Journal of Process Control 47 (2016) 175–190

Contents lists available at ScienceDirect

Journal of Process Control

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

Reconstruction of missing data using compressed sensing techniques with adaptive dictionary

Satheesh K. Perepu, Arun K. Tangirala*

Dept. of Chemical Engineering, IIT Madras, Chennai 600036, India

ARTICLE INFO

Article history:
Received 28 May 2016
Received in revised form 9 August 2016
Accepted 25 August 2016

Keywords:
Compressed sensing
Random sampling
Missing data
Adaptive dictionary
Irregular sampling
Block segmentation
Dictionary learning

ABSTRACT

Missing data is a commonly encountered and challenging issue in data-driven process analysis. Several methods that attempt to estimate missing observations for the purpose of control, identification, etc. have been developed over the decades. However, existing methods tend to produce erroneous estimates when the percentage of missing data is high and mostly do not exploit the benefit of parsimonious or sparse signal representations. Recently developed compressed sensing (CS) techniques are naturally suited to handle the problem of missing data recovery since they provide powerful signal recovery methods that take advantage of sparse representations of signals in a set of functions, known as the *overcomplete dictionary*. A majority of these signal recovery algorithms assume that the dictionary is known beforehand. This paper presents a method to estimate missing observations using CS ideas, but with an adaptive learning of the overcomplete dictionary from data. The method is particularly devised for signals that have a block-diagonal sparse representation, an assumption that is not too restrictive. An iterative optimization method, consisting of an iterative CS problem on block-segmented data, for discovering this sparsifying dictionary is presented. Further, we present theoretical and practical guidelines for the segmentation size. It is shown that the error at each iteration is bounded for the exact, i.e., zero model mismatch and noise-free, case. Demonstrations on five different systems illustrate the efficacy of the proposed method with respect to recovery of missing data and convergence properties. Finally, the method is observed to require fewer observations than a fixed dictionary for a given reconstruction accuracy.

1. Introduction

Control strategies, both classical and modern, critically depend on the availability of measurements. Industrial measuring devices (sensors) are known to experience intermittent failures and/or produce outliers [19]. On the other hand, it is also a widely accepted fact that measurements may be only available on an irregular basis in several applications, especially where manual measuring mechanisms are deployed; for example, in the measurement of molecular weight of a polymer, finesses of cement, etc. The foregoing three apparently different scenarios can be, in fact, perceived as a single problem of missing data since both the cases of outliers (after identifying their locations) and irregular samples can be viewed as that of regular sampling with missing data. In order to ensure active, uninterrupted and quality control, it is therefore highly essential to be equipped with methods that handle missing data. Consequently, developing methods for missing data reconstruction has been an active topic of research for nearly a century now [29]. Recent literature, especially in the last two decades, has witnessed the development of a good deal of control and identification algorithms in presence of irregularly sampled data [34,42,7,47,46,18]. One of the earliest communicated methods was by Yates [45], which assumed missing data in the response variable of a process. It involves first estimating the parameters of a regression model (for the response) from a block of complete observations and replacing the missing data with the predictions from that model. However, in the presence of noise, this method tends to produce estimates with large errors, which is one of its main drawbacks. Bartlett [3] proposed an iterative method that primarily minimized the errors between statistical properties of blocks of complete data and incomplete data. Although this method gives unbiased estimates, it suffers from certain shortcomings that are similar to that of Yates' method. Following these initial developments, several methods that produce unbiased and efficient estimates of the missing observations have appeared on the forefront. These methods include the well-known maximum likelihood-based expectation maximization (EM) [9] and multiple imputation methods, see [29,22,27] for a detailed exposition. In another work, Folch et al. [15] developed

* Corresponding author.
E-mail address: arunkt@iitm.ac.in (A.K. Tangirala).

http://dx.doi.org/10.1016/j.jprocont.2016.08.008
0959-1524/

a MATLAB toolbox for missing data estimation based on principal component analysis. Despite their successes, the aforementioned methods suffer from one or a few serious shortcomings, namely, obtaining inefficient and unbiased estimates, supplying initial guess for estimates and very limited applicability to the small size of available observations. Yang et al. [43] developed a method to estimate missing data based on matrix completion. Although this method yields unbiased estimates for lower percentage of missing data, it cannot handle high percentage of missing data.

In this work, we propose to develop a novel missing data reconstruction method based on CS ideas, that addresses most of the above mentioned shortcomings. The origins of CS-based signal recovery (CSSR) methods are rooted in the problem of recovering a long signal $\{s[\cdot]\}$ of length N from a fewer set of $M \ll N$ measurements $\{y[\cdot]\}$ [10]. These measurements could be in the domain of the signal or in a different basis – for example, measurements of a time-domain signal in the frequency domain or in the wavelet domain. There are usually two key assumptions involved in the traditional formulation of CSSR methods: (i) the measurement and signal are linearly related and (ii) there exists a "dictionary" in which the long unknown signal **s** is *sparse*, i.e., it has a signal representation with very few K non-zero coefficients in that dictionary. Mathematically, a dictionary is a generalization of basis to include linearly dependent functions, which is also the idea in frame theory.

Recovering signals from limited and irregularly spaced data is an old problem; however, the CS-based formulation in [10,6,38] is novel because it aims to reconstruct the signal by first reconstructing its representation in a *sparsifying dictionary*. This approach is in stark deviation from the traditional reconstruction formulations, which solely focus on reconstructing the signal in the domain of observation.

A characteristic feature of CSSR problems is that they are *underdetermined* problems, i.e., situations where the number of unknowns (typically N) is higher than the knowns, which are the M measurements. CS methods work around this hurdle by casting the underdetermined problem in the raw domain as an *overdetermined* problem in the sparsifying dictionary, providing $M > K$. This is the key idea underlying the success of these methods. Fig. 1 schematically illustrates this idea. It is typically assumed that the sparsifying dictionary, represented by the matrix **B**, and hence the overcomplete dictionary **A**, is known. In this case, various algorithms such as basis pursuit (BP), orthogonal matching pursuit (OMP), etc. can be used to reconstruct sparse representation of the signal from available data, see [13,30] for a detailed discussion of these algorithms. However, in many applications, it is unrealistic to assume that the appropriate sparsifying dictionary is known a priori.

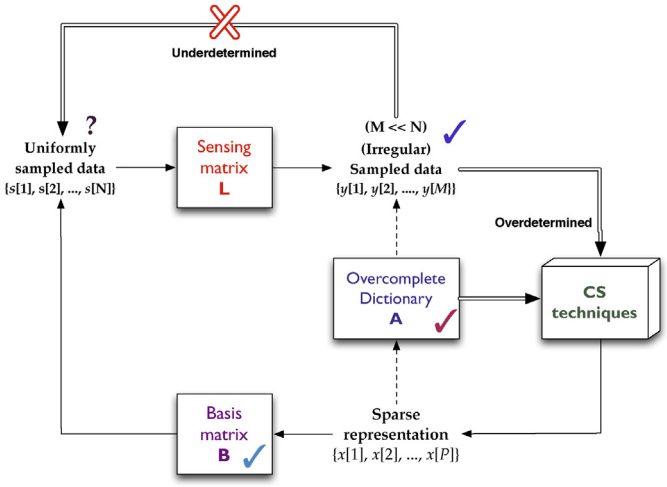

Fig. 1. Overview of compressed sensing.

Given this situation, an approach that is widely followed is to pre-select or fix the dictionary based on some mathematical considerations. For instance, Lustig et al. [23] demonstrated a faster way of reconstructing magnetic resonance images than the traditional methods using CS techniques assuming these images are sparse in Fourier basis. In another application of missing data reconstruction, Stankovic et al. [35] performed statistical analysis for efficient detection of signal components from irregularly sampled data by assuming the sparsifying dictionary to be the Fourier basis. However, in reality, a large class of signals are different from a pure mixture of sinusoids. More importantly, as it is known in several signal processing applications, a fixed dictionary is not necessarily and usually the most appropriate dictionary for a given process. The reason is that, a fixed dictionary, while being mathematically suitable, does not necessarily conform to the physics of the process. Furthermore, in majority of the cases, the sparsifying dictionary is not known a priori. Therefore, it becomes necessary to determine the sparsifying dictionary from data. It must be remarked that in this pursuit, one rarely discovers a closed-form expression for the dictionary. Consequently, efforts are usually directed towards directly determining the dictionary matrix **B** or the overcomplete dictionary **A**, assuming that the sensing matrix **L** is known. The resulting dictionary is then said to be *adaptive*, i.e., derived from data. A strikingly similar scenario exists in the class of *source separation* problems, where it is required to recover "source signatures" from mixture measurements [1]. The CS problem is analogous to the case of underdetermined source separation problem, where one has fewer mixtures than sources, but it is known that only a few (sparse) sources that actually participate in the mixture generation. The fixed dictionary scenario corresponds to known source signature problem, whereas the adaptive dictionary corresponds to discovering the source signatures from the data as well. Wang et al. [40] used fixed dictionary to estimate the missing observations in a wireless sensor network using the ideas of CS. Spoorthy et al. [33] also used the fixed dictionary to estimate missing data in sales gathering using CS techniques. The limitation of both the methods is that they use fixed dictionary instead of adaptive dictionary to estimate missing data. The advantages of using adaptive dictionary over fixed dictionary are explained in the subsequent paragraphs.

Adaptive dictionary estimation methods not only aid in extracting the dictionary that is appropriate to a given application, but also offer the provision of imposing the amount of sparsity demanded by a particular application. The adaptive dictionary estimation problem has received appreciable attention, in the general literature [2,14,25,28,44] and in image processing and source separation applications [26,1]. In general, these methods implement two-step approach that alternate between optimizing the sparse representation in a given dictionary and finding an optimal sparsifying dictionary for a given representation. The method of optimal directions or dictionaries (MOD) algorithm, developed by Engan et al. [14] and the K-SVD due to Aharon et al. [2] suffer from shortcomings in that they require a large number of training sets of **y** to estimate sparsifying dictionary, which is not the case of missing data, whereas a more realistic situation is that only a single set of **y** is available. In another work, Duarte-Carvajalino and Sapiro [11] optimized both measurement matrix **L** and dictionary **B** simultaneously such that the both **L** and **B** are incoherent. In the case of present work, the method cannot be used since the sensing matrix **L** is fixed. The work of Peyré et al. [26] is especially devised for morphology component analysis of images and works with the variational energy principle.

The main contribution of this work is towards developing a method to reconstruct missing data using CS techniques with an adaptively sparsifying dictionary, a method for which is also proposed. The sensing matrix is essentially assumed to consist of ones