Chapter 9
Nonlinear Partial Least Squares:
An Overview

Roman Rosipal
Medical University of Vienna, Austria & Pacific Development and Technology, LLC, USA

ABSTRACT

In many areas of research and industrial situations, including many data analytic problems in chemistry, a strong nonlinear relation between different sets of data may exist. While linear models may be a good simple approximation to these problems, when nonlinearity is severe they often perform unacceptably. The nonlinear partial least squares (PLS) method was developed in the area of chemical data analysis. A specific feature of PLS is that relations between sets of observed variables are modeled by means of latent variables usually not directly observed and measured. Since its introduction, two methodologically different concepts of fitting existing nonlinear relationships initiated development of a series of different nonlinear PLS models. General principles of the two concepts and representative models are reviewed in this chapter. The aim of the chapter is two-fold i) to clearly summarize achieved results and thus ii) to motivate development of new computationally efficient nonlinear PLS models with better performance and good interpretability.

INTRODUCTION

Two-block linear partial least squares (PLS) has been proven to be a valuable method for modeling relationships between two data sets (data blocks). This method was developed in chemometrics and has received a great deal of attention in the fields of analytical chemistry, organic and bio-organic chemistry, medicinal chemistry and chemical engineering. PLS has also been successfully applied in other scientific areas including bioinformatics (Boulesteix & Strimmer, 2007), food research (Martens & Martens, 1986), medicine (Worsley, 1997), pharmacology (Leach & Gillet, 2003, Nilsson, Jong, & Smilde, 1997), social sciences (Hulland, 1999), physiology and neurophysiology (Lobaugh, West, & McIntosh, 2001, Trejo, Rosipal, & Matthews, 2006), to name a few.

DOI: 10.4018/978-1-61520-911-8.ch009
PLS models relationships between sets of observed variables by means of latent variables. It can serve for regression and classification tasks as well as dimension reduction techniques and modeling. The underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables. This projection of the observed data onto a subspace of usually orthogonal latent variables has been shown to be a powerful technique when observed variables are highly correlated, noisy and the ratio between the number of observations (data samples) and observed variables is low. The basic assumption of linear PLS is that the studied relation between observed data sets is linear and the same assumption of linearity then holds for modeling the relation in the projected subspace; that is, between latent variables.

However, in many areas of research and industrial situations a strong nonlinear relation between sets of data may exist. Although linear PLS can be used to approximate this nonlinearity, in many situations such approximation may not be adequate and the use of a nonlinear model is needed.

This chapter introduces the main concepts of nonlinear PLS and provides an overview of its application to different data analysis problems. The aim is to present a concise introduction that is a valuable guide for anyone who is concerned with nonlinear data analysis.

BACKGROUND

The concept of nonlinear PLS modeling was introduced by S. Wold, Kettaneh-Wold, and Skagerberg (1989). Already in this seminal work, the authors distinguished and described two basic principles for modeling curved relationships between sets of observed data. The first principle, here denoted as Type I, is well-known and used in mathematical statistics and other research fields. The principle applies first a nonlinear transformation to observed variables. In the new representation a linear model is constructed. This principle can be easily applied to PLS, and indeed several different nonlinear PLS models were proposed and applied to real data sets. The first nonlinear PLS models in this category were constructed by using simple polynomial transformations of the observed data (Berglund & Wold, 1997, 1999). However, the proposed polynomial transformation approach possesses several computational and generalization limitations. To overcome these limitations, a computationally elegant kernel PLS method was proposed by Rosipal and Trejo (2001). The powerful concept of a kernel mapping function allows to construct highly flexible but still computationally simple nonlinear PLS models. However, in spite of the ability of kernel PLS to fit highly complex nonlinear data relationships, the model represents a ‘black-box’ with limited possibility to interpret the results with respect to the original data.

It is the second, here denoted as Type II, general principle for constructing nonlinear PLS models which overcomes the problem of loss of interpretability, but this is achieved at the expense of computational cost and optimization complexity. In contrast to the Type I principle, a nonlinear relation between latent variables is assumed and modeled, while the extracted latent vectors themselves are kept to be a linear combination of the original, not transformed, data. A quadratic function was used to fit relationship between latent variables in the first Type II nonlinear PLS approaches (Höskuldsson, 1992, S. Wold et al., 1989). Later, smoothing splines (Frank, 1990, S. Wold, 1992), artificial neural networks (Baffi, Martin, & Morris, 2000, Qin & McAvoy, 1992), radial basis neural networks (Wilson, Irwin, & Lightbody, 1997) or genetic programming methods (Hiden, McKay, Willis, & Montague, 1998) were used to fit more complex nonlinear relationships. Computational and optimization difficulties of the approach arise at the point when initially estimated weights for projecting observed data into latent vectors need to be corrected. The initial weights are estimated...