Signal Processing 90 (2010) 2822–2834

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Data compression under constraints of causality and variable finite memory

A. Torokhti *, S.J. Miklavcic

University of South Australia, School of Mathematics and Statistics, 1 Mawson Lakes Blv, 5095, Mawson Lakes, SA, Australia

ARTICLE INFO

Article history:
Received 22 May 2009
Received in revised form
5 January 2010
Accepted 5 April 2010
Available online 13 April 2010

Keywords:
Data compression
Wiener filtering
Causality
memory

ABSTRACT

Data compression techniques mainly consist of two operations, data compression itself and a consequent data de-compression. In real time, the compressor and de-compressor are causal and, at a given time, may process (or 'remember') only a fragment of the input signal. In the latter case, we say that such a filter has a finite memory. We study a new technique for optimal real-time data compression. Our approach is based on a specific formulation of two related problems so that one problem is stated for data compression and another one for data de-compression. A compressor and de-compressor satisfying conditions of causality and memory are represented by matrices with special forms, A and B, respectively. A technique for the solution of the problems is developed on the basis of a reduction of minimization problems, in terms of matrices A and B, to problems in terms of specific blocks of A and B. The solutions represent data compressor and data de-compressor in terms of blocks of those matrices that minimize associated error criteria. The analysis of the associated errors is also provided.

1. Introduction

A study of data compression methods is motivated by the necessity to reduce expenditures incurred with the transmission, processing and storage of large data arrays [1,2]. Such methods have also been applied successfully to the solution of problems related, e.g., to clustering [3], feature selection [4,5], forecasting [6,7] and estimating the medium from transmission data [8].

Data compression techniques are often performed on the basis of the Karhunen–Loève transform (KLT),[1] which is closely related to the principal component analysis (PCA). A basic theory for the KLT-PCA can be found, for example, in [1,4,9]. In short, the KLT-PCA produces a linear operator of a given rank that minimizes an associated error over all linear operators of the same rank. In a standard KLT–PCA application (e.g., presented in [4]), an observable signal and a reference signal are the same. In other words, the standard KLT–PCA provides data compression only and no noise filtering.

Scharf [1,10] presented an extension of the PCA–KLT[2] for the case when an observable signal $\mathbf{y}$ and a reference signal $\mathbf{x}$ are different and no explicit analytic representation of $\mathbf{y}$ in terms of $\mathbf{x}$ is known. In particular, $\mathbf{y}$ can be a noisy version of $\mathbf{x}$. The method [1,10] assumes that covariance matrix E_{yy}, formed from $\mathbf{y}$, is nonsingular. Yamashita and Ogawa [11] studied and justified a version of the PCA–KLT for the case where the covariance matrix E_{yy} may be singular and $\mathbf{y}=\mathbf{x}+\mathbf{w}$ with $\mathbf{w}$ an additive noise. Hua and Liu [12] considered an extension of the PCA–KLT to the case when E_{yy} is singular, and $\mathbf{y}$ and $\mathbf{x}$ are as in [1,10]. Torokhti and Friedland [15] studied a

* Corresponding author. Tel.: +61 8 8302 3812; fax: +61 8 8302 5785.
E-mail address: anatoli.torokhti@unisa.edu.au (A. Torokhti).

[1] The KLT is also known as Hotelling Transform and eigenvector transform.

[2] The list of references related to the PCA–KLT is very long. For example, a Google search for 'Karhunen-Loève transform and principal component analysis' gives 9230 items. Here, we mention only the most relevant references to the problem under consideration.

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.04.001

weighted version of the PCA–KLT. Torokhti and Howlett [16,17] extended the PCA–KLT to the case of optimal non-linear data compression. Torokhti and Manton [18] further advanced results in [15–17] to the so-called generic weighted filtering of stochastic signals. Advanced computational aspects of the PCA–KLT were provided, in particular, by Hua, Nikpour and Stoica [2], Hua and Nikpour [19], Stoica and Viberg [20], and Zhang and Golub [21]. Other relevant references can be found, e.g. in the bibliographies of the books by Scharf [1], and Torokhti and Howlett [9].

While the topics of data compression have been intensively studied (in particular, in the references mentioned above), a number of related fundamental questions remain open. One of them concerns real-time data processing. In this paper, the real-time aspect of the data compression problem is the dominant motivation for considering specific restrictions associated with causality and memory. Similar observations motivated studies in works by Gastpar, Dragotti and Vetterli [13], Roy and Vetterli [14], Fomin and Ruzhansky [22], Torokhti and Howlett [23], and Howlett, Torokhti and Pearce [24]. In [13,14], distributed signal estimation has been studied for the case of multiple sensors, each observing only a part of the input signal. In [13,14], the statement of the data compression problem and associated assumptions are different from those considered below. In [22–24], a causal Weiner-like filtering with memory has been considered, but not in the context of data compression.

We note that conditions of causality and memory make the problem very specific and difficult. In Section 3.2, we provide a new approach to the problem solution and give an analysis of the associated error. In more detail, motivations to consider the problem are as follows.

First motivation: causality and memory. Data compression techniques mainly consist of two operations, data compression itself and a subsequent data de-compression (or reconstruction). In real time, the compressor and de-compressor are causal and may be performed with a memory.

A causality constraint follows from the observation that in practice, the present value of the output of a filter [3] is not affected by future values of the input [25]. To determine the output signal at time t_k, with $k=1,\dots,m$, *the causal filter* should 'remember' the input signal up to time t_k, i.e., at times t_k, t_{k-1}, …, t_1. Such a situation is typical, for example, in a medical computer diagnostic [5].

A memory constraint is motivated as follows. The output of the compressor and/or de-compressor at time t_k with $k=0,1,\dots,m$, may only be determined from a 'fragment' of the input defined at times $t_k, t_{k-1}, \dots, t_{k-(\eta_k-1)}$ with $\eta_k = 1,\dots,k$. In other words, compressor and de-compressor should 'remember' that fragment of the input. The 'length' of the fragment for a given k, i.e. the number η_k, is called a local memory. The local memory could be different for different k, therefore, we also say that η_k is a local *variable* memory. A collection of the local memories, $\{q_1,\dots, q_m\}$, is called the *variable finite memory* or simply *variable memory*. A formalization of these concepts is given in Section 3.1. Matrices that form filter models with a variable memory possess special structure. Some related examples are given in Section 3.1.

Thus, our first motivation, to consider the problem in the form presented in Section 2.4 below, comes from the observation that the compressor and de-compressor, used in real time, should be causal with variable finite memory.

Second motivation: specific formulation of the problem. In reality, the compression and de-compression are separated in time. Therefore, it is natural to pose optimization problems for them separately, one specific problem for each operation, compression and de-compression. Associated optimization criteria could be formulated in many ways. Some of them are discussed in Appendix A, and we show that those criteria lead to significant difficulties. To avoid the difficulties considered in Appendix A, a new approach to the solution of the data compression problem is presented here. The approach is based on a specific formulation of two related problems given in Section 2. Solutions of those problems represent an associated optimal compressor and optimal de-compressor, respectively. It is shown in Section 3.1 that the optimal compressor and de-compressor satisfying conditions of causality and variable finite memory must have special forms. This implies that signals processed by these operators should be presented in special forms as well. In Sections 3.1 and 3.2 this issue is discussed in detail.

Next, traditionally, the data compression problem is studied in terms of linear operators, mainly due to the simplicity of their implementation. See, for example, [1–4] and [10–15] and references herein. Here, we extend the approaches of linear data compression proposed in [1–4] and [10–15]. A case of non-linear compression and de-compression with causality and memory is more complicated, and it can be studied on the basis of results obtained below combined, e.g., with the approaches to optimal non-linear filtering presented in [9,16–18,24].

2. Basic idea and statement of the problem

2.1. Informal statement of problem

In an informal way, the data compression problem we consider can be expressed as follows.

Let (Ω,Σ,μ) be a probability space, where $\Omega=\{\omega\}$ is the set of outcomes, Σ a σ-field of measurable subsets in Ω and $\mu:\Sigma\to[0,1]$ an associated probability measure on Σ with $\mu(\Omega)=1$. Let $\mathbf{y}\in L^2(\Omega,\mathbb{R}^n)$ be observable data and $\mathbf{x}\in L^2(\Omega,\mathbb{R}^m)$ be a reference signal that is to be estimated from $\mathbf{y}$ in such a way that

(i) first, data $\mathbf{y}$ should be compressed to a shorter vector $\mathbf{z}\in L^2(\Omega,\mathbb{R}^r)$[4] with $r<\min\{m,n\}$,
(ii) then $\mathbf{z}$ should be decompressed (reconstructed) to a signal $\tilde{\mathbf{x}}\in L^2(\Omega,\mathbb{R}^m)$ so that $\tilde{\mathbf{x}}$ is 'close' to $\mathbf{x}$ in some appropriate sense, and

[3] When a context is clear, we use the term 'filter' for both compressor and de-compressor.

[4] Components of $\mathbf{z}$ are often called *principal components* [4].