

Handling correlated and repeated measurements with the smoothed multivariate square-root Lasso

Quentin Bertrand

Joint work with:

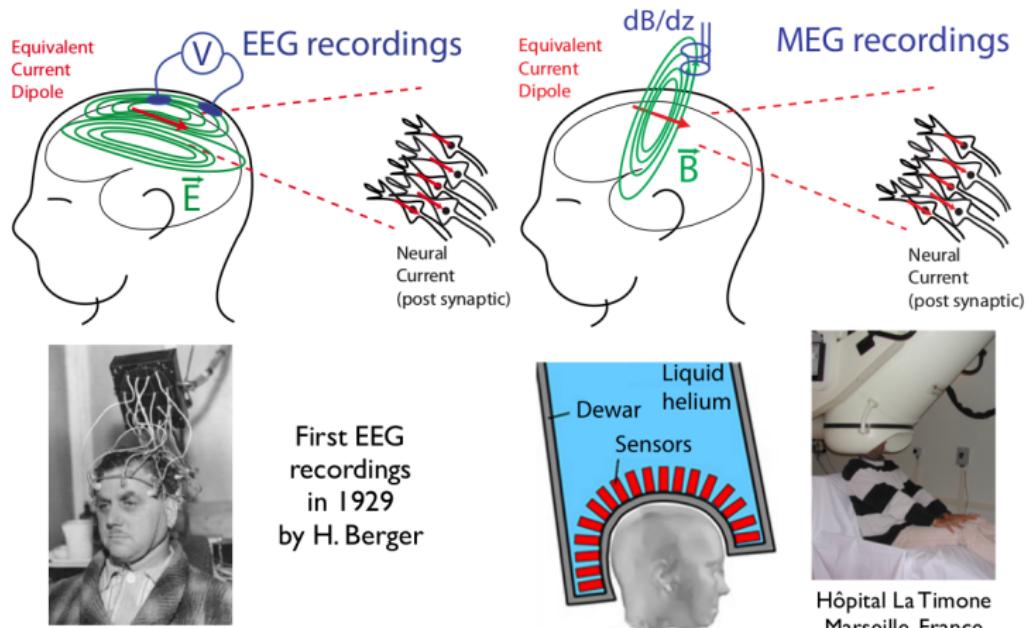
Mathurin Massias (INRIA)

Alexandre Gramfort (INRIA)

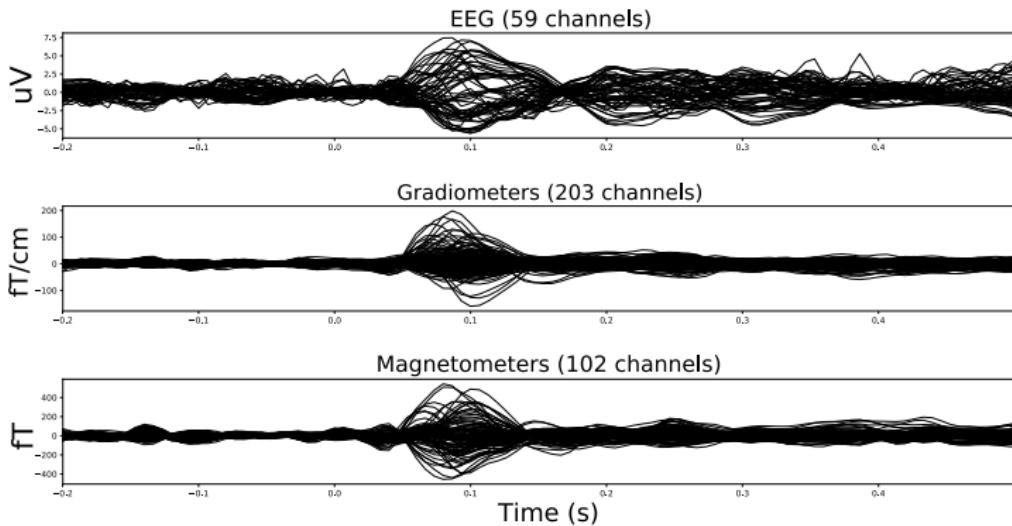
Joseph Salmon (IMAG, Univ Montpellier, CNRS)

M/EEG inverse problem for brain imaging

- ▶ sensors: electric and magnetic fields during a cognitive task
- ▶ goal: which parts of the brain are responsible for the signals?
- ▶ applications: epilepsy treatment, brain aging, anesthesia risks

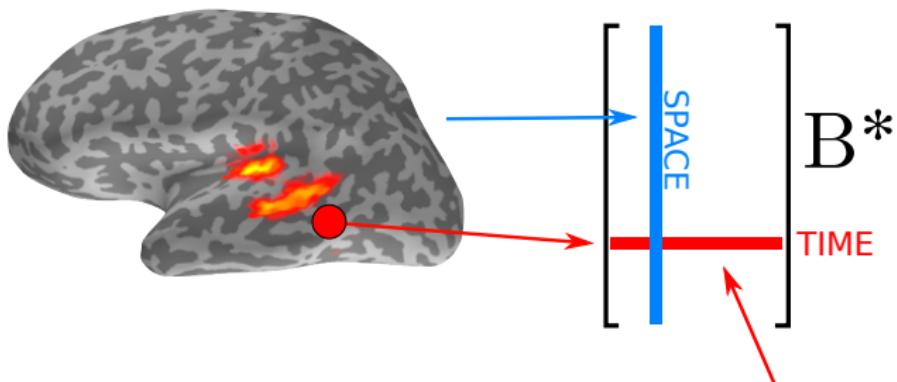


M/EEG data

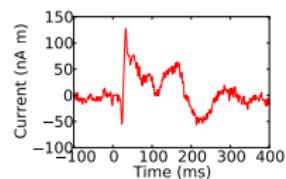


- ▶ 3 different types of sensor

Source modeling (discretization with voxels)

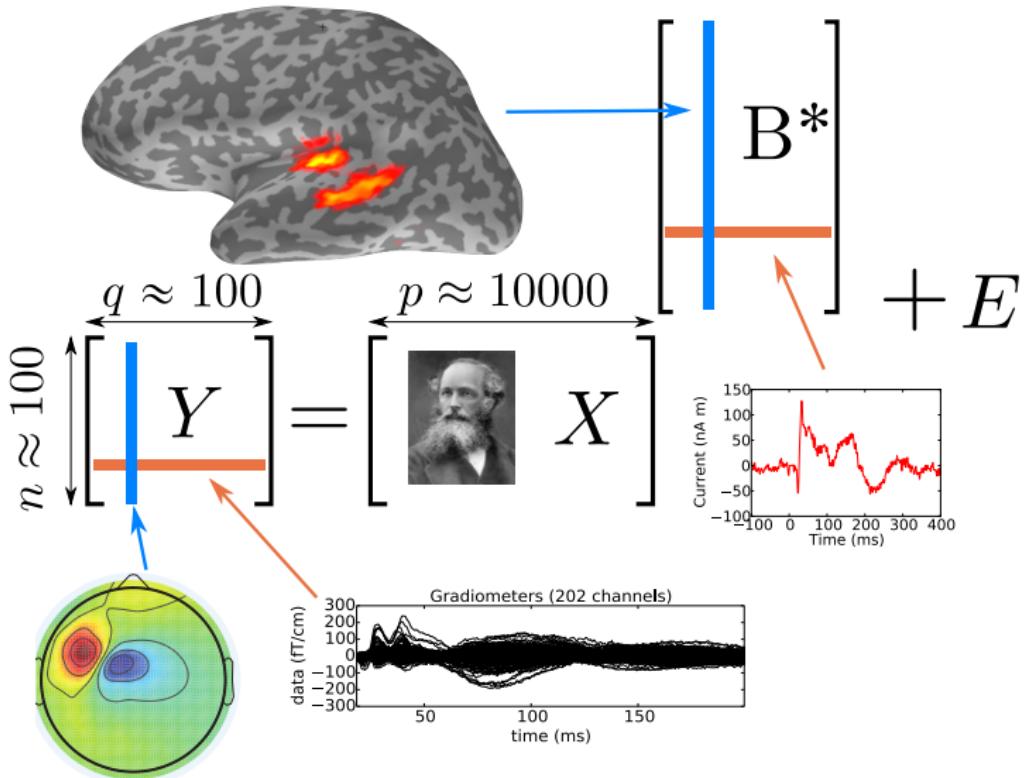


Position a few thousands candidate sources over the brain (e.g., every 5mm)



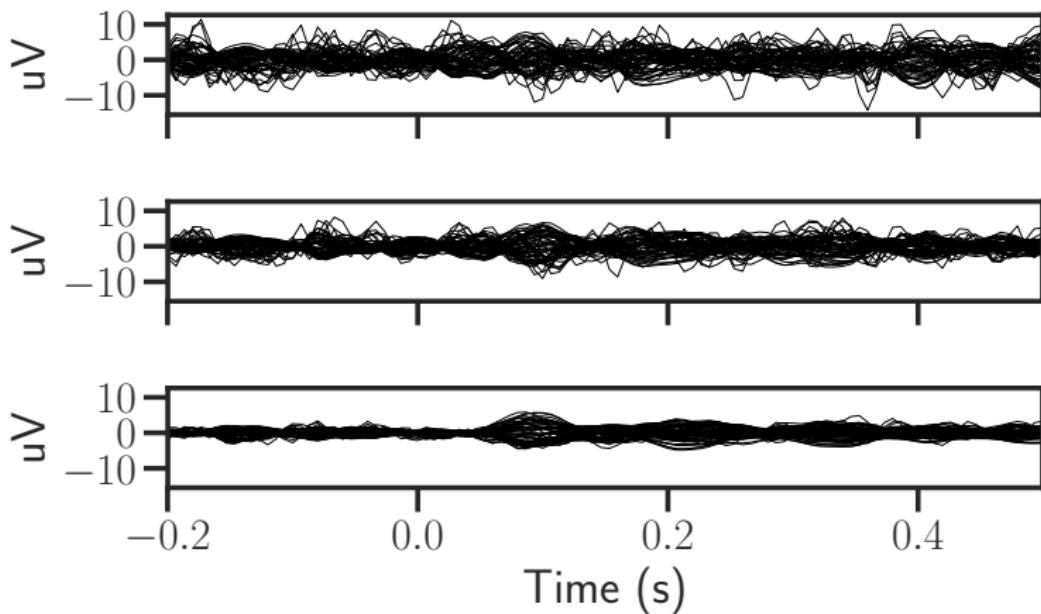
$$B^* \in \mathbb{R}^{p \times q}$$

The M/EEG inverse problem: modeling

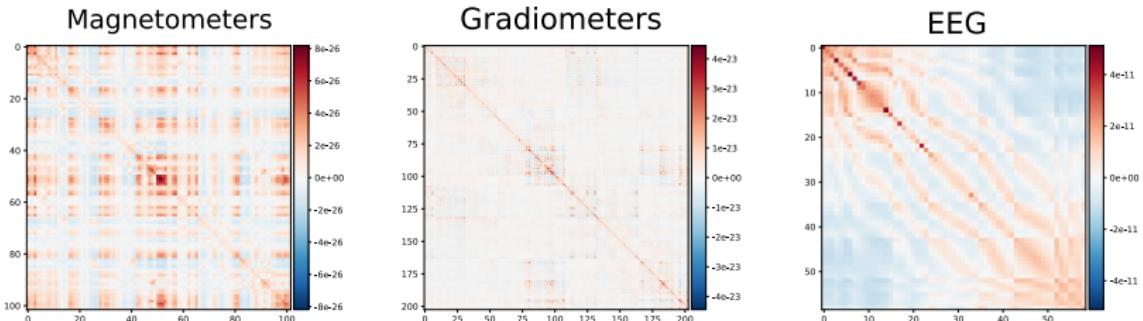


Very noisy data: must repeat recordings

- ▶ average of 5 (top) / 10 (middle) / 50 (bottom) repetitions



Noise covariance for each type of sensor



- ▶ 3 different sensors \implies 3 different noise structures

A Multi-Task framework

Multi-Task regression notation:

- ▶ n observations (e.g., number of sensors)
- ▶ q tasks (e.g., temporal information)
- ▶ p features
- ▶ r number of repetitions
- ▶ $Y^{(1)}, \dots, Y^{(r)} \in \mathbb{R}^{n \times q}$ observation matrices; $\bar{Y} = \frac{1}{r} \sum_l Y^{(l)}$
- ▶ $X \in \mathbb{R}^{n \times p}$ design matrix (known)

$$Y^{(l)} = XB^* + SE^{(l)}$$

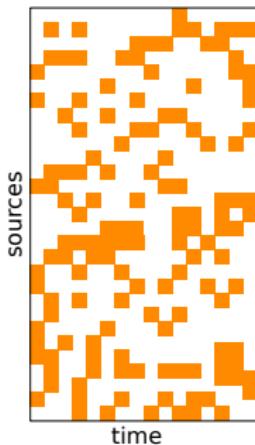
where

- ▶ $B^* \in \mathbb{R}^{p \times q}$: true source activity matrix (unknown)
- ▶ $S \in \mathbb{S}_{++}^n$ co-standard deviation matrix (unknown)
- ▶ $E^{(1)}, \dots, E^{(r)} \in \mathbb{R}^{n \times q}$: white Gaussian noise

Multi-Task penalties⁽¹⁾

Popular convex penalties considered:

$$\hat{\mathbf{B}} \in \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \left(\frac{1}{2nq} \|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_F^2 + \lambda \Omega(\mathbf{B}) \right)$$



Sparse support: no structure

Penalty: **Lasso type**

$$\Omega(\mathbf{B}) = \|\mathbf{B}\|_1 = \sum_{j=1}^p \sum_{k=1}^q |\mathbf{B}_{j,k}|$$

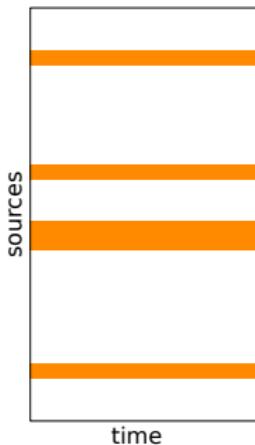
Parameter $\hat{\mathbf{B}} \in \mathbb{R}^{p \times q}$

⁽¹⁾ G. Obozinski, B. Taskar, and M. I. Jordan. "Joint covariate selection and joint subspace selection for multiple classification problems". In: *Statistics and Computing* 20.2 (2010), pp. 231–252.

Multi-Task penalties⁽¹⁾

Popular convex penalties considered: Multi-Task Lasso (MTL)

$$\hat{\mathbf{B}} \in \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \left(\frac{1}{2nq} \|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_F^2 + \lambda \Omega(\mathbf{B}) \right)$$



Parameter $\hat{\mathbf{B}} \in \mathbb{R}^{p \times q}$

Sparse support: group structure

Penalty: **Group-Lasso type**

$$\Omega(\mathbf{B}) = \|\mathbf{B}\|_{2,1} = \sum_{j=1}^p \|\mathbf{B}_{j,:}\|_2$$

where $\mathbf{B}_{j,:}$ the j -th row of \mathbf{B}

⁽¹⁾ G. Obozinski, B. Taskar, and M. I. Jordan. "Joint covariate selection and joint subspace selection for multiple classification problems". In: *Statistics and Computing* 20.2 (2010), pp. 231–252.

Multi-Task data-fitting term

- ▶ Classical Multi-Task estimator: use averaged signal
$$\hat{\mathbf{B}} \in \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \left(\frac{1}{2nq} \|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_F^2 + \lambda \Omega(\mathbf{B}) \right)$$
- ▶ $\|\cdot\|_F^2$ is not designed to handle correlated noise

Multi-Task data-fitting term

- ▶ Classical Multi-Task estimator: use averaged signal

$$\hat{\mathbf{B}} \in \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \left(\frac{1}{2nq} \|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_F^2 + \lambda \Omega(\mathbf{B}) \right)$$

- ▶ $\|\cdot\|_F^2$ is not designed to handle correlated noise
- ▶ can we find and use a better data-fitting term to take the correlation in account?

Multi-Task data-fitting term

- ▶ Classical Multi-Task estimator: use averaged signal
$$\hat{\mathbf{B}} \in \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \left(\frac{1}{2nq} \|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_F^2 + \lambda \Omega(\mathbf{B}) \right)$$
- ▶ $\|\cdot\|_F^2$ is not designed to handle correlated noise
- ▶ can we find and use a better data-fitting term to take the correlation in account?

Reminder on the Lasso theory⁽²⁾ (i.i.d. case, Single-Task)

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

Theorem

- ▶ i.i.d. Gaussian noise
- ▶ + X satisfying the “Restricted Eigenvalue” property
- ▶ + $\lambda = 2\sigma_* \sqrt{\frac{2 \log(p/\delta)}{n}}$
- ▶ \implies with probability $1 - \delta$:

$$\frac{1}{n} \|X\beta^* - X\hat{\beta}^{(\lambda)}\|^2 \leq \frac{18}{\kappa_{s^*}^2} \frac{\sigma_*^2 s^*}{n} \log\left(\frac{p}{\delta}\right)$$

BUT σ_* is unknown in practice !

⁽²⁾P. J. Bickel, Y. Ritov, and A. B. Tsybakov. "Simultaneous analysis of Lasso and Dantzig selector". In: *Ann. Statist.* 37.4 (2009), pp. 1705–1732.

Reminder on the Lasso theory⁽²⁾ (i.i.d. case, Single-Task)

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

Theorem

- ▶ i.i.d. Gaussian noise
- ▶ + X satisfying the “Restricted Eigenvalue” property
- ▶ + $\lambda = 2\sigma_* \sqrt{\frac{2 \log(p/\delta)}{n}}$
- ▶ \implies with probability $1 - \delta$:

$$\frac{1}{n} \|X\beta^* - X\hat{\beta}^{(\lambda)}\|^2 \leq \frac{18}{\kappa_{s^*}^2} \frac{\sigma_*^2 s^*}{n} \log\left(\frac{p}{\delta}\right)$$

BUT σ_* is unknown in practice !

⁽²⁾P. J. Bickel, Y. Ritov, and A. B. Tsybakov. “Simultaneous analysis of Lasso and Dantzig selector”. In: *Ann. Statist.* 37.4 (2009), pp. 1705–1732.

Reminder on the square-root Lasso⁽³⁾⁽⁴⁾⁽⁵⁾

(i.i.d. case, Single-Task)

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{\sqrt{n}} \|y - X\beta\|_2 + \lambda \|\beta\|_1$$

Theorem

- ▶ i.i.d. Gaussian noise
- ▶ + X satisfying the “Restricted Eigenvalue” property
- ▶ + $\lambda = 2\sqrt{\frac{2 \log(p/\delta)}{n}}$
- ▶ \implies with high probability:

$$\frac{1}{n} \|X\beta^* - X\hat{\beta}^{(\lambda)}\|^2 \leq \frac{18}{\kappa_{\beta^*}^2} \frac{\sigma_*^2 s^*}{n} \log\left(\frac{p}{\delta}\right)$$

λ does not depend on σ_* anymore!

⁽³⁾ A. Belloni, V. Chernozhukov, and L. Wang. “Square-root Lasso: pivotal recovery of sparse signals via conic programming”. In: *Biometrika* 98.4 (2011), pp. 791–806.

⁽⁴⁾ T. Sun and C.-H. Zhang. “Scaled sparse linear regression”. In: *Biometrika* 99.4 (2012), pp. 879–898.

⁽⁵⁾ C. Giraud. *Introduction to high-dimensional statistics*. Vol. 138. CRC Press, 2014.

The Smoothed Concomitant Lasso⁽⁶⁾ (i.i.d. case, Single-Task)

$$\hat{\beta}^{(\lambda)} \in \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{\sqrt{n}} \underbrace{\|y - X\beta\|_2}_{\text{non-smooth}} + \lambda \underbrace{\|\beta\|_1}_{\text{non-smooth}}$$

Idea: replacing $\|\cdot\|_2$ by $\|\cdot\|_2 \square \underbrace{\sigma \omega\left(\frac{\cdot}{\sigma}\right)}_{\text{smooth}}(z) = \min_{\sigma \geq \underline{\sigma}} \left(\frac{\|z\|_2^2}{2\sigma} + \frac{\sigma}{2} \right)$

$$(\hat{\beta}^{(\lambda)}, \hat{\sigma}^{(\lambda)}) \in \arg \min_{\beta \in \mathbb{R}^p, \sigma \geq \underline{\sigma}} \frac{\|y - X\beta\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda \|\beta\|_1$$

- ▶ jointly convex: alternate minimization

⁽⁶⁾ E. Ndiaye et al. "Efficient Smoothed Concomitant Lasso Estimation for High Dimensional Regression". In: *Journal of Physics: Conference Series* 904.1 (2017), p. 012006.

The Smoothed Concomitant Lasso (i.i.d. case, Single-Task)

$$(\hat{\beta}^{(\lambda)}, \hat{\sigma}^{(\lambda)}) \in \arg \min_{\beta \in \mathbb{R}^p, \sigma \geq \underline{\sigma}} \frac{\|\mathbf{y} - \mathbf{X}\beta\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda \|\beta\|_1$$

Questions:

- ▶ what is the impact of the smoothing on statistical properties?
(sup-norm convergence? support recovery?)
- ▶ can this estimator (with unknown σ^*) generalize for correlated Gaussian noise?

Statistical properties

$$(\hat{\beta}^{(\lambda)}, \hat{\sigma}^{(\lambda)}) \in \arg \min_{\beta \in \mathbb{R}^p, \sigma \geq \underline{\sigma}} \frac{\|y - X\beta\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda \|\beta\|_1$$

Proposition

- ▶ i.i.d. Gaussian noise
- ▶ + X satisfying the “mutual incoherence” property
- ▶ + $\lambda \sim \frac{\sqrt{\log p}}{\sqrt{n}}$
- ▶ + $\underline{\sigma} \leq \sigma_* / \sqrt{2}$
- ▶ \Rightarrow with high probability:

$$\|\beta^* - \hat{\beta}\|_\infty \leq C \sigma_* \sqrt{\frac{\log p}{n}}$$

SCL support recovery performance as a function of λ and σ

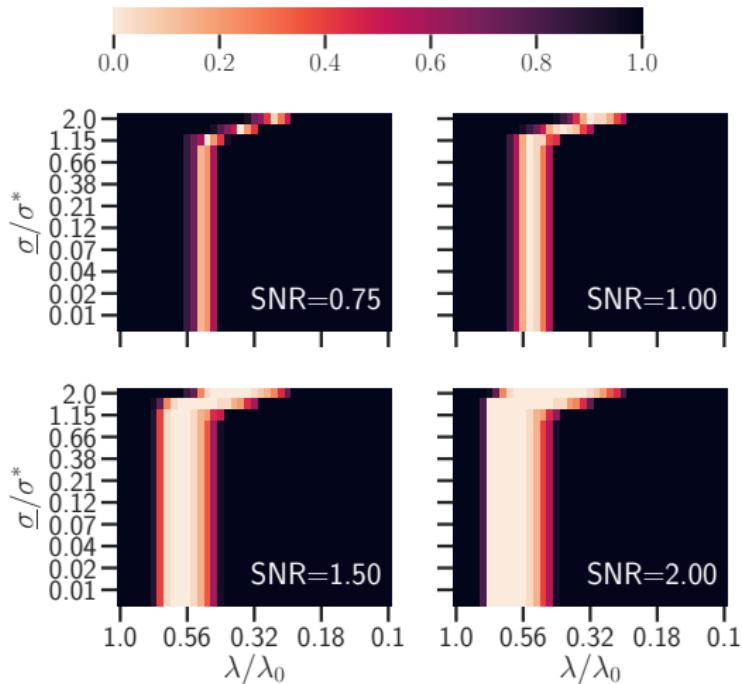


Figure: (Synthetic data, $n = 50$, $p = 1000$, $q = 20$) Hard recovery loss for different values of SNR for the multitask SCL.

Generalization ? Yes ! (correlated Gaussian noise, Multi-Task)

$$\text{SGCL}^{(7)}: (\hat{\mathbf{B}}^{\text{SGCL}}, \hat{\mathbf{S}}^{\text{SGCL}}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \mathbf{S} \in \mathbb{S}_{++}^n, \mathbf{S} \succeq \sigma}} \underbrace{\frac{\|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_{\mathbf{S}^{-1}}^2}{2nq}}_{\text{smooth}} + \frac{\text{Tr}(\mathbf{S})}{2n} + \underbrace{\lambda \|\mathbf{B}\|_{2,1}}_{\text{separable}}$$

Benefits

- ▶ jointly convex formulation

Drawbacks:

- ▶ Statistically: $\mathcal{O}(n^2)$ parameters to estimate for \mathbf{S} only nq observations

⁽⁷⁾ M. Massias et al. "Generalized Concomitant Multi-Task Lasso for Sparse Multimodal Regression". In: AISTATS. vol. 84. 2018, pp. 998–1007.

Generalization ? Yes ! (correlated Gaussian noise, Multi-Task)

SGCL⁽⁷⁾:

$$(\hat{\mathbf{B}}^{\text{SGCL}}, \hat{\mathbf{S}}^{\text{SGCL}}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \mathbf{S} \in \mathbb{S}_{++}^n, \mathbf{S} \succeq \sigma}} \underbrace{\frac{\|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_{\mathbf{S}^{-1}}^2}{2nq}}_{\text{smooth}} + \underbrace{\frac{\text{Tr}(\mathbf{S})}{2n}}_{\text{separable}} + \lambda \|\mathbf{B}\|_{2,1}$$

Benefits

- ▶ jointly convex formulation

Drawbacks:

- ▶ Statistically: $\mathcal{O}(n^2)$ parameters to estimate for \mathbf{S} only nq observations

Question: can this estimator take advantage of the number of repetitions?

⁽⁷⁾ M. Massias et al. "Generalized Concomitant Multi-Task Lasso for Sparse Multimodal Regression". In: AISTATS. vol. 84. 2018, pp. 998–1007.

Generalization ? Yes ! (correlated Gaussian noise, Multi-Task)

SGCL⁽⁷⁾:

$$(\hat{\mathbf{B}}^{\text{SGCL}}, \hat{\mathbf{S}}^{\text{SGCL}}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \mathbf{S} \in \mathbb{S}_{++}^n, \mathbf{S} \succeq \sigma}} \underbrace{\frac{\|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_{\mathbf{S}^{-1}}^2}{2nq}}_{\text{smooth}} + \underbrace{\frac{\text{Tr}(\mathbf{S})}{2n}}_{\text{separable}} + \lambda \|\mathbf{B}\|_{2,1}$$

Benefits

- ▶ jointly convex formulation

Drawbacks:

- ▶ Statistically: $\mathcal{O}(n^2)$ parameters to estimate for \mathbf{S} only nq observations

Question: can this estimator take advantage of the number of repetitions?

⁽⁷⁾ M. Massias et al. "Generalized Concomitant Multi-Task Lasso for Sparse Multimodal Regression". In: AISTATS. vol. 84. 2018, pp. 998–1007.

Statistical properties⁽⁸⁾

$$\arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \mathbf{S} \in \mathbb{S}_{++}^n, \mathbf{S} \succeq \underline{\sigma}}} \underbrace{\frac{\|\bar{\mathbf{Y}} - \mathbf{X}\mathbf{B}\|_{\mathbf{S}^{-1}}^2}{2nq}}_{\text{Term 1}} + \underbrace{\frac{\text{Tr}(\mathbf{S})}{2n}}_{\text{Term 2}} + \underbrace{\lambda \|\mathbf{B}\|_{2,1}}_{\text{Term 3}}$$

Proposition

- ▶ i.i.d. Gaussian noise
- ▶ + \mathbf{X} satisfying the “mutual incoherence” property
- ▶ + $\lambda \sim \frac{\sqrt{\log p}}{q\sqrt{n}}$
- ▶ + $\underline{\sigma} \leq \sigma_* / \sqrt{2}$
- ▶ \implies with high probability: at least $1 - \dots - 2ne^{-cq/n}$

$$\frac{1}{q} \|\mathbf{B}^* - \hat{\mathbf{B}}\|_{2,\infty} \leq C\sigma_* \frac{1}{q} \sqrt{\frac{\log p}{n}}$$

⁽⁸⁾ Q. Bertrand et al. “Support recovery and sup-norm convergence rates for sparse pivotal estimation”. In: (Submitted) (2019).

Can take advantage of repetitions? Yes!

CLaR:

$$(\hat{\mathbf{B}}^{\text{CLaR}}, \hat{\mathbf{S}}^{\text{CLaR}}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \mathbf{S} \in \mathbb{S}_{++}^n, \mathbf{S} \succeq \boldsymbol{\sigma}}} \frac{\sum_{l=1}^r \|Y^{(l)} - \mathbf{X}\mathbf{B}\|_{\mathbf{S}^{-1}}^2}{2nqr} + \frac{\text{Tr}(\mathbf{S})}{2n} + \lambda \|\mathbf{B}\|_{2,1}$$

- ▶ Statistically: $\mathcal{O}(n^2)$ parameters to estimate for \mathbf{S} with nqr observations ($r = \text{number of repetitions}$)

Proposition

Link with the Trace norm⁽⁹⁾

$$\hat{\mathbf{B}}^{\text{CLaR}} = \arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} (\|\cdot\|_{\text{Tr}} \square \omega_{\sigma}) (Z) + \lambda n \|\mathbf{B}\|_{2,1} .$$

where $Z = \frac{1}{\sqrt{q}} [Y^{(1)} - XB | \dots | Y^{(r)} - XB]$.

- ▶ justification for the estimator introduced heuristically
- ▶ smoothed and generalized version of the multivariate square-root Lasso⁽¹⁰⁾:

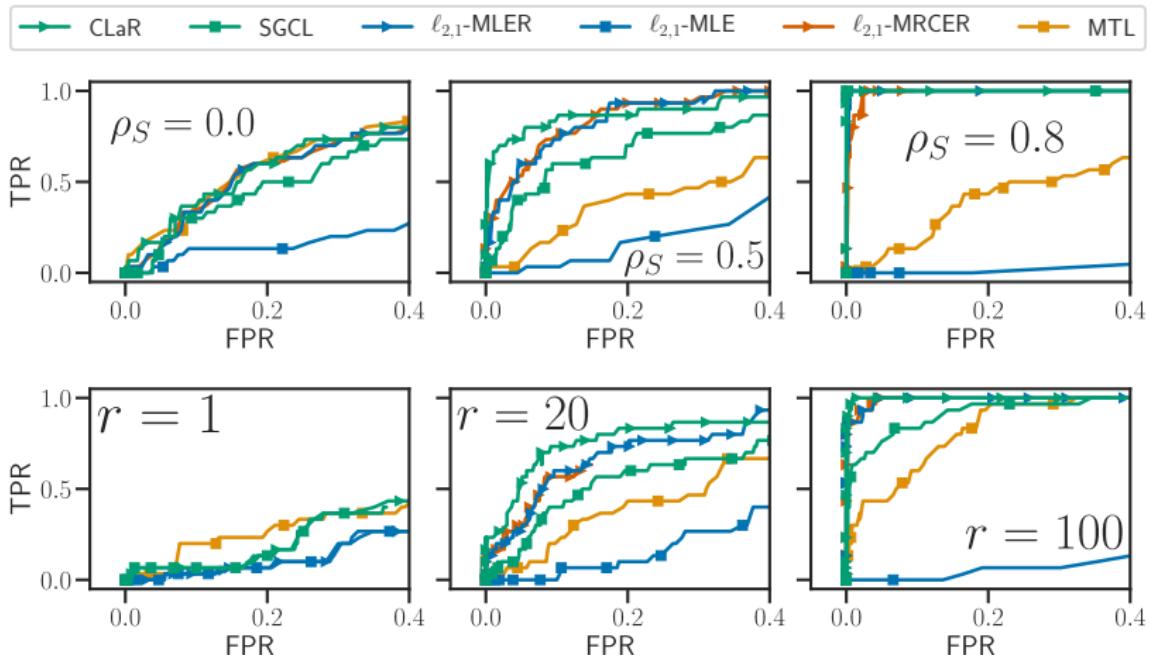
$$\arg \min_{\mathbf{B} \in \mathbb{R}^{p \times q}} \frac{1}{n\sqrt{q}} \| [Y^{(1)} - XB | \dots | Y^{(r)} - XB] \|_{\text{Tr}} + \lambda \|\mathbf{B}\|_{2,1} .$$

⁽⁹⁾ Q. Bertrand et al. "Handling correlated and repeated measurements with the smoothed multivariate square-root Lasso". In: *NeurIPS* (2019).

⁽¹⁰⁾ S. van de Geer. *Estimation and testing under sparsity*. Vol. 2159. Lecture Notes in Mathematics. Lecture notes from the 45th Probability Summer School held in Saint-Flour, 2015, École d'Été de Probabilités de Saint-Flour. Springer, 2016, pp. xiii+274.

Simulated scenarios

- ▶ $n = 150, p = 500, q = 100$
- ▶ X Toeplitz-correlated
- ▶ S^* Toeplitz matrix: $S^*_{i,j} = \rho_{S^*}^{|i-j|}, \rho_{S^*} \in]0, 1[$



Real data

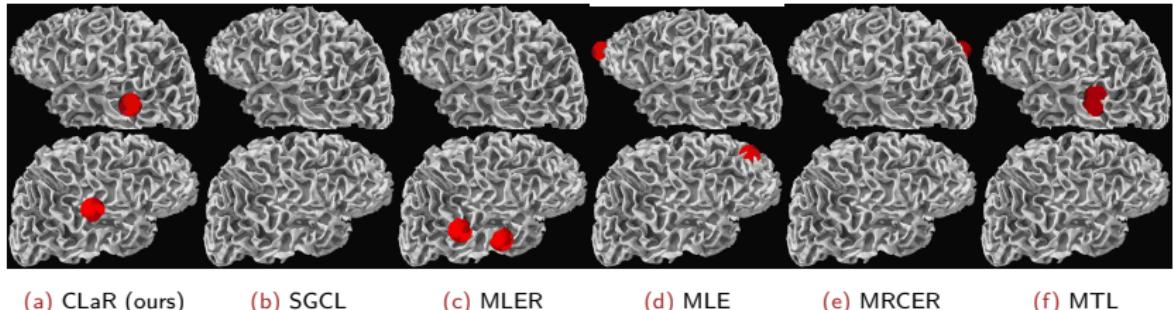


Figure: Real data, left auditory stimulations ($n = 102$, $p = 7498$, $q = 76$, $r = 63$) Sources found in the left hemisphere (top) and the right hemisphere (bottom) after left auditory stimulations.

- ▶ expected: 2 sources (one in each auditory cortex)
- ▶ λ chosen such that $\|\hat{\mathbf{B}}\|_{2,0} = 2$
- ▶ deep sources for SGCL and $\ell_{2,1}$ -MRCER (not visible)

Conclusion and perspectives

- ▶ New estimator to handle correlated noise and repetitions in Multi-Task
- ▶ Improved support identification

Conclusion and perspectives

- ▶ New estimator to handle correlated noise and repetitions in Multi-Task
- ▶ Improved support identification
- ▶ Numerical cost "similar" to classical Multi-Task Lasso

Conclusion and perspectives

- ▶ New estimator to handle correlated noise and repetitions in Multi-Task
- ▶ Improved support identification
- ▶ Numerical cost "similar" to classical Multi-Task Lasso
- ▶ Ongoing work: non-convex penalties, statistical analysis.

Conclusion and perspectives

- ▶ New estimator to handle correlated noise and repetitions in Multi-Task
- ▶ Improved support identification
- ▶ Numerical cost "similar" to classical Multi-Task Lasso
- ▶ Ongoing work: non-convex penalties, statistical analysis.

Merci!

“All models are wrong but some come with good open source implementation and good documentation to use these.”

A. Gramfort

- ▶ Python code online <https://github.com/QB3/CLaR>
- ▶ Papers: arXiv^{(11), (12) (13)}

⁽¹¹⁾ M. Massias et al. “Generalized Concomitant Multi-Task Lasso for Sparse Multimodal Regression”. In: *AISTATS*. vol. 84. 2018, pp. 998–1007.

⁽¹²⁾ Q. Bertrand et al. “Handling correlated and repeated measurements with the smoothed multivariate square-root Lasso”. In: *NeurIPS* (2019).

⁽¹³⁾ Q. Bertrand et al. “Support recovery and sup-norm convergence rates for sparse pivotal estimation”. In: *(Submitted)* (2019).

Competitors

- ▶ (smoothed) $\ell_{2,1}$ -MLE

$$(\hat{\mathbf{B}}, \hat{\Sigma}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \Sigma \succeq \underline{\sigma}^2 / r^2}} \left\| \bar{\mathbf{Y}} - X\mathbf{B} \right\|_{\Sigma^{-1}}^2 - \log \det(\Sigma^{-1}) + \lambda \|\mathbf{B}\|_{2,1} ,$$

- ▶ and its repetitions version ($\ell_{2,1}$ -MLER):

$$(\hat{\mathbf{B}}, \hat{\Sigma}) \in \arg \min_{\substack{\mathbf{B} \in \mathbb{R}^{p \times q} \\ \Sigma \succeq \underline{\sigma}^2}} \sum_1^r \left\| \mathbf{Y}^{(l)} - X\mathbf{B} \right\|_{\Sigma^{-1}}^2 - \log \det(\Sigma^{-1}) + \lambda \|\mathbf{B}\|_{2,1} .$$

- ▶ $\ell_{2,1}$ -MLE and $\ell_{2,1}$ -MLER are bi-convex but not jointly convex

Smoothing of matrix norm

Huber-like formula for the Frobenius norm

$$\|\cdot\|_F \square_{\underline{\sigma}} \omega \left(\frac{\cdot}{\underline{\sigma}} \right) (Z) = \begin{cases} \frac{\|Z\|_F^2}{2\underline{\sigma}} + \frac{\underline{\sigma}}{2}, & \text{if } \|Z\|_F \leq \underline{\sigma} \\ \|Z\|_F, & \text{if } \|Z\|_F > \underline{\sigma} \end{cases}$$
$$= \min_{\sigma \geq \underline{\sigma}} \left(\frac{\|Z\|_F^2}{2\sigma} + \frac{\sigma}{2} \right)$$

What about other norms ?

Smoothing of matrix norm

Huber-like formula for the Frobenius norm

$$\|\cdot\|_F \square_{\underline{\sigma}} \omega \left(\frac{\cdot}{\underline{\sigma}} \right) (Z) = \begin{cases} \frac{\|Z\|_F^2}{2\underline{\sigma}} + \frac{\underline{\sigma}}{2}, & \text{if } \|Z\|_F \leq \underline{\sigma} \\ \|Z\|_F, & \text{if } \|Z\|_F > \underline{\sigma} \end{cases}$$
$$= \min_{\sigma \geq \underline{\sigma}} \left(\frac{\|Z\|_F^2}{2\sigma} + \frac{\sigma}{2} \right)$$

What about other norms ?

Huber-like formula for the nuclear/trace norm

$$\|\cdot\|_{s,1} \square \omega_{\underline{\sigma}} (Z) = \begin{cases} \frac{\|Z\|_F^2}{2\underline{\sigma}} + \frac{\underline{\sigma}}{2} n \wedge q, & \text{if } \|Z\|_\infty \leq \underline{\sigma} \\ \frac{1}{2\underline{\sigma}} \sum_i \gamma_i^2 - (\gamma_i \wedge \underline{\sigma} - \gamma_i)^2, & \text{if } \|Z\|_\infty > \underline{\sigma} \end{cases}$$
$$= \min_{S \succeq \underline{\sigma}} \frac{1}{2} \|Z\|_{S^{-1}}^2 + \frac{1}{2} \text{Tr}(S)$$

γ_i : singular values of Z

$\|Z\|_{S^{-1}}^2 := \text{Tr}(Z^\top S^{-1} Z)$ Mahalanobis distance

Smoothing of matrix norm

Huber-like formula for the Frobenius norm

$$\|\cdot\|_F \square_{\underline{\sigma}} \omega \left(\frac{\cdot}{\underline{\sigma}} \right) (Z) = \begin{cases} \frac{\|Z\|_F^2}{2\underline{\sigma}} + \frac{\underline{\sigma}}{2}, & \text{if } \|Z\|_F \leq \underline{\sigma} \\ \|Z\|_F, & \text{if } \|Z\|_F > \underline{\sigma} \end{cases}$$
$$= \min_{\sigma \geq \underline{\sigma}} \left(\frac{\|Z\|_F^2}{2\sigma} + \frac{\sigma}{2} \right)$$

What about other norms ?

Huber-like formula for the nuclear/trace norm

$$\|\cdot\|_{s,1} \square \omega_{\underline{\sigma}} (Z) = \begin{cases} \frac{\|Z\|_F^2}{2\underline{\sigma}} + \frac{\underline{\sigma}}{2} n \wedge q, & \text{if } \|Z\|_\infty \leq \underline{\sigma} \\ \frac{1}{2\underline{\sigma}} \sum_i \gamma_i^2 - (\gamma_i \wedge \underline{\sigma} - \gamma_i)^2, & \text{if } \|Z\|_\infty > \underline{\sigma} \end{cases}$$
$$= \min_{S \succeq \underline{\sigma}} \frac{1}{2} \|Z\|_{S^{-1}}^2 + \frac{1}{2} \text{Tr}(S)$$

γ_i : singular values of Z

$\|Z\|_{S^{-1}}^2 := \text{Tr}(Z^\top S^{-1} Z)$ **Mahalanobis distance**