Learning best wavelet packet bases for compressed sensing of classes of images: application to brain MR imaging

Michal P. Romaniuk,[1] Anil W. Rao,[1] Robin Wolz,[1] Joseph V. Hajnal[2] and Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Department of Computing, Imperial College London, United Kingdom [2]Division of Imaging Sciences and Biomedical Engineering, King's College London, United Kingdom

Abstract We study the feasibility of learning an adapted wavelet packet basis for a class of signals and subsequently using this basis for reconstruction of unseen signals of the same class. In particular, we learn an adapted basis for a set of brain MR images and then use this basis as a sparsifying transform for image reconstruction from undersampled MR data. Simulation results show that there is a significant improvement in sparse approximation accuracy when comparing an adapted wavelet packet basis to a wavelet basis, and a modest improvement in image reconstruction from undersampled k-space data.

1 Introduction

Compressed sensing (CS) provides a mathematical framework [4,8] for reconstruction of signals sampled at sub-Nyquist rates, provided that those signals can be represented sparsely using either an orthogonal transform, or a dictionary of signals [3]. CS led to recent advances in medical imaging, and in particular magnetic resonance imaging (MRI), starting with the work of Lustig *et al.* [11,10]. While in most imaging situations wavelets provide accurate sparse approximations of signals, recently there has been much interest in finding representations adapted to particular signals. Examples include patch-based dictionaries [1] successful in CS MRI [17] as well as dictionaries adapted specifically for compressed sensing [9].

Similarly, the adaptive framework of wavelet packets [6,7] has also been explored for CS [14,15]. While patch-based dictionaries and wavelet packets are similar in that both approaches adapt a set of atoms to efficiently represent training signals, they are also very different in other respects. In particular, wavelet packets provide a basis for the whole image, rather than small regions. They also naturally have a multiscale structure and representation coefficients can be computed efficiently using standard wavelet algorithms.

The work in [14,15] explores finding the best basis while reconstructing the undersampled signal, *i.e.* without learning from prior examples. In contrast, our algorithm learns the best basis from a collection of example images. Our method is based on well known principles and algorithms for wavelet packets and approximation in bases [6,7,16,18,12]. The main contribution of our work consists of designing a best basis search cost function that emphasises the criteria that are important in compressed sensing. We also choose a suitable algorithm for optimising this cost function and validate our approach on images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [13] database.

2 Compressed sensing

A general model of a compressed sensing system is of the form

$$y = Ax + z \tag{1}$$

where x is the sparse representation vector of the signal of interest, A is the information operator that transforms signals from the sparse representations to the actual measurements, and z represents random noise inherent in instrumentation. The essence of compressed sensing is that A has less rows than columns, *i.e.* the number of measurements in y is less than the number of components of the sparse representation vector x. In compressed sensing MRI, typically $A = RFD$, where D is a basis (or dictionary) that is used to sparsely represent the image, F is a multidimensional Fourier transform that represents the MRI acquisition process, and R is a matrix that selects a subset of measurements. The estimate $\hat{x}$ of x is computed by solving the problem [2,3]

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_1 \quad \text{subject to} \quad \|A\tilde{x} - y\|_2^2 \leq \varepsilon^2 \tag{2}$$

where ε^2 is an upper limit on the noise energy $\|z\|_2^2$. If A satisfies the *Restricted Isometry Property* [5] with an appropriate isometry constant, then the solution $\hat{x}$ to the problem (2) obeys [2,3]

$$\|\hat{x} - x\|_2 \leq C_0 \cdot \frac{\|x - x_s\|_1}{\sqrt{s}} + C_1\varepsilon \tag{3}$$

for some constants C_0 and C_1, where x_s is the best s-sparse approximation of x (an approximation with s non-zero components).

One of the principal design goals in compressed sensing is to find a dictionary that leads to the sparsest representations possible, *i.e.* reducing s. At the same time, eq. (3) implies that the modelling error induced by neglecting small coefficients will degrade the quality of reconstruction.

3 Wavelet packets

Wavelet packets [6,7] are an extension of wavelets. In a wavelet framework, a single-level discrete wavelet transform consists of filtering the source signal with a dual-branch orthogonal filter bank, followed by downsampling of the resultant sequences. A multi-resolution analysis (wavelet decomposition) then consists of recursively applying this process on the lowpass branch of the filter bank, producing a tree-like structure. With wavelet packets, the filter bank can be applied on the highpass branches as well, which means that a variety of tree-like filter structures can be generated. All possible choices of which branches to decompose further give rise to a large set of admissible trees, or in other words a library of bases [12]. The leaves of each admissible tree define a complete set of basis vectors.

With careful choice of basis from a wavelet packet library, one might be able to find sparser approximations than with the standard wavelet basis. The Coifman-Wickerhauser (CW) basis selection algorithm [7] can efficiently find an optimal tree, in the sense that the associated basis minimises a cost function with respect to the resultant representation vector, provided that the cost function satisfies the criterion of an *additive information cost function*:

Definition 1. [7] *A map $\mathcal{M}$ from sequences $\{x_i\}$ to $\mathbb{R}$ is called an* additive information cost function *if $\mathcal{M}(0) = 0$ and $\mathcal{M}(\{x_i\}) = \sum_i \mathcal{M}(x_i)$.*

When a basis defined by an orthogonal matrix $B \in \mathbb{R}^{n\times n}$ (with atoms as columns) is used to represent a signal $x \in \mathbb{R}^n$, the associated information cost is $\mathcal{M}(B^\intercal x)$, *i.e.* the information cost function is applied to the vector of transform coefficients.

4 Best wavelet packet bases and compressed sensing

4.1 Single signal case

Our analysis in this section is based on [12] (pp. 450-452, 611-614), but we have to adapt it for approximation error measured in ℓ_1 norm. Given a vector (signal) $y \in \mathbb{R}^n$, we express basis search as minimising the following Lagrangian with respect to the wavelet packet basis B_{wp} and the vector of approximation coefficients $x_\Lambda \in \mathbb{R}^n$.

$$\mathcal{L}(y, B_{wp}, T, x_\Lambda) = \left\|B_{wp}^\intercal y - x_\Lambda\right\|_1 + T\left\|x_\Lambda\right\|_0 \tag{4}$$

where T is the Lagrangian multiplier. $\|.\|_0$ denotes the ℓ_0 pseudo-norm and $A^\intercal$ denotes the transpose of matrix A (we consider real-valued images only). We assume a predefined filter bank for the wavelet packet decomposition. Our choice to minimise the transform domain ℓ_1 error is based on the conclusions of our discussion of compressed sensing: based on eq. (3) we reason that the compressed sensing reconstruction error should depend on transform domain ℓ_1 approximation error (among other factors). Note that, since we are working with orthonormal bases, approximations generated by hard-thresholding transform coefficients are always optimal, in the sense that, as long as the basis is not changed, improving sparsity would require reducing accuracy and improving accuracy would require less sparse approximation. It follows by looking at eq. (4) that T is also the threshold at which the entries of $x = B^\intercal y$ are set to zero in

x_Λ (see [12] pp. 612-613 for the precise argument with error defined in ℓ_2 norm). Therefore, we can write eq. (4) in alternative form

$$\mathcal{L}(y, B_{wp}, T) = \sum_i \inf(|x[i]|, T) \quad \text{where} \quad x = B^\intercal y \tag{5}$$

where $x[i]$ denotes the i-th entry of the vector x.

If T is fixed, the cost function in eq. (5) is a valid additive information cost function, so the CW algorithm can be used for best basis search. Note also that the form of $\mathcal{L}$ in eq. (5) is independent of x_Λ. This is a result of the fact that B_{wp} is an orthogonal basis: by fixing T and selecting B_{wp}, we implicitly assign $x_\Lambda = \rho_T\left(B_{wp}^\intercal y\right)$ where $\rho_T(a)$ is an operator that sets to zero all entries of a that are smaller than T.

So far we can find optimal approximations for any given coefficient threshold T but this threshold is difficult to choose. It would be more practical to specify either the number of coefficients or the approximation error. A closely related problem of finding best wavelet packet bases in rate-distortion sense for source coding applications was considered in [16]. In the following we take a similar approach to our problem.

Consider two solutions to minimising $\mathcal{L}$, with respective thresholds $T_1 < T_2$. The basis for $T = T_1$ will favour accuracy at the cost of sparsity, while the trade-off will move in the opposite direction for $T = T_2$. Since both bases are optimal at their respective thresholds, we have $\|B_1^\intercal y - \rho_{T_1}(B_1^\intercal y)\|_1 \leq \|B_2^\intercal y - \rho_{T_2}(B_2^\intercal y)\|_1$ and $\|\rho_{T_1}(B_1^\intercal y)\|_0 \geq \|\rho_{T_2}(B_2^\intercal y)\|_0$. So if for any T the approximation is not sparse enough, we can increase T and repeat basis search to trade some accuracy for sparsity. Similarly, if the approximation is not accurate enough, we can trade some sparsity for accuracy by reducing T. Therefore, the optimal threshold for a required level of sparsity or approximation error can be found by bisection search.

4.2 Extension to multiple signals

For our application the signal to be sparsified is not fully known: we only have partial Fourier data. We can, however, search for a basis that minimises the expected information cost over a set of reference images. In this context, we choose the cost function as the mean over the whole training data set, *i.e.*

$$\bar{\mathcal{L}}\left(\{y_i\}_{i=1,\ldots,N}, B_{wp}, T\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\left\|B_{wp}^\intercal y_i - \rho_T\left(B_{wp}^\intercal y_i\right)\right\|_1 + T\left\|\rho_T\left(B_{wp}^\intercal y_i\right)\right\|_0\right) \tag{6}$$

where y_i are the training examples and N is the number of signals in the training set. Since $\rho_T\left(B_{wp}^\intercal y_i\right) = (x_\Lambda)_i$, this is the mean of (4) over all training examples. This cost function can be optimised by decomposing each signal in a full wavelet packet tree, constructing a joint tree [18] where each node is assigned a cost by taking the average of the same node over all individual signal trees, and then applying the CW algorithm to the joint tree to find an optimal basis [18].

Note that the same threshold T is used for wavelet packet decompositions of all images. It is therefore necessary to ensure that images are pre-processed so that one value of T is suitable for all of them. While designing an optimal pre-processing step is beyond the scope of this work, we decided to follow a simple approach of scaling image intensities so that the 99^{th} percentile of image histograms matches across the whole dataset.

4.3 Compressed sensing reconstruction

The proposed algorithm finds a sparsifying basis which can be used with standard compressed sensing reconstruction algorithms. In this work, we have integrated it with the framework of [10]. The main difference is replacing the wavelet transform with the optimised wavelet packet tree transform.

5 Experiments

Experiments were conducted using 826 MR images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`http://www.loni.ucla.edu/ADNI`). The images were resliced to 1mm voxel size using the tools from J. Shen's NIfTI toolbox, version 2011-09-21 (`http://www.rotman-baycrest.on.ca/~jimmy/NIfTI/`), and then either cropped or zero-padded to $256\times256\times256$ size. Image intensities were then scaled to make the 99^{th} percentile of all image histograms match. Our study was done on 2D slices of images pre-processed in this way (one slice from each image), but our method can be easily extended to 3D images. Computations were done using MATLAB R2010b (The MathWorks, Natick, MA).

5.1 Approximation of brain MR images

The aim of this experiment was to measure the trade-off between sparsity and accuracy for an adapted wavelet packet basis and compare it with a wavelet basis. We followed a standard ten-fold cross-validation approach. For each fold we trained a basis to best approximate the out-of-fold data with 10% of coefficients, and used the fold data for validation. We compared the results with approximations generated with a wavelet basis. Both wavelet and wavelet packet decompositions were done to four levels with the Daubechies 4 filter bank. To produce an empirical estimate of average approximation accuracy over one fold, we computed the mean and sample standard deviation over the means extracted from individual folds. Results are presented in Fig. 1.

These results show that an adapted wavelet packet basis offers consistent improvement over standard wavelets. Although the basis was optimised for a 10% sparsity target, it performs better over a wide range of sparsities. Interestingly, in the ten-fold cross-validation process the same basis was learned each time. This indicates that our approach is quite stable and robust.

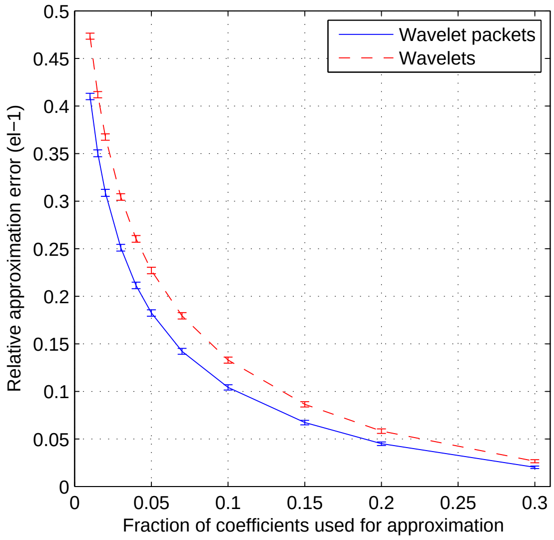

Figure 1. Approximation of brain MR images with wavelets and wavelet packets. The basis was trained for optimal representation with 10% of coefficients.

5.2 Compressed sensing reconstruction of brain MR images

We reconstructed images from fold 1 using the basis learned from folds 2–10. Undersampling mask generation and CS reconstruction were done using SparseMRI V0.2 (`http://www.stanford.edu/~mlustig/SparseMRI.html`). We made some changes to the original software: we modified the k-space mask generation code to enable lower sampling densities and we removed the k-space density compensation step in reconstruction. The method of [10] uses a combination of wavelet domain sparsity and total variation (TV) to regularise the reconstruction. We left the weights of transform domain sparsity and TV unchanged (0.05 and 0.02 respectively), for both wavelets and wavelet packets.

Table 1. Compressed sensing reconstruction PSNR with wavelets (W) and wavelet packets (P). The figures are means over the 83 images in fold 1. m/n is the ratio of the number of k-space samples to the number of pixels in the image. $s_n^\star$ is the target sparsity optimised for in the training process. All figures (except m/n) are in decibels.

m/n	W	P,$s_n^\star = 0.1$	P,$s_n^\star = 0.05$	P, $s_n^\star = 0.02$	P, $s_n^\star = 0.01$	P, $s_n^\star = 0.005$
0.4	34.0999	**34.3187**	34.0599	33.8154	33.7118	33.7118
0.35	31.4394	**31.5990**	31.3284	31.1503	31.0973	31.0973
0.3	27.4338	**27.5222**	27.4048	27.2641	27.2068	27.2068
0.25	24.5557	**24.6713**	24.5758	24.4548	24.4156	24.4156
0.2	23.7575	**23.8077**	23.7507	23.6528	23.6355	23.6355

An example image with its wavelet and wavelet packet reconstructions is displayed in Fig. 2. Table 1 presents the peak signal–to–noise ratio (PSNR) over a range of configurations.

These results show that substituting the wavelet transform with a pre-trained wavelet packet transform in compressed sensing reconstruction with sparsity and total variation regularisation leads to a modest improvement in PSNR.

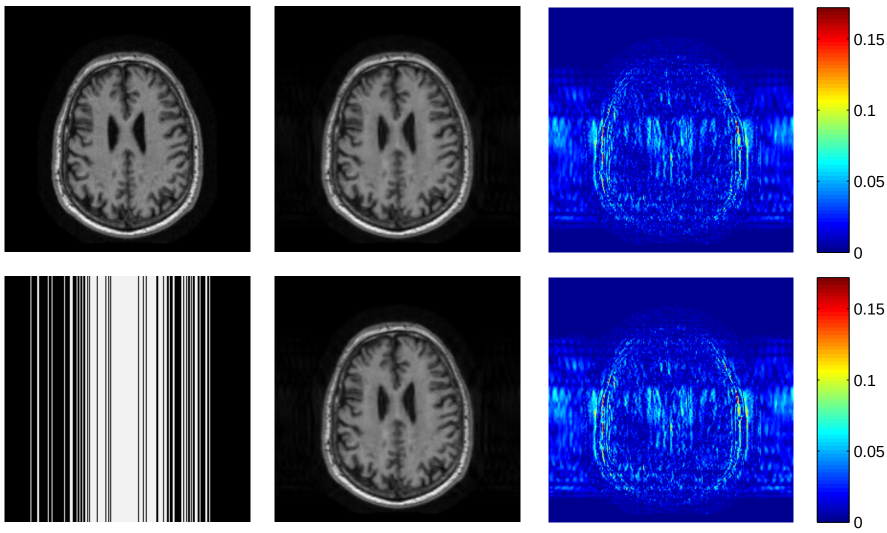

Figure 2. Example reconstructions with wavelets and wavelet packets. Top row (left to right): original image, wavelet reconstruction from 40% k-space sampling and the associated error map. Bottom row (left to right): k-space mask for 40% sampling (sampled frequencies in white), wavelet packet reconstruction from 40% sampling and the associated error map.

6 Conclusions

We propose a method for learning a wavelet packet basis from a set of images, with a cost function selected specifically for compressed sensing. We evaluate its performance in approximation and compressed sensing reconstruction of unseen images. Our results show that a wavelet packet basis learned from example brain images can yield more accurate sparse approximations of unseen brain images. However, despite this significant improvement, in reconstruction of brain images from partial k-space data the difference between the learned basis and a fixed wavelet basis is quite modest. Further work will be required to explain this behaviour.

References

1. M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311 –4322, 2006.
2. E.J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346:589 – 592, 2008.
3. E.J. Candes, Y.C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.
4. E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489 – 509, 2006.
5. E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203 – 4215, 2005.
6. R.R. Coifman, Y. Meyer, S. Quake, and M.V. Wickerhauser. Signal processing and compression with wavelet packets. Technical report, Numerical Algorithms Research Group, Department of Mathematics, Yale University, New Haven, Connecticut 06520, 1990.
7. R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713 –718, 1992.
8. D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289 –1306, 2006.
9. J.M. Duarte-Carvajalino and G. Sapiro. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 18(7):1395 –1408, 2009.
10. M. Lustig, D. Donoho, and J.M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
11. M. Lustig, J.M. Santos, D.L. Donoho, and J.M. Pauly. k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In *Proceedings of the 13th Annual Meeting of ISMRM, Seattle*, page 2420, 2006.
12. S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Elsevier/Academic Press, Amsterdam, 3rd edition, 2009.
13. S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett. The Alzheimer's Disease Neuroimaging Initiative. *Neuroimaging Clinics of North America*, 15(4):869–877, 2005. Alzheimer's disease: 100 years of progress.
14. G. Peyre. Best basis compressed sensing. In *Proceedings of the 1st International Conference on Scale Space and Variational Methods in Computer Vision*, SSVM'07, pages 80–91, Berlin, Heidelberg, 2007. Springer-Verlag.
15. G. Peyre. Best basis compressed sensing. *IEEE Transactions on Signal Processing*, 58(5):2613 –2622, 2010.
16. K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Transactions on Image Processing*, 2(2):160 –175, 1993.
17. S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028 –1041, 2011.
18. M.V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. AK Peters, Ltd., Wellesley, Massachusetts, 1994.