Electromagnetic neural source imaging under sparsity constraints with SURE-based hyperparameter tuning

Pierre-Antoine Bannier, Quentin Bertrand, Joseph Salmon, Alexandre Gramfort

Electromagnetic neural source imaging under sparsity constraints with SURE-based hyperparameter tuning

Pierre-Antoine Bannier
Université Paris Saclay, Inria, CEA
Palaiseau, 91120, France
pierreantoine.bannier@gmail.com

Quentin Bertrand
Université Paris Saclay, Inria, CEA
Palaiseau, 91120, France
quentin.bertrand@inria.fr

Joseph Salmon
IMAG, Univ. Montpellier, CNRS
Institut Universitaire de France (IUF)
Montpellier, France
joseph.salmon@umontpellier.fr

Alexandre Gramfort
Université Paris Saclay, Inria, CEA
Palaiseau, 91120, France
alexandre.gramfort@inria.fr

Abstract

Estimators based on non-convex sparsity-promoting penalties were shown to yield state-of-the-art solutions to the magneto-/electroencephalography (M/EEG) brain source localization problem. In this paper we tackle the model selection problem of these estimators: we propose to use a proxy of the Stein's Unbiased Risk Estimator (SURE) to automatically select their regularization parameters. The effectiveness of the method is demonstrated on realistic simulations and 30 subjects from the Cam-CAN dataset. To our knowledge, this is the first time that sparsity promoting estimators are automatically calibrated at such a scale. Results show that the proposed SURE approach outperforms cross-validation strategies and state-of-the-art Bayesian statistics methods both computationally and statistically.

1 Introduction

Magneto- and electroencephalography (M/EEG, [6, 13]) are non-invasive technologies tailored for monitoring electrical brain activity with a high temporal resolution (milliseconds). Yet, reconstructing the spatial cortical current density at the origin of M/EEG data is a challenging high-dimensional ill-posed linear inverse problem [16, 3]. The M/EEG inverse problem is typically addressed using Lasso-type estimators [42], and more precisely group-Lasso penalties [31, 23]. The latter approaches can be refined with non-convex penalties [11, 40] that exhibit several advantages: they yield sparser physiologically-plausible solutions, mitigate the intrinsic Lasso amplitude bias, and rely on iterative convex optimization problems which can be solved efficiently with coordinate descent [44, 37, 28].

The major practical bottleneck of these techniques remains the calibration of the regularization parameter, *i.e.,* the parameter trading the data-fitting term against the sparsity-promoting prior. State-of-the-art hyperparameter selection techniques for the M/EEG source localization problem include hierarchical Bayesian modelling [30, 35] and hyperparameter optimization (HO) [14, 26, 22]. Hierarchical Bayesian approaches require to specify a prior distribution on the regularization hyperparameter. Regression coefficients and the regularization parameter can then be inferred with multiple techniques [43, 32, 4, 45]. The idea of HO is to select the regularization parameter such that the regression coefficients minimize a given criterion. This then boils down to a bilevel optimization problem can then be solved using zero-order methods [7, 27, 1] or first-order methods [21, 8, 9]. Popular statistical criteria include K-fold cross-validation (folds are created across sensors [15]),

which is however not well-suited for the M/EEG inverse problem: samples are not *i.i.d.* due to spatially-correlated sensors [16]. Therefore, the crucial question revolves around finding a criterion to properly identify the neural generators at the origin of the observed signal. Stein [38] proposed an unbiased estimator of the quadratic risk of estimators: Stein's Unbiased Risk Estimator (SURE), which has proved to be well-suited for inverse problem hyperparameter selection [10, 34, 33].

In this paper, we combine reweighting techniques with a SURE-based model selection to identify active brain sources. The main contributions of this paper are as follows:

1. We combined the SURE criterion with a non-convex estimator. This yields a fast and parameter-free approach to select the regularization parameter for the `irMxNE` algorithm.
2. With experiments on more than 30 subjects from the Cam-CAN biobank [41], we extensively show that the proposed SURE-based hyperparameter selection technique for `irMxNE` achieves state-of-the-art results on real M/EEG data.
3. Code is available at https://github.com/PABannier/automatic_hp_selection_for_meg and already disseminated via the popular brain imaging package MNE [24].

Notation. The Frobenius norm of $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_F$. For any integer $d \in \mathbb{N}$, we denote by $[d]$ the set $\{1, \ldots, d\}$. With M/EEG data, N is the number of sensors, T the number of time instants of the measurements, and S the number of source locations positioned on the cortical mantle.

2   Method

Model. The M/EEG source localization problem can be cast as a high-dimensional inverse problem:

$$\mathbf{M} = \mathbf{G}\mathbf{X}^* + \mathbf{E} \ , \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{N \times T}$ is a measurement matrix, $\mathbf{G} \in \mathbb{R}^{N \times P}$ is the (known) design matrix, $\mathbf{X}^* \in \mathbb{R}^{P \times T}$ is the unknown true regression parameters, $\mathbf{E} \in \mathbb{R}^{N \times T}$ is the noise matrix, which is assumed Gaussian *i.i.d.* : $\mathbf{E}_{i,t} \sim \mathcal{N}(0,1), \forall i \in [N], \forall t \in [T]$. The coefficient matrix $\mathbf{X}^* \in \mathbb{R}^{P \times T}$ is to be recovered from the observation of $\mathbf{M}$ and $\mathbf{G}$. While simpler models assume the source orientations are normal to the cortical mesh, we rely on *free orientation* models, where amplitudes and orientations are jointly inferred. It boils down to reconstructing the amplitudes of three orthogonal sources at each spatial location: $P = 3S$ and $\mathbf{X} \in \mathbb{R}^{3S \times T}$ is partitioned into S blocks $\mathbf{X}_s \in \mathbb{R}^{3 \times T}$, where each $\mathbf{X}_s$ is the block corresponding to the sth source location [25].

Estimator. Given a regularization parameter $\lambda > 0$, the `irMxNE` optimization problem reads:

$$\widehat{\mathbf{X}}^{(\lambda)} \in \underset{\mathbf{X} \in \mathbb{R}^{P \times T}}{\arg\min} \frac{1}{2}\|\mathbf{M} - \mathbf{G}\mathbf{X}\|_F^2 + \lambda \sum_{s=1}^{S} \sqrt{\|\mathbf{X}_s\|_F} \ . \tag{2}$$

Non-convex problem (2) is solved by iterativaly solving convex problems [11, 40], see Alg. 1.

Model selection. The parameter λ is often chosen such that the corresponding regression coefficients $\widehat{\mathbf{X}}^{(\lambda)}$ (in Eq. (2)) minimize a given criterion. [38] proposed a criterion to avoid overcomplex models:

$$\text{SURE}(\widehat{\mathbf{X}}^{(\lambda)}) = -NT\sigma^2 + \|\mathbf{M} - \mathbf{G}\widehat{\mathbf{X}}^{(\lambda)}\|_F^2 + 2\sigma^2 \text{dof}(\widehat{\mathbf{X}}^{(\lambda)}) \ , \tag{3}$$

where $\text{dof}(\widehat{\mathbf{X}})$ is the degrees of freedom of $\widehat{\mathbf{X}}$ [18] and σ is the true noise level of the measurements. The non-convex penalty used in Eq. (2) prevents the derivation of a closed-form formula for the `irMxNE` degree of freedom. To circumvent this issue, multiple SURE proxies have been proposed [34]. In this work we use the Finite-Difference Monte-Carlo SURE (FDMC SURE, [17], Alg. 2). Due to the non-convexity of the problem, the dof term can only be evaluated numerically [29].

3   Experiments

Several strategies have been proposed to calibrate the regularization parameter in Eq. (2): SURE (proposed, Alg. 2 in App. B.1), spatial cross-validation [39], and hierarchical Bayesian model with a Gamma hyperprior on λ (λ-MAP, [4]). It is worth mentioning that λ-MAP is not a fully automatic method: the default value of the Gamma hyperprior parameter β given in [5] yields poor results in our experiments. We had to handtune $\beta = 10$ for simulated data, and $\beta = 5$ for real data. Data is assumed to be whitened, therefore the true noise level σ is assumed to be known and equal to 1 [19].
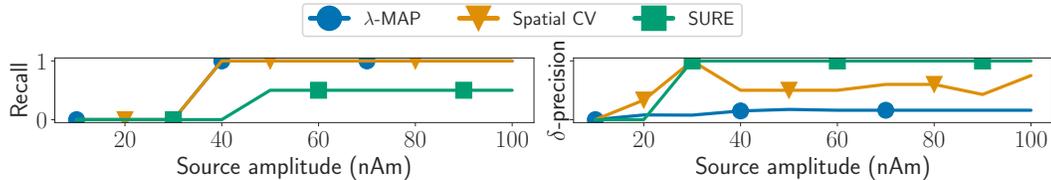
Figure 1 – **Simulated data, statistics on the active set recovery.** Statistics for each model selection procedure on the reweighted Lasso, fitted on a simulated setup from the left auditory task.

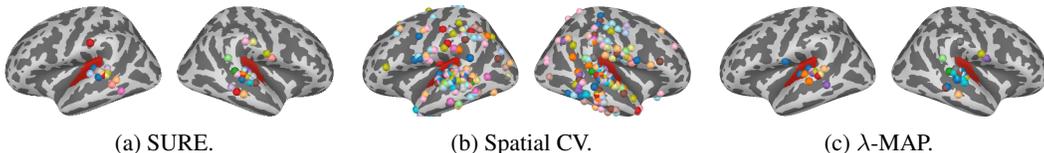(a) SURE. (b) Spatial CV. (c) λ-MAP.

Figure 2 – **Real data, brain source locations aggregated across subjects.** Brain source locations reconstructed in each hemisphere after an auditory stimulation for each model selection procedure.

Experiments on simulated data (Fig. 1). We compared the robustness of each hyperparameter selection technique in various signal-to-noise ratio regimes. We simulated two sources, one in each auditory cortex and varied their amplitude. We computed δ-statistics [12] on the recovery of the active set, and chose an extent with 7 mm of radius. SURE correctly reconstructs the active sources (1 of δ-precision) but fails at identifying all of them. Since both spatial CV and λ-MAP tend to overfit the data, they predict larger supports than expected and yield high recall values.

Experiments on real data (Tab. 1 and Fig. 2). We compared hyperparameter selection for the regularization parameter λ of the reweighted multitask Lasso (Eq. (2)) on 30 subjects of the Cam-CAN biobank dataset [41]. Data consists in the recording of $N = 306$ magnetometers and gradiometers after a left auditory stimulation ($T = 71$ time samples). We used Eq. (2) to estimate the active sources among the $P = 24,582$ source candidates. For each hyperparameter selection technique, we obtain active sources for each subject that we represent on an average brain in Fig. 2. Summary statistics of the experiment are provided in Tab. 1. Regarding computational efficiency, λ-MAP, spatial CV and SURE took respectively 8, 1184 and 492 seconds per subproblem for the 30 subjects.

Tab. 1 shows that spatial CV consistently overfits the dataset by choosing a too small regularization hyperparameter. As predicted by theory [36, 2], spatial CV yields overly large supports, most often with more than two sources. λ-MAP provides an average number of recovered sources similar to SURE. Nonetheless, SURE yields a significantly better explained variance.

Fig. 2 shows the reconstructed sources after a left auditory stimulation on 30 subjects, registered on an average template brain. Each dot represents an active source in the brain, each color corresponds to a subject. The areas colored in red are the auditory cortices obtained using a functional atlas [20]. Ideally, we expect one reconstructed source in each auditory cortex, and a low dispersion of the sources across subjects. Spatial CV yields sources all over the brain surface. λ-MAP correctly localizes the sources but fails to recover any sources on almost one third of the subjects (see Tab. 1). SURE always recovers correctly at least one source, and often the correct two sources in both cortices.

Table 1 – Real data, aggregated results.

Average metrics	Spatial CV	SURE	λ-MAP
$\lambda/\lambda_{\max}$	0.30	0.61	0.9
Explained variance	**0.67**	0.32	0.06
# of sources	9.30	**1.44**	1.3
% of zero sources	0	**0**	29.63
% of one source	0	**55.56**	40.74
% of two sources	3.7	**44.44**	18.52
% of > 2 sources	96.3	**0**	11.11

3

4 Broader impact

This paper paves the way to a wider adoption of recent results in machine learning in the context of non-invasive brain source imaging commonly employed in cognitive and clinical neuroscience. Since the code is already available in the `MNE` package, the proposed algorithm could soon be impactful in the neuroimaging community.

Acknowledgments

References

[1] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. 2019.

[2] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.

[3] S. Baillet, J.C. Mosher, and R.M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30, 2001.

[4] Y. Bekhti, R. Badeau, and A. Gramfort. Hyperparameter Estimation in Maximum a Posteriori Regression Using Group Sparsity with an Application to Brain Imaging. *25th European Signal Processing Conference (EUSIPCO)* , pages 256–260, 2017.

[5] Y. Bekhti, F. Lucka, J. Salmon, and A. Gramfort. A hierarchical bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging. *Inverse Problems*, 34(8):085010, jun 2018.

[6] H. Berger. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.

[7] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 2012.

[8] Q. Bertrand, Q. Klopfenstein, M. Blondel, S. Vaiter, A. Gramfort, and J. Salmon. Implicit differentiation of Lasso-type models for hyperparameter optimization. *ICML*, 2020.

[9] Q. Bertrand, Q. Klopfenstein, M. Massias, M. Blondel, S. Vaiter, A. Gramfort, and J. Salmon. Implicit differentiation for fast hyperparameter selection in non-smooth convex learning. *arXiv preprint arXiv:2105.01637*, 2021.

[10] T. Blu and F. Luisier. The SURE-LET approach to image denoising. *IEEE Trans. Image Process.*, 16(11):2778–2786, 2007.

[11] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l_1 minimization. *J. Fourier Anal. Applicat.*, 14(5-6):877–905, 2008.

[12] J.-A. Chevalier, J. Salmon, and B. Thirion. Statistical inference with ensemble of clustered desparsified lasso. *MICCAI*, 2018.

[13] D. Cohen. Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786, 1968.

[14] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256, 2007.

[15] P. Craven and G. Wahba. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. 1979.

[16] A. M. Dale and M. I. Sereno. Improved localizadon of cortical activity by combining eeg and meg with mri cortical surface reconstruction: A linear approach. *Journal of cognitive neuroscience*, 5:162—176, 1993.

[17] C.-A. Deledalle, S. Vaiter, J. Fadili, and G. Peyré. Stein Unbiased GrAdient estimator of the Risk (SUGAR) for multiple parameter selection. *SIAM J. Imaging Sci.*, 7(4):2448–2487, 2014.

[18] B. Efron. How biased is the apparent error rate of a prediction rule? *J. Amer. Statist. Assoc.*, 81(394):461–470, 1986.

[19] D. A. Engemann and A. Gramfort. Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals. *NeuroImage*, 108:328 – 342, 2015.

[20] B. Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[21] L. Franceschi, M. Donini, P. Frasconi, and M. Pontil. Forward and reverse gradient-based hyperparameter optimization. In *ICML*, pages 1165–1173, 2017.

[22] L. Franceschi, P. Frasconi, S. Salzo, and M. Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *ICML*, pages 1563–1572, 2018.

[23] A. Gramfort, M. Kowalski, and M. Hämäläinen. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.*, 57(7):1937–1961, 2012.

[24] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen. MNE software for processing MEG and EEG data. *NeuroImage*, 86:446 – 460, 2014.

[25] 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*, 70:410–422, 2013.

[26] G. Kunapuli. *A bilevel optimization approach to machine learning*. Rensselaer Polytechnic Institute, 2008.

[27] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *Journal of Machine Learning Research*, 2017.

[28] M. Massias, S. Vaiter, A. Gramfort, and J. Salmon. Dual extrapolation for sparse generalized linear models. *Journal of Machine Learning Research*, 2020.

[29] R. Mazumder, J. H. Friedman, and T. J. Hastie. *SparseNet*: coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.*, 106(495):1125–1138, 2011.

[30] R. Molina, A. K. Katsaggelos, and J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans. Image Process.*, 8(2):231–246, 1999.

[31] G. Obozinski, M. J. Wainwright, and M. Jordan. Support union recovery in high-dimensional multivariate regression. *Ann. Statist.*, 39(1):1–47, 2011.

[32] M. Pereyra, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Maximum-a-posteriori estimation with unknown regularisation parameters. *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 230–234, 2015.

[33] J. C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE Approach for Digital Signal/Image Deconvolution Problems. *IEEE Transactions on Signal Processing*, 57:4616–4632, 2009.

[34] S. Ramani, T. Blu, and M. Unser. Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.*, 17(9):1540–1554, 2008.

[35] M. Sato, T. Yoshioka, S. Kajihara, K. Toyama, N. Goda, K. Doya, and M. Kawato. Hierarchical bayesian estimation for meg inverse problem. *NeuroImage*, 23(3):806–826, 2004.

[36] J. Shao. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88(422):486–494, 1993.

[37] H.-J. M. Shi, S. Tu, Y. Xu, and W. Yin. A primer on coordinate descent algorithms. *ArXiv e-prints*, 2016.

[38] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6):1135–1151, 1981.

[39] L. R. A. Stone and J.C. Ramer. Estimating WAIS IQ from Shipley Scale scores: Another cross-validation. *Journal of clinical psychology*, 21(3):297–297, 1965.

[40] D. Strohmeier, Y. Bekhti, J. Haueisen, and A. Gramfort. The iterative reweighted Mixed-Norm Estimate for spatio-temporal MEG/EEG source reconstruction. *IEEE Transactions on Medical Imaging*, 2016.

[41] J. R. Taylor, W. Nitin, C. Rhodri, A. Tibor, M. A. Shafto, Marie Dixon, Lorraine K. Tyler, Cam-CAN, and Richard N. Henson. The cambridge centre for ageing and neuroscience (cam-can) data repository: Structural and functional mri, meg, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144:262–269, 2017.

[42] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 58(1):267–288, 1996.

[43] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, 1:211–244, 2001.

[44] P. Tseng and S. Yun. Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization. *J. Optim. Theory Appl.*, 140(3):513, 2009.

[45] A. F. Vidal, V. De Bortoli, M. Pereyra, and A. Durmus. Maximum likelihood estimation of regularization parameters in high-dimensional inverse problems: An empirical bayesian approach part i: Methodology and experiments. *SIAM Journal on Imaging Sciences*, 13(4):1945–1989, 2020.

A MxNE

We recall that $\mathbf{G} \in \mathbb{R}^{N \times P}$ denotes the design matrix, $\mathbf{M} \in \mathbb{R}^{N \times T}$ is the measurement matrix. Given a regularization parameter $\lambda > 0$, the MxNE optimization problem [23] reads:

$$\widehat{\mathbf{X}}^{(\lambda)} \in \underset{\mathbf{X} \in \mathbb{R}^{P \times T}}{\arg\min} \frac{1}{2} \|\mathbf{M} - \mathbf{G}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda \sum_{s=1}^{S} \|\mathbf{X}_s\|_{\mathrm{F}} \ . \tag{4}$$

It relies on a convex optimization problem that can be solved using block coordinate descent solvers [44, 37, 28].

B Algorithms details

B.1 FDMC SURE

Evaluating the FDMC SURE consists in solving the following bilevel optimization problem [17]:

$$\hat{\lambda} \in \underset{\lambda \in \mathbb{R}}{\arg\min} \frac{1}{2} \|\mathbf{M} - \mathbf{G}\widehat{\mathbf{X}}^{(\lambda,1)}\|_{\mathrm{F}}^2 + \frac{2\sigma^2}{\epsilon} \langle \mathbf{G}(\widehat{\mathbf{X}}^{(\lambda,2)} - \widehat{\mathbf{X}}^{(\lambda,1)}), \boldsymbol{\Delta} \rangle$$

$$\text{s.t. } \widehat{\mathbf{X}}^{(\lambda,1)} \in \underset{\mathbf{X} \in \mathbb{R}^{P \times T}}{\arg\min} \frac{1}{2} \|\mathbf{M} - \mathbf{G}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda \sum_{s=1}^{S} \sqrt{\|\mathbf{X}_s\|_{\mathrm{F}}} \tag{5}$$

$$\widehat{\mathbf{X}}^{(\lambda,2)} \in \underset{\mathbf{X} \in \mathbb{R}^{P \times T}}{\arg\min} \frac{1}{2} \|\mathbf{M} + \epsilon \boldsymbol{\Delta} - \mathbf{G}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda \sum_{s=1}^{S} \sqrt{\|\mathbf{X}_s\|_{\mathrm{F}}} \ ,$$

with $\epsilon > 0$ and $\boldsymbol{\Delta} \in \mathbb{R}^{N \times T}$ a matrix which coefficients are independent and identically distributed from a normal distribution of mean 0 and of variance 1. Since data is assumed to be whitened, σ is set to 1. The finite difference step ϵ is chosen using the heuristic from [17]: $\epsilon = \frac{2\sigma}{N^{0.3}}$. Alg. 1 shows how to solve the inner problem of Eq. (5). The computation of the FDMC SURE can be found in Alg. 2. Below $\odot$ stands for the element-wise multiplication.

Algorithm 1 IRMxNE [40]

input : $\mathbf{G} \in \mathbb{R}^{N \times P}, \mathbf{M} \in \mathbb{R}^{N \times T}, \lambda > 0, K \in$
    $\mathbb{N}$
init : $\widetilde{\mathbf{X}} = \mathbf{0}_{\mathbb{R}^{P \times T}}, \mathbf{W} = \mathbf{I}_P, \mathbf{w} = \mathbf{0}_S$
    $\epsilon = 10^{-8}$
for $k = 1, \dots, K$ **do**
    $\widetilde{\mathbf{G}} = \mathbf{G}\mathbf{W}$
    // Iteratively solve convex problems
    $\widetilde{\mathbf{X}} \leftarrow \mathrm{MxNE}(\widetilde{\mathbf{G}}, \mathbf{M})$ // Using $\widetilde{\mathbf{X}}$ to warm
    start
    $\mathbf{X} = \mathbf{W}\widetilde{\mathbf{X}}$
    $\mathbf{w} = ((2\|\mathbf{X}_s\|_{\mathrm{F}} + \epsilon)^{-1})_{s \in [S]}$
    $\mathbf{W} = \mathrm{diag}(\mathbf{w} \otimes \mathbf{1}_{(3)})$
return $\mathbf{X}$

Algorithm 2 COMPUTE FDMC SURE (adapted from [17])

input : $\mathbf{G} \in \mathbb{R}^{N \times P}, \mathbf{M} \in \mathbb{R}^{N \times T}, \lambda > 0, K \in$
    $\mathbb{N}$
init : $\boldsymbol{\Delta} \in \mathbb{R}^{N \times T}, \epsilon > 0, \sigma > 0$
    $\forall (n,t) \in [n] \times [T], \boldsymbol{\Delta}_{n,t} \sim \mathcal{N}(0,1)$
$\mathbf{X}^{(\lambda,1)} \leftarrow \mathrm{Alg. 1}(\mathbf{G}, \mathbf{M}, \lambda, K)$
$\mathbf{X}^{(\lambda,2)} \leftarrow \mathrm{Alg. 1}(\mathbf{G}, \mathbf{M} + \epsilon\boldsymbol{\Delta}, \lambda, K)$
// Degree of freedom computation
$\mathbf{J} = \mathbf{G}(\mathbf{X}^{(\lambda,2)} - \mathbf{X}^{(\lambda,1)}) \odot \boldsymbol{\Delta}$
// Finite-difference
$\mathrm{dof} = \frac{1}{\epsilon} \sum_{i=1}^{N} \sum_{j=1}^{T} \mathbf{J}_{i,j}$
// FDMC SURE computation
$\mathrm{SURE} = \|\mathbf{M} - \mathbf{G}\mathbf{X}^{(\lambda,1)}\|_{\mathrm{F}}^2 - NT\sigma^2 + 2\sigma^2\mathrm{dof}$
return SURE

Combining both algorithms, we propose in Alg. 3 the procedure to automatically select $\lambda > 0$ for irMxNE.

Algorithm 3 GRID-SEARCH TO SOLVE PROBLEM 5

input : $\mathbf{G} \in \mathbb{R}^{N \times P}, \mathbf{M} \in \mathbb{R}^{N \times T}, \lambda_1, \ldots, \lambda_n > 0, K \in \mathbb{N}$
init : $\text{SURE}_{\text{opt}} = +\infty$
for $\lambda = \lambda_1, \ldots, \lambda_n$ **do**
    $\text{SURE} \leftarrow$ Alg. 2$(\mathbf{G}, \mathbf{M}, \lambda, K)$
    if $\text{SURE} < \text{SURE}_{\text{opt}}$ **then**
        $\lambda_{\text{opt}} = \lambda; \text{SURE}_{\text{opt}} = \text{SURE}$
return λ_{opt}

B.2 Cross-validation

Cross-validation consists in partitioning data $(\mathbf{G}, \mathbf{M})$ into $V \in \mathbb{N}^*$ hold-out datasets $(\mathbf{G}^{\text{train}_v}, \mathbf{M}^{\text{train}_v})$, $v \in [V]$. The matrices are partitioned along the row axis, a setup referred to as spatial cross-validation when dealing with M/EEG data. The regularization parameter λ is then chosen to minimize the averaged squared norm of the errors:

$$\hat{\lambda} \in \underset{\lambda \in \mathbb{R}}{\arg\min} \, \frac{1}{V} \sum_{v=1}^{V} \|\mathbf{M}^{\text{val}_v} - \mathbf{G}^{\text{val}_v} \widehat{\mathbf{X}}^{(\lambda,v)}\|_{\text{F}}^2$$

$$\text{s.t. } \widehat{\mathbf{X}}^{(\lambda,v)} \in \underset{\mathbf{X} \in \mathbb{R}^{P \times T}}{\arg\min} \, \frac{1}{2} \|\mathbf{M}^{\text{train}_v} - \mathbf{G}^{\text{train}_v} \mathbf{X}\|_{\text{F}}^2 + \lambda \sum_{s=1}^{S} \sqrt{\|\mathbf{X}_s\|_{\text{F}}} \quad , \quad \forall v \in [V] \ . \tag{6}$$

B.3 Hierarchical Bayesian modelling

Algorithm 4 λ-MAP [4]

input : $\mathbf{G} \in \mathbb{R}^{N \times P}, \mathbf{M} \in \mathbb{R}^{N \times T}, \lambda^{(0)} \in \mathbb{R}^+, n_{\text{iter}} \in \mathbb{N}^*, \beta > 0, \epsilon > 0$
init : $\lambda_{\max} = \|\mathbf{G}^\top \mathbf{M}\|_{2,\infty}, m = \lambda_{\max}/2, \alpha = m\beta + 1$
for $i = 1, \ldots, n_{\text{iter}}$ **do**
    $\mathbf{X} \leftarrow$ Alg. 1 ($\mathbf{G}, \mathbf{M}, \lambda^{(i-1)}$) `// Solve ixMxNE problem`
    $\lambda^{(i)} = (2ST + \alpha - 1)/(\sum_{s=1}^{S} \sqrt{\|\mathbf{X}_s\|_{\text{F}}} + \beta)$
    if $|\lambda^{(i)} - \lambda^{(i-1)}| < \epsilon$ **then**
        return $\lambda^{(i)}$
return $\lambda^{n_{\text{iter}}}$

For λ-MAP, we had to fine-tune by hand the hyperprior parameter. We set $\beta = 10$ for simulated data, and $\beta = 5$ for real data. While λ-MAP is robust to the initialization of λ, it remains highly dependent of β, making it not fully automatic. Some experiments required an order of magnitude larger β to prevent the iterate scheme to converge over $\lambda_{\max}$. Indeed, we have noticed that λ-MAP often selects a hyperparameter λ larger than $\lambda_{\max}$, due to a poorly-chosen β.

B.4 Warm start

To accelerate the grid-search procedure, we sequentially solve the first convex subproblems (before the first reweighting) and initialize the weights of the i-th Lasso estimator $\widehat{\mathbf{X}}^{(\lambda_i)}$ with $\widehat{\mathbf{X}}^{(\lambda_{i-1})}$. This computational trick is known as warm start (Alg. 5).

Algorithm 5 EFFICIENT WARM START FOR REWEIGHTED LASSO WITH GRID-SEARCH

---

input : $\mathbf{G} \in \mathbb{R}^{N \times P}, \mathbf{M} \in \mathbb{R}^{N \times T}, \lambda_1, \ldots, \lambda_n > 0, K \in \mathbb{N}^*$
init : $\mathbf{X}^{(\lambda_0)} = \mathbf{0}_{N \times T}$
```
// First solve problems without reweighting
```
for $i \in [n]$ **do**
    ```// This leads a more efficient warm start```
    $\mathbf{X}^{(\lambda_i)} \leftarrow$ `MxNE`$(\mathbf{G}, \mathbf{M}, \lambda_i)$ `// Solve MxNE using` $\mathbf{X}^{(\lambda_{i-1})}$ `to warm start`
```
// Then solve the remaining subproblems
```
for $i \in [n]$ **do**
    $\mathbf{X}^{(\lambda_i)} \leftarrow$ `irMxNE`$(\mathbf{G}, \mathbf{M}, \lambda_i, K-1)$ `// Solve irMxNE using` $\mathbf{X}^{(\lambda_i)}$ `to warm start`
return $\mathbf{X}^{(\lambda_1)}, \ldots, \mathbf{X}^{(\lambda_n)}$