

Une technique de lissage de données basée sur la théorie de la régularisation

José-Maria Fullana

Service de biophysique, Laboratoires Innothera, 7–9 av. Francois Vincent Raspail, BP 12, 94111 Arcueil cedex, France

Reçu le 10 octobre 2001 ; accepté après révision le 22 juillet 2002

Note présentée par Évariste Sanchez-Palencia.

Résumé Une méthode simple de lissage basée sur la théorie de la régularisation est présentée. On considère le problème de lissage de données comme un problème d'optimisation dont la fonctionnelle à minimiser reflète les caractéristiques du système. On fait appel à la théorie de la régularisation pour définir la courbe optimale qui donne les données filtrées. Des simulations numériques permettent de comparer la méthode à une méthode spectrale classique. *Pour citer cet article : J.-M. Fullana, C. R. Mecanique 330 (2002) 647–652.*
© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

mécanique des fluides / lissage de données / optimisation / théorie de la régularisation

A filter technique based on regularization theory

Abstract A simple filter technique based on the regularization theory is presented. We consider the problem as an optimization one. The regularization theory gives us a suitable theoretical framework to define a functional to minimize. We make a numerical comparison between this method and a classical Fourier technique. *To cite this article: J.-M. Fullana, C. R. Mecanique 330 (2002) 647–652.*
© 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

fluid mechanics / data filter / optimization / regularization theory

Abridged English version

1. Introduction

In medical research on the venous system, experimental data of the vessels (flow and area) are obtained from the echographs. The interpretation of these experimental data contributes to the knowledge of the hemodynamic of the venous network. In spite of technical progress, these data are often very difficult to exploit because of the experimental limits: added to the problem of resolution, the flows are disturbed by the measurement conditions. Moreover, the diameter of the vessels is small and the average velocity low.

Fig. 1 presents the spatial velocity blood profile taken by an echograph at time t following a pinching. The rise phase can be used to calculate the physical parameters of the vessels and, once the profile established, it can be used to calculate the shear at the wall. By observing the experimental data, one realizes the importance to have realistic and reliable data, and thus to have techniques for smoothing the data or reducing the noise. Many techniques for smoothing and filtering exist in the literature (for example, the Kalman filter [2] or AR and ARMA models [3]). In particular, in medical research, spectral filtering is often used.

Adresse e-mail : jose-maria.fullana@innothera.com (J.-M. Fullana).

This Note thus presents a simple method of smoothing data, based on ideas resulting from the theory of the regularization of the curves, and it is compared numerically with a powerful spectral method, the optimal Wiener filter.

2. Regularization method

Given $V_i \pm \delta V_i$ ($i = 1, N$) the velocity profile at time t , and δV_i the experimental error, the objective is to find a smooth curve C , from a series of values \hat{V}_i , which passes through the experimental data $V_i \pm \delta V_i$. It is simple to deduce that an infinite number of smooth curves satisfy this condition, and therefore the problem is ill posed in the sense defined by Hadamard: a problem is well posed when a unique solution exists, and when it depends continuously on the initial data. To limit the space of the possible solutions, one tries to restore the unfulfilled criteria by using the a priori knowledge of the system. This approach defines the regularization theory [6].

One thus seeks a functional H , function of the experimental data V_i and new data \hat{V}_i , satisfying the following a priori information: the resulting curve is rather regular, and we know the noise level of the data V_i . Following [7] we use Eq. (1).

The first term penalizes the difference between two consecutive values of \hat{V} (the local smoothing) and the second penalizes the non-fidelity with respect to the initial data. The free parameter λ controls the relative importance of each term. Since it is a quadratic form of V_i and \hat{V}_i , H has only one minimum, which is a function of the parameter λ . The value of λ is fundamental because the solution depends on it. From a practical point of view the choice is heuristic and fixed for each experiment. From the theoretical point of view we are in the framework of the central problem of the Tikhonov regularization: how to choose λ to produce a solution close to the noise free solution. Much work has been done recently using different techniques like the cross-validation principle, the visualization of residuals, the singular components plots (see the reference [8]). The optimal λ is computed here using a good estimation of the residual error $\|e\|_2$, so we choose λ such that $\hat{\mathbf{V}}$ satisfies $\|\hat{\mathbf{V}} - \mathbf{V}\|_2 = \|e\|_2$ where $\hat{\mathbf{V}} = \{\hat{V}_1, \dots, \hat{V}_N\}$.

The formal problem is the following: given $H(\hat{V}_1 \dots \hat{V}_N)$ defined in the N dimensional space, $\hat{\mathbf{V}} = \{\hat{V}_1, \dots, \hat{V}_N\}$; we seek for the global minimum starting from an initial condition on H . The evolution of the vector $\hat{\mathbf{V}}$ is defined by the rule of the gradient: the speed of $\hat{\mathbf{V}}$ is proportional to its slope and it is calculated using the Gateaux derivative. We note that τ is not connected to physical time; it is an iteration parameter which manages the convergence rate. The equation of evolution of \hat{V}_i is given by Eq. (2).

3. Numerical application

The Wiener optimal filter is a powerful spectral technique which uses the Fast Fourier transform. One can define, starting from the signal spectrum $W(f)$, a filter $\phi(f)$ in the Fourier space such as the spectrum of $W_w(f) = W(f)\phi(f)$ is close to the original one. The signal is then transformed in the real space, V_w , and used for numerical calculations. An introduction to the filter methods is presented in [9]. To compare the two methods, we have generated synthetic data U , similar to those of Fig. 1, representing a parabolic profile established after a vein pinching. A uniform noise (between 0 and 1) is added to the simulated signal. The value $T = 0.4$ allows us to generate data comparable with the experimental data of Fig. 1. The studied variable is the shear at the wall obtained by $\dot{\gamma} = dV/dr$, for $r = R$, R being the radius of the vessel, if one supposes a parabolic profile. Fig. 2 presents the temporal evolution (in arbitrary units) of shear at the wall computed for the noisy data \mathbf{V} , and after regularization, $\hat{\mathbf{V}}_{\text{opt}}$. One observes a great difference before and after regularization: in the second case the shear at the wall presents a well defined plateau which makes possible its determination.

To compare the evolution of the shear at wall, several sets of data are generated for values of $T = 0.20, 0.40, 0.60, 0.80, 1.00$. For each data series one calculates the average plateau value and the standard deviation, starting from the data \mathbf{V} , the data \mathbf{V}_w after Wiener filtering, and the regularized data $\hat{\mathbf{V}}_{\text{opt}}$. Table 1 gives the numerical results.

For the regularized data, the evolution of the standard deviations as a function of noise level (columns ec. type of Table 1) is almost linear, with a weak slope; this is an indicator of the robustness of the technique relative to noise. It is observed that for a high level of noise ($T = 1.00$) the standard deviation is about than 10% of the average value. It should be noticed that the Wiener filter is penalized because we compute the shear at the wall, which poses convergence problems.

4. Conclusion

We present a practical application of the theory of the regularization on the high frequencies filtering problem, a classical phenomenon in experiments on veins. The principal interest is this facility of use on wall sided data. The next development way is to study the influence of the optimal parameter λ on the computed solutions.

1. Introduction

En recherche et développement médical sur le système veineux, les données en débit et en calibre des vaisseaux sont issues des échographes. L'utilisation et l'interprétation de ces données expérimentales contribuent à la connaissance de l'hémodynamique du réseau veineux. Malgré les avancées techniques, ces données sont souvent très difficiles à exploiter à cause des limites expérimentales, car outre le problème de résolution des appareils, les écoulements sont perturbés par les conditions de mesure mêmes. Ce phénomène est d'autant plus important que le diamètre des vaisseaux est petit et la vitesse débitante basse. Par exemple, la vitesse débitante moyenne dans une veine chez un sujet en orthostatisme est à la limite de la résolution des appareils (de l'ordre de quelques cm/s).

L'analyse de données expérimentales est beaucoup plus répandue dans le système artériel, et ceci en partie parce que les vitesses sont plus élevées et les diamètres des artères plus grands ce qui permet l'utilisation aisée des échographes, par exemple pour l'identification des paramètres représentatifs de l'état artériel par des méthodes non invasives [1]. Dans le domaine du système sanguin, ce type d'approche est moins développé. La Fig. 1 présente le profil spatial des vitesses d'un écoulement sanguin pris par un échographe à un instant t suite à un pincement. L'écoulement passe, lors du relâchement du pincement, de l'arrêt à un profil développé. La phase de montée peut servir à calculer les paramètres physiques du vaisseau et de l'écoulement par des méthodes inverses, et une fois le profil établi, il peut servir à calculer le cisaillement à la paroi. En observant les données expérimentales, on réalise l'importance d'avoir des données réalistes et fiables, et donc de posséder des techniques de lissage de données ou de réduction de bruit qui puissent les fournir. De nombreuses techniques de lissage et de filtrage existent dans la littérature et peuvent être appliquées sur des données expérimentales issues de la recherche veineuse comme par exemple à base du filtrage de Kalman [2] ou de la modélisation et de l'identification auto-régressive (AR et ARMA) [3]. D'autres techniques déjà utilisées dans le domaine artériel comme la déconvolution des profils de vitesses [4] peuvent aussi servir au filtrage des signaux, mais, d'une manière générale, il n'y pas de technique spécifique aux données issues de la recherche veineuse. En particulier, dans le domaine médical, le filtrage spectral (linéaire) est très utilisé. Cette approche pose néanmoins des problèmes de convergence pour les données proches des parois.

Parce que la compréhension des phénomènes à basse vitesse, et le développement de nouvelles théories—croissance des sténoses, calcul de la contrainte maximale à la paroi [5], dommages causés sur les parois des vaisseaux par le cisaillement à la paroi—passent par une amélioration de la qualité des données expérimentales, une technique simple de lissage de données est proposée. Cette note présente donc une méthode de lissage basée sur des idées issues de la théorie de la régularisation des courbes, et comparée numériquement à une méthode spectrale puissante, le filtre optimal de Wiener.

2. Méthode de régularisation

Soit $V_i \pm \delta V_i$ ($i = 1, N$) un profil de vitesses à l'instant t (l'indice temporel t est éliminé par souci de lisibilité), δV_i étant l'erreur expérimentale. L'objectif est de trouver une courbe lisse C , composée d'une série de valeurs \widehat{V}_i , qui passe par le jeu de données expérimentales $V_i \pm \delta V_i$. Il est simple de déduire qu'un nombre infini de courbes lisses remplissent cette condition, donc le problème est mal posé dans le sens défini par Hadamard : un problème est bien posé quand une solution unique existe, et qu'elle dépend continuellement des données initiales. Pour limiter l'espace des solutions possibles, on essaye de rétablir les critères insatisfaits en utilisant des connaissances a priori sur le système. Cette approche définit un domaine de recherche appelé la théorie de régularisation [6].

L'idée de base consiste à attaquer le problème comme un problème d'optimisation, en définissant une fonctionnelle avec des contraintes qui reflètent les exigences physiques du système. On cherche donc une fonctionnelle H , fonction des données expérimentales V_i et des nouvelles données \widehat{V}_i , guidée par la connaissance a priori sur la régularité du profil de vitesses et du niveau de bruit du signal V_i . Pour cela on utilise l'équation suivante [7] :

$$H = \frac{1}{2} \sum_i (\widehat{V}_i - \widehat{V}_{i+1})^2 + \frac{1}{2} \lambda \sum_i (\widehat{V}_i - V_i)^2 \quad (1)$$

Le premier terme contrôle localement le lissage en forçant le choix de la solution dans l'ensemble des fonctions continues dérivables deux fois, et le deuxième pénalise la non-fidélité aux données initiales. Le paramètre λ donne la mesure de l'importance relative de chaque terme et il doit être ajusté. Puisqu'elle est une forme quadratique de V_i et \widehat{V}_i , H n'a qu'un minimum, qui est une fonction du paramètre λ .

Du point de vue théorique on se place dans le cadre du problème central de la régularisation de Tikhonov : comment choisir λ pour produire une solution proche de la solution non bruitée. De nombreux travaux ont été réalisés récemment en utilisant des techniques variées comme le principe de la validation croisée, la visualisation des résiduels, les graphiques des composantes singulières (voir la référence [8] pour détails). Dans cette étude la valeur de λ est calculée grâce à la bonne estimation de l'erreur résiduelle $\|e\|_2$, c'est à dire, on choisit λ tel que la norme résiduelle satisfait $\|\widehat{\mathbf{V}} - \mathbf{V}\|_2 = \|e\|_2$ où $\widehat{\mathbf{V}} = \{\widehat{V}_1, \dots, \widehat{V}_N\}$.

Le problème formel est le suivant : soit $H(\widehat{V}_1 \dots \widehat{V}_N)$ définie dans l'espace de dimension N , nous cherchons le minimum global à partir d'une condition initiale sur H . L'évolution du vecteur $\widehat{\mathbf{V}}$ est définie par la règle du gradient en utilisant la dérivation de Gâteaux : la vitesse de $\widehat{\mathbf{V}}$ est proportionnelle à sa pente ce qui assure qu'au minimum $d\widehat{\mathbf{V}}/d\tau = 0$. On remarque que τ n'est nullement relié au temps physique, c'est un paramètre d'itération qui gère la vitesse de convergence. L'équation d'évolution de \widehat{V}_i est donnée par

$$\frac{d\widehat{V}_i}{d\tau} = -\frac{\partial H}{\partial \widehat{V}_i} \quad (2)$$

donc,

$$\frac{d\widehat{V}_i}{d\tau} = (\widehat{V}_{i+1} + \widehat{V}_{i-1} - 2\widehat{V}_i) + \lambda(V_i - \widehat{V}_i)$$

Le premier terme est la version discrète du Laplacien. Si $\lambda = 0$ on a simplement une équation de diffusion et les \widehat{V}_i vont évoluer vers un état où \widehat{V}_i est constant pour tout i . En revanche, si $\lambda \neq 0$, le deuxième terme pousse les valeurs de \widehat{V}_i vers les données expérimentales. L'algorithme pour l'équation d'évolution est finalement

$$\begin{aligned} \widehat{V}_i^{\tau+\Delta\tau} &= \widehat{V}_i^\tau + \Delta\tau (\widehat{V}_{i+1}^\tau + \widehat{V}_{i-1}^\tau - 2\widehat{V}_i^\tau) + \lambda\Delta\tau (V_i - \widehat{V}_i^\tau), \quad i = 2, \dots, N-1 \\ V_i - \delta V_i &< \widehat{V}_i < V_i + \delta V_i, \quad i = 2, \dots, N-1 \\ \widehat{V}_1 &= \widehat{V}_N = 0 \\ \widehat{\mathbf{V}}^0 &= \mathbf{V} \end{aligned}$$

Au minimum on trouve le vecteur optimal \hat{V}_{opt} . Il faut remarquer que seules les hautes fréquences sont rejetées lors du processus. La Fig. 1 présente l'application de la méthode de régularisation sur des données expérimentales.

3. Application numérique

Le filtre optimal de Wiener est une technique spectrale très puissante qui utilise la transformée rapide de Fourier. On peut définir, à partir du spectre du signal $W(f)$, un filtre $\phi(f)$ dans l'espace de Fourier tel que le spectre de $W_w(f) = W(f)\phi(f)$ soit proche du spectre original dans le sens des moindres carrés. Le signal est ensuite transformé dans l'espace réel, V_w , et utilisé pour les calculs numériques. Une introduction aux méthodes utilisant la transformée rapide de Fourier est présentée en [9].

Pour comparer les deux méthodes on génère des données synthétiques U semblables à celles de la Fig. 1, représentant un profil parabolique établi après un pincement de veine. Un bruit T est ajouté pour simuler complètement le signal, soit $V = U + T(0,5 - \text{unif}(0, 1))$, où $\text{unif}(0, 1)$ est un générateur de bruit uniforme entre 0 et 1. La valeur $T = 0,4$ permet de générer des données comparables aux données expérimentales de la Fig. 1. La variable dynamique étudiée est le cisaillement à la paroi obtenu par $\dot{\gamma} = dV/dr$ supposant un profil parabolique avec $r = R$, où R est le rayon du vaisseau. La Fig. 2 présente l'évolution temporelle (en unités arbitraires) du cisaillement à la paroi calculé pour des données synthétiques bruitées V pour $T = 0,4$, et pour les mêmes données après régularisation, \hat{V}_{opt} . On y observe une grande différence avant et après régularisation : dans le deuxième cas, le cisaillement à la paroi présente un plateau bien défini qui permet de le déterminer sans ambiguïté.

Figure 1. Données expérimentales et données après régularisation montrant le profil des vitesses d'un écoulement sanguin, la lumière spatiale est de l'ordre de 1 cm, le pic maximum de vitesse est de l'ordre de 2 cm/s.

Figure 1. Experimental and regularized velocity profile, spatial width is about 1 cm and the maximum velocity is roughly 2 cm/s.

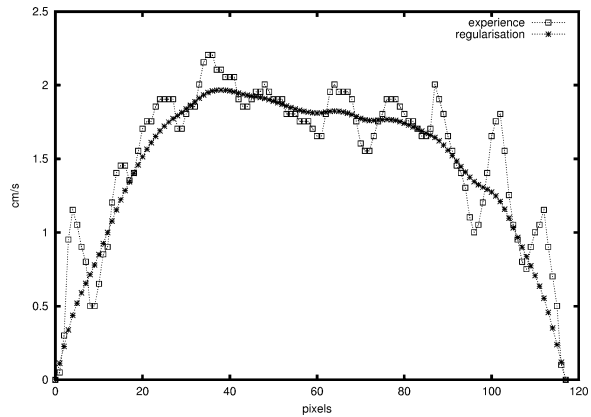


Figure 2. Évolution temporelle du cisaillement à la paroi pour des données synthétiques bruitées ($T = 0,4$) avant et après régularisation. Le temps est donné en unités arbitraires.

Figure 2. Temporal evolution of the shear at the wall for synthetic data ($T = 0.4$) before and after regularization. Time is given in arbitrary units.

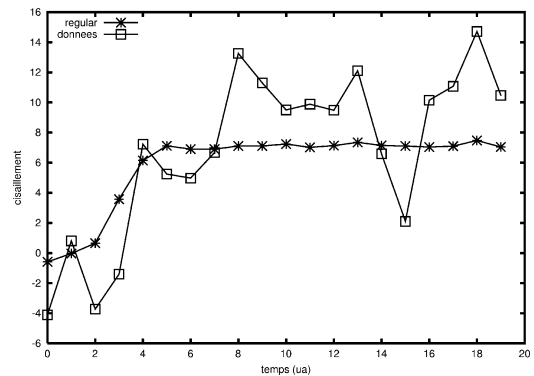


Tableau 1. L'évolution de l'écart type du cisaillement à la paroi en fonction du bruit T pour les données brutes, après le filtrage de Wiener et après régularisation. La valeur optimale de λ est entre parenthèses

Table 1. Shear wall as function of noise T for experimental data, after Weiner filtering and after regularization

| T | ec. type données | ec. type wiener | ec. type régularisation (λ_{opt}) |
|------|------------------|-----------------|---|
| 0,20 | 1,76 | 1,35 | 0,49 (0,039) |
| 0,40 | 3,53 | 2,38 | 0,67 (0,048) |
| 0,60 | 5,29 | 3,76 | 0,80 (0,064) |
| 0,80 | 7,06 | 5,11 | 0,90 (0,095) |
| 1,00 | 8,83 | 6,07 | 0,98 (0,188) |

Pour comparer l'évolution du cisaillement à la paroi en fonction du bruit T , plusieurs séries de données sont générées pour des valeurs de $T = 0,20, 0,40, 0,60, 0,80, 1,00$. Pour chaque série on calcule dans le plateau la valeur moyenne et l'écart type (ec. type) à partir des données \mathbf{V} , des données \mathbf{V}_w après le filtrage de Wiener, et après régularisation $\hat{\mathbf{V}}_{opt}$. La Table 1 présente les résultats numériques : le niveau de bruit T , les écarts type pour chaque méthode (la valeur théorique du cisaillement étant de 7,80) et la valeur optimale de λ .

Pour les données régularisées, l'évolution des écarts type en fonction du niveau de bruit T (les colonnes ec. type de la Table 1) est presque linéaire avec une pente assez faible ; ce qui est un indicateur de la robustesse de cette technique par rapport au bruit. On observe que pour un grand niveau de bruit ($T = 1,00$) l'écart type est de l'ordre de 10% par rapport à la valeur moyenne. Il faut remarquer que le filtre optimal de Wiener est pénalisé car le calcul du cisaillement utilise les données proches de la paroi, ce qui pose, en général, des problèmes de convergence pour les méthodes spectrales.

4. Conclusion

Nous présentons une application pratique de la théorie de la régularisation au problème de filtrage des hautes fréquences, phénomène typique rencontré dans l'analyse du système veineux. L'intérêt principal de l'approche est sa simplicité et son utilisation possible sur des données proche la paroi des vaisseaux. Un axe de recherche futur est l'étude théorique et numérique de l'influence de la valeur du paramètre λ optimal sur les profils optimaux.

Remerciements. L'auteur remercie P. Fourgeau (Laboratoires Innothera) pour les données expérimentales.

Références bibliographiques

- [1] F. Bruni, Méthodes d'extraction de paramètres caractéristiques de l'état artériel par des méthodes non invasives, Ph.D. thesis, Université Paris 7, 1998.
- [2] R.E. Kalman, A new approach to linear filter and prediction problems, Trans. ASME Control Ser. A 1 (1960) 152–192.
- [3] H. Tong, Non-Linear Time Series: A Dynamical System Approach, Clarendon Press, Oxford, 1990.
- [4] P. Flaud, A. Bensalah, P. Peronneau, Deconvolution process in measurement of arterial velocity profiles via an ultrasonic pulsed doppler velocimeter for evaluation of the wall shear rate, Ultrasound in Med. and Biol. 23 (3) (1997) 425–436.
- [5] S. Lorthois, P.-Y. Lagree, Écoulement dans un convergent axisymétrique : calcul de la contrainte de cisaillement pariétale maximale, C. R. Acad. Sci. Paris, Serie Iib 328 (2000) 33–40.
- [6] T. Poggio, V. Torre, Ch. Koch, Computational vision and regularization theory, Nature 317 (1985) 314–319.
- [7] J. Hertz, A. Krogh, R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, New York, 1991.
- [8] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, PA, 1998.
- [9] J. Max, Méthodes et techniques de traitement du signal et applications aux mesures physiques, Masson, Paris, 1989.