Signal Processing 106 (2015) 73–87

Contents lists available at ScienceDirect

Signal Processing

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

Wavelet shrinkage using adaptive structured sparsity constraints

Diego Tomassi [a,c,d,*], Diego Milone [d], James D.B. Nelson [b]

[a] Instituto de Matemática Aplicada del Litoral, UNL–CONICET, Argentina
[b] Department of Statistical Science, University College London, United Kingdom
[c] Departamento de Matemática, Facultad de Ingeniería Química, Universidad Nacional del Litoral, Argentina
[d] Centro de Investigación en Señales, Sistemas e Inteligencia Computacional, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Argentina

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form
29 June 2014
Accepted 1 July 2014
Available online 10 July 2014

Keywords:
Structured sparsity
Regularised regression
Denoising
Dual-tree complex wavelet transform

ABSTRACT

Structured sparsity approaches have recently received much attention in the statistics, machine learning, and signal processing communities. A common strategy is to exploit or assume prior information about structural dependencies inherent in the data; the solution is encouraged to behave as such by the inclusion of an appropriate regularisation term which enforces structured sparsity constraints over sub-groups of data. An important variant of this idea considers the tree-like dependency structures often apparent in wavelet decompositions. However, both the constituent groups and their associated weights in the regularisation term are typically defined a priori. We here introduce an adaptive wavelet denoising framework whereby a sparsity-inducing regulariser is modified based on information extracted from the signal itself. In particular, we use the same wavelet decomposition to detect the location of salient features in the signal, such as jumps or sharp bumps. Given these locations, the weights in the regulariser associated to the groups of coefficients that cover these time locations are modified in order to favour retention of those coefficients. Denoising experiments show that, not only does the adaptive method preserve the salient features better than the non-adaptive constraints, but it also delivers significantly better shrinkage over the signal as a whole.

1. Introduction

A key attraction of wavelets is their compressive representation of data. This is fundamental to powerful nonlinear processing methods such as wavelet shrinkage [1–3]. Early approaches often regarded wavelet coefficients as statistically independent. Further developments, however, showed that for many applications involving real-world signals and images, performance improved when the dependencies between coefficients were taken into account [4–9]. Most of such methods typically focussed on the persistency property which is often apparent across wavelet scales. The simplest models account for such statistical dependencies between parent coefficients at a given level of the decomposition and their child coefficients at the following level of finer resolution. Although methods based on these models proved successful in many applications such as denoising, compression, and classification, some concerns remained about the preservation of salient features in the signal, such as jumps or sharp bumps [10]. In applications such as denoising or

* Corresponding author at: Instituto de Matemática Aplicada del Litoral, UNL–CONICET, Colectora RN168 km. 472, Paraje El Pozo, 3000 Santa Fe, Argentina. Tel.: +54 342 4511370.
E-mail address: diegot@santafe-conicet.gov.ar (D. Tomassi).

http://dx.doi.org/10.1016/j.sigpro.2014.07.001
0165-1684/

deconvolution these features are typically over-smoothed which compromises the quality of the estimates. Some attempts to improve performance under these conditions explore total variation filtering [11,12], combined Tikhonov and total variation regularisation [10] and decompositions based on footprints of the discontinuities in the signal [13].

In this work we take advantage of the latest developments in regularised least-squares regression to promote tree-structured sparsity on the denoised estimates. Unlike previous tree-structured estimators [5,8,14–16], the method proposed here uses a lasso-like algorithm with a mixed-norm regulariser that induces structured sparsity over an overcomplete representation. A novel signal-driven approach is introduced to adapt the weights of the regulariser. The ability of shift-invariant complex wavelet transforms to detect salient features in the signal is exploited to design a penalisation term which favours estimated jumps or sharp bumps during the optimisation process. We show that this results in a denoising approach with better preservation of salient features.

The manuscript is organised as follows. In the remainder of the current section we provide motivation and discuss the specific contributions of our work in the context of the current literature. In Section 2 we offer an overview of structured sparsity approaches and the dual-tree complex wavelet transform. The proposed method is introduced in Section 3. This considers both an oracle and a practical approach to account for the occurrence of salient features. In Section 4, results obtained in denoising experiments show the advantage of the proposed adaptive scheme over structured sparsity estimates set a priori. We then close with the main conclusions and a discussion of further work.

1.1. Motivation

Sparse representations have been at the core of many signal processing methods in recent years [17,18]. Early algorithms such as basis pursuit [19] and matching pursuit [20] regarded coefficients as mutually independent, meaning that each atom in the decomposition is selected or discarded independently of its neighbours. In the signal processing community, efforts to introduce structured sparsity constraints were spurred by the compressed sensing paradigm [21,22] which used prior knowledge to reconstruct signals with fewer samples than classical sampling theorems allowed. Model-based compressed sensing has showed promise in this context [23–25]. These early attempts, however, were based on non-convex or greedy optimisation approaches. To achieve scalability without compromising consistency, non-greedy convex approaches are often desirable. To this end, regularised approaches using mixed-norms have proven successful in obtaining sparse estimates that retain an assumed dependence structure [26,27].

It is important to note that most of the existing wavelet/structured models deal with the persistency property of the coefficients without taking into account any additional information provided by the specific choice of transformation or dictionary used to obtain the representation [27–29]. Moreover, all of these dependence structures are set a priori, and no further information from the signal is used to adapt them. In denoising applications, features with strong local high frequency content are often over-smoothed by such methods. This is due to the erroneous shrinkage or elimination of coefficients at finer scale levels. On the other hand, when regularisation parameters are set to favour data-fitting much more than sparsity, the resulting estimates often retain too many fine-scale coefficients and remain noisy.

1.2. Contribution

In this work, a new signal processing method is developed that uses additional information, extracted directly from the signal, to reinforce the a priori structured sparsity constraints. To do so, we use the dual-tree complex wavelet transform (DTCWT) as the sparsity inducing transform together with a hierarchical mixed norm regulariser. The weights in the regulariser are adaptively modified in order to help preserve salient features of the analysed signal. This adaptive modification is driven by a detection stage which aggregates information from the different scales of the wavelet decomposition to infer the locations of salient features in the signal. In this way, the mixed norm regulariser, defined a priori, is tailored to the observed signal.

1.3. Related work

Tree-structured estimators have been proposed earlier for wavelet decompositions, both in the signal processing and statistics communities [5,8,14–16]. They often rely on orthonormal transforms and hard-thresholding approaches. Following their success in machine learning and statistics, generalised lasso-type algorithms have received recent and growing attention for signal processing applications. The closest works are [29,30]. In [29], the parent–child dependence of wavelet coefficients is coded into overlapping groups, each of which comprises a parent–child pair. A variable replication approach is taken into account for different instances of a given coefficient appearing in different groups and a regularisation term is added to account for the dissimilarity of the replicates of the same variable. Unlike the present work, their approach uses the standard discrete wavelet transform (DWT) without adding any additional information onto the structure assumed a priori. In [30], a chain structure is assumed to model the spectrogram of audio signals obtained from their short-time Fourier transform representation. This simple structure gives rise to a regularisation term that is bounded above by a quantity which is simpler to compute, allowing for an efficient minimisation–majorisation algorithm. It should be noted, however, that it is suited for signals with emphasised band structures in their spectrogram. On the other hand, edge information has been used to aid image denoising [31] and reconstruction under compressed sensing applications [32]. To the best of our knowledge, however, such information has not been used to adapt a structured regulariser as proposed here.