Problème inverse d'EEG :
TP2 - Régularisation de Tikhonov et régularisation TV-L1

Master 2 SISEA (ISTIC)

Université de Rennes1

1 Introduction

Une approche populaire pour obtenir une solution au problème inverse consiste à régulariser le problème en résolvant le problème d'optimisation suivante :

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2 + \lambda f(\mathbf{s}). \tag{1}$$

Le premier terme assure que la solution reconstruit bien les données et le deuxième terme, appelé terme de régularisation, inclut les hypothèses sur les sources. Le paramètre de régularisation λ permet de gérer l'équilibre entre les deux termes. La proposition de différents termes de régularisation a donné lieu à différents algorithmes. Dans ce TP, nous nous concentrons sur des termes de régularisation de type norme L2 ($f(\mathbf{s}) = ||\mathbf{s}||_2^2$; régularisation de Tikhonov), exploité dans l'algorithme MNE (minimum norm estimate), et de type norme TV-L1 ($f(\mathbf{s}) = ||\mathbf{T}\mathbf{s}||_1 + \alpha||\mathbf{s}||_1$), utilisé par l'algorithme SISSY (source imaging based on structured sparsity). Ici, $\mathbf{T}$ est un opérateur linéaire qui implémente le gradient sur la surface du maillage modélisant le cortex. Les éléments $T_{e,d}$ de $\mathbf{T}$, $e = 1, \ldots, E$, $d = 1, \ldots, D$, où E est le nombre des arêtes du maillage, sont définis par :

$$T_{e,d} = \begin{cases} 1 & \text{si } d = d_{e,1} \\ -1 & \text{si } d = d_{e,2} \\ 0 & \text{sinon} \end{cases} \tag{2}$$

où $d_{e,1}$ et $d_{e,2}$ sont les indices des dipôles partageant la e-ième arête.

2 Préparation

L'objectif des méthodes de la localisation de sources (distribuées) en EEG consiste à résoudre le problème d'optimisation (1). Pour la plupart des termes de régularisation, comme, par exemple, pour la norme TV-L1, ceci nécessite de recourir à un algorithme itératif. Par contre, pour la norme L2, une solution analytique peut être déduite.

2.1 Algorithme MNE

Montrer que la solution analytique au problème de minimisation

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2 + \lambda||\mathbf{s}||_2^2 \tag{3}$$

peut s'écrire sous la forme :

$$\mathbf{s} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}} + \lambda\mathbf{I})^{-1}\mathbf{x}. \tag{4}$$

Remarque : utiliser le lemme d'inversion $(\mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{B} + \mathbf{P}^{-1})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{P}\mathbf{B}^{\mathrm{T}} + \mathbf{R})^{-1}$.

2.2 Algorithme SISSY

A partir du problème d'optimisation

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2 + \lambda(||\mathbf{z}||_1 + \alpha||\mathbf{y}||_1) \quad \text{s. t.} \quad \mathbf{z} = \mathbf{T}\mathbf{s},\ \mathbf{y} = \mathbf{s} \tag{5}$$

montrer que les équations de mise à jour de l'algorithme ADMM pour les variables $\mathbf{s}$, $\mathbf{z}$, $\mathbf{y}$ et les multiplicateurs de Lagrange $\mathbf{u}$ et $\mathbf{v}$ sont donnés par :

$$\mathbf{s}^{(k+1)} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \rho(\mathbf{T}^{\mathrm{T}}\mathbf{T} + \mathbf{I})\right)^{-1} \left(\mathbf{A}^{\mathrm{T}}\mathbf{x} + \rho\mathbf{T}^{\mathrm{T}}\mathbf{z}^{(k)} + \mathbf{T}^{\mathrm{T}}\mathbf{u}^{(k)} + \rho\mathbf{y}^{(k)} + \mathbf{v}^{(k)}\right) \tag{6}$$

$$\mathbf{z}^{(k+1)} = \mathrm{prox}_{\lambda/\rho}\left(\mathbf{T}\mathbf{s}^{(\mathrm{k}+1)} - \frac{1}{\rho}\mathbf{u}^{(\mathrm{k})}\right) \tag{7}$$

$$\mathbf{y}^{(k+1)} = \mathrm{prox}_{\lambda\alpha/\rho}\left(\mathbf{s}^{(\mathrm{k}+1)} - \frac{1}{\rho}\mathbf{v}^{(\mathrm{k})}\right) \tag{8}$$

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \rho\left(\mathbf{z}^{(k+1)} - \mathbf{T}\mathbf{s}^{(k+1)}\right) \tag{9}$$

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} + \rho\left(\mathbf{y}^{(k+1)} - \mathbf{s}^{(k+1)}\right) \tag{10}$$

où $\mathrm{prox}_\beta(\mathbf{y}) = \min_{\mathbf{x}} \frac{1}{2}||\mathbf{x} - \mathbf{y}||_2^2 + \beta||\mathbf{x}||_1$ est l'opérateur proximal associé avec la norme L1 et dont la solution est connue :

$$x_i = \begin{cases} y_i - \beta & \text{si } y_i > \beta \\ y_i + \beta & \text{si } y_i < -\beta \\ 0 & \text{sinon} \end{cases} \tag{11}$$

et ρ est un paramètre de pénalité (utiliser $\rho = 1$).

3 Manipulations

3.1 Implémentation de l'algorithme MNE

Ecrire une fonction `MNE(x,A,lambda)` qui implémente l'algorithme MNE. Tester l'algorithme pour $\lambda = 1$. Visualiser la solution obtenue à l'aide de la fonction `trisurf`.

3.2 Etude de l'algorithme MNE

1. Faire varier le paramètre de régularisation (entre 0.1 et 1000) et observer l'influence de ce paramètre sur le résultat de l'algorithme en comparant la solution inverse obtenue avec la configuration de source originale.
2. Faire varier le RSB (entre 0.1 et 10) et faire varier le paramètre de régularisation (entre 1 et 1000). Comparer les résultats. Que pouvez-vous conclure sur le choix du paramètre de régularisation adéquat en fonction du RSB ?
3. Dans la littérature, on trouve plusieurs heuristiques qui sont utilisées pour fixer le paramètre de régularisation. Par la suite, nous allons considérer trois de ces critères :

   L-curve criterion Cette heuristique consiste à tracer l'erreur de reconstruction $||\mathbf{x} - \mathbf{A}\mathbf{s}||_2$ en fonction de la norme L2 de la solution $||\mathbf{s}||_2$ pour différents paramètres de régularisation sur une double échelle logarithmique. La courbe ainsi obtenue est souvent en forme de L. Le paramètre de régularisation est alors choisi de manière à correspondre au pli du L.

   Discrepancy principle Cette heuristique est basée sur l'idée que l'erreur de reconstruction est due au bruit. Le paramètre de régularisation est alors choisie tel que la puissance de l'erreur de reconstruction correspond à peu près à la puissance du bruit, c'est-à-dire $||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2 \approx ||\mathbf{n}||_2^2$. La puissance du bruit doit, en pratique, être estimée préalablement à partir des données (mais elle est connue pour notre exemple de données simulées).

Generalized cross-validation Cette heuristique est basée sur l'idée qu'un bon choix du paramètre de régularisation devrait permettre de bien prédire des observations exclues dans le calcul de la solution inverse. Suivant le principe de validation croisée, le paramètre de régularisation est alors choisi tel qu'il minimise la fonction suivante qui caractérise l'erreur de prédiction généralisée :

$$\mathrm{GCV}(\lambda) = \frac{||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2}{(\mathrm{trace}\{\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^{\mathrm{T}}\})^2} = \frac{||\mathbf{x} - \mathbf{A}\mathbf{s}||_2^2}{(\mathrm{trace}\{\mathbf{I} - \mathbf{A}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}} + \lambda\mathbf{I})^{-1}\})^2}.$$

Tester les trois heuristiques pour le choix du paramètre de régularisation en fixant le RSB à 1 et en faisant varier le paramètre de régularisation entre 1e-10 et 1e10. Conclure sur l'efficacité des trois heuristiques dans le contexte du problème inverse en EEG.

3.3 Implémentation de l'algorithme SISSY

Ecrire une fonction `SISSY(x,A,T,lambda,alpha)` qui implémente l'algorithme ADMM pour résoudre le problème d'optimisation (5). La matrice `T` peut être calculé à l'aide de la fonction `variation_operator (mesh,'face')`. Pour diminuer la complexité numérique dans le calcul de la mise à jour du vecteur **s**, une décomposition matricielle (décomposition de Cholesky) et un lemme d'inversion sont utilisés. De ce fait, insérer les lignes suivantes au début du code (exécutées une fois au début de l'algorithme) :

```
P=sparse(rho*(T.'*T+speye(size(T,2))));
APi=A/P;
L=chol(eye(size(A,1))+APi*A.','lower');
s1=A.'*x;
```

et inclure les lignes suivantes dans la boucle pour la mise à jour du vecteur **s** à chaque itération de l'ADMM :

```
b=s1+rho*(T.'*(z+u/rho)+y+v/rho);
s=P\b-APi'*(L'\(L\(APi*b)));
```

Fixer le nombre d'itérations de l'ADMM à 60.
Tester l'algorithme pour RSB=10, $\lambda = 1$, $\alpha = 0.1$.

3.4 Etude de l'algorithme SISSY

1. Faire varier le paramètre de régularisation λ (entre 0.01 et 1000) et observer l'influence de ce paramètre sur le résultat de l'algorithme.
2. Choisir une valeur approprié pour le paramètre λ et faire varier le paramètre α de 0 à 1. Comparer la solution inverse avec la configuration de sources originale.
3. Conclure sur l'influence des deux paramètres de régularisation sur la solution inverse et du choix approprié de ces paramètres.
4. Pour choisir un paramètre de régularisation approprié, trouver le paramètre λ pour lequel la contrainte

$$f(\mathbf{s}) = ||\mathbf{T}\mathbf{s}||_0 + \alpha||\mathbf{s}||_0$$

que nous souhaitons réellement imposer (basée sur la norme L0 et non la norme L1) est minimale. Pour calculer la norme L0, considérer que tous les éléments pour lesquels $|s_d| < 0.01 \max_{d\in\{1,\ldots,D\}}(|s_d|)$ sont zéros. Tester cette heuristique pour le choix de λ et comparer avec le résultat obtenu en utilisant le discrepancy principle.