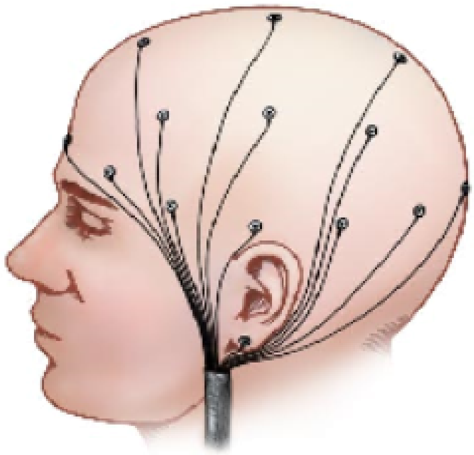
Problème Inverse en ElectroEncéphaloGraphie (EEG)



Amar Kachenoura[1,2]

[1]INSERM, U1099, Rennes, F-35000, France

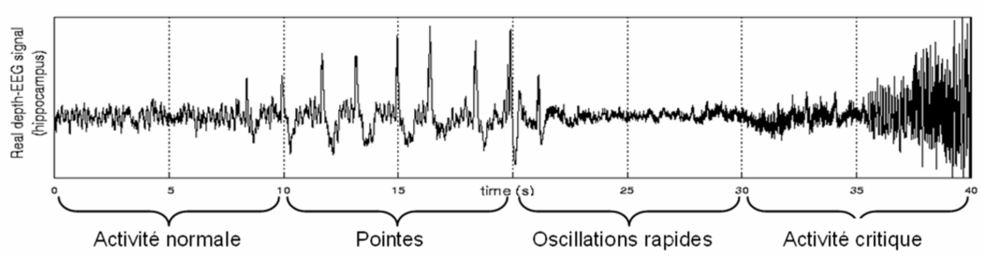
[2]Univ de Rennes , LTSI, Rennes, F-35000, France

Qu'est-ce que l'épilepsie ?

- Maladie neurologique caractérisée par la répétition de crises
  → conséquences socioprofessionnelles importantes

Trouble neurologique le plus courant
1% de la population

- Fonctionnement anormal, aigu et transitoire de l'**activité électrique du cerveau**

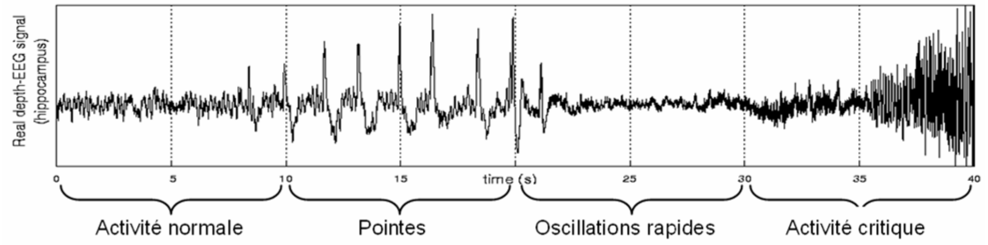


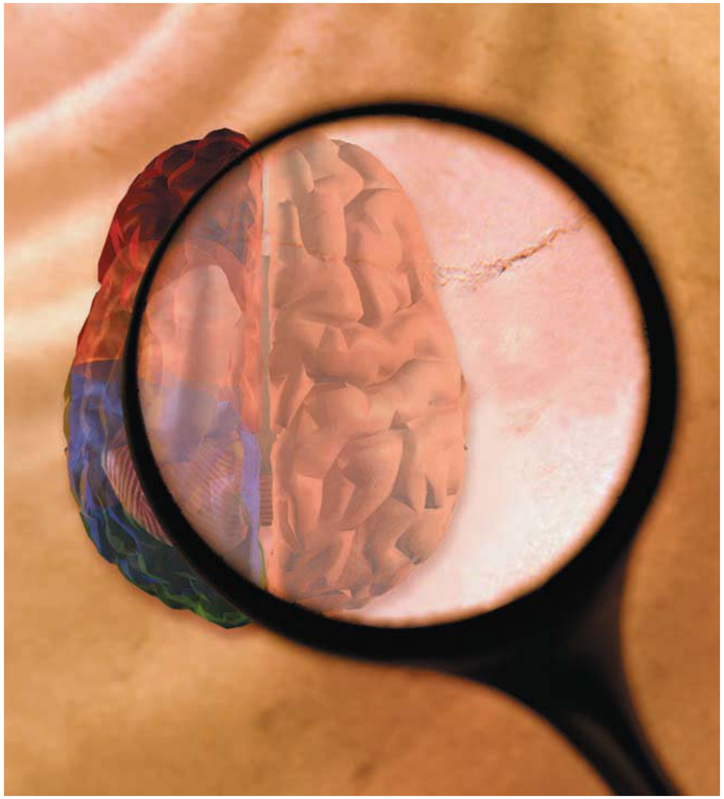
- Epilepsies **pharmaco-résistantes (30%)** → foyers épileptiques

Approches thérapeutiques?

- **Chirurgie :** résection du foyer épileptique
- **Prérequis :** localisation du ou des foyers épileptiques

  → utilisation de mesures intracérébrales de l'activité neuronale (**caractère invasif**)



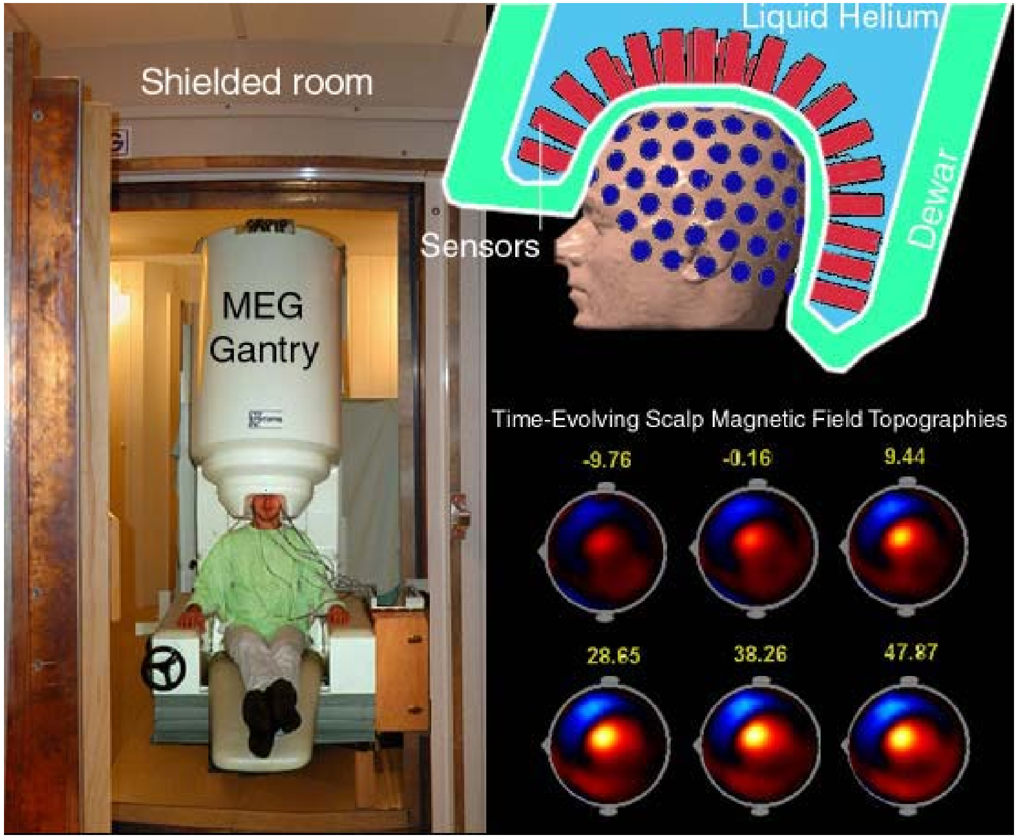


- Coût onéreux du processus occasionnant une liste d'attente considérable

  → ~300 bilans préopératoires seulement par an pour environ 10 000 candidats à la chirurgie

- **Méthodes non invasives :** l'EEG ou la MEG?

La MagnétoEncéphaloGraphie

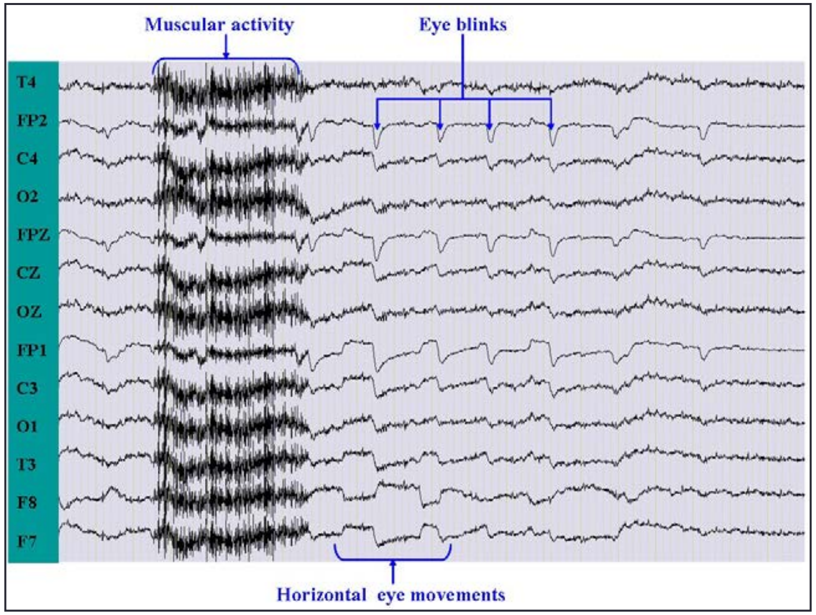


S. Baillet, J. C. Mosher, and R.M. Leahy. Electromagnetic brain mapping.
IEEE Signal Processing Magazine, 18(6):14–30, Nov. 2001. Avec permission.

- Champs magnétique issu de l'activité cérébrale enregistré en 1968 par David Cohen (physicien du MIT, spécialisé dans les blindages magnétiques)
- De l'ordre de 100 femtoteslas à la surface de la tête, soit 10 milliards de fois plus faible que le champs magnétique terrestre

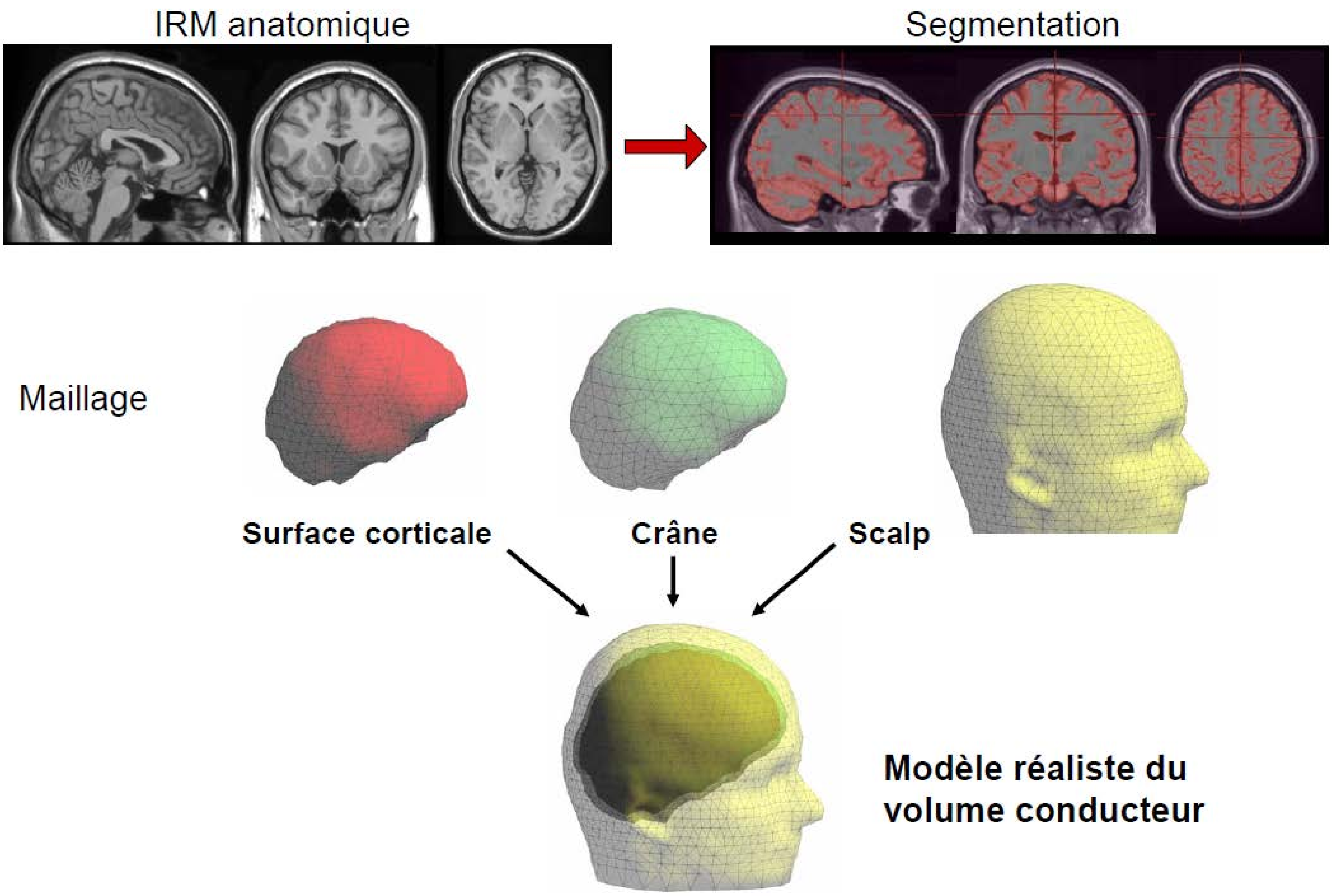
- Capteurs SQUID (Superconducting QUantum Interference Device, inventé par Jim Zimmerman en 1965) plongés dans un cryostat rempli d'hélium liquide à -269° C

L'ElectroEncéphaloGraphie

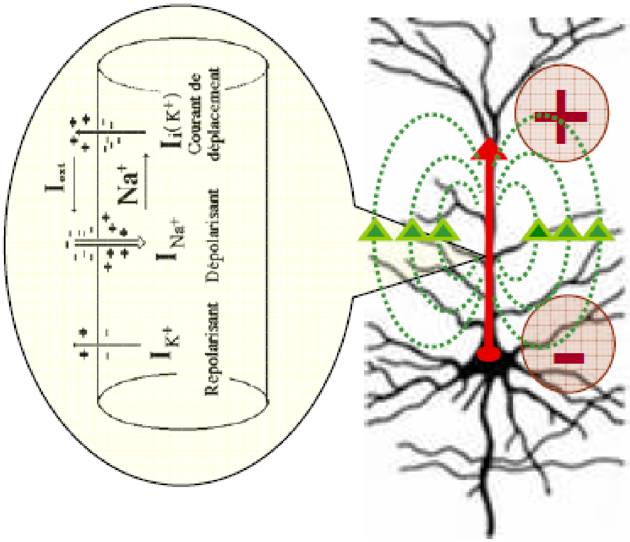


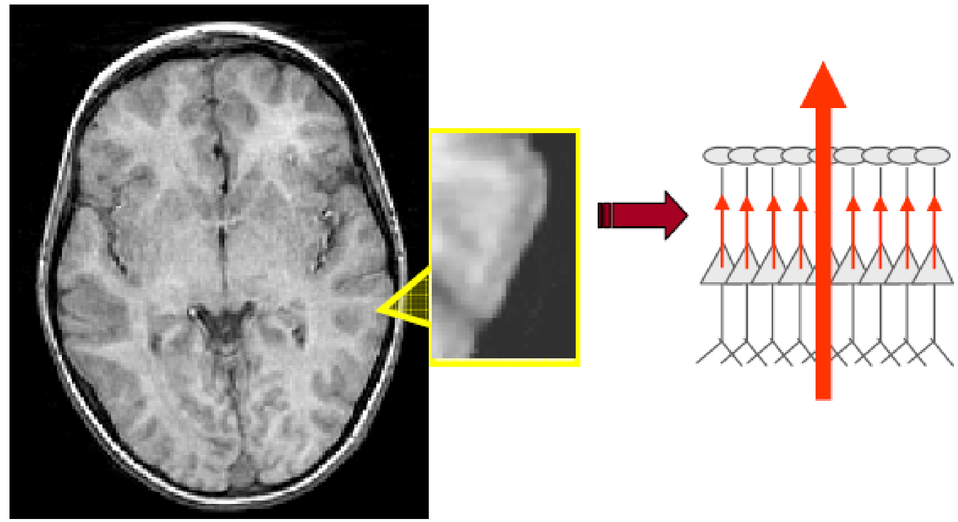
- Différences de **potentiel électrique** de l'ordre de la dizaine de microvolts
- Modalité la plus **rudimentaire** et la plus **ancienne** avec aujourd'hui au plus 256 électrodes
- **Premières expériences** menées sur des lapins et singes à même le cortex fin du XIXème
- Première **étude sur l'homme** en 1920 par Hans Berger (phys. allemand)

Modélisation du milieu : la tête



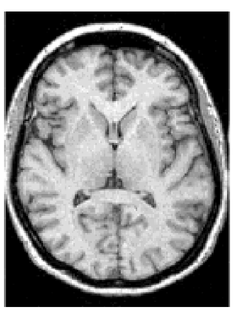
Modélisation de l'activité cérébrale : sources dipolaires distribuées





Maillage réaliste de la surface corticale

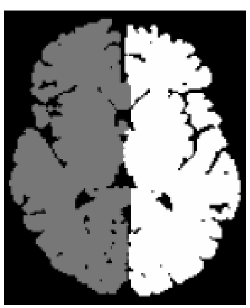
1 triangle = 1 dipôle



IRM anatomique

Segmentation (gris/blanc)

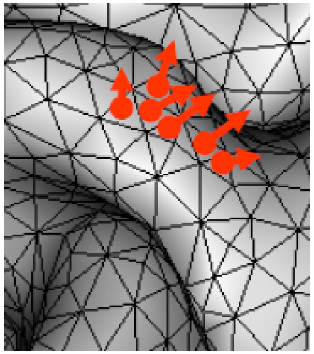
Séparation des 2 hémisphères



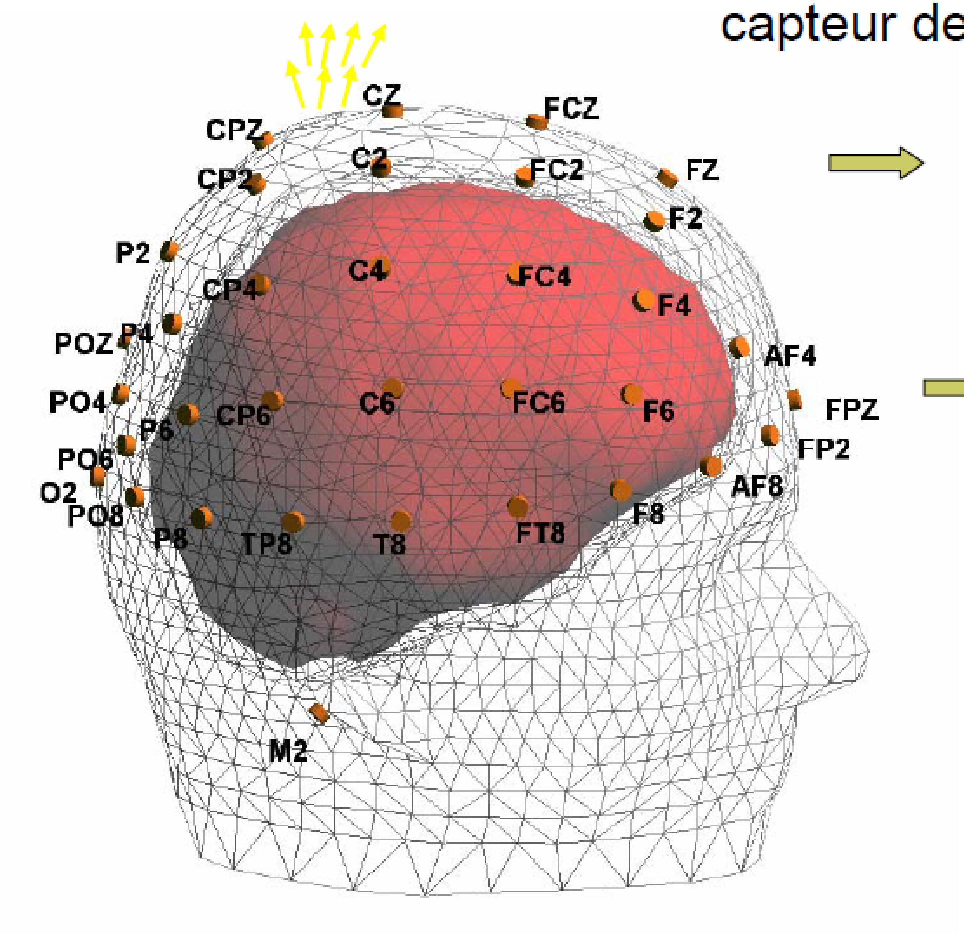
Volume binaire

Reconstruction de surface

Maillage surfacique (triangles)



Modélisation du transfert de l'activité cérébrale vers les électrodes EEG de surface

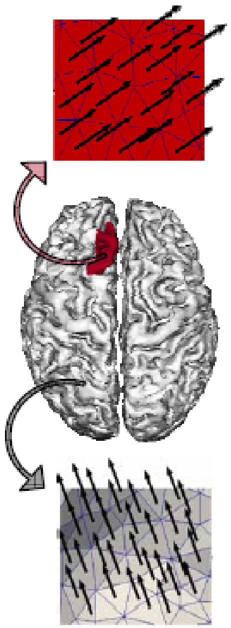


Calcul du transfert entre un dipôle de courant situé à la surface du cortex et un capteur de scalp

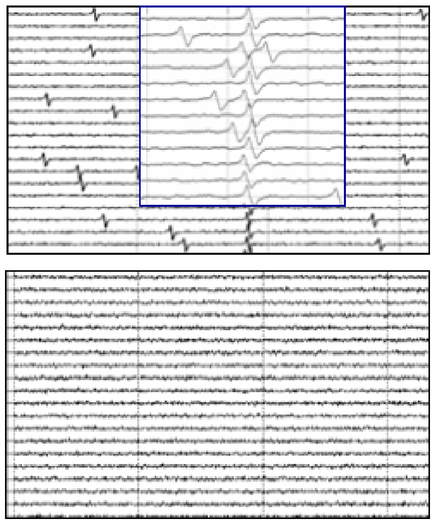
- Approximation quasi-statique des équations de Maxwell
- Calcul numérique complexe d'intégrales dans le cas d'un modèle de tête réaliste
- Obtention du vecteur de transfert $\mathbf{a}(\theta)$ pour chaque position θ

Localisation de générateurs épileptiques

Activité à la surface du cortex



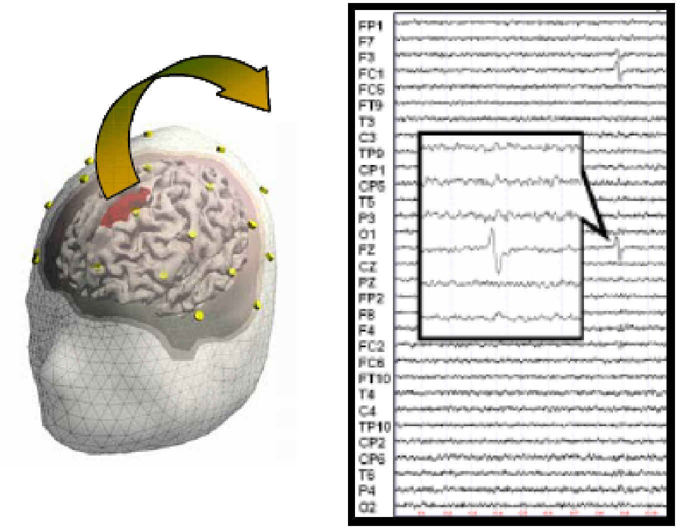
Activité épileptique quasi-similaire au sein d'une même source distribuée.



Activité de fond quasi-gaussienne.

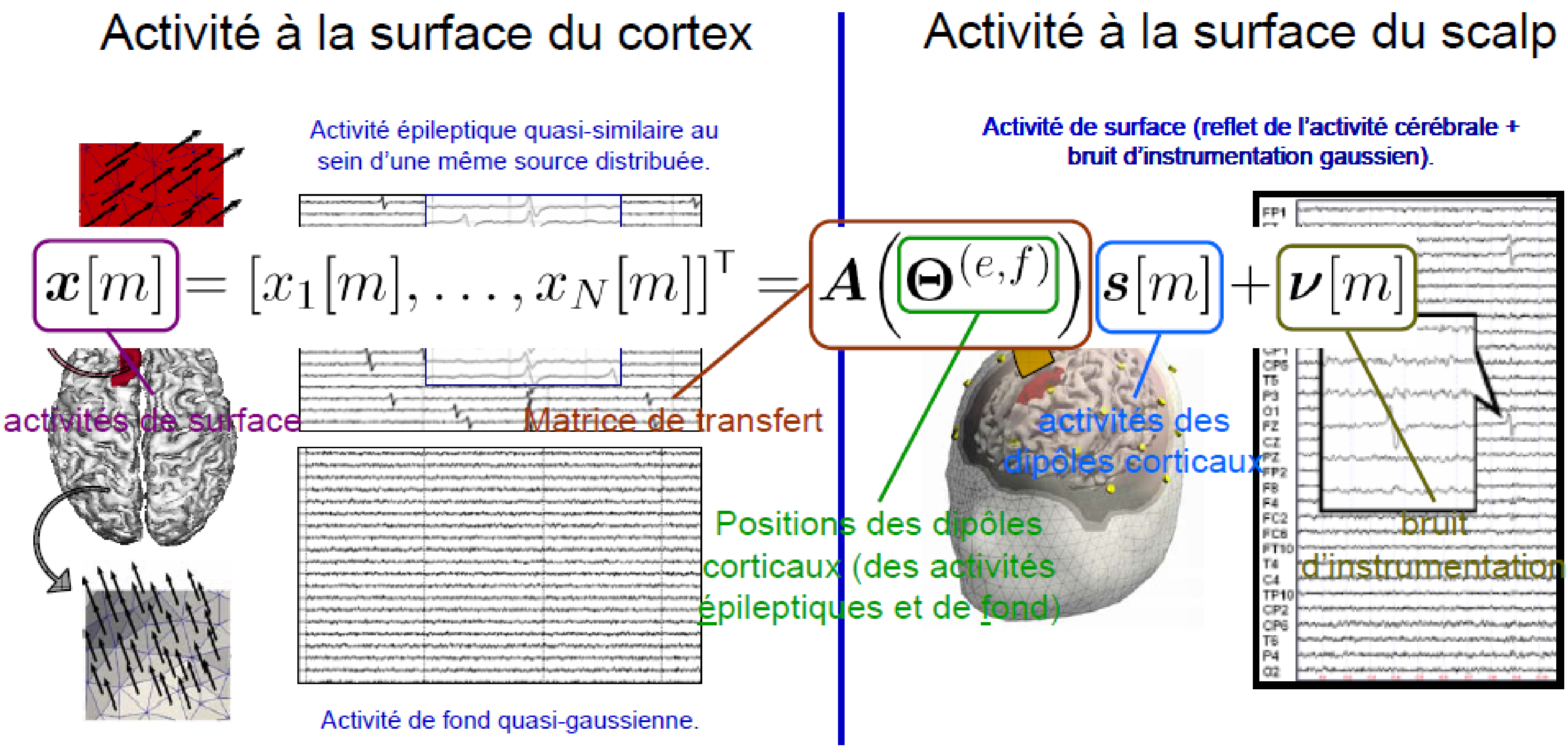
Activité à la surface du scalp

Activité de surface (reflet de l'activité cérébrale + bruit d'instrumentation gaussien).



- **Objectif :** localiser les générateurs cérébraux d'activité électrique pathologique et spatialement distribuée à partir de signaux de scalp
- **Difficultés :** problème mal posé, erreurs de modèle, faible SNR, etc,

Localisation de générateurs épileptiques



- **Objectif :** localiser les générateurs cérébraux d'activité électrique pathologique et spatialement distribuée à partir de signaux de scalp
- **Difficultés : problème inverse** mal posé, erreurs de modèle, faible SNR, etc,

Problème inverse mal posé

- **Problème direct :** le modelé physique adopté donne $\boldsymbol{x}$ connaissant $\boldsymbol{s}$ de manière implicite ($\boldsymbol{G}(\boldsymbol{s}\ ,\ \boldsymbol{x}) = \mathbf{0}$) ou explicite ($\boldsymbol{x} = \boldsymbol{G}(\boldsymbol{s})$), sous forme d'équations différentielles, aux dérivées partielles, variationnelles, intégrales, etc.

- **Nécessite d'inverser le modèle physique :** l'inversion est un problème mathématique mal posé

- **Problème mathématique bien posé au sens d'Hadamard :** problème dont la solution ***i)*** existe, ***ii)*** est unique et ***iii)*** est stable (i.e. la solution dépend continûment de la donnée).

- **Problème mathématique mal pose :** une au moins des conditions ***i)***, ***ii)*** et ***iii)*** n'est pas satisfaite

Méthode déterministes pour résoudre un PI

Problème inverse linéaire

$$x = Gs + n$$

Mesures : x

Paramètres inconnus : s

Opérateur linéaire connu (obtenu par discrétisation et connaissance en physique ou par calibration) : G

Bruit : n

Exemples :

- vectorisation d'images
- Localisation de sources (EEG, MEG, ECG, Télécommunication, …)

Méthode déterministes pour résoudre un PI

3 cas différents en absence de bruit $\boldsymbol{x} = \boldsymbol{G}\boldsymbol{s}$ avec $\boldsymbol{G} \in \mathbb{R}^{m \times n}$ de rang plein

$\rightarrow rang\,(\boldsymbol{G}) = min(m, n)$

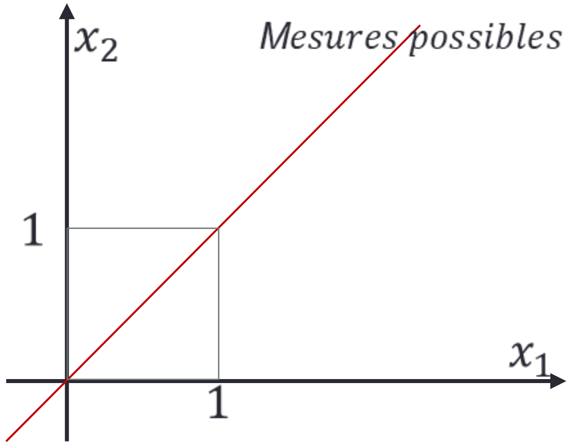
Cas 1 $m = n$ **:**

- Problème bien posé
- Solution unique $\boldsymbol{s} = \boldsymbol{G}^{-1}\boldsymbol{x}$

Cas 2 $m > n$ **:**

- Problème surdéterminé (plus de mesures que d'inconnues)
  $\rightarrow$ souvent pas de solution exacte

Exemple $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} s$



- Solution la plus proche au sens des moindres carrés
  $\min_{\boldsymbol{s}} \|\boldsymbol{x} - \boldsymbol{G}\boldsymbol{s}\|_F{}^2 + \lambda \|\boldsymbol{s}\|_F{}^2$

Méthode déterministes pour résoudre un PI

Solution

$$\min_{s} \|x - Gs\|_F^2 = \min_{s} (x - Gs)^T (x - Gs)$$

$$\Rightarrow J = x^T x + s^T G^T G s - 2 s^T G^T x$$

$$\frac{\partial J}{\partial s} = +2 G^T G s - 2 G^T x = 0$$

$$\Rightarrow 2 (G^T G) s = +2 G^T x$$

$$\Rightarrow s = (G^T G)^{-1} G^T x \quad \Big\| \quad \underbrace{(G^T G)^{-1} G^T}_{} = G^{\#}$$

$\hookrightarrow$ Pseudo Inverse

Solution Instable en présence de Bruit

→ Définition par SVD

$$G = U \Delta V^T = [U_s \; U_n] \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_s^T \\ V_n^T \end{bmatrix}$$

$$= U_s \Sigma_s V_s^T$$

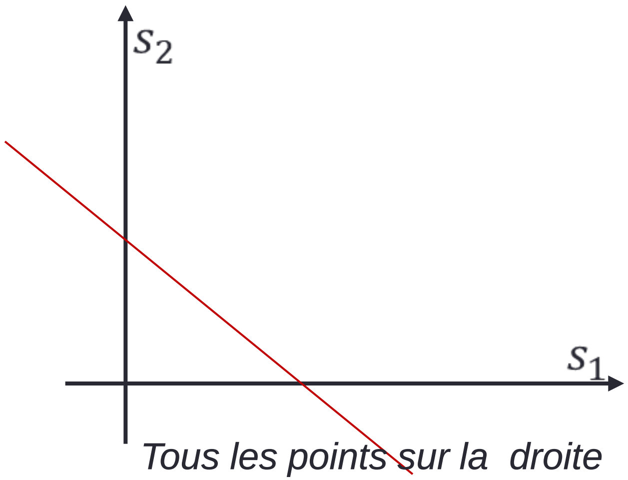
$$\boxed{\Rightarrow G^{\#} = V_s \Sigma_s^{-1} U_s^T}$$

Méthode déterministes pour résoudre un PI

Cas 3 $m < n$ ***:***

- Problème sous-déterminé (moins de mesures que d'inconnues) $\rightarrow$ une infinité de de solutions

Exemple $x = [1\ 1]\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = s_1 + s_2$ (pour x=1)



Tous les points sur la droite rouge sont des solutions

- Pour obtenir une solution il faut de l'information supplémentaire (s_1=1, s_1=s_2).

Notion de régularisation

- Transformer un problème mal posé en un problème bien posé en ajoutant de l'information supplémentaires (a priori) / des contraintes $\rightarrow$ en pratique nous avons souvent des informations supplémentaires sur la solution attendue,
- Différentes méthodes (avec diffèrent a priori) existent, le choix de la méthode dépend de l'application,

Méthode déterministes pour résoudre un PI

Régularisation de TiKhonov (Tychonoff)

Idée : trouver la solution à norme minimale $\rightarrow$ en pratique, pour des raisons physiques, la solution la plus plausible est souvent celle qui est associées à une énergie minimale,

Problème d'optimisation modifié

$$\min_{s} \underbrace{\|\boldsymbol{x} - \boldsymbol{G}\boldsymbol{s}\|_F^{\,2}}_{\text{Erreur de reconstruction (data fit - attache aux données)}} + \underbrace{\lambda}_{\text{Paramètre de régularisation}} \underbrace{\|\boldsymbol{s}\|_F^{\,2}}_{\text{Terme de régularisation (prior)}}$$

Terme de régularisation (prior)

Erreur de reconstruction (data fit - attache aux données)

Paramètre de régularisation qui permet de trouver un compromis entre l'erreur de reconstruction et l'influence du prior

Formulation alternative

$$\min\|\boldsymbol{s}\|_F^{\,2} \ \text{s.t.}\ \|\boldsymbol{x} - \boldsymbol{G}\boldsymbol{s}\|_F^{\,2} \leq \boldsymbol{\beta}$$

Paramètre de régularisation

Méthode déterministes pour résoudre un PI

Solution

$$J = \|x - Gs\|_F^2 + \lambda \|s\|_F^2 = (x - Gs)^T (x - Gs) + \lambda s^T s$$

$$J = x^T x - 2 s G^T x + s^T G^T G s + \lambda s^T s$$

$$\Rightarrow \frac{\partial J}{\partial s} = -2 G^T x + 2 G^T G s + 2 \lambda s = 0$$

$$\Rightarrow (G^T G + \lambda I) s = G^T x$$

$$\Rightarrow s = \underbrace{(G^T G + \lambda I)}_{n \times n}{}^{-1} G^T x$$

↳ Souvent de très grande Dimension

↓ Contraignent à la Pseudo inverse $G^\#$,

λI assure la stabilité de la solution.

Méthode déterministes pour résoudre un PI

→ Lemme d'inversion :

$$\left(B^T R^{-1} B + P^{-1}\right)^{-1} B^T R^{-1} = P B^T \left(B P B^T + R\right)^{-1}$$

Donc

$B^T = G^T$, $R^{-1} = I$

$P^{-1} = I$.

$\cancel{s = (G^T G + \lambda I)^{-1} G^T x = \lambda I G^T (G}$

$$\Rightarrow \quad s = \left(G^T G + \lambda I\right)^{-1} G^T x$$

$$s = G^T \underbrace{\left(G G^T + \lambda I\right)^{-1}}_{m \times m} x$$

↳ Souvent de Taille Bien plus raisonnable.

Choix du paramètre de régularisation λ

En pratique, on peut tester différentes valeur de λ et choisir celle qui semble donner un bon résultats. Cependant, il est souvent nécessaire de choisir λ automatiquement. Pour ce faire, on utilise des heuristiques (règles pratiques empiriques).

Critère basé L-Curve

Tracer $\|\boldsymbol{s}\|_F$ vs $\|\boldsymbol{x}-\boldsymbol{G}\boldsymbol{s}\|_F$ pour différents λ sur une échelle double logarithmique

- Pour montrer le compromis entre l'erreur de reconstruction et le terme régularisation;
- Fonction à décroissance monotone;
- Choisir λ au point associé avec la plus grande courbature (max de f'').

Principe d'incertitude (discrepancy principle)

L'idée est que l'erreur de reconstruction est due au bruit, Il faut choisir λ tel que :

$$\|\boldsymbol{x}-\boldsymbol{G}\boldsymbol{s}\|_F^{\,2} \approx \lambda \|\boldsymbol{n}\|_F^{\,2}$$

$\rightarrow$ nécessite l'information sur le niveau du bruit.

Méthode déterministes pour résoudre un PI

Lien entre la régularisation de TiKhonov et le modèle probabiliste (TD)

Exercise : Soit $\boldsymbol{x} \in \mathbb{R}^{\mathrm{N}}$ un vecteur d'observations recueillies à partir d'un réseau de N électrodes tel que :

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\nu} \tag{1}$$

où $\boldsymbol{s} \in \mathbb{R}^P$ est un vecteur de sources que l'on cherche à estimer, $\boldsymbol{A} \in \mathbb{R}^{N \times P}, N \leq P$ est la matrice qui représente le transfert entre l'espace sources est l'espace observations et $\nu \in \mathbb{R}^N$ est le vecteur d'un bruit additif supposé gaussien $\mathcal{N}(0, \boldsymbol{\Sigma}_\nu)$.

a. Estimer le vecteur source $\boldsymbol{s}$ au sens de moindre carrée sous l'hypothèse énergie minimale de cette dernière (principe de régularisation de Tikhonov)

b. Comparer le résultat obtenu de la question précédente avec celui obtenu sans considérer l'hypothèse de l'énergie minimale de sources (vu en cours)

c. Supposons maintenant que $\boldsymbol{x}$ et $\boldsymbol{s}$ sont deux vecteurs aléatoires issus des processus stochastiques gaussiennes $\mathcal{X} \rightsquigarrow \mathcal{N}(\boldsymbol{x}, \boldsymbol{\Sigma}_{\mathcal{X}})$ et $\mathcal{S} \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathcal{S}})$, respectivement. Montrer que l'estimation bayésienne de $\boldsymbol{s}$ est égale à celle trouvée au sens de moindre carrée avec la régularisation de Tikhonov.

Méthode déterministes pour résoudre un PI

Solution

c. D'un point de vu probabiliste, $\boldsymbol{x}$ et $\boldsymbol{s}$ sont deux vecteurs aléatoires vus comme deux réalisations de deux processus stochastiques, $\mathcal{X}$ et $\mathcal{S}$, respectivement. L'estimation bayésienne de $\boldsymbol{s}$ par la probabilité à posteriori $p(\mathcal{S}|\mathcal{X})$ est basée, comme son nom l'indique, sur la théorème de Bayes :

$$p(\mathcal{S}|\mathcal{X}) = \frac{p(\mathcal{X}|\mathcal{S})p(\mathcal{S})}{p(\mathcal{X})} \tag{3}$$

où $p(\mathcal{X}|\mathcal{S})$ est la vraisemblance de $\mathcal{X}$ par rapport au paramètre $\mathcal{S}$. $p(\mathcal{S})$ est l'a priori sur la paramètre $\mathcal{S}$ et $p(\mathcal{X})$ est l'a priori sur les observations. Normalement $1/p(\mathcal{X})$ est considéré comme un facteur de normalisation. La solution recherchée $\boldsymbol{s}$, pourra être considérée comme la moyenne de la processus $\{\mathcal{S}\}$. Ainsi, estimer $\boldsymbol{s}$ est possible par un calcul de la moyenne de la distribution à posteriori $p(\mathcal{S}|\mathcal{X})$. On sait maintenant que : $\mathcal{X} \rightsquigarrow \mathcal{N}(\boldsymbol{x}, \boldsymbol{\Sigma}_{\mathcal{X}})$ et $\mathcal{S} \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathcal{S}})$, alors :

$$\begin{aligned} p(\mathcal{S}) &\approx exp(-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S}) && (4) \\ p(\mathcal{X}) &\approx exp(-\frac{1}{2}(\mathcal{X}-\boldsymbol{x})^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}(\mathcal{X}-\boldsymbol{x})) && (5) \\ \Rightarrow p(\mathcal{S}|\mathcal{X}) &\approx exp(-\frac{1}{2}(\mathcal{X}-\boldsymbol{x})^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}(\mathcal{X}-\boldsymbol{x}) - \frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S}) && (6) \end{aligned}$$

Méthode déterministes pour résoudre un PI

On sait aussi que $\mathcal{X} = \boldsymbol{A}\mathcal{S}$, alors :

$$
\begin{aligned}
\Rightarrow p(\mathcal{S}|\mathcal{X}) &\approx exp(-\frac{1}{2}(\boldsymbol{A}\mathcal{S}-\boldsymbol{x})^{\intercal}\boldsymbol{\Sigma}_{\boldsymbol{A}\mathcal{S}}^{-1}(\boldsymbol{A}\mathcal{S}-\boldsymbol{x})-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S}) && (7)\\
&\approx exp(-\frac{1}{2}(\boldsymbol{A}\mathcal{S}-\boldsymbol{x})^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}(\boldsymbol{A}\mathcal{S}-\boldsymbol{x})-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S} && (8)\\
&\approx exp(-\frac{1}{2}(\mathcal{S}^{\intercal}\boldsymbol{A}^{\intercal}-\boldsymbol{x}^{\intercal})(\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}\mathcal{S}-\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x})-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S}) && (9)\\
&\approx exp(-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}\mathcal{S}+\mathcal{S}^{\intercal}\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}+\frac{1}{2}\boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}-\frac{1}{2}\mathcal{S}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}\mathcal{S}) && (10)\\
&\approx exp(-\frac{1}{2}\mathcal{S}^{\intercal}(\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1})\mathcal{S}+\mathcal{S}^{\intercal}\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}+\frac{1}{2}\boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}) && (11)\\
&\approx exp(-\frac{1}{2}\mathcal{S}^{\intercal}(\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1})\mathcal{S}+\boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}\mathcal{S}+\frac{1}{2}\boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}) && (12)
\end{aligned}
$$

Il est claire que la matrice $\boldsymbol{B} = \boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1}$ est une matrice symétrique. D'après l'algèbre linéaire, on a l'égalité suivante :

$$
-\frac{1}{2}\boldsymbol{z}^{\intercal}\boldsymbol{B}\boldsymbol{z}+\boldsymbol{h}^{\intercal}\boldsymbol{z} = -\frac{1}{2}(\boldsymbol{z}-\boldsymbol{B}^{-1}\boldsymbol{h})^{\intercal}\boldsymbol{B}(\boldsymbol{z}-\boldsymbol{B}^{-1}\boldsymbol{h})+\frac{1}{2}\boldsymbol{h}^{\intercal}\boldsymbol{A}^{-1}\boldsymbol{h} \qquad (13)
$$

Alors en substituant (13) dans (12) avec $\boldsymbol{h}^{\intercal} = \boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}$ et $\boldsymbol{z} = \mathcal{S}$, on obtient :

$$
p(\mathcal{S}|\mathcal{X}) \approx exp(-\frac{1}{2}(\mathcal{S}-(\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1})^{-1}\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x})^{\intercal}(\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1})(\mathcal{S}-(\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{A}+\boldsymbol{\Sigma}_{\mathcal{S}}^{-1})^{-1}\boldsymbol{A}^{\intercal}\boldsymbol{\Sigma}_{\mathcal{X}}^{-1}\boldsymbol{x}))
$$

Méthode déterministes pour résoudre un PI

Il est claire que la distribution $p(\mathcal{S}|\mathcal{X})$ est aussi gaussienne avec une moyenne

$$\hat{\boldsymbol{s}} = (\boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{A} + \boldsymbol{\Sigma}_{\mathcal{S}}^{-1})^{-1} \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} \tag{14}$$

Si $\boldsymbol{\Sigma}_{\mathcal{S}} = \sigma_{\mathcal{S}}^2 \mathbf{I}$ et $\boldsymbol{\Sigma}_{\mathcal{X}} = \sigma_{\mathcal{X}}^2 \mathbf{I}$, alors :

$$\hat{\boldsymbol{s}} = (\boldsymbol{A}^\intercal \boldsymbol{A} + \frac{\sigma_{\mathcal{X}}^2}{\sigma_{\mathcal{S}}^2} \mathbf{I})^{-1} \boldsymbol{A}^\intercal \boldsymbol{x} = (\boldsymbol{A}^\intercal \boldsymbol{A} + \lambda \mathbf{I})^{-1} \boldsymbol{A}^\intercal \boldsymbol{x} \tag{15}$$

avec $\lambda = \sigma_{\mathcal{X}} / \sigma_{\mathcal{S}}$.

Une autre manière pour trouver cette estimation de $\boldsymbol{s}$ est de maximiser $p(\mathcal{S}|\mathcal{X})$, alors :

$$\begin{aligned}
&\max_{\mathcal{S}} p(\mathcal{S}|\mathcal{X}) \Leftrightarrow \min_{\mathcal{S}} f(\mathcal{S}) = -\frac{1}{2}\mathcal{S}^\intercal \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{A} \mathcal{S} + \mathcal{S}^\intercal \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} + \frac{1}{2} \boldsymbol{x}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} - \frac{1}{2} \mathcal{S}^\intercal \boldsymbol{\Sigma}_{\mathcal{S}}^{-1} \mathcal{S} \\
&\Rightarrow \frac{\partial f}{\partial \mathcal{S}} = -\boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{A} \mathcal{S} + \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} - \boldsymbol{\Sigma}_{\mathcal{S}}^{-1} \mathcal{S} = 0 \qquad (16) \\
&\Rightarrow \frac{\partial f}{\partial \mathcal{S}} = -(\boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{A} + \boldsymbol{\Sigma}_{\mathcal{S}}^{-1}) \mathcal{S} + \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} = 0 \qquad (17) \\
&\Rightarrow \hat{\mathcal{S}} = (\boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{A} + \boldsymbol{\Sigma}_{\mathcal{S}}^{-1})^{-1} \boldsymbol{A}^\intercal \boldsymbol{\Sigma}_{\mathcal{X}}^{-1} \boldsymbol{x} \qquad (18)
\end{aligned}$$

On note que la maximum d'une gaussienne coincïde avec sa moyenne.