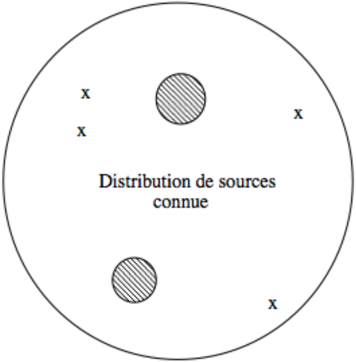
Une courte introduction aux problèmes inverses

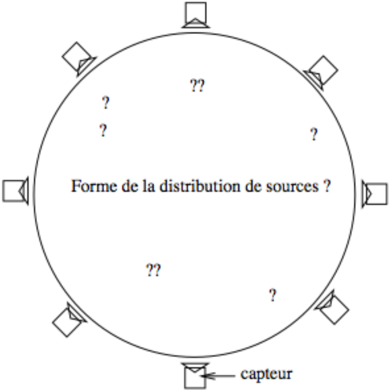
Matthieu Kowalski

L2S – CNRS - Supelec - Univ. Paris Sud

Problème inverse : définition a contrario

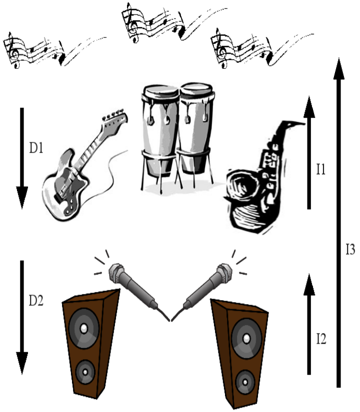


Problème direct



Problème inverse

Exemple : problèmes inverses en audio



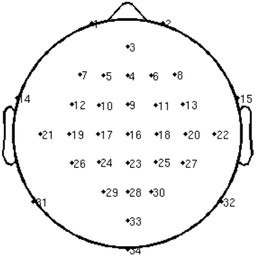
Problèmes directs

1. Synthèse de sons et d'instruments
2. Mélange des signaux

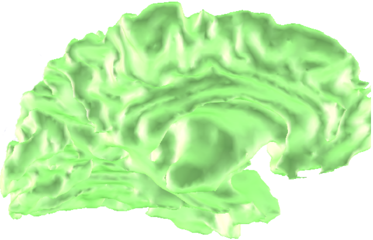
Problèmes inverses

1. Transcription automatique monophonique
2. Séparation de sources
3. Transcription automatique polyphonique

Exemple (suite) : problème inverse en MEEG



? ⇓?



À partir d'un enregistrement EEG sur le scalp comment localiser les sources dans le cerveau ?

Problème inverse : formalisation

$\mathcal{A}($	$\mathbf{y}$,	$\mathbf{x}$,	$\mathbf{z}$,	$\mathbf{b})$	$= 0$
modèle	mesures	grandeurs inconnues	grandeurs intermédiaires	erreurs et bruits	

Relation explicite : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}, \mathbf{b})$

Erreur en sortie : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}) \diamond \mathbf{b}$

Erreur additive : $\mathbf{y} = \mathcal{A}(\mathbf{x}, \mathbf{z}) + \mathbf{b}$

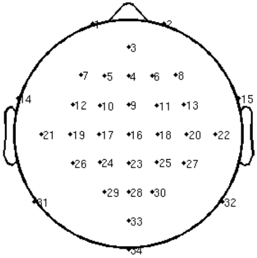
Relation entre **x** et **z** : $\begin{cases} \mathbf{y} = \mathcal{A}_1(\mathbf{x}, \mathbf{z}) + \mathbf{b} \\ \mathcal{A}_2(\mathbf{x}, \mathbf{z}) = 0 \end{cases}$

Modèle non linéaire simple : $\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b}$

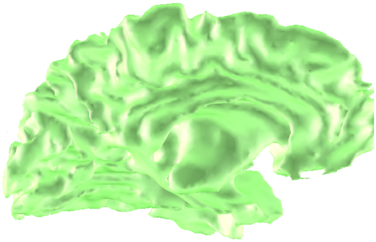
Modèle linéaire : $\mathbf{y} = \mathcal{A}\mathbf{x} \diamond \mathbf{b}$

Modèle linéaire + bruit additif : $\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{b}$

Exemple (suite) : problème inverse en MEEG



$? \Downarrow ?$



À partir d'un enregistrement EEG sur le scalp comment localiser les sources dans le cerveau ?

enregistrements : $M \in \mathbb{R}^{N \times KT}$

Matrice de "diffusion" connue : $G \in \mathbb{R}^{N \times I}$

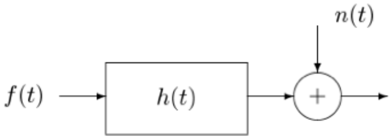
sources : $X \in \mathbb{R}^{I \times KT}$

$$M = GX + E$$

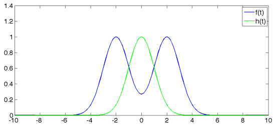
avec $N \approx 150$ et $I \approx 18000$

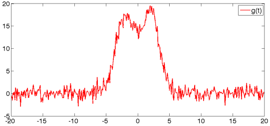
Problème **très** sous-déterminé

Déconvolution

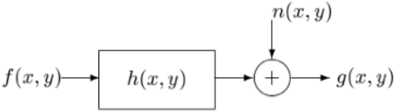


$$g(t) = h(t) \star f(t) + n(t)$$
$$= \int_{-\infty}^{+\infty} f(t) h(x-t) \,\mathrm{d}\, t + n(t)$$



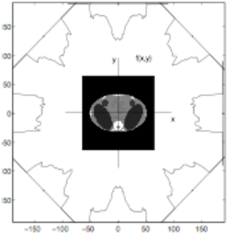


Restauration d'image



$$g(x,y) = h(x,y) \star f(x,y) + n(x,y)$$

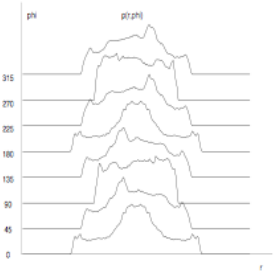
Tomographie



$$f(x,y) \longrightarrow \boxed{\text{TR}} \longrightarrow p(r,\phi)$$

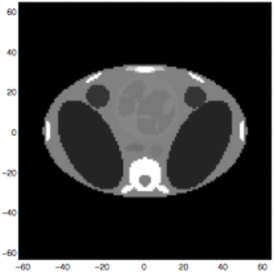
$$p(r,\phi) = \int_{L_{r,\phi}} f(x,y)\ \mathrm{d}l$$

$$p(r,\phi) = \iint_D f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,\mathrm{d}x\,\mathrm{d}y$$



?

$\Longrightarrow$



Exemple élémentaire (1)

But : trouver le spectre ($|\hat{x}(\nu)|^2$) d'un signal.

On observe : modèle continu

$$y(t) = h(t)x(t) + b(t)$$

Après échantillonage : modèle discret

$$y_n = h_n x_n + b_n \ , \ \ n = 1, \ldots, N$$

1ère difficulté : indépendament de h, le caractère discret fini des données.

Conséquence

- Périodisation de la transformée de Fourier.
- $\hat{x}$ doit être à support fini et il faut éviter les repliements.

Exemple élémentaire (2)

Le modèle s'écrit :

$$y_n = \int_0^1 \hat{h} \star \hat{x}(\nu) e^{2i\pi\nu n} \, \mathrm{d}\nu + b_n \; \forall n = 1, \ldots, N$$

- Perte de résolution due au noyau de convolution $\hat{h}$.
- on observe "seulement" N echantillons

2ème difficulté : impossibilité d'inverser mathématiquement le modèle, même si h est connue.

Demo Matlab

Problème inverse « mal posé »(1)

On a identifié plusieurs problèmes :

1. On peut trouver une infinité de solutions (ici on a fait du "zeros padding") ;
2. Une petite pertubation due au bruit peut avoir de très grosses conséquences sur l'estimation. (Amplification des erreurs en $1/h_n$ dans le problème précédent)

Problème inverse « mal posé »(2)

Hadamard défini un problème « **bien posé** »si :

1. Pour chaque donnée y dans une classe définie $\mathcal{Y}$, il existe une solution dans une classe prescrite $\mathcal{X}$ **(existence)**
2. La solution est unique dans $\mathcal{X}$ **(unicité)**
3. La dépendance de x par rapport à y est continue, c'est-à-dire que lorsque l'erreur δy sur la donnée y tend vers 0, l'erreur induite δx induite sur x tend aussi vers 0 **(continuité)**

La notion de continuité correspond aux notions de **stabilité** ou de **robustesse** de la solution.

Inversion directe - deconvolution par filtrage inverse

$$y(t) = h(t) \star x(t) + b(t) \longrightarrow \text{TF} \longrightarrow Y(\omega) = H(\omega)\,X(\omega) + B(\omega) \longrightarrow$$

$$\hat{X}(\omega) = \frac{1}{H(\omega)}\,Y(\omega) \longrightarrow \text{TF inverse} \longrightarrow \hat{x}(t)$$

Problème : Amplification du bruit

$$\hat{X}(\omega) = X(\omega) + \frac{B(\omega)}{H(\omega)}$$

Méthode analytique

On essaie d'inverser “directement le modèle” :

- Inversion direct des observations ;
- Inversion dans un espace dual (e.g. Fourier) ;
- Inversion par décomposition tronquée sur une base.

Limitations :

- Méthode limité aux modèles simples ;
- Bruit non pris en compte ;
- Nécessite un jeu complet de donnée (non bruitée).

Méthodes algébriques

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

$$\mathbf{y} \in \mathbb{R}^M \quad \mathbf{x} \in \mathbb{R}^N \quad \mathbf{A} \in \mathbb{R}^{M \times N} \quad \mathbf{b} \in \mathbb{R}^M$$

- Rétrojection ou filtre adapté : $\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{y}$
- Inversion directe lorsque possible : $\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{y}$
- Inverse généralisée :
  - $M > N$ et $\text{rang}(A) = N \longrightarrow \hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$
  - $M < N$ et $\text{rang}(A) = M \longrightarrow \hat{\mathbf{x}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$
  - Décomposition en valeurs singulières (SVD) : $\hat{\mathbf{x}} = \sum_{k=1}^{r} \frac{\langle \mathbf{u}_k, \mathbf{y} \rangle}{\lambda_k} \mathbf{v}_k$

Pseudo Inverse - cas sur-déterminé

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

$$\mathbf{y} \in \mathbb{R}^M \quad \mathbf{x} \in \mathbb{R}^N \quad \mathbf{A} \in \mathbb{R}^{M \times N} \quad \mathbf{b} \in \mathbb{R}^M$$

Un problème est dit **sur-déterminé** si on a plus d'obervations que d'inconnus, i.e.

$$M > N$$

Problème : il n'existe en général pas de solution vérifiant exactement $\mathbf{y} = \mathbf{A}\mathbf{x}$

Chercher x revient en fait à un problème de *régression linéaire* : on cherche à passer « *au mieux* »des données.

Cette solution dites des **moindres carrés** minimise l'erreur quadratique entre les obervations et le signal cherché

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$$

Pseudo Inverse - cas sous-déterminé

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

$$\mathbf{y} \in \mathbb{R}^M \quad \mathbf{x} \in \mathbb{R}^N \quad \mathbf{A} \in \mathbb{R}^{M \times N} \quad \mathbf{b} \in \mathbb{R}^M$$

Un problème est dit **sous-déterminé** si on a plus d'obervations que d'inconnus, i.e.

$$M < N$$

Problème : Il existe une infinité de solution au problème $\mathbf{y} = \mathbf{A}\mathbf{x}$

On peut définir une inverse généralisée, qui correspond en fait à la solution d'énergie minimale :

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y} \\ &= \operatorname*{argmin}_{x \in \mathbb{R}^N} \|\mathbf{x}\|_2 \\ &\text{s.c. } \mathbf{y} = \mathbf{A}\mathbf{x} \end{aligned}$$

Pseudo Inverse - SVD

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

$$\mathbf{y} \in \mathbb{R}^M \quad \mathbf{x} \in \mathbb{R}^N \quad \mathbf{A} \in \mathbb{R}^{M\times N} \quad \mathbf{b} \in \mathbb{R}^M$$

Dans tous les cas, calculer la persudo inverse peut se faire par la SVD :

$$\hat{\mathbf{x}} = \sum_{k=1}^{r} \frac{\langle \mathbf{u}_k, \mathbf{y}\rangle}{\lambda_k}\mathbf{v}_k$$

où

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^*$$

et **S** matrice diagonale comportant les $\lambda_1, \ldots, \lambda_{min(M,N)}$

Conséquence : Une valeur singulière proche de 0 risque de faire « exploser »l'inverse généralisée (notion de conditionnement)

Inversion régularisée

Dans le cas sous-déterminée, l'inverse généralisée peut s'écrire comme

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{x \in \mathbb{R}^N} \|\mathbf{x}\|_2 \quad \text{s.c. } \mathbf{y} = \mathbf{A}\mathbf{x}$$

Si l'on veut prendre le bruit en compte, on généralise comme :

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{x \in \mathbb{R}^N} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{x}\|_2^2$$

où $\lambda > 0$ est un “hyper-paramètre” qui permet de balancer entre

- l'attache aux données $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$
- la régularisation $\|\mathbf{x}\|_2^2$

ce problème est équivalent à

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{x \in \mathbb{R}^N} \|\mathbf{x}\|_2^2 \quad \text{s.c. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \sigma^2$$

qui s'interprète comme la solution *d'énergie minimale*, pour un bruit blanc gaussien de variance σ^2 .

Deconvolution par energie minimale

Cas continu :

$$g(t) = h(t) \star f(t) + b(t)$$

$$J(f) = \|g - h \star f\|^2 \longrightarrow \frac{\partial J}{\partial f} = -2h(-t) \star [g(t) - h(t) \star f(t)] = 0$$

$$[h(-t) \star h(t)] \star f(t) = h(-t) \star g(t)$$

Passant dans le domaine de Fourier on obtient :

$$\left[|H(\omega)|^2\right] F(\omega) = H^*(\omega)G(\omega) \longrightarrow F(\omega) = \frac{H^*(\omega)G(\omega)}{|H(\omega)|^2} = \frac{1}{H(\omega)} G(\omega)$$

Remarque :

$$J(f) = \|g - h \star f\|^2 + \lambda \|d \star f\|^2 \longrightarrow$$

$$F(\omega) = \frac{H^*(\omega)G(\omega)}{|H(\omega)|^2 + \lambda |D(\omega)|^2} = \frac{1}{H(\omega)} \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda |D(\omega)|^2} G(\omega)$$

Solution "optimale" d'un critère

Définition de la solution comme le minimiseur d'une fonction coût :

$$\mathbf{x} = \operatorname*{argmin}_{\mathbf{x}} J(\mathbf{x})$$

- Moindres carrés (MC) : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2$
- MC de norme minimale (MCNM) : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \lambda\|\mathbf{x}\|^2$
- Régularisation classique : $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}(\mathbf{x})\|^2 + \lambda\|\mathbf{D}\mathbf{x}\|$
- Régularisation plus générale :

$$J(\mathbf{x}) = \mathcal{Q}(\mathbf{y} - \mathbf{A}(\mathbf{x})) + \lambda\Phi(\mathbf{D}\mathbf{x})$$

ou encore

$$J(\mathbf{x}) = \Delta_1(\mathbf{y}, A(\mathbf{x})) + \lambda\Delta_2(\mathbf{x}, \mathbf{x}_\infty)$$

Solution "optimale" d'un critère

$$\mathbf{x} = \operatorname*{argmin}_{\mathbf{x}} J(\mathbf{x})$$

Avantages

- Nombreux modèles d'attaches aux données
- Nombreux critères de régularisation (e.g. les normes)
- Algorithmes performants pour l'optimisation convexe

Limitations

- Bruit implicitement considéré blanc et gaussien
- Prise en compte d'information *a priori* sur la solution limitée
- Manque d'outil pour la détermination des hyperparamètres

Inversion probabiliste

On se place dans le cadre de l'estimation :

- Maximum de vraisemblance (MV)
- Maximum d'entropie (ME)
- Maximum d'entropie sur la moyenne (MEM)
- Estimation bayésienne (EB)

Avantages

- Prise en compte de la nature du bruit
- Prise en compte d'information *a priori* sur la solution
- Cadre cohérent permettant aussi la détermination des hyperparamètres

Limitations

Mise en œuvre plus difficile en pratique

Conclusion

- Lorsqu'on a des mesures et un modèle les liant aux inconnues, on se trouve devant un problème inverse.
- Un grand nombre de problèmes de restauration et de reconstruction des signaux et des images se formalisent comme des problèmes inverses linéaires.
- Les problèmes inverses sont souvent des problèmes mal-posés.
- Les méthodes "allant de soi" ne donnent que rarement satisfaction.