ADAPTIVE DENOISING IN ORTHONORMAL LIBRARIES

S. E. FERRANDO, L. A. KOLASA, AND N. KOVAĆEVIČ

Abstract. This note is a brief introduction to the subject of denoising in orthonormal libraries. This algorithm is implemented in the file `wave++/demos/demoWavPack.cc`. Donoho and Johnstone ([2]) introduced an adaptive algorithm that extends nonlinear thresholding denoising in a fixed orthonormal basis to a multiple bases setting. In that paper a search for an optimal basis from a large collection of tree-structured orthonormal bases (i.e. a *library*) was proposed. That technique gives the, so-called, best ortho-basis estimate.

1. Introduction

Suppose we have noisy data

$$y_i = f_i + \sigma . z_i \quad i = 1, \dots , N, \tag{1}$$

where $f = (f_i)$ is a given signal and $z = (z_i)$ are samples from i.i.d. random variables with $\mathcal{N}(0, 1)$ distributions. Thresholding schemes on *fixed* orthonormal basis are well known estimators for the underlying signal. It is well known that the quality of such denoising is related to how well is the signal compressed in the orthonormal basis. Reference [2], building on ideas from [1], proposes an algorithm for *adaptive* denoising, namely an orthonormal basis is chosen from a library of such bases which satisfies certain optimality conditions for denoising. This scheme will be called *Best Ortho-Basis denoising* and reviewed in the next section. The purpose of this note is to review, for the benefits of the user of $wave++$, the necessary background to understand the algorithm implemented in `wave++/demos/demoWavPack.cc`.

2. Best Orthonormal Basis for Denoising

Here we summarize the main result from [2] as presented in [3]. Results for a single basis and for libraries will be presented simultaneously and contrasted. Suppose we have available a library $\mathcal{L}$ of orthonormal bases, such as the Wavelet Packet bases or the Cosine Packet bases of Coifman and Meyer. Let $\mathcal{B} \in \mathcal{L}$ and $\theta(x, \mathcal{B})$ denote the vector of coefficients of a vector x in the basis $\mathcal{B}$. Consider the family $\hat{\Phi}_{\mathcal{B}}$ of estimators defined by keeping or killing empirical coefficients in some basis $\mathcal{B} \in \mathcal{L}$. Such estimators $\hat{f}(y; w, \mathcal{B})$ in the coefficient domain are of the form

$$\theta_i(\hat{f}, \mathcal{B}) = w_i . \theta_i(y, \mathcal{B}). \tag{2}$$

where each weight w_i is either 0 or 1. Formally, the set of estimators associated to $\mathcal{B}$ are $\hat{\Phi}_{\mathcal{B}} = \{\hat{f}(., w, \mathcal{B}) : w \in \{0, 1\}^N\}$ and the ones associated to $\mathcal{L}$ are $\hat{\Phi}_{\mathcal{L}} =$

Research partially supported by NSERC.

$\bigcup_{\mathcal{B}\in\mathcal{L}} \hat{\Phi}_{\mathcal{B}} = \{\hat{f}(.,w,\mathcal{B}) : w \in \{0,1\}^N, \mathcal{B} \in \mathcal{L}\}$. Given an estimator $\hat{f}$ of f, quality of estimation is mesured in terms of its risk

$$R\left(\hat{f}(w,\mathcal{B}), f\right) = \mathbf{E}\left(||\hat{f}(y,w,\mathcal{B}) - f||^2\right). \tag{3}$$

Define the ideal risk for the case of a library by

$$\mathcal{R}_{\mathcal{L}}(f) = \inf_{\hat{f}\in\hat{\Phi}_{\mathcal{L}}} R\left(\hat{f}(w,\mathcal{B}), f\right) \tag{4}$$

In the single basis setting, $\hat{\Phi}_{\mathcal{L}}$ should be replaced by $\hat{\Phi}_{\mathcal{B}}$ in (4). Notice that in order to attain this ideal risk, knowledge of the vector f is required. The estimate attaining this ideal risk is therefore not an empirical estimate but an *oracle* one ([2]). Associated to the ideal risk is the ideal basis $\mathcal{B}_{\mathcal{L}} \in \mathcal{L}$ defined by

$$\mathcal{B}_{\mathcal{L}} = \arg \inf_{\hat{f}\in\hat{\Phi}_{\mathcal{L}}} R\left(\hat{f}(w,\mathcal{B}), f\right) = \arg \inf_{\hat{\Phi}_{\mathcal{B}}\subset\hat{\Phi}_{\mathcal{L}}} \mathcal{R}_{\mathcal{B}}(f) \tag{5}$$

In order to introduce empirical estimates, define, for $\lambda > 0$ given, the vector $\delta(y,\mathcal{B},\lambda) = (\delta_i(y,\mathcal{B},\lambda)$ by

$$\delta_i(y,\mathcal{B},\lambda) = 1_{\{|\theta(y,\mathcal{B}| > \sigma.\sqrt{\lambda}\}}\ \mathrm{sign}(y) \tag{6}$$

Set $\lambda = \lambda_{\mathcal{B}} = 2\log(N)$. The empirical estimate $\hat{f}_{\mathcal{B}}(y;\delta(\mathcal{B},\lambda_{\mathcal{B}}))$ relative to $\mathcal{B}$ is given, in the coefficient domain, by:

$$\theta_i(\hat{f}_{\mathcal{B}},\mathcal{B}) = \delta_i(y,\mathcal{B},\lambda_{\mathcal{B}}).\theta_i(y,\mathcal{B}). \tag{7}$$

The following oracle inequality holds for all f and $N > 4$

Theorem 1. *If* $\hat{f}_{\mathcal{B}} = \hat{f}_{\mathcal{B}}(y;\delta(\mathcal{B},\lambda_{\mathcal{B}}))$

$$R(\hat{f}^*, f) \leq 2\log(N).\left(\sigma^2 + \mathcal{R}_{\mathcal{B}}(f)\right). \tag{8}$$

Equation (8) applied to $\mathcal{B} = \mathcal{B}_{\mathcal{L}}$ suggests that this ideal basis will deliver a better estimate that in any other basis $\mathcal{B} \in \mathcal{L}$. The main point of [2] is to give an algorithm to select a basis $\hat{\mathcal{B}}_{\mathcal{L}} \in \mathcal{L}$ such that behaves similarly to $\mathcal{B}_{\mathcal{L}}$ with respect to oracle inequalities. In order to describe these results, set $M_{\mathcal{L}}$ equal to the number of distinct vectors occuring among all bases in $\mathcal{L}$ and $t_{\mathcal{L}} = \sqrt{2log_e(M_{\mathcal{L}})}$. Choose $\xi > 8$ and set the threshold parameter to be $\lambda = \lambda_{\mathcal{L}} = (\xi.(1 + t_{\mathcal{L}}))^2$. Define now the entropy functional

$$\mathcal{E}_{\lambda}(y,\mathcal{B}) = \sum_{i=1}^{N} \min\left(\theta_i^2(y,\mathcal{B}), \sigma^2\,\lambda_{\mathcal{L}}\right). \tag{9}$$

Let $\hat{\mathcal{B}}_{\mathcal{L}}$ be the best (empirical) orthonormal basis relative to this entropy:

$$\hat{\mathcal{B}}_{\mathcal{L}} = \arg\min_{\mathcal{B}\in\mathcal{L}} \mathcal{E}_{\lambda_{\mathcal{L}}}(y,\mathcal{B}). \tag{10}$$

The risk of the empirical estimate satisfies the following oracle inequality for all f

Theorem 2. *If* $\hat{f}_{\hat{\mathcal{B}}_{\mathcal{L}}} = \hat{f}_{\hat{\mathcal{B}}_{\mathcal{L}}}\left(y;\delta(\hat{\mathcal{B}}_{\mathcal{L}},\lambda_{\mathcal{L}})\right)$ *then*

$$R(\hat{f}_{\hat{\mathcal{B}}_{\mathcal{L}}}, f) \leq A(\xi).\lambda_{\mathcal{L}}.\left(\sigma^2 + \mathcal{R}_{\mathcal{L}}(f)\right) \tag{11}$$

where $A(\xi) = 6.(1 - 8/\xi)^{-1}$.

Remark 1. *In* [2] *a similar result is proved that holds with high probability, this is in contrast to the above result that holds in the mean.*

An algorithm follows from the above results. We briefly describe it here.

Algorithm

• Compute

$$\hat{\mathcal{B}}_{\mathcal{L}} = \arg\min_{\mathcal{B}\in\mathcal{L}} \mathcal{E}_{\lambda_{\mathcal{L}}}(y, \mathcal{B}). \tag{12}$$

This computation can be done by means of the Best Wavelet Basis algorithm given that the above minimization involves an additive functional. To simplify the notation set $\hat{\mathcal{B}} = \hat{\mathcal{B}}_{\mathcal{L}}$.

• Compute

$$\theta_i(\hat{f}_{\hat{\mathcal{B}}}, \hat{\mathcal{B}}) = \delta_i(y, \hat{\mathcal{B}}, \lambda_{\mathcal{L}}).\theta_i(y, \hat{\mathcal{B}}). \tag{13}$$

• Finally apply the inverse transform (associated to the basis $\hat{\mathcal{B}}$) to the coefficients given by (13) to get

$$\hat{f}_{\mathcal{B}}(y; \delta(\mathcal{B}, \lambda_{\mathcal{B}})).$$

2.1. **Practical Considerations.** The parameters involved in Theorem 2 are not substantially larger than those appearing in Theorem 1. In practical situations though, the discrepancy in the values of λ may cause the best ortho basis estimate to be poorer than single basis estimates. The following remarks are intended to remedy this problem. The threshold parameter $\lambda_{\mathcal{L}}$ plays a double role in the constructions described above. First, it is used to select $\hat{\mathcal{B}}$ and then it is used to threshold $\theta_i(y, \hat{\mathcal{B}})$. In this second application, $\lambda_{\mathcal{L}}$ is too large for the method to be competitive with thresholding in a single basis. In [2] it is mentioned that the parameter $\xi > 8$ could be made smaller. We have found that a good performance for the best wavelet basis denoising algorithm to perform well is to use $\xi = 1$.

References

[1] R.R. Coifman, M.V. Wickerhauser, *Entropy-based algorithms for best basis selection.* IEEE transactions on Information Theory, Vol. **38**, N 2, March 1992.

[2] D. L. Donoho, I.M. Johnstone, *Ideal denoising in an orthonormal basis chosen from a library of bases.* Comptes Rendus Aca. Sci. Paris A , **319**, 1327-1322, (1994).

[3] D. L. Donoho, *CART and Best-OrthoBasis: a connection.* Preprint.

Department of Mathematics, Physics and Computer Science, Ryerson Polytechnic University, Toronto, Ontario M5B 2K3, Canada.

E-mail address: ferrando@acs.ryerson.ca

E-mail address: lkolasa@acs.ryerson.ca

E-mail address: natasak@home.com