Part 3: Euclidean Distance Optimization

Tim Duff

February 14, 2025

For our purposes, a *real algebraic variety* $X \subset \mathbb{R}^n$ is any set defined implicitly by polynomial equations,

$$X = \{(x_1, \ldots, x_n) \mid f_1(x_1, \ldots, x_n) = \ldots = f_k(x_1, \ldots, x_n) = 0\}. \tag{1}$$

We shall be interested in the following **nearest-point problem:** given $u \in \mathbb{R}^n$ and a real algebraic variety $X \subset \mathbb{R}^n$, find $x \in X$ minimizing the (squared) Euclidean distance $\|x - u\|^2$.
Two remarks are in order. First, note that the minimum is always attained by some $x \in X$, since X is closed in the Euclidean topology on $\mathbb{R}^n$. Second, note that the same minimum is attained if we work with the non-squared distance $\|x - u\|$. Working with the squared distance is better for us because it extends to a polynomial over $\mathbb{C}$.

Example 1. Fix a matrix $A \in \mathbb{R}^{m\times n}$, and consider, for any $r \le \min(m, n)$, the set of matrices of rank at most r,

$$X_{\le r} = \{B \in \mathbb{R}^{m\times n} \mid \text{rk}(B) \le r\}.$$

This set is a real algebraic variety: note that $B \in X_{\le r}$ if and only if all of its $r \times r$ minors vanish, and each of these minors is a polynomial in the entries of B. Let $A = U\Sigma V^T$ be the SVD of A, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and the nonzero entries of $\Sigma \in \mathbb{R}^{m\times n}$ are $\sigma_1, \ldots, \sigma_{\text{rk}(A)}$ going down the diagonal. Letting $u_1, \ldots, u_m \in \mathbb{R}^m$ and $v_1, \ldots, v_n \in \mathbb{R}^n$ be the columns of U and V, respectively, this implies

$$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_{\text{rk}(A)} u_{\text{rk}(A)} v_{\text{rk}(A)}^T$$

Theorem 1 (Eckart-Young Theorem)**.** If $\text{rk}(A) > r$, then

$$\min_{B\in X_{\le r}} \|A - B\|^2 = \sigma_{r+1}^2 + \ldots + \sigma_{\text{rank}(A)}^2, \tag{2}$$

which is attained at

$$B_* = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T \in X_{\le r}. \tag{3}$$

In particular, when $m = n = r + 1$, we have

$$\min_{B\in X_{\le n-1}} \|A - B\|^2 = \sigma_n^2 = 1/\kappa[\text{inv}](A). \tag{4}$$

Equation (4) offers an interpretation of the Eckart-Young theorem as a *condition number theorem*, in which the squared distance from a problem to some locus of ill-posed problems is inversely propositional to the condition number of that problem. (Note: more generally, truncating the SVD as in Equation (3) still gives us the global optimimum if we replace the Frobenius norm $\| \bullet \|$ in 2 with any other orthogonally invariant matrix norm. For example, if we use the operator norm, than the minimum squared distance is just σ_{r+1}^2.)

In calculus class, you learned how to optimize a function by first computing its critical points. Later on, you probably learned about the method of Lagrange multipliers for solving constrained optimization problems. These same ideas can be used to solve nearest-point problems, although there are a few subtleties.

Example 2. Consider the parabola

$$X = \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}.$$

If we are given a data point $u = (u_1, u_2) \in \mathbb{R}^2$, how do we minimize its distance from X? To answer this question, it is helpful to leverage an alternative, *rational parametric description* of X as the image of the map

$$\begin{aligned} \phi : \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto (t, t^2). \end{aligned}$$

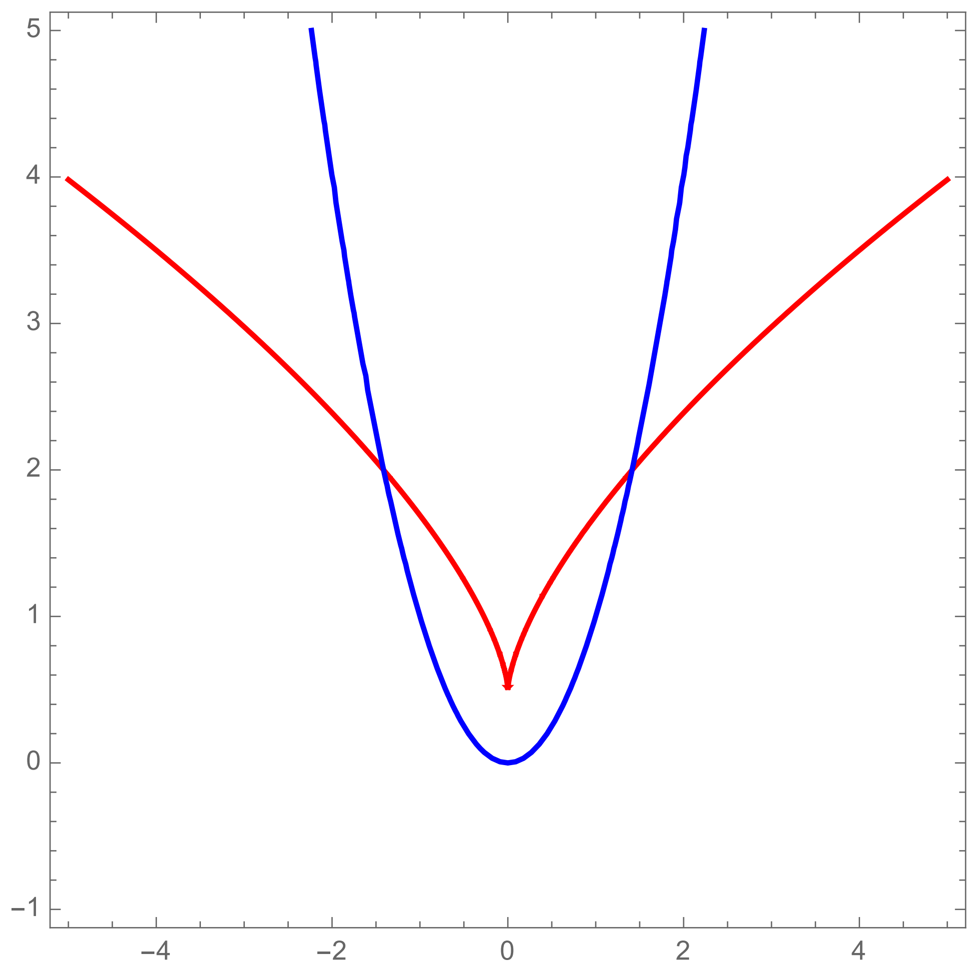

Figure 1: The parabola of Example 2 its evolute.

Using the parametrization ϕ, our problem reduces to minimizing a function of a single variable, namely

$$\ell(t;u) = (u_1 - t)^2 + (u_2 - t^2)^2,$$

and evaluating ϕ at t attaining the minimum. So let's differentiate:

$$\ell'(t;u) = 2(u_1 - t) + 2(u_2 - t^2)(-2t) = 4t^3 + (2 - 4u_2)t + 2u_1 = 0. \tag{5}$$

In principal, this can be solved algebraically[1] But it will be more helpful to look at specific examples, eg.

$$(u_1, u_2) = (0, 1) \quad \rightsquigarrow \quad \ell'(t;u) = 4t^3 - 2t = 2t(2t^2 - 1).$$

Thus, there are three critical points. For $t = 0$, we obtain $(x, y) = (0, 0)$, at distance 1 from u. For $t = \pm 1/\sqrt{2}$, we obtain $(x, y) = (\pm 1/\sqrt{2}, 1/2)$, which both attain the minimum squared distance

$$(1/\sqrt{2})^2 + (1 - 1/2)^2 = 3/4 < 1.$$

Note that the critical point $(x, y) = (0, 0)$ is a local (not global) maximum.

Let's try another data point:

$$(u_1, u_2) = (3, 0) \quad \rightsquigarrow \quad \ell'(t;u) = 4t^3 + 2t - 6 = 2(t - 1)(2t^2 + 2t + 3).$$

The quadratic factor $2t^2 + 2t + 3$ has discriminant $2^2 - 4 \cdot 2 \cdot 3 < 0$. Thus, there is only one critical point.

Already, Example 2 illustrates the following fact: *the number of real critical points is not constant over the space of data, but the number of complex critical points is.*

Analogously to the case of univariate polynomials, for any nearest point problem there is a set of "degenerate problem instances" $\vec{u} \in \mathbb{R}^n$ where we may expect the number of real solutions to change. In the case of the parabola, this degenerate set is a cubic curve with a cusp. This is the simplest example of an *evolute*; its defining equation may be obtained by taking the discriminant in t of the polynomial 5. See Figure 1 for an illustration.

Example 3. For the cuspidal cubic

$$X = \{(x, y) \mid x^3 - y^2 = 0\},$$

we have the rational parametrization $t \mapsto (x(t), y(t) = (t^2, t^3)$. To solve the nearest-point problem on X for data (u_1, u_2), we look at the roots of the polynomial

$$\frac{d}{dt}\left((u_1 - t^2)^2 + (u_2 - t^3)^2\right) = 6t^5 + 4t^3 - 6u_2t^2 - 4u_1t \tag{6}$$

Notice that $t = 0$ is always a root of this polynomial; in other words, the cusp point $(x(0), y(0)) = (0, 0)$ is always a critical point! The additional four roots may or may not be real.

Example 4. In general, an algebraic variety need not have a rational parametrization. The simplest example is an elliptic curve in the plane

$$X = \{(x, y) \mid \underbrace{y^2 - x(x + 1)(x - 1)}_{f(x,y)}\}.$$

Note that, also unlike the last example, X has no singular points. To find the critical points for the distance function $d_u(x, y) = (x - u_1)^2 + (y - u_2)^2$, we can use Lagrange multipliers: for some λ, we must have

$$\nabla f + \lambda \nabla d_u = 0.$$

In fact, we can get by without introducing a third variable λ by constructing an *augmented Jacobian* which must be rank-deficient,

$$\begin{bmatrix} x - u_1 & \partial_x f \\ y - u_2 & \partial_y f \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \end{bmatrix} \quad \Rightarrow \quad \det \begin{bmatrix} x - u_1 & \partial_x f \\ y - u_2 f & \partial_y f \end{bmatrix} = 0. \tag{7}$$

Thus, we need to solve a system of 2 equations in 2 variables:

$$\begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix} = \begin{bmatrix} f(x, y) \\ \det \begin{bmatrix} x - u_1 & y - u_2 \\ \partial_x f & \partial_y f \end{bmatrix} \end{bmatrix} \tag{8}$$

[1] Many authors, especially outside of math, refer to "analytic" solutions which are really "algebraic". What's up with that?

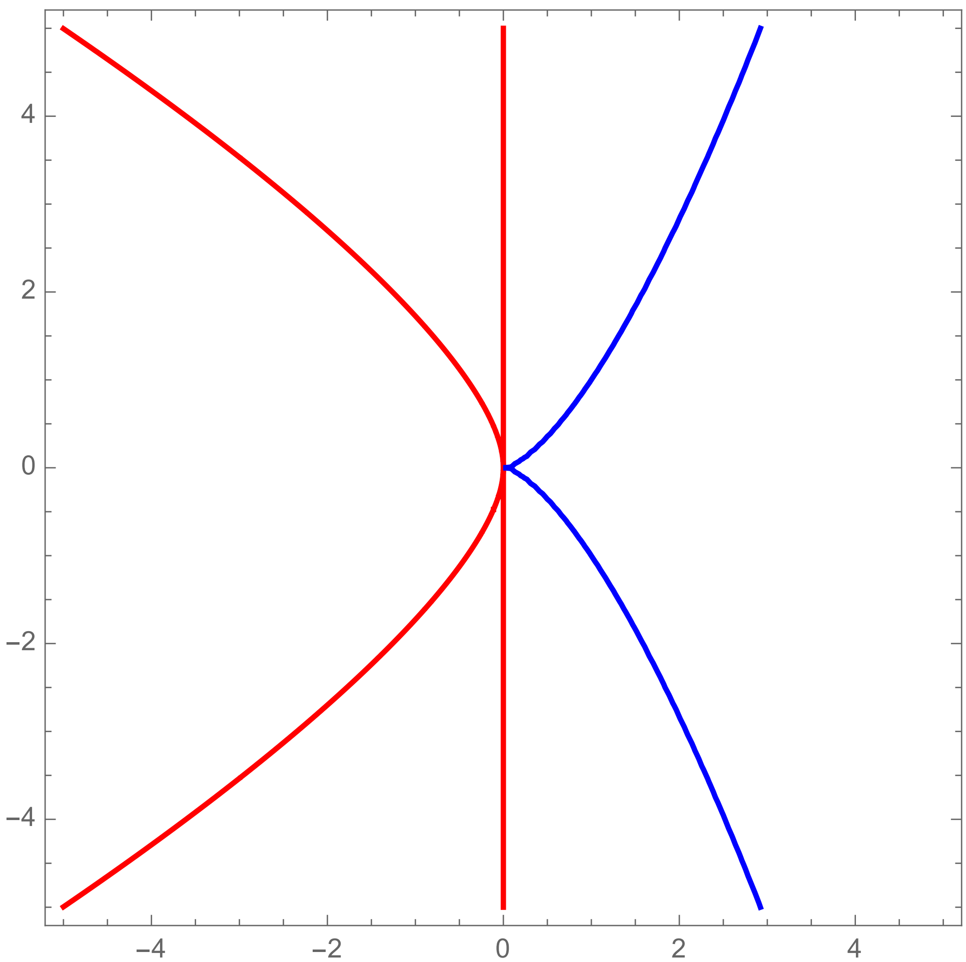

Figure 2: The cuspidal cubic (blue) of Example 3 and a discriminant for its nearest point problem (red), consisting of a straight-line component and a quartic curve (the latter of which is the evolute of the cubic.)

Notice that both g_1 and g_2 have total degree 3. Thus, to solve the system, it is tempting to use homotopy continuation with the following *total degree start system*, constructed analogously to our treatment of the univariate case:

$$\begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x^3 - 1 \\ y^3 - 1 \end{bmatrix}. \tag{9}$$

This start system has $3 \cdot 3 = 9$ solutions, so we might expect that the same is true for the target system 8. Surprisingly, though, the target system has only 7 complex solutions![2]

In general, the complexity of a nearest-point problem may be quantified by study of the so-called *Euclidean distance degree*, which we now define. To do so, consider the *complexification* of the real algebraic variety X in 1, defined as follows:

$$X_{\mathbb{C}} = \{\mathbf{x} \in \mathbb{C}^n \mid f_1(\mathbf{x}) = \ldots = f_k(\mathbf{x}) = 0\}. \tag{10}$$

(We'll also write $X_{\mathbb{R}} = X$ if we need to emphasize the field.) We will make some assumptions about X and the equations that define it:

A1 $f_1, \ldots, f_k$ generate a prime ideal in the polynomial ring $\mathbb{C}[\mathbf{x}]$.

A2 X contains a smooth real point.

We talked a bit about **A1** in class: a proper definition will appear in our notes later on. For **A2**, consider the $k \times n$ Jacobian matrix $J(f)$ whose entries are the partial derivatives of $f_1, \ldots, f_k$ with respect to the variables $x_1, \ldots, x_n$. For $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, we may define the $\mathbb{K}$-dimension of X

$$\dim(X_{\mathbb{K}}) = \min_{x \in X_{\mathbb{K}}} n - \operatorname{rk}(Jf(\mathbf{x})).$$

A point $\mathbf{x} \in X_{\mathbb{C}}$ is said to be *smooth* if $\dim(X_{\mathbb{C}}) = n - \operatorname{rk} Jf(\mathbf{x}))$. An example which fails to satisfy **A2** is the real variety $X = \{(x_1, \ldots, x_n) \mid x_1^2 + \ldots + x_n^2\}$ for any $n > 1$.

Definition / Theorem 1.

Our previous examples may be summarized as follows.

1. The ED degree of the parabola in Example 2 is 3.
2. The ED degree of the cuspidal cubic in Example 3 is 4 (since we must ignore the singular point.)
3. The ED degree of the elliptic curve is 7.

In all of the above examples, we may above complex critical points. However, there do exist examples where all critical points are real. For example, if $X_{r,n} \subset \mathbb{R}^{n \times n}$ is the variety of all $n \times n$ matrices of rank at most r, then

$$\operatorname{ED}(X_{n,r}) = \binom{n}{r}.$$

Furthermore, if A has SVD

$$A = \sum_{k=1}^{n} \sigma_k u_k v_k^T,$$

then an "enhanced" version of the Eckart-Young theorem implies that all critical points have the form

$$\sum_{j \in S} \sum_{k=1}^{n} \sigma_k u_k v_k^T, \qquad \text{for some } S \subset \{1, \ldots, n\} \text{ w/ } \#S = r.$$

References: Breiding, P, Kohn, K, and Sturmfels, B (2024). Metric Algebraic Geometry (Chapter 2).

[2] If you're familiar with Bezout's theorem: check that the system obtained by homogenizing $g_1 = g_2 = 0$ has a multiplicity-2 solution on the hyperplane at infinity in $\mathbb{P}^2$.