

COMPLEX TENSORS ALMOST ALWAYS HAVE BEST LOW-RANK APPROXIMATIONS

YANG QI, MATEUSZ MICHAŁEK, AND LEK-HENG LIM

ABSTRACT. Low-rank tensor approximations are plagued by a well-known problem — a tensor may fail to have a best rank- r approximation. Over \mathbb{R} , it is known that such failures can occur with positive probability, sometimes with certainty: in $\mathbb{R}^{2 \times 2 \times 2}$, every tensor of rank 3 fails to have a best rank-2 approximation. We will show that while such failures still occur over \mathbb{C} , they happen with zero probability. In fact we establish a more general result with useful implications on recent scientific and engineering applications that rely on sparse and/or low-rank approximations: Let V be a complex vector space with a Hermitian inner product, and X be a closed irreducible complex analytic variety in V . Given any complex analytic subvariety $Z \subseteq X$ with $\dim Z < \dim X$, we prove that a general $p \in V$ has a unique best X -approximation $\pi_X(p)$ that does not lie in Z . In particular, it implies that over \mathbb{C} , any tensor almost always has a unique best rank- r approximation when r is less than the generic rank. Our result covers many other notions of tensor rank: symmetric rank, alternating rank, Chow rank, Segre–Veronese rank, Segre–Grassmann rank, Segre–Chow rank, Veronese–Grassmann rank, Veronese–Chow rank, Segre–Veronese–Grassmann rank, Segre–Veronese–Chow rank, and more — in all cases, a unique best rank- r approximation almost always exist. It applies also to block-terms approximations of tensors: for any r , a general tensor has a unique best r -block-terms approximations. When applied to sparse-plus-low-rank approximations, we obtain that for any given r and k , a general matrix has a unique best approximation by a sum of a rank- r matrix and a k -sparse matrix with a fixed sparsity pattern; this arises in, for example, estimation of covariance matrices of a Gaussian hidden variable model with k observed variables conditionally independent given r hidden variables.

1. INTRODUCTION

There are numerous problems in scientific and engineering applications that may ultimately be put in the following form: Given a real or complex-valued function $f : \Omega \rightarrow \mathbb{K}$ (with $\mathbb{K} = \mathbb{R}$ or \mathbb{C} respectively), find a *best approximation* of f by a sum of functions f_1, \dots, f_r with some special structure, i.e.,

$$(1) \quad \min_{f_1, \dots, f_r \in Y} \|f - (f_1 + \dots + f_r)\|,$$

where Y is a subset of functions possessing that special structure. The problem, called the *best r term approximation problem* or *best rank- r approximation problem* is ubiquitous by its generality and simplicity. A slight generalization involves

$$(2) \quad \min_{f_1, \dots, f_r \in Y_1; g_1, \dots, g_s \in Y_2} \|f - (f_1 + \dots + f_r + g_1 + \dots + g_s)\|,$$

where Y_1 and Y_2 denote two subsets of functions each with a different structure (e.g., this arises in so-called sparse plus low-rank approximations). More generally, the approximation could involve even more subsets Y_1, \dots, Y_k of functions, each capturing a different structure in the target function f .

In practice, if the problem is not already discrete, then it has to be discretized for the purpose of computations. This usually involves discretizing Ω (e.g., sampling points, triangulation, mesh

2010 *Mathematics Subject Classification.* 15A69, 41A50, 41A52, 41A65, 97N50, 51M35.

Key words and phrases. Tensors, tensor ranks, best rank- r approximations, best k -term approximations, sparse-plus-low-rank approximations.

generation) or finding a finite-dimensional approximation of the function spaces (e.g., Galerkin method, quadrature) or both (e.g., collocation methods). The result of which is that we may in effect assume that Ω is a finite set or that the function space $L_{\mathbb{K}}^2(\Omega)$ is a finite-dimensional vector space (as our notation implies, we assume that the norm in (1) is an l^2 -norm). In which case, up to a choice of basis, $L_{\mathbb{K}}^2(\Omega) \cong \mathbb{K}^n$, where $n = \dim L_{\mathbb{K}}^2(\Omega)$ (or, if Ω is finite, $n = \#\Omega$).

The main issue with (1) is that in many common scenarios, the approximation problem in (1) does not have a solution: Let $f \in L_{\mathbb{K}}^2(\Omega)$ and $Y \subseteq L_{\mathbb{K}}^2(\Omega)$ be a closed subset (under the metric topology induced by $\|\cdot\|$), the infimum

$$\inf_{f_1, \dots, f_r \in Y} \|f - (f_1 + \dots + f_r)\|$$

is often not attainable when $r > 1$. The reason being that the set of r -term approximants

$$\Sigma_r^\circ(Y) := \{f \in L_{\mathbb{K}}^2(\Omega) \mid f = f_1 + \dots + f_r \text{ for some } f_1, \dots, f_r \in Y\}$$

is often not closed when $r > 1$. Let $\Sigma_r(Y)$ denote the closure of $\Sigma_r^\circ(Y)$ under the metric topology. Each element in $\Sigma_r(Y)$ is a limit of a sequence of r -term approximants but may not itself be an r -term approximant. While

$$\varphi_* \in \operatorname{argmin}_{\varphi \in \Sigma_r(Y)} \|f - \varphi\|$$

always exist, it would not in general be an r -term approximant, i.e., φ_* may fail to be of the required form $f_1 + \dots + f_r$ with $f_i \in Y$ — this happens when $\varphi_* \in \Sigma_2(Y) \setminus \Sigma_2^\circ(Y)$. The main result of this article is to show that it makes a vast difference whether $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ — for the latter, this failure almost never happens under some mild conditions.

We will elaborate this phenomenon more concretely by studying the case of separable approximations. In this case, $\Omega = \Omega_1 \times \dots \times \Omega_d$ and

$$(3) \quad Y = \{\varphi \in L_{\mathbb{K}}^2(\Omega) \mid \varphi = \varphi_1 \otimes \dots \otimes \varphi_d \text{ where } \varphi_j \in L_{\mathbb{K}}^2(\Omega_j)\}$$

is the set of separable functions, i.e., functions $\varphi : \Omega_1 \times \dots \times \Omega_d \rightarrow \mathbb{K}$ of the form

$$\varphi(x_1, x_2, \dots, x_d) = \varphi_1(x_1)\varphi_2(x_2) \dots \varphi_d(x_d), \quad x_j \in \Omega_j, \quad j = 1, \dots, d.$$

Take $d = 3$ for simplicity. When $\mathbb{K} = \mathbb{R}$, it has been shown [10, Theorem 8.4] that there is a nonempty open subset $\mathcal{O} \subseteq L_{\mathbb{R}}^2(\Omega)$ such that for any $f \in \mathcal{O}$, the infimum

$$\inf_{\varphi_i, \psi_i \in L_{\mathbb{R}}^2(\Omega_i)} \|f - \varphi_1 \otimes \varphi_2 \otimes \varphi_3 - \psi_1 \otimes \psi_2 \otimes \psi_3\|$$

fails to be attainable. In particular the set of such failures, given that it contains an open set, must have positive volume, i.e., such failures occur with positive probability and cannot be ignored in practice. Given any $f \in L_{\mathbb{K}}^2(\Omega)$, the *rank* of f is the integer r such that f is a sum of r separable functions, i.e.,

$$(4) \quad \operatorname{rank}(f) = \min \left\{ r \mid \sum_{i=1}^r \varphi_{1,i} \otimes \dots \otimes \varphi_{d,i} \right\},$$

and so $\Sigma_r^\circ(Y) = \{f \in L_{\mathbb{K}}^2(\Omega) \mid \operatorname{rank}(f) \leq r\}$. Therefore the preceding discussion says that the set of f that fails to have a best rank-two approximation has positive volume. In fact, there are more extreme examples [10, Theorem 8.1] where *every* $f \in L_{\mathbb{R}}^2(\Omega)$ of rank > 2 fails to have a best rank-two approximation. Our main result in this article will show that such failures are rare, in fact, they almost never happen, when $\mathbb{K} = \mathbb{C}$.

Up until this point, we have presented our discussions entirely in the language of function approximation in order to put the problems (1) and (2) within the context they most frequently arise, which traditionally goes under the heading of *nonlinear approximation*. Nevertheless, the subject has witnessed many recent breakthroughs and is now studied under a number of different names [2, 4, 6, 8, 11, 12, 19, 21, 22, 26]. In this article we will undertake an approach via complex algebraic geometry and real analytic geometry. From our perspective, the problems in (1) and (2)

are respectively about *secant varieties* and *join varieties*. We will henceforth drop the function approximation description in order to focus on the crux of the issue.

In the rest of this article, we will let U denote a real vector space and V denote a complex vector space. In geometric terms, let X be a closed semianalytic subset of an \mathbb{R} -vector space U , and $Z \subseteq X$ be a semianalytic subset of X with $\dim Z < \dim X$. The set Z here represents the ‘bad points’ to be avoided — given $p \in U$, we seek a best approximation $x_* \in X$ that attains $\min_{x \in X} \|p - x\|$ but we also want $x_* \notin Z$. The example we discussed above has $X = \Sigma_2(Y)$, $Z = \Sigma_2(Y) \setminus \Sigma_2^\circ(Y)$, and $f \in L_{\mathbb{K}}^2(\Omega)$ in the role of $p \in V$, with Y the set of separable functions as defined in (3). By this example, we see that there can be a nonempty open subset $\mathcal{O} \subseteq U$ such that for $p \in \mathcal{O}$, any best approximation of p lies in Z , which in this example represents the functions in X that cannot be written as a sum of two separable functions.

We shall show that this does not happen when $\mathbb{K} = \mathbb{C}$: If X is a closed irreducible complex analytic variety in a \mathbb{C} -vector space V , and Z is a complex analytic subvariety of X with $\dim Z < \dim X < \dim V$, then any general $p \in V$ will have its best approximation in X but not in Z , and moreover this best approximation will be unique. We will apply this to obtain a number of existence and uniqueness results for various nonlinear approximations common in applications.

2. SUBANALYTIC GEOMETRY

We will study our approximation problems using selected tools from subanalytic geometry [1], which we review in the following. A *semianalytic* subset of \mathbb{R}^n is a set that is locally of the form

$$X = \bigcup_{i=1}^m \bigcap_{j=1}^k X_{ij},$$

where each X_{ij} takes the form

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f_{ij}(x_1, \dots, x_n) = 0\} \quad \text{or} \quad \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid f_{ij}(x_1, \dots, x_n) > 0\}$$

for some real analytic function f_{ij} . Let W be a linear subspace of \mathbb{R}^n , and $\pi: \mathbb{R}^n \rightarrow W$ be a linear projection. A subset $X \subseteq W$ is called *subanalytic* if X is locally of the form $\pi(Y)$ for some relatively compact semianalytic subset $Y \subseteq \mathbb{R}^n$. Given subanalytic subsets $X \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^n$, a continuous map $f: X \rightarrow Y$ is called *subanalytic* if the graph of f is subanalytic in $\mathbb{R}^m \times \mathbb{R}^n$.

We will also make significant use of Whitney stratifications, which we recall here. Let \mathcal{I} be a partially ordered set with order relation $<$, and X a closed subset of a real smooth manifold M . A *Whitney stratification* of X is a locally finite collection of disjoint locally closed smooth submanifolds $S_i \subseteq X$ called *strata* that satisfy:

- (i) $X = \bigcup_{i \in \mathcal{I}} S_i$.
- (ii) If $i < j$, then $S_i \cap \overline{S_j} \neq \emptyset$ if and only if $S_i \subseteq \overline{S_j}$, where $\overline{S_j}$ denotes the Euclidean closure of S_j .
- (iii) Let $S_\alpha \subseteq \overline{S_\beta}$. Let $\{x_i\}_{i=1}^\infty \subseteq S_\beta$ and $\{y_i\}_{i=1}^\infty \subseteq S_\alpha$ be two sequences converging to the same $y \in S_\alpha$. If the secant lines $\text{span}\{x_i, y_i\}$ converge to some limiting line ℓ and the tangent spaces $T_{x_i}(S_\beta)$ converge to some limiting space τ , then $T_y(S_\alpha) \subseteq \tau$ and $\ell \subseteq \tau$.

For a proper subanalytic map $f: X \rightarrow Y$, there are Whitney stratifications of X and Y into analytic manifolds such that f is a *stratified map*, i.e., for each stratum $B \subseteq Y$, $f^{-1}(B)$ is a union of connected components of strata of X . We refer the reader to [1] for further details.

The following result is well-known in the semialgebraic setting [15, Theorem 3.3] but we need it in a subanalytic setting; fortunately a similar proof yields the required analogous result.

Lemma 2.1. *Let U_1 and U_2 be real vector spaces. Let $X \subseteq U_1$ be a k -dimensional subanalytic subset and $f: X \rightarrow U_2$ be a subanalytic map. Then the set of points of X where f is not differentiable is contained in a subanalytic subset of dimension strictly smaller than k .*

Proof. Let $\Gamma \subseteq U_1 \times U_2$ be the graph of f , and $\pi_1: X \times U_2 \rightarrow X$ be the projection. Given a Whitney stratification $\Gamma = \bigcup_{i \in \mathcal{I}} \Gamma_i$ of Γ , since each $\pi_1(\Gamma_i)$ is subanalytic, there is a Whitney stratification $X = \bigcup_{j \in \mathcal{J}} X_j$ such that $\pi_1(\Gamma_i)$ is a union of strata, namely $f|_{X_j}: X_j \rightarrow U_2$ is a subanalytic map whose graph is an analytic submanifold of $U_1 \times U_2$. For simplicity, we denote $f|_{X_j}$ by f_j and the graph of f_j by Δ_j . Then the set of points of X_j where f_j is not differentiable is contained in the critical values of $\pi_1: \Delta_j \rightarrow X_j$. Hence the set of nondifferentiable points of f in X is contained in

$$\{x \in X_j \mid \dim X_j < k\} \cup \{x \in X_l \mid \dim X_l = k \text{ and } x \text{ is a critical value of } \pi_1: \Delta_l \rightarrow X_l\}$$

which is a subanalytic subset whose dimension is less than k . \square

In this article we prove most of our results for *complex analytic varieties*, i.e., defined locally by the common zero loci of finitely many holomorphic functions; although the examples in Section 5 are all *complex algebraic varieties*, i.e., those holomorphic functions are polynomials.¹ Since a subset of \mathbb{C}^n may be regarded as a subset of \mathbb{R}^{2n} , we have the following relations between (semi)algebraic and (semi)analytic sets:

$$\begin{aligned} \text{complex algebraic variety} &\subseteq \text{real semialgebraic set} \subseteq \text{real semianalytic set}, \\ \text{complex algebraic variety} &\subseteq \text{complex analytic variety} \subseteq \text{real semianalytic set}. \end{aligned}$$

So for example any property of real semianalytic sets will automatically be satisfied by a complex analytic variety. In particular, the following important theorem [20] also holds true in subanalytic, semialgebraic, or complex algebraic contexts; but for our purpose, it will be stated for complex analytic sets.

Theorem 2.2. *Let M be a complex manifold and $Z \subseteq M$ be a closed complex analytic subset. Then there is a Whitney stratification $Z = \bigcup_{i \in \mathcal{I}} S_i$ of Z such that*

- (a) *each S_i is a complex submanifold of M ;*
- (b) *if $S_i \subseteq \overline{S_j}$, there is a vector bundle $E_i \rightarrow S_i$, a neighborhood $U_i \subseteq E_i$ of the zero section S_i , and a homeomorphism $h: U_i \rightarrow \overline{S_j}$ of U_i to an open subset $h(U_i)$ in $\overline{S_j}$.*

Notations and terminologies. In the next sections, we will frame our discussion over an abstract real vector space U or an abstract complex vector space V . By a semialgebraic or subanalytic subset $X \subseteq U$ we mean that X can be identified with some semialgebraic or subanalytic subset in \mathbb{R}^n when we identify $U \cong \mathbb{R}^n$ by fixing a basis of U . Likewise, by an algebraic or analytic variety $X \subseteq V$ we mean that X can be identified with some algebraic or analytic variety in \mathbb{C}^n when we identify $V \cong \mathbb{C}^n$ by fixing a basis of V .

The main reason we prefer such coordinate-free descriptions instead of assuming at the outset that $U = \mathbb{R}^n$ or $V = \mathbb{C}^n$ is that we will ultimately be applying the results to cases where U and V have additional structures, e.g., $\mathbb{K}^{n_1} \otimes \cdots \otimes \mathbb{K}^{n_d}$, $\mathcal{S}^d(\mathbb{K}^n)$, $\Lambda^k(\mathbb{K}^n)$, $\mathcal{S}^d(\Lambda^k(\mathbb{K}^n))$, $\mathcal{S}^{d_1}(\mathbb{K}^n) \otimes \cdots \otimes \mathcal{S}^{d_n}(\mathbb{K}^n)$, $\Lambda^{k_1}(\mathbb{K}^n) \otimes \cdots \otimes \Lambda^{k_n}(\mathbb{K}^n)$, etc.

The results in this article will be proved for *general points*. A point in a semianalytic set X is said to be *general* with respect to a property if the subset of points that do not have that property is contained in a semianalytic subset whose dimension is strictly smaller than $\dim X$. In particular, a point in a real vector space is said to be *general* with respect to a property if the set of points that do not have that property is contained in a real analytic hypersurface. The same notion applies to a point in a complex vector space, regarded as a real vector space of real dimension twice its complex dimension.

Establishing that a result holds true for all general points is far stronger than the common practice in applied and computational mathematics of establishing it ‘with high probability’ or ‘almost surely’ or ‘almost everywhere’. If a result is valid for all general points, then not only do

¹While every complex *projective* analytic variety is algebraic by Chow’s theorem, there exist complex analytic varieties that are not algebraic.

we know that it is valid almost surely/everywhere but also that the invalid points are all limited to a subset of strictly smaller dimension. In fact, when the result is about points in a vector space, which is the case in our article, this subset of invalid points can be described by a single equation that may in principle be determined, so that we know where the invalid points lie.

3. BEST APPROXIMATION BY POINTS IN A CLOSED SUBANALYTIC SET

Let U be an n -dimensional real vector space with an inner product $\langle \cdot, \cdot \rangle$ and corresponding ℓ^2 -norm $\|\cdot\|$. Let $X \subseteq U$ be a closed subanalytic set. We define the squared distance function d by

$$d: U \rightarrow \mathbb{R}, \quad p \mapsto \min_{q \in X} \|p - q\|^2.$$

For $p \notin X$, a *best X -approximation* of p is a minimizer $q \in X$ that attains $\min_{q \in X} \|p - q\|^2$, and it is customary to write $\operatorname{argmin}_{q \in X} \|p - q\|^2$ for the set of all such minimizers. Let

$$(5) \quad \mathcal{C}(X) := \{p \in U \setminus X \mid p \text{ does not have a unique best approximation in } X\}.$$

We have the following subanalytic analogue of [17, Theorem 3.7].

Lemma 3.1. *Let $m, n \in \mathbb{N}$. For any closed subanalytic set $X \subseteq U$, the set $\mathcal{C}(X) \cap \overline{B}(0, m)$ is subanalytic, and $\mathcal{C}(X)$ has dimension less than n . Here $\overline{B}(0, m) := \{x \in U \mid \|x\| \leq m\}$.*

Proof. We start by defining the two maps:

$$\begin{aligned} f: U \times \mathbb{R} \times X &\rightarrow \mathbb{R}, & (p, t, q) &\mapsto \|p - q\|^2 - t, \\ g: U \times \mathbb{R} \times X \times (0, \infty) &\rightarrow \mathbb{R}, & (p, t, q, \varepsilon) &\mapsto t + \varepsilon - \|p - q\|^2, \end{aligned}$$

and the two projections:

$$\begin{aligned} \pi_1: U \times \mathbb{R} \times X &\rightarrow U \times \mathbb{R}, \\ \pi_2: U \times \mathbb{R} \times X \times (0, \infty) &\rightarrow U \times \mathbb{R} \times X. \end{aligned}$$

Let Γ be the graph of d . Then

$$\Gamma = [U \times [0, \infty)] \cap [U \times \mathbb{R} \setminus \pi_1(f^{-1}(-\infty, 0))] \cap [\pi_1(U \times \mathbb{R} \times X \setminus \pi_2(g^{-1}(-\infty, 0)))]$$

because

$$(p, t) \in U \times \mathbb{R} \setminus \pi_1(f^{-1}(-\infty, 0)) \iff \|p - q\|^2 \geq t \text{ for all } q \in X,$$

and

$$\begin{aligned} (p, t) \in \pi_1(U \times \mathbb{R} \times X \setminus \pi_2(g^{-1}(-\infty, 0))) \\ \iff \text{for any } \varepsilon > 0, \text{ there is a } q \in X \text{ with } \|p - q\|^2 < t + \varepsilon. \end{aligned}$$

Therefore $d|_{\overline{B}(0, m)}$, the squared distance function restricted to $\overline{B}(0, m)$, is a subanalytic function for each $m \in \mathbb{N}$. By [17, Theorem 2.1], the set $\mathcal{C}(X)$ in (5) comprises precisely the nonsmooth points:

$$\mathcal{C}(X) = \{p \in U \setminus X \mid d \text{ is not differentiable at } p\}.$$

Hence, by Lemma 2.1, $\mathcal{C}(X) \cap \overline{B}(0, m)$ is a subanalytic subset with dimension less than n . Since

$$\mathcal{C}(X) = \bigcup_{m=1}^{\infty} (\mathcal{C}(X) \cap \overline{B}(0, m)),$$

$\mathcal{C}(X)$ must also have dimension less than n . □

4. BEST APPROXIMATION BY POINTS IN A CLOSED COMPLEX VARIETY

We will now switch our discussion from \mathbb{R} to \mathbb{C} . Let V be an n -dimensional complex vector space with a Hermitian inner product $\langle \cdot, \cdot \rangle$ and corresponding ℓ^2 -norm $\|\cdot\|$. For any closed irreducible complex analytic variety $X \subseteq V$, again we let $\mathcal{C}(X)$ denote the set of points which do not have a unique best approximation in X , i.e., with V in place of U in (5). Since V may be regarded as a real vector space of real dimension $2n$, X is naturally a real analytic variety. In particular, best X -approximations are as defined in Section 3.

By Lemma 3.1, $\mathcal{C}(X)$ has real dimension strictly less than $2n$ and we easily deduce the following.

Corollary 4.1. *Let X be a closed complex analytic variety in a complex vector space V . Then a general $p \in V$ has a unique best X -approximation.*

For any $p \in V \setminus (\mathcal{C}(X) \cup X)$, p has a unique best X -approximation, which we will denote by $\pi_X(p)$. Thus this gives us a map $\pi_X: V \setminus (\mathcal{C}(X) \cup X) \rightarrow X$ that sends p to its unique best X -approximation $\pi_X(p)$, i.e.,

$$\|p - \pi_X(p)\|^2 = \min_{q \in X} \|p - q\|^2.$$

The map π_X is also clearly subanalytic when restricted to $\overline{B}(0, m)$. With this observation, we will state and prove our main result.

Theorem 4.2. *Let X be a closed irreducible complex analytic variety in V . Let $Z \subseteq X$ be any complex analytic subvariety with $\dim Z < \dim X$. For a general $p \in V$, its unique best X -approximation $\pi_X(p)$ does not lie in Z .*

Proof. We proceed by contradiction. Suppose there is a nonempty open subset $\mathcal{O} \subseteq V \setminus (\mathcal{C}(X) \cup X)$ such that for any $p \in \mathcal{O}$, the best X -approximation $\pi_X(p)$ lies in Z . Without loss of generality, we may assume that Z is a minimum subvariety of X containing $\pi_X(\mathcal{O})$ in the sense that any subvariety containing $\pi_X(\mathcal{O})$ has dimension at least $\dim_{\mathbb{C}} Z$. For simplicity, we choose a basis of V so that we may write $u = (u_1, \dots, u_n)$, $v = (v_1, \dots, v_n) \in V$, and the inner product is given by $\langle u, v \rangle = u_1 \bar{v}_1 + \dots + u_n \bar{v}_n$ under this basis. With this assumption, the norm is $\|v\| = (|v_1|^2 + \dots + |v_n|^2)^{1/2}$.

First we assume $\dim Z > 0$. By taking a Whitney stratification, there is a nonsingular point $z \in Z$ and an open ball $B(z, \rho) \subseteq V$ around z , such that $B(z, \rho) \cap Z$ is a connected complex submanifold in V , and $\pi_X^{-1}(B(z, \rho) \cap Z)$ contains a nonempty open subset $\mathcal{O}' \subseteq \mathcal{O}$. For notational convenience, we will still denote $B(z, \rho) \cap Z$ by Z and \mathcal{O}' by \mathcal{O} .

Given $p \in \mathcal{O}$, let $B(p, \varepsilon) \subseteq \mathcal{O}$ be a small open ball of p , and let $q = \pi_X(p)$. Then

$$\left. \frac{d}{dt} \langle p - q(t), p - q(t) \rangle \right|_{t=0} = 0$$

for any real analytic curve $q(t) \subseteq Z$ with $q(0) = q$, which implies that

$$\operatorname{Re} \langle p - q, v \rangle = 0 \quad \text{for any } v \in T_q(Z),$$

where $T_q(Z)$ is the tangent space of Z at q . Without loss of generality, let $q = 0 \in V$, and let

$$W := \{x \in V \mid \operatorname{Re} \langle x - q, v \rangle = 0 \text{ for any } v \in T_q(Z)\}.$$

Since $T_q(Z) = T_0(Z)$ is a complex vector space, we in fact have

$$W = \{x \in V \mid \langle x, v \rangle = 0 \text{ for any } v \in T_0(Z)\},$$

i.e., W is a complex vector subspace of V whose complex codimension is $\dim_{\mathbb{C}} Z$. Thus

$$W \cap Z \cap B(0, \delta) = \{0\}$$

for some small open ball $B(0, \delta) \subseteq V$. Since X is irreducible and the set of nonsingular points of X is dense in X , $(X \setminus Z) \cap B(0, \delta) \neq \emptyset$. By Theorem 2.2, Z has a small tubular neighborhood in X . Thus

$W \cap X \cap B(0, \delta)$ is of positive dimension. Hence it suffices to show that $\pi_{W \cap X \cap B(0, \delta)}(W \cap B(p, \varepsilon))$ cannot be just $\{0\}$.

We have reduced the problem to showing that in the complex vector space W , given a positive-dimensional complex analytic variety $W \cap X \cap B(0, \delta)$ and a point $0 \in W \cap X \cap B(0, \delta)$,

$$\pi_{W \cap X \cap B(0, \delta)}(W \cap B(p, \varepsilon)) \neq \{0\}$$

for a nonempty open ball $W \cap B(p, \varepsilon)$ in W . Again for notational simplicity, we will still use X to denote $W \cap X \cap B(0, \delta)$, and $B(p, \varepsilon)$ to denote $W \cap B(p, \varepsilon)$. One should note that this is in fact the case $\dim Z = 0$.

Pick a complex analytic curve $C \subseteq X$ passing through 0. We will show by contradiction that $\pi_C(B(p, \varepsilon)) \neq \{0\}$, which implies that $\pi_X(B(p, \varepsilon)) \neq \{0\}$. Suppose $\pi_C(B(p, \varepsilon)) = \{0\}$, i.e.,

$$|x_1|^2 + \cdots + |x_n|^2 < |x_1 - y_1|^2 + \cdots + |x_n - y_n|^2$$

for all $x = (x_1, \dots, x_n) \in B(p, \varepsilon)$ and all $q = (y_1, \dots, y_n) \in C \cap B^\times(0, \eta)$, where $B^\times(0, \eta) := B(0, \eta) \setminus \{0\}$ and $\eta > 0$ is some small number. By applying elimination theory, we may let $(y_1(t), \dots, y_n(t))$ be a local parametrization of C such that $y_1(0) = \cdots = y_n(0) = 0$, and each $y_j(t)$ is holomorphic in t around 0, i.e.,

$$y_j(t) = \sum_{k=m_j}^{\infty} a_{j,k} t^k, \quad \text{where } m_j \geq 1 \text{ and } a_{j,m_j} \neq 0, \quad j = 1, \dots, n.$$

Let

$$f(t) := \sum_{j=1}^n (|x_j|^2 - |x_j - y_j(t)|^2)$$

and let $m := \min\{m_1, \dots, m_n\}$. Then

$$f(t) = 2 \sum_{j=1}^n (\operatorname{Re}(a_{j,m_j} t^{m_j} \bar{x}_j) + O(t^{m_j+1})) = \operatorname{Re}(t^m g(x)) + O(t^{m+1}),$$

where

$$g(x) := \sum_{m_j=m} 2a_{j,m_j} \bar{x}_j \quad \text{and} \quad O(t^{m+1}) = \text{terms of degrees } \geq m+1.$$

Note that $g(x)$ is a sum over terms with $m_j = m$. So $g(x) \neq 0$ for a general x . Let $t^m = \lambda \overline{g(x)}$ for some small $\lambda > 0$. Then $f(t) > 0$, which contradicts our assumption that $f(t) < 0$ for all $x = (x_1, \dots, x_n) \in B(p, \varepsilon)$ and all $q = (y_1, \dots, y_n) \in C \cap B^\times(0, \eta)$. So $\pi_C(B(p, \varepsilon))$ cannot be just $\{0\}$, which allows us to conclude that $\pi_X(\mathcal{O}) \not\subseteq Z$. \square

Since the singular locus of a complex analytic variety is a complex analytic subvariety of smaller dimension, we obtain the following corollary of Theorem 4.2.

Corollary 4.3. *Let X be a closed irreducible complex analytic variety in V . For a general $p \in V$, its best X -approximation $\pi_X(p)$ is a nonsingular point of X .*

5. APPLICATIONS TO FUNCTION, MATRIX, AND TENSOR APPROXIMATIONS

We will now apply the results in Section 4 back to the functional approximation problems described in Section 1. Since our results require that we work over \mathbb{C} (in fact, they are false over \mathbb{R}), we will write $L^2(\Omega) = L^2_{\mathbb{C}}(\Omega)$ in the following. A point to recall is that

$$L^2(\Omega_1 \times \cdots \times \Omega_d) = L^2(\Omega_1) \otimes \cdots \otimes L^2(\Omega_d)$$

and so multivariate functions are in fact tensors. In many, if not most applications, the multivariate target function (i.e., the function to be approximated) possesses various forms of symmetries, the most common being *full* invariance or skew-invariance under all permutations of arguments:

$$f(x_{\tau(1)}, \dots, x_{\tau(d)}) = f(x_1, \dots, x_d) \quad \text{or} \quad g(x_{\tau(1)}, \dots, x_{\tau(d)}) = \text{sgn}(\tau)g(x_1, \dots, x_d),$$

for all $\tau \in \mathfrak{S}_d$. In other words, f is a symmetric tensor and g an alternating tensor:

$$f \in \mathcal{S}^d(L^2(\Omega)) \quad \text{or} \quad g \in \Lambda^d(L^2(\Omega)).$$

Note that in these cases, we need $\Omega_1 = \dots = \Omega_d = \Omega$ in order to permute arguments. There are many types of *partial* symmetries as well. For example, we may have functions that satisfy

$$\begin{aligned} f(x_{\rho(1)}, \dots, x_{\rho(d)}, y_{\tau(1)}, \dots, y_{\tau(k)}) &= f(x_1, \dots, x_d, y_1, \dots, y_k), \\ g(x_{\rho(1)}, \dots, x_{\rho(d)}, y_{\tau(1)}, \dots, y_{\tau(k)}) &= \text{sgn}(\rho) \text{sgn}(\tau)g(x_1, \dots, x_d, y_1, \dots, y_k) \end{aligned}$$

for all $\rho \in \mathfrak{S}_d$ and $\tau \in \mathfrak{S}_k$. These correspond respectively to functions

$$f \in \mathcal{S}^d(L^2(\Omega_1)) \otimes \mathcal{S}^k(L^2(\Omega_2)) \quad \text{and} \quad g \in \Lambda^d(L^2(\Omega_1)) \otimes \Lambda^k(L^2(\Omega_2)).$$

There are also more intricate symmetries. For example, a complex-valued function $f(X)$ of a matrix variable $X = (x_{ij})_{i,j=1}^{k,d}$ that satisfies

$$\begin{aligned} f(x_{\rho_1(1),\tau(1)}, \dots, x_{\rho_1(k),\tau(1)}, \dots, x_{\rho_d(1),\tau(d)}, \dots, x_{\rho_d(k),\tau(d)}) \\ = \text{sgn}(\rho_1) \dots \text{sgn}(\rho_d) f(x_{1,1}, \dots, x_{k,1}, \dots, x_{1,d}, \dots, x_{k,d}) \end{aligned}$$

for all $\rho_1, \dots, \rho_d \in \mathfrak{S}_k$ and $\tau \in \mathfrak{S}_d$. This corresponds to $f \in \mathcal{S}^d(\Lambda^k(L^2(\Omega)))$. We will see more examples later.

Each of these classes of functions comes with its own nonlinear r -term approximation problem that preserves the respective symmetries or skew-symmetries, full or partial. For example, for $f \in L^2(\Omega \times \dots \times \Omega)$, we may just want

$$(6) \quad \inf \left\| f - \sum_{i=1}^r \varphi_{1,i} \otimes \dots \otimes \varphi_{d,i} \right\|,$$

as discussed in Section 1, but for a symmetric $f \in \mathcal{S}^d(L^2(\Omega))$ or an alternating $g \in \Lambda^d(L^2(\Omega))$, the more natural approximation problems are respectively

$$(7) \quad \inf \left\| f - \sum_{i=1}^r \varphi_{1,i} \circ \dots \circ \varphi_{d,i} \right\| \quad \text{or} \quad \inf \left\| g - \sum_{i=1}^r \varphi_{1,i} \wedge \dots \wedge \varphi_{d,i} \right\|,$$

with $\varphi_{j,i} \in L^2(\Omega)$, $i = 1, \dots, r$, $j = 1, \dots, d$. In fact, there are other natural options for a symmetric f , where we may instead seek an approximation of the form

$$(8) \quad \inf \left\| f - \sum_{i=1}^r \varphi_i^{\otimes d} \right\|,$$

with $\varphi_i \in L^2(\Omega)$, $i = 1, \dots, r$. The symbols \circ and \wedge denote symmetric product and exterior product respectively; for those who have not encountered them, these and other notations will be defined in Section 5.1. Then in Section 5.2, we will describe other classes of r -term approximation problems for functions possessing more complicated symmetries than the ones in (6), (7), (8). All these approximation problems have two features in common:

- (i) Each of them is associated with a complex algebraic variety; e.g., (6) with the Segre variety, (7) with the Chow variety and Grassmann variety respectively, (8) with the Veronese variety.
- (ii) All of them may fail to have a solution; e.g., there exist target functions where the infima in (6), (7), (8) cannot be attained.

We will later see how our results in Section 4 combined with (i) allow us to rectify (ii).

We would like to mention a slightly different, perhaps more common, alternative for framing the function approximation problems above. As we described in Section 1, for computational purposes, $L^2(\Omega_i)$'s are all finite-dimensional (e.g., when Ω_i 's are all finite sets) and in this case, $L^2(\Omega_i) \cong \mathbb{C}^{n_i}$ with a choice of basis, and thus

$$\begin{aligned} L^2(\Omega_1 \times \Omega_2) &= L^2(\Omega_1) \otimes L^2(\Omega_2) \cong \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \cong \mathbb{C}^{n_1 \times n_2}, \\ L^2(\Omega_1 \times \Omega_2 \times \Omega_3) &= L^2(\Omega_1) \otimes L^2(\Omega_2) \otimes L^2(\Omega_3) \cong \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} \otimes \mathbb{C}^{n_3} \cong \mathbb{C}^{n_1 \times n_2 \times n_3}, \\ &\vdots \\ L^2(\Omega_1 \times \cdots \times \Omega_d) &= L^2(\Omega_1) \otimes \cdots \otimes L^2(\Omega_d) \cong \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_d} \cong \mathbb{C}^{n_1 \times \cdots \times n_d}. \end{aligned}$$

The set $\mathbb{C}^{n_1 \times \cdots \times n_d}$ denotes the vector space of d -dimensional hypermatrices

$$(a_{i_1, \dots, i_d})_{i_1, \dots, i_d=1}^{n_1, \dots, n_d}$$

(when $d = 2$, this reduces to a usual matrix in linear algebra), which of course is nothing more than a convenient way to represent a function $f : \Omega_1 \times \cdots \times \Omega_d \rightarrow \mathbb{C}$ by storing its value

$$f(i_1, \dots, i_d) = a_{i_1, \dots, i_d}$$

at $(i_1, \dots, i_d) \in \Omega_1 \times \cdots \times \Omega_d$, assuming that Ω_i 's are all finite sets. In fact, practitioners in computational mathematics overwhelmingly regard a tensor as such a hypermatrix. The symmetries described earlier carry verbatim to hypermatrices with the permutations acting on the indices.

We may either frame our results in this section in the form of function approximations or in the form of tensor approximations (or more accurately, matrix/hypermatrix approximations), but instead of favoring one over the other, we would simply resort to stating them for abstract vector spaces. So implicitly, $V = L^2(\Omega_1 \times \cdots \times \Omega_d)$ for function approximations and $V = \mathbb{C}^{n_1 \times \cdots \times n_d}$ for tensor approximations.

5.1. Segre/Veronese/Grassmann varieties and friends. We begin by reviewing some basic tensor constructions. The varieties that we define will be subsets of tensor spaces constructed via one or more of the following ways. We write $V_1 \otimes \cdots \otimes V_d$ for the tensor product of vector spaces V_1, \dots, V_d . A rank-one tensor is a nonzero decomposable tensor $v_1 \otimes \cdots \otimes v_d$, where $v_i \in V_i$. When $V_1 = \cdots = V_d = V$, we use the abbreviations $V^{\otimes d} = V \otimes \cdots \otimes V$ and $v^{\otimes d} = v \otimes \cdots \otimes v$. In this case, the symmetric group \mathfrak{S}_d acts on rank-one tensors by

$$\tau(v_1 \otimes \cdots \otimes v_d) = v_{\tau(1)} \otimes \cdots \otimes v_{\tau(d)},$$

and this action extends linearly to an action on $V^{\otimes d}$. The subspaces

$$\begin{aligned} \mathcal{S}^d(V) &= \{T \in V^{\otimes d} \mid \tau(T) = T \text{ for all } \tau \in \mathfrak{S}_d\}, \\ \mathcal{A}^d(V) &= \{T \in V^{\otimes d} \mid \tau(T) = \text{sgn}(\tau)T \text{ for all } \tau \in \mathfrak{S}_d\}, \end{aligned}$$

are called the spaces of *symmetric d -tensors* and *alternating d -tensors* respectively. We may construct tensor spaces with more intricate symmetries and skew-symmetries by combining these, e.g., $\mathcal{S}^d(V) \otimes \mathcal{S}^k(V)$, $\mathcal{A}^d(V) \otimes \mathcal{A}^k(V)$, $\mathcal{S}^d(\mathcal{A}^k(V))$, $\mathcal{S}^d(\mathcal{S}^k(V))$, etc.

The *symmetric product* and *alternating product* of $v_1, \dots, v_d \in V$ are respectively defined by

$$\begin{aligned} v_1 \circ \cdots \circ v_d &:= \frac{1}{d!} \sum_{\tau \in \mathfrak{S}_d} v_{\tau(1)} \otimes \cdots \otimes v_{\tau(d)} \in \mathcal{S}^d(V), \\ v_1 \wedge \cdots \wedge v_d &:= \frac{1}{d!} \sum_{\tau \in \mathfrak{S}_d} \text{sgn}(\tau) v_{\tau(1)} \otimes \cdots \otimes v_{\tau(d)} \in \mathcal{A}^d(V). \end{aligned}$$

In finite dimensions, each of the approximation problems we have encountered earlier and yet others that we will see in Section 5.2 comes with an associated complex algebraic variety. These varieties are usually defined as complex projective varieties, i.e., subsets of projective spaces defined

by the common zero loci of a finite collection of homogeneous polynomials, and we will not deviate from this standard practice in our definitions below:

Segre variety. This is the image $\sigma(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d)$ of the Segre embedding:

$$\sigma: \mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d \rightarrow \mathbb{P}(V_1 \otimes \cdots \otimes V_d), \quad ([v_1], \dots, [v_d]) \mapsto [v_1 \otimes \cdots \otimes v_d].$$

Veronese variety. This is the image $\nu(\mathbb{P}V)$ of the Veronese embedding:

$$\nu: \mathbb{P}V \rightarrow \mathbb{P}\mathcal{S}^d(V), \quad [v] \mapsto [v^{\otimes d}].$$

Chow variety. This is the image² $\text{Ch}_d(V) := \kappa((\mathbb{P}V)^d)$ of the Chow map:

$$\kappa: \mathbb{P}V \times \cdots \times \mathbb{P}V \rightarrow \mathbb{P}\mathcal{S}^d(V), \quad ([v_1], \dots, [v_d]) \mapsto [v_1 \circ \cdots \circ v_d].$$

Grassmann variety. This is the image $\psi(\text{Gr}_k(V))$ of the *Grassmannian* $\text{Gr}_k(V)$, the set of k -dimensional linear subspaces of V , under the *Plücker embedding*:

$$\psi: \text{Gr}_k(V) \rightarrow \mathbb{P}\Lambda^k(V), \quad \text{span}\{v_1, \dots, v_k\} \mapsto [v_1 \wedge \cdots \wedge v_k].$$

Segre–Veronese variety. This is the image $\sigma\nu(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_m)$ of the Segre–Veronese embedding:

$$\begin{aligned} \sigma\nu: \mathbb{P}V_1 \times \cdots \times \mathbb{P}V_m &\rightarrow \mathbb{P}(\mathcal{S}^{d_1}(V) \otimes \cdots \otimes \mathcal{S}^{d_m}(V)), \\ ([v_1], \dots, [v_m]) &\mapsto [v_1^{\otimes d_1} \otimes \cdots \otimes v_m^{\otimes d_m}]. \end{aligned}$$

Segre–Chow variety. This is the image $\sigma\kappa(\text{Ch}_{d_1}(V) \times \cdots \times \text{Ch}_{d_m}(V))$ of the Segre–Chow map:

$$\begin{aligned} \sigma\kappa: \text{Ch}_{d_1}(V) \times \cdots \times \text{Ch}_{d_m}(V) &\rightarrow \mathbb{P}(\mathcal{S}^{d_1}(V) \otimes \cdots \otimes \mathcal{S}^{d_m}(V)), \\ ([v_{1,1} \circ \cdots \circ v_{d_1,1}], \dots, [v_{1,m} \circ \cdots \circ v_{d_m,m}]) &\mapsto [(v_{1,1} \circ \cdots \circ v_{d_1,1}) \otimes \cdots \otimes (v_{1,m} \circ \cdots \circ v_{d_m,m})]. \end{aligned}$$

Segre–Grassmann variety. This is the image $\sigma\psi(\text{Gr}_{k_1}(V) \times \cdots \times \text{Gr}_{k_m}(V))$ of the Segre–Grassmann map:

$$\begin{aligned} \sigma\psi: \text{Gr}_{k_1}(V) \times \cdots \times \text{Gr}_{k_m}(V) &\rightarrow \mathbb{P}(\Lambda^{k_1}(V) \otimes \cdots \otimes \Lambda^{k_m}(V)), \\ ([v_{1,1} \wedge \cdots \wedge v_{k_1,1}], \dots, [v_{1,m} \wedge \cdots \wedge v_{k_m,m}]) &\mapsto [(v_{1,1} \wedge \cdots \wedge v_{k_1,1}) \otimes \cdots \otimes (v_{1,m} \wedge \cdots \wedge v_{k_m,m})]. \end{aligned}$$

Veronese–Chow variety. This is the image $\nu\kappa(\text{Ch}_d(V))$ of the Veronese–Chow map:

$$\nu\kappa: \text{Ch}_d(V) \rightarrow \mathbb{P}(\mathcal{S}^k(\mathcal{S}^d(V))), \quad [v_1 \circ \cdots \circ v_d] \mapsto [(v_1 \circ \cdots \circ v_d)^{\otimes k}].$$

Veronese–Grassmann variety. This is the image $\nu\psi(\text{Gr}_k(V))$ of the Veronese–Grassmann map:

$$\nu\psi: \text{Gr}_k(V) \rightarrow \mathbb{P}(\mathcal{S}^d(\Lambda^k(V))), \quad [v_1 \wedge \cdots \wedge v_k] \mapsto [(v_1 \wedge \cdots \wedge v_k)^{\otimes d}].$$

Segre–Veronese–Chow variety. This is the image $\sigma\nu\kappa(\text{Ch}_{d_1}(V) \times \cdots \times \text{Ch}_{d_m}(V))$ of the Segre–Veronese–Chow map:

$$\begin{aligned} \sigma\nu\kappa: \text{Ch}_{d_1}(V) \times \cdots \times \text{Ch}_{d_m}(V) &\rightarrow \mathbb{P}(\mathcal{S}^{k_1}(\mathcal{S}^{d_1}(V)) \otimes \cdots \otimes \mathcal{S}^{k_m}(\mathcal{S}^{d_m}(V))), \\ ([v_{1,1} \circ \cdots \circ v_{d_1,1}], \dots, [v_{1,m} \circ \cdots \circ v_{d_m,m}]) &\mapsto [(v_{1,1} \circ \cdots \circ v_{d_1,1})^{\otimes k_1} \otimes \cdots \otimes (v_{1,m} \circ \cdots \circ v_{d_m,m})^{\otimes k_m}]. \end{aligned}$$

²This is usually called the *Chow variety of zero cycles* in $\mathbb{P}V^*$. It is a subvariety of $\mathbb{P}\mathcal{S}^d(V)$ whose affine cone is the set of degree- d homogeneous polynomials on the dual space V^* that can be decomposed into a product of linear forms [18, Chapter 4, Proposition 2.1].

Segre–Veronese–Grassmann variety. This is the image $\sigma\nu\psi(\mathrm{Gr}_{k_1}(V) \times \cdots \times \mathrm{Gr}_{k_m}(V))$ of the Segre–Veronese–Grassmann map:

$$\begin{aligned} \sigma\nu\psi: \mathrm{Gr}_{k_1}(V) \times \cdots \times \mathrm{Gr}_{k_m}(V) &\rightarrow \mathbb{P}(\mathbb{S}^{d_1}(\Lambda^{k_1}(V)) \otimes \cdots \otimes \mathbb{S}^{d_m}(\Lambda^{k_m}(V))), \\ ([v_{1,1} \wedge \cdots \wedge v_{k_1,1}], \dots, [v_{1,m} \wedge \cdots \wedge v_{k_m,m}]) &\mapsto [(v_{1,1} \wedge \cdots \wedge v_{k_1,1})^{\otimes d_1} \otimes \cdots \otimes (v_{1,m} \wedge \cdots \wedge v_{k_m,m})^{\otimes d_m}]. \end{aligned}$$

Note that $\sigma\nu, \sigma\kappa, \sigma\psi, \nu\kappa, \nu\psi, \sigma\nu\kappa, \sigma\nu\psi$ are all compositions of their respective constituent maps. In fact we may use successive compositions of the embeddings σ and ν to define more complicated algebraic varieties of the same nature; it is not possible to exhaust all such constructions. We refer the readers to [18, 23, 25] for basic properties of the Segre, Veronese, Grassmann, and Chow varieties from an algebraic geometric perspective, although none of which would be required for our subsequent discussions.

While we have defined all these varieties as projective varieties, it is important to note that for approximation problems, we will need to work in vector spaces rather than projective spaces. As such, when we discuss these varieties in the context of approximations, we would in fact be referring to their *affine cones*: For any $X \subseteq \mathbb{P}V$, its affine cone is $\widehat{X} := \pi^{-1}(X) \cup \{0\}$, where $\pi: V \setminus \{0\} \rightarrow \mathbb{P}V$ is the quotient map taking a vector space V onto its projective space $\mathbb{P}V$. We will write $[v] := \pi(v)$ for the projective equivalence class of $v \in V \setminus \{0\}$.

5.2. Best low-rank approximations of tensors. A variety $X \subseteq \mathbb{P}V$ is said to be *nondegenerate* if X is not contained in any hyperplane. An implication is that its affine cone \widehat{X} would span V , i.e., $\mathrm{span}(\widehat{X}) = V$, and so any $p \in V$ can be expressed as a linear combination $p = \alpha_1 x_1 + \cdots + \alpha_r x_r$ with $x_1, \dots, x_r \in \widehat{X}$. Since, by the definition of affine cone, $x \in \widehat{X}$ iff $\alpha x \in \widehat{X}$ for any $\alpha \neq 0$, we may in fact replace the linear combination by a sum, i.e., every $p \in V$ may be expressed as $p = x_1 + \cdots + x_r$ for some $x_1, \dots, x_r \in \widehat{X}$.

Given an irreducible nondegenerate complex projective variety $X \subseteq \mathbb{P}V$, the X -rank of a nonzero $p \in V$ is defined as

$$\mathrm{rank}_X(p) := \min\{r \in \mathbb{N} : p = x_1 + \cdots + x_r, x_i \in \widehat{X}\},$$

and $\mathrm{rank}_X(0) := 0$. The X -border rank of p , denoted by $\overline{\mathrm{rank}}_X(p)$, is the minimum integer r such that p is a limit of a sequence of X -rank- r points. For irreducible nondegenerate complex projective varieties $X_1, \dots, X_r \subseteq \mathbb{P}V$, the *join map* is defined by

$$J: \widehat{X}_1 \times \cdots \times \widehat{X}_r \rightarrow V, \quad (x_1, \dots, x_r) \mapsto x_1 + \cdots + x_r.$$

The Euclidean closure of the image $J(\widehat{X}_1 \times \cdots \times \widehat{X}_r)$ in V is called the *join variety* of X_1, \dots, X_r . In particular, when $X_1 = \cdots = X_r = X$, we denote the image of J by $\Sigma_r^\circ(X)$, and its Euclidean closure by $\Sigma_r(X)$, often called the *r th secant variety* of X . Note that

$$\Sigma_r^\circ(X) = \{p \in V : \mathrm{rank}_X(p) \leq r\} \quad \text{and} \quad \Sigma_r(X) = \{p \in V : \overline{\mathrm{rank}}_X(p) \leq r\}.$$

A notion closely related to secant varieties is that of a *tangent variety*, defined for a nonsingular projective variety, or more generally, for any projective variety as

$$\tau(X) := \bigcup_{x \in X} \widehat{T}_x(X) \quad \text{or} \quad \tau(X) := \bigcup_{x \in X \setminus X_{\mathrm{sing}}} \widehat{T}_x(X)$$

respectively. Here $\widehat{T}_x(X)$ is the affine tangent space of X at x [25, Section 8.1.1] and X_{sing} is the *singular locus*, i.e., the subvariety of singular points in X , which has measure zero (in fact, positive codimension). When X is nonsingular, $\tau(X)$ is an algebraic variety. When X is singular, $\tau(X)$ is a quasiaffine variety. By abusing terminologies slightly, we call $\tau(X)$ a tangent variety in both cases. Among the varieties in Section 5.1 that we are interested in, the Chow variety $\mathrm{Ch}_d(V)$ is singular when $\dim V > 2$ and thus so are the Segre–Chow, Veronese–Chow, Segre–Veronese–Chow varieties; the rest are all nonsingular varieties.

In fact, for each of the varieties listed in Section 5.1, an element $A \in \tau(X)$ has a *normal form* that is either well-known or straightforward to determine, i.e., we may write down an expression for an arbitrary $A \in \tau(X)$.³

Proposition 5.1. *As each variety X in Section 5.1 is defined by a map, we will denote a tensor in the tangent space of that variety with a subscript given by the respective map. With this notation, the following is a list of normal forms for $\tau(X)$:*

$$\begin{aligned}
A_\sigma &= \sum_{i=1}^d v_1 \otimes \cdots \otimes v_{i-1} \otimes w_i \otimes v_{i+1} \otimes \cdots \otimes v_d, \\
A_\nu &= v^{\circ(d-1)} \circ w, \\
A_\kappa &= \sum_{i=1}^d v_1 \circ \cdots \circ v_{i-1} \circ w_i \circ v_{i+1} \circ \cdots \circ v_d, \\
A_\psi &= \sum_{i=1}^k v_1 \wedge \cdots \wedge v_{i-1} \wedge w_i \wedge v_{i+1} \wedge \cdots \wedge v_k, \\
A_{\sigma\nu} &= \sum_{i=1}^m v_1^{\otimes d_1} \otimes \cdots \otimes v_{i-1}^{\otimes d_{i-1}} \otimes (v_i^{\circ(d_i-1)} \circ w_i) \otimes v_{i+1}^{\otimes d_{i+1}} \otimes \cdots \otimes v_m^{\otimes d_m}, \\
A_{\sigma\kappa} &= \sum_{i=1}^m (v_{1,1} \circ \cdots \circ v_{d_1,1}) \otimes \cdots \otimes (v_{1,i-1} \circ \cdots \circ v_{d_{i-1},i-1}) \\
&\quad \otimes \left(\sum_{j=1}^{d_i} v_{1,i} \circ \cdots \circ v_{j-1,i} \circ w_{j,i} \circ v_{j+1,i} \circ \cdots \circ v_{d_i,i} \right) \\
&\quad \otimes (v_{1,i+1} \circ \cdots \circ v_{d_{i+1},i+1}) \otimes \cdots \otimes (v_{1,m} \circ \cdots \circ v_{d_m,m}), \\
A_{\sigma\psi} &= \sum_{i=1}^m (v_{1,1} \wedge \cdots \wedge v_{k_1,1}) \otimes \cdots \otimes (v_{1,i-1} \wedge \cdots \wedge v_{k_{i-1},i-1}) \\
&\quad \otimes \left(\sum_{j=1}^{k_i} v_{1,i} \wedge \cdots \wedge v_{j-1,i} \wedge w_{j,i} \wedge v_{j+1,i} \wedge \cdots \wedge v_{k_i,i} \right) \\
&\quad \otimes (v_{1,i+1} \wedge \cdots \wedge v_{k_{i+1},i+1}) \otimes \cdots \otimes (v_{1,m} \wedge \cdots \wedge v_{k_m,m}), \\
A_{\nu\kappa} &= (v_1 \circ \cdots \circ v_d)^{\circ(k-1)} \circ \left(\sum_{i=1}^d v_1 \circ \cdots \circ v_{i-1} \circ w_i \circ v_{i+1} \circ \cdots \circ v_d \right), \\
A_{\nu\psi} &= (v_1 \wedge \cdots \wedge v_k)^{\circ(d-1)} \circ \left(\sum_{i=1}^k v_1 \wedge \cdots \wedge v_{i-1} \wedge w_i \wedge v_{i+1} \wedge \cdots \wedge v_k \right), \\
A_{\sigma\nu\kappa} &= \sum_{i=1}^m (v_{1,1} \circ \cdots \circ v_{d_1,1})^{\otimes k_1} \otimes \cdots \otimes (v_{1,i-1} \circ \cdots \circ v_{d_{i-1},i-1})^{\otimes k_{i-1}} \\
&\quad \otimes \left[(v_{1,i} \circ \cdots \circ v_{d_i,i})^{\circ(k_i-1)} \circ \left(\sum_{j=1}^{d_i} v_{1,i} \circ \cdots \circ v_{j-1,i} \circ w_{j,i} \circ v_{j+1,i} \circ \cdots \circ v_{d_i,i} \right) \right]
\end{aligned}$$

³In general this is not possible for higher order tensors, i.e., we do not have an analogue of Jordan normal form for all d -tensors when $d > 2$; but in our case, we are restricting to a very small subset, namely, only the tensors in $\tau(X)$.

$$\begin{aligned}
 & \otimes (v_{1,i+1} \circ \cdots \circ v_{d_{i+1},i+1})^{\otimes k_{i+1}} \otimes \cdots \otimes (v_{1,m} \circ \cdots \circ v_{d_m,m})^{\otimes k_m}, \\
 A_{\sigma\nu\psi} = & \sum_{i=1}^m (v_{1,1} \wedge \cdots \wedge v_{k_1,1})^{\otimes d_1} \otimes \cdots \otimes (v_{1,i-1} \wedge \cdots \wedge v_{k_{i-1},i-1})^{\otimes d_{i-1}} \\
 & \otimes \left[(v_{1,i} \wedge \cdots \wedge v_{k_i,i})^{\circ(d_i-1)} \circ \left(\sum_{j=1}^{k_i} v_{1,i} \wedge \cdots \wedge v_{j-1,i} \wedge w_{j,i} \wedge v_{j+1,i} \wedge \cdots \wedge v_{k_i,i} \right) \right] \\
 & \otimes (v_{1,i+1} \wedge \cdots \wedge v_{k_{i+1},i+1})^{\otimes d_{i+1}} \otimes \cdots \otimes (v_{1,m} \wedge \cdots \wedge v_{k_m,m})^{\otimes d_m}.
 \end{aligned}$$

Proof. Let X be one of the varieties in Section 5.1. Then any $A \in \tau(X)$ is of the form

$$A = \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t},$$

where γ is any complex analytic curve $\gamma(t) \subseteq \widehat{X} \setminus \{0\}$ such that $\gamma(t)$ is nonconstant around 0 and $\gamma(0)$ is a nonsingular point in \widehat{X} . In fact, such a curve $\gamma(t)$ can be written down explicitly for any X in Section 5.1. Let $*$ denote either \circ or \wedge . Then any point in $\widehat{X} \setminus \{0\}$ takes the form

$$(v_{1,1} * \cdots * v_{k_1,1})^{\otimes d_1} \otimes \cdots \otimes (v_{1,m} * \cdots * v_{k_m,m})^{\otimes d_m},$$

and any curve $\gamma(t) \subset \widehat{X} \setminus \{0\}$ is of the form

$$\gamma(t) = (v_{1,1}(t) * \cdots * v_{k_1,1}(t))^{\otimes d_1} \otimes \cdots \otimes (v_{1,m}(t) * \cdots * v_{k_m,m}(t))^{\otimes d_m},$$

where $v_{j,i}(t) \in V_j$ is a complex analytic curve with $v_{j,i}(0) = v_{j,i}$. Thus

$$\begin{aligned}
 A &= \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t} \\
 &= \frac{d}{dt} (v_{1,1}(t) * \cdots * v_{k_1,1}(t))^{\otimes d_1} \Big|_{t=0} \otimes (v_{1,2} * \cdots * v_{k_2,2})^{\otimes d_2} \otimes \cdots \\
 &\quad \otimes (v_{1,m} * \cdots * v_{k_m,m})^{\otimes d_m} + \cdots + (v_{1,1} * \cdots * v_{k_1,1})^{\otimes d_1} \otimes \cdots \\
 &\quad \otimes (v_{1,m-1} * \cdots * v_{k_{m-1},m-1})^{\otimes d_{m-1}} \otimes \frac{d}{dt} (v_{1,m}(t) * \cdots * v_{k_m,m}(t))^{\otimes d_m} \Big|_{t=0} \\
 &= \sum_{i=1}^m (v_{1,1} * \cdots * v_{k_1,1})^{\otimes d_1} \otimes \cdots \otimes (v_{1,i-1} * \cdots * v_{k_{i-1},i-1})^{\otimes d_{i-1}} \\
 &\quad \otimes \left[(v_{1,i} * \cdots * v_{k_i,i})^{\circ(d_i-1)} \circ \left(\sum_{j=1}^{k_i} v_{1,i} * \cdots * v_{j-1,i} * w_{j,i} * v_{j+1,i} * \cdots * v_{k_i,i} \right) \right] \\
 &\quad \otimes (v_{1,i+1} * \cdots * v_{k_{i+1},i+1})^{\otimes d_{i+1}} \otimes \cdots \otimes (v_{1,m} * \cdots * v_{k_m,m})^{\otimes d_m}
 \end{aligned}$$

for some $w_{j,i} \in V_j$. □

The simplest examples with $\overline{\text{rank}}_X(A) \neq \text{rank}_X(A)$ may be found in $\tau(X)$. Note that since every tangent is a limit of 2-secants, we clearly have $\tau(X) \subseteq \Sigma_2(X)$; but in general $\tau(X) \not\subseteq \Sigma_2^\circ(X)$. The next proposition gives sufficient conditions for a tensor A in Proposition 5.1 so that $A \in \tau(X) \setminus \Sigma_2^\circ(X)$, i.e.,

$$\overline{\text{rank}}_X(A) = 2 < \text{rank}_X(A).$$

As each variety X in Section 5.1 is defined by a map, we will denote the X -rank and border X -rank of a variety with a subscript given by the respective map.

Proposition 5.2. *Let $d, k, m \geq 3$. The tensors in $\tau(X)$ for any X that is one of the varieties in Section 5.1 have X -border-rank two and X -rank strictly greater than two given the respective sufficient conditions:*

(i) If $\{v_i, w_i\}$ is linearly independent for $i = 1, \dots, d$, then

$$\overline{\text{rank}}_{\sigma}(A_{\sigma}) = 2 < \text{rank}_{\sigma}(A_{\sigma}).$$

(ii) If $\{v, w\}$ is linearly independent, then

$$\overline{\text{rank}}_{\nu}(A_{\nu}) = 2 < \text{rank}_{\nu}(A_{\nu}).$$

(iii) If $[v_1], \dots, [v_d], [w_1], \dots, [w_d]$ are distinct in $\mathbb{P}V$, then

$$\overline{\text{rank}}_{\kappa}(A_{\kappa}) = 2 < \text{rank}_{\kappa}(A_{\kappa}).$$

(iv) If $v_1 \wedge \dots \wedge v_k \wedge w_1 \wedge \dots \wedge w_k \neq 0$, then

$$\overline{\text{rank}}_{\psi}(A_{\psi}) = 2 < \text{rank}_{\psi}(A_{\psi}).$$

(v) If $[v_1], \dots, [v_m], [w_1], \dots, [w_m]$ are distinct in $\mathbb{P}V$, then

$$\overline{\text{rank}}_{\sigma\nu}(A_{\sigma\nu}) = 2 < \text{rank}_{\sigma\nu}(A_{\sigma\nu}).$$

(vi) If $[v_{1,1}], \dots, [v_{d_m,m}], [w_{1,1}], \dots, [w_{d_m,m}]$ are distinct in $\mathbb{P}V$, then

$$\overline{\text{rank}}_{\sigma\kappa}(A_{\sigma\kappa}) = 2 < \text{rank}_{\sigma\kappa}(A_{\sigma\kappa}).$$

(vii) If $v_{1,i} \wedge \dots \wedge v_{k_i,i} \wedge w_{1,i} \wedge \dots \wedge w_{k_i,i} \neq 0$ for $i = 1, \dots, m$, then

$$\overline{\text{rank}}_{\sigma\psi}(A_{\sigma\psi}) = 2 < \text{rank}_{\sigma\psi}(A_{\sigma\psi}).$$

(viii) If $[v_1], \dots, [v_d], [w_1], \dots, [w_d]$ are distinct in $\mathbb{P}V$, then

$$\overline{\text{rank}}_{\nu\kappa}(A_{\nu\kappa}) = 2 < \text{rank}_{\nu\kappa}(A_{\nu\kappa}).$$

(ix) If $v_1 \wedge \dots \wedge v_k \wedge w_1 \wedge \dots \wedge w_k \neq 0$, then

$$\overline{\text{rank}}_{\nu\psi}(A_{\nu\psi}) = 2 < \text{rank}_{\nu\psi}(A_{\nu\psi}).$$

(x) If $[v_{1,1}], \dots, [v_{d_m,m}], [w_{1,1}], \dots, [w_{d_m,m}]$ are distinct in $\mathbb{P}V$, then

$$\overline{\text{rank}}_{\sigma\nu\kappa}(A_{\sigma\nu\kappa}) = 2 < \text{rank}_{\sigma\nu\kappa}(A_{\sigma\nu\kappa}).$$

(xi) If $v_{1,i} \wedge \dots \wedge v_{k_i,i} \wedge w_{1,i} \wedge \dots \wedge w_{k_i,i} \neq 0$ for $i = 1, \dots, m$, then

$$\overline{\text{rank}}_{\sigma\nu\psi}(A_{\sigma\nu\psi}) = 2 < \text{rank}_{\sigma\nu\psi}(A_{\sigma\nu\psi}).$$

Proof. Since each of these tensors in Proposition 5.1 is a limit of the form

$$A = \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t},$$

we must have $\overline{\text{rank}}_X(A) \leq 2$. On the other hand, because \widehat{X} is complete as a metric space, $\overline{\text{rank}}_X(A) = 1$ if and only if $\text{rank}_X(A) = 1$. Hence requiring that $\text{rank}_X(A) > 1$ ensures that $\overline{\text{rank}}_X(A) > 1$. That the respective sufficient condition guarantees $\text{rank}_X(A) > 2$ for each X in Section 5.1 follows from straightforward linear algebra. \square

Note that for each X in Section 5.1, the respective sufficient condition in Proposition 5.2 is satisfied by a general $A \in \tau(X)$.

The *generic X -rank*, denoted $r_g(X)$, is the minimum integer r such that $\Sigma_r(X) = V$. Let $r, s \in \mathbb{N}$ be such that $1 < r < s \leq r_g(X)$, and let $p \in V$ with $\text{rank}_X(p) = s$. Then p does not necessarily have a best X -rank- r approximation. While such failures are well-known for Segre variety [10] and Veronese variety [7], we deduce from Proposition 5.2 that they occur for all varieties in Section 5.1 when the orders of the tensors are higher than two.

Theorem 5.3. *Let X be any of the projective varieties listed in Section 5.1 and let $d, k, m \geq 3$. Then there exist a tensor $A \in \text{span}(X)$, i.e., A has the required symmetry and/or skew-symmetry, and an $r < r_g(X)$ such that the best X -rank- r approximation of A does not exist, i.e.,*

$$\inf_{\text{rank}_X(B) \leq r} \|A - B\|$$

is not attained by any tensor $B \in \text{span}(X)$ with $\text{rank}_X(B) \leq r$.

Proof. For each projective variety X in Section 5.1, we need to exhibit a tensor $A \in \text{span}(X)$ with different X -rank and X -border rank; in which case A will not have a best X -rank- r approximation for $r = \overline{\text{rank}}_X(A)$. By Proposition 5.2, we see that by setting $r = 2$ and picking a general point $A \in \tau(X)$ that satisfies the sufficient conditions in Proposition 5.2, we obtain a tensor with no best X -rank-two approximation for each of the varieties X in Section 5.1. \square

While Theorem 5.3 shows that the phenomenon where a tensor fails to have a best X -rank- r approximation can and does happen for every variety in Section 5.1, our next result is that such failures occur with probability zero, a consequence of Theorem 4.2. A high-level explanation goes as follows: When $r < r_g(X)$, the set of points in $\Sigma_r(X)$ whose X -rank is r contains a Zariski open subset, and so the set of “bad points” in $\Sigma_r(X)$ — those whose X -rank is not r — is contained in a subvariety; Theorem 4.2 then guarantees that these “bad points” are avoided almost always. A formal statement follows next.

Theorem 5.4. *Let $X \subseteq \mathbb{P}V$ be an irreducible nondegenerate complex projective variety. Then a general point $p \in V$ has a unique best X -rank- r approximation whenever $r < r_g(X)$.*

When applied to the varieties in Section 5.1, we obtain the following corollary.

Corollary 5.5. *Let V and V_1, \dots, V_d be complex vector spaces.*

- (i) *Any general tensor $A \in V_1 \otimes \dots \otimes V_d$ has a unique best rank- r approximation when $r < r_g(\sigma(\mathbb{P}V_1 \times \dots \times \mathbb{P}V_d))$, i.e.,*

$$\inf_{v_{j,i} \in V_j} \left\| A - \sum_{i=1}^r v_{1,i} \otimes \dots \otimes v_{d,i} \right\|$$

can be attained.

- (ii) *Any general symmetric tensor $A \in \mathcal{S}^d(V)$ has a unique best symmetric rank- r approximation when $r < r_g(\nu(\mathbb{P}V))$, and a unique best Chow rank- r approximation when $r < r_g(\text{Ch}_d(V))$, i.e.,*

$$\inf_{v_i \in V} \left\| A - \sum_{i=1}^r v_i^{\otimes d} \right\| \quad \text{and} \quad \inf_{v_{j,i} \in V} \left\| A - \sum_{i=1}^r v_{1,i} \circ \dots \circ v_{d,i} \right\|$$

can be attained.

- (iii) *Any general alternating k -tensor $A \in \Lambda^k(V)$ has a unique best alternating rank- r approximation when $r < r_g(\text{Gr}_d(V))$, i.e.,*

$$\inf_{v_{j,i} \in V} \left\| A - \sum_{i=1}^r v_{1,i} \wedge \dots \wedge v_{k,i} \right\|$$

can be attained.

- (iv) *Any general $A \in \mathcal{S}^{d_1}(V) \otimes \dots \otimes \mathcal{S}^{d_m}(V)$ has a unique best Segre–Veronese rank- r approximation when $r < r_g(\sigma\nu(\mathbb{P}V_1 \times \dots \times \mathbb{P}V_m))$, and a unique best Segre–Chow rank- r approximation when $r < r_g(\sigma\kappa(\text{Ch}_{d_1}(V) \times \dots \times \text{Ch}_{d_m}(V)))$, i.e.,*

$$\inf_{v_{j,i} \in V} \left\| A - \sum_{i=1}^r v_{1,i}^{\otimes d_1} \otimes \dots \otimes v_{m,i}^{\otimes d_m} \right\|$$

and

$$\inf_{v_{j,l,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,1,i} \circ \cdots \circ v_{d_1,1,i}) \otimes \cdots \otimes (v_{1,m,i} \circ \cdots \circ v_{d_m,m,i}) \right\|$$

can be attained.

- (v) Any general tensor $A \in \Lambda^{k_1}(V) \otimes \cdots \otimes \Lambda^{k_m}(V)$ has a unique best Segre–Grassmann rank- r approximation when $r < r_g(\sigma\psi(\text{Gr}_{k_1}(V) \times \cdots \times \text{Gr}_{k_m}(V)))$, i.e.,

$$\inf_{v_{j,l,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,1,i} \wedge \cdots \wedge v_{k_1,1,i}) \otimes \cdots \otimes (v_{1,m,i} \wedge \cdots \wedge v_{k_m,m,i}) \right\|$$

can be attained.

- (vi) Any general tensor $A \in \mathbf{S}^k(\mathbf{S}^d(V))$ has a unique best Veronese–Chow rank- r approximation when $r < r_g(\nu\kappa(\text{Ch}_d(V)))$, i.e.,

$$\inf_{v_{j,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,i} \circ \cdots \circ v_{d,i})^{\otimes k} \right\|$$

can be attained.

- (vii) Any general tensor $A \in \mathbf{S}^d(\Lambda^k(V))$ has a unique best Veronese–Grassmann rank- r approximation when $r < r_g(\nu\psi(\text{Gr}_k(V)))$, i.e.,

$$\inf_{v_{j,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,i} \wedge \cdots \wedge v_{k,i})^{\otimes d} \right\|$$

can be attained.

- (viii) Any general tensor $A \in \mathbf{S}^{k_1}(\mathbf{S}^{d_1}(V)) \otimes \cdots \otimes \mathbf{S}^{k_m}(\mathbf{S}^{d_m}(V))$ has a unique best Segre–Veronese–Chow rank- r approximation when $r < r_g(\sigma\nu\kappa(\text{Ch}_{d_1}(V) \times \cdots \times \text{Ch}_{d_m}(V)))$, i.e.,

$$\inf_{v_{j,l,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,1,i} \circ \cdots \circ v_{d_1,1,i})^{\otimes k_1} \otimes \cdots \otimes (v_{1,m,i} \circ \cdots \circ v_{d_m,m,i})^{\otimes k_m} \right\|$$

can be attained.

- (ix) Any general tensor $A \in \mathbf{S}^{d_1}(\Lambda^{k_1}(V)) \otimes \cdots \otimes \mathbf{S}^{d_m}(\Lambda^{k_m}(V))$ has a unique best Segre–Veronese–Grassmann rank- r approximation when $r < r_g(\sigma\nu\psi(\text{Gr}_{k_1}(V) \times \cdots \times \text{Gr}_{k_m}(V)))$, i.e.,

$$\inf_{v_{j,l,i} \in V} \left\| A - \sum_{i=1}^r (v_{1,1,i} \wedge \cdots \wedge v_{k_1,1,i})^{\otimes d_1} \otimes \cdots \otimes (v_{1,m,i} \wedge \cdots \wedge v_{k_m,m,i})^{\otimes d_m} \right\|$$

can be attained.

5.3. Join approximations. Our general existence and uniqueness result in Theorem 4.2 has implications beyond approximation by secant varieties in tensor spaces. In this section, we will look at some other approximation problems that arise in practical applications but do not fall naturally into the class of approximation problems considered in Section 5.2

5.3.1. Sparse-plus-low-rank approximations. Let V and W be complex vector spaces. Let S be the linear subspace of k -sparse matrices with a fixed sparsity pattern in $V \otimes W$. Consider also $\Sigma_r(\sigma(\mathbb{P}V \times \mathbb{P}W))$, the set of matrices of rank not more than r . Given a matrix $A \in V \otimes W$, a best k -sparse-plus-rank- r approximation [3, 5] of A is a solution to the problem:

$$(9) \quad \inf_{\text{rank}(B) \leq r, C \in S} \|A - (B + C)\|.$$

In other words, we would like to find a best approximation of A by a matrix in $J(\Sigma_r(\sigma(\mathbb{P}V \times \mathbb{P}W)), \mathbb{P}S)$. Recall that this is a join variety, as defined in Section 5.2. By Theorem 4.2, we may deduce the following existence and uniqueness result for such approximation problems.

Corollary 5.6. *For a general complex matrix A , the problem (9) has a unique solution $B+C$ such that $\text{rank}(B) = r$ and $C \in S$ is a k -sparse matrix.*

5.3.2. *Gaussian r -factor analysis model with k observed variables.* Consider a Gaussian hidden variable model with k observed variables X_1, \dots, X_k and r hidden variables Y_1, \dots, Y_r where $(X_1, \dots, X_k, Y_1, \dots, Y_r)$ follows a joint multivariate normal distribution with positive definite covariance matrix. If the observed variables are conditionally independent given the hidden variables, this model is called the *Gaussian r -factor analysis model with k observed variables* [13] and denoted by $\mathbf{F}_{k,r}$. In fact, by [13, Proposition 1], $\mathbf{F}_{k,r}$ is the family of multivariate normal distributions $\mathcal{N}(\mu, \Sigma)$ on \mathbb{R}^k with $\mu \in \mathbb{R}^k$ and Σ belonging to

$$F_{k,r} := \{\Psi + LL^T \in \mathbb{R}^{k \times k} \mid \Psi \succ 0 \text{ and diagonal, } L \in \mathbb{R}^{k \times r}\}.$$

The standard approach in *algebraic statistics* [14] and also that in [13] is to drop any semialgebraic conditions and complexify. In this case, it means to drop the condition $\Sigma \succ 0$ and regard all quantities as complex valued, i.e.,

$$F_{k,r}(\mathbb{C}) = \{\Psi + LL^T \in \mathbb{C}^{k \times k} \mid \Psi \text{ diagonal, } L \in \mathbb{C}^{k \times r}\}.$$

This is the set of complexified covariance matrices for the model $\mathbf{F}_{k,r}$ and the algebraic approach undertaken in [13] effectively treats $F_{k,r}(\mathbb{C})$ as the parameter space of the Gaussian r -factor analysis model with k observed variables.

If we replace the space of matrices $V \otimes W$ in Section 5.3.1 by the space of symmetric matrices $S^2(V)$ and let $S \subseteq S^2(V)$ be the subspace of diagonal matrices, then we see that

$$F_{k,r}(\mathbb{C}) = J(\Sigma_r(\nu_2(\mathbb{P}V)), \mathbb{P}S).$$

It follows from Corollary 5.6 that every general complexified covariance matrix $\Sigma \in S^2(V)$ has a unique best approximation by $\Psi + LL^T$, a complexified covariance matrix for the model $\mathbf{F}_{k,r}$. To provide context, readers unfamiliar with factor analysis [24, Chapter 9] should note that its main parameter estimation problem is to determine the matrix of *loadings* L and the diagonal matrix of *specific variances* $\Psi = \text{diag}(\psi_1, \dots, \psi_k)$ from a sample covariance matrix Σ .

5.3.3. *Block-term tensor approximations.* Our final example of finding a best approximation in a join variety brings us back to tensors. A class of tensor approximation problems with groundbreaking applications in signal processing is the so-called *block-term decompositions* [9]. Unlike the factor analysis application in Section 5.3.2, which is really a problem over \mathbb{R} but is complexified to allow techniques of algebraic statistics, the use of block-term decompositions as a model in signal processing is naturally and necessarily over \mathbb{C} .

Any tensor $A \in V_1 \otimes \dots \otimes V_d$ induces a linear map $A_i: V_i^* \rightarrow V_1 \otimes \dots \otimes V_{i-1} \otimes V_{i+1} \otimes \dots \otimes V_d$. The rank of this linear map is of course just the dimension of its image $\dim A_i(V_i^*)$. The *multilinear rank* of A is then defined to be the d -tuple

$$\mu\text{rank}(A) := (\dim A_1(V_1^*), \dots, \dim A_d(V_d^*))$$

and the set

$$\text{Sub}_{r_1, \dots, r_d}(V_1 \otimes \dots \otimes V_d) := \{A \in V_1 \otimes \dots \otimes V_d \mid \mu\text{rank}(A) \leq (r_1, \dots, r_d)\}$$

is a complex algebraic variety called a *subspace variety*.

Given $(r_{1,1}, \dots, r_{1,d}), \dots, (r_{k,1}, \dots, r_{k,d})$, a tensor $B \in V_1 \otimes \dots \otimes V_d$ is said to have a block-term decomposition if

$$B = B_1 + \dots + B_k, \quad \mu\text{rank}(B_i) \leq (r_{i,1}, \dots, r_{i,d}), \quad i = 1, \dots, k.$$

It is known [9] that the best block-term approximation problem

$$(10) \quad \inf \{ \|A - (B_1 + \dots + B_k)\| \mid \mu\text{rank}(B_i) \leq (r_{i,1}, \dots, r_{i,d}), i = 1, \dots, k \}$$

does not have a solution for certain choices of A , i.e., the infimum above cannot be attained, much like the tensor approximation problems we discussed in Section 5.2.

Theorem 4.2, when applied to the join variety of k subspace varieties, gives us the following.

Corollary 5.7. *Let V_1, \dots, V_d be complex vector spaces. A general tensor $A \in V_1 \otimes \dots \otimes V_d$ has a unique best approximation in the join variety*

$$J(\text{Sub}_{r_{1,1}, \dots, r_{1,d}}(V_1 \otimes \dots \otimes V_d), \dots, \text{Sub}_{r_{k,1}, \dots, r_{k,d}}(V_1 \otimes \dots \otimes V_d)),$$

i.e., the best block-term approximation problem in (10) has a unique solution.

6. CONCLUSION

Our study in this article demonstrates a significant difference between tensor approximation problems over \mathbb{R} and over \mathbb{C} . This revelation is consistent with our knowledge of other aspects of tensors — tensor rank, tensor spectral norm, tensor nuclear norm are all known to be dependent on the base field [10, 16]; even the fact that the Grothendieck constants have different values over \mathbb{R} and over \mathbb{C} is a manifestation of this phenomenon [27].

The knowledge that over \mathbb{C} , the best rank- r approximation problems for tensors and, more generally, any best r -term approximation problems, are well-posed except on a set of measure zero should provide a reasonable amount of justification for applications that depend on such approximations. In fact, such “almost everywhere” guarantees are about as much as one may hope for, since “everywhere” is certainly false — as we saw, for every single approximation problem that we considered in this article, there are indeed instances where best approximations do not exist.

ACKNOWLEDGMENT

We are grateful to J. M. Landsberg for posing the question about general existence of best low-rank approximations of tensors over \mathbb{C} to us (when LHL visited College Station in 2010, and when he visited YQ in Grenoble in 2015). MM thanks W. Hackbusch for inspiring discussions on the topic.

The work in this article is generously supported by DARPA D15AP00109 and NSF IIS 1546413. LHL is supported by a DARPA Director’s Fellowship.

REFERENCES

- [1] E. BIERSTONE AND P. D. MILMAN, *Semianalytic and subanalytic sets*, Inst. Hautes Études Sci. Publ. Math., (1988), pp. 5–42, http://www.numdam.org/item?id=PMIHES_1988__67__5_0.
- [2] E. J. CANDÈS, *Compressive sampling*, in International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.
- [3] E. J. CANDÈS, X. LI, Y. MA, AND J. WRIGHT, *Robust principal component analysis?*, J. ACM, 58 (2011), pp. Art. 11, 37, <https://doi.org/10.1145/1970392.1970395>.
- [4] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080, <https://doi.org/10.1109/TIT.2010.2044061>.
- [5] V. CHANDRASEKARAN, S. SANGHAVI, P. A. PARRILO, AND A. S. WILLSKY, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2011), pp. 572–596, <https://doi.org/10.1137/090761793>.
- [6] A. COHEN, W. DAHMEN, AND R. DEVORE, *Compressed sensing and best k -term approximation*, J. Amer. Math. Soc., 22 (2009), pp. 211–231, <https://doi.org/10.1090/S0894-0347-08-00610-3>.
- [7] P. COMON, G. GOLUB, L.-H. LIM, AND B. MOURRAIN, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279, <https://doi.org/10.1137/060661569>.
- [8] G. CYBENKO, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Systems, 2 (1989), pp. 303–314, <https://doi.org/10.1007/BF02551274>.
- [9] L. DE LATHAUWER, *Decompositions of a higher-order tensor in block terms. II. Definitions and uniqueness*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1033–1066, <https://doi.org/10.1137/070690729>.
- [10] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127, <https://doi.org/10.1137/06066518X>.

- [11] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306, <https://doi.org/10.1109/TIT.2006.871582>.
- [12] D. L. DONOHO AND M. ELAD, *Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 2197–2202, <https://doi.org/10.1073/pnas.0437847100>.
- [13] M. DRTON, B. STURMFELS, AND S. SULLIVANT, *Algebraic factor analysis: tetrads, pentads and beyond*, Probab. Theory Related Fields, 138 (2007), pp. 463–493, <https://doi.org/10.1007/s00440-006-0033-2>.
- [14] M. DRTON, B. STURMFELS, AND S. SULLIVANT, *Lectures on algebraic statistics*, vol. 39 of Oberwolfach Seminars, Birkhäuser Verlag, Basel, 2009, <https://doi.org/10.1007/978-3-7643-8905-5>.
- [15] A. H. DURFEE, *Neighborhoods of algebraic sets*, Trans. Amer. Math. Soc., 276 (1983), pp. 517–530, <https://doi.org/10.2307/1999065>.
- [16] S. FRIEDLAND AND L.-H. LIM, *Nuclear norm of higher-order tensors*, Math. Comp., (2017), <https://doi.org/10.1090/mcom/3239>. To appear.
- [17] S. FRIEDLAND AND M. STAWISKA, *Some approximation problems in semi-algebraic geometry*, in Constructive approximation of functions, vol. 107 of Banach Center Publ., Polish Acad. Sci. Inst. Math., Warsaw, 2015, pp. 133–147, <https://doi.org/10.4064/bc107-0-9>.
- [18] I. M. GEL'FAND, M. M. KAPRANOV, AND A. V. ZELEVINSKY, *Discriminants, resultants, and multidimensional determinants*, Mathematics: Theory & Applications, Birkhäuser Boston, Inc., Boston, MA, 1994, <https://doi.org/10.1007/978-0-8176-4771-1>.
- [19] A. C. GILBERT, S. MUTHUKRISHNAN, AND M. J. STRAUSS, *Approximation of functions over redundant dictionaries using coherence*, in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), ACM, New York, 2003, pp. 243–252.
- [20] M. GORESKY AND R. MACPHERSON, *Stratified Morse theory*, vol. 14 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Springer-Verlag, Berlin, 1988, <https://doi.org/10.1007/978-3-642-71714-7>.
- [21] R. GRIBONVAL AND M. NIELSEN, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory, 49 (2003), pp. 3320–3325, <https://doi.org/10.1109/TIT.2003.820031>.
- [22] D. GROSS, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Inform. Theory, 57 (2011), pp. 1548–1566, <https://doi.org/10.1109/TIT.2011.2104999>.
- [23] J. HARRIS, *Algebraic geometry*, vol. 133 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1992, <https://doi.org/10.1007/978-1-4757-2189-8>. A first course.
- [24] R. A. JOHNSON AND D. W. WICHERN, *Applied multivariate statistical analysis*, Pearson Prentice Hall, Upper Saddle River, NJ, sixth ed., 2007.
- [25] J. M. LANDSBERG, *Tensors: geometry and applications*, vol. 128 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2012.
- [26] V. TEMLYAKOV, *Greedy approximation*, vol. 20 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2011, <https://doi.org/10.1017/CB09780511762291>.
- [27] J. ZHANG AND L.-H. LIM, *Grothendieck constant is norm of strassen matrix multiplication tensor*, (2017), <https://arxiv.org/abs/1711.04427>. Preprint.

COMPUTATIONAL AND APPLIED MATHEMATICS INITIATIVE, DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL, 60637-1514.

E-mail address: yangqi@galton.uchicago.edu

MAX PLANCK INSTITUTE FOR MATHEMATICS IN THE SCIENCES, 04103 LEIPZIG, GERMANY AND POLISH ACADEMY OF SCIENCES, 00656 WARSAW, POLAND.

E-mail address: Mateusz.Michalek@mis.mpg.de

COMPUTATIONAL AND APPLIED MATHEMATICS INITIATIVE, DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.

E-mail address, corresponding author: lekhang@galton.uchicago.edu