KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Antonello N., De Sena E., Moonen M., Naylor P. A., van Waterschoot T. (2016) **SPARSE SIGNAL RECONSTRUCTION VIA PARAMETERIZED HYPERBOLIC REGULARIZATION AND GRADUATED NON-CONVEXITY**
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	
Journal homepage	
Author contact	niccolo.antonello@esat.kuleuven.be *+ 32 (0)16 321855*
IR	

(article begins on next page)

SPARSE SIGNAL RECONSTRUCTION VIA PARAMETERIZED HYPERBOLIC REGULARIZATION AND GRADUATED NON-CONVEXITY

Niccolò Antonello[1], Enzo De Sena[1], Marc Moonen[1], Patrick A. Naylor[2] and Toon van Waterschoot[1,3]

[1] KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), 3001 Leuven, Belgium
[2] Imperial College London, Dept. of Electrical Engineering, SW7 2AZ, London, United Kingdom
[3] KU Leuven, Dept. of Electrical Engineering (ESAT-ETC), 2440 Geel, Belgium

ABSTRACT

One of the most used algorithms in sparse signal reconstruction is the least absolute shrinkage and selection operator (LASSO). A weakness of LASSO is that the non-zero components of the reconstructed sparse signal are biased towards smaller values. Here a novel family of sparsity-inducing parameterized regularization functions is constructed. Rather than defining the cost functions themselves, the design is carried out on their gradient. The resulting cost function is twice-differentiable and thus allows to use typical Newton iterations. Furthermore, it enables to remove the bias present in the LASSO solution. The family of regularization functions is parameterized by two variables, which enable to control the convexity of the function. Starting from the solution to a convex optimization problem, a series of optimization problems is iteratively solved while increasing the non-convexity of the regularization function until a sparse solution is obtained. The algorithm is shown to outperform LASSO in an acoustic single-channel equalization problem.

Index Terms— Graduated Non-Convexity, sparse optimization, LASSO, sparse regularization, non-convex optimization

1. INTRODUCTION

Sparse signal reconstruction appears in many fields of engineering. This problem can be formulated as follows. A sparse signal $\mathbf{x} \in \mathbb{R}^{N_x}$, is to be recovered from noisy measurements $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^{N_y}$ is additive noise, $\mathbf{y}$ is an N_y long vector and $\mathbf{H}$ is an $N_x \times N_y$ matrix. One way to carry out the recovery consists in solving the following optimization problem:

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}), \tag{1}$$

where $\phi(\mathbf{x})$ is a sparsity-inducing regularization function and λ is a scalar. The optimal solution of this problem, $\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} f(\mathbf{x})$, is the reconstructed sparse signal. A typical choice of $\phi(\mathbf{x})$ is the l_1-norm, which leads to the least absolute shrinkage and selection operator (LASSO) [1]. The cost function of LASSO is non-differentiable at the origin due to the absolute value operation of the l_1-norm. Typically LASSO is solved using iterative shrinkage-thresholding [2], quadratic programming [1], or convex optimization [3]. One of the weaknesses of LASSO is that it tends to bias the non-zero components of the reconstructed sparse signal towards smaller values. Different approaches have been proposed to compensate for this weakness [4, 5].

In this paper, an alternative approach is presented: a novel family of parameterized regularization functions $\phi_h(\mathbf{x})$ is proposed. As in [6], these regularization functions are designed by their gradient and not by their cost function. Their gradient is defined using differentiable hyperbolic functions. This is done in such a way that the resulting regularization function approximates the l_1-norm around zero and the l_0-norm elsewhere. Acting in this way it is possible to remove the bias of non-zero components leading to an improved accuracy of the reconstructed signal. The proposed regularization functions are non-convex but their non-convexity is controlled by two parameters. These parameters can be adjusted to improve the sparsity promotion, at the expense of an increased non-convexity. A similar sparsity-inducing regularization function, i.e. a family of arctangent regularization functions, was used in [5]. In [5] convex optimization was used to solve (1). Since this family of regularization functions can, in general, be non-convex, an additional optimization problem was solved in order to obtain the set of parameters that maintain problem (1) convex. The arctangent regularization function was still non-differentiable at zero. The novel family of regularization functions presented in this paper is motivated by the observation that having a differentiable gradient of the sparsity-inducing regularization functions makes it possible to use Newton optimization algorithms. Newton optimization algorithms are easy to implement and have fast convergence speed. A twice differentiable approximation of the l_1-norm has also been studied in [7]. However, since it was based on an approximation of the LASSO cost function, the solution remained biased. In this paper, problem (1) is solved using a variant of the graduated non-convexity (GNC) method [8–10]. The idea is to start from a convex problem and to solve iteratively a number of optimization problems by gradually increasing the non-convexity of $\phi_h(\mathbf{x})$. Each problem is initialized with the solution of the previous iteration. The main difference with the original GNC is the new family of regularization functions. The paper is structured as follows: in Sec.2 the novel family of regularization functions is presented. In Sec.3 the conditions for which problem (1) can be kept convex while using these regularization functions are studied. Sec.4 describes how the parameters of the regularization functions should be chosen to avoid ill-conditioning problems. In Sec.5 the proposed variant of the GNC method is described. Finally, in Sec.6 simulation results are presented.

This research work was carried out at the ESAT Laboratory of KU Leuven, the frame of the FP7-PEOPLE Marie Curie Initial Training Network "Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)", funded by the European Commission under Grant Agreement no. 316969, KU Leuven Research Council CoE PFV/10/002 (OPTEC), the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office IUAP P7/19 "Dynamical systems control and optimization" (DYSCO) 2012-2017, KU Leuven Impulsfonds IMP/14/037 and was supported by a Postdoctoral Fellowship (F+/14/045) of the KU Leuven Research Fund. The scientific responsibility is assumed by its authors.

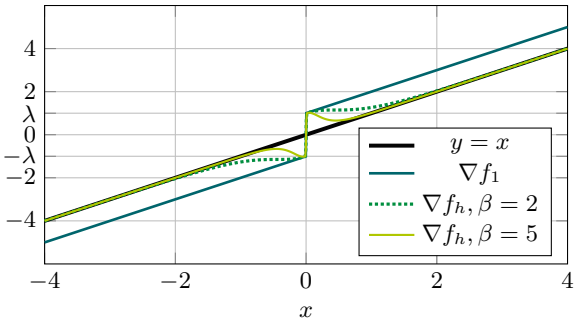

Fig. 1. Sub-gradient of $f_{l_1}(x)$ and gradient of $f_h(x)$ for $\gamma = 100$, $\lambda = 1$ and $y = 0$ for the scalar case.

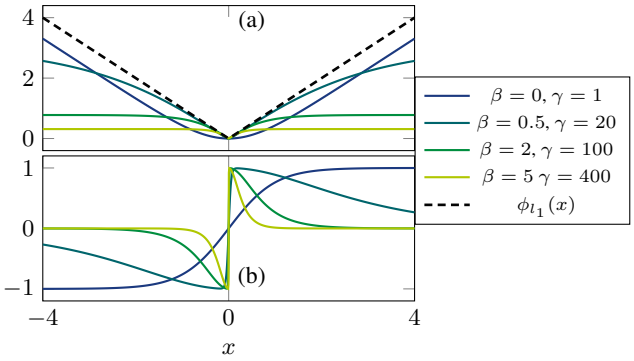

Fig. 2. Sparsity-inducing regularization $\phi_h(x)$ (a) and its derivative $\nabla\phi_h(x)$ (b) for values of β and γ.

2. SPARSITY-INDUCING REGULARIZATION

The gradient of the cost function of $f(\mathbf{x})$ of problem (1) can be written as:

$$\nabla f(\mathbf{x}) = \mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y}) + \lambda\nabla\phi(\mathbf{x}). \tag{2}$$

This section introduces the novel family of regularization functions. In order to explain the rationale behind the choice of regularization functions in simple terms, this section focuses on the simpler case of $\mathbf{H} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. In this case the problem is effectively reduced to a simple thresholding problem. The parameter λ now represents a threshold. What is sought here is to cancel out from the signal $\mathbf{y}$ all the components that are below the threshold λ. *Hard thresholding* represent one of the solutions to this problem: if the m-th component of $\mathbf{y}$ is $y_m < \lambda$ then the m-th component of reconstructed signal $\hat{\mathbf{x}}$ is $\hat{x}_m = 0$, otherwise $\hat{x}_m$ is equal to y_m. One of the weaknesses of the hard thresholding is that when a component of $\mathbf{y}$ is close to λ small perturbations result in big changes of $\hat{\mathbf{x}}$.

In order to avoid this, *soft thresholding* can be used instead. Soft thresholding is actually the particular case of LASSO where $\mathbf{H} = \mathbf{I}$. The sparsity-inducing regularization used here is the l_1-norm

$$\phi_{l_1}(\mathbf{x}) = \|\mathbf{x}\|_1, \tag{3}$$

which is non-differentiable at the origin and has the sub-gradient $\nabla\phi_{l_1}(\mathbf{x}) \in \mathbb{R}^{N_x}$ with the m-th component defined as

$$\frac{\partial\phi_{l_1}(\mathbf{x})}{\partial x_m} = \text{sign}(x_m). \tag{4}$$

Fig.1 shows the sub-gradient of the cost function when $\nabla\phi_{l_1}(\mathbf{x})$ is used in (2) for the scalar case and $y = 0$. For different values of y the graph would simply be shifted up or down by y. The point where the gradient is null is the optimal solution $\hat{x}$. For example, if $y = 4$ this would result in shifting the graph down, i.e. on Fig.1 the label of 4 would be substituted with 0 and the others accordingly, resulting in the gradient vanishing at $4 - \lambda$ and not at 4 as sought. Hence the optimal solution $\hat{x}$ is biased towards smaller components respect to y.

In order to remove this bias, $\phi_{l_1}(\mathbf{x})$ will be replaced by $\phi_h(\mathbf{x})$ with its gradient $\nabla\phi_h(\mathbf{x}) \in \mathbb{R}^{N_x}$ having its m-th component defined as

$$\frac{\partial\phi_h(\mathbf{x})}{\partial x_m} = \tanh(\gamma x_m)\,\text{sech}(\beta x_m), \tag{5}$$

where β and γ are scalars. It can be noticed in Fig.1 that for large values of γ the gradient $\nabla f_h(x)$ will approximate well the sub-gradient $\nabla f_{l_1}(x)$ around zero and will approach $(y - x)$ elsewhere, hence removing the bias. The parameters β and γ control the non-convexity of this regularization function. For $\beta = 2$, the gradient $\nabla f_h(\mathbf{x})$ is monotonically increasing, thus avoiding the problem of having multiple local minima and resulting in a convex problem. However it can be noticed that if e.g. $y = 1 + \epsilon$ with ϵ being small, the graph of Fig.1 would be shifted down of this value and the gradient would not vanish at $x = 1 + \epsilon$ as sought. Therefore a bias would be still present. This actually happens for all $1 < |y| < 2$. If β is further increased, i.e. $\beta = 5$ in Fig.1, this is no longer the case but then convexity is lost.

For $\gamma = \beta = 1$, $\nabla\phi_h(\mathbf{x})$ can be integrated into the hyperbolic secant $\phi_h(\mathbf{x}) = -\,\text{sech}(\mathbf{x}) + C$, where C is a constant. For specific β and γ, expression (5) cannot be integrated analytically. However, it can be integrated numerically. The resulting $\phi_h(\mathbf{x})$ can be seen for the scalar case in Fig.2. The choice of using two parameters instead of one will be motivated in Sec.4. Clearly for relatively large β and γ this function is non-convex, but as it was shown describing Fig.1, for a specific set of β and λ the overall problem can be kept convex.

3. CONVEXITY

In this section the convexity of $f_h(\mathbf{x})$ is studied, initially for the simpler case where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N_x}$ and $\mathbf{H} = \mathbf{I}$. The Hessian of $\phi_h(\mathbf{x})$ is a diagonal matrix $\nabla^2\phi_h(\mathbf{x})$ having its m-th diagonal components defined as

$$\begin{aligned}\frac{\partial^2\phi_h(\mathbf{x})}{\partial x_m^2} = \;& \gamma\,\text{sech}^2(\gamma x_m)\,\text{sech}(\beta x_m) \\ & - \beta\tanh(\gamma x_m)\tanh(\beta x_m)\,\text{sech}(\beta x_m).\end{aligned} \tag{6}$$

Notice that the first term is bounded between 0 and γ since $0 \leq \text{sech}(x_m) \leq 1$. Since for $x_m \geq 0$, $\tanh(x_m) \geq 0$ and for $x_m \leq 0$, $\tanh(x_m) \leq 0$, then $0 \leq \tanh(\gamma x_m)\tanh(x_m) \leq 1$. Hence $-1 \leq -\tanh(\gamma x_m)\tanh(\beta x_m)\,\text{sech}(\beta x_m) \leq 0$ although it can be observed empirically that the lower bound is actually $-\frac{1}{2}$. This implies that the second term is bounded between $-\frac{\beta}{2}$ and 0. Hence the Hessian is bounded as follows

$$-\frac{\beta}{2}\mathbf{I} \preceq \nabla^2\phi_h(\mathbf{x}) \preceq \gamma\mathbf{I}. \tag{7}$$

The Hessian of the cost function is

$$\nabla^2 f_h(\mathbf{x}) = \mathbf{I} + \lambda\nabla^2\phi_h(\mathbf{x}), \tag{8}$$

and due to (7) if $\lambda \leq \frac{2}{\beta}$ it follows that

$$\nabla^2 f_h(\mathbf{x}) \succeq 0 \tag{9}$$

which proves the convexity of $f_h(\mathbf{x})$.

In the more general case where $\mathbf{H} \neq \mathbf{I}$, problem (1) cannot be interpreted as a thresholding problem any longer. Obtaining a convex problem is now more difficult since in (8) the identity matrix must be replaced by $\mathbf{H}^T\mathbf{H}$. However, if $\mathbf{H}^T\mathbf{H}$ is not singular and β is chosen to be small enough the problem is still convex since

$$\lim_{\beta \to 0} \frac{\partial \phi_h(\mathbf{x})}{\partial x_m} = \tanh(\gamma x_m) \tag{10}$$

which means that $\phi_h(\mathbf{x})$ has a monotonically increasing derivative at each component. This can be seen in Fig.2 for the scalar case with $\beta = 0$.

4. SPARSITY-INDUCING HESSIAN CONDITION NUMBER

The Newton method requires the inversion of the Hessian matrix. It is therefore important for the Hessian matrix to be well-conditioned. The reason why the family of regularization functions is parameterized over two variables instead of a single one is that if $\beta = \gamma$, the Hessian can easily become ill-conditioned. Consider, for instance, the case where $x_m = 0$, then from (6) the m-th component of the diagonal of $\nabla^2\phi_h(\mathbf{x})$ will be equal to γ, independently of the value of β. On the other hand, if x_m is equal to 1, the m-th component will be very close to zero for $\beta = \gamma$ being large. The diagonal of $\nabla^2\phi_h(\mathbf{x})$ will therefore contain very large numbers where x_m is close to zero and very small ones where $x_m \neq 0$ resulting in $\nabla^2\phi_h(\mathbf{x})$ having an extremely high condition number. If $\mathbf{x}$ is sparse this condition is is rather common. If $\beta \ll \gamma$ instead, with γ large and $x_m = 1$, the m-th component of the diagonal $\nabla^2\phi_h(\mathbf{x})$ will be $-\beta \tanh(\beta)\,\mathrm{sech}(\beta)$ which is still a small number, but non-zero. This ensures that the condition number of $\nabla^2\phi_h(\mathbf{x})$ does not grow too large.

5. HYPERBOLIC REGULARIZED GRADUATED NON-CONVEXITY

For the general case where $\mathbf{H}$ is not an identity matrix and the sparsity-inducing regularization function is non-convex, keeping problem (1) to be convex is non-trivial. In this paper, non-convex optimization is therefore used. This implies that optimal solution computed by the optimization algorithm is not guaranteed to be a global minimum. Nevertheless, local minima can still represent valid sparse solutions. A countermeasure that is often used with Newton methods is to incorporate a line-search step in the algorithm [11]. This approach, however, is inconvenient in this context since $\phi_h(\mathbf{x})$ is not known analytically. A GNC method represents a suitable alternative [8].

The idea behind GNC is to start with a convex problem which, while not inducing sparsity, produces a solution that is then used to initialize a new optimization problem where the non-convexity is slightly increased. In this way, the new cost function changes only partially and the previous solution is close to the minimum of the new cost function. Due to this fact, only few Newton iterations are necessary to converge to the new solution and a line-search algorithm is no longer necessary. This procedure is performed iteratively until β and γ are large enough to promote sparsity effectively. Algorithm 1 shows the pseudo code of the proposed HRGNC. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ represent the *relaxation parameter sequences* of the increasing parameters, e.g. for β, $\boldsymbol{\beta} = [\beta_0 \dots \beta_{k_{\max}}]^T$. Notice here that if the inner loop reaches the maximum number of iterations, i.e. the Newton optimization diverges, the solution of the previous problem is given as output. As simulations will show in Sec.6, this situation occurs only for few cases. Divergence happens only for relatively high values of β and γ where the Hessian has a high condition number.

Algorithm 1 Hyperbolic Regularized Graduated Non-Convexity

Input: $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\mathbf{y}$, λ, TOL
Output: $\hat{\mathbf{x}}$

Set $\hat{\mathbf{x}}_{0,0} = \mathbf{0}$
Compute $\mathbf{H}^T\mathbf{H}$
for $k = 0, 1, \dots, k_{\max}$ **do**
  for $l = 0, 1, \dots, l_{\max}$ **do**
    if $l = l_{\max}$ **then return** $\mathbf{x}_{k-1,l}$ ▷ Maximum iterations reached
    else
      if $\|\nabla f(\hat{\mathbf{x}}_{k,l})\|_2 \leq \text{TOL}$ **then**
        Break ▷ Exit inner loop
      else
        $\nabla f(\hat{\mathbf{x}}_{k,l}) = \mathbf{H}^T(\mathbf{H}\hat{\mathbf{x}}_{k,l} - \mathbf{y}) + \nabla\phi_h(\hat{\mathbf{x}}_{k,l})$
        $\nabla^2 f(\hat{\mathbf{x}}_{k,l}) = \mathbf{H}^T\mathbf{H} + \nabla^2\phi_h(\hat{\mathbf{x}}_{k,l})$
        $\mathbf{p}_l = -\nabla^2 f(\hat{\mathbf{x}}_{k,l})^{-1}\nabla f(\hat{\mathbf{x}}_{k,l})$ ▷ Newton Step
        $\hat{\mathbf{x}}_{k,l+1} \leftarrow \mathbf{p}_l + \hat{\mathbf{x}}_{k,l}$
      end if
    end if
  end for
  $\beta \leftarrow \beta_k$, $\gamma \leftarrow \gamma_k$ ▷ Outer loop, increase β and γ
  $\hat{\mathbf{x}}_{k+1,0} \leftarrow \hat{\mathbf{x}}_{k,l}$ ▷ Initialize new problem with previous solution
end for
return $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}}_{k,l}$

6. SIMULATION RESULTS

The algorithm is tested for a single-channel equalization problem. Here, a 0.5 seconds ($N_y = 4 \cdot 10^3$) room impulse response (RIR) of a $4 \times 4 \times 4$ m^3 room is generated using the randomized image method [12] with a reverberation time of 0.5 seconds and sampling frequency of $Fs = 8$ kHz. The positions of source and receiver are chosen randomly. The sparse signal $\mathbf{x}$ has $N_x = 500$ components and is also chosen randomly. More specifically, each component assumes a non-zero value with independent probability 0.07, and, in case it is non-zero, its amplitude is normally distributed with variance 1. The sparse signal $\mathbf{x}$ is then convolved with the RIR: the matrix $\mathbf{H}$ is a $4 \cdot 10^3 \times 500$ convolution matrix which means an overdetermined system has to be solved. Additive white noise $\mathbf{w}$ is then added to $\mathbf{H}\mathbf{x}$ to obtain $\mathbf{y}$. The additive white noise has the same energy as $\mathbf{H}\mathbf{x}$, i.e. an SNR of 0 dB is obtained. HRGNC, LASSO and LS are then used to reconstruct $\mathbf{x}$ from $\mathbf{y}$. LASSO is solved using CVX [13]. For both HRGNC and LASSO λ is set to 4. The normalized mean square error (NMSE) is defined as

$$\text{NMSE} = \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2}, \tag{11}$$

and the support error (SE) as

$$\text{SE} = \|s(\mathbf{x}) - s(\hat{\mathbf{x}})\|_0, \tag{12}$$

where $s(\mathbf{x})$ is a hard thresholding function which returns 1 if x_m is larger than 10^{-2} and 0 elsewhere.

Fig.4 shows the relaxation parameter sequences used in HRGNC. The final values are chosen to be $\beta = 5$ and $\gamma = 400$. These parameters are gradually increased in an alternating fashion in $k_{\max} = 100$

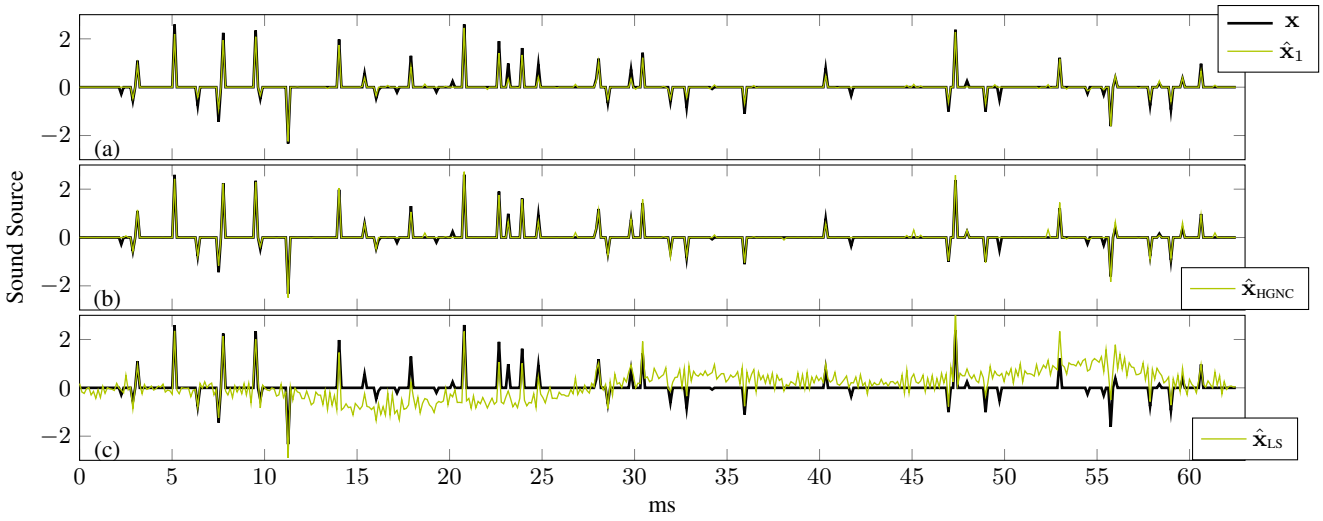

Fig. 3. Original and reconstructed sparse signals using (a) LASSO (NMSE $= 0.246$, SE $= 36$), (b) HRGNC (NMSE $= 0.167$, SE $= 24$) and (c) LS (NMSE $= 1.42$).

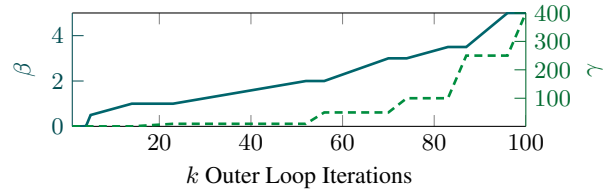

Fig. 4. Relaxation parameter sequences $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

iterations. These curves were found empirically and they can be further optimized to reduce the number of iterations and ensure convergence of the inner loop of Algorithm 1.

Fig.3 shows an example of a reconstructed sparse signal $\hat{\mathbf{x}}$ using the different algorithms. It can be observed how LS reconstructs a noisy signal with a low frequency oscillation due to the ill-conditioning of $\mathbf{H}^T\mathbf{H}$. On the other hand, this is not the case for both LASSO and HRGNC due to the regularization. It can also be noticed how the non-zero components are more biased towards lower values in the case of LASSO as compared to HRGNC. This leads to an improved NMSE for HRGNC. The SE also shows that smaller spurious non-zero components are present in the LASSO solution.

	LS	LASSO	HRGNC
Average NMSE	5.9	0.399	0.255
Average SE	-	24.17	27.3
Median SE	-	19.5	16

Table 1. Average of the error measures for the different methods for 100 different realizations of noise and random positions in the room.

In order to test the algorithms, Monte Carlo simulations are run for 100 different realizations of noise and random positions in the room. Table 1 shows the error measures for the different methods. HRGNC performs better than LASSO having an average NMSE that is 0.14 smaller. The median of the SE is also lower for HRGNC than all other methods. The mean, on the other hand, is higher. This is due to five cases where the algorithm diverged, and, as explained in Algorithm 1, the solution was taken as the one of the previous subproblem. Interestingly, in the case with highest SE, the solution has an NMSE of only 0.15, compared to 0.30 for LASSO. In other words, even though the solution is less sparse, the NMSE is particularly small.

7. CONCLUSIONS

A novel algorithm for sparse signal reconstruction has been presented. A least squares problem is regularized using a new family of parameterized sparsity-inducing regularization functions whose gradient is designed using hyperbolic functions. The sparsity-inducing regularization functions are constructed in such a way as to avoid the bias of the non-zero components of the solution towards smaller values. This is a problem which affects the well known LASSO. Thanks to the differentiability of the proposed sparsity-inducing regularization functions, Newton algorithms can be used leading to an algorithm of easy implementation and fast convergence. Since the analytical expression of $\phi_h(\mathbf{x})$ is in general not available, line-search optimization should be avoided. For this reason an algorithm based on GNC has been proposed to solve the non-convex optimization problem. Starting from the solution of a convex relaxation of the final cost function, a series of optimization problems is solved iteratively. Each problem provides a new solution that is fed to the next iteration. Two parameters, which control the non-convexity of the sparsity-inducing regularization functions, are slowly increased at each iteration creating a new cost function. The solution of each problem is reached using only few Newton iterations since the new minimum is very close to the one of the previous cost function. This makes it possible to efficiently solve the non-convex problem. The proposed HRGNC has been tested in the case of a single-channel room acoustic equalization. Compared to LASSO, HRGNC achieves in general better results that promote sparsity more efficiently and remove the non-zero component bias effectively. Future work will focus on the design of optimal relaxation parameter sequences which minimize the number of iterations and ensure convergence more robustly.

8. REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J.Roy. Statist. Soc.*, pp. 267–288, 1996.

[2] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[3] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l 1-regularized least squares," *IEEE J. of Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, 2007.

[4] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. of Fourier anal. app.*, vol. 14, no. 5-6, pp. 877–905, 2008.

[5] I. W. Selesnick and I. Bayram, "Sparse signal estimation by maximally sparse convex optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1078–1092, 2014.

[6] R. Chartrand, "Shrinkage mappings and their induced penalty functions," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2014, pp. 1026–1029.

[7] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for l1 regularization: A comparative study and two new approaches," in *Machine Learning: ECML 2007*, pp. 286–297. Springer, 2007.

[8] A. Blake and A. Zisserman, *Visual reconstruction*, vol. 2, MIT press Cambridge, 1987.

[9] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, 2009.

[10] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise gaussian mrf," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 571–585, 1998.

[11] J. Nocedal and S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.

[12] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[13] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, "Convex optimization in julia," in *High Performance Technical Computing Dynamic Languages (HPTCDL)*. IEEE, 2014, pp. 18–28.