# 3D Geometry for Computer Graphics – Exercise 3 Solution

1. Suppose you have a set of points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ that were sampled from some real-world data. For example, these could be points coming from some sensor, collected in a physical experiment, statistical data or data coming from a 3D scanner. You are a scientist that wants to perform principal component analysis on the data you collected and to find a line/hyperplane that approximates your points. However, it is known that the dataset contains *noise* and *outliers* – some points might be far off (see Figure 1). How do errors in the samples affect the result of the PCA? (Hint: outliers will confuse/divert the PCA. Explain why, based on the quadratic expression that PCA minimizes.)

   *Answer:* The cost function that PCA minimizes is quadratic:

   $$\min \sum_i \|\mathbf{x}_i - \mathbf{x}'_i\|^2,$$

   where $\mathbf{x}_i$ are the original points and $\mathbf{x}'_i$ are the projected points. This expression highly punishes points whose projection is far away from the original location: if $\|\mathbf{x} - \mathbf{x}'\| = d$ then the above some is increased by $d^2$. Thus, PCA will try to move the principal line closer to the outliers – it will cost less than having the line close to most of the points and having several outliers.

2. Try to suggest ways to overcome the problem in the previous question.

   *Answer:* This is where you were supposed to be creative... In general, it is a very important area of research, called *robust statistics*. How to decide what is an outlier and to ignore it. Some ways to overcome the problem are to choose another cost function; to randomize the PCA; to make the points weighted.

3. One of the everyday applications of PCA in computer graphics is computation of tight bounding boxes for complex objects, as mentioned in class. However, in general PCA does not give the minimal-volume bounding box. Explain why. (Hint: think how PCA would be affected by varying density of the points on the object, for example, if in some part of the object there is a very high concentration of points.)

   *Answer:* Suppose we have an object with many points in its interior. The amount of inner points doesn't change the surface of the object, and thus does not change the minimal bounding box. But it does change the result of PCA: if we have a lot of inner points, they will affect the scatter matrix $S$ and divert the principal components.

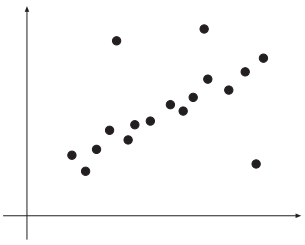

Figure 1: The points represent the measured data. There is noise and some points erroneously fall far off (they are *outliers*).