

# Traitement du signal pour l'amélioration de la qualité de mesure en spectroscopie

Laboratoire PRISME  
Laboratoire GREMI

Date

## Contenu

1. Problème
2. Modèle
3. Manières de traiter le problème
  - (a) Modèle d'Identification
  - (b) Modèle d'Inversion
4. Algorithme de Richardson-Lucy
5. Méthode L-curve
6. Critère de performances
7. Code

## 0.1 Problème

En spectroscopie, les signaux mesurés sont souvent affectés par divers types de bruit et de distorsion, rendant difficile l'interprétation précise des caractéristiques spectrales.

Le problème de récupération du signal réel en spectroscopie peut être perçu de différentes manières, mais il est généralement modélisé comme un problème de déconvolution plutôt que comme un problème d'identification ou d'inversion pour plusieurs raisons spécifiques à la nature des données et des processus en jeu.

Le signal mesuré en spectroscopie est intrinsèquement le résultat d'une convolution entre le signal réel et la réponse instrumentale (fonction de transfert ou PSF). Cette convolution représente la manière dont l'instrument distord le signal.

## 0.2 Modèle

Dans le contexte des mesures spectroscopiques, le modèle de récupération de signal le plus utilisé est basé sur la déconvolution, où le signal mesuré est modélisé comme la convolution du signal original avec la fonction de réponse de l'instrument (PSF, Point Spread Function) :

$$m = s * b + n$$

Cette approche permet de conceptualiser le processus de distorsion du signal par l'instrument et d'élaborer des stratégies pour inverser cet effet. Les modèles mathématiques tels que la déconvolution de Wiener et la régularisation de Tikhonov sont fréquemment employés pour stabiliser le processus de récupération de signal en présence de bruit. Ces méthodes utilisent respectivement un filtrage adaptatif et l'ajout d'un terme de régularisation pour contrôler l'amplification du bruit durant la déconvolution.

De plus, les méthodes basées sur l'apprentissage profond, comme les réseaux de neurones convolutifs, sont de plus en plus explorées pour leur capacité à extraire des signaux utiles de fonds bruités complexes sans une connaissance explicite de la PSF.

Dans certains modèles de traitement du signal en spectroscopie, le terme de bruit peut être omis du modèle mathématique pour plusieurs raisons. L'omission du terme de bruit dans les modèles de convolution en spectroscopie peut être motivée par plusieurs facteurs clés. Premièrement, dans de nombreux cas, les modèles peuvent être simplifiés pour se concentrer uniquement sur la réponse de l'instrument (la PSF), surtout lorsque la fonction de transfert est bien comprise et que le bruit est relativement faible par rapport au signal mesuré. Cela permet une modélisation plus directe et moins complexe, ce qui est particulièrement avantageux dans les applications où la rapidité de calcul est essentielle.

## 0.3 Manières de traiter le problème

Le problème de récupération du signal en spectroscopie peut être abordé à l'aide de différents modèles, dont les modèles d'identification et d'inversion. Ces approches se concentrent sur la manière dont le signal réel est transformé en signal mesuré et sur la résolution de ce processus.

### 0.3.1 Modèle d'Identification

En identification, nous cherchons à déterminer les paramètres d'un modèle qui décrivent notre système. Pour le cas de la spectroscopie, nous devons modéliser le processus par lequel le signal réel est transformé en signal mesuré. Voici comment cela pourrait se formuler :

**Modèle :** Supposons que le signal réel  $s$  est décrit par un ensemble de paramètres  $\theta$ . La relation entre  $s$  et  $m$  est donnée par une fonction  $f$  qui inclut la convolution avec la réponse instrumentale et l'ajout de bruit :

$$m = f(s, \theta) + n$$

**Objectif :** Identifier les paramètres  $\theta$  qui définissent  $s$  de manière à minimiser la différence entre  $m$  et  $f(s, \theta)$ . Cela pourrait impliquer des techniques d'estimation comme les moindres carrés, l'estimation de maximum de vraisemblance, ou des méthodes bayésiennes.

**Méthode :** Définir une fonction de coût ou de vraisemblance  $L$  :

$$L(\theta) = \|m - f(s, \theta)\|^2$$

Optimiser cette fonction pour trouver les meilleurs paramètres  $\theta$ .

### 0.3.2 Modèle d'Inversion

L'inversion concerne généralement la résolution d'une équation directe pour obtenir la variable d'intérêt. Dans le contexte de la spectroscopie :

**Modèle :** Nous savons que le signal mesuré  $m$  est le résultat d'une convolution du signal réel  $s$  avec la réponse instrumentale PSF  $b$  :

$$m = s * b + n$$

**Objectif :** Trouver  $s$  directement en inversant l'opération de convolution.

**Méthode :** Si nous ignorons le bruit  $n$  pour simplifier, la formule d'inversion peut être écrite (en théorie) comme :

$$s = \mathcal{F}^{-1} \left( \frac{\mathcal{F}(m)}{\mathcal{F}(b)} \right)$$

où  $\mathcal{F}$  représente la transformée de Fourier et  $\mathcal{F}^{-1}$  sa transformée inverse.

## Limitations des Modèles d'Identification et d'Inversion

- **Identification :**

- **Complexité du Modèle :** Identifier les bons paramètres peut être très complexe, surtout si le modèle de signal réel  $s$  est non trivial.
- **Sensibilité au Bruit :** Les méthodes d'identification peuvent être sensibles au bruit, rendant difficile la détermination précise des paramètres.

- **Inversion :**

- **Instabilité Numérique :** L'inversion directe de la convolution, surtout en présence de bruit, peut être numériquement instable. Les petites perturbations dans le signal mesuré peuvent conduire à de grandes erreurs dans le signal réel reconstruit.
- **Amplification du Bruit :** Le processus d'inversion a tendance à amplifier le bruit, ce qui rend les solutions obtenues moins fiables sans régularisation adéquate.

Modéliser ce processus comme une déconvolution est la manière la plus naturelle et la plus précise de refléter le processus physique.

## 0.4 Algorithme de Richardson-Lucy

L'algorithme de Richardson-Lucy est un choix privilégié pour la déconvolution en spectroscopie en raison de plusieurs avantages :

**Méthode :** Nous considérons le problème de la restauration d'un signal réel  $s$  à partir d'un signal mesuré  $m$ , dégradé par une fonction de réponse impulsionnelle (PSF)  $b$ . L'objectif est de retrouver  $s$  en connaissant  $m$  et  $b$ .

### Application du Théorème de Bayes

Pour un état donné  $s_i$  du signal réel, l'application du théorème de Bayes nous donne la probabilité conditionnelle de  $s_i$  étant donné un état observé  $m_k$  du signal mesuré :

$$P(s_i|m_k) = \frac{P(m_k|s_i)P(s_i)}{\sum_j P(m_k|s_j)P(s_j)} \quad (1)$$

### Probabilité Marginale de $m$

La probabilité de voir un état spécifique  $m_k$  à travers tous les états possibles de  $s$  est donnée par :

$$P(s_i) = \sum_k P(s_i|m_k)P(m_k) \quad (2)$$

### Solution Itérative (Algorithme de Richardson-Lucy)

Sans connaissance de la probabilité a priori  $P(s_i)$ , nous utilisons une méthode itérative :

$$P_{r+1}(s_i) = P_r(s_i) \sum_k \frac{P(m_k|s_i)P(m_k)}{\sum_j P(m_k|s_j)P_r(s_j)} \quad (3)$$

où  $P_r(s_i)$  est l'estimation de  $P(s_i)$  à l'itération  $r$ .

### Reformulation Pratique avec la PSF

En utilisant la convolution pour lier  $s$  et  $m$  à travers  $b$ , on simplifie l'itération comme suit :

$$s_{i,r+1} = s_{i,r} \sum_k \frac{b_{i,k}m_k}{\sum_j b_{j,k}s_{j,r}} \quad (4)$$

### Convergence

Le processus itératif se poursuit jusqu'à ce que les modifications entre les estimations successives de  $s$  soient minimales, indiquant une convergence vers l'estimation du signal original  $s$ .

**Avantages :**

- **Robustesse :** Il est particulièrement efficace pour traiter les données avec un faible rapport signal/bruit, ce qui est courant en spectroscopie. Dans [?], la robustesse et les bonnes performances de ce critère sont démontrées par des simulations approfondies réalisées pour différents scénarios de correction de bande passante.
- **Préservation des non-négativités :** Cet algorithme garantit que les valeurs du signal déconvolué restent non négatives, ce qui est physiquement réaliste pour les spectres.
- **Convergence vers une solution :** L'algorithme itératif de Richardson-Lucy converge souvent vers une solution qui maximise la vraisemblance du signal réel, en particulier pour les distributions de Poisson, couramment rencontrées dans les mesures spectroscopiques.

**Inconvénient :** Un inconvénient de la méthode Richardson-Lucy est qu'elle est itérative et que ses résultats peuvent dépendre du nombre d'itérations employées. La qualité des résultats peut varier en fonction du nombre d'itérations et peut nécessiter une évaluation minutieuse pour éviter des effets de sur-apprentissage ou des erreurs de reconstruction.

## 0.5 Méthode L-curve

La méthode L-curve est un outil graphique utilisé pour trouver un paramètre de régularisation optimal pour résoudre des problèmes discrets mal posés, tels que ceux rencontrés dans les problèmes inverses ou dans les situations où la solution d'un système d'équations linéaires est sensible aux erreurs dans les données. La régularisation permet de stabiliser la solution en introduisant des informations ou des contraintes supplémentaires. La méthode L-curve consiste à tracer la norme de la solution régularisée en fonction de la norme des résidus pour différentes valeurs du paramètre de régularisation et à sélectionner le paramètre correspondant au « coin » du L-curve.

### Étapes pour utiliser la méthode de la courbe L pour l'algorithme de Richardson-Lucy

**1. Exécution de l'algorithme de Richardson-Lucy :**

Effectuez la déconvolution de Richardson-Lucy pour une série d'itérations  $k$ , disons de 1 à  $k_{\max}$ . L'algorithme met à jour l'estimation de l'image itérativement comme suit :

$$s_{k+1} = s_k \cdot \left( \frac{m}{b * s_k} * b \right)$$

où  $s_k$  est l'image estimée à l'itération  $k$ ,  $m$  est l'image floue observée,  $b$  est la fonction de répartition du point (PSF), et  $*$  indique une convolution.

**2. Calcul des normes résiduelles et des normes de solution :**

Pour chaque itération  $k$ , calculez :

- La norme résiduelle :  $\|b * s_k - m\|$
- La norme de solution :  $\|s_k\|$

**3. Tracez la courbe L :**

Tracez  $\log(\|b * s_k - m\|)$  contre  $\log(\|s_k\|)$  pour les itérations  $k = 1, 2, \dots, k_{\max}$ .

**4. Identifiez le coin :**

Le tracé forme généralement une courbe en forme de L. Le coin de cette courbe représente un équilibre entre l'ajustement des données observées (petite norme résiduelle) et le maintien de la stabilité de la solution (petite norme de solution). Le nombre optimal d'itérations est choisi à ce point de coin.



## 0.6 Critère de performances

Pour évaluer quantitativement les performances des spectres déconvolués, on utilise principalement ces métriques :

1. **Erreur Quadratique Moyenne Normalisée (NMSR)** :  $\text{NMSR} = \frac{\|s-m\|^2}{\|f\|^2}$ , où  $s$  est le spectre de référence et  $m$  le spectre déconvolué. Cette mesure nécessite un spectre de référence et est donc principalement utilisée dans des expériences simulées.

Une valeur proche de zéro indique que le spectre déconvolué est très proche du spectre de référence, signifiant une excellente restauration.

2. **Ratio de Suppression du Bruit (NSR)** :  $\text{NSR} = \frac{\sum |m|}{\sum |s|}$ , avec  $m$  comme le spectre mesuré et  $s$  le spectre latent. Le NSR mesure l'efficacité de l'algorithme à supprimer le bruit dans le spectre.

Une valeur élevée signifie une suppression efficace du bruit, reflétant une amélioration significative de la clarté du spectre.

3. **Ratio de Largeur à Mi-Hauteur (FWHMR)** :  $\text{FWHMR} = \frac{\text{FWHM}_m}{\text{FWHM}_s}$ , où FWHM représente la largeur à mi-hauteur des bandes dans les spectres mesuré et latent, respectivement.

Une valeur supérieure à 1 indique une amélioration de la résolution du spectre déconvolué par rapport au spectre mesuré.

Des valeurs élevées de NSR et de FWHMR indiquent une qualité supérieure du spectre résultant, avec une amélioration de la résolution et une réduction efficace du bruit. Ces métriques offrent un cadre robuste pour comparer la fidélité des spectres déconvolués par rapport aux spectres originaux ou simulés.

**Erreur quadratique moyenne (rms) de la reconstruction :** Cette mesure évalue la précision de la reconstruction du spectre sous-jacent. L'erreur rms est calculée en fonction du nombre d'essais de Monte Carlo ( $L$ ) et du nombre de valeurs du spectre mesuré ( $K$ ). Elle est définie par la formule :

$$\text{rms} = \sqrt{\frac{1}{L} \sum_{\ell=1}^L \frac{1}{K-4} \sum_{k=3}^{K-2} (s^{(\ell)}(\lambda_k) - m(\lambda_k))^2}$$

où  $s_i$  représente la valeur réelle du spectre à l'index  $i$  et  $\hat{s}_i$  la valeur reconstruite.

**Incertitude moyenne :** En plus de l'erreur rms, l'incertitude moyenne est calculée. En théorie, le rapport entre l'erreur rms et les incertitudes moyennes devrait être autour de 1. Ce rapport est donné par :

$$u_{\text{mean}} = \sqrt{\frac{1}{K-4} \sum_{k=3}^{K-2} (V_s)_{kk}}$$

$$V_s = \frac{1}{L-1} \sum_{\ell=1}^L (S^{(\ell)} - \hat{S})(S^{(\ell)} - \hat{S})^T$$

$$S = \frac{1}{L} \sum_{\ell=1}^L S^{(\ell)}$$

**Interprétation des valeurs :**

- **Valeurs  $< 1$  :** Cela indique que les incertitudes sont plus grandes que les erreurs quadratiques moyennes de la reconstruction. Cela peut suggérer que les incertitudes sont surestimées ou que les erreurs méthodologiques systématiques ont été négligées dans l'analyse des incertitudes.
- **Valeurs  $> 1$  :** Cela signifie que les incertitudes sont sous-estimées par rapport aux erreurs rms de la reconstruction. Cela peut indiquer que des erreurs systématiques ont été négligées dans l'évaluation des incertitudes.