Question 5

Anshika Raman
Roll No: 210050014

Kushal Agarwal
Roll No: 210100087

Kavan Vavadiya
Roll No: 210100166

November 7, 2024

Problem Setup:

We are given two sets of points $P_1 \in \mathbb{R}^{2\times N}$ and $P_2 \in \mathbb{R}^{2\times N}$ such that

$$P_1 = RP_2 + E,$$

where R is an orthonormal transformation matrix we aim to find, and $E \in \mathbb{R}^{2\times N}$ represents noise or error. Our goal is to determine R such that it minimizes the error E, subject to the constraint that R is orthonormal.

(a) Why does the standard least squares solution $R = P_1P_2^T(P_2P_2^T)^{-1}$ fail?

The standard least squares solution,

$$R = P_1P_2^T(P_2P_2^T)^{-1},$$

minimizes the Frobenius norm $\|E\|_F = \|P_1 - RP_2\|_F$, but it does not guarantee that R will be orthonormal.

To see why, we can examine whether the resulting matrix R obtained from this solution satisfies the orthonormality condition, $R^TR = I$.

1. Analyzing Orthonormality of R

An orthonormal matrix R should satisfy:

$$R^TR = I \quad \text{and} \quad RR^T = I.$$

We can check these conditions for $R = P_1P_2^T(P_2P_2^T)^{-1}$.

2. Computing R^TR

Let us compute R^TR:

$$R^TR = \left(P_1P_2^T(P_2P_2^T)^{-1}\right)^T \left(P_1P_2^T(P_2P_2^T)^{-1}\right).$$

Expanding and simplifying, we get:

$$R^TR = (P_2P_2^T)^{-1}P_2P_1^TP_1P_2^T(P_2P_2^T)^{-1}.$$

However, due to the presence of noise E in the equation $P_1 = RP_2 + E$, we have that $P_1P_2^T \neq RP_2P_2^T$. Therefore, $P_2P_1^TP_1P_2^T \neq P_2P_2^T$, meaning $R^TR \neq I$ in general. This implies that R does not satisfy the orthonormality constraint.

3. Computing RR^T
Similarly, we can compute RR^T as follows:

$$RR^T = \left(P_1 P_2^T (P_2 P_2^T)^{-1}\right) \left(P_1 P_2^T (P_2 P_2^T)^{-1}\right)^T.$$

Expanding, we get:

$$RR^T = P_1 P_2^T (P_2 P_2^T)^{-1} (P_2 P_2^T)^{-1} P_2 P_1^T.$$

Due to the noise matrix E, this expression does not simplify to I. Therefore, $RR^T \neq I$.

4. Conclusion
The fact that $R^T R \neq I$ and $RR^T \neq I$ demonstrates that the least squares solution $R = P_1 P_2^T (P_2 P_2^T)^{-1}$ fails to satisfy the orthonormality constraint required for a valid transformation matrix R. This failure necessitates an alternative approach that incorporates the orthonormality constraint, which we will explore in subsequent parts of this assignment.

(b) Deriving the Expression for $E(R)$

We are given:

$$E(R) = \|P_1 - RP_2\|_F^2$$

Expanding $E(R)$ in terms of the trace function, we obtain:

$$E(R) = \mathrm{trace}((P_1 - RP_2)^T (P_1 - RP_2)).$$

Expanding this trace expression using the distributive property of the transpose, we get:

$$E(R) = \mathrm{trace}(P_1^T P_1 + P_2^T R^T R P_2 - P_2^T R^T P_1 - P_1^T R P_2).$$

Since R is orthonormal, $R^T R = I$, so we have:

$$E(R) = \mathrm{trace}(P_1^T P_1 + P_2^T P_2 - P_2^T R^T P_1 - P_1^T R P_2).$$

We apply the **cyclic property of the trace**, which states that for any matrices A and B of compatible dimensions, $\mathrm{trace}(AB) = \mathrm{trace}(BA)$, allowing us to rewrite terms without changing their values. Thus, we simplify further as follows:

$$E(R) = \mathrm{trace}(P_1^T P_1 + P_2^T P_2) - 2\,\mathrm{trace}(P_1^T R P_2).$$

(c) Why is minimizing $E(R)$ with respect to R equivalent to maximizing trace$(P_1^T R P_2)$?

The term $\mathrm{trace}(P_1^T P_1 + P_2^T P_2)$ is constant with respect to R and does not affect the optimization. Therefore, minimizing $E(R)$ reduces to minimizing $-2\mathrm{trace}(P_1^T R P_2)$, which is equivalent to maximizing $\mathrm{trace}(P_1^T R P_2)$.

(d) Justifying Trace Manipulations

We express trace($P_1^T R P_2$) as:

$$\text{trace}(P_1^T R P_2) = \text{trace}(R P_2 P_1^T).$$

The justification for this step lies in the cyclic property of the trace function. Since trace(AB) = trace(BA), we can rearrange the terms in the trace without changing the value, allowing us to rewrite trace($P_1^T R P_2$) as trace($R P_2 P_1^T$).

Given $P_2 P_1^T = U' S' V'^T$, the SVD of $P_2 P_1^T$, we substitute to get:

$$\text{trace}(R P_2 P_1^T) = \text{trace}(R U' S' V'^T).$$

Using the cyclic property of the trace, we can rewrite this as:

$$\text{trace}(R U' S' V'^T) = \text{trace}(S' V'^T R U') = \text{trace}(S' X),$$

where $X = V'^T R U'$.

(e) Maximizing the Trace with Respect to X

We now need to maximize trace($S'X$), where S' is diagonal. Given that S' has non-negative entries (as it is derived from the SVD), the trace will be maximized when each entry of S' is paired with an entry of 1 in X, thus achieving the maximum product. Therefore, to maximize trace($S'X$), we require $X = I$, meaning X should be the identity matrix. This ensures that each diagonal element of S' contributes fully to the trace, giving us the maximum possible sum.

(f) Determining R

To extend the reasoning from part (e), we aim to derive the optimal matrix R that maximizes the expression trace($P_1^T R P_2$).

In part (e), we reached:

$$\text{trace}(P_1^T R P_2) = \text{trace}(S' X),$$

where S' is a diagonal matrix containing the singular values of $P_2 P_1^T$, and $X = V'^T R U'$.

Since S' is diagonal, the best way to maximize the trace expression trace($S'X$) is to set $X = I$, the identity matrix. This is because the trace of S' will be maximized when X aligns perfectly with S'.

Thus, to achieve $X = I$, we require:

$$X = V'^T R U' = I.$$

Rearranging this equation gives:

$$R = V' U'^T.$$

(g) Additional Constraint for R as a Rotation Matrix

If we specifically need R to represent a rotation matrix, we must impose an additional constraint: the determinant of R must be $+1$. A rotation matrix not only preserves lengths and angles (orthonormality) but also preserves orientation. Therefore, the condition $\det(R) = +1$ is essential to ensure that R is a proper rotation matrix rather than a reflection.

To enforce this constraint, we select the sign of U' or V' in the SVD decomposition such that $\det(U'V'^T) = +1$, which guarantees that R represents a rotation.