Ellipsoids, Rotations, and Singular Value Decomposition

Orestis Plevrakis

1 Revisiting Ellipsoids

In the second lecture, we saw that ellipsoids are sets of the form $\{x \in \mathbb{R}^n : (x-x_0)^\top A(x-x_0) \leq 1\}$, for some $A \succ 0$ and $x_0 \in \mathbb{R}^n$. Here is an equivalent definition:

Theorem 1. *Let E be an ellipsoid in $\mathbb{R}^n$. Then, $E = \{x \in \mathbb{R}^n : (x-x_0)^\top A^{-1}(x-x_0) \leq 1\}$, for some $A \succ 0$ and $x_0 \in \mathbb{R}^n$.*

Proof. First, observe that if $A \succ 0$, then A is invertible and $A^{-1} \succ 0$. This follows immediately from the diagonalization of A. Now, it follows that every set that has the form given in the theorem's statement is an ellipsoid. To show that any ellipsoid has this form, it suffices to use that $(A^{-1})^{-1} = A$ for any invertible matrix A. □

There is an alternative way of describing an ellipsoid, that is not an inequality. It is based on the concept of matrix "square-root".

Definition. Let $A \succcurlyeq 0$, and let $A = U\Lambda U^\top$ be its eigendecomposition, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with all $\lambda_i \geq 0$, and $UU^\top = I$. We define the square-root of A as $A^{1/2} := U\Lambda^{1/2}U^\top$, where $\Lambda^{1/2} := \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.

Observe that for $A \succcurlyeq 0$, we have $(A^{1/2})^2 = A$. Also, for $A \succ 0$, we have $(A^{-1})^{1/2} = (A^{1/2})^{-1}$. We will use $A^{-1/2}$ to denote this last matrix. We will also need the following notation:

Notation. For a set S in $\mathbb{R}^n$ and a matrix $A^{n\times n}$, we denote by $A(S)$ the image of S under A, i.e., the set $\{Ax : x \in S\}$. For any $x_0 \in \mathbb{R}^n$ we denote by $S + x_0$ the set $\{x + x_0 : x \in S\}$, i.e., the translation of S by x_0. Finally, we denote by B the unit ball $\{x \in \mathbb{R}^n : \|x\| \leq 1\}$.

Theorem 2. *Let $A \succ 0, x_0 \in \mathbb{R}^n$, and $E = \{x \in \mathbb{R}^n : (x-x_0)^\top A^{-1}(x-x_0) \leq 1\}$. Then, $E = A^{1/2}(B) + x_0$.*

Proof. Note that

$$(x-x_0)^\top A^{-1}(x-x_0) = (x-x_0)^\top A^{-1/2}A^{-1/2}(x-x_0) = \|A^{-1/2}(x-x_0)\|^2$$

Thus, $x \in E$ is equivalent to $A^{-1/2}(x-x_0) = y$ for some $y \in B$, which is equivalent to $x = A(y)+x_0$ for some $y \in B$. □

This theorem implies that the ellipsoids are exactly the images of affine functions with PD matrices, applied on the unit ball. A natural question is: what is the image of a general affine function $x \mapsto Ax + x_0$, applied on the unit ball? The following theorem says that if the function is invertible (equivalently, A is invertible) then we still get an ellipsoid.

Theorem 3. *A subset of $\mathbb{R}^n$ is an ellipsoid if and only if it is the image of the unit ball under an invertible affine function.*

Proof. One direction is implied by Theorem 2. For the other, let $A \in \mathbb{R}^{n\times n}$ invertible and $x_0 \in \mathbb{R}^n$. We will show that $A(B) + x_0$ is an ellipsoid.

$$x \in A(B) + x_0 \iff \|A^{-1}(x - x_0)\| \le 1 \iff (x - x_0)^\top (A^{-1})^\top A^{-1}(x - x_0) \le 1$$

In Homework 2, we saw that all matrices of the form $MM^\top$ are PSD. Now, if M is invertible, then $MM^\top$ is PD (why?). This concludes the proof. $\square$

It can be shown that for a non-invertible affine function, we will get a lower-dimensional ellipsoid, but we will not go into this. Now, ellipsoids are geometric objects and we just saw three new different ways of describing them. We proved the equivalence of these alternatives with short algebraic arguments. However, *where is the geometry?* We will now sketch a different proof for each of the above theorems. These proofs will be longer, but they will illuminate the underlying geometry and deepen our understanding.

Theorem 1

Let $A \succ 0, x_0 \in \mathbb{R}^n$. We consider the eigendecomposition of $A = U\Lambda U^\top$, and we have

$$\begin{aligned}(x - x_0)^\top A^{-1}(x - x_0) &= (x - x_0)^\top U\Lambda^{-1}U^\top(x - x_0) = (x - x_0)^\top U\Lambda^{-1/2}\Lambda^{-1/2}U^\top(x - x_0) \\ &= \|\Lambda^{-1/2}U^\top(x - x_0)\|^2 = \sum_{i=1}^{n} \frac{((x - x_0)\cdot u_i)^2}{\left(\sqrt{\lambda_i}\right)^2}\end{aligned}$$

where u_i is the i^{th} column of U. As we saw in Lecture 2, the inequality $\sum_{i=1}^{n} \frac{((x-x_0)\cdot u_i)^2}{\left(\sqrt{\lambda_i}\right)^2} \le 1$ describes an ellipsoid with axes along $u_1, \ldots, u_n$ and axis-lengths $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$. $\square$

To understand deeper what Theorem 2 is saying and why it holds, we need to understand first the geometry of orthogonal maps, i.e., maps of the form $x \mapsto Ux$, where U is orthogonal [1].

1.1 Orthogonal maps, isometries, and rotations

Theorem 4. *Orthogonal maps preserve norms and distances.*

Proof. Let U be an orthogonal matrix. For any $x \in \mathbb{R}^n$, $\|Ux\|^2 = x^\top U^\top Ux = x^\top x = \|x\|^2$, and so norms are preserved. Let $x, y \in \mathbb{R}^n$. Then, $\|Ux - Uy\|^2 = \|U(x - y)\|^2 = \|x - y\|^2$. $\square$

We can say even more about orthogonal maps.

The case of $\mathbb{R}^3$

Definition 5. A function $R : \mathbb{R}^3 \to \mathbb{R}^3$ is called a rotation if there is a line ℓ (the axis) passing through 0, and an angle $\phi \in [0, 2\pi)$ such that for all $x \in \mathbb{R}^3$, $R(x)$ is the rotation of x by angle ϕ around ℓ.

In general, the definition of a rotation does not include that the axis passes through the origin. However, in this course we will only deal with such rotations.

Theorem 6. *Let $U \in \mathbb{R}^{3\times 3}$ be an orthogonal matrix with columns u_1, u_2, u_3. Then, exactly one of the following holds:*

[1] Orthogonal matrix and orthonormal matrix mean the same thing.

- *The map $x \mapsto Ux$ is a rotation.*
- *The map* $x \mapsto \left[-u_1 \vdots u_2 \vdots u_3 \right] x$ *is a rotation.*

Proof. Assume first that the columns of U satisfy the right-hand-rule (RHR). Remember that this is equivalent to $det(U)$ being positive. We will use a very intuitive theorem of Euler.

Theorem. *Let v_1, v_2, v_3 and u_1, u_2, u_3 be two orthonormal bases of $\mathbb{R}^3$. Suppose they both satisfy the RHR. Then, there is a rotation R sending the first base to the second, i.e., $R(v_i) = u_i$ for $i = 1, 2, 3$.*

We apply the theorem with $v_i \leftarrow e_i$ and for u_i our own u_i. Let R be the rotation the theorem gives us. I claim that $R(x) = Ux$ for all x. Here is why: imaging continuously rotating x together with e_1, e_2, e_3 until they reach $R(x)$ and u_1, u_2, u_3. Then, during this motion, the projections of the rotating x on the vectors of the rotating basis do not change; they are x_1, x_2, x_3. Thus, when the rotation is completed, the final point is $R(x) = x_1u_1 + x_2u_2 + x_3u_3$. Now, if u_1, u_2, u_3 do not satisfy the RHR, then $x \mapsto Ux$ is definitely not a rotation (why?). However, $-u_1, u_2, u_3$ will satisfy it, and this completes the argument. $\square$

Back to $\mathbb{R}^n$

The concept of rotation can be generalized in $\mathbb{R}^n$, and most properties that we expect them to have, they do have them. For example, if $R : \mathbb{R}^n \to \mathbb{R}^n$ is a rotation, then it maps any ellipsoid, centered at 0 with axes along an orthonormal basis $u_1, \ldots, u_n$, onto another ellipsoid centered at 0 with axes along $R(u_1), \ldots, R(u_n)$, while maintaining the axis-lengths. In particular, it maps any ball centered at 0 onto itself. We will not give here a formal definition of rotations in $\mathbb{R}^n$, as it will not be needed to get the intuition I want to convey.[2] It turns out that Theorem 6 transfers identically in $\mathbb{R}^n$:

Theorem 7. *Let $U \in \mathbb{R}^{n\times n}$ be an orthogonal matrix with columns $u_1, u_2, \ldots, u_n$. Then, exactly one of the following holds:*

- *The map $x \mapsto Ux$ is a rotation.*
- *The map* $x \mapsto \left[-u_1 \vdots u_2 \vdots \ldots \vdots u_n \right] x$ *is a rotation.*

We won't prove this. Now, note that

$$U = \left[-u_1 \vdots u_2 \vdots \ldots \vdots u_n \right] \cdot \begin{bmatrix} -1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

where the second matrix is a reflection-matrix since when it acts on a vector x, it flips the sign of its first coordinate. Thus, here is how orthogonal maps act geometrically: *they are either rotations or compositions of a rotation with a reflection that flips the sign of the first coordinate.*

[2] The interested reader can check here to see what rotation means in $\mathbb{R}^n$.

Proof of Theorem 2

With our understanding of orthogonal maps, Theorem 2 becomes now self-evident: fix an $A \succ 0$ and $x_0 \in \mathbb{R}^n$. Consider the diagonalization of $A = U\Lambda U^\top$. Then, $A^{1/2}(B) + x_0 = (U\Lambda^{1/2}U^\top)(B) + x_0 = U(\Lambda^{1/2}(U^\top(B))) + x_0$. Let's ignore for the moment that U and $U^\top$ are transposes and inverses of each other and let's treat them as arbitrary orthogonal matrices. Now, the first (and most important case) is when both are rotations. Then, $U^\top(B) = B$. Furthermore, $\Lambda^{1/2}(B) = \left\{x \in \mathbb{R}^n : \sum_{i=1}^n \frac{x_i^2}{(\sqrt{\lambda_i})^2} \leq 1\right\}$, i.e., the ellipsoid centered at 0 with the standard axes, and axis-lengths $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ (why?). Now, again from the properties of rotations that I mentioned, we have that $U(\Lambda^{1/2}(B))$ is the same ellipsoid but with axes $u_1, \ldots, u_n$. The addition of x_0 just transfers the center. So, in the case of two rotations, we have proven the theorem. For the other cases, observe that for the reflection $F(x_1, x_2, \ldots, x_n) = (-x_1, x_2, \ldots, x_n)$, we have $F(B) = B$ and $F(\Lambda^{1/2}(B)) = \Lambda^{1/2}(B)$ (why?). This concludes the proof. □

Proof Theorem 3

From our previous discussion, it follows that if $A = U_1 \Sigma U_2$, where U_1, U_2 are orthogonal matrices and Σ is diagonal with positive diagonal elements, then $A(B) + x_0$ is an ellipsoid. It turns out that this not much to ask from a matrix:

Theorem 8. *Let $A \in \mathbb{R}^{n \times n}$. Then, there exists orthogonal $U, V \in \mathbb{R}^{n \times n}$ and diagonal $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$ with $\sigma_i \geq 0$, such that $A = U\Sigma V^\top$. If A is invertible, then all σ_i are positive.*

This decomposition is called *singular value decomposition* (SVD). Observe that Theorem 3 is now proven. □