

COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 26 au 30 AVRIL 1977

SEPARATION DE DEUX PICS VISUELLEMENT CONFONDUS

SEPARATION OF TWO VISUALLY CONFOUNDED PEAKS

Hervé RIX et Jean-Pierre MALENGE

Département I.M.A.N. et LA 190
I.M.S.P. Université de Nice, Parc Valrose - F06034 NICE CEDEX

RESUME

Les expériences physico-chimiques fournissent fréquemment des signaux ayant la forme de pics :

- forme gaussienne, éventuellement convoluée par un filtre exponentiel, en chromatographie ou en chimie-physique nucléaire;
- forme lorentzienne en diffusion de la lumière.

Un pic est défini mathématiquement comme une fonction $f(x)$ positive, intégrable, possédant un maximum et seulement deux zéros pour sa dérivée seconde. Un signal $y(x)$ sera dit de la forme $f(x)$ s'il s'écrit :

$$y(x) = k f \left(\frac{x - d}{a} \right)$$

On s'intéresse au cas : $a > 0$, $k \geq 0$.

Il peut arriver qu'un signal ayant une allure de pic résulte de la contribution de deux signaux de la forme $f(x)$, pour certaines valeurs des paramètres k , d , a . C'est dans ces conditions que l'on présente un test permettant de décider si le pic observé est ou non la somme de deux pics.

La méthode proposée utilise, comme information essentielle, l'intégrale du signal : elle est donc peu sensible au bruit. Le test est particulièrement efficace lorsque les deux pics ont des surfaces notablement différentes (rapport de la plus petite à la plus grande de l'ordre de quelques %).

SUMMARY

Physical and chemical experiments usually give signals with a peak shape :

- gaussian shape, sometimes convoluted by an exponential filter, in chromatography or nuclear chemical physics;
- Lorentz shape in light diffusion.

A peak is mathematically defined as a positive and integrable function with one maximum and only two zeros for its second derivative. A signal $y(x)$ will be described as having the shape $f(x)$ if it can be written :

$$y(x) = k f \left(\frac{x - d}{a} \right)$$

We shall be concerned with the case: $a > 0$, $k \geq 0$.

A signal having a peak aspect may happen to result from the contribution of two signals with the f shape, for some values of the parameters k , d , a . In this case we present a test which allows to decide if the observed peak is or is not the sum of two peaks.

The proposed method makes use of the signal integral as essential data : therefore it is very little sensitive to noise. The test is particularly efficient when the two peaks have quite different surfaces (smaller to larger ratio of the order of a few percents).



1 - INTRODUCTION

L'information délivrée à la sortie d'un appareil de mesure est souvent représentée par une suite de signaux ayant la forme de pics. Citons, parmi les exemples les plus courants, tous les spectres (d'émission, d'absorption ou de diffusion) et les chromatogrammes.

Si à un phénomène "pur" (raie pure dans le cas des spectres, pic d'un corps pur dans le cas de la chromatographie) est associé un pic, la réciproque, vraie la plupart du temps, peut être mise en défaut : en dessous d'un certain degré de résolution, un signal ayant une forme de pic peut résulter de la composition de deux ou plusieurs phénomènes.

Le but de notre travail est justement de présenter une méthode permettant de trancher la question dans le cas particulier de l'alternative "pic simple" (représentant un phénomène unique) ou "pic double" (superposition de deux pics simples) dans le cas où, visuellement, le profil est celui d'un pic, c'est-à-dire représentable mathématiquement par une fonction positive ayant un zéro pour sa dérivée première et seulement deux zéros pour sa dérivée seconde.

La méthode proposée suppose que l'on connaisse un modèle pour la forme d'un pic simple et que tous les signaux ayant la forme $f(x)$ soient ceux dont le profil peut s'écrire $kf\left(\frac{x-d}{a}\right)$. On se ramène à un problème de détection à seuil après avoir défini une distance entre un signal quelconque et la classe des signaux de la forme f . Dans le calcul numérique n'interviennent que les valeurs échantillonnées du signal et de son intégrale; aucune dérivation n'est nécessaire. La méthode est donc, indépendamment de tout filtrage, moins sensible au bruit que celles utilisant le calcul de la dérivée seconde (analyse de pente [1],[2]) ou le calcul des moments (méthode des moments [2],[3]). On ne peut cependant pas comparer globalement ces méthodes et la nôtre : celles-là visent à estimer les paramètres du deuxième pic éventuel, alors que celle-ci ne cherche qu'à détecter sa présence. Concrètement, dans le cas de la chromatographie, les méthodes d'estimation cherchent à déterminer la nature et la concentration des corps associés au signal observé alors que la nôtre est destinée à détecter la présence de traces dans un produit.

Nous décrivons ci-dessous le principe de la méthode dans le cas général d'une forme de pic quelconque $f(x)$ convoluée par un filtre exponentiel connu car l'observation d'un signal se fait couramment après son passage dans un filtre RC.

Comme application nous avons simulé numériquement la somme de deux pics de forme $f(x)$ gaussienne convoluée par un filtre dont la constante de temps est égale à l'écart-type de la gaussienne. Pour une valeur donnée du rapport S/B nous étudions la zone de détectabilité en fonction de la séparation des pics et du rapport des surfaces (on suppose des pics de même largeur). Le principal domaine d'application visé en utilisant ce modèle est la chromatographie en phase gazeuse. L'exposé de la méthode dans le cas du modèle gaussien convolué par une exponentielle, ainsi que des résultats numériques montrant son efficacité dans le cas d'un petit pic situé dans la trainée d'un pic principal, ont déjà été publiés dans une revue spécialisée de chromatographie [4].

Outre une présentation plus générale de la méthode, l'originalité du présent travail réside dans l'utilisation d'une nouvelle distance entre un signal et la classe des signaux de référence s'avérant plus efficace pour séparer les pics. De plus on s'intéresse au cas où le plus petit des deux pics est en tête aussi bien qu'au cas où il est en queue du pic principal.

2 - EXPOSE DE LA METHODE

Que la variable soit une fréquence ou bien le temps, les pics observés peuvent s'interpréter - aussi bien du point de vue de la physique que d'un point de vue purement mathématique - comme des densités de probabilité (après avoir été normés). Leurs intégrales sont donc des fonctions de répartition.

La classe des signaux de forme f est l'ensemble des signaux dont le profil peut s'écrire, en fonction de la variable x (temps ou fréquence) : $kf\left(\frac{x-d}{a}\right)$

où k et a sont des nombres positifs et d un nombre quelconque. La fonction f étant normée s'interprète comme la densité d'une variable aléatoire X ; la classe des signaux de forme f s'identifie à l'ensemble des lois du "type" de f (cf. FORTET [5]). Soit F la fonction de répartition de X associée à f ; étant donnée la forme de pic de f , F est une fonction strictement croissante et continue, donc possède une fonction réciproque F^{-1} . Soit G la fonction de répartition d'un signal de la forme f , c'est à dire associée

SEPARATION DE DEUX PICS VISUELLEMENT CONFONDUS

à une v.a. $Y = aX + d$ ($a > 0$); on a :

$$G(y) = F \left(\frac{y - d}{a} \right)$$

ou, puisque F est inversible :

$$F^{-1}(G(y)) = \frac{y - d}{a}$$

La représentation de la quantité $y^* = F^{-1}(G(y))$ en fonction de y est donc une droite lorsque G est la fonction de répartition d'un signal de la forme f . Le principe de la méthode est d'apprécier l'écart d'un signal à une forme f en mesurant l'écart à la linéarité de la fonction y^* . Cela n'est autre qu'une généralisation du test de Henry couramment utilisé pour juger graphiquement du caractère gaussien d'une variable aléatoire. La différence essentielle est qu'au lieu de nous contenter d'une appréciation visuelle nous proposons une mesure quantitative de l'écart, fondée sur l'observation que la déformation créée par un deuxième pic n'est pas la même que celle créée par du bruit.

Convolution par un filtre exponentiel

Un signal de forme f , après avoir été filtré par un filtre exponentiel de constante de temps τ , a la forme

$$g(x) = f * EU(x) \tag{1}$$

où :

- U est la fonction d'Heaviside
- $E(x) = \mu e^{-\mu x}$ ($\mu = 1/\tau$)

La fonction y^* définie plus haut à partir de la forme f , peut alors se calculer de la même manière à partir de la forme g . On peut aussi, après avoir déconvolué le signal de sa partie exponentielle, prendre la forme f comme référence. Cette seconde stratégie permet de conserver la même fonction de répartition lorsque les variations des conditions expérimentales ne se manifestent que sur le filtre exponentiel.

Si l'on dérive la relation (1) au sens des distributions on obtient :

$$g'(x) = -\mu g(x) + \mu f(x) \tag{2}$$

d'où

$$f(x) = g(x) + \frac{1}{\mu} g'(x) \tag{3}$$

La relation (3) montre que la forme f s'exprime en fonction de g et de sa dérivée; donc l'intégrale de f fera intervenir l'intégrale de g et g elle-même :

$$F(x) = E(x) + \frac{1}{\mu} g(x) \tag{4}$$

Soit alors $V(x)$ un signal observé ayant une allure de pic; soit S sa surface. On calcule la fonction :

$$W(x) = F^{-1} \left(\frac{1}{S} \int_{-\infty}^x V(u) du + \frac{1}{\mu} \frac{V(x)}{S} \right) \tag{5}$$

- Si V est de la forme g , c'est à dire si $V(x) = g \left(\frac{x - d}{a} \right)$ ou, pour simplifier, puisque l'on s'intéressera à des pics de même largeur $V(x) = g(x - d)$ ($a = 1$), on a :

$$W(x) = x - d$$

-Si V n'est pas de la forme g , $W(x)$ n'est pas une droite.



SEPARATION DE DEUX PICS VISUELLEMENT CONFONDUS

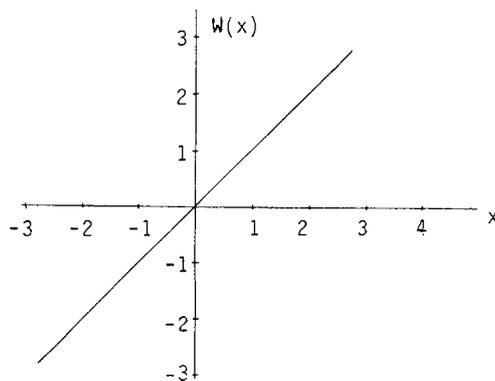
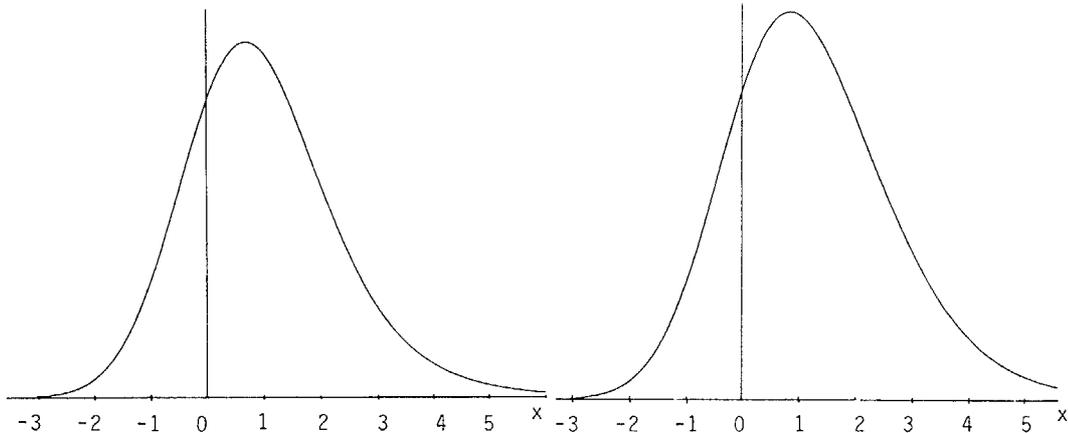


Fig. 1a
Pic simple

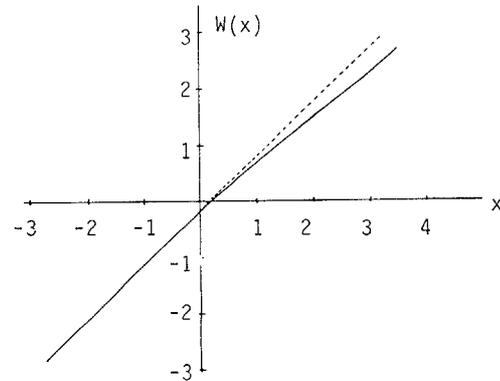


Fig. 1b
Pic double; la droite en pointillé est ajustée sur la première partie de la courbe $W(x)$.

Les figures (1a) et (1b) illustrent le comportement de W pour un "pic simple" et un "pic double". La forme f est une gaussienne dont l'écart-type est pris comme unité d'abscisse; la forme g est la convolution de f par un filtre exponentiel dont la constante de temps est égale à cet écart type; le "pic double" de la figure 1b a été obtenu en ajoutant au "pic simple" de la figure 1a un pic de même largeur, de surface égale à 20 % de celle du premier et à une distance $d = 1,5$.

C'est ici que se pose le problème du choix d'une distance Δ permettant de mesurer l'écart à la linéarité.

Définition de la distance Δ

Soit V un signal quelconque ayant une allure de pic et $\mathcal{L}(f)$ la classe des signaux de forme f . Après avoir appliqué au signal V la transformation définie en (5) on est ramené à définir la distance de W à la famille \mathcal{D} des droites (affines) du plan. Désignons par x_i (i varie de 1 à n) les abscisses des points d'échantillonnage. L'utilisation de la distance quadratique moyenne classique de W à \mathcal{D} est mal adaptée à notre problème : elle n'utilise pas le fait que la déformation de W , due à un deuxième pic, n'est sensible que sur une partie de W . Lorsque le plus petit pic est en queue la déformation se situe essentiellement dans la deuxième moitié de W (cf. fig. 1b). C'est pourquoi, pour les résultats que nous avons déjà publiés [4], Δ représentait l'écart quadratique moyen des $W(x_i)$, pour $i = p + 1$ à n , par rapport à la droite des moindres carrés ajustée sur les p premiers points (correspondant à la montée du pic).

La définition de Δ que nous proposons ici permet de tirer parti au maximum de la remarque précédente (au prix d'un temps de calcul un peu plus élevé) et cela quelle que soit la position relative des deux composantes du signal observé.

SEPARATION DE DEUX PICS VISUELLEMENT CONFONDUS

Soit a et b deux valeurs particulières des abscisses x_i . Etant donné un signal V , appelons Δ_a^b l'écart quadratique moyen des $W(x_i)$, pour x_i n'appartenant pas à $[a,b]$, par rapport à la droite des moindres carrés ajustée sur les points dont l'abscisse est dans $[a,b]$. La distance Δ de V à $\mathcal{L}(f)$ est définie par :

$$\Delta(V, \mathcal{L}(f)) = \max_{a \in \{x_1, \dots, x_n\}} \Delta_a^{a+\delta}$$

où δ est une longueur fixe représentant approximativement la largeur d'un pic (deux fois l'écart-type pour une gaussienne).

Avant d'examiner les résultats obtenus avec la nouvelle définition de Δ , nous allons expliciter la stratégie qui nous permettra de dire, en présence de bruit, s'il y a un ou deux pics - cette stratégie étant indépendante de la définition de Δ .

Stratégie de détection

a) Détermination du seuil Δ_0 : On calcule Δ pour un signal composé d'un pic simple et d'un bruit gaussien additif blanc, de rapport S/B donné ($S/B = \text{Hauteur maximale du signal}/4 \times \text{écart-type du bruit}$) et pour une série de réalisations (20 en général) de ce bruit. De cette statistique on associe à cette valeur de S/B un seuil Δ_0 (par exemple l'écart maximal observé) au-dessus duquel on décidera la présence d'un deuxième pic.

b) Détection de la présence d'un second pic : étant donné un signal réel dont on suppose connu le rapport S/B on calcule la valeur de Δ pour ce signal et on décide qu'il y a un deuxième pic si $\Delta > \Delta_0$.

c) Zones de détectabilité : afin de déterminer les limites dans lesquelles notre méthode risque d'être valable, nous avons simulé un signal composé de deux pics de forme g (forme gaussienne f d'écart-type unité convoluée par une exponentielle de constante de temps unité) :

$$V(x) = g(x) + k g(x - d)$$

pour plusieurs valeurs des paramètres k et d .

3 - RESULTATS

Les résultats sont représentés ci-dessous :

a) dans un tableau résumant la statistique sur Δ et permettant d'associer un seuil Δ_0 aux valeurs du rapport S/B étudiées;

S/B	Δ_{moyen}	$\Delta_0 = \Delta_{\text{max}}$
30	$6 \cdot 10^{-2}$	$1,5 \cdot 10^{-1}$
100	$1,6 \cdot 10^{-2}$	$4 \cdot 10^{-2}$
250	$6,8 \cdot 10^{-3}$	$1 \cdot 10^{-2}$
10^4	$2,2 \cdot 10^{-3}$	$2,2 \cdot 10^{-3}$

b) dans le plan (d,k) sous forme de courbes "iso Δ " (fig. 2). Les nombres près des courbes sont les valeurs du rapport S/B correspondant à $2 \times \Delta_0$. En effet si l'on détecte avec une probabilité d'erreur jugée suffisamment faible pour $\Delta > \Delta_0$, on doit, pour déterminer la région du plan (k,d) dans laquelle on est (presque) sûr de détecter un deuxième pic tracer les "iso Δ " avec $\Delta = 2\Delta_0$. Les courbes en trait plein sont relatives à la nouvelle définition de Δ , les courbes en pointillé sont relatives à l'ancienne définition qui ajustait la droite sur la première partie de $W(x)$ et calculait l'écart quadratique moyen de la deuxième partie de $W(x)$ par rapport à cette droite.

On constate que les statistiques sur Δ avec un seul pic bruité, donc les valeurs des seuils Δ_0 , sont les mêmes pour les deux définitions de Δ alors que l'écart créé par un deuxième pic est supérieur ou égal (souvent supérieur) avec la nouvelle définition de Δ . Ceci tient au fait que l'ancien Δ permettait d'apprécier la déformation maximale créée par le bruit seul. En effet, considérons la transformation définie



SEPARATION DE DEUX PICS VISUELLEMENT CONFONDUS

en (5) où le signal $V(x)$ contient un bruit blanc gaussien additif. L'intégration de $V(x)$ introduit un processus de Wiener-Lévy (non stationnaire au second ordre et dont la variance est proportionnelle au nombre de pas d'intégration). On peut montrer, dans le cas où F est la fonction intégrale de Gauss que les quantités $W(x)$ sont entachées d'un bruit non stationnaire au premier ordre créant une déformation par rapport à une droite surtout sensible dans la deuxième partie de la courbe $W(x)$.

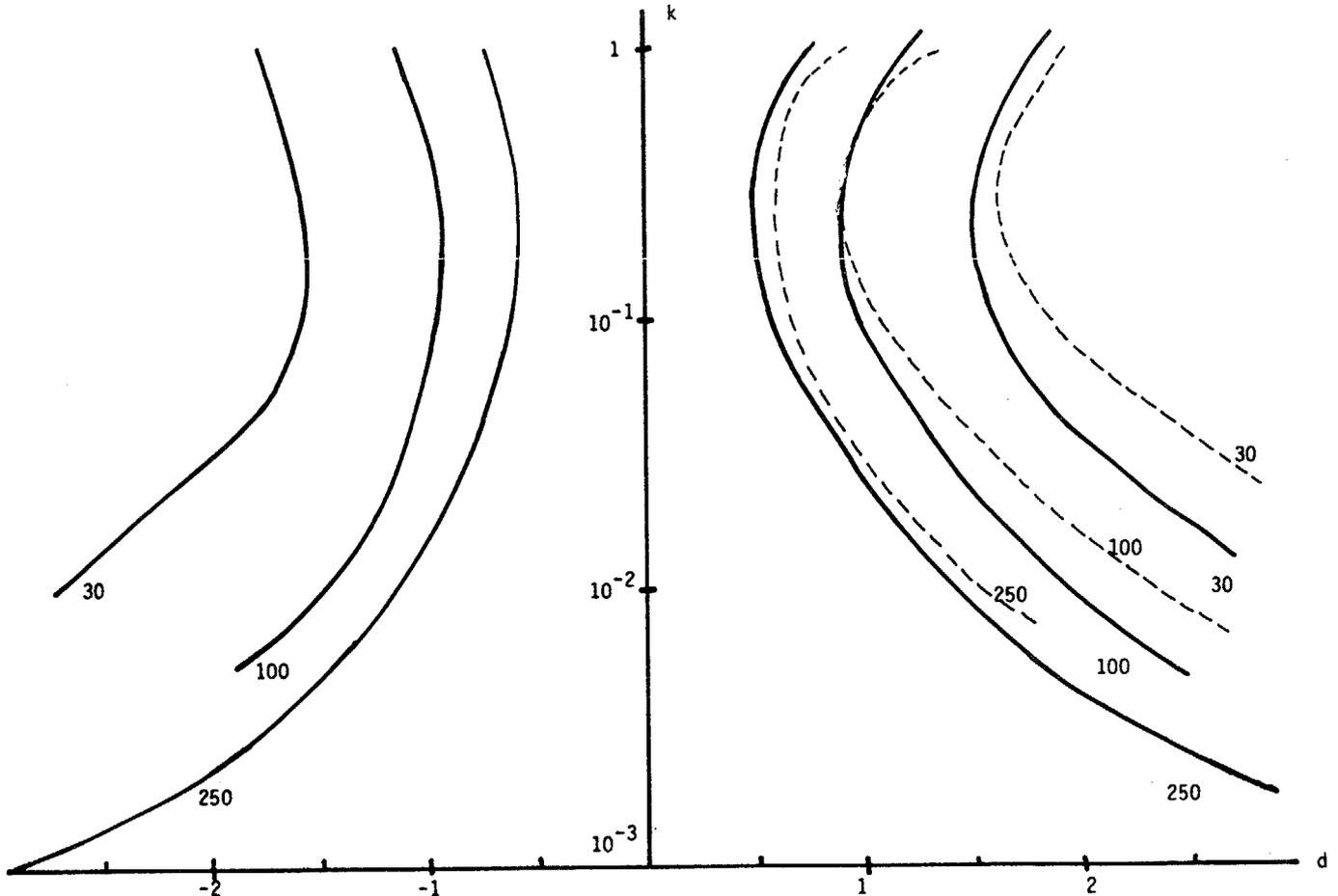


Fig. 2. Courbes "iso Δ "

Les nombres près des courbes représentent le rapport S/B.

— Nouvelle définition de Δ

--- Ancienne définition de Δ

4 - CONCLUSION

Nous avons exposé une méthode permettant de mesurer l'écart d'un signal à une forme donnée en nous ramenant à un problème de régression linéaire et en définissant la distance d'une fonction $W(x)$, transformée du signal observé, à l'ensemble des droites du plan.

La variation de notre distance Δ est beaucoup plus sensible à la présence d'un second pic qu'à la présence de bruit. Il reste à savoir, avant d'améliorer encore éventuellement le traitement informatique, si l'écart entre la forme réelle d'un pic et son modèle peut être rendu suffisamment petit pour que notre simulation soit l'image de la réalité ! C'est à cette question que nous essaierons de répondre en utilisant des pics chromatographiques réels.

REFERENCES

- (1) E. GRUSHKA, *Anal. Chem.* **44**, 1733 (1972)
- (2) E. GRUSHKA, G.C. MONACELLI, *Anal. Chem.* **44**, 484 (1972)
- (3) E. GRUSHKA, M.N. MYERS, I.C. GIDDINGS, *Anal. Chem.* **42**, 21 (1970)
- (4) H. RIX, J.P. MALENGE, *Chromatographia*, **11**, 554 (1976)
- (5) R. FORTET in "Elements du calcul des probabilités" tome 1. Ed. du CNRS (1965)