GROUP SPARSE LMS FOR MULTIPLE SYSTEM IDENTIFICATION

Lei Yu[1*], *Chen Wei*[1], *Gang Zheng*[2]

1. School of Electronic and Information, Wuhan University, China
2. Group Non-A, INRIA Lille, France

ABSTRACT

Armed with structures, group sparsity can be exploited to extraordinarily improve the performance of adaptive estimation. In this paper, a group sparse regularized least-mean-square (LMS) algorithm is proposed to cope with the identification problems for multiple/multi-channel systems. In particular, the coefficients of impulse response function for each system are assumed to be sparse. Then, the dependencies between multiple systems are considered, where the coefficients of impulse responses of each system share the same pattern. An iterative online algorithm is proposed via proximal splitting method. At the end, simulations are carried out to verify the superiority of our proposed algorithm to the state-of-the-art algorithms.

Index Terms— LMS, Multiple system identification, Group sparsity, Proximal splitting method.

1. INTRODUCTION

System identification has its roots in standard statistical techniques. Many of the basic routines have direct interpretations as well known statistical methods such as Least Squares and maximum likelihood. And least mean square (LMS) algorithm is a classical method in adaptive system identification due to its good performance, easy implementation, and high robustness [1].

In many scenarios, some prior information of the unknown systems might be used to improve the accuracy and efficiency of adaptive estimation algorithms. In this paper, sparsity priori on system coefficients, where the system response contains a large set of zeros, will be exploited. Sparsity is actually a very common feature existed in many system, such as the echo paths [2] and digital TV transmission channels [3]. And many algorithms have been proposed recently based on LMS to utilize the prior information of sparsity, such as ℓ_0-LMS [4], ZA-LMS [5], RZA-LMS [6], OLBI-LMS [7] and so on, where a new term related to sparsity priori is usually plugged into the cost function by introducing ℓ_1 norm or ℓ_0 norm (indicates the number of non-zero elements), for instance ZA-LMS and ℓ_0-LMS.

This work is funded by National Natural Science Foundation of China under Grant no. 61401315.

Besides system sparsity, the correlations behind elements of sparse vectors exhibit a consistent pattern, i.e. structured sparsity, which has been proved to promote the accuracy of the estimation [8, 9, 10]. the correlation between systems is also a very useful information that can be exploited to improve identification performance, but rarely considered along with sparsity. Consequently, in this paper, system sparsity and correlation between the considered systems are both assumed in multiple / multi-channel systems, as shown in Fig. 1. Particularly, we can assume that the systems sharing the same condition and environment will exhibit similar sparsity pattern, i.e. zero or nonzero elements appear in the same locations, as shown in Fig. 2. This assumption can be verified in many applications such as multiple echo channel and underwater acoustic channel[11]. And in order to exploit both sparsity and system correlation, the structured sparsity of the multiple systems is assumed to identify these systems.

This paper is organized as follows. In section 2, we briefly describe the multiple system to identify and explain its structured sparse pattern. In section 3, the group sparse LMS is proposed to solve the identification problems for multiple / multi-channel systems. Section 4 provides numerical simulation results and section 5 ends the paper with a conclusion.

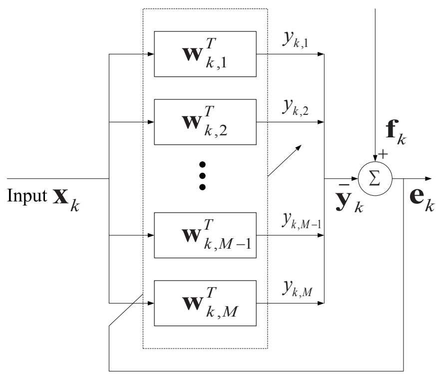


Fig. 1. Adaptive multiple systems.

2. MULTIPLE SYSTEM IDENTIFICATION

Conventional system identification or LMS filter usually considers only one system. However, in many application we face multiple systems with strong correlation. As shown in Fig. 1, at the time step $k(k = 1, 2, \cdots, t)$ the input $\mathbf{x}_k$ produces M outputs $f_{k,1}, f_{k,2}, \cdots, f_{k,M}$ through the designed adaptive system. The difference $\mathbf{e}_k$ between the output $\mathbf{y}_k$ and the desired output $\mathbf{f}_k$ adjusts the coefficients $\mathbf{w}_{k,1}, \mathbf{w}_{k,2}, \cdots, \mathbf{w}_{k,M}$ of the filters as feedback.

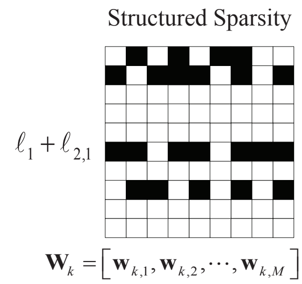


Fig. 2. Weight matrix of multiple systems with structured sparsity.

When the weight vectors are sparse, only few coefficients are large (black patches in Fig. 2) within the most zero or negligible small coefficients (blank patches in Fig. 2). We consider the case that the non-zero coefficients of the multiple systems only appear at the same entry in weight vectors as shown in Fig. 2. The weight vectors $\mathbf{w}_{k,1}, \mathbf{w}_{k,2}, \cdots, \mathbf{w}_{k,M}$ are merged into a weight matrix $\mathbf{W}_k$ as columns. This pattern of structured sparsity feature in $\mathbf{W}_k$ can be induced by constraining the sum of ℓ_1 norm and $\ell_{2,1}$ norm of $\mathbf{W}_k$ [12]. Denoting $w[i,j]$ the element of the matrix $\mathbf{W} \in \mathbb{R}^{N\times M}$ in the i-th row and the j-th column, the ℓ_1 norm and $\ell_{2,1}$ norm of $\mathbf{W}$ are expressed as

$$\|\mathbf{W}\|_1 = \sum_{i=1}^{N}\sum_{j=1}^{M}|w[i,j]|, \|\mathbf{W}\|_{2,1} = \sum_{i=1}^{N}\sqrt{\sum_{j=1}^{M}|w[i,j]|^2}. \quad (1)$$

3. STRUCTURED SPARSE LMS

In this section we firstly derive the final form of the optimization problem and then state the algorithm to solve the problem. We update the weight matrix $\mathbf{W}_{k+1}$ at time step $k+1$ by minimizing the cost function, i.e.

$$\mathbf{W}_{k+1} = \arg\min_{\mathbf{W}} \begin{pmatrix} \delta \sum\limits_{s=1}^{k} l_s(\mathbf{W}) + \frac{1}{2}\|\mathbf{W}\|_F^2 \\ +\lambda\|\mathbf{W}\|_1 + \mu\|\mathbf{W}\|_{2,1} \end{pmatrix}, \quad (2)$$

where $l_s(\mathbf{W}) = \frac{1}{2}\left\|\mathbf{W}^T\mathbf{x}_s - \mathbf{f}_s\right\|_2^2$ is the loss sum of the multiple system at time step s, $\delta > 0$ is the learning rate. The second term is the regularizer (Frobenius norm) and the last two terms induce $\mathbf{W}$ to generate the priori structured sparsity. For the optimization problem (2), the convex loss function $l_s(\mathbf{W})$ can be approximated by a linear form [13] $g_s(\mathbf{W}_s) \odot (\mathbf{W} - \mathbf{W}_s)$, with $g_s(\mathbf{W})$ its gradient

$$\begin{aligned} g_s(\mathbf{W}) &= \nabla_{\mathbf{W}} l_s(\mathbf{W}) \\ &= \mathbf{x}_s\mathbf{x}_s^T\mathbf{W} - \mathbf{x}_s\mathbf{f}_s^T. \end{aligned}$$

Denoting $su(\mathbf{W})$ the sum of all elements of $\mathbf{W}$, then (2) turns into

$$\mathbf{W}_{k+1} = \arg\min_{\mathbf{W}} \begin{pmatrix} \delta \sum\limits_{s=1}^{k} su(g_s(\mathbf{W}_s) \odot (\mathbf{W} - \mathbf{W}_s)) \\ +\frac{1}{2}\|\mathbf{W}\|_F^2 + \lambda\|\mathbf{W}\|_1 + \mu\|\mathbf{W}\|_{2,1} \end{pmatrix}, \quad (3)$$

"$\odot$" denotes the Hadamard product of two matrixes.

By adjusting constant term, optimization problem (2) is actually equivalent to the following:

$$\begin{aligned} \mathbf{W}_{k+1} &= \arg\min_{\mathbf{W}} \begin{pmatrix} \delta \sum\limits_{s=1}^{k} su(g_s(\mathbf{W}_s) \odot \mathbf{W}) + \lambda\|\mathbf{W}\|_1 \\ +\frac{1}{2}su(\mathbf{W} \odot \mathbf{W}) + \mu\|\mathbf{W}\|_{2,1} \end{pmatrix} \\ &= \arg\min_{\mathbf{W}} \begin{pmatrix} \frac{1}{2}su(\mathbf{W} \odot \mathbf{W} - 2\mathbf{Z}_{k+1} \odot \mathbf{W}) \\ +\lambda\|\mathbf{W}\|_1 + \mu\|\mathbf{W}\|_{2,1} \end{pmatrix} \\ &= \arg\min_{\mathbf{W}} \left(\frac{1}{2}\|\mathbf{W} - \mathbf{Z}_{k+1}\|_F^2 + \lambda\|\mathbf{W}\|_1 + \mu\|\mathbf{W}\|_{2,1}\right), \end{aligned} \quad (4)$$

where $\mathbf{Z}_{k+1} = -\delta \sum\limits_{s=1}^{k} g_s(\mathbf{W}_s) = \mathbf{Z}_k - \delta g_k(\mathbf{W}_k)$. The optimization problem (4) is a typical convex problem with singular terms, which can be coupled with proximal splitting method[14, 15], and each element of matrix $\mathbf{W}_{k+1}$ can be estimated as following:

$$\begin{aligned} w_{k+1}[i,j] &= \operatorname{sgn}(z_{k+1}[i,j]) \cdot (|z_{k+1}[i,j]| - \lambda)_+ \\ &\cdot \left(1 - \mu \Big/ \sqrt{\sum_{j=1}^{M}(|z_{k+1}[i,j]| - \lambda)_+^2}\right)_+, \end{aligned} \quad (5)$$

where for $x \in \mathbb{R}, (x)_+ = \max(x, 0)$. Based on the above, we propose a group sparse LMS (GS-LMS) algorithm dedicated to identifying the multiple systems with system parameters sharing same sparsity pattern. The iterative algorithm determines an adaptive filter system by minimizing the cost function (2), or equivalently (4), and the proposed algorithm can be concluded as Algorithm 1.

To ensure the convergence of the proposed algorithm, the learning rate δ should satisfy $0 < \delta < 1/\lambda_{\max}$, where $\lambda_{\max}$ is the greatest eigenvalue of the input covariance matrix $\mathbf{C} = E\left(\mathbf{x}_k\mathbf{x}_k^T\right)$. Parameters λ and μ is much related to the actual problem considered in practice, and they can be tuned according to [9], or simply fixed empirically (as presented in the simulation parts).

Note that, when parameter $\mu = 0$, Algorithm 1 will degenerate to OLBI-LMS [7]. Thus OLBI-LMS is a particular

Algorithm 1 GS-LMS algorithm

Input:
The input and desired output of multiple system at each time step $\mathbf{x}_k$ and $\mathbf{f}_k$;

Output:
The weight matrix designed at each time step $\mathbf{W}_k$;

Initialize $\mathbf{W}_1 = \mathbf{0}$ and $\mathbf{Z}_1 = \mathbf{0}$;
for $k = 1, 2, \cdots$ **do**
  $\mathbf{Z}_{k+1} = \mathbf{Z}_k + \delta\left(\mathbf{x}_k\mathbf{f}_k^T - \mathbf{x}_k\mathbf{x}_k^T\mathbf{W}_k\right)$;
  Compute $\mathbf{W}_{k+1}$ by Eq.(5) from $\mathbf{Z}_{k+1}$;
end for

case of our proposed GS-LMS algorithm. In the following section, simulations will be given to demonstrate the superiority of GS-LMS to OLBI-LMS.

4. SIMULATIONS

In this section, numerical simulations are carried out to analyze the effects of algorithm parameters and then verify effectiveness of the proposed algorithm, GS-LMS.

The simulations configuration is set as follows. The input signal $\mathbf{x}_k$ is drawn from Gaussian distribution with zero mean and unit variance. The measurement noise ε_k is also a zero-mean white Gaussian process with variance σ_ε^2 adjusted to achieve SNR=15dB. The filter length (order) is set to be $n = 500$, thus the learning rate $\delta = 8 \times 10^{-4}$. The non-zero coefficients in each column of $\mathbf{W}_*$ are drawn from a Gaussian distribution with zero mean and unit variance, and their locations are randomly assigned. The performance is evaluated using the mean square deviation (MSD) and the recovered sparse support (SSU), respectively defined as

$$\text{MSD}(k) = E(\|\mathbf{W}_k - \mathbf{W}_*\|_F^2), k = 1, 2, \cdots, t$$

and $\text{SSU}(k) = \|\mathbf{W}_k\|_0$, where $\mathbf{W}_*$ is the true filters weight matrix and k is the iteration number.

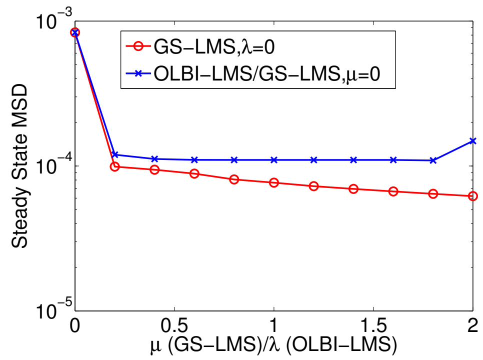
4.1. Empirical analysis for algorithm parameters

In the proposed GS-LMS algorithm, it has two un-fixed parameters, which will be empirically analyzed in this section.

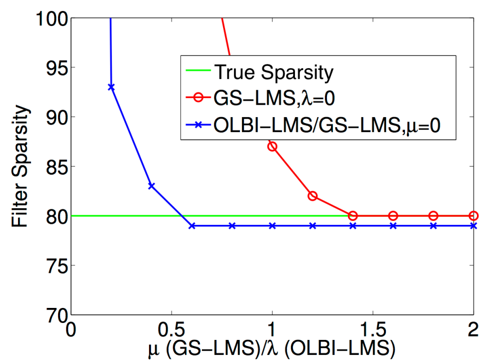
4.1.1. Estimation Accuracy

Firstly, we fix the parameter $\lambda = 0$ and then evaluate the effects of μ to the final performance of the multiple systems identification, where the filter sparsity is equal to 80 and with order of 500. The value of μ is varying from 0 to 2, then the proposed GS-LSM is exploited to identify the simulated systems, with the observed noisy signals. 100 trials are implemented and averaged to reach the expectation value of MSD, as shown in Fig. 3a (the circled line). One can easily find that the steady state MSD is decreasing. At the same time, the recovered filter SSU with respect to different μ is also plotted in Fig. 3b (the circled line). The result reflects that when μ is greater than 1.4, GS-LMS will identify a filter with same sparsity support as the original true filter.

On the other hand, the similar simulations related to parameter λ is carried out, where μ is set to be 0. The results are shown in Fig. 3(a) and (b) as starred lines. One can find that the steady state MSD is almost constant when $\lambda \in [0.5, 1.8]$ and increasing when $\lambda > 1.8$. Moreover, the recovered SP is less than the true filter SSU, i.e. over sparsify the filter coefficients. Consequently, we set this parameter as small as possible to ensure the correct sparsity support identification, for instance, $\lambda = 0.5$.



(a)



(b)

Fig. 3. Steady State MSD (a) and recovered filter SSU (b) with respect to different parameter settings.

4.1.2. Convergence Rate

In this section, the convergence rate related to parameter settings is analyzed. The system filter sparsity is set to 80 with order of 500, the parameter $\lambda = 0.5$, vary μ in $\{0.6, 1.4, 2\}$, then one can plot the MSD evolution and SSU evolution with respect to iteration numbers, as shown in Fig. 4.

Fig. 4(a) presents the evolution of MSD with different values for μ. Clearly, when μ is small, i.e. $\mu = 0.6$, the GS-LMS converges faster than higher values for μ, but the accuracy (represented by steady state MSD) is lower. On the other hand, when μ is large, i.e. $\mu = 2$, the GS-LMS converges slower than lower value for $\mu = 1.4$, but the accuracy cannot be improved significantly.

Fig. 4(b) gives the evolution of SSU with different values for μ. We can find that the convergence speed is slowing down as increasing the value of μ. However, when μ is small, i.e. $\mu = 0.6$, it gives a good convergence rate for GS-LMS, but the recovered SSU does not equal to the true filter sparse support.

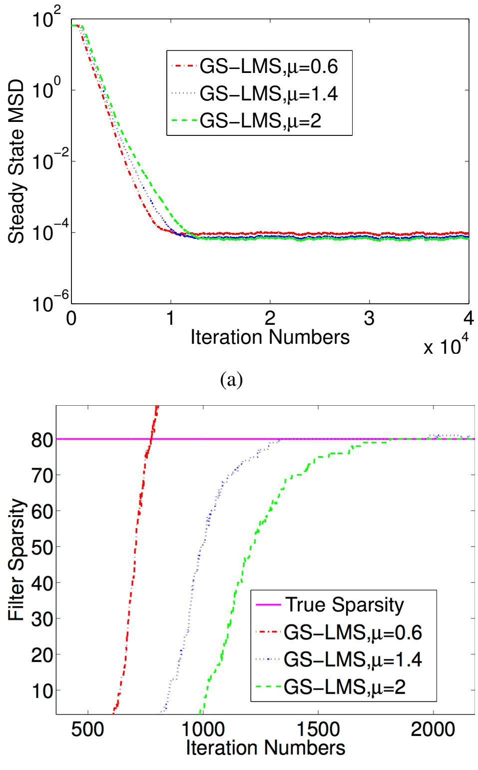


Fig. 4. MSD (a) and recovered filter SSU (b) with respect to iteration numbers.

As a conclusion, we can empirically set the parameters $\mu = 1.4$ and $\lambda = 0.5$, which leverage the accuracy and convergence speed.

4.2. Convergence rate comparisons

In this section, comparisons are made to OLBI-LMS proposed in [7] and the standard LMS according the convergence rate. MSD and SSU are both considered. As mentioned before that OLBI-LMS is actually a special case of GS-LMS, where $\mu = 0$. Thus, in this simulation, we set $\lambda = 0.5$ for OLBI-LMS according to [7] and parameters for GS-LMS are set according to the last section. On the other hand, we set the sparsity level for each system channel as 80, which is different from[7] that with a very small sparsity level. Then, 100 trials are carried out and then plot the evolution of MSD and sparse support number with respect to iterations in Fig. 5.

Fig. 5(a) shows that GS-LMS and the standard LMS have almost same convergence speed, but GS-LMS is more accurate than the standard LMS. On the other hand, comparing to OLBI-LMS, the convergence speed of GS-LMS is much faster. Fig. 5(b) presents the evolution of SSU for different algorithms. One can clearly find that the proposed GS-LMS algorithm can correctly and quickly converge to the true sparsity support, while the other two algorithms cannot do this task very well: the standard LMS cannot correctly identify the filter sparsity support, while OLBI-LMS converges very slowly.

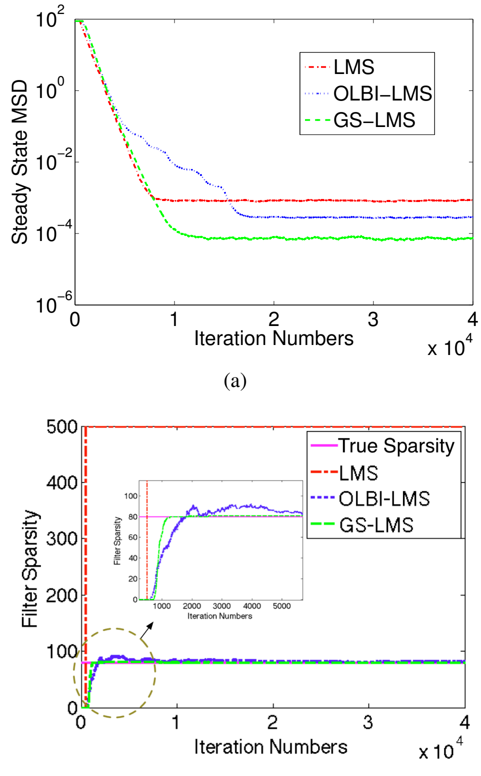


Fig. 5. MSD (a) and recovered filter SSU (b) with respect to iteration numbers.

4.3. Accuracy comparison to state-of-the-art algorithms

In this section, the simulations are carried out to compare the performance with respect to different sparsity levels. Comparisons are made to the state-of-the-art algorithms, such as ZA-LMS [5], RZA-LMS[6], OLBI-LMS[7]. The configuration of the simulation is as follows: the learning rate is set to $\delta = 8 \times 10^{-3}$, the OLBI-LMS threshold is $\lambda = 0.5$, and the GS-LMS thresholds are $\lambda = 0.5, \mu = 1.4$. Then the simulations are carried out with $m = 10$ system channels, each channel is with filter length (order) $n = 100$, and the filters weight vectors sparsity for each channel $\|\mathbf{w}_{k,j}\|_0 (j = 1, 2, \cdots, M)$ varies from 10 to 100, then one can plot the steady state MSD with respect to the filter sparsity for different algorithms, as shown in Fig. 6. From the result, one can clearly find that the proposed GS-LMS method has superior performance (i.e. smallest steady state MSD) to other state-of-the-art algorithms.

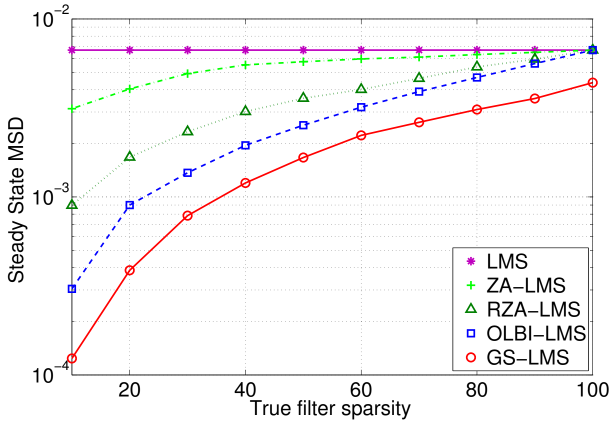


Fig. 6. Comparison to state-of-the-art algorithms with respect to steady state MSD for different filter sparsity.

5. CONCLUSION

This paper proposed an algorithm, namely, GS-LMS that dedicated to the multiple system identification problems. In particular, the sparsity and intra-system correlation are simultaneously considered. The simulation results verify that, when encountering with the multiple system identification problems, the proposed GS-LMS is superior to the-state-of-the-art algorithms, with improved accuracy and high convergence speed. Future works will focus on the theoretical analysis to the proposed GS-LMS and extensions to practical applications, for instance, echo cancellation.

6. REFERENCES

[1] Bernard Widrow and Samuel D Stearns, "Adaptive signal processing," *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 491 p.*, vol. 1, 1985.

[2] Donald L Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[3] W Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, 1995.

[4] Yuantao Gu, Jian Jin, and Shunliang Mei, "l_0 norm constraint lms algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, 2009.

[5] Yilun Chen, Yuantao Gu, and Alfred O Hero, "Sparse lms for system identification," in *ICASSP*. IEEE, 2009, pp. 3125–3128.

[6] Yilun Chen, Yuantao Gu, and Alfred O Hero, "Regularized least-mean-square algorithms," *arXiv preprint arXiv:1012.5066*, 2010.

[7] Tao Hu and Dmitri B Chklovskii, "Sparse lms via online linearized bregman iteration," *arXiv preprint arXiv:1210.0563*, 2012.

[8] L. Yu, J.-P. Barbot, G. Zheng, and H. Sun, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.

[9] Zhilin Zhang and Bhaskar D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[10] L. Yu, H. Sun, G. Zheng, and J.-P. Barbot, "Model based Bayesian compressive sensing via Local Beta Process," *Signal Processing*, vol. 108, no. 3, pp. 259–271, 2015.

[11] Christian R Berger, Shengli Zhou, James C Preisig, and Peter Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1708–1721, 2010.

[12] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the m/eeg inverse problem using accelerated gradient methods," *Physics in medicine and biology*, vol. 57, no. 7, pp. 1937, 2012.

[13] Shai Shalev-Shwartz, *Online Learning: Theory, Algorithms, and Applications*, Ph.D. thesis, Hebrew University, 2007.

[14] Patrick L Combettes and Jean-Christophe Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[15] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, and M. Kowalski, "Time-frequency mixed-norm estimates: sparse m/eeg imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410–422, 2013.