

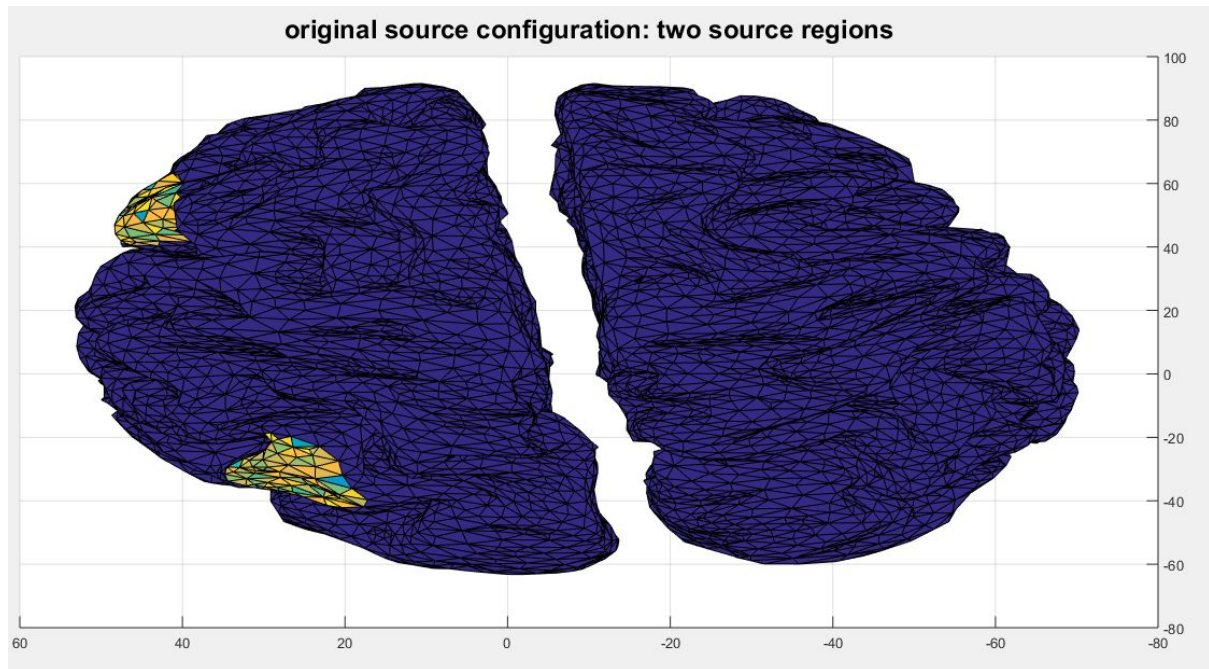
Master SISEA - Problèmes Inverses

Gonzalo Quintana

TP2 - Régularisation de Tikhonov et régularisation TV-L1	2
2.1) Algorithme MNE	3
2.2) Algorithme SISSY	4
3.1) Implémentation de l'algorithme MNE	6
3.2) Étude de l'algorithme MNE	7
Variation de lambda (paramètre de régularisation)	7
Variation de la RSB et de lambda	8
Choix du paramètre de régularisation	9
L-curve criterion	9
Discrepancy Principle	10
Generalized Cross Validation	11
3.3) Implémentation de l'algorithme SISSY	12
3.4) Etude de l'algorithme SISSY	13
Variation de lambda (paramètre de régularisation)	13
Variation d'alpha	13
Choix du paramètre de régularisation	14
TP3 - Analyse des performances et comparaison des algorithmes étudiés	16
2.1) Comparaison qualitative	16
Bruit gaussien non corrélé	16
Bruit gaussien spatialement corrélé (activité de fond)	18
2.2) Comparaison quantitative	19
Graphique de type boxplot	20
DLE vs RSB	20

TP2 - Régularisation de Tikhonov et régularisation TV-L1

Le but de ce tp est de résoudre le problème inverse de localisation des sources dans la cas d'un signal d'EEG, où la plupart des dipôles sont inactives. Dans le graphique suivante, on voit en jaune les dipôles actifs et en bleu les inactives (pour l'instant temporelle où on trouve l'amplitude de signal maximale).



2.1) Algorithme MNE

$$\min_s \|X - AS\|_2^2 + \lambda \|S\|_2^2 = \min_s \varphi(s)$$

$$\begin{aligned}\varphi(s) &= (X - AS)^T (X - AS) + \lambda S^T S \\ &= X^T X - 2X^T AS + S^T A^T AS + \lambda S^T S\end{aligned}$$

$$\begin{aligned}\nabla \varphi(s) &= -2A^T X + 2A^T AS + 2\lambda S = 0 \quad (\text{condition nécessaire}) \\ (A^T A + \lambda I)S &= A^T X\end{aligned}$$

$$S = (A^T A + \lambda I)^{-1} A^T X$$

En utilisant le lemme d'inversion avec $R^{-1} = I$, $P^{-1} = \lambda I$ et $B = A$

$$\begin{aligned}(A^T A + \lambda I)^{-1} A^T &= \frac{1}{\lambda} I A^T (\underbrace{\lambda I}_{\lambda} A^T + \underbrace{I}_{\lambda})^{-1} = \lambda^{-1} A^T [\lambda^{-1} (A A^T + \lambda I)]^{-1} \\ &= \lambda^{-1} A^T \lambda (A A^T + \lambda I)^{-1} = A^T (A A^T + \lambda I)^{-1}\end{aligned}$$

$$\boxed{\therefore S = A^T (A A^T + \lambda I)^{-1} X}$$

Comme le problème est convexe par rapport à s , cette solution est aussi suffisante. condition

2.2) Algorithm SISSY

2.2) Algorithm SISSY

$$\min_{\lambda} \|x - A\lambda\|_2^2 + \lambda (\|z\|_1 + \alpha \|y\|_1) \quad \text{s.t. } z = T\lambda, y = \lambda$$

(regularisation TV-L1)

Lagrangian augmented:

$$L(\lambda, z, \gamma, u, v) = \frac{1}{2} \|x - A\lambda\|_2^2 + \lambda (\|z\|_1 + \alpha \|y\|_1) + u^T (z - T\lambda) + v^T (x - \lambda) + \rho/2 \|y - \lambda\|_2^2 + \rho/2 \|z - T\lambda\|_2^2$$

$$\bullet \lambda = \arg \min_{\lambda} L(\lambda, z, \gamma, u, v)$$

$$L(\dots) = \frac{1}{2} x^T x - x^T A \lambda + \frac{1}{2} A^T A \lambda + \lambda (\|z\|_1 + \alpha \|y\|_1) + u^T (z - T\lambda) + v^T (x - \lambda) + \rho/2 (y^T y - 2y^T \lambda + \lambda^T \lambda) + \rho/2 (z^T z - 2z^T T\lambda + \lambda^T T^T T \lambda)$$

$$\frac{\partial L(\dots)}{\partial \lambda} = -A^T x + A^T A \lambda - T^T u - v - \rho y + \rho \lambda - \rho T^T z + \rho T^T T \lambda = 0$$

$$(A^T A + \rho I + \rho T^T T) \lambda = A^T x + T^T u + v + \rho y + \rho T^T z$$

$$\boxed{\lambda = (A^T A + \rho(I + T^T T))^{-1} (A^T x + T^T u + \rho T^T z + \rho y + v)}$$

$$\boxed{\lambda^{(k+1)} = (A^T A + \rho(I + T^T T))^{-1} (A^T x^{(k)} + T^T u^{(k)} + \rho T^T z^{(k)} + \rho y^{(k)} + v^{(k)})}$$

$$\cdot \mathbf{z} = \arg \min_{\mathbf{z}} L_{\mathbf{z}}(\mathbf{z})$$

$$\begin{aligned} L_{\mathbf{z}}(\mathbf{z}) &= \lambda \|\mathbf{z}\|_1 + \mathbf{u}^T \mathbf{z} + \rho/2 \|\mathbf{z} - \mathbf{T}\mathbf{A}\|_2^2 + \text{cte} \\ &= \lambda \|\mathbf{z}\|_1 + \mathbf{u}^T \mathbf{z} + \rho/2 (\mathbf{z}^T \mathbf{z} - 2 \mathbf{z}^T \mathbf{T}\mathbf{A} + \mathbf{A}^T \mathbf{T}^T \mathbf{T}\mathbf{A}) + \text{cte} \\ &= \lambda \|\mathbf{z}\|_1 + \frac{\rho}{2} \mathbf{z}^T \mathbf{z} + (\mathbf{u}^T - \rho \mathbf{A}^T \mathbf{T}^T) \mathbf{z} + \text{cte} \\ &= \lambda \|\mathbf{z}\|_1 + \frac{\rho}{2} (\mathbf{z}^T \mathbf{z} + \frac{2 \mathbf{u}^T - 2 \mathbf{A}^T \mathbf{T}^T}{\rho} \mathbf{z} + 4 (\frac{\mathbf{u}}{\rho} - \mathbf{A}^T \mathbf{T})^T (\frac{\mathbf{u}}{\rho} - \mathbf{A}^T \mathbf{T})) \\ &= \frac{\rho}{2} \|\mathbf{z} - (\mathbf{A}^T \mathbf{T} - \frac{\mathbf{u}}{\rho})\|_2^2 + \lambda \|\mathbf{z}\|_1 \end{aligned}$$

on voit que $\mathbf{z} = \arg \min_{\mathbf{z}} L_{\mathbf{z}}(\mathbf{z}) = \arg \min_{\mathbf{z}} \frac{\rho}{2} \|\mathbf{z} - (\mathbf{A}^T \mathbf{T} - \frac{\mathbf{u}}{\rho})\|_2^2 + \lambda \|\mathbf{z}\|_1$

$$\boxed{\mathbf{z} = \text{prox}_{\lambda/\rho}(\mathbf{A}^T \mathbf{T} - \frac{\mathbf{u}}{\rho})} \Rightarrow \boxed{\mathbf{z}^{(k+1)} = \text{prox}_{\frac{\lambda}{\rho}}(\mathbf{A}^{(k+1)T} \mathbf{T} - \frac{\mathbf{u}^{(k)}}{\rho})}$$

$$\cdot \mathbf{y} = \arg \min_{\mathbf{y}} L_{\mathbf{y}}(\mathbf{y})$$

$$\begin{aligned} L_{\mathbf{y}}(\mathbf{y}) &= \lambda_2 \|\mathbf{y}\|_1 + \mathbf{V}^T \mathbf{y} + \rho/2 \|\mathbf{y} - \mathbf{A}\|_2^2 + \text{cte} \\ &= \lambda_2 \|\mathbf{y}\|_1 + \mathbf{V}^T \mathbf{y} + \rho/2 (\mathbf{y}^T \mathbf{y} - 2 \mathbf{y}^T \mathbf{A} + \mathbf{A}^T \mathbf{A}) + \text{cte} \\ &= \lambda_2 \|\mathbf{y}\|_1 + \rho/2 (\mathbf{y}^T \mathbf{y} - 2 (\mathbf{A}^T - \frac{\mathbf{V}}{\rho})^T \mathbf{y} + (\mathbf{A} - \mathbf{V}/\rho)^T (\mathbf{A} - \mathbf{V}/\rho)) + \text{cte} \\ &= \frac{\rho}{2} \|\mathbf{y} - (\mathbf{A} - \mathbf{V}/\rho)\|_2^2 + \lambda_2 \|\mathbf{y}\|_1 = \frac{1}{2} \|\mathbf{y} - (\mathbf{A} - \mathbf{V}/\rho)\|_2^2 + \frac{\lambda_2}{\rho} \|\mathbf{y}\|_1 \end{aligned}$$

$$\boxed{\mathbf{y} = \text{prox}_{\frac{\lambda_2}{\rho}}(\mathbf{A} - \frac{1}{\rho} \mathbf{V})} \Rightarrow \boxed{\mathbf{y}^{(k+1)} = \text{prox}_{\frac{\lambda_2}{\rho}}(\mathbf{A}^{(k+1)} - \frac{1}{\rho} \mathbf{V}^{(k)})}$$



$$u = \underset{u}{\operatorname{argmin}} L_u(u)$$

$L_u(u) = \|u\|^2 (z - T\delta) \rightarrow$ on utilise le gradient ascendant

$$u^{(k+1)} = u^{(k)} + \rho \nabla L_u(u) \quad \nabla L_u(u) = z - T\delta$$

$$u^{(k+1)} = u^{(k)} + \rho (z - T\delta) \Rightarrow \cancel{u^{(k+1)} = u^{(k)} + \rho (T\delta^{(k+1)} - z^{(k+1)})}$$

$$\cancel{u^{(k+1)} = u^{(k)} + \rho (T\delta^{(k+1)} - z^{(k+1)})}$$

$$\boxed{u^{(k+1)} = u^{(k)} + \rho (z^{(k+1)} - T\delta^{(k+1)})}$$

$$v = \underset{v}{\operatorname{argmin}} L_v(v)$$

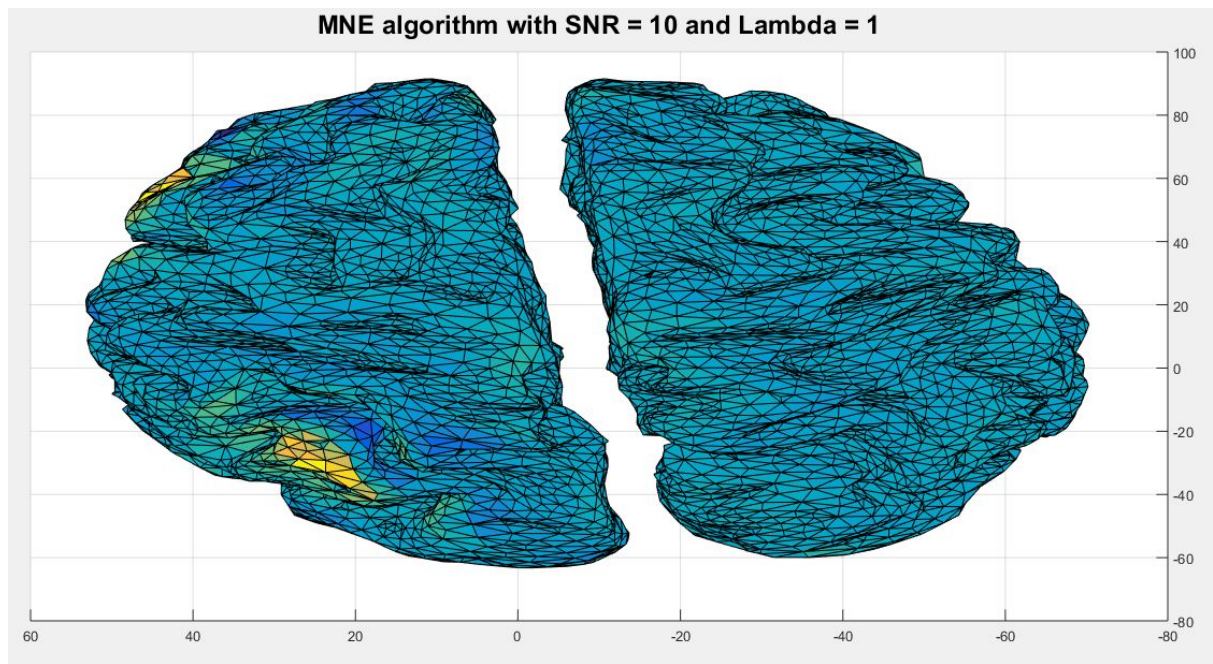
On faisant la même chose : $L_v(v) = v^T (y - \delta)$

$$\nabla L_v(v) = y - \delta$$

$$\boxed{v^{(k+1)} = v^{(k)} + \rho (y^{(k+1)} - \delta^{(k+1)})}$$

3.1) Implémentation de l'algorithme MNE

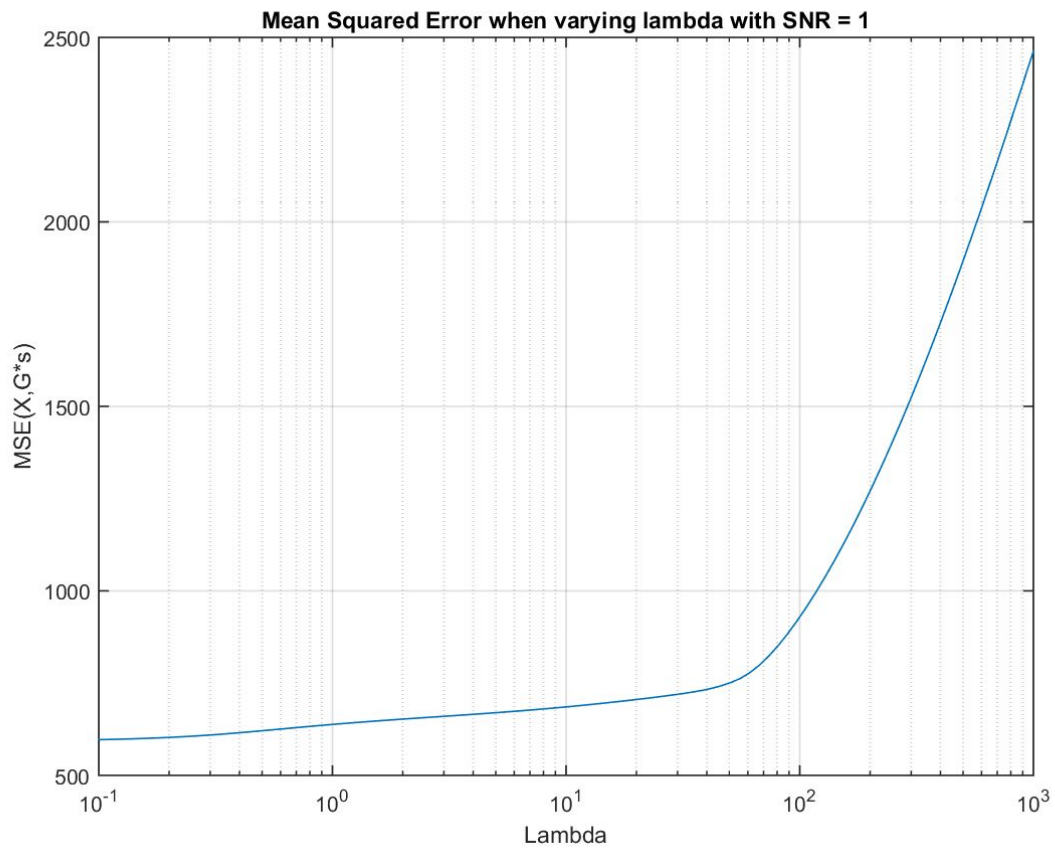
On implémente l'algorithme MNE avec $\lambda=1$ et on visualise la solution inverse pour le même instant temporelle d'avant. En comparant avec la localisation originale des sources, on note qu'on arrive à identifier les deux zones actives.



3.2) Étude de l'algorithme MNE

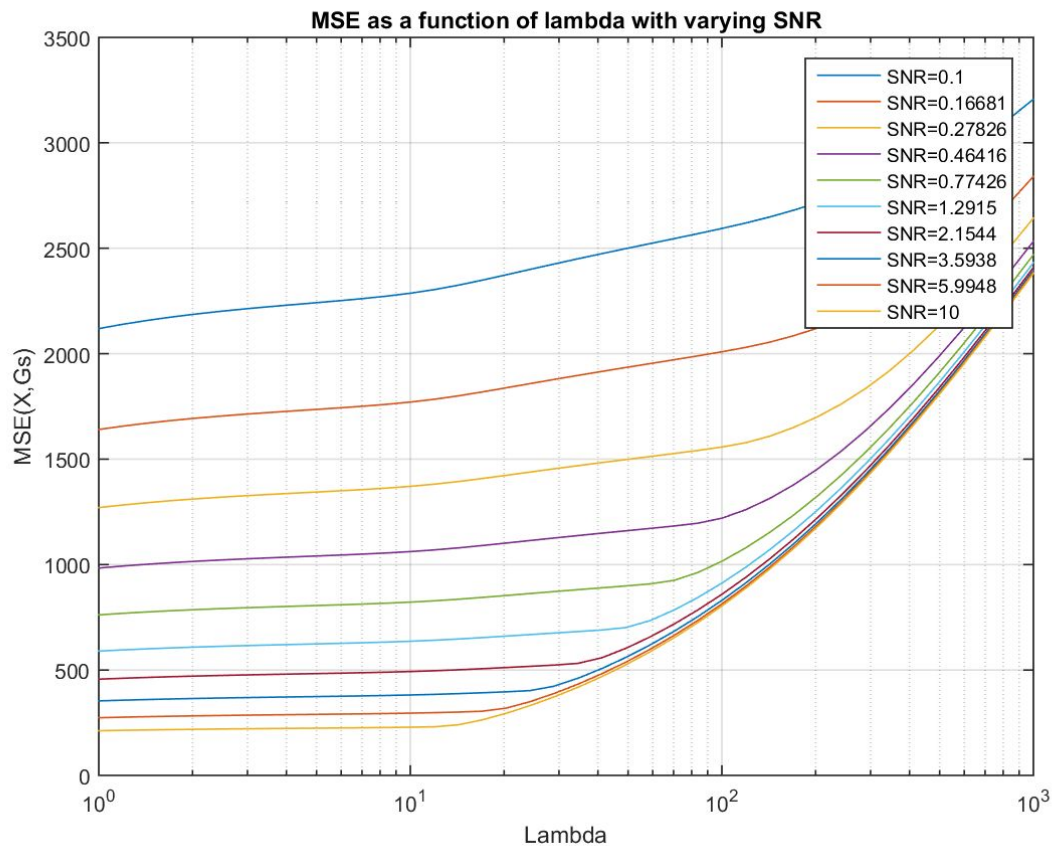
Variation de lambda (paramètre de régularisation)

On graphique l'erreur quadratique moyenne entre X et $G*s$ - étant s l'estimation donnée par l'algorithme MNE- en fonction du paramètre de régularisation Λ . On observe que pour des λ trop grandes l'EQM est grande, parce qu'on donne trop d'importance à la régularisation et non pas au vrai problème. Au contraire, pour des λ plus faibles l'EQM est plus petite parce qu'on donne plus d'importance au problème original.



Variation de la RSB et de lambda

On varie la RSB (relation signal bruit) et on calcule l'erreur quadratique moyenne pour des différents lambdas (paramètre de régularisation). On observe que pour des RSBs plus faibles, l'erreur est plus grande.

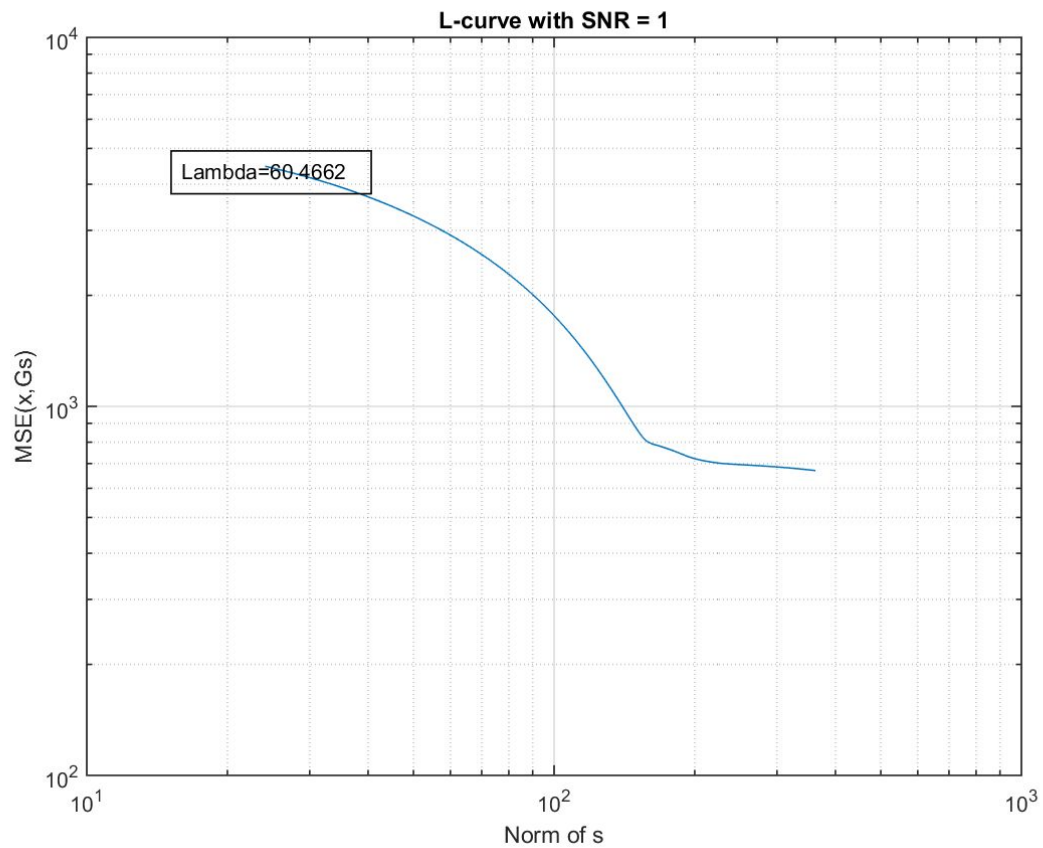


Choix du paramètre de régularisation

Pour choisir le paramètre de régularisation, on considère trois heuristiques différentes: le L-curve criterion, le discrepancy principle et le Generalized Cross Validation. On implémente ces trois heuristiques avec $RSB=1$.

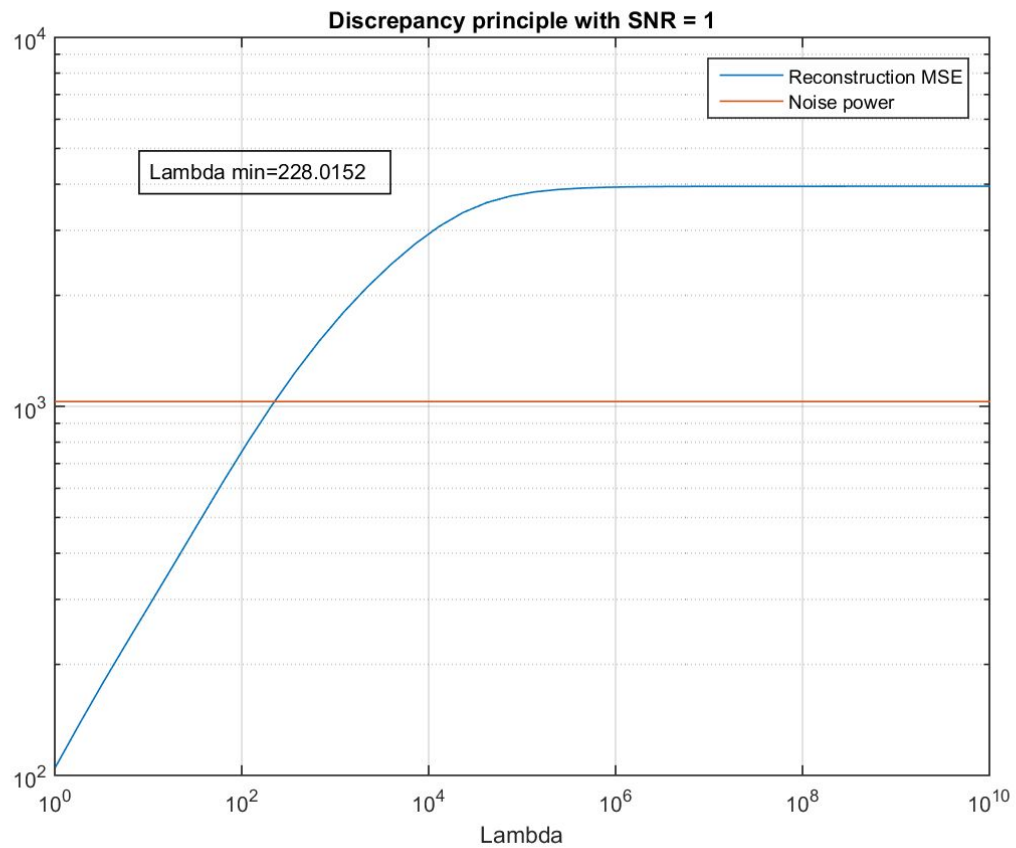
L-curve criterion

Pour des différentes lambdas, on graphique l'erreur quadratique moyenne entre x et G_s en fonction de la norme de s . Le lambda optimal correspond au pli de la courbe.



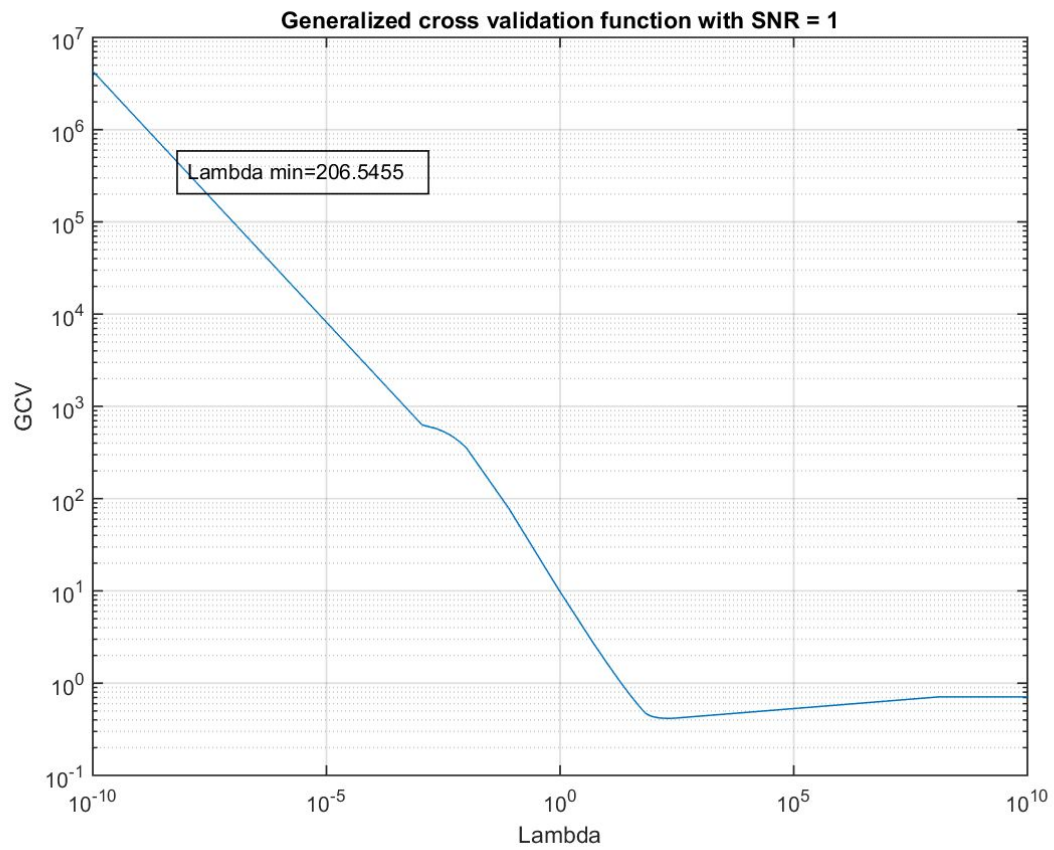
Discrepancy Principle

Le lambda optimal est donné par le point d'intersection entre les courbes de puissance de l'erreur de reconstruction (erreur entre x et Gs) et puissance du bruit (qui ne dépend pas de lambda).

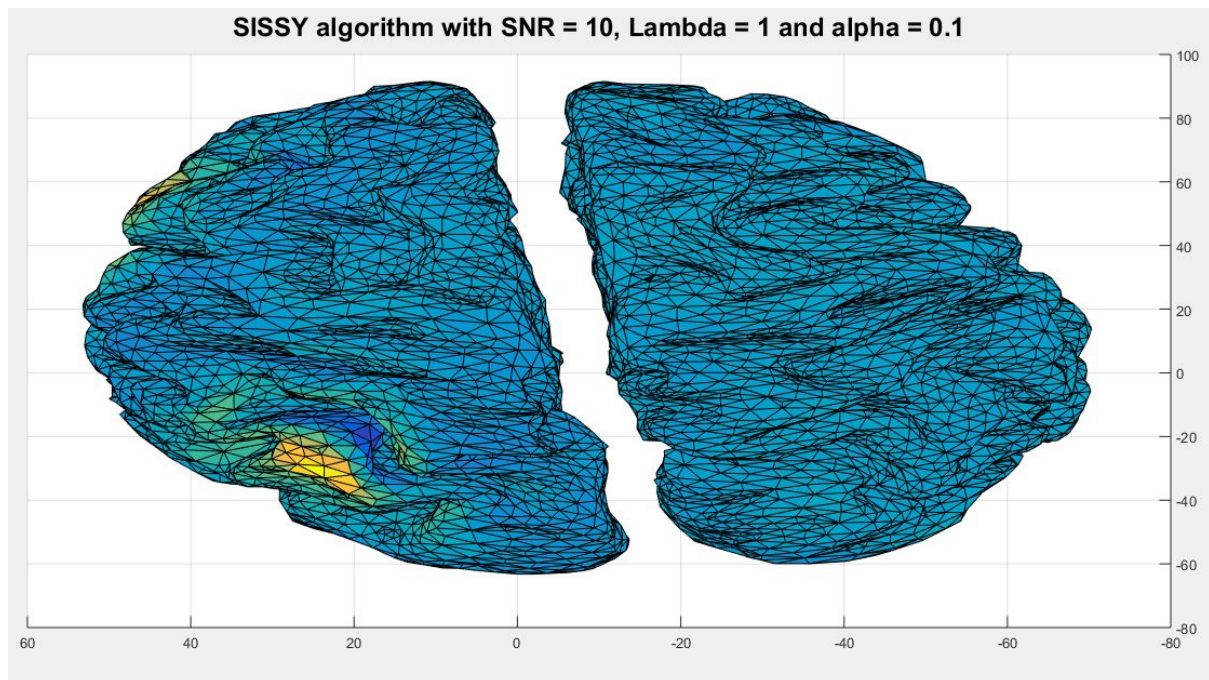


Generalized Cross Validation

Le paramètre de régularisation est choisi tel qu'il minimise la courbe GCV.



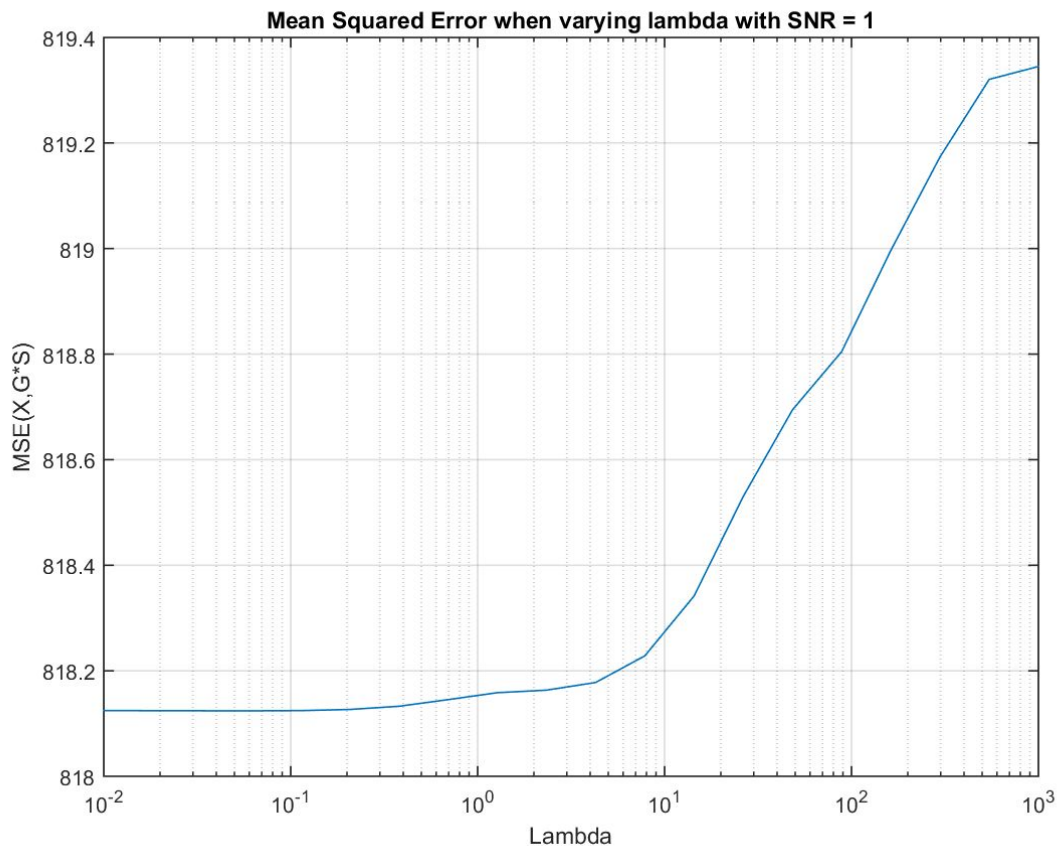
3.3) Implémentation de l'algorithme SISSY



3.4) Etude de l'algorithme SISSY

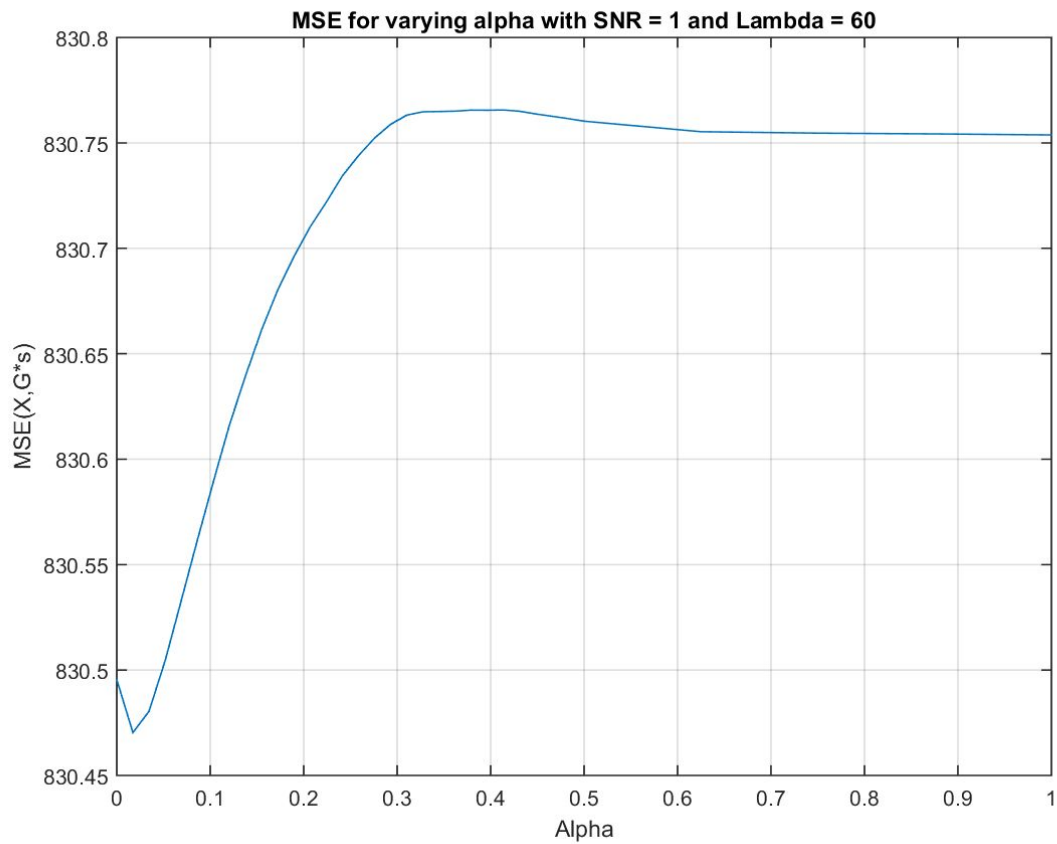
Variation de lambda (paramètre de régularisation)

On graphique l'erreur quadratique moyenne entre X et G^*s en fonction du paramètre de régularisation Λ (pour $\alpha=0.1$). On retrouve une courbe similaire à celle de l'algorithme MNE.



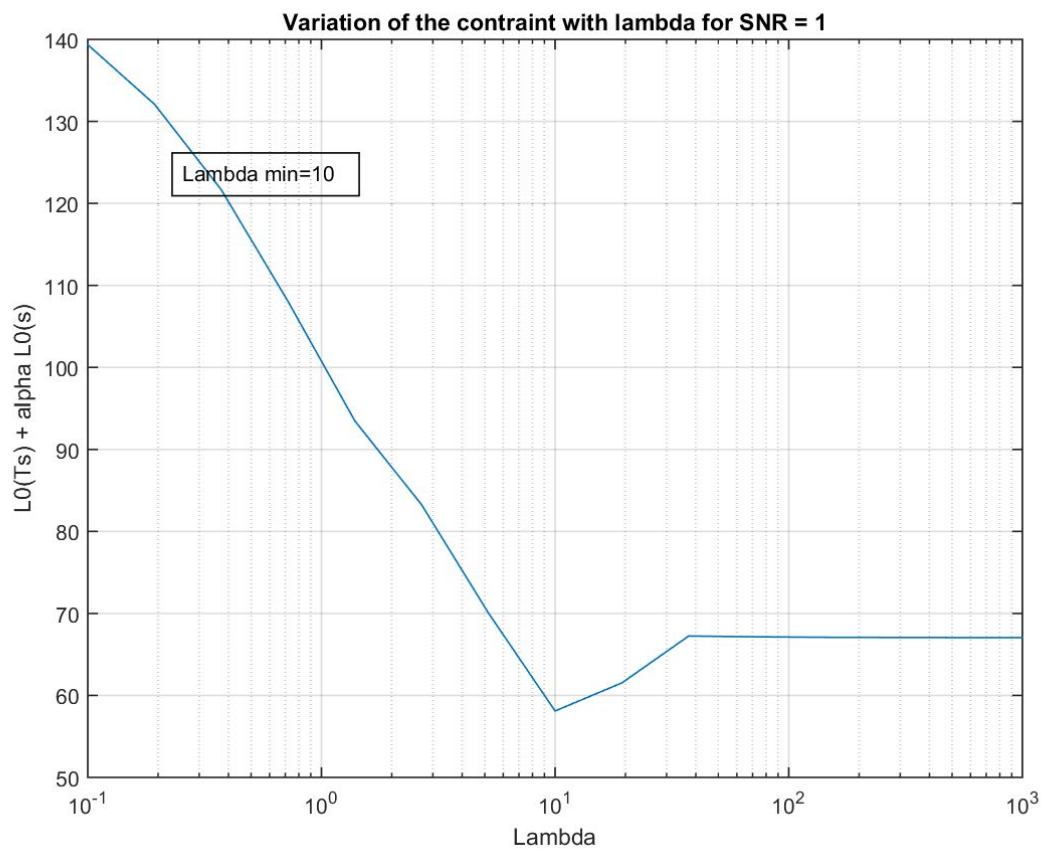
Variation d'alpha

On calcule l'EQM en fonction du paramètre α , en choisissant $\lambda=6$ (ce qui correspond au pli de la courbe antérieure). On choisit cette valeur parce que c'est la plus grande (solution plus *sparse*) qui permet d'avoir des bonnes performances (selon la courbe antérieure). On obtient la courbe suivante, où on note que les valeurs petites d' α sont les plus optimales.



Choix du paramètre de régularisation

On calcule la valeur de la contrainte qu'on souhaite vraiment imposer (basée sur la norme L_0) pour des différents lambdas et on choisit celui qui la minimise. Le lambda obtenu est égal à 10, qui est dans le même ordre de magnitude que celui obtenu avec le *discrepancy principle* (on avait obtenu lambda égal à 6). On conclut, donc, que ces deux heuristiques sont cohérentes.

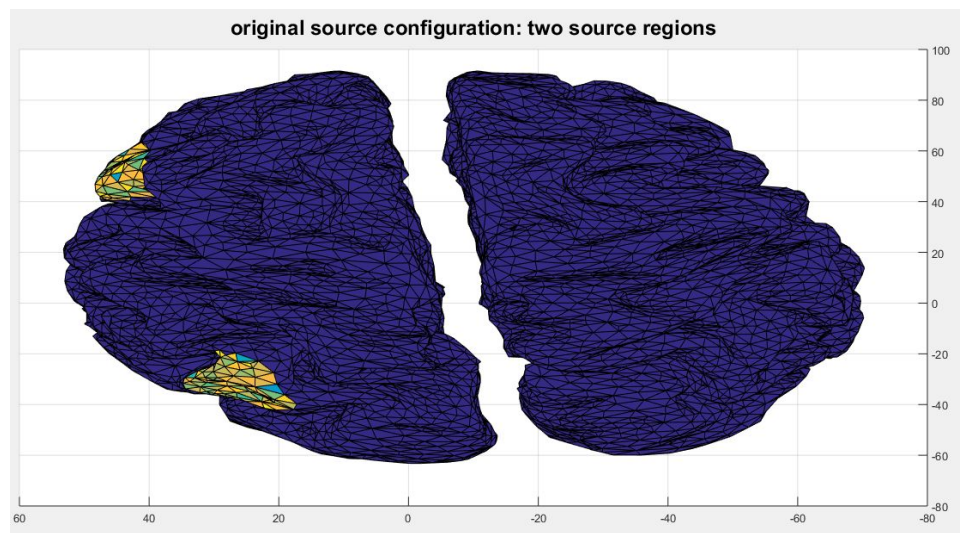


TP3 - Analyse des performances et comparaison des algorithmes étudiés

Le but de ce tp est de comparer les deux algorithmes d'une façon qualitative et quantitative. On utilise $\lambda=228$ pour le MNE (choisi selon le discrepancy principle) et $\alpha=0.1$ et $\lambda=10$ pour le SISSY (demandé dans l'énoncé).

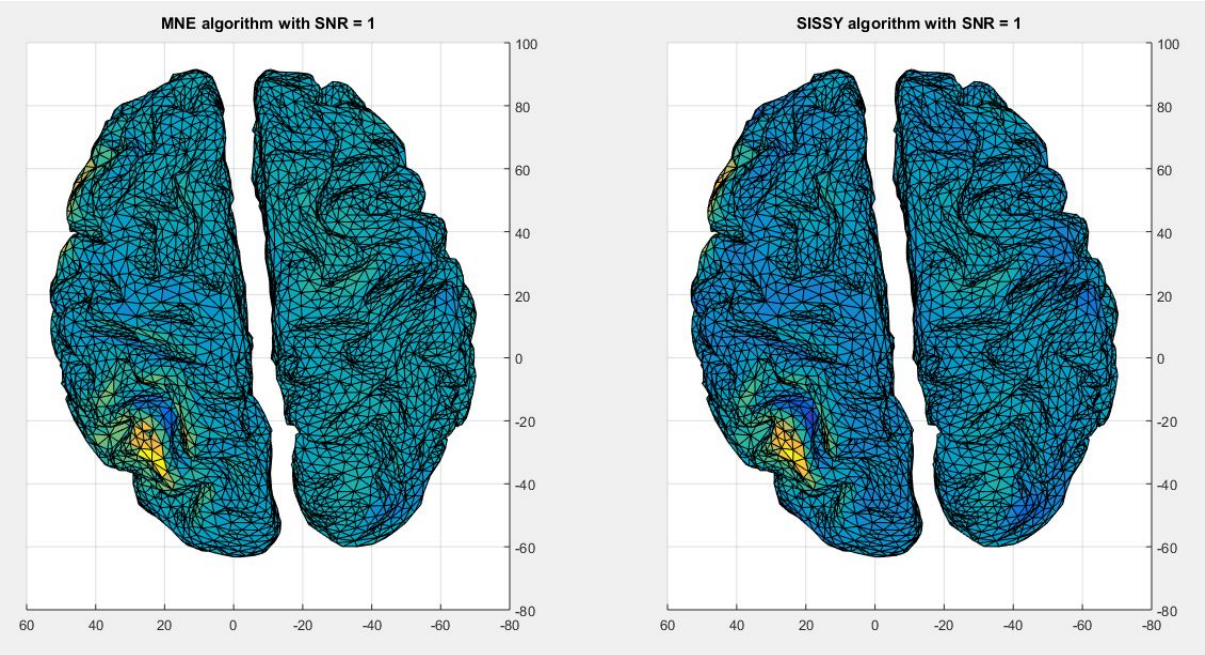
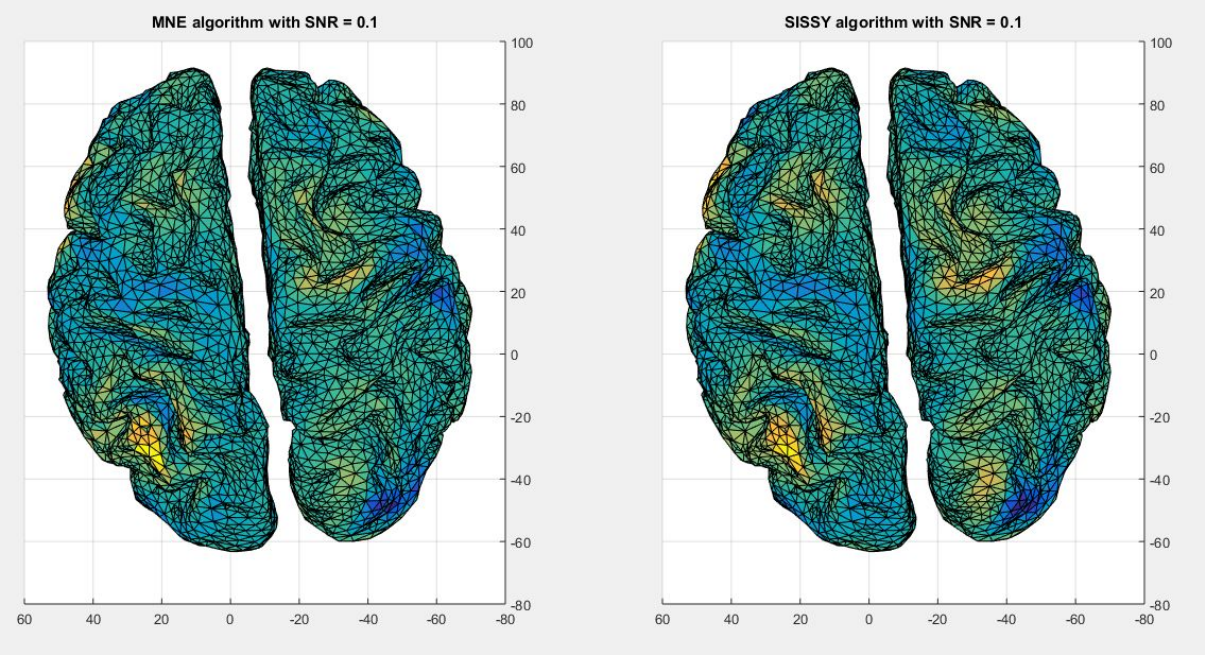
2.1) Comparaison qualitative

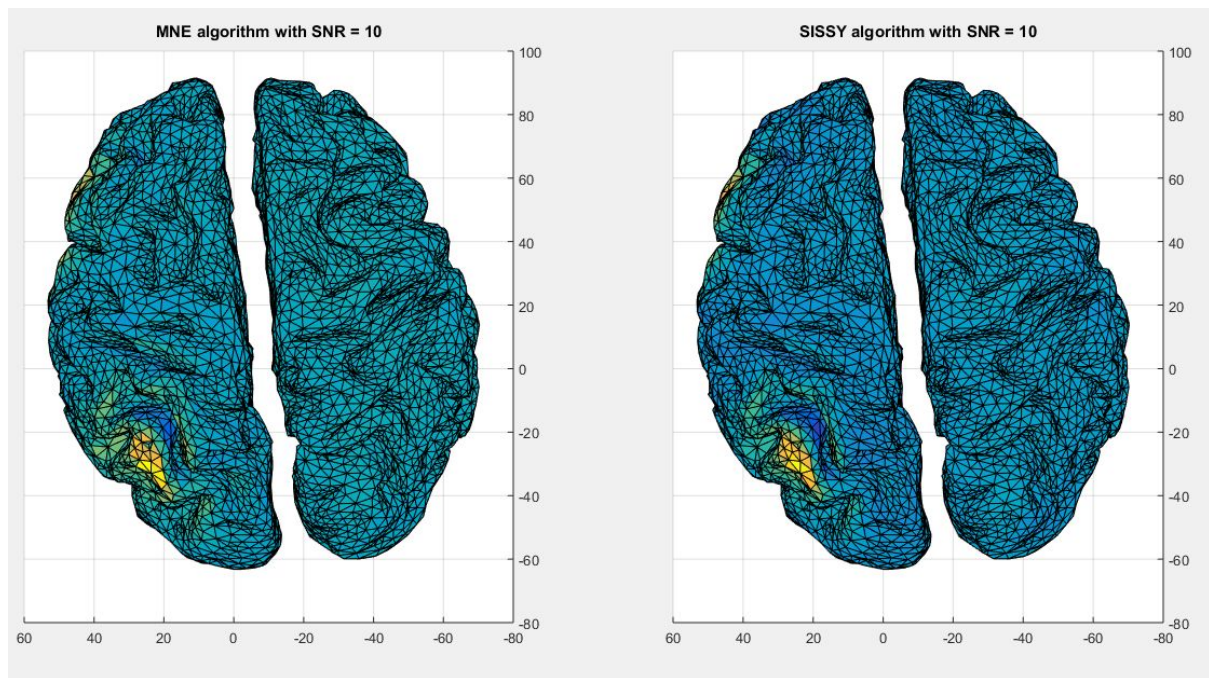
Dans cette partie, le but est de comparer les résultats des deux algorithmes d'une façon qualitative. Pour cela, on va comparer les solutions inverses avec la configuration original des sources, en utilisant la fonction trisurf. A continuation, on montre la configuration des sources originale à l'instant du pic du signal d'intérêt. On peut voir en jaune les deux parties du cerveau qui sont actives.



Bruit gaussien non corrélé

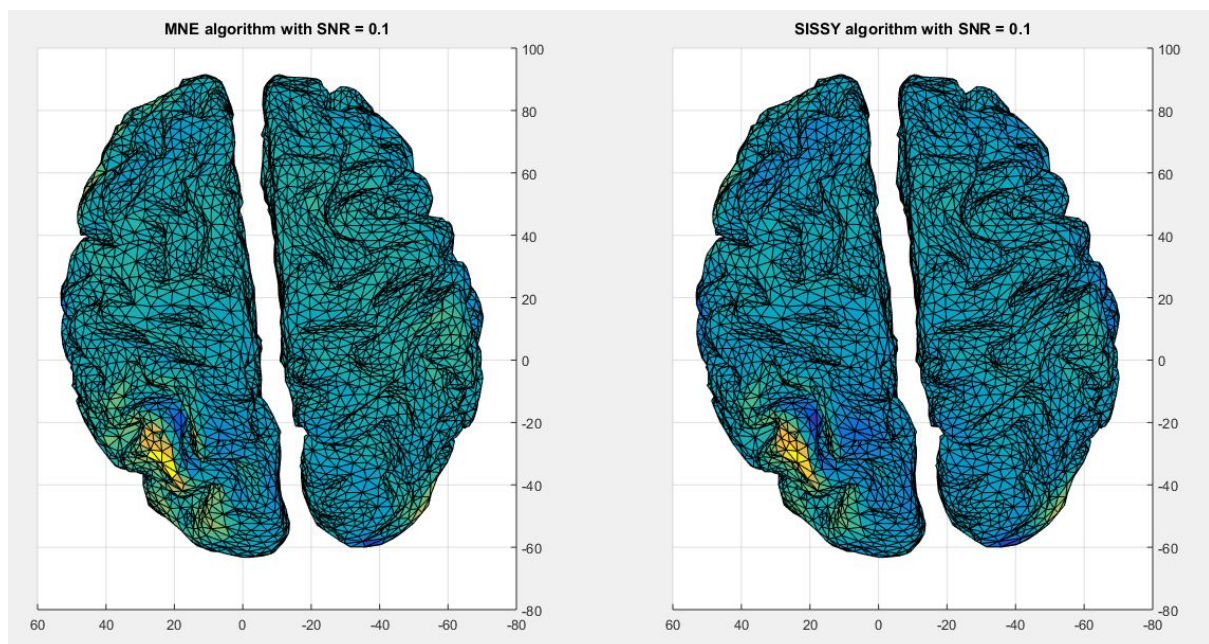
On visualise les solutions inverses des algorithmes SISSY et MNE pour des différentes valeurs de RSB (ou SNR dans les graphiques). Le bruit ajouté au niveau des capteurs est décorrélé. On peut facilement voir que la solution s'améliore en augmentant le RSB, comme prévu.

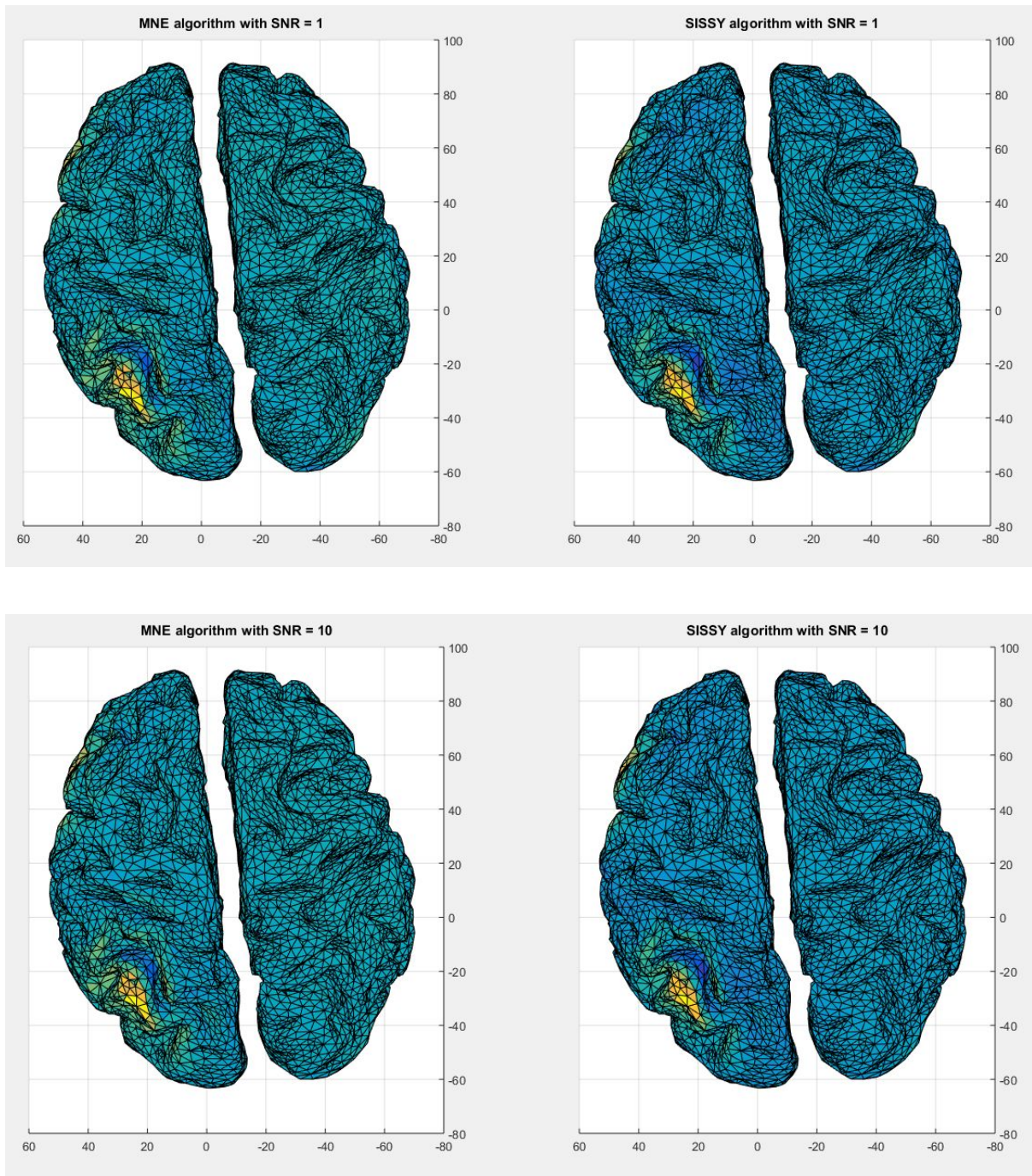




Bruit gaussien spatialement corrélé (activité de fond)

On visualise les solutions inverses des algorithmes mais, cette fois, en ajoutant du bruit spatialement corrélé. Cela correspond au modèle activité de fond et considère une activité gaussienne des dipôles qui ne sont pas actifs.





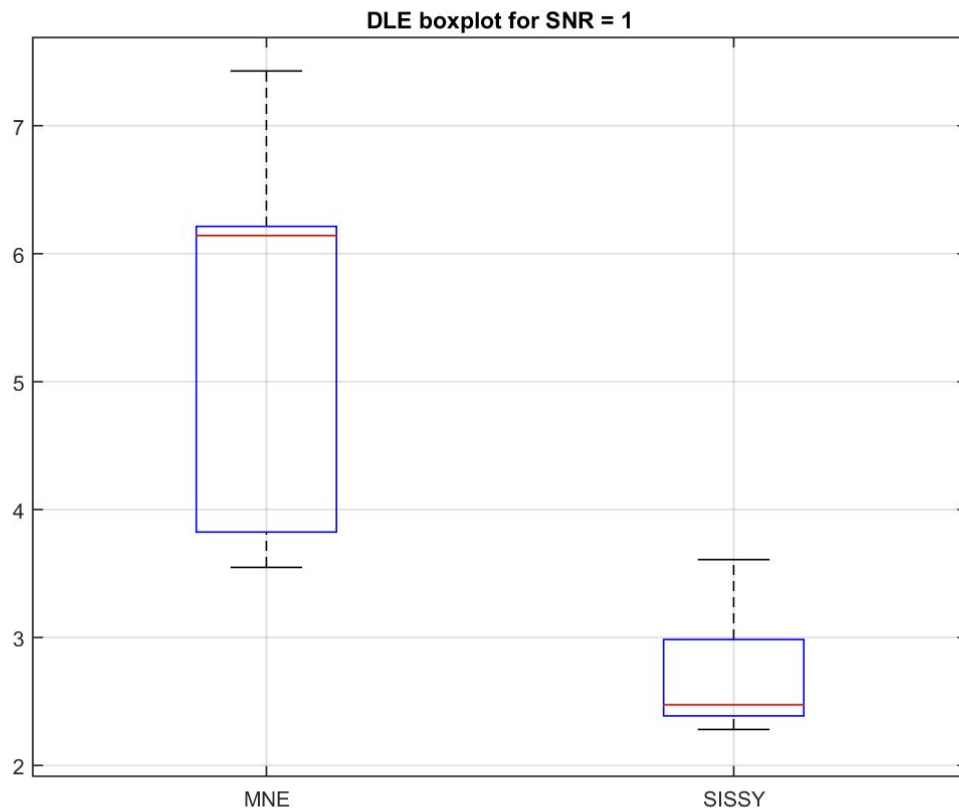
On note que les résultats sont meilleurs, ce qui est logique parce que le bruit corrélé donne plus d'information que le bruit blanc. Au même temps, les résultats s'améliorent quand la RSB croît.

2.2) Comparaison quantitative

Pour évaluer quantitativement les deux algorithmes, on utilise le critère DLE (Dipole Localization Error), détaillé dans l'énoncé du TP.

Graphique de type boxplot

En considérant le modèle d'activité de fond, on génère 10 réalisations du bruit corrélé et on estime la distribution des sources avec les deux algorithmes. On représente les résultats dans un graphique de type boxplot. Dans ce graphique, la médiane est représentée par la ligne rouge et les limites de la boîte indiquent les percentiles de 25 et 75. Les lignes horizontales noires représentent les valeurs maximales et minimales obtenues, sans inclure les *outliers*, ceux qui sont représentés avec des symboles '+').



On voit clairement que la méthode SISSY a des meilleurs résultats que la méthode MNE (DLE plus faible): elle a une médiane et une variance plus faibles que la méthode MNE. On souligne que dans aucune des méthodes on a obtenu des *outliers*.

DLE vs RSB

On répète l'expérience antérieure pour des différentes valeurs de RSB (SNR dans les graphiques) et on graphique la moyenne du DLE en fonction de la RSB. On obtient les courbes suivantes, ou on note que le DLE décroît avec le RSB (comme attendu) et que, pour des valeurs de RSB plus grandes qu'environ 0.3, le DLE de SISSY est plus petite que celui de MNE. Cela est cohérent avec la théorie et avec le graphique de type *boxplot* (parce que dans cette graphique, on utilise RSB=1 et on observe que le DLE de SISSY est plus faible que celui de MNE).

