

## Improved MRI Reconstruction via Non-Convex Elastic Net

Angshul Majumdar<sup>1</sup> and Rabab K. Ward<sup>2</sup>

<sup>1</sup>Indraprastha Institute of Information Technology, Delhi

<sup>2</sup>Department of Electrical and Computer Engineering, University of British Columbia, Vancouver

### ABSTRACT

This work proposes the use of an elastic-net to reconstruct Magnetic Resonance Images from their partially sampled K-space. The resulting elastic-net formulation of this problem is composed of two terms – the first term promotes sparsity and the other one promotes a grouping effect. The advantage of using an elastic-net for MRI reconstruction is that it can recover the hierarchically correlated sparse wavelet coefficients of the image. We develop two reconstruction methods via two elastic-net formulations - the synthesis prior and the analysis prior. We also impose non-convex sparsity penalties. There are no existing algorithms that solve such problems; hence we derive efficient algorithms for solving them. The experimental results show that our proposed analysis prior method outperforms state-of-the-art in MRI reconstruction.

**Index Terms**— elastic net, compressed sensing, MRI.

### 1. INTRODUCTION

In Magnetic Resonance Imaging (MRI), the acquisition model is expressed as,

$$y = Fx + \eta, \quad \eta : N(0, \sigma^2) \quad (1)$$

where  $x$  is the underlying MRI signal that needs to be reconstructed,  $F$  is the Fourier mapping from the image domain to the frequency space,  $y$  is the acquired data in the frequency space (K-space) and  $\eta$  is the noise.

When the K-space is fully sampled on the uniform Cartesian grid, the reconstruction is trivial – i.e. an inverse FFT is simply applied on the collected K-space samples  $y$  to reconstruct the image  $x$ . However, the full sampling of the K-space on a uniform Cartesian grid is time consuming and is the main source of delay in MRI acquisition. The straightforward way to accelerate the MRI scan time is to partially sample the K-space. The partially sampled K-space data acquisition is modeled as,

$$y = RFx + \eta \quad (2)$$

where  $R$  is the sampling mask.

Such partial K-space sampling reduces the acquisition time, but the reconstruction becomes challenging. Equation (2) is an under-determined linear inverse problem since the number of collected K-space samples is less than the number of pixels in the image.

Presently the most obvious way to solve (2) is to employ compressed sensing techniques e.g. as in [1-3]. Therefore to reconstruct the MR image, its sparsity in the wavelet transform domain is exploited. The Wavelet transforms that are employed in MRI reconstruction are either orthogonal or tight-frame<sup>1</sup>. For both cases, the transform coefficient vector  $\alpha$  and the image domain signal  $x$  are represented by the analysis-synthesis equations:

$$\text{analysis} : \alpha = Wx \quad (3a)$$

$$\text{synthesis} : x = W^T \alpha \quad (3b)$$

Incorporating the synthesis equation (3b) into the partial K-space data acquisition model (2), one gets

$$y = RFW^T \alpha + \eta \quad (4)$$

The sparse transform coefficients are easily solved by  $l_p$ -norm minimization [1-3],

$$\hat{\alpha} = \min_{\alpha} \|\alpha\|_p^p \quad \text{subject to} \quad \|y - RFW^T \alpha\|_2^2 \leq \epsilon \quad (5)$$

where  $\|\alpha\|_p^p = \sum_i |\alpha_i|^p$  and  $\epsilon = n\sigma^2$ ,  $n$  being the number of K-space samples.

Once the wavelet coefficients  $\alpha$  are recovered, the image is obtained by applying the synthesis equation (3b).

The aforementioned synthesis prior formulation (5) is the most common technique for CS based MRI reconstruction; it is applicable for both orthogonal and tight-frame transforms. However, it was observed in [3-5] that better results are obtained when tight-frames are used with the alternate analysis prior formulation:

$$\hat{x} = \min_x \|Wx\|_p^p \quad \text{subject to} \quad \|y - RFx\|_2^2 \leq \epsilon \quad (6)$$

For the orthogonal transforms, the analysis and the synthesis prior are equivalent theoretically; but they are not the same for tight-frames.

The wavelet transform of piecewise smooth images have a tree sparse structure, i.e. if there is a high valued coefficient at a higher wavelet scale, then there is a high valued coefficient at the corresponding positions at the lower scales as well. Several studies have analysed the tree structured sparsity [6, 7] of the wavelet transform. They show that, incorporating the structural information into the sparsity assumption yields better reconstruction results.

<sup>1</sup> Orthogonal :  $W^T W = I = W W^T$

Tight - frame :  $W^T W = I \neq W W^T$

Solving the inverse problem (4) by assuming tree structured sparsity is a challenging task since the algorithm requires searching for the tree structure at every iteration. This is a computationally intensive process and hence the algorithms proposed for solving the tree-structured sparsity problem [8, 9] are not scalable enough for MRI reconstruction.

A smarter alternative was proposed in [10]. It does not explicitly exploit the tree-structure, but groups the wavelet coefficients into overlapping groups by following this structure. The groups are defined by the finest scale indices, and the indices at lower resolutions belong to multiple groups. Such a group-sparsity problem can be solved efficiently, and as shown in [10], this formulation leads to very good MRI reconstruction results.

In this work, we do not impose such strong group-sparsity constraints. What we observe is that the sparse wavelet coefficients for a piece-wise smooth signal are not really independent. If the leaf node is of high value, it is likely that the root nodes will be of high values as well. High valued coefficients do not occur independently in the wavelet domain but are hierarchically correlated.

Based on this assumption, we formulate an elastic-net problem for MRI reconstruction. The elastic-net [11, 12] is a popular convex regression technique in the machine learning community. Our work is influenced by studies in non-convex CS [1-3]; therefore instead of the convex elastic-net [11, 12] we propose a non-convex version of it. Moreover, the original elastic-net is a synthesis prior formulation; motivated by the success of analysis prior formulation for MRI reconstruction [3-5] we propose a novel analysis prior elastic net formulation. Non-convex versions of synthesis and analysis prior elastic-net regularization are new; thus there are no efficient algorithms to solve them. We derive simple and efficient algorithms to solve these problems.

The rest of the paper is organized into several sections. The following section briefly discusses the elastic-net formulation. In section 3, we frame MRI reconstruction as an elastic-net problem. The algorithms for solving the optimization problems are derived in section 4. The experimental results are discussed in section 5. Finally, the conclusions of the work are shown in section 6.

## 2. ELASTIC-NET REGULARIZATION

Consider the classical regression problem:

$$y = Ax + \eta, \quad \eta : N(0, \sigma^2) \quad (7)$$

where  $y$  is a vector of the collected data, the  $A$  matrix consists of the explanatory variables,  $x$  is the unknown weight vector which interprets the data in terms of the explanatory variables and  $\eta$  is the noise.

Since the noise is assumed to be Normally distributed, one needs to minimize the least squares cost function. In most cases, the problem is not well conditioned and needs to be regularized.

The most straightforward regularization is the ridge regression, which is expressed as:

$$x_{ridge} = \min_x \|y - Ax\|_2^2 + \lambda \|x\|_2^2 \quad (8)$$

Unfortunately ridge regression results in a dense solution, i.e. it explains the data in terms of all the explanatory variables. The outcome lacks interpretability. In order to overcome this issue, the LASSO (least angle shrinkage and selection operator) was proposed in [13]. LASSO replaces the  $l_2$ -norm constraint by an  $l_1$ -norm:

$$x_{lasso} = \min_x \|y - Ax\|_2^2 + \lambda \|x\|_1 \quad (9)$$

The  $l_1$ -norm penalty promotes selection of very few variables, i.e. the weight vector  $x$  is sparse.

The selection of few variables improves interpretability. One can now analyze and interpret the data with only a few explanatory variables.

However, LASSO suffers from a serious shortcoming. In most cases the explanatory variables are not independent from each other, they are correlated. In such a situation, the LASSO selects only one variable from the group of correlated ones. This is the result of enforcing too much sparsity on the variable selection operation. One would like to know all the variables which have contributed to the outcome ( $y$ ); but LASSO ignores the correlated variables and thus loses out on correct interpretability. In order to promote the grouping of correlated variables, the elastic net regularization was proposed in [11]. The optimization problem is framed as follows:

$$x_{net} = \min_x \|y - Ax\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|x\|_2^2 \quad (10)$$

Here the  $l_1$ -norm constraint promotes sparsity (as in LASSO) but the  $l_2$ -norm constraint promotes selection of correlated variables.

## 3. PROPOSED FORMULATION

We are interested in MRI reconstruction, where the task is to recover the sparse wavelet transform coefficients from the partial K-space data (3). We argued above that the high valued wavelet coefficients do not occur independently; rather the wavelet coefficients at the different scales are hierarchically correlated.

Generally  $l_p$ -minimization such as (5) and (6) above, is employed to recover the sparse wavelet transform coefficients [1-3]. But such  $l_p$ -minimization suffers from the same shortcomings as the LASSO - there is always a chance that the  $l_p$ -minimization will not recover all the hierarchically correlated wavelet coefficients. In order to address this issue, we propose an elastic-net like regularization term for both the synthesis and analysis priors respectively :

$$\hat{\alpha} = \min_{\alpha} \|\alpha\|_p^p + \eta \|\alpha\|_2^2 \quad \text{subject to} \quad \|y - RFW^T \alpha\|_2^2 \leq \epsilon \quad (11)$$

$$\hat{x} = \min_x \|Wx\|_p^p + \eta \|Wx\|_2^2 \quad \text{subject to} \quad \|y - RFx\|_2^2 \leq \epsilon \quad (12)$$

Here the  $l_p$ -norm enforce sparsity in the selected variables, but the  $l_2$ -norm promotes selection of grouped variables.

Such non-convex elastic-net problems have not been formulated before, especially the analysis prior formulation (12) is entirely new. There are no efficient algorithms to solve these problems. In the following

section, we derive the algorithms for solving these optimization problems.

#### 4. SOLVING THE OPTIMIZATION PROBLEMS

The task is to solve (11) and (12). Solving the constrained optimization problems is difficult. Therefore we will derive algorithms to that minimize their unconstrained counterparts.

$$\text{synthesis} : \frac{1}{2} \|y - Ax\|_2^2 + \frac{\lambda}{p} \|x\|_p^p + \frac{\lambda \cdot \eta}{2} \|x\|_2^2 \quad (13)$$

$$\text{analysis} : \frac{1}{2} \|y - Ax\|_2^2 + \frac{\lambda}{p} \|Hx\|_p^p + \frac{\lambda \cdot \eta}{2} \|Hx\|_2^2 \quad (14)$$

The formulations (13), (14) are equivalent to (11), (12), for correct choice of the parameters  $\lambda$  and  $\varepsilon$ . Unfortunately the relationship between these two parameters is not analytical. However, when  $\varepsilon$  is known, it is possible to find  $\lambda$  by global cross validation or via the L-curve method.

The first simplification for both the analysis and the synthesis priors is to assemble all the  $l_2$ -norm terms together, so that (13) is recast as (15).

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{\lambda}{p} \|x\|_p^p \quad (15)$$

$$\text{where } y' = \begin{pmatrix} y \\ 0 \end{pmatrix} \text{ and } A' = \begin{pmatrix} A \\ \sqrt{\lambda \cdot \eta} I \end{pmatrix}.$$

Similarly, the analysis prior form (14) can be recast as (16) using similar notations:

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{\lambda}{p} \|Hx\|_p^p \quad (16)$$

$$\text{where } A' = \begin{pmatrix} A \\ \sqrt{\lambda \cdot \eta} H \end{pmatrix}.$$

In the second step, we introduce a variable splitting, as is done in alternating directions methods [14]; this will decompose the original problem into two easy sub-problems.

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{1}{2} \|x - w\|_2^2 + \frac{\lambda}{p} \|w\|_p^p \quad (17)$$

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{1}{2} \|x - w\|_2^2 + \frac{\lambda}{p} \|Hw\|_p^p \quad (18)$$

Here  $w$  is a proxy for  $x$ . We proceed by alternately fixing one variable and solving for the other (i.e., alternating direction), and iterating. For both (17) and (18), with fixed  $w$ , the subproblems in  $x$  are quadratic:

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{1}{2} \|x - w\|_2^2 \quad (19)$$

This can be recast as:

$$\frac{1}{2} \|y'' - A''x\|_2^2 \quad (20)$$

$$\text{where } y'' = \begin{pmatrix} y' \\ w \end{pmatrix} \text{ and } A'' = \begin{pmatrix} A' \\ I \end{pmatrix}$$

and has a closed form solution, but in practice (20) is easily solved by Conjugate Gradient (CG) methods.

In the next step of the iteration, we fix  $x$ , and solve the following subproblems:

$$\frac{1}{2} \|x - w\|_2^2 + \frac{\lambda}{p} \|w\|_p^p \quad (21)$$

$$\frac{1}{2} \|x - w\|_2^2 + \frac{\lambda}{p} \|Hw\|_p^p \quad (22)$$

Solving (21), is straightforward, via the p-shrinkage operator [15, 16]. For each element  $i$ , this is defined as:

$$w_i = \text{signum}(x_i) \max(0, |x_i| - \frac{\lambda}{p} |x_i|^{p-1}) \quad (23)$$

Solving the subproblem (22) is slightly more complicated. Differentiating (22) yields,

$$x - w + \lambda H^T DHx, \text{ where } D = \text{diag}(|Hx|^{p-2}) \quad (24)$$

Setting the gradient to zero, one gets

$$(I + \lambda H^T DH)w = x \quad (25)$$

Using the matrix inversion lemma,

$$(I + \lambda H^T DH)^{-1} = I - H^T \left( \frac{1}{\lambda} D^{-1} + H^T H \right)^{-1} H$$

we have the following identity,

$$w = x - H^T \left( \frac{1}{\lambda} D^{-1} + H^T H \right)^{-1} Hx \quad (26)$$

Or equivalently,

$$z = \left( \frac{1}{\lambda} D^{-1} + H^T H \right)^{-1} Hx \quad (27)$$

where  $w = x - H^T z$

Solving  $z$  requires solving the following,

$$\left( \frac{1}{\lambda} D^{-1} + H^T H \right) z = Dx \quad (28)$$

Adding  $cz$  to both sides of (28) and subtracting  $H^T Hz$  gives,

$$z = \left( \frac{1}{\lambda} D^{-1} + cI \right)^{-1} (cz + H(x - H^T z)) \quad (29)$$

where  $w = x - H^T z$

where  $c$  is the maximum eigenvalue of  $W^T W$ .

Inverting  $\left( \frac{1}{\lambda} D^{-1} + cI \right)$  is easy since  $D$  is a diagonal matrix. This is a coupled equation, in practice it is solved iteratively, i.e. at the  $k^{\text{th}}$  iteration,

$$z_k = \left( \frac{1}{\lambda} D^{-1} + cI \right)^{-1} (cz_{k-1} + H(x - H^T z_{k-1}))$$

The last step of the algorithm is to relax the equality constraint, between  $x$  and  $w$  so that instead of having  $\|x - w\|_2^2$  as in (19), we have a relaxed version of it,

$$\frac{1}{2} \|y' - A'x\|_2^2 + \frac{1}{2} \|x + \beta - w\|_2^2 \quad (30)$$

where  $\beta$  is the dual variable, which is updated as,

$$\beta \leftarrow \beta + x - w \quad (31)$$

Owing to limitations in space, we cannot write the algorithms concisely; therefore in this section we have only outlined the major milestones of the algorithm.

## 5. EXPERIMENTAL EVALUATION

In this work, we have performed an experimental study on three different MR images (Fig. 1). The ground-truth data is collected by fully sampling the k-space on a uniform Cartesian grid. All the images are of size 256 x 256. For the experiments, the variable density random sampling with 3 fold acceleration factor (i.e. 33% sampling) is used for simulating the partial sampling of the K-space.

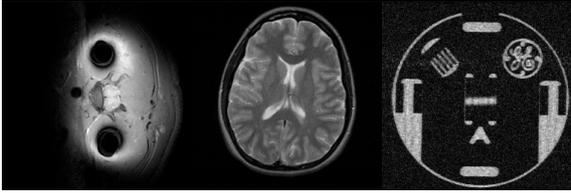


Fig. 1. Left to Right: Spine (rat), Brain and Phantom.

Our method requires specification of three parameters:  $p$ ,  $\eta$  and  $\lambda$ . We found that  $p = 0.8$  yields the best results. The values of  $\eta = 0.25$  and  $\lambda = 0.1$  were found via the L-curve method. We used the complex dual-tree wavelet as the sparsifying transform.

The results of quantitative evaluation are given in Table 1. Normalized Mean Squared Error (NMSE) is the standard metric used for evaluating MRI reconstruction. We compared our approach with the sparse recovery method [1] and the overlapping group-sparse recovery method [10].

Table 1: Reconstruction Results

Dataset	Sparse [1]	Group-sparse [10]	Synthesis prior elastic-net	Analysis prior elastic-net
Spine	0.12	0.08	0.09	0.05
Brain	0.18	0.13	0.15	0.09
Phantom	0.23	0.15	0.18	0.11

The results are as expected. The sparse recovery technique [1] does not assume anything apart from sparsity in the wavelet domain. It uses the most generalized formulation and undoubtedly yields the worst results. The overlapping group-sparsity prior [10] enforces the most stringent constraints on the reconstruction problem and yields very good results. Our elastic-net formulation does not enforce strict group-sparsity but encourages grouping effect. Both [1] and [10] are synthesis prior formulations. Our synthesis prior formulation is only marginally worse than [10]; but the analysis prior formulation yields considerably superior results than [10]. The fact that analysis prior yields better results than the synthesis prior have been observed before [3-5].

For qualitative evaluation, the 4 reconstructed and the 4 difference images are shown in Fig. 1. Owing to limitations in space, we only show the results for the brain image. The difference images are contrast enhanced by 10 times for visual clarity.

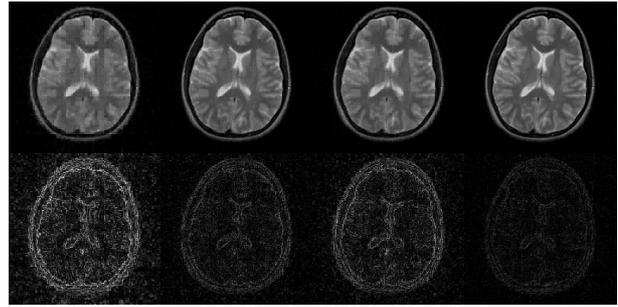


Fig. 2. Reconstructed and difference images. Left to Right: sparse recovery [1], group-sparse recovery [10], proposed synthesis prior and proposed analysis prior.

The qualitative results corroborate the quantitative observations. The difference images (between ground-truth and reconstructed) show that sparse recovery method yields the worst results, followed by our proposed synthesis prior elastic-net. The best results are obtained by our analysis prior elastic-net formulation, followed by the over-lapping group sparse recovery technique [10]. This is also evident from the reconstructed images. The images reconstructed by the Sparse recovery and our synthesis prior elastic-net methods have considerable reconstruction artifacts. But these artifacts are virtually non-existent in the group-sparse recovery technique [10] and in our analysis prior formulation.

## 6. CONCLUSION

Generally,  $l_p$ -norm minimization methods are employed for reconstructing MR images from partially sampled K-space data. Such methods exploit the sparsity of images in the wavelet domain for reconstruction. It is well known that the wavelet coefficients of such images show a tree-structure, i.e. if a coefficient at a higher scale has a high value, the corresponding wavelet coefficient at each of the lower scales will also exhibit high values. Standard  $l_p$ -minimization fail to explicitly capture this hierarchical correlation.

Algorithms that explicitly exploit the wavelet tree-structure are not scalable for practical MRI problems. A recent work has proposed an overlapping group-sparse formulation based on this tree-structure [10] and has shown significant improvements over the standard  $l_p$ -minimization methods. In this work, rather than assuming a strong group-structure, we propose to impose a grouping effect via elastic-net formulation. The resulting synthesis prior formulation yields better results than standard  $l_0$ -minimization [1] but is slightly worse than [10]. Our analysis prior formulation however, yields even better reconstruction than the group-sparsity formulation [10].

Single channel MRI reconstruction is the simplest problem in this domain. This work shows how it can be improved via the elastic-net formulation. In the future, we plan to exploit this hierarchical correlation in the wavelet domain for solving more complex MRI reconstruction problems, such as those arising in multi-channel parallel MRIs and multi-echo MRIs.

## REFERENCES

- [1] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l0-minimization," *IEEE Trans. Med. Imaging*, Vol. 28 (1), pp. 106–121, 2009.
- [2] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data", in *IEEE International Symposium on Biomedical Imaging*, 2009.
- [3] A. Majumdar and R. K. Ward, "Under-determined Non-Cartesian MR Reconstruction", *MICCAI*, pp. 513-520, 2010.
- [4] A. Majumdar and R. K. Ward, "On the Choice of Compressed Sensing Priors: An Experimental Study", *Signal Processing: Image Communication*, Vol. 27 (9), pp. 1035–1048, 2012.
- [5] A. Majumdar and R. K. Ward, "Accelerating Multi-Echo T2 Weighted MR Imaging: Analysis Prior Group Sparse Optimization", *Journal of Magnetic Resonance*, Vol. 210 (1), pp. 90-97, 2011.
- [6] L. He and L. Carin, "Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing", *IEEE Transactions on Signal Processing*, Vol. 57 (9), pp. 3488-3497, 2009.
- [7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing", *IEEE Transactions in Information Theory*, Vol. 56, pp. 1982-2001, 2010.
- [8] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Wavelet-Domain Compressive Signal Reconstruction Using a Hidden Markov Tree Model", *IEEE ICASSP*, pp. 5137-5140, 2008.
- [9] M. N. Do and C. N. H. La, "Tree-based majorize-minimize algorithm for compressed sensing with sparse-tree prior," *Int. Workshop Computat. Adv. Multi-Sensor Adapt. Processing*, pp. 129-132, 2007.
- [10] C. Chen and J. Huang, "The Benefit of Tree Sparsity in Accelerated MRI", *MICCAI workshop on Sparsity Techniques in Medical Imaging*, 2012.
- [11] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net", *J. Royal Statist. Soc. B.*, Vol. 67 (2), pp. 301-320, 2005.
- [12] Van der Kooij, A.J. & Meulman, J.J. (2006). Regularization with Ridge penalties, the Lasso, and the Elastic Net for Regression with Optimal Scaling Transformations.
- [13] Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso", *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pp. 267-288
- [14] R. Chartrand and B. Wohlberg, "A Nonconvex ADMM Algorithm for Group Sparsity with Sparse Groups", *IEEE ICASSP*, 2013.
- [15] S. Voronin and R. Chartrand, "A new generalized thresholding algorithm for inverse problems with sparsity constraints", in *IEEE ICASSP*, 2013
- [16] R. Chartrand, "Generalized shrinkage and penalty functions", in *IEEE Global Conference on Signal and Information Processing*, 2013