Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Statistics & Data Analysis 46 (2004) 11–31
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Partial least-squares vs. Lanczos bidiagonalization—I: analysis of a projection method for multiple regression

Lars Eldén*

Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden

Received 1 December 2002; received in revised form 3 June 2003; accepted 10 June 2003

Abstract

Multiple linear regression is considered and the partial least-squares method (PLS) for computing a projection onto a lower-dimensional subspace is analyzed. The equivalence of PLS to Lanczos bidiagonalization is a basic part of the analysis. Singular value analysis, Krylov subspaces, and shrinkage factors are used to explain why, in many cases, PLS gives a faster reduction of the residual than standard principal components regression. It is also shown why in some cases the dimension of the subspace, given by PLS, is not as small as desired.

Keywords: Partial least-squares; Lanczos bidiagonalization; Singular value decomposition; Principal components regression; Krylov subspace; Chemometrics; Shrinkage factors

1. Introduction

Consider the minimization problem

$$\min_{\beta} \|y - X\beta\|, \tag{1}$$

where X is an $m \times n$ real matrix, and the norm is the Euclidean vector norm. This is the *linear least-squares problem* in numerical linear algebra, and the *multiple linear regression problem* in statistics. In the latter case the vector y consists of observations of a response variable, and the columns of X contain the values of the explanatory variables. Often the matrix is large and ill-conditioned: the column vectors are (almost) linearly

* Tel. +46-13-282183; fax: +46-13-136053.
E-mail address: laeld@math.liu.se (L. Eldén).

0167-9473/$ - see front matter
doi:10.1016/S0167-9473(03)00138-5

dependent, which is equivalent to the explanatory variables being (almost) collinear. In such cases the straightforward solution of (1) may be physically meaningless (from the point of view of the application at hand) and difficult to interpret. Then one may want to express the solution by projecting it onto a lower-dimensional subspace: let V be an $m \times k$ matrix with orthonormal columns. Using this as a basis for the subspace, one considers the approximate minimization

$$\min_{\beta} \|y - X\beta\| \approx \min_{z} \|y - XVz\|.$$

One obvious method for projecting the solution onto a low-dimensional subspace is *principal components regression* (*PCR*) (Massy, 1965). In numerical linear algebra this method is called *truncated singular value decomposition* (*TSVD*). Another such method, the *partial least-squares* (*PLS*) method (Wold, 1975), is used in chemometrics for determining the projection matrix (Wold et al., 2001). It has been known for quite some time (Wold et al., 1984) (see also Helland, 1988; Di Ruscio, 2000; Phatak and de Hoog, 2002) that PLS is equivalent to Lanczos (Golub-Kahan) bidiagonalization (Golub and Kahan, 1965; Paige and Saunders, 1982) (we will refer to this as LBD).

PLS is a standard method in chemometrics since at least two decades (Wold et al., 2001), and it is often used for visualization of scores and loadings in practical data analysis, complemented with other diagnostics. However, its properties seem not to be completely understood, partly because *the solution β is a nonlinear function of the data y*. The purpose of this paper is to analyze and illustrate some properties of PLS, using the equivalence to LBD and a notion called *shrinkage factors* (Frank and Friedman, 1993), which is a technique based on the singular value decomposition (principal components). We give an example that explains why PLS often leads to a faster reduction of the residual than traditional PCR (Phatak and de Hoog, 2002). In particular, we claim that *singular value analysis is essential in understanding the properties of PLS*.

In order to set the stage for our analysis of PLS, we first give a brief introduction to PCR in Section 2. In Section 3 we present the NIPALS version of the PLS method, followed by a description of Lanczos bidiagonalization. Then we demonstrate the equivalence between the two methods by proving that they give the same sequence of orthogonal basis vectors. An analysis of the behavior of PLS, in terms of shrinkage factors is given in Section 5.

The emphasis of this paper is to demonstrate the properties of the methods using examples. More theoretical discussions are given in, e.g., (Burnham et al., 1996; Frank and Friedman, 1993; Di Ruscio, 2000; Helland, 2001; Phatak and de Hoog, 2002).

In spite of the fact that PLS reduces the residual faster than traditional PCR, there are problems where the dimension of the subspace produced is considered to be inconveniently large (Wold et al., 1998) (in the sense that there is an initial phase of the PLS algorithm, where the residual is not reduced substantially). Actually, our examples demonstrate why this happens. For such problems, a preprocessing approach is often used, so called *orthogonal signal correction*. One such method is called O-PLS (Trygg and Wold, 2002). In a sequel paper (Eldén and Ranjbar, 2003) we will demonstrate