3D Geometry for Computer Graphics - Exercise 3

18/03/2004

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.)

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

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.)

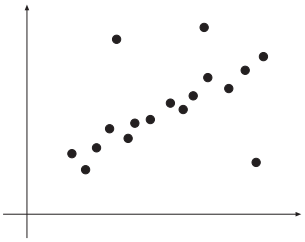

Figure 1: The points represent the measured data. There is noise and some points erroneously fall far off (they are *outliers*).