

Paper Number: 812

High dimensional data visualization helps data mining from well log data

Dong, S.Q.¹, Zeng, L.B.¹

¹ China University of Petroleum-Beijing, 18 Fuxue Road Changping, Beijing, China and dshaoqun@qq.com.

Reservoir properties are reflected in different well log responses. A better level of reservoir characterisation can be achieved if the many different types of responses used for interpretation can be integrated. However, high-dimensional datasets can be very difficult to visualize, which increases the difficulty for data mining from well logs, such as searching for criterion for lithology identification. Problems associated with visualization of, and understanding of relationships in high-dimensional datasets hinders reservoir interpretation.

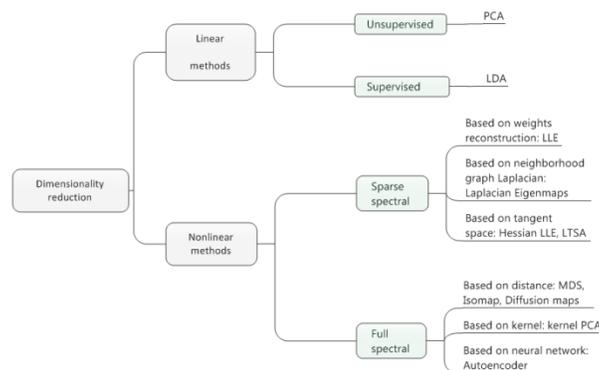


Figure 1: Taxonomy of dimensionality reduction methods.

Reduction of dimensionality is considered to be a solution since it facilitates classification, visualization, and compression of high dimensional data [1]. Recently, a large number of linear and nonlinear techniques for dimensionality reduction have been proposed. Figure.1 shows the taxonomy of techniques for dimensionality reduction. It is subdivided into linear and nonlinear techniques. These techniques can convert the high-dimensional data set $X = \{x_1, x_2, \dots, x_n\}$ into two or three-dimensional data $Y = \{y_1, y_2, \dots, y_m\}$ that can be displayed in a scatterplot [2].

The implementation of nonlinear techniques using well logs is suggested here. The aim is to obtain a projection similar to that shown in Figure 2, which employs a nonlinear, convex and sparse spectral method called locally linear embedding (LLE) that preserves the local structure of data. In this way the general topological structure of the original high dimensional space can be visualized. The novel dimensionality reduction techniques are compared with ordinary linear ones and summarize inherent weaknesses of the nonlinear dimensionality reduction techniques by both a theoretical and an empirical

evaluation. This allows geologists to choose a proper *dimensionality reduction* method for their specific issues.

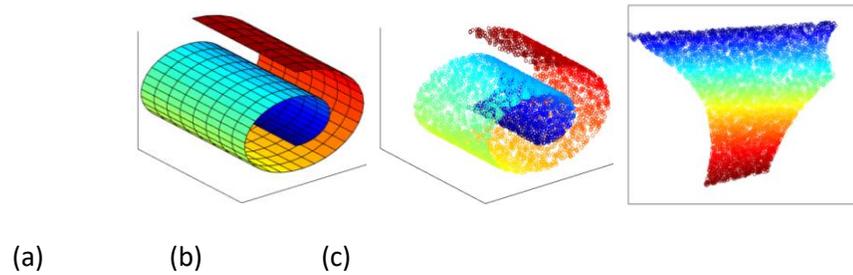


Figure 2: The process of dimensionality reduction by LLE. (a) The intrinsic function for the real-world problem. (b) The samples collected. (c) The feature distribution of (b) after dimensionality reduction.

References:

[1] Laurens van der Maaten et al. (2009) In: *Dimensionality reduction: A comparative review*: Journal of Machine Learning Research, 66-71.

[2] Laurens van der Maaten et al. (2008) In: *Visualizing data using t-SNE*: Journal of Machine Learning Research, 2579-2605.

