

INVERSE PROBLEMS WITH TIME-FREQUENCY DICTIONARIES AND NON-WHITE GAUSSIAN NOISE

Matthieu Kowalski^{*†*}, Alexandre Gramfort^{‡*}

^{*} Laboratoire des Signaux et Systèmes (Univ Paris-Sud, CNRS, CentraleSupélec) Gif-sur-Yvette, France

[†] INRIA, Saclay-Île-de-France, Parietal team, France

[‡] Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France

^{*} Neurospin bât 145, CEA, DSV, I²BM, Gif-Sur-Yvette, France

ABSTRACT

Sparse regressions to solve ill-posed inverse problems have been massively investigated over the last decade. Yet, when noise is present in the model, it is almost exclusively considered as Gaussian and white. While this assumption can hold in practice, it rarely holds when observations are time series as they are corrupted by auto-correlated and colored noise. In this work we study sparse regression under the assumption of non white Gaussian noise and explain how to run the inference using proximal gradient methods. We investigate an application in brain imaging: the problem of source localization using magneto- and electroencephalography which allow functional brain imaging with high temporal resolution. We use a time-frequency representation of the source waveforms and a sparse regularization which promotes focal sources with smooth and transient activations. Our approach is evaluated using simulations comparing it to strategies that assume the noise is white or to simple prewhitening.

Index Terms— Deconvolution, Inverse Problem, Time-Frequency Whitening, Non white Gaussian Noise, Sparse recovery, Denoising

1. INTRODUCTION

Sparse regression consists in estimating unknown coefficients in a linear regression framework under the assumption that only a few regressors, also called variables or features, are predictive. In a word, most features are non-informative and their associated coefficients should be zero. The forward problem we consider in this work is the one proposed in [1] in the context of M/EEG. MEG and EEG are brain imaging methods that record the electromagnetic signals produced by active neurons using an array of sensors. The ambition is to use M/EEG to localize active brain regions non-invasively,

This work benefited from the support of the "FMJH Program Gaspard Monge in optimization and operation research", and from the support to this program from EDF. This work was supported by the ANR grant THALAM-EEG, ANR-14-NEUC-0002-01.

and therefore obtain images of the active brain at a millisecond time-scale, which is particularly useful for clinical and cognitive neuroscience.

The M/EEG data are multivariate time series. The model reads:

$$M = GX + n \quad (1)$$

where

- $M \in \mathbb{R}^{NT}$ is the measurement matrix (N signals of length T);
- $G \in \mathbb{R}^{NP}$ is the known instantaneous mixing matrix, a.k.a. lead field matrix or gain matrix obtained by solving numerically Maxwell's equations;
- $X \in \mathbb{R}^{PT}$ is the unknown matrix containing the neuronal source waveforms (P signals of length T);
- $n \in \mathbb{R}^{NT}$ is the noise, supposed to be independent in space, i.e. across lines, with each line being distributed according a (possibly correlated) normal Law.

The rest of the paper is organized as follows. In next Section 2 we expose a brief state of the art to deal with non white Gaussian noise. Section 3 exposes an efficient time-frequency whitening strategy to take into account the non white noise in the model. In Section 4 we perform a numerical comparison of five strategies to deal with the non white noise. finally, Section 5 concludes the paper.

2. STATE OF THE ART

The state of the art of M/EEG inverse problem with white Gaussian noise using a Lasso/Basis pursuit denoising [2, 3] approach reads [1]:

$$\min_Z \frac{1}{2} (M - GZ\Phi^*)^* \Sigma^{-1} (M - GZ\Phi^*) + \lambda \|Z\|_1, \quad (2)$$

where $\|Z\|_1$ stands for the ℓ_1 norm for the flattened matrix Z , $\Phi \in \mathbb{C}^{TQ}$ is a matrix of Q time-frequency atoms and $Z \in \mathbb{C}^{PQ}$ are the corresponding synthesis coefficients. The dimension Q is the product of the number of time bins Q_t and frequency bins Q_f . Estimation of the sources are then

given by resynthesis $X = Z\Phi^*$. $\|Z\|_1$ may eventually be replaced by a mixed-norm or a combination of mixed-norm and ℓ_1 norm [1]. For the sake of simplicity and clarity, we stick here to the ℓ_1 norm.

A classic approach to deal with non white Gaussian noise, is to use the variance-covariance matrix Σ of the noise in the data-fidelity term:

$$\min_Z \frac{1}{2}(M - GZ\Phi^*)^* \Sigma^{-1} (M - GZ\Phi^*) + \lambda \|Z\|_1. \quad (3)$$

However, this approach supposes to have a nice estimation of Σ^{-1} and depends of the length T . Here, we use the fact that any Gaussian noise $n \sim \mathcal{N}(0, \Sigma)$ (with autocorrelation in $\ell^1(\mathbb{Z})$) can be obtained by filtering a white Gaussian noise $n = w \star \tilde{n}$, where $\tilde{n} \sim \mathcal{N}(0, 1)$. Then w is a whitening filter such that $w^{-1} \star n \sim \mathcal{N}(0, 1)$, and the whitening process can be done directly on the Fourier domain by “flattening” the power spectrum.

Another classic approach is to pre-whiten the data M [4–6]. Denoting by $\tilde{M} = M \star w^{-1}$ such a pre-whitening matrix, one solves

$$\min_Z \frac{1}{2} \|\tilde{M} - GZ\Phi^*\|^2 + \lambda \|Z\|_1. \quad (4)$$

Then, the estimated sources are given by $X = Z\Phi^* \star w$. While this approach is useless with a ℓ_2 regularizer (the minimizer remaining the same as in the original problem), it is justified when using an ℓ_1 regularization thanks to the thresholding step: pre-whiten the data M enhances the frequencies which are the less corrupted by the non-white noise.

3. MODEL AND ALGORITHM

We present here two approaches in order to take the non-white noise into account. The first one uses the “true” direct model, supposing that one knows perfectly the power spectrum of the noise. The second one is more realistic to practical situation, where the power spectrum of the noise can only be estimated on the data.

3.1. True whitening in the frequency domain

Supposing one has a whitening filter w^{-1} such that the noise satisfies $n = w^{-1} \star \tilde{n}$ with $\tilde{n} \sim \mathcal{N}(0, 1)$. One can rewrite the forward model (1)

$$\begin{aligned} M &= GX + n \\ &= GX + w \star \tilde{n} \\ M \star w^{-1} &= GX \star w^{-1} + \tilde{n} \end{aligned} \quad (5)$$

where the filtering operations $\bullet \star w^{-1}$ are always applied to each rows of the matrices. Then, if one knows the whitening filter w^{-1} , the non-white forward model can be written as a classic forward model corrupted by a white noise. However,

we will see that only the spectrum of w^{-1} has to be known in practice. Indeed, using Equation (5), the sparse approach to estimate X reads

$$\min_Z \frac{1}{2} \|M \star w^{-1} - GZ\Phi^* \star w^{-1}\|^2 + \lambda \|Z\|_1, \quad (6)$$

which can be written in the Fourier domain

$$\min_Z \frac{1}{2} \|(M \star w^{-1} - GZ\Phi^* \star w^{-1})F\|^2 + \lambda \|Z\|_1, \quad (7)$$

where F is the orthogonal complex valued matrix corresponding to the Fourier transform.

Minimizing (7) using a first order optimization scheme such as (F)ISTA [7, 8] requires to compute the gradient of the data term, which is the smooth term,

$$\frac{1}{2} \|(M \star w^{-1} - GZ\Phi^* \star w^{-1})F\|^2 = \frac{1}{2} \|(\hat{M} - GZ\Phi^* F) \otimes \hat{w}^{-1}\|^2, \quad (8)$$

where \hat{M} denotes the Fourier transform of each of the rows of M and $\hat{M} \otimes \hat{w}^{-1}$ denotes the element-wise multiplication of each row of \hat{M} by \hat{w}^{-1} . The gradient step is then followed by the soft-thresholding operator, which corresponds to the proximal operator of the ℓ_1 norm. The gradient step on the data term reads:

$$\begin{aligned} Z^{k+1/2} &= Z^k \\ &+ \frac{1}{L} G^T \left((\hat{M} - GZ^k \Phi^* F) \otimes |\hat{w}^{-1}|^2 \right) F^* \Phi \end{aligned} \quad (9)$$

Interestingly, one can see that only the power spectrum of w^{-1} appears, without any phase information. The scalar $L > 0$ depends on the Lipschitz constant of the data fidelity term and can be upper bounded by $\|\Phi\|^2 \|G\|^2 \|\hat{w}^{-1}\|_\infty^2$. Using FISTA [8], this whitening approach in the Fourier domain is summarized in Algorithm 1. We choose here the relaxation given in [9], which ensures the convergence of the iterates.

Algorithm 1: FISTA with true whitening in the Fourier domain

Input: $Z^{(0)}, Y^{(0)} \in \mathbb{R}^N, |\hat{w}^{-1}|^2$,
 $L = \|\Phi\|^2 \|G\|^2 \|\hat{w}^{-1}\|_\infty^2$

repeat

$$\begin{aligned} Z^{k+1/2} &= \\ Y^k + \frac{1}{L} G^T \left((\hat{M} - GY^k \Phi^* F) \otimes |\hat{w}^{-1}|^2 \right) F^* \Phi; \\ Z^{k+1} &= Z^{k+1/2} \left(1 - \frac{\lambda/L}{|Z^{k+1/2}|} \right)^+; \\ Y^{(k+1)} &= Z^{(k+1)} + \frac{k+4}{k+5} (Z^{(k+1)} - Z^{(k)}); \end{aligned}$$

until convergence;

3.2. Whitening in the time-frequency domain

A practical challenge is to obtain an estimation of the noise spectrum. In practice, one can use a Welch estimator [10] to

obtain an estimate of the spectrum, which depends of the size of the analysis window. Using the same window as in Φ , we denote by $|\tilde{w}|^2$ the Welch estimator of the noise spectrum. We use here a classical approximation widely use in audio-source separation [11, 12]. That is, for a signal $x \in \mathbb{R}^T$ and a convolution kernel $w \in \mathbb{R}^T$, denoting by φ_{kf} the time-frequency atoms constituting Φ , the convolution can be approximated in the time-frequency domain by the complex valued product

$$\langle w \star x, \varphi_{kf} \rangle \simeq \hat{w}[f] \langle x, \varphi_{kf} \rangle. \quad (10)$$

This approximation is studied in [13, Lemma 1]. One could be tempted to use this approximation in order to simplify the gradient evaluation (9). However, using such an approximation in ISTA can lead to a divergent algorithm, as the error could be not summable [7].

We then propose to consider the following function, where the convolution is directly approximated in the time-frequency domain

$$\min_{\tilde{Z}} \frac{1}{2} \|M\Phi \otimes \tilde{w}^{-1}\Phi^* - G(Z \otimes \tilde{w}^{-1})\Phi^*\|^2 + \lambda \|Z\|_1. \quad (11)$$

Then, using the changes of variables $\tilde{M} = M\Phi \otimes \tilde{w}^{-1}\Phi^*$ and $\tilde{Z} = Z \otimes \tilde{w}^{-1}$, one has

$$\min_{\tilde{Z}} \frac{1}{2} \|\tilde{M} - G\tilde{Z}\Phi^*\|^2 + \lambda \|\tilde{Z}\|_{\tilde{w};1}, \quad (12)$$

with $\|\tilde{Z}\|_{\tilde{w};1} = \sum_{kf} w_f |\tilde{Z}_{kf}|$. Minimization Problem (11) reduces then to a simple Basis-Pursuit Denoising with a weighted ℓ_1 norm. The gradient step reduces to a classic gradient descent as with the white noise problem (1), with a data matrix \tilde{M} prewhitened in the time-frequency domain. The proximal step is a soft-thresholding operation, depending on the amount of the noise in each frequency bin. Using FISTA to minimize (12) gives Algorithm 2. In practice $\tilde{w}^{-1} = (1/\sqrt{(|w|^2)_1}, \dots, 1/\sqrt{(|w|^2)_{Q_f}})$, $|w|^2$ being the noise power spectrum estimated by the Welch estimator [10]. The sources are then re-synthesized in the time domain such that $X = (Z \otimes \tilde{w})\Phi^*$

Algorithm 2: FISTA with time-frequency whitening

Input: $Z^{(0)}, Y^{(0)} \in \mathbb{R}^N$, \tilde{w}^{-1} , $L = \|\Phi\|^2 \|G\|^2$

repeat

$$Z^{k+1/2} = Y^{(k)} + \frac{1}{L} G^T \left(\tilde{M} - G Y^{(k)} \Phi^* \right) \Phi;$$

$$Z^{k+1} = Z^{k+1/2} \left(1 - \frac{\tilde{w}\lambda/L}{|Z^{k+1/2}|} \right)^+$$

$$Y^{(k+1)} = Z^{(k+1)} + \frac{k+4}{k+5} (Z^{(k+1)} - Z^{(k)})^+;$$

until convergence;

Notice that, when initialized by 0, the first step of this algorithm can be seen as a classic spectral subtraction [14].

4. NUMERICAL EXPERIMENTS

Application context. In the present work, we explore the effect of non white Gaussian noise in the context of functional brain imaging using magnetoencephalography (MEG) and electroencephalography (EEG) signals. Source imaging with MEG and EEG delivers insights into the active brain at a millisecond time scale in a non-invasive way. To achieve this, one needs to solve the bioelectromagnetic inverse problem. It is a high dimensional ill-posed regression problem which requires proper regularization. As it is natural to assume that a limited set of brain foci are active during a cognitive task, sparse focal source configurations are commonly promoted using convex sparse priors [15, 16].

For the problem at stake the noise is known to be non-white with a signal-to-noise ratio (SNR) that depends on the frequency [17]. The amplitude of power spectrum of the noise is commonly modeled as a “1/f” function: the energy of the noise decreases as the inverse of the frequency.

Method comparison setup. We propose to compare five strategies in order to solve the M/EEG inverse problem in the presence of pink noise:

1. A “True whitening” in the Fourier domain, using the true noise spectrum in $1/f$. This corresponds to Algorithm 1. Method is referred to as *TrueWhite* in the Figures and provide an oracle result in order to compare other approaches.
2. An Estimated whitening in the Fourier domain, using the Welch estimator of the noise spectrum. This corresponds to Algorithm 2. Method is referred to as *TFWhiteEst* in the Figures.
3. A “True Pre-Whitening” in the Fourier domain. The matrix M is whitened in the Fourier domain, and a classic ℓ_2 data term is used in the cost function. Method is referred to as *PreWhiteTrue* in the Figures and provide an oracle results for the prewhitening approaches.
4. An Estimated Pre-Whitening in the Fourier domain. Method is referred to as *PreWhiteEst* in the Figures.
5. A baseline approach without any (pre)whitening process. Method is referred to as *L1* in the Figures.

Simulation setup. We simulated the signals such that $T = 1024$ samples, Φ is a Gabor matrix with Hann window of length 64 samples and a hop-size of 16 samples. The matrix $G \in \mathbb{R}^{30 \times 200}$ was generated randomly with highly correlated columns such that correlation between column i and j is $0.95^{|i-j|}$. Z has 10 non-zeros lines, each non zero line is a random sparse coefficient vector (from 70% to 90% of zeros coefficients), and n is a Pink noise (noise spectrum in $1/f$).

Results are summarized on Figure 1. For each method the quality of the estimate is evaluated using the reconstruction error in dB. As the ground truth is here known and the discussion on the setting of the regularization parameter is outside of the scope of this paper, for each method the regularization parameter yielding the smallest reconstruction error

was used. One can observe that all strategies taking into account the color of the noise outperform the $L1$ method that just assumes an additive white noise. As expected, the “true whitening” approach outperforms all the other approach, but is unusable in practice, as the true spectrum of the noise is unattainable. This justifies its use as an oracle. The proposed strategy $TFWhiteEst$ yields a systematic improvement that is the highest when the SNR is low and Z is the less sparse. This observation is particularly relevant for the application at stake where the SNR is low and the number of sources can increase with complexity of the cognitive experiments. More interestingly, it still outperforms the “true pre-whitening” approach, and justifies to take into account the profile of the noise in the direct problem. Finally, the simple approach without any (pre-)whitening process gives the the worst results as expected.

5. CONCLUSION

The contribution of this work is to propose a principled approach to take into account colored Gaussian noise in sparse regressions solved using first order optimization methods. The method is particularly adapted to inverse problems where data are time series, as they are generally corrupted by an autocorrelated noise. The approach proposed is tested on a neuroscience application. Using simulated MEG data and realistic noise models our results demonstrate the benefit of the approach.

Future work will be on the extension of this contribution to sparse structured regularizations, beyond ℓ_1 , as well as non-convex regularizations, such as SCAD [18] or empirical Wiener [19], for which proximal operators exist.

REFERENCES

- [1] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämmäläinen, and M. Kowalski, “Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations,” *NeuroImage*, vol. 70, pp. 410–422, 2013.
- [2] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [3] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Serie B*, vol. 58, no. 1, pp. 267–288, 1996.
- [4] J. da Costa, K. Liu, H. C. So, M. Roemer, and S. Schwarz, “Generalized multidimensional prewhitening for enhanced signal reconstruction and parameter estimation,” *Signal Processing*, 2013.
- [5] A.-K. Seghouane and Y. Saad, “Prewhitening high-dimensional fMRI data sets without eigendecomposition,” *Neural computation*, vol. 26, no. 5, pp. 907–919, 2014.

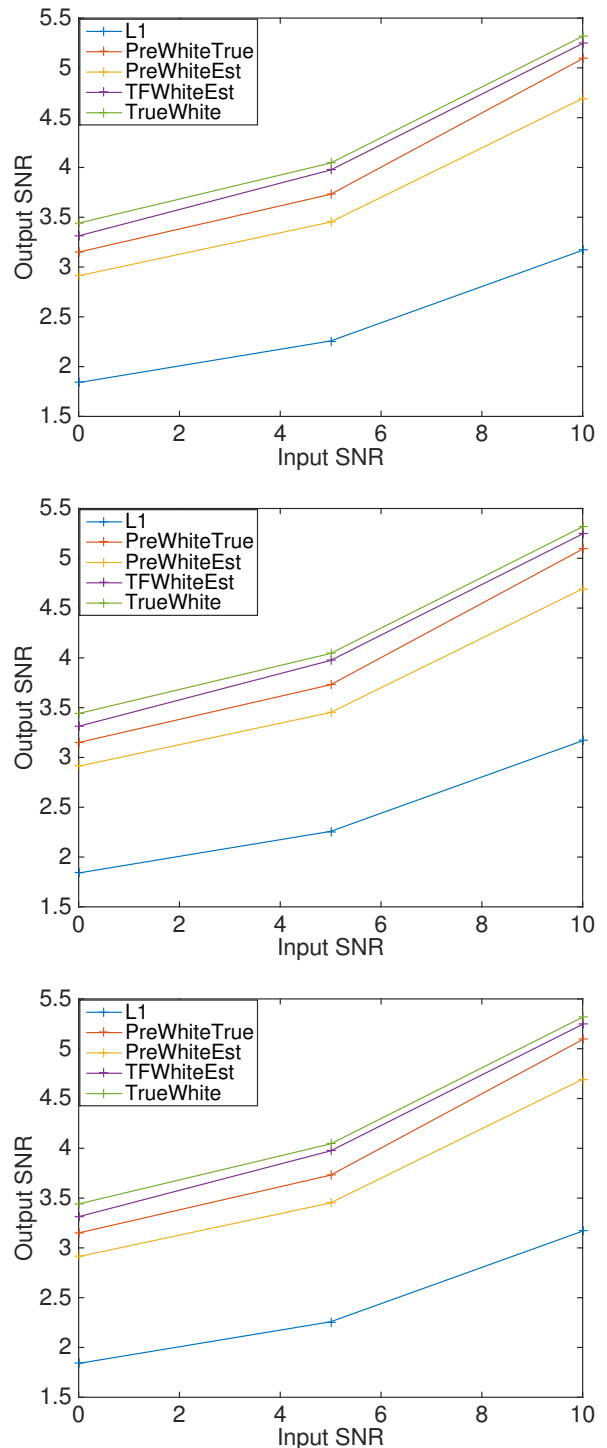


Fig. 1. Comparison of the five methods. Up to bottom: from the less sparse signal to the sparsest (resp. 70%, 80% and 90% of zero coefficients on each non-zeros lines). One can observe that the baseline approach that assumes Gaussian white noise is outperformed by all methods that attempts at modeling the noise color. The oracle approach that is informed by the true noise spectrum offers the best performance closely followed by the method that has only access to the estimated noise spectrum.

- [6] A. S. Dalalyan, M. Hebiri, K. Meziani, and J. Salmon, "Learning heteroscedastic models by convex programming under group sparsity," in *Proc. ICML*, 2013.
- [7] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.
- [8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [9] A. Chambolle and C. Dossal, "On the convergence of the iterates of fista." *Preprint hal-01060130*, September, 2014.
- [10] P. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. 15, pp. 70–73, 1967.
- [11] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and ℓ_1 -norm minimization," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007, article ID 24717.
- [12] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1592–1604, July 2007.
- [13] M. Kowalski, E. Vincent, and R. Gribonval, "Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1818–1829, 2010.
- [14] K. Lebart, J.-M. Boucher, and P. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 359–366, 2001.
- [15] S. Haufe, V. V. Nikulin, A. Ziehe, K.-R. Müller, and G. Nolte, "Combining sparsity and rotational invariance in EEG/MEG source reconstruction," *NeuroImage*, vol. 42, no. 2, pp. 726–738, Aug. 2008.
- [16] A. Gramfort, M. Kowalski, and M. S. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods." *Phys. Med. Biol.*, vol. 57, no. 7, pp. 1937–1961, Apr. 2012.
- [17] R. R. Ramirez, B. H. Kopell, C. R. Butson, B. C. Hiner, and S. Baillet, "Spectral signal space projection algorithm for frequency domain meg and eeg denoising, whitening, and source imaging," *NeuroImage*, vol. 56, no. 1, pp. 78 – 92, 2011.
- [18] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [19] K. Siedenburg, "Persistent empirical wiener estimation with adaptive threshold selection for audio denoising," in *Proc. of the 9th Sound and Music Computing Conference*, 2012, pp. 426–433.