X)^\tau$ for all X in that neighborhood. While this establishes existence, the exponent τ is generally non-constructive and can be arbitrarily close to zero, providing a weak guarantee.

The primary contribution of this paper is to establish an explicit, local Hölderian error bound for the set (1). We consider a natural residual function that combines the affine and rank constraint violations:

$$f(X) := \sum_{i=n-r+1}^n \sigma_i^2(X) + \frac{1}{2} \|\mathcal{A}(X) - b\|^2,$$

where $\sigma_i(X)$ denotes the i -th singular value of X . Our main result, Proposition 3.6, shows that for any compact set \mathcal{K} , there exist constants $c > 0$ and $\tau > 0$, with τ explicitly given in terms of the problem dimensions, such that

$$c \cdot \text{dist}(X, \mathcal{S}) \leq f(X)^\tau \quad \text{for all } X \in \mathcal{K}.$$

This result provides a concrete, quantitative measure of the geometric regularity of the rank-constrained affine set. By providing an explicit exponent, our work moves beyond the abstract existence guarantees of classical semi-algebraic geometry and provides a foundational tool that can be used for the convergence analysis of any iterative method aiming to find a point in (1).

2. Notation and Preliminaries

Throughout this paper, we let $\mathbb{R}^{m \times n}$ denote the space of $m \times n$ real matrices. This space is equipped with the trace inner product $\langle X, Y \rangle := \text{trace}(X^T Y)$ and its induced norm, the Frobenius norm, denoted by $\|\cdot\|$. The singular values of a matrix $X \in \mathbb{R}^{m \times n}$ are denoted by $\sigma_1(X) \geq \sigma_2(X) \geq \dots \geq \sigma_{\min\{m, n\}}(X)$ in non-increasing order. The distance from a point X to a set $A \subset \mathbb{R}^{m \times n}$ is defined as $\text{dist}(X, A) := \inf_{Y \in A} \|X - Y\|$. The open ball centered at X with radius ϵ is denoted by $\mathbb{B}(X, \epsilon)$, and its closure is $\bar{\mathbb{B}}(X, \epsilon)$.

The open (closed) unit ball centered at the origin is denoted by \mathbb{B} ($\bar{\mathbb{B}}$). We use I_k to denote the $k \times k$ identity matrix. For a linear map $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^l$, its adjoint is denoted by \mathcal{A}^* .

Our analysis relies on tools from variational analysis [10]. For a lower semicontinuous function $\phi : \mathbb{R}^l \rightarrow \mathbb{R}$, the Fréchet subdifferential at $x \in \mathbb{R}^l$ is defined by

$$\hat{\partial}\phi(x) := \left\{ v \in \mathbb{R}^l \mid \liminf_{y \rightarrow x, y \neq x} \frac{\phi(y) - \phi(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

The limiting (or Mordukhovich) subdifferential, denoted by $\partial\phi(x)$, is defined as

$$\partial\phi(x) := \left\{ v \in \mathbb{R}^l \mid \exists x_k \xrightarrow{\phi} x, v_k \rightarrow v \text{ with } v_k \in \hat{\partial}\phi(x_k) \text{ for all } k \right\},$$

where $x_k \xrightarrow{\phi} x$ means $x_k \rightarrow x$ and $\phi(x_k) \rightarrow \phi(x)$. If $\bar{x} \in \mathbb{R}^l$ is a local minimum of ϕ , then $0 \in \hat{\partial}\phi(\bar{x})$ by definition. For our nonsmooth residual function $f(X)$, we define its minimal norm subgradient (or slope) as $m_f(X) := \text{dist}(0, \partial f(X))$.

A key concept is that of semi-algebraic sets and functions. A set $A \subset \mathbb{R}^d$ is called semi-algebraic if it can be represented as a finite union of sets defined by a finite number of polynomial equalities and inequalities. A function is semi-algebraic if its graph is a semi-algebraic set. The semi-algebraic property is fundamental because it guarantees the existence of a local error bound via the Łojasiewicz inequality. Our primary tool is an effective version of this inequality for polynomials, which provides an explicit exponent.

Lemma 2.1 ([2, Theorem 4.2]). *Let $P : \mathbb{R}^l \rightarrow \mathbb{R}$ be a polynomial of degree $\deg(P) \leq d$. Let $\bar{x} \in \mathbb{R}^l$ be a critical point, i.e., $\nabla P(\bar{x}) = 0$. Then for any $r_0 > 0$ there exists $\epsilon > 0$ and $c > 0$ such that*

$$\|\nabla P(x)\| \geq c|P(x) - P(\bar{x})|^{1-\tau}$$

for any $x \in \mathbb{B}(\bar{x}, r_0)$ with $|f(x)| < \epsilon$, where the Łojasiewicz exponent τ is given by $\tau = R(l, d)^{-1}$, with $R(l, d) := d(3d - 3)^{l-1}$.

For a polynomial of degree $d = 4$ in l variables, this exponent is $\tau = (4 \cdot 9^{l-1})^{-1}$.

3. Deriving the Error Bound

Without loss of generality, assume that $\mathcal{S} \neq \emptyset$, $\bar{X} \in \mathcal{S}$ and $m \geq n$. The following analysis mainly follows the analysis in [4]. Consider the function $g : \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times (n-r)}$, defined by

$$g(X, V) = \langle XV, XV \rangle + \frac{1}{2} \|\mathcal{A}(X) - b\|_F^2.$$

Then g is a polynomial in $n(m + n - r)$ variables of degree 4. Moreover, we have

$$\min_{V: V^\top V = I_{n-r}} g(X, V) = \sum_{i=n-r+1}^n \sigma_i^2(X) + \frac{1}{2} \|\mathcal{A}(X) - b\|_F^2 = f(X),$$

where $\sigma_1(X) \geq \sigma_2(X) \geq \dots \geq \sigma_n(X)$ are the singular values of X . Therefore, it is easy to know both f and g are semi-algebraic functions and S is a semi-algebraic set. For each $X \in \mathbb{R}^{m \times n}$, denote

$$E(X) = \{V \in \mathbb{R}^{n \times (n-r)} \mid g(X, V) = f(X), V^\top V = I_{n-r}\},$$

which is a nonempty and compact set. For simplicity, denote $\mathcal{O}_{n,r} = \{V \in \mathbb{R}^{n \times (n-r)} \mid V^\top V = I_{n-r}\}$. Then, we have the following lemma.

Lemma 3.1. *The set-valued mapping $E : \mathbb{R}^{m \times n} \rightrightarrows \mathbb{R}^{n \times (n-r)}$, $X \mapsto E(X)$ is locally Hölder stable, i.e., for any fixed $\bar{X} \in \mathbb{R}^{m \times n}$ and $\epsilon > 0$, there exist $c, \alpha > 0$ such that*

$$E(X) \subset E(\bar{X}) + c\|X - \bar{X}\|^\alpha \bar{\mathbb{B}} \quad \text{for all } X \in \mathbb{B}(\bar{X}, \epsilon).$$

Proof. We set the function $H(X, V) := |g(X, V) - f(X)| + \|V^\top V - I_{n-r}\|_F^2$. Then the function H is semi-algebraic and locally Lipschitz. Additionally, we have $E(X) = \{V \mid H(X, V) = 0\}$. Since $\mathcal{O}_{n,r}$ is a compact set, it follows from the Łojasiewicz inequality that there exist constants $c, \alpha > 0$ such that

$$c \cdot \text{dist}(V, E(\bar{X})) \leq H(\bar{X}, V)^\alpha \quad \text{for all } V \in \mathcal{O}_{n,r}.$$

On the other hand, since H is locally Lipschitz, it is therefore globally L -Lipschitz on the compact set $\mathbb{B}(\bar{X}, \epsilon) \times \mathcal{O}_{n,r}$ for some $L > 0$. Thus, for $X \in \mathbb{B}(\bar{X}, \epsilon)$ and $V \in E(X)$, we have

$$\begin{aligned} c \cdot \text{dist}(V, E(\bar{X})) &\leq H(\bar{X}, V)^\alpha = |H(\bar{X}, V) - H(X, V)|^\alpha \\ &\leq (L\|X - \bar{X}\|)^\alpha = L^\alpha \|X - \bar{X}\|^\alpha. \end{aligned}$$

This completes the proof. \square

From Lemma 3.1, the following proposition is obvious.

Proposition 3.2. *For each $\epsilon' > 0$ and fixed \bar{X} , there exists a $\epsilon < \epsilon'$ such that for all $X \in \mathbb{B}(\bar{X}, \epsilon)$ and all $V \in E(X)$, we have*

$$\text{dist}(V, E(\bar{X})) < \epsilon'.$$

Next, we investigate the relationship between the generalized differentials of $g(X, V)$ and $f(X)$. Since g is a continuously differentiable polynomial, it is natural to expect that certain structural properties of $g(X, V)$ can be transferred to $f(X)$.

Lemma 3.3. *For all $X \in \mathbb{R}^{m \times n}$ and $V \in E(X)$, the following statements hold:*

(i) $\hat{\partial}f(X) \subset \{\nabla_X g(X, V)\}$. Moreover,

$$\emptyset \neq \partial f(X) \subset \bigcup_{U \in E(X)} \{\nabla_X g(X, U)\} \quad \text{and} \quad m_f(X) \geq \inf_{U \in E(X)} \|\nabla_X g(X, U)\|.$$

(ii) $\|\nabla_V g(X, V)\| \leq 2f(X)$.

Proof. (i) Take arbitrary $Z \in \hat{\partial}f(X)$. By definition, for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$f(X + H) - f(X) - \langle H, Z \rangle \geq -\epsilon \|H\| \quad \text{for all } H \in \mathbb{B}(0, \delta).$$

We define the function

$$\phi(H) := g(X + H, V) - \langle Z, H \rangle + \epsilon \|H\|.$$

Then for all $H \in \mathbb{B}(0, \delta)$, we have

$$\begin{aligned} \phi(H) &\geq f(X + H) - \langle Z, H \rangle + \epsilon \|H\| \\ &\geq f(X) = g(X, V) = \phi(0). \end{aligned}$$

That is to say, 0 is the local minima of ϕ , therefore we have

$$0 \in \hat{\partial}\phi(0) \subseteq \nabla_X g(X, V) - Z + \epsilon \bar{\mathbb{B}}.$$

Thus, we have $\|\nabla_X g(X, V) - Z\| \leq \epsilon$, this holds for any $\epsilon > 0$. Then $Z = \nabla_X g(X, V)$, this holds for any $Z \in \hat{\partial}f(X)$, then $\hat{\partial}f(X) \subset \{\nabla_X g(X, V)\}$.

On the other hand, since f is semi-algebraic, it is therefore continuously differentiable on a semi-algebraic dense open set $\mathcal{U} \subset \mathbb{R}^{m \times n}$ from the Cell Decomposition Theorem [12, Theorem 4.2]. Then for all $X \in \mathcal{U}$, we have

$$\hat{\partial}f(X) = \partial f(X) = \{\nabla_X g(X, V)\}.$$

Since \mathcal{U} is dense, the function g is C^∞ , the multi-function $E(X)$ is compact-valued and is locally Hölder stable (Lemma 3.1), we conclude that

$$\emptyset \neq \partial f(X) \subset \bigcup_{U \in E(X)} \{\nabla_X g(X, U)\},$$

and the last inequality in (i) follows immediately.

(ii) We have $g(X, V) = f(X) = \min_{U^\top U = I_{n-r}} g(X, U)$. Then there exists a multiplier $Y \in I_{n-r}$ such that $X^\top X V = V Y$. Multiplying both sides of the equation on the left by V^\top , we obtain $Y = (X V)^\top X V$. Therefore,

$$\begin{aligned} \|\nabla_V g(X, V)\| &= \|2X^\top X V\| = 2\|V Y\| \\ &\leq \|V\|_2 \|Y\|_F = \|Y\|. \end{aligned} \tag{2}$$

Additionally, we have

$$2f(X) \geq 2\langle X V, X V \rangle = 2\langle V Y, V \rangle = 2\text{trace}(Y). \tag{3}$$

Since $Y = (X V)^\top X V$ is a positive semidefinite matrix, we have

$$\|Y\| = \sqrt{\text{trace}(Y^2)} \leq \text{trace}(Y).$$

Combined with (2) and (3), the proof is completed. \square

Having established a clear linkage between $f(X)$ and $g(X, V)$, our focus now shifts to deriving the error bound for g since it is a polynomial.

Proposition 3.4. *There exist constants $c > 0$ and $\epsilon' > 0$ such that*

$$\|\nabla g(X, V)\| \geq c \cdot g(X, V)^{1-\tau}$$

for all $X \in \mathbb{B}(\bar{X}, \epsilon')$ and $V \in \mathbb{R}^{n \times (n-r)}$ with $\text{dist}(V, E(\bar{X})) < \epsilon'$, where

$$\tau := \frac{1}{R(n(m+n-r), 4)} := \frac{1}{4 \cdot 9^{n(m+n-r)-1}}.$$

Proof. We know $f(\bar{X}) = 0$, and $E(\bar{X})$ is a compact set. For any $\bar{V} \in E(\bar{X})$, we have $g(\bar{X}, \bar{V}) = 0$ and $\nabla g(\bar{X}, \bar{V}) = 0$ by simple observation. Recall that g is a polynomial in $n(m+n-r)$ variables of degree $d = 4$. From the Łojasiewicz inequality (Lemma 2.1), we have constants $c > 0, \epsilon' > 0$ such that

$$\|\nabla g(X, V)\| \geq c \cdot g(X, V)^{1-\tau}$$

for all $X \in \mathbb{B}(\bar{X}, \epsilon')$ and $V \in \mathbb{B}(\bar{V}, \epsilon')$. The conclusion then follows easily from the compactness of $E(\bar{X})$. \square

Now we are ready for the Łojasiewicz inequality of the nonsmooth function f . Specifically, we have the following result.

Theorem 3.5. *There exist constants $c > 0$ and $\epsilon > 0$ such that*

$$m_f(X) \geq c \cdot f(X)^{1-\tau}$$

for all $X \in \mathbb{B}(\bar{X}, \epsilon)$, where $\tau := \frac{1}{R(n(m+n-r), 4)}$.

Proof. Let c, ϵ' , and $\epsilon < \epsilon'$ be the positive constants such that Propositions 3.2 and 3.4 hold. Take arbitrary $X \in \mathbb{B}(\bar{X}, \epsilon)$, we have $\text{dist}(E(X), E(\bar{X})) \leq \epsilon'$. Since $E(X)$ is compact, there exists a $V \in E(X)$ such that

$$\|\nabla_X g(X, V)\| = \inf_{U \in E(X)} \|\nabla_X g(X, U)\|.$$

It follows from Lemma 3.3 that

$$m_f(X) \geq \|\nabla_X g(X, V)\| \quad \text{and} \quad \|\nabla_V g(X, V)\| \leq 2f(X).$$

Since all norms on finitely dimensional normed vector spaces are equivalent, we have

$$\begin{aligned} m_f(X) + 2f(X) &\geq \|\nabla_X g(X, V)\| + \|\nabla_V g(X, V)\| \\ &\geq c_1 \|\nabla g(X, V)\| \\ &\stackrel{(a)}{\geq} c_1 c \cdot g(X, V)^{1-\tau} = c_1 c \cdot f(X)^{1-\tau} \end{aligned}$$

for some c_1 only determined by m, n, r , and (a) comes from $\text{dist}(V, E(\bar{X})) < \epsilon'$ and Proposition 3.4. Note that $f(\bar{X}) = 0$, we can diminish ϵ until we may assume

$$f(X)^\tau < \frac{c_1 c}{4}, \quad \forall X \in \mathbb{B}(\bar{X}, \epsilon)$$

if necessary. Consequently, we obtain

$$m_f(X) \geq (c_1 c - 2f(X)^\tau) f(X)^{1-\tau} \geq \frac{c_1 c}{2} f(X)^{1-\tau}, \quad \forall X \in \mathbb{B}(\bar{X}, \epsilon).$$

This completes the proof. \square

The above Łojasiewicz inequality leads to an error bound of \mathcal{S} .

Proposition 3.6. *For any compact set $\mathcal{K} \subset \mathbb{R}^{m \times n}$, there exists a constant $c > 0$ such that for all $X \in \mathcal{K}$, the inequality*

$$c \cdot \text{dist}(X, \mathcal{S}) \leq f(X)^\tau$$

holds with $\tau := \frac{1}{R(n(m+n-r), 4)}$.

Proof. This can be derived by Theorem 3.5, [4, Lemma 2.2], and the compactness of \mathcal{K} . \square

The potential looseness of the error bound in Proposition 3.6 is a direct consequence of its reliance on the Łojasiewicz exponent for polynomials, whose tightness still remains unknown. When the function f satisfies certain regularity condition, a global bound can be established in the following. The proof of Theorem 3.8 is omitted as it follows the same arguments used for [4, Proposition 6.1].

Assumption 3.7. *For the function f , there exist constants $c > 0$ and $R > 0$ such that for all $\|X\| \geq R$, the inequality*

$$m_f(X) \geq c$$

holds.

Theorem 3.8. *If Assumption 3.7 holds, then the set \mathcal{S} is compact and there exists a constant $c > 0$ such that for any $X \in \mathbb{R}^{m \times n}$, the inequality*

$$c \cdot \text{dist}(X, \mathcal{S}) \leq f(X)^\tau + f(X)$$

holds with $\tau := \frac{1}{R(n(m+n-r), 4)}$.

4. Conclusion

In this paper, we constructed a concrete error bound for rank-constrained affine matrix sets. Our approach involved first defining a polynomial auxiliary function and establishing a key relationship between it and the problem's residual function. By applying the Łojasiewicz inequality and its known exponent for polynomials, we then derived an explicit form for the error bound. We acknowledge that a limitation of this current result is its looseness. Therefore, future work will focus on identifying and verifying sufficient conditions under which a tighter, more practical error bound can be obtained. This remains a primary direction for our further research.

References

- [1] E. Candès, B. Recht, Exact matrix completion via convex optimization, *Commun. ACM* 55 (2012) 111–119.
- [2] D. D'Acunto, K. Kurdyka, Explicit bounds for the Łojasiewicz exponent in the gradient inequality for polynomials, *Ann. Polon. Math.* 87 (2005) 51–61.
- [3] M.A. Davenport, J. Romberg, An overview of low-rank matrix recovery from incomplete observations, *IEEE J. Sel. Top. Signal Process.* 10 (2016) 608–622.
- [4] S.-T. Dinh, T.-S. Phạm, Łojasiewicz inequalities with explicit exponents for smallest singular value functions, *J. Complexity* 41 (2017) 58–71.
- [5] M. Fazel, H. Hindi, S.P. Boyd, A rank minimization heuristic with application to minimum order system approximation, In: *Proceedings of the American Control Conference*, vol. 6, IEEE, 2001, pp. 4734–4739.
- [6] P. Jain, P. Netrapalli, S. Sanghavi, Low-rank matrix completion using alternating minimization, In: *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, 2013, pp. 665–674.
- [7] J. Bochnak, M. Coste, M.-F. Roy, *Real Algebraic and Analytic Geometry*, Springer, Berlin, 1998.
- [8] J.-S. Pang, Error bounds in mathematical programming, *Math. Program.* 79 (1997) 299–332.

- [9] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, *SIAM Rev.* 52 (2010) 471–501.
- [10] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
- [11] J. Sun, Q. Qu, J. Wright, A geometric analysis of phase retrieval, *Found. Comput. Math.* 18 (2018) 1131–1198.
- [12] L. Van den Dries, C. Miller, Geometric categories and o-minimal structures, *Duke Math. J.* 84 (1996) 497–540.
- [13] L. Vandenberghe, S. Boyd, Semidefinite programming, *SIAM Rev.* 38 (1996) 49–95.