COUNTING REAL CRITICAL POINTS OF THE DISTANCE TO ORTHOGONALLY INVARIANT MATRIX SETS

DMITRIY DRUSVYATSKIY, HON-LEUNG LEE, AND REKHA R. THOMAS

ABSTRACT. Minimizing the Euclidean distance to a set arises frequently in applications. When the set is algebraic, a measure of complexity of this optimization problem is its number of critical points. In this paper we provide a general framework to compute and count the real smooth critical points of a data matrix on an orthogonally invariant set of matrices. The technique relies on "transfer principles" that allow calculations to be done in the space of singular values of the matrices in the orthogonally invariant set. The calculations often simplify greatly and yield transparent formulas. We illustrate the method on several examples, and compare our results to the recently introduced notion of *Euclidean distance degree* of an algebraic variety.

1. INTRODUCTION

Finding an element of a subset $\mathcal{V}$ in $\mathbb{R}^n$ closest to a specified point y is a common task in computational mathematics, often called the *Euclidean distance (ED) minimization problem*:

$$(1.1) \qquad \text{minimize } \sum_{i=1}^{n}(x_i - y_i)^2 \text{ subject to } x \in \mathcal{V}.$$

Our current work is motivated by the systematic study of the "critical points" of the problem (1.1) in an algebraic setting, initiated in [12] and continued in [11, 19, 26]. There, the basic assumption is that $\mathcal{V}$ is a *real variety* — zero set of finitely many polynomials with real coefficients. Consider now the set $\mathcal{V}_{\mathbb{C}}$ of complex points satisfying the defining equations of $\mathcal{V}$. Then a *critical point* of $y \in \mathbb{C}^n$ with respect to (1.1) is any smooth (possibly complex) point x of $\mathcal{V}_{\mathbb{C}}$ such that $y - x$ lies in the normal space of $\mathcal{V}_{\mathbb{C}}$ at x, meaning that $y - x$ lies in the span of the gradients of the defining equations at x. It was shown in [12] that for a general data point $y \in \mathbb{C}^n$, the number of (complex) critical points of (1.1) is a constant. This constant is called the *Euclidean distance degree* of $\mathcal{V}$, denoted by EDdegree($\mathcal{V}$), and is a measure of the algebraic complexity of expressing a minimizer of (1.1) as a function of y.

The work in this paper is geared towards understanding the (real) critical points of *orthogonally invariant* matrix sets. Our theme is best illustrated with an example. Fix positive integers $r \leq n \leq t$ and consider the matrix set

$$\mathbb{R}_r^{n \times t} := \{X \in \mathbb{R}^{n \times t} \ : \ \text{rank}(X) \leq r\}.$$

Finding the closest matrix of rank at most r to a given matrix Y arises in many applications. The set $\mathbb{R}_r^{n \times t}$ is a real variety, and the authors of [12] established that

EDdegree($\mathbb{R}_r^{n \times t}$) = $\binom{n}{r}$. They also provide a recipe for all the critical points of Y on $\mathbb{R}_r^{n \times t}$, which may be viewed as a generalization of the *Eckart-Young theorem*.

In the context of this paper, what is important about $\mathbb{R}_r^{n \times t}$ is that it is *orthogonally invariant*, meaning that if $X \in \mathbb{R}_r^{n \times t}$ then for all orthogonal matrices U and V of appropriate sizes, $UXV^\top$ is also in $\mathbb{R}_r^{n \times t}$. Alternately, membership of X in the set $\mathbb{R}_r^{n \times t}$ is determined solely by the vector of singular values $\sigma(X) = (\sigma_1(X), \ldots, \sigma_n(X))$. Indeed, a matrix X lies in $\mathbb{R}_r^{n \times t}$ if and only if its vector of singular values $\sigma(X)$ lies in the set

$$\mathbb{R}_r^n := \{x \in \mathbb{R}^n : \operatorname{rank}(x) \leq r\},$$

where $\operatorname{rank}(x)$ denotes the number of nonzero coordinates of x. Observe that the geometry of the piecewise linear set $\mathbb{R}_r^n$ is much simpler than that of the highly nonlinear set $\mathbb{R}_r^{n \times t}$. In particular, $\mathbb{R}_r^n$ is also a variety and EDdegree($\mathbb{R}_r^n$) = $\binom{n}{r}$. The equality EDdegree($\mathbb{R}_r^n$) = EDdegree($\mathbb{R}_r^{n \times t}$) is not accidental; in this paper, we elucidate this phenomenon.

As alluded to above, our focus in this paper is on a (naturally defined) real analog of critical points for (1.1). Namely, we say that $x \in \mathbb{R}^n$ is a (real) *ED critical point* of $y \in \mathbb{R}^n$ relative to a set $\mathcal{V} \subset \mathbb{R}^n$ (not necessarily a variety) if x is a smooth point of $\mathcal{V}$ and $y - x$ lies in the normal space to $\mathcal{V}$ at x; see Definition 2.1 for details. Our main result shows that one can always obtain the ED critical points of an orthogonally invariant matrix set by restricting to diagonal matrices, or equivalently, to an *absolutely symmetric* set obtained from the singular values of the matrices. In the last section, we explore the connection between ED critical points and the critical points in the sense of [12] that happen to be real, a surprisingly subtle topic. In particular, ED critical point calculations often help to understand EDdegree($\mathcal{V}$), when $\mathcal{V}$ is a variety.

Sets constrained via their singular values are numerous in applications. Define $\mathcal{E} \subseteq \mathbb{R}^{3 \times 3}$ to be the set of rank deficient matrices with two equal singular values:

(1.2) $\mathcal{E} := \{X \in \mathbb{R}^{3 \times 3} : \sigma_1(X) = \sigma_2(X), \sigma_3(X) = 0\}.$

A matrix $X \in \mathcal{E}$ is called an *essential matrix* in 3D computer vision and represents a pair of calibrated pinhole cameras [17, Chapter 9]. The set $\mathcal{E}$ also happens to be a real variety cut out by the following ten cubic equations in the entries of X [14, Proposition 4]:

(1.3) $\det(X) = 0, \ 2XX^\top X - \operatorname{tr}(XX^\top)X = 0.$

Observe that $\mathcal{E}$ is orthogonally invariant as membership of a matrix X in $\mathcal{E}$ depends only on its singular values. We will show that one can obtain the ED critical points of $\mathcal{E}$ from the simpler set $E_{3,2}$ of vectors in $\mathbb{R}^3$ with one coordinate zero and the other two equal in absolute value.

The idea of studying orthogonally invariant matrix sets $\mathcal{M} \subseteq \mathbb{R}^{n \times t}$ via their diagonal restrictions $S = \{x : \operatorname{Diag} x \in \mathcal{M}\}$ is not new, and goes back at least to von Neumann's theorem on *unitarily invariant matrix norms* [30]. In recent years, the general theme has become clear: various analytic properties of $\mathcal{M}$ and S are in one-to-one correspondence. This philosophy is sometimes called the "transfer principle"; see for instance, [8]. For example, $\mathcal{M}$ is C^p-smooth around a matrix X if and only if S is C^p-smooth around $\sigma(X)$ [6, 21, 27, 29]. Other properties, such as *convexity* [9], *positive reach* [7], *partial smoothness* [6], and *Whitney conditions* [13] follow the same paradigm. We note in passing that the setting of eigenvalue constrained sets

of nonsymmetric matrices is more complicated; see e.g. [3, 23]. In the current work, we derive a transfer theorem for the ED critical points of an orthogonally invariant matrix set $\mathcal{M}$ and illustrate it on several examples. The manuscripts [7, 22, 24] play a central role in our work. The transfer paradigm has many useful features. The first is that in many instances, the calculation over the set $S \subseteq \mathbb{R}^n$ is simpler than the original one over $\mathcal{M}$. This can make the formulas for the number of ED critical points of $\mathcal{M}$ much more transparent as compared to the direct calculation in matrix space. The case of $\mathbb{R}_r^{n \times t}$ is an example of this. Secondly, our method focuses on (real) ED critical points as opposed to all complex critical points in the setting of algebraic varieties. This is useful in applications and allows us to gauge the difference between the real and complex situations (under appropriate conditions).

This paper is structured as follows. In Section 2 we establish basic notation and define the key notion of ED critical points of a set in $\mathbb{R}^n$ with respect to a data point y. Then in Section 3 we derive our main theorem that transfers the study of ED critical points of an orthogonally invariant set of matrices to that of its diagonal restriction. Section 4 illustrates the technique on some concrete examples deriving formulas for the number of ED critical points in each case. Finally in Section 5 we establish the relationships between our work and the results in [12], by restricting to algebraic varieties. In particular, we compare our formulas to those for EDdegree in several instances. We note in passing, that many of the results in the paper also hold for complex matrices with respect to the standard Hermitian inner product.

2. ED CRITICAL POINTS OF SUBSETS OF $\mathbb{R}^n$

Throughout this paper we consider the n-dimensional Euclidean space $\mathbb{R}^n$, with a fixed orthonormal basis. Let $\langle \cdot, \cdot \rangle$ denote the inner product in this setting, and $\| \cdot \|$ denote the induced norm. The *distance* and the *projection* of a (data) point $y \in \mathbb{R}^n$ onto a subset $S \subseteq \mathbb{R}^n$, are defined by

$$\operatorname{dist}_S(y) := \inf_{x \in S} \|y - x\|, \qquad \text{and}$$

$$\operatorname{proj}_S(y) := \{x \in S \,:\, \operatorname{dist}_S(y) = \|y - x\|\}.$$

Computing the distance of y to S amounts to solving the optimization problem:

$$(2.1) \qquad \inf_{x \in S} \frac{1}{2} \|y - x\|^2 = \inf_{x \in S} \frac{1}{2} \sum_{i=1}^{n} (y_i - x_i)^2.$$

A classical first-order necessary condition for a putative point $x \in S$ to be optimal for (2.1) is that the gradient of the object function, namely $x - y$, makes an acute angle with every vector v in the *tangent cone*[1]

$$\mathcal{T}_S(x) := \mathbb{R}_+ \left\{ \lim_{z_i \to x} \frac{z_i - x}{\|z_i - x\|} \,:\, z_i \in S \right\}.$$

One can regard such points as generalized critical points of the distance minimization problem (2.1). On the other hand, in order to compare and unify our work with that in [12], we will impose an extra smoothness condition on a point $x \in S$ in order for it to be considered critical for (2.1). To this end, throughout the manuscript, we fix $p \in \{2, 3, \ldots, \infty, \omega\}$, and say that a point $x \in S$ is C^p-*smooth* if there is a neighborhood Ω of x such that $S \cap \Omega$ is an embedded C^p-smooth manifold (recall that C^ω means *real analytic*). In this case, the tangent cone $\mathcal{T}_S(x)$ is the usual

---

[1]If x is an isolated point of S, then $\mathcal{T}_S(x)$ is declared to consist only of the origin.
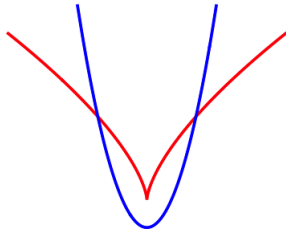
FIGURE 1. A parabola in the plane with its evolute or ED discriminant.

tangent space in the sense of differentiable manifolds, and the criticality condition above amounts to the inclusion $y - x \in \mathcal{N}_S(x)$, where $\mathcal{N}_S(x)$ denotes the *normal space* to S at x — the orthogonal complement of $\mathcal{T}_S(x)$.

From now on we abbreviate "C^p-smooth" to "smooth". We will use S^* to denote the set of all smooth points of S. Here is then our main definition, attuned to the one considered in [12].

Definition 2.1 (ED critical points). Consider a set $S \subseteq \mathbb{R}^n$ and a point y in $\mathbb{R}^n$. A point $x \in S$ is an *ED critical point* of y on S if the following conditions hold:

 (i) $x \in S^*$, and
 (ii) $y - x \in \mathcal{N}_S(x)$.

The symbol $\mathrm{C}_S(y)$ will denote the set of all ED critical points of y on S, while the cardinality of $\mathrm{C}_S(y)$ will be denoted by $C_S^{\#}(y)$.

By definition, all ED critical points of y on S are real and smooth. The number of ED critical points $C_S^{\#}(y)$ varies with y. For example, consider the parabola S shown in Figure 1, with an additional curve called its *ED discriminant* or *evolute*. All points y above the evolute have three ED critical points while below the evolute they have one ED critical point. Since S is an algebraic variety, we can compute its EDdegree which is three, i.e., $S_{\mathbb{C}}$ has three distinct regular complex critical points almost everywhere. We will comment more on the ED discriminant in the Appendix.

Letting Θ be the collection of all Lebesgue null subsets of $\mathbb{R}^n$, the following worst-case measure of criticality arises naturally:

$$C^{\#}(S) := \inf_{\Gamma \in \Theta} \sup_{y \in \Gamma^c} C_S^{\#}(y),$$

where Γ^c is the set complement of Γ in the ambient space. Indeed, intuitively if one believes that any single zero measure set Γ can be discarded, then $C^{\#}(S)$ measures the maximal value of $C_S^{\#}(y)$ that has a non-negligible chance of being encountered. For our purposes, one could think of $C^{\#}(S)$ as a real analog of the EDdegree considered in [12]. We discuss this further in Section 5.

We note in passing that this criticality measure can be infinite in pathological situations (e.g. union of countably many co-centric circles in $\mathbb{R}^2$). On the other hand, for structured sets, such as those that are semi-algebraic, this number is finite.

Proposition 2.2 (ED critical points of semi-algebraic sets). *For any semi-algebraic set S in $\mathbb{R}^n$, the number $C^{\#}(S)$ is finite.*

Proof. Standard quantifier elimination shows that the set S^* is semi-algebraic. Define the manifolds $M_i := \{x \in S^* : \dim \mathcal{T}_S(x) = i\}$ for $i = 1, \ldots, n$. Again,

quantifier elimination shows that the normal bundle

$$\Omega := \bigsqcup_i \{(x, v) : x \in M_i, \, v \in \mathcal{N}_{M_i}(x)\}$$

is a semi-algebraic subset of $\mathbb{R}^n \times \mathbb{R}^n$ having dimension n. Consider now the mapping $\phi \colon \Omega \to \mathbb{R}^n$ defined by $\phi(x, v) = x + v$. Notice that $C_S(y)$ coincides with the projection of the preimage $\phi^{-1}(y)$ onto x. For dimensional reasons, there is a full-measure set $D \subseteq \mathbb{R}^n$ so that for every $y \in D$ the preimage $\phi^{-1}(y)$ has finite cardinality. Moreover, since preimages of a semi-algebraic map have a uniformly bounded number of connected components (see e.g. [4, Theorem 3.12]), the quantity $C_S^{\#}(y)$ is uniformly bounded on D. The result follows. $\qquad\square$

Next we consider ED critical points on a union of finitely many sets. Consider a finite collection of sets $\{S_i\}_{i \in \mathcal{I}}$ in $\mathbb{R}^n$ and define the union $U := \bigcup_{i \in \mathcal{I}} S_i$. In general, the two sets $C_U(y)$ and $\bigcup_{i \in \mathcal{I}} C_{S_i}(y)$ can be vastly different because of the way the sets S_i intersect. For instance, think of two half spaces whose union is $\mathbb{R}^n$. A simple situation in which more can be said is when locally around each critical point $x \in C_U(y)$, the set U coincides with S_i for some $i \in \mathcal{I}$. In that situation, if $x \in C_U(y)$ then x is also in $C_{S_i}(y)$. An important situation in this paper is the case of all S_i being affine subspaces. We say that a finite collection of sets $\{S_i\}_{i \in \mathcal{I}}$ in $\mathbb{R}^n$ is *minimally defined* if no S_i is contained in any S_j for distinct indices i and j.

Proposition 2.3 (ED critical points of affine complexes). *Consider a finite collection of affine subspaces $\{S_i\}_{i \in \mathcal{I}}$ in $\mathbb{R}^n$, that is minimally defined, and let $U := \bigcup_{i \in \mathcal{I}} S_i$. Then we have*

$$(2.2) \qquad U^* = \{x \in \mathbb{R}^n : \text{ there exists unique } i \in \mathcal{I} \text{ with } x \in S_i\}.$$

Consequently for any $y \in \mathbb{R}^n$, we have

$$(2.3) \qquad \mathrm{C}_U(y) = \bigsqcup_{i \in \mathcal{I}} \left(\mathrm{proj}_{S_i}(y) \cap U^*\right),$$

and the equality

$$C^{\#}(U) = |\mathcal{I}|.$$

Proof. The inclusion $\supseteq$ in (2.2) follows since the sets S_i are affine. To see the reverse inclusion, observe that the tangent cone to U at any point x coincides with the union $\bigcup_{i \in \mathcal{I} : x \in S_i} (S_i - x)$. For $x \in U^*$, the cone $\mathcal{T}_U(x)$ is itself a linear subspace, and hence, by the minimality of the collection $\{S_i\}_{i \in \mathcal{I}}$, there exists a unique index $i \in \mathcal{I}$ satisfying $x \in S_i$. This establishes (2.2). Equation (2.3) then follows immediately.

To see the last claim, consider the set

$$Z := \{y \in \mathbb{R}^n : \exists i \in \mathcal{I} \text{ with } \mathrm{proj}_{S_i}(y) \cap U^* = \emptyset\}.$$

We will show that Z is a finite union of proper affine subspaces of $\mathbb{R}^n$. To see this, consider a point $y \in Z$ along with an index $i \in \mathcal{I}$ satisfying $\mathrm{proj}_{S_i}(y) \cap U^* = \emptyset$. From (2.2), we conclude that there exists $j \in \mathcal{I}$, distinct from i, satisfying $\mathrm{proj}_{S_i}(y) \cap S_j \neq \emptyset$. Thus y lies in the set $S_i^{\perp} + (S_i \cap S_j)$. Since the collection $\{S_i\}$ is minimally defined, the intersection $S_i \cap S_j$ has dimension strictly smaller than that of S_i. Consequently the affine space $S_i^{\perp} + (S_i \cap S_j)$ has dimension strictly smaller than n. Taking the union over all pairs of distinct indices $i, j \in \mathcal{I}$, we deduce that Z is a finite union of proper affine subspaces of $\mathbb{R}^n$. Therefore, by (2.3), $C_U^{\#}(y) = |\mathcal{I}|$ for all $y \notin Z$ which proves that $C^{\#}(U) = |\mathcal{I}|$. $\qquad\square$
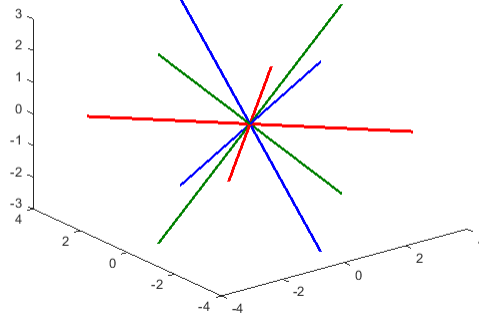
FIGURE 2. the set $E_{3,2}$

The following two elementary examples illustrate Proposition 2.3; these are essentially the piecewise linear examples alluded to in the introduction. In the next section, we will use them to obtain the ED critical points of the (nonlinear) matrix sets $\mathbb{R}_r^{n \times t}$ and $\mathcal{E}$. In what follows we set $[n] := \{1, \ldots, n\}$.

Example 2.4 (Union of r-dimensional coordinate subspaces). For $x \in \mathbb{R}^n$, define the rank of x, denoted by $\text{rank}(x)$, to be the number of nonzero coordinates of x. Fix an integer $r \in [n]$ and recall the set

(2.4) $$\mathbb{R}_r^n = \{x \in \mathbb{R}^n \,:\, \text{rank}(x) \leq r\}.$$

Define $\mathcal{I}$ to be the collection of distinct cardinality r subsets of $[n]$. Then $|\mathcal{I}| = \binom{n}{r}$, and

$$\mathbb{R}_r^n = \bigcup_{S \in \mathcal{I}} \text{span}\,\{e_i\}_{i \in S},$$

where e_i denotes the i'th coordinate vector in $\mathbb{R}^n$. This representation of $\mathbb{R}_r^n$ is minimal and therefore, Proposition 2.3 implies the equality $C^{\#}(\mathbb{R}_r^n) = |\mathcal{I}| = \binom{n}{r}$.

Example 2.5 (k nonzero entries equal up to sign). Fix an integer $k \in [n]$, and let $E_{n,k}$ be the set of points in $\mathbb{R}^n$ with the property that k of their coordinates are equal in absolute value and the other $n - k$ coordinates are zero. Note that $E_{n,k}$ can be written as the union of $2^{k-1}\binom{n}{k}$ linear subspaces (minimally defined). Proposition 2.3 then implies that $C^{\#}(E_{n,k}) = 2^{k-1}\binom{n}{k}$. As a special case, the set $E_{3,2}$, which is a union of six lines in $\mathbb{R}^3$ (see Figure 2), satisfies $C^{\#}(E_{3,2}) = 6$.

3. ED CRITICAL POINTS OF ORTHOGONALLY INVARIANT MATRIX SETS

In this section we describe our main result which yields an elegant technique for counting the ED critical points of orthogonally invariant matrix sets, based on the tools established in [7, 22, 24]. Setting the notation, let $\mathbb{R}^{n \times t}$ denote the set of real $n \times t$ matrices, where we assume without loss of generality that $n \leq t$. The singular value map $\sigma \colon \mathbb{R}^{n \times t} \to \mathbb{R}^n$ assigns to each matrix $X \in \mathbb{R}^{n \times t}$ the vector of its singular values $\sigma(X) := (\sigma_1(X), \ldots, \sigma_n(X))$ arranged in non-increasing order.

The corresponding inverse map is defined as

$$\sigma^{-1}(S) := \{X \in \mathbb{R}^{n \times t} : \sigma(X) \in S\} \qquad \text{for any subset } S \text{ of } \mathbb{R}^n.$$

We will be interested in subsets of $\mathbb{R}^{n \times t}$ that are invariant under multiplication on the left and right by orthogonal matrices. In what follows, we let $\mathcal{O}^s$ denote the group of real $s \times s$ orthogonal matrices.

Definition 3.1 (Orthogonal invariance).
A set $\mathcal{M} \subseteq \mathbb{R}^{n \times t}$ is *orthogonally invariant* if

$$\mathcal{M} = U\mathcal{M}V^\top \qquad \text{for all } U \in \mathcal{O}^n \text{ and } V \in \mathcal{O}^t.$$

An example of an orthogonally invariant matrix set is the essential variety

$$\mathcal{E} = \{X \in \mathbb{R}^{3 \times 3} \ : \ \sigma_1(X) = \sigma_2(X), \sigma_3(X) = 0\}$$

considered in the introduction. At first sight, $\mathcal{E}$ is a complicated set; it is a highly nonlinear real variety cut out by the cubic polynomials in (1.3). Since orthogonally invariant matrix sets are precisely those matrix sets for which membership is determined solely by the singular values of its elements, the restriction of such sets to the subspace of diagonal matrices plays an important role. Our strategy will be to exploit this observation to calculate the ED critical points of orthogonally invariant matrix sets.

A mapping $\pi : [n] \to \{\pm 1, \dots, \pm n\}$ is a *signed permutation* if the assignment $i \mapsto |\pi(i)|$ is a permutation on $[n]$ in the usual sense. We let $\Pi_n^\pm$ denote the set of signed permutations on $[n]$. Note that any signed permutation $\pi \in \Pi_n^\pm$ induces a linear map $\mathbb{R}^n \to \mathbb{R}^n$ which we also denote by π.

Definition 3.2 (Absolute symmetry). A set $S \subseteq \mathbb{R}^n$ is said to be *absolutely symmetric* if

$$S = \pi S \qquad \text{for all } \pi \in \Pi_n^\pm.$$

For any set $S \subseteq \mathbb{R}^n$, we call $\Pi_n^\pm S := \{\pi x \ : \ \pi \in \Pi_n^\pm \text{ and } x \in S\}$ the *absolute symmetrization* of S.

For a vector $x \in \mathbb{R}^n$, let $\mathrm{Diag}\,x \in \mathbb{R}^{n \times t}$ denote the matrix with x in its principal diagonal and zeros elsewhere. If $\mathcal{M} \subseteq \mathbb{R}^{n \times t}$ is orthogonally invariant, then the set $\{x \in \mathbb{R}^n : \mathrm{Diag}\,x \in \mathcal{M}\}$ is absolutely symmetric, and we have the following basic observation (see e.g. [24, Proposition 5.1]).

Theorem 3.3 (Diagonal Correspondence). *A set $\mathcal{M} \subseteq \mathbb{R}^{n \times t}$ is orthogonally invariant if and only if there exists an absolutely symmetric set $S \subseteq \mathbb{R}^n$ such that $\mathcal{M} = \sigma^{-1}(S)$.*

Thus the assignment σ^{-1} from the family of absolutely symmetric sets in $\mathbb{R}^n$ to the family of orthogonally invariant sets in $\mathbb{R}^{n \times t}$ is a bijection. As discussed in the introduction, it so happens that various analytic properties of absolutely symmetric sets S and orthogonally invariant sets $\sigma^{-1}(S)$ are in one-to-one correspondence; see e.g. [6–9, 13, 21, 27, 29, 30]. This may seem somewhat surprising since σ is a highly nonsmooth mapping; it is the absolute symmetry of the underlying set S that makes up for the fact. As an illustration, we show in Theorem 3.4 how the transfer principle for algebraicity can be used to see that the set of essential matrices $\mathcal{E}$ is a variety, even without explicitly knowing its defining equations (1.3). The proof is entirely analogous to the symmetric case in [8, Proposition 1.1]; we provide an argument for completeness.

Recall that a set $\mathcal{V} \subseteq \mathbb{R}^n$ is called a *real variety* if there exist polynomials $f_1, \ldots, f_s \in \mathbb{R}[x_1, \ldots, x_n]$ such that $\mathcal{V} = \{x \in \mathbb{R}^n : f_1(x) = \ldots = f_s(x) = 0\}$. Note that if $x \in \mathcal{V}$, then $\sum_{i=1}^{s} f_i^2(x) = 0$, and conversely if a point $x \in \mathbb{R}^n$ satisfies $\sum_{i=1}^{s} f_i^2(x) = 0$, then $f_i(x) = 0$ for all $i = 1, \ldots, s$, and hence, $x \in \mathcal{V}$. Therefore, $\mathcal{V}$ can be described as the set of real zeros of a single sum of squares polynomial with real coefficients, which we call a defining polynomial of $\mathcal{V}$.

Theorem 3.4 (Transfer of algebraicity). *Suppose $S \subseteq \mathbb{R}^n$ is an absolutely symmetric set. Then $\sigma^{-1}(S) \subseteq \mathbb{R}^{n \times t}$ is a real variety if and only if S is a real variety.*

Proof. If $\sigma^{-1}(S)$ is a real variety, then $S = \{x \in \mathbb{R}^n : \text{Diag}\, x \in \sigma^{-1}(S)\}$ is the real variety cut out by the equations defining $\sigma^{-1}(S)$ when x_{ij} are set to zero for all $i \neq j$. Suppose conversely that S is a real variety with defining polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$. Consider the polynomial $\hat{f}(x) := \sum_{\pi \in \Pi_n^{\pm}} f^2(\pi x)$. Note that $\hat{f}(x)$ is also a defining polynomial of S, and hence, $\sigma^{-1}(S)$ is the zero level set of $\hat{f} \circ \sigma$. To finish the proof, we just need to show that $\hat{f} \circ \sigma(X)$ is a polynomial in the entries of X. To this end, since $\hat{f}$ is invariant under sign changes, it is easy to see that $\hat{f}$ is a symmetric polynomial in the squares $x_1^2, \ldots, x_n^2$, that is we may write $\hat{f}(x) = g(x_1^2, \ldots, x_n^2)$ for some symmetric polynomial g. By the fundamental theorem of symmetric polynomials (see e.g. [25]) we may write g as a polynomial of elementary symmetric polynomials $\epsilon_1, \ldots, \epsilon_n$. On the other hand, the expressions $\epsilon_i(\sigma_1^2(X), \ldots, \sigma_n^2(X))$ coincide with the coefficients of the characteristic polynomial of $X^\top X$ and are hence, polynomial expressions in the entries of X. $\qquad\square$

Another useful illustration of transfer principles concerns the distance to orthogonally invariant matrix sets. Recall that for a set $S \subseteq \mathbb{R}^n$, the *distance* and *projection* of a point $y \in \mathbb{R}^n$ to (respectively, onto) S are defined by $\text{dist}_S(y) := \inf_{x \in S} \|y - x\|$, and $\text{proj}_S(y) := \{x \in S : \text{dist}_S(y) = \|y - x\|\}$. (The distance and the projection in the matrix space $\mathbb{R}^{n \times t}$ are defined analogously with respect to the *Frobenius norm* $\|Y\| := \sqrt{\sum_{i,j} Y_{ij}^2}$.) Then for an absolutely symmetric set $S \subseteq \mathbb{R}^n$, the following holds [7, Proposition 8]:

$$(3.1) \qquad\qquad \text{dist}_{\sigma^{-1}(S)}(Y) = \text{dist}_S(\sigma(Y)).$$

This in turn implies the following result, which was essentially proved in [7, Proposition 8], though not formally recorded. We provide a proof sketch for completeness.

Proposition 3.5 (Projections onto orthogonally invariant matrix sets). *If $S \subseteq \mathbb{R}^n$ is an absolutely symmetric set, then for any matrix $Y \in \mathbb{R}^{n \times t}$, the projection $\text{proj}_{\sigma^{-1}(S)}(Y)$ is precisely the set*

$$(3.2) \qquad \left\{ U(\text{Diag}\, x)V^\top : U \in \mathcal{O}^n, V \in \mathcal{O}^t \text{ where } \begin{matrix} Y = U(\text{Diag}\, \sigma(Y))V^\top, \\ x \in \text{proj}_S(\sigma(Y)) \end{matrix} \right\}.$$

Proof. Consider first matrices $U \in \mathcal{O}^n, V \in \mathcal{O}^t$ with $Y = U(\text{Diag}\, \sigma(Y))V^\top$ and a vector $x \in \text{proj}_S(\sigma(Y))$. Define $X := U(\text{Diag}\, x)V^\top$ and observe the equalities:

$$\|X - Y\| = \|x - \sigma(Y)\| = \text{dist}_S(\sigma(Y)) = \text{dist}_{\sigma^{-1}(S)}(Y),$$

where the last equality follows from (3.1). Hence the inclusion $X \in \text{proj}_{\sigma^{-1}(S)}(Y)$ is valid, as claimed. Conversely, for any matrix $X \in \text{proj}_{\sigma^{-1}(S)}(Y)$ observe

$$\text{dist}_{\sigma^{-1}(S)}(Y) = \|X - Y\| \geq \|\sigma(X) - \sigma(Y)\| \geq \text{dist}_S(\sigma(Y)),$$

where the first inequality follows from the Von Neumann's trace inequality [18, p. 182]; see also [24, Theorem 4.6]. Equation (3.1) implies equality throughout. In particular we get $\sigma(X) \in \text{proj}_S(\sigma(Y))$. Moreover, applying the equality characterization in the trace inequality [24, Theorem 4.6], we conclude that X and Y admit a simultaneous ordered singular value decomposition, that is, there exist matrices $U \in \mathcal{O}^n, V \in \mathcal{O}^t$ with $Y = U(\text{Diag } \sigma(Y))V^\top$ and $X = U(\text{Diag } \sigma(X))V^\top$. This completes the proof. $\square$

A consequence of Proposition 3.5 is a convenient representation of the normal space at a point in an orthogonally invariant manifold. First of, the set $\sigma^{-1}(S)$ is smooth around X if and only if S is smooth around $\sigma(X)$ (see e.g. [6, Theorem 2.4]). Next recall that at any C^2-smooth point x of a set M, the following equivalence holds:

$$z \in \mathcal{N}_M(x) \quad \Longleftrightarrow \quad \{x\} = \text{proj}_M(x + \lambda z) \quad \text{for some} \quad \lambda > 0.$$

Proposition 3.5 and a short computation implies that at any smooth point X of $\sigma^{-1}(S)$ the following formula holds:

$$\mathcal{N}_{\sigma^{-1}(S)}(X) = \left\{ U(\text{Diag } z)V^\top \; : \; U \in \mathcal{O}^n, V \in \mathcal{O}^t \text{ with } \begin{array}{l} X = U(\text{Diag } \sigma(X))V^\top, \\ z \in \mathcal{N}_S(\sigma(X)) \end{array} \right\}.$$

A more general expression without smoothness assumptions can be found in [24, Theorem 7.1]. Using the above facts, we now derive a transfer principle for ED critical points which is the main result of this section. The proof uses the following lemma.

Lemma 3.6 (Transfer of Lebesgue null sets). *Consider an absolutely symmetric set $S \subseteq \mathbb{R}^n$. Then S is Lebesgue null if and only if $\sigma^{-1}(S)$ is Lebesgue null.*

We have placed the proof of the lemma above at the end of the section so as to not stray from the narrative.

Theorem 3.7. (ED critical points of orthogonally invariant matrix sets)
Consider an absolutely symmetric set $S \subseteq \mathbb{R}^n$ and a matrix $Y \in \mathbb{R}^{n \times t}$ along with a singular value decomposition $Y = \overline{U}(\text{Diag } \sigma(Y))\overline{V}^\top$. Suppose moreover, that Y has all distinct singular values. Then for the orthogonally invariant set $\mathcal{M} := \sigma^{-1}(S)$, we have

$$(3.3) \qquad C_{\mathcal{M}}(Y) = \{\overline{U}(\text{Diag } \omega)\overline{V}^\top \; : \; \omega \in C_S(\sigma(Y))\}.$$

Consequently, we also obtain the equality $C^\#(\mathcal{M}) = C^\#(\mathcal{S})$.

Proof. Consider a critical point $X \in C_{\mathcal{M}}(Y)$. Then there exist orthogonal matrices U and V, and a vector $z \in \mathcal{N}_S(\sigma(X))$ satisfying

$$X = U(\text{Diag } \sigma(X))V^\top \quad \text{and} \quad Y = X + U(\text{Diag } z)V^\top.$$

Hence we deduce

$$Y = U(\text{Diag } (\sigma(X) + z))V^\top.$$

By uniqueness of singular values and singular vectors, there exists a signed permutation matrix $\pi \colon \mathbb{R}^n \to \mathbb{R}^n$ satisfying

$$\sigma(Y) = \pi(\sigma(X)) + \pi(z).$$

On the other hand, it is easy to verify the inclusion $\pi(z) \in \mathcal{N}_S(\pi\sigma(X))$. Therefore, we obtain $\pi(\sigma(X)) \in C_S(\sigma(Y))$.

Define now the extended signed permutation matrix $\hat{\pi} := \begin{bmatrix} \pi & 0 \\ 0 & I \end{bmatrix} \in \mathbb{R}^{t \times t}$. Observe then the equality

$$Y = U(\mathrm{Diag}\,\pi^{-1}(\sigma(Y)))V^\top = (U\pi)(\mathrm{Diag}\,\sigma(Y))(V\hat{\pi})^\top.$$

Changing the left and the right singular vector pairs in U and in V by the same sign change in the first place, if needed, we can ensure $U\pi = \overline{U}$ and $V\hat{\pi} = \overline{V}$. Hence,

$$X = U(\mathrm{Diag}\,\sigma(X))V^\top = \overline{U}(\mathrm{Diag}\,\pi(\sigma(X)))\overline{V}^\top,$$

as claimed.

Now to prove the reverse inclusion, consider a point $\omega \in \mathrm{C}_S(\sigma(Y))$. Thus there exists a vector $z \in \mathcal{N}_S(\omega)$ satisfying

$$\sigma(Y) - \omega = z.$$

Choose a signed permutation π so that $\pi(\omega)$ is nonnegative and nonincreasing. Observe then that

$$\pi(\sigma(Y)) - \pi(\omega) = \pi(z) \in \mathcal{N}_S(\pi(\omega)).$$

Defining

$$X := \overline{U}(\mathrm{Diag}\,\omega)\overline{V}^\top = (\overline{U}\pi)(\mathrm{Diag}\,\pi(\omega))(\overline{V}\hat{\pi})^\top,$$

we deduce

$$X + (\overline{U}\pi)(\mathrm{Diag}\,\pi(z))(\overline{V}\hat{\pi})^\top = (\overline{U}\pi)(\mathrm{Diag}\,\pi(\sigma(Y)))(\overline{V}\hat{\pi})^\top = Y.$$

The inclusion $X \in \mathrm{C}_S(Y)$ follows. This establishes equation (3.3).

Define now the set

$$\Delta := \{x \in \mathbb{R}^n : |x_i| = |x_j| \text{ for some indices } i \neq j\}.$$

Consider a Lebesgue null set $\Gamma \subseteq \mathbb{R}^n$. Then clearly the absolutely symmetric set $\widehat{\Gamma} := \Delta \cup \Pi_n^\pm \Gamma$ is Lebesgue null as well. Lemma 3.6 shows that $\sigma^{-1}(\widehat{\Gamma})$ is also Lebesgue null. Note for all $\pi \in \Pi_n^\pm$ and $y \in \mathbb{R}^n$, one has $C_S(\pi y) = \pi C_S(y)$ and hence $C_S^\#(\pi y) = C_S^\#(y)$. Combining these observations, we obtain

$$C^\#(\mathcal{M}) \leq \sup_{Y \in \sigma^{-1}(\widehat{\Gamma}^c)} C_\mathcal{M}^\#(Y) = \sup_{y \in \widehat{\Gamma}^c} C_S^\#(y) \leq \sup_{y \in \Gamma^c} C_S^\#(y),$$

where the middle equality follows from equation (3.3). Since Γ was an arbitrary null set, we deduce $C^\#(\mathcal{M}) \leq C^\#(S)$.

To see the reverse inequality, consider a Lebesgue null set $\Gamma \subseteq \mathbb{R}^{n \times t}$. Consider the following subset of Γ:

$$\widehat{\Gamma} := \{Y \in \Gamma : UYV^\top \in \Gamma \text{ for all } U \in \mathcal{O}^n, V \in \mathcal{O}^t\}.$$

Clearly $\widehat{\Gamma}$ is an orthogonally invariant Lebesgue null set. Now trivially we have

$$\sup_{Y \in \widehat{\Gamma}^c} C_\mathcal{M}^\#(Y) \geq \sup_{Y \in \Gamma^c} C_\mathcal{M}^\#(Y).$$

On the other hand, it is easy to verify the equation $C_\mathcal{M}(UYV^\top) = U(C_\mathcal{M}(Y))V^\top$ and thus, $C_\mathcal{M}^\#(UYV^\top) = C_\mathcal{M}^\#(Y)$ for all $Y \in \mathbb{R}^{n \times t}$ and $U \in \mathcal{O}^n, V \in \mathcal{O}^t$. Hence

$$\sup_{Y \in \widehat{\Gamma}^c} C_\mathcal{M}^\#(Y) = \sup_{Y \in \Gamma^c} C_\mathcal{M}^\#(Y).$$

Notice $\widetilde{\Gamma} := \widehat{\Gamma} \cup \sigma^{-1}(\Delta)$ is also a Lebesgue null orthogonally invariant set. Hence we deduce

$$\sup_{Y \in \Gamma^c} C_{\mathcal{M}}^{\#}(Y) \geq \sup_{Y \in \widetilde{\Gamma}^c} C_{\mathcal{M}}^{\#}(Y) = \sup_{y \in \sigma(\widetilde{\Gamma})^c} C_S^{\#}(y) \geq C^{\#}(S),$$

where the middle equality follows from equation (3.3) and last inequality follows from Lemma 3.6. Since Γ was an arbitrary Lebesgue null set, we deduce the reverse inequality $C^{\#}(\mathcal{M}) \geq C^{\#}(S)$, and hence equality, as claimed. $\qquad\square$

The assumption that singular values of the data point Y are distinct is essential in Theorem 3.7 as the following example shows.

Example 3.8 (Nondistinct singular values). Consider the orthogonally invariant set $\mathcal{M} = \{X \in \mathbb{R}^{2 \times 2} \; : \; \det(X) = 0\}$. We may write $\mathcal{M} = \sigma^{-1}(S)$ where $S := \{(x_1, x_2) \in \mathbb{R}^2 \; : \; x_1 x_2 = 0\}$ is absolutely symmetric. Then $C_S((1,1)) = \{(1,0), (0,1)\}$. On the other hand, one can verify that $uu^{\top} \in C_{\mathcal{M}}(I_2)$ for each $u \in \mathbb{R}^2$ of norm one, where I_2 is the 2×2 identity matrix.

We now apply Theorem 3.7 to our two running examples.

Example 3.9 (Matrices of rank at most r). Fix $r \leq n$ and recall the orthogonally invariant set

$$\mathbb{R}_r^{n \times t} = \{X \in \mathbb{R}^{n \times t} \; : \; \text{rank}(X) \leq r\}.$$

This set is a determinantal variety cut out by the $(r+1) \times (r+1)$ minors of a symbolic $n \times t$ matrix. One sees that $\sigma^{-1}(\mathbb{R}_r^n) = \mathbb{R}_r^{n \times t}$ where $\mathbb{R}_r^n$ is the absolutely symmetric set defined in (2.4). Theorem 3.7 and Example 2.4 together imply that

$$C^{\#}(\mathbb{R}_r^{n \times t}) = C^{\#}(\mathbb{R}_r^n) = \binom{n}{r}.$$

This number is also the ED degree of $\mathbb{R}_r^{n \times t}$ (cf. [12, Example 2.3]); we will revisit this point in Section 5. In particular, the ED critical points of a general data matrix Y, with singular value decomposition $Y = U(\text{Diag } \sigma(Y))V^{\top}$, are the $\binom{n}{r}$ matrices obtained by setting to zero all possible choices of $n - r$ singular values in this singular value decomposition. The matrix

$$U \left(\text{Diag}\left(\sigma_1(Y), \ldots, \sigma_r(Y), \underbrace{0, \ldots, 0}_{n-r}\right)\right) V^{\top}$$

is a nearest element of $\mathbb{R}_r^{n \times t}$ to Y. This is precisely the statement of the classical *Eckart-Young theorem*.

Example 3.10 (Essential variety). The essential variety $\mathcal{E}$ defined in (1.2) is orthogonally invariant with $\mathcal{E} = \sigma^{-1}(E_{3,2})$ where $E_{3,2}$ is the absolutely symmetric set defined in Example 2.5. Then by Theorem 3.7 and Example 2.5,

$$C^{\#}(\mathcal{E}) = C^{\#}(E_{3,2}) = 6.$$

Furthermore, for a matrix $Y \in \mathbb{R}^{3 \times 3}$ along with a singular value decomposition $Y = U(\text{Diag } \sigma(Y))V^{\top}$, by (3.2), the matrix

$$U \left(\text{Diag}\left(\frac{\sigma_1(Y) + \sigma_2(Y)}{2}, \frac{\sigma_1(Y) + \sigma_2(Y)}{2}, 0\right)\right) V^{\top}$$

is a nearest element of $\mathcal{E}$ to Y. This is precisely Hartley's result [16, Theorem 5] which is well known in the computer vision community. The six ED critical points

of a general $Y = U(\text{Diag } \sigma(Y))V^\top$ are the matrices $U(\text{Diag } x)V^\top$ where x varies over the following vectors:

$$\left(\frac{\sigma_1(Y)+\sigma_2(Y)}{2}, \frac{\sigma_1(Y)+\sigma_2(Y)}{2}, 0\right), \quad \left(\frac{\sigma_1(Y)-\sigma_2(Y)}{2}, \frac{-\sigma_1(Y)+\sigma_2(Y)}{2}, 0\right),$$

$$\left(\frac{\sigma_1(Y)+\sigma_3(Y)}{2}, 0, \frac{\sigma_1(Y)+\sigma_3(Y)}{2}\right), \quad \left(\frac{\sigma_1(Y)-\sigma_3(Y)}{2}, 0, \frac{-\sigma_1(Y)+\sigma_3(Y)}{2}\right),$$

$$\left(0, \frac{\sigma_2(Y)+\sigma_3(Y)}{2}, \frac{\sigma_2(Y)+\sigma_3(Y)}{2}\right), \quad \left(0, \frac{\sigma_2(Y)-\sigma_3(Y)}{2}, \frac{-\sigma_2(Y)+\sigma_3(Y)}{2}\right).$$

We end the section with the proof of the lemma evoked in the proof of Theorem 3.7.

Proof of Lemma 3.6. Before we get to the main part of the proof, we need some notation. To this end, define $\mathcal{O}_n^t$ to be the set of $t \times n$ matrices with orthonormal columns. Note that $\mathcal{O}_n^t$ is a smooth manifold of dimension $nt - \frac{n(n+1)}{2}$, and is called a *Stiefel manifold*; see for example [2, Proposition A.4]. In particular $\mathcal{O}^n$ is a manifold of dimension $\frac{n(n-1)}{2}$. We deduce that the product manifold $\mathcal{O}^n \times \mathcal{O}_n^t \times \mathbb{R}^n$ has dimension exactly nt, same as $\mathbb{R}^{n \times t}$. For any vector $x \in \mathbb{R}^n$ let sDiag x denote the $n \times n$ diagonal matrix with x on the diagonal. Define now the mapping $\Gamma \colon \mathcal{O}^n \times \mathcal{O}_n^t \times \mathbb{R}^n \to \mathbb{R}^{n \times t}$ by setting $\Gamma(U, V, x) = U(\text{sDiag } x)V^\top$. Notice that for the absolutely symmetric set S, we have equality $\Gamma(\mathcal{O}^n \times \mathcal{O}_n^t \times S) = \sigma^{-1}(S)$.

Suppose now that S is a Lebesgue null set. Then clearly the set $\mathcal{O}^n \times \mathcal{O}_n^t \times S$ is Lebesgue null in the manifold $\mathcal{O}^n \times \mathcal{O}_n^t \times \mathbb{R}^n$. Since Γ is smooth, the image $\Gamma(\mathcal{O}^n \times \mathcal{O}_n^t \times S)$, which coincides with $\sigma^{-1}(S)$, is Lebesgue null $\mathbb{R}^{n \times t}$, as claimed. Conversely, suppose that $\sigma^{-1}(S)$ is Lebesgue null. It is well-known that Γ is a local diffeomorphism on an open full-measure subset of the domain manifold $\mathcal{O}^n \times \mathcal{O}_n^t \times \mathbb{R}^n$. To see this quickly, define the set

$$\Delta = \{x \in \mathbb{R}^n : |x_i| = |x_j| \text{ for some indices } i \neq j\}.$$

Observe that since Γ is a semi-algebraic map, we may stratify $\mathcal{O}^n \times \mathcal{O}_n^t \times \Delta^c$ into finitely many smooth manifolds $\mathcal{M}_i$ so that the restriction of Γ to each $\mathcal{M}_i$ has constant rank. Suppose now for the sake of contradiction that the derivative of Γ has deficient rank on some maximal dimensional manifold $\mathcal{M}_i$. Then by the constant rank theorem [20, Theorem 5.13], there exists a nontrivial smooth path $(U(t), V(t), x(t))$ in $\mathcal{M}_i$ so that Γ is constant on the path, meaning that the matrices $U(t)(\text{sDiag } x(t))V(t)^\top$ are equal for all t. On the other hand, taking into account that the coordinates of $x(t)$ are distinct and appealing to uniqueness of singular values and singular vectors, we obtain a contradiction. Hence Γ is a local diffeomorphism on an open full measure subset of $\mathcal{O}^n \times \mathcal{O}_n^t \times \mathbb{R}^n$. It follows immediately that S is Lebesgue null, since otherwise the image $\Gamma(\mathcal{O}^n \times \mathcal{O}_n^t \times S)$ would fail to be Lebesgue null. This completes the proof. $\square$

4. APPLICATIONS

We now apply the techniques developed in the last section to calculate and count the ED critical points of several orthogonally invariant matrix sets.

Example 4.1 (Matrices orthogonally equivalent to a given matrix)**.** Fix a matrix $A \in \mathbb{R}^{n \times t}$. We say $X \in \mathbb{R}^{n \times t}$ is *orthogonally equivalent* to A if $X = UAV^\top$ for some $U \in \mathcal{O}^n$ and $V \in \mathcal{O}^t$. Let $\mathcal{M}_A \subseteq \mathbb{R}^{n \times t}$ be the set of matrices that are orthogonally equivalent to A. Then $\mathcal{M}_A$ is orthogonally invariant and $\mathcal{M}_A = \sigma^{-1}(\sigma(A))$. By

Theorem 3.7,
$$C^{\#}(\mathcal{M}_A) = |\Pi_n^{\pm}\{\sigma(A)\}| = 2^n \frac{n!}{n_1! \ldots n_k!}$$
where $\sigma(A) = (\underbrace{\sigma_1, \ldots, \sigma_1}_{n_1}, \ldots, \underbrace{\sigma_k, \ldots, \sigma_k}_{n_k})$.

Example 4.2 (Real orthogonal group). The real orthogonal group $\mathcal{O}^n$ is a special case of the previous example because $\mathcal{O}^n$ is the set of $n \times n$ real matrices that are orthogonally equivalent to the $n \times n$ identity matrix, namely,
$$\mathcal{O}^n = \sigma^{-1}((\underbrace{1, \ldots, 1}_{n})).$$

Hence $C^{\#}(\mathcal{O}^n) = 2^n$ which is also the ED degree of $\mathcal{O}^n$ (cf. [11, Theorem 3.2]).

By Proposition 3.5, for any $Y \in \mathbb{R}^{n \times n}$ with singular value decomposition $Y = U(\operatorname{Diag} \sigma(Y))V^{\top}$ where $U, V \in \mathcal{O}^n$, one has $UV^{\top} \in \operatorname{proj}_{\mathcal{O}^n}(Y)$. Moreover, if the singular values of Y are distinct, then by Theorem 3.7, the set of critical points of Y on $\mathcal{O}^n$ is $\{U(\operatorname{Diag} x)V^{\top} : x_i = \pm 1 \ \forall \ i\}$. These results were also obtained by Draisma and Baaijens [11] using algebraic methods.

Example 4.3 (Unit sphere of Schatten d-norm). The Schatten d-norm of a matrix $X \in \mathbb{R}^{n \times t}$ is defined as
$$\|X\|_d := \left[\sum_{i=1}^{n}(\sigma_i(X))^d\right]^{1/d}.$$

Note the following fact: $\|UXV^{\top}\|_d = \|X\|_d$ for any $X \in \mathbb{R}^{n \times t}$, $U \in \mathcal{O}^n$, $V \in \mathcal{O}^t$. Hence the unit sphere of the Schatten d-norm
$$\mathcal{F}_{n,t,d} := \{X \in \mathbb{R}^{n \times t} : \|X\|_d = 1\}$$
is orthogonally invariant. Assume $d \geq 2$ is an even integer. One has $\mathcal{F}_{n,t,d} = \sigma^{-1}(F_{n,d})$ where $F_{n,d}$ is the *affine Fermat hypersurface*
$$F_{n,d} := \left\{x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i^d = 1\right\}$$
which is absolutely symmetric. Therefore, $C^{\#}(\mathcal{F}_{n,t,d}) = C^{\#}(F_{n,d})$ for all t.

Consider the special case $n = 2$. For simplicity, we set $F_d := F_{2,d}$. Then F_2 is the unit circle and $C^{\#}(F_2) = 2$. Suppose $d \geq 4$. For any $y \in \mathbb{R}^2$, a point x is an ED critical point of y on F_d if and only if x lies on both F_d and the curve

(4.1) $$\gamma_d := \{x \in \mathbb{R}^2 : x_1^{d-1}(x_2 - y_2) = x_2^{d-1}(x_1 - y_1)\}.$$

The graph of γ_4 is shown in Figure 3. By symmetry, we may assume that $y_1 > y_2 > 0$. When y_1 is small, the "optimal" points, m_1, m_2, m_3, m_4, on the pieces of γ_d in the coordinate directions lie inside the curve F_d, namely,
$$m_i \in \{x \in \mathbb{R}^2 : x_1^d + x_2^d < 1\}$$
for $i = 1, \ldots, 4$. Hence, $C_{F_d}^{\#}(y) \leq 8$ for any point y except on a null set, and there is an open set V in $\mathbb{R}^2$ such that $C_{F_d}^{\#}(y) = 8$ for any $y \in V$. To sum up, by Theorem 3.7, we have
$$C^{\#}(\mathcal{F}_{2,t,d}) = C^{\#}(F_d) = \begin{cases} 2 & \text{if } d = 2 \\ 8 & \text{if } d \geq 4. \end{cases}$$
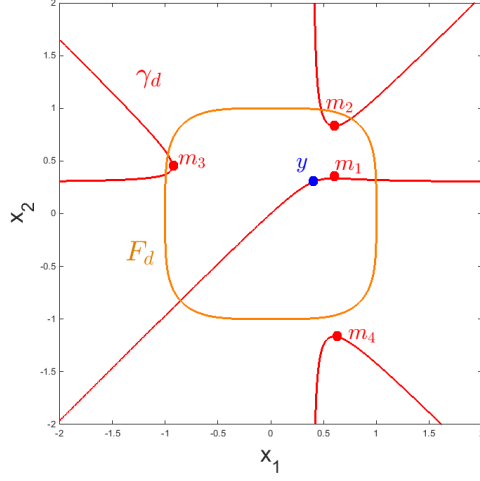
FIGURE 3. The curves F_d and γ_d in Example 4.3 for $d = 4$

Example 4.4 ($SL_n^\pm$). This example was considered in [11, Section 4]. Define the orthogonally invariant set

$$SL_n^\pm := \{X \in \mathbb{R}^{n \times n} \ : \ \det(X) = \pm 1\} = \sigma^{-1}(H_n)$$

where H_n is the absolutely symmetric set defined as

$$H_n := \{x \in \mathbb{R}^n \ : \ x_1 \cdots x_n = \pm 1\}.$$

Consider the special case $n = 2$. We evaluate $C^\#(SL_2^\pm)$ by computing $C^\#(H_2)$. The set H_2 is a disjoint union of two hyperbolas:

$$H_2^\pm := \{x \in \mathbb{R}^2 \ : \ x_1 x_2 = \pm 1\}.$$

Since these hyperbolas are disjoint, for any y, one has

$$C_{H_2}^\#(y) = C_{H_2^+}^\#(y) + C_{H_2^-}^\#(y).$$

Notice that each of the sets $C_{H_2^\pm}(y)$ is defined by the roots of a univariate quartic. Indeed,

$$C_{H_2^+}(y) = \left\{ \left(x, \frac{1}{x}\right) \in \mathbb{R}^2 \ : \ q_y^+(x) := x^4 - x^3 y_1 + x y_2 - 1 = 0 \right\}$$
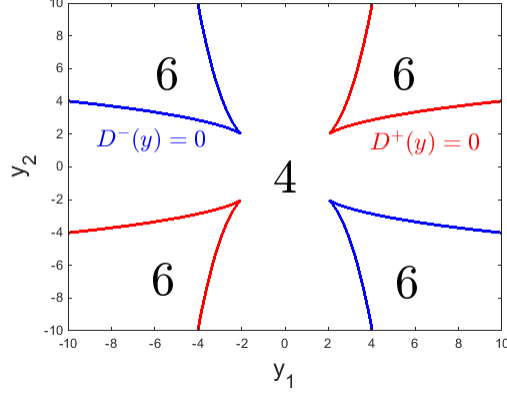
and

$$C_{H_2^-}(y) = \left\{ \left(x, \frac{-1}{x}\right) \in \mathbb{R}^2 \ : \ q_y^-(x) := x^4 - x^3 y_1 - x y_2 - 1 = 0 \right\}.$$

By computing the trace forms of these quartics and applying Sylvester's criterion (see e.g. [28, Corollary 2.9]) we acquire these results:

$$C_{H_2}^\#(y) = \begin{cases} 6 & \text{if } D^+(y) > 0 \text{ or } D^-(y) > 0 \\ 4 & \text{if } D^+(y) < 0 \text{ and } D^-(y) < 0 \end{cases}$$

where the bivariate polynomials

$$D^+(y) := -256 + 192 y_1 y_2 + 6 y_1^2 y_2^2 + 4 y_1^3 y_2^3 - 27 y_1^4 - 27 y_2^4$$

FIGURE 4. Values of $C^{\#}_{H_2}(y)$ for general y, in Example 4.4

and
$$D^-(y) := -256 - 192y_1y_2 + 6y_1^2y_2^2 - 4y_1^3y_2^3 - 27y_1^4 - 27y_2^4$$
are the discriminants of q_y^+ and q_y^- respectively; see Figure 4. Then by Theorem 3.7, we know
$$C^{\#}(SL_2^{\pm}) = 6.$$

5. CONNECTIONS TO THE EUCLIDEAN DISTANCE DEGREE

As mentioned in the introduction, this work was inspired by the general framework in [12] for counting the critical points of the squared Euclidean distance function on an algebraic variety. In this section we comment on the relationships between our results and those in [12], beginning with some basic facts about real varieties [1, 31].

Consider a real variety $\mathcal{V} \subseteq \mathbb{R}^n$ whose vanishing ideal is generated by the polynomials $f_1, \ldots, f_s \in \mathbb{R}[x_1, \ldots, x_n]$. Then the *Zariski closure* of $\mathcal{V}$ is $\mathcal{V}_{\mathbb{C}} := \{x \in \mathbb{C}^n : f_1(x) = \ldots = f_s(x) = 0\}$, the smallest complex variety containing $\mathcal{V}$. The real part of $\mathcal{V}_{\mathbb{C}}$, i.e., $\mathcal{V}_{\mathbb{C}} \cap \mathbb{R}^n$, is precisely $\mathcal{V}$ and hence, the vanishing ideal of $\mathcal{V}_{\mathbb{C}}$ in $\mathbb{C}[x_1, \ldots, x_n]$ is also generated by $f_1, \ldots, f_s$. Also, both $\mathcal{V}$ and $\mathcal{V}_{\mathbb{C}}$ have the same dimension over $\mathbb{R}$ and $\mathbb{C}$ respectively. Recall that if $\mathcal{V}_{\mathbb{C}} = \bigcup_{j=1}^t \mathcal{W}^j$ is a minimal irreducible decomposition of $\mathcal{V}_{\mathbb{C}}$, then for any j, the real part of $\mathcal{W}^j$, denoted by $\mathcal{U}^j$, is an irreducible real variety whose Zariski closure is $\mathcal{W}^j$. Moreover, $\mathcal{V} = \bigcup_{j=1}^t \mathcal{U}^j$ is a minimal irreducible decomposition of $\mathcal{V}$. For simplicity, in the rest of this section we will assume that $\mathcal{V}_{\mathbb{C}}$ is irreducible.

Denote by $J(f)$ the $s \times n$ *Jacobian* matrix whose (i, j)-entry is $\frac{\partial f_i}{\partial x_j}$. Then a point $x \in \mathcal{V}_{\mathbb{C}}$ is said to be *regular* if the rank of the Jacobian matrix evaluated at x, $\mathrm{rank}(J(f)(x))$, equals the codimension of $\mathcal{V}_{\mathbb{C}}$. Let $\mathcal{V}_{\mathbb{C}}^{\mathrm{reg}}$ denote the regular points of $\mathcal{V}_{\mathbb{C}}$. It is known that $\mathcal{V}_{\mathbb{C}}^{\mathrm{reg}}$ is a Zariski open subset of $\mathcal{V}_{\mathbb{C}}$, and its complement is a proper subvariety in $\mathcal{V}_{\mathbb{C}}$, denoted as $\mathrm{Sing}(\mathcal{V}_{\mathbb{C}})$, and called the *singular locus* of $\mathcal{V}_{\mathbb{C}}$. The set $\mathcal{V}^{\mathrm{reg}}$ of regular points in $\mathcal{V}$ is precisely the set of real points in $\mathcal{V}_{\mathbb{C}}^{\mathrm{reg}}$. However, while $\mathcal{V}_{\mathbb{C}}^{\mathrm{reg}}$ is dense in $\mathcal{V}_{\mathbb{C}}$ in the Euclidean topology, $\mathcal{V}^{\mathrm{reg}}$ may not be dense in $\mathcal{V}$ or even $\mathcal{V}^*$, the set of smooth (i.e., C^p-smooth) points in $\mathcal{V}$. The following example illustrates this behavior.
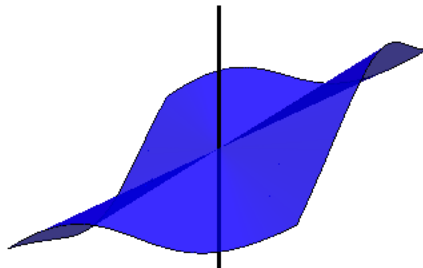
FIGURE 5. The Cartan umbrella

Example 5.1 (Cartan umbrella). The *Cartan umbrella* is the real variety

$$\mathcal{V} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,:\, x_3(x_1^2 + x_2^2) - x_1^3 = 0\}.$$

Here $\mathcal{V}_{\mathbb{C}}$ is irreducible, as is $\mathcal{V}$. On the other hand, $\mathcal{V}$ is also the union of a surface (which is not a real variety) and the x_3-axis (Figure 5). No point on the x_3-axis is regular while all points on the x_3-axis, except the origin, are smooth.

This example shows an important distinction between smooth and regular points on a real variety. Our goal in this section is to elaborate on this difference in the context of critical points as defined in [12]. Recall that the normal space at a regular point $x \in \mathcal{V}_{\mathbb{C}}$, denoted as $\mathcal{N}_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{alg}}(x)$, is the row space of $J(f)(x)$. Given a data point $y \in \mathbb{C}^n$, and a real variety $\mathcal{V}$, the authors of [12] study the set of critical points

(5.1) $$C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y) := \left\{ x \in \mathcal{V}_{\mathbb{C}}^{\mathrm{reg}} \,:\, y - x \in \mathcal{N}_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{alg}}(x) \right\}.$$

Note that in the above algebraic setting, even though our input is a real variety $\mathcal{V} \subseteq \mathbb{R}^n$, there is a natural passage into the complex numbers; the real variety $\mathcal{V}$ is replaced by the complex variety $\mathcal{V}_{\mathbb{C}}$, the data point can be any $y \in \mathbb{C}^n$, and the critical points of y are regular points in $\mathcal{V}_{\mathbb{C}}$. This complexification yields the first key result in [12], namely that for a fixed $\mathcal{V}$, the number of critical points of a general data point $y \in \mathbb{C}^n$ is a constant. In other words, the cardinality of $C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y)$, which we will denote as $C_{\mathcal{V}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y)$, is a constant. This constant is called the *Euclidean distance degree (EDdegree)* of $\mathcal{V}$ in [12]. The phrase "general data point" refers to the fact that $C_{\mathcal{V}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y)$ is different from EDdegree($\mathcal{V}$) only if y lies on certain proper subvarieties in $\mathbb{C}^n$ which we will describe in the Appendix. We remark that if $\mathcal{V}_{\mathbb{C}}$ is reducible, then EDdegree($\mathcal{V}$) is the sum of the ED degrees of its irreducible components.

Our results in this paper provide a method to compute the (smooth) ED critical points of a real orthogonally invariant set of matrices. We did not require this set to be a real variety, and the notion of smoothness (of critical points) was geometric and not algebraic. Therefore, in order to connect our work to that in [12], we need to restrict to real varieties $\mathcal{V} \subseteq \mathbb{R}^n$, data points $y \in \mathbb{R}^n$, and understand the relationships between the following two sets of critical points of y:

- $C_{\mathcal{V}}(y)$, the set of ED critical points of y on $\mathcal{V}$ as in Section 2, and
- $C_{\mathcal{V}}^{\mathrm{reg}}(y)$, the set of regular critical points of y on $\mathcal{V}$.

As usual we denote the cardinalities of these sets by $C_{\mathcal{V}}^{\#}(y)$ and $C_{\mathcal{V}}^{\#\,\mathrm{reg}}(y)$ respectively. Also, note that $C_{\mathcal{V}}^{\mathrm{reg}}(y) = C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y) \cap \mathbb{R}^n = C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y) \cap \mathcal{V}$. Since a regular point of $\mathcal{V}$ is C^p-smooth for $p \in \{\infty, \omega\}$, for any y we have,

$$(5.2) \qquad C_{\mathcal{V}}^{\mathrm{reg}}(y) \subseteq C_{\mathcal{V}}(y) \quad \text{and so,} \quad C_{\mathcal{V}}^{\#\,\mathrm{reg}}(y) \leq C_{\mathcal{V}}^{\#}(y).$$

However, for a given y, this inequality can be strict since there can be smooth points on $\mathcal{V}$ that are not regular as we saw in the Cartan umbrella. In fact, in the Cartan umbrella there is a dense set of y for which the inequality in (5.2) is strict since for any y with $y_3 \neq 0$, there will be an ED critical point on the x_3-axis.

Moving on to invariants of the entire set $\mathcal{V}$, on the one hand, we have the quantity $C^{\#}(\mathcal{V}) = \inf_{\Gamma \in \Theta} \sup_{y \in \Gamma^c} C_{\mathcal{V}}^{\#}(y)$ from Section 2, and on the other hand, we have the constant $\mathrm{EDdegree}(\mathcal{V})$ which is an upper bound to $C_{\mathcal{V}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y)$ whenever $C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y)$ is finite. Since the inequality in (5.2) can be strict, it is not clear that the number $C^{\#}(\mathcal{V})$ is always a lower bound to $\mathrm{EDdegree}(\mathcal{V})$. However, it is true under certain conditions.

Theorem 5.2. *If $\mathcal{V}$ is a real variety such that $\mathcal{V}^{\mathrm{reg}}$ is Euclidean dense in $\mathcal{V}^*$, then we have the inequality $C^{\#}(\mathcal{V}) \leq \mathrm{EDdegree}(\mathcal{V})$.*

This is an immediate consequence of the following lemma. Note that the Cartan umbrella does not satisfy the assumptions of Theorem 5.2.

Lemma 5.3. *Consider a real variety $\mathcal{V} \subseteq \mathbb{R}^n$ (possibly reducible) such that $\mathcal{V}^{\mathrm{reg}}$ is Euclidean dense in $\mathcal{V}^*$. Then there is an open set of points $y \in \mathbb{R}^n$ such that*

$$(5.3) \qquad \text{for any irreducible component } \mathcal{W} \text{ of } \mathcal{V}, \; C_{\mathcal{W}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y) = \mathrm{EDdegree}(\mathcal{W}),$$

$$(5.4) \qquad C_{\mathcal{V}}^{\#}(y) = C^{\#}(\mathcal{V}), \quad \text{and}$$

$$(5.5) \qquad C_{\mathcal{V}}(y) \subseteq C_{\mathcal{V}}^{\mathrm{reg}}(y).$$

Proof. Fix an irreducible component $\mathcal{W}$ of $\mathcal{V}$. Then from [12, Theorem 4.1] the set of all $y \in \mathbb{C}^n$ for which $C_{\mathcal{W}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y) \neq \mathrm{EDdegree}(\mathcal{W})$ lies in a proper subvariety D in $\mathbb{C}^n$. Consider the real variety $D_{\mathbb{R}} := D \cap \mathbb{R}^n$ which contains all the real points y with $C_{\mathcal{W}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y) \neq \mathrm{EDdegree}(\mathcal{W})$. The Zariski closure of $D_{\mathbb{R}}$ is a subvariety of D with complex dimension equal to the real dimension of $D_{\mathbb{R}}$. Therefore, the real dimension of $D_{\mathbb{R}}$ is smaller than n, and hence $D_{\mathbb{R}}$ is a proper subvariety in $\mathbb{R}^n$. In particular, the complement of $D_{\mathbb{R}}$ in $\mathbb{R}^n$ is an open full measure set. Taking the intersection of all such open full measure sets as we vary over the irreducible components of $\mathcal{V}$, we still have an open full measure set, and any point y in this set satisfies (5.3).

Since $\mathcal{V}$ is a variety, one has $C^{\#}(\mathcal{V}) < \infty$ by Proposition 2.2. Then by quantifier elimination, the set $\{y \in \mathbb{R}^n \; : \; C_{\mathcal{V}}^{\#}(y) = C^{\#}(\mathcal{V})\}$ is semialgebraic. Moreover, this set cannot be Lebesgue null by the definition of $C^{\#}(\mathcal{V})$. Therefore, it must contain an open subset. Hence (5.4) holds for all y in some open set.

We now show that (5.5) holds for any point y in an open full measure set, which will prove the lemma. For any $i = 0, 1, \ldots, n$ consider the submanifold

$$M_i := \{x \in \mathcal{V}^* \; : \; \dim(\mathcal{T}_{\mathcal{V}}(x)) = i\}.$$

These submanifolds M_i clearly partition $\mathcal{V}^*$. We first claim $M_i \cap \mathrm{cl}(M_j) = \emptyset$ for all distinct pairs $i, j \in [n]$. Note the symbol "cl" here denotes the Euclidean closure. Indeed, if there existed some point $x \in M_i \cap \mathrm{cl}(M_j)$, then we would deduce for all points $y \in M_j$ sufficiently near x, the equality $j = \dim \mathcal{T}_{\mathcal{V}}(y) = \dim \mathcal{T}_{\mathcal{V}}(x) = i$, a

contradiction. The middle equality follows from the fact that $\dim \mathcal{T}_{\mathcal{V}}(\cdot)$ is constant in a neighborhood of x in $\mathcal{V}$.

Now since $\mathcal{V}^{\mathrm{reg}}$ is dense in $\mathcal{V}^*$, the inclusion $\mathcal{V}^* \subseteq \mathrm{cl}(\mathcal{V}^{\mathrm{reg}}) = \bigcup_i \mathrm{cl}(\mathcal{V}^{\mathrm{reg}} \cap M_i)$ holds. We claim that for any i, the set $\mathcal{V}^{\mathrm{reg}} \cap M_i$ is Euclidean dense in M_i, namely, $M_i \subseteq \mathrm{cl}(\mathcal{V}^{\mathrm{reg}} \cap M_i)$. To see this, it suffices to establish $M_i \cap \mathrm{cl}(\mathcal{V}^{\mathrm{reg}} \cap M_j) = \emptyset$ for any $j \neq i$. This follows immediately by observing $M_i \cap \mathrm{cl}(\mathcal{V}^{\mathrm{reg}} \cap M_j) \subseteq M_i \cap \mathrm{cl}(M_j) = \emptyset$. It follows that for any i, the strict inequality

$$\dim(M_i \setminus (\mathcal{V}^{\mathrm{reg}} \cap M_i)) < \dim M_i$$

holds (see [5, Proposition 3.16]). By dimension arguments, the dimension of the set

$$\bigcup_{x \in M_i \setminus (M_i \cap \mathcal{V}^{\mathrm{reg}})} (x + \mathcal{N}_{\mathcal{V}}(x))$$

is strictly less than n. In particular, its interior is empty. Taking the union over i,

$$\bigcup_{x \in \mathcal{V}^* \setminus \mathcal{V}^{\mathrm{reg}}} (x + \mathcal{N}_{\mathcal{V}}(x))$$

also has empty interior. Any y which is not in this set satisfies (5.5). □

We conclude the paper by comparing $C^{\#}(\mathcal{V})$ and $\mathrm{EDdegree}(\mathcal{V})$ in some of the examples we saw in Section 3 and Section 4.

Example 5.4 ($SL_n^{\pm}$)**.** Every point in the Zariski closure of $SL_n^{\pm}$ (c.f. Example 4.4) is regular. Therefore by Theorem 5.2, we obtain

$$C^{\#}(SL_n^{\pm}) \leq \mathrm{EDdegree}(SL_n^{\pm}) = n2^n.$$

The formula for EDdegree was derived in [11]. We saw in Example 4.4 that strict inequality holds when $n = 2$.

Example 5.5 (Unit sphere of Schatten d-norm)**.** Consider the set $\mathcal{F}_{2,2,d} \subseteq \mathbb{R}^{2 \times 2}$ given in Example 4.3. For any d the Zariski closure of $\mathcal{F}_{2,2,d}$ is regular everywhere. By Theorem 5.2,

$$\mathrm{EDdegree}(\mathcal{F}_{2,2,d}) \geq C^{\#}(\mathcal{F}_{2,2,d}) = 8,$$

for $d \geq 4$. From `Macaulay2` [15] computations, $\mathrm{EDdegree}(\mathcal{F}_{2,2,d})$ equals $16, 34, 64, 98$ when $d = 4, 6, 8, 10$ respectively. This suggests the gap between $C^{\#}(\mathcal{F}_{2,2,d})$ and $\mathrm{EDdegree}(\mathcal{F}_{2,2,d})$ increases with d.

Example 5.6. Let $\mathcal{E}$ be the essential variety from (1.2). It is an irreducible variety of codimension three in $\mathbb{R}^{3 \times 3}$ [10, Proposition 3.6]. By the transfer of smoothness (see the paragraph after the proof of Proposition 3.5), we know $\mathcal{E}^* = \mathcal{E} \setminus \{0\}$. Moreover, a straightforward computation verifies that any non-zero essential matrix is a regular point of $\mathcal{E}$. Thus $\mathcal{E}^{\mathrm{reg}} = \mathcal{E}^* = \mathcal{E} \setminus \{0\}$, and by Theorem 5.2, we have

$$(5.6) \qquad\qquad \mathrm{EDdegree}(\mathcal{E}) \geq C^{\#}(\mathcal{E}) = 6.$$

With completely different tools, but entirely motivated by this work, one can prove the equality $\mathrm{EDdegree}(\mathcal{E}) = 6$. The details will be described in a forthcoming paper.

APPENDIX A

The Cartan umbrella illustrates a yet unexplored feature of the EDdegree concerning the exceptional loci of points y at which $C_{\mathcal{V}_{\mathbb{C}}}^{\#\,\mathrm{reg}}(y)$ can be different from EDdegree($\mathcal{V}$). We comment briefly on this topic which deserves further investigation.

For each $y \in \mathbb{C}^n$, the set $C_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{reg}}(y)$ is the variety of a polynomial ideal called the *critical ideal* of y (see (2.1) in [12]), and typically, this variety has EDdegree($\mathcal{V}$)-many distinct complex solutions. There are two ways in which the critical ideal of y can have fewer distinct roots; the first is because of roots with multiplicity, and the second is because a root may wander off into $\mathrm{Sing}(\mathcal{V}_{\mathbb{C}})$ due to closure issues. These situations create two exceptional loci in the space of data points y. The locus of $y \in \mathbb{C}^n$ for which the critical ideal has roots with multiplicity is called the *ED discriminant* of $\mathcal{V}_{\mathbb{C}}$, and is denoted as $\Sigma_{\mathcal{V}_{\mathbb{C}}}$. The ED discriminant is typically a hypersurface in $\mathbb{C}^n$ and can be computed from the equations of $\mathcal{V}$ (see Section 7 in [12] for algorithms and examples). We saw the ED discriminant of the parabola $\{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1^2\}$ in Figure 1. It is the curve defined by

$$16y_2^3 - 27y_1^2 - 24y_2^2 + 12y_2 - 2 = 0.$$

The second type of exceptional locus has not been studied in [12] and was suggested to us by Bernd Sturmfels. To describe it, let us denote the copy of $\mathbb{C}^n$ that contains $\mathcal{V}_{\mathbb{C}}$ as $\mathbb{C}_x^n$ and the copy that contains the data points y as $\mathbb{C}_y^n$. The *ED correspondence* of $\mathcal{V}_{\mathbb{C}}$, denoted as $\mathbb{E}_{\mathcal{V}_{\mathbb{C}}}$ and described in [12, Section 4], is the Zariski closure of:

$$(A.1) \qquad \left\{ (x, y) \in \mathbb{C}_x^n \times \mathbb{C}_y^n \ : \ x \in \mathcal{V}_{\mathbb{C}}^{\mathrm{reg}}, \ y \in \mathbb{C}^n, \ x - y \in \mathcal{N}_{\mathcal{V}_{\mathbb{C}}}^{\mathrm{alg}}(x) \right\}.$$

By [12, Theorem 4.1], the ED correspondence is an irreducible variety of dimension n in $\mathbb{C}_x^n \times \mathbb{C}_y^n$. It admits two natural projections π_x and π_y into $\mathbb{C}_x^n$ and $\mathbb{C}_y^n$ respectively. Over general $y \in \mathbb{C}^n$, the projection $\pi_y : \mathbb{E}_{\mathcal{V}_{\mathbb{C}}} \to \mathbb{C}_y^n$ has finite fibers of cardinality equal to EDdegree($\mathcal{V}$). The exceptional locus we are looking for is the set of data points $y \in \mathbb{C}^n$ that have critical points that fall into the singular locus $\mathrm{Sing}(\mathcal{V}_{\mathbb{C}})$. This is precisely the Zariski closure of the set

$$(A.2) \qquad\qquad\qquad \pi_y\left(\mathbb{E}_{\mathcal{V}_{\mathbb{C}}} \cap (\mathrm{Sing}(\mathcal{V}_{\mathbb{C}}) \times \mathbb{C}_y^n)\right)$$

which we call the *ED data singular locus* of $\mathcal{V}_{\mathbb{C}}$, and denote by $D_{\mathcal{V}_{\mathbb{C}}}$. This affine scheme deserves further study, and we illustrate both $D_{\mathcal{V}_{\mathbb{C}}}$ and $\Sigma_{\mathcal{V}_{\mathbb{C}}}$ for the Cartan umbrella in Example A.1. Note that $D_{\mathcal{V}_{\mathbb{C}}}$ is empty for the parabola in Figure 1 since its singular locus is empty. The upshot of the above discussion is that for all $y \in \mathbb{R}^n \setminus D_{\mathcal{V}_{\mathbb{C}}}$ but in a connected component of $\mathbb{R}_y^n \setminus \Sigma_{\mathcal{V}_{\mathbb{C}}}$, the number of regular critical points on $\mathcal{V}$, $C_{\mathcal{V}}^{\#\,\mathrm{reg}}(y)$, is a constant.

Example A.1 (Cartan umbrella continued)**.** The ED discriminant of the Cartan umbrella in Example 5.1 is a degree 12 surface in the space of data points given by the following equation:
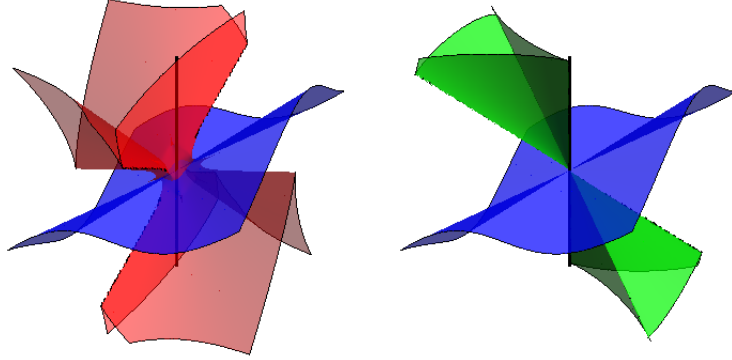
FIGURE 6. The Cartan umbrella with its ED discriminant on the left and ED data singular locus on the right.

$$256y_1^{12} - 35328y_1^{10}y_2^2 - 108984y_1^8y_2^4 - 111867y_1^6y_2^6 - 93975y_1^4y_2^8 - 9216y_1^2y_2^{10}$$
$$-2048y_2^{12} - 2304y_1^{11}y_3 - 2112y_1^9y_2^2y_3 - 149280y_1^7y_2^4y_3 - 116868y_1^5y_2^6y_3$$
$$+53532y_1^3y_2^8y_3 + 34560y_1y_2^{10}y_3 + 6912y_1^{10}y_3^2 + 14016y_1^8y_2^2y_3^2 - 28764y_1^6y_2^4y_3^2$$
$$+41502y_1^4y_2^6y_3^2 - 86430y_1^2y_2^8y_3^2 - 768y_2^{10}y_3^2 - 7936y_1^9y_3^3 + 150720y_1^7y_2^2y_3^3$$
$$-200148y_1^5y_2^4y_3^3 - 411728y_1^3y_2^6y_3^3 + 1476y_1y_2^8y_3^3 + 9216y_1^8y_3^4 - 46656y_1^6y_2^2y_3^4$$
$$+31908y_1^4y_2^4y_3^4 + 110817y_1^2y_2^6y_3^4 + 4953y_2^8y_3^4 - 27648y_1^7y_3^5 + 23808y_1^5y_2^2y_3^5$$
$$+91236y_1^3y_2^4y_3^5 - 40284y_1y_2^6y_3^5 + 28672y_1^6y_3^6 - 196992y_1^4y_2^2y_3^6 - 240480y_1^2y_2^4y_3^6$$
$$-2592y_2^6y_3^6 - 9216y_1^5y_3^7 + 14208y_1^3y_2^2y_3^7 + 28800y_1y_2^4y_3^7 + 27648y_1^4y_3^8$$
$$+39168y_1^2y_2^2y_3^8 + 2304y_2^4y_3^8 - 27648y_1^3y_3^9 - 27648y_1y_2^2y_3^9 = 0.$$

The EDdegree of the Cartan umbrella is seven. The ED discriminant partitions $\mathbb{R}^3$ into two regions where the real regular critical points are either one or three, while the number of ED critical points is two or four. In particular, for all $y \in \mathbb{R}^3$ for which the polynomial defining the ED discriminant is positive, we get three real regular critical points.

The ED data singular locus, $D_{\mathcal{V}_{\mathbb{C}}}$, is the reducible surface in $\mathbb{C}^3$ defined by

$$(y_1^2 + y_2^2) \cdot (4y_1^4 + 8y_1^2y_2^2 + 4y_2^4 + 4y_1^3y_3 + 36y_1y_2^2y_3 + 27y_2^2y_3^2) = 0.$$

(see Figure 6) The data point $y = (-2, -1, 2)$ lies on $D_{\mathcal{V}_{\mathbb{C}}}$ but not on $\Sigma_{\mathcal{V}_{\mathbb{C}}}$. The critical ideal of this point has seven distinct complex roots but one of them is singular. Among the remaining six roots, two are real and, $(-2, -1, 2)$ lies in the region where the polynomial defining $\Sigma_{\mathcal{V}_{\mathbb{C}}}$ is positive.

REFERENCES

[1] J. Bochnak, M. Coste, and M.-F. Roy. *Real Algebraic Geometry*, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1998. Translated from the 1987 French original, Revised by the authors.

[2] P. Bürgisser and F. Cucker. *Condition: The Geometry of Numerical Algorithms*. Springer Science & Business Media, 2013.

[3] J.V. Burke and M.L. Overton. Variational analysis of non-Lipschitz spectral functions. *Math. Program.*, 90(2, Ser. A):317–351, 2001.

[4] M. Coste. *An Introduction to O-minimal Geometry*. Pisa: Istituti editoriali e poligrafici internazionali, 2000.

[5] M. Coste. An introduction to semialgebraic geometry. *RAAG network school*, 145, 2002.

[6] A. Daniilidis, D. Drusvyatskiy, and A.S. Lewis. Orthogonal invariance and identifiability. *SIAM J. Matrix Anal. Appl.*, 35(2):580–598, 2014.

[7] A. Daniilidis, A.S. Lewis, J. Malick, and H. Sendov. Prox-regularity of spectral functions and spectral sets. *J. Convex Anal.*, 15(3):547–560, 2008.

[8] A. Daniilidis, J. Malick, and H.S. Sendov. Locally symmetric submanifolds lift to spectral manifolds. *Preprint U.A.B.* **23**/*2009, 43 p., arXiv:1212.3936 [math.OC]*.

[9] C. Davis. All convex invariant functions of Hermitian matrices. *Arch. Math.*, 8:276–278, 1957.

[10] M. Demazure. Sur deux problemes de reconstruction. 1988.

[11] J. Draisma and J. A. Baaijens. Euclidean distance degrees of real algebraic groups. *Linear Algebra and its Applications*, 467:174–187, 2015.

[12] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels, and R.R. Thomas. The Euclidean distance degree of an algebraic variety. *Foundations of Computational Mathematics, to appear*. arXiv:1309.0049.

[13] D. Drusvyatskiy and M. Larsson. Approximating functions on stratified sets. *Trans. Amer. Math. Soc.*, 367(1):725–749, 2015.

[14] O.D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 4(3):225–246, 1990.

[15] D.R. Grayson and M.E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[16] R. Hartley. An investigation of the essential matrix. *GE internal report*, 1995.

[17] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge university press, 2003.

[18] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[19] H. Lee. The Euclidean distance degree of Fermat hypersurfaces. *arXiv:1409.0684*, 2014.

[20] J.M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *GTM*. Springer, New York, second edition, 2013.

[21] A.S. Lewis. Derivatives of spectral functions. *Math. Oper. Res.*, 21(3):576–588, 1996.

[22] A.S. Lewis. Nonsmooth analysis of eigenvalues. *Math. Program.*, 84(1, Ser. A):1–24, 1999.

[23] A.S. Lewis. The mathematics of eigenvalue optimization. *Math. Program.*, 97(1-2, Ser. B):155–176, 2003. ISMP, 2003 (Copenhagen).

[24] A.S. Lewis and H.S. Sendov. Nonsmooth analysis of singular values. I. Theory. *Set-Valued Anal.*, 13(3):213–241, 2005.

[25] I.G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[26] G. Ottaviani, P. Spaenlehauer, and B. Sturmfels. Exact solutions in structured low-rank approximation. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1521–1542, 2014.

[27] M. Šilhavý. Differentiability properties of isotropic functions. *Duke Math. J.*, 104(3):367–373, 2000.

[28] B. Sturmfels. *Solving Systems of Polynomial Equations*. American Mathematical Soc., 2002.

[29] J. Sylvester. On the differentiability of O(n) invariant functions of symmetric matrices. *Duke Math. J.*, 52(2):475–483, 1985.

[30] J. von Neumann. Some matrix inequalities and metrization of matrix-space. *Tomck. Univ. Rev.*, 1:286–300, 1937.

[31] H. Whitney. Elementary structure of real algebraic varieties. *Annals of Mathematics*, 66(3):545–556, 1957.

Department of Mathematics, University of Washington, Box 354350, Seattle, WA 98195-4350

E-mail address: [ddrusv, hllee, rrthomas]@uw.edu