Rapport sur un nouvel algorithme de correction de la ligne de base pour THEM Geophysics

Splines, interpolation des moyennes et moindres carrés (SIM)

THEM1999-02

Daniel Lemire, Ph.D.

Montréal, 12 juin 1999

Définitions et quelques suppositions

On définit ici une *fenêtre* comme étant l'intervalle entre le début d'une impulsion et le début de l'impulsion suivante ($k_{\max}$ = 512 échantillons). On prend pour acquis qu'il existe un court temps à la fin de la fenêtre où le signal devrait être très faible (*late window*) débutant à k_1. De plus, comme le signal mesure en fait la dérivée du champs, sa masse devrait être nulle.

Un peu de mathématique

Alors qu'il est utile de procéder à tâtons dans un premier temps, ceci est de moins en moins acceptable avec un système ayant la maturité de celui de THEM Geophysics. Il est donc nécessaire de recourir, le plus possible, à un cadre formel. Il ne s'agit pas ici de faire des mathématiques pour faire des mathématiques, mais bien de chercher à pouvoir démontrer clairement les propriétés optimales de l'algorithme plutôt que de simplement affirmer : "Ça semble bien marcher...".

On doit calculer, pour chaque fenêtre un polynôme qui représente la ligne de base (ou correction) du signal. Pour ce faire on pose 3 conditions :

1. il doit y avoir continuité de la ligne de base d'une fenêtre à l'autre;
2. la masse du signal corrigé devrait être nulle de 0 à k_1;
3. l'énergie[1] résiduelle du signal de k_1 à $k_{\max}$ devrait être minimale.

La dernière condition est une variante de l'algorithme de correction de la ligne déjà utilisé par THEM Geophysics qui prenait plutôt la forme suivante :

3 (précédent algo.). la masse du signal de k_1 à $k_{\max}$ devrait être nulle.

Cette condition, qui limite le degré du polynôme à 2, a l'inconvénient de permettre des oscillations assez forte dans le signal (voir Figure 1). Ces oscillations vont rendre l'algorithme instable dans des conditions où il y a un fort bruit dans les basses fréquences (autour de 60 Hz) i.e. l'algorithme ne corrigera pas correctement la ligne de base.

---

[1] L'énergie d'un signal digital est défini comme la somme de ses valeurs au carré.

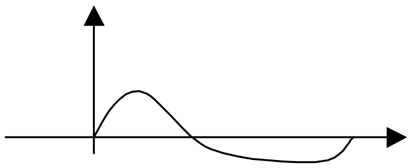


Figure 1. Exemple d'un signal ayant une masse nulle

Nous avons donc un algorithme qui s'appuie à la fois sur la théorie des splines (conditions 1), l'interpolation des moyennes (condition 2) et un algorithme des moindres carrés (condition 3). Nous appelerons donc cet algorithme SIM_{n} où n est le degré de l'algorithme.

Utilisons un polynôme $p(x)$ de degré n

$$p(x) = \sum_{j=0}^{n} a_j x^j \ .$$

Le signal original est noté par $\{s_k\}_{k=0}^{k_{\max}}$. La condition 1 peut simplement être satisfaite en posant $p(0) = y_0$ où y_0 est une valeur imposée par la ligne de base calculée dans le signal qui précède la fenêtre courante. La condition 2 quant à elle équivaut à l'équation

$$\sum_{k=0}^{k_1} p(k) = \sum_{k=0}^{k_1} s_k \ .$$

Finalement, la dernière condition peut être résolu en utilisant la méthode de Lagrange. Soit le Lagrangien

$$L(a_0, \ldots, a_n, \lambda_0, \lambda_1) = \sum_{k=k_1}^{k_{\max}} (p(k) - s_k)^2 - \lambda_0 l_0 - \lambda_1 l_1$$

où

$$l_0(a_0, \ldots, a_n) = p(0) - y_0$$

et

$$l_1(a_0, \ldots, a_n) = \sum_{k=0}^{k_1} p(k) - \sum_{k=0}^{k_1} s_k \ .$$

Nous obtenons alors n + 1 conditions données par

$$\frac{\partial L(a_0, \ldots, a_n, \lambda_0, \lambda_1)}{\partial a_i} = 0 \ \text{pour } i = 0, \ldots, n.$$

Après simplification, cette dernière équation devient

$$2\sum_{k=k_1}^{k_{\max}} k^i p(k) - \lambda_0 \delta_{0,i} - \lambda_1 \sum_{k=0}^{k_1} k^i = 2\sum_{k=k_1}^{k_{\max}} k^i s_k \ \text{ où } \ \delta_{i,0} = \begin{Bmatrix} 1 \ si \, i = 0 \\ 0 \, autrement \end{Bmatrix}.$$

On reconnaît dans cette équation les moments du signal et du polynôme recherché :

$$M_i(p) = \sum_{k=k_1}^{k_{\max}} k^i p(k)$$

et

$$M_i(s) = \sum_{k=k_1}^{k_{\max}} k^i s_k \ .$$

On constate que alors que le problème peut être résolu par un ensemble d'équations linéaires et donc, par la simple inversion d'une matrice n + 3 par n + 3. Il est à noter que seul le membre de droite est modifié et donc le calcul de l'inverse de la matrice n'a à être fait qu'une seule fois ce qui permet une implémentation rapide.

Degré de l'algorithme

Il est nécessaire de choisir avec soin le degré des polynômes utilisés. Le choix peut se faire d'après le nombre d'extremums de la ligne de base à l'intérieur d'une fenêtre. Le degré optimal devrait être exactement le nombre d'extremums plus un permis dans la ligne de base d'une fenêtre (512 échantillons). Comme on trouve parfois jusqu'à 2 extremums clairement définis, un minimum et un maximum (les dos d'âne tels qu'observés par Bernard Kremer), un algorithme du 3e degré devrait donner des résultats pratiquement optimaux. Avec un algorithme du 2e degré, on pourrait montrer que l'on retrouve pratiquement le précédent algorithme (celui utilisé par THEM Geophysics jusqu'à présent).

L'inconvénient à l'utilisation de polynômes de trop haut degré est similaire à ce qui se produit si on utilise un filtre passe-bas avec un seuil trop élevé. On risque de perdre de l'information et de rendre invisible certains détails qui sont pourtant significatifs. Alors que l'algorithme fournit à THEM Geophysics permettra des polynômes d'un degré arbitraire, il faut prendre gare ! Des polynômes de degré 4 ou plus risquent de rendre le code instable numériquement (valeurs aberrantes) et risque de faire perdre des informations qu'on ne pourra pas récupérer plus tard dans le traitement.

Aspect informatique

Cet algorithme est extrêmement rapide. Aussi rapide en fait que l'algorithme précédent (même complexité et temps de calcul du même ordre).

Résultats

Nous avons comparé le schéma SIM_3 avec le schéma présentement utilisé par THEM Geophysics (SIM_2). Sur des données peu bruitées, les résultats sont très similaires (résultats non présentés), mais sur des données bruitées (avec le bruit dit "*de 60 Hz*"), le nouveau schéma tend à être plus stable. On observe une amélioration parfois significative (voir Figure 1 et 2).

Figure 1. Signal traité par l'interpolation des moyennes avec splines (schéma déjà utilisé par THEM Geophysics : correspond environ à SIM_2). ($k_1 = k_{max}$ - 48)

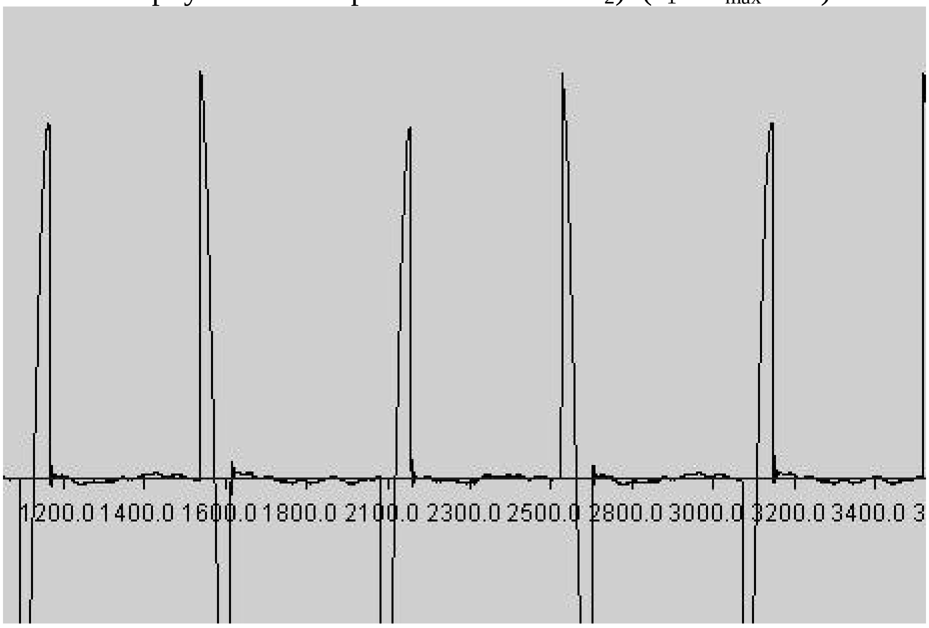
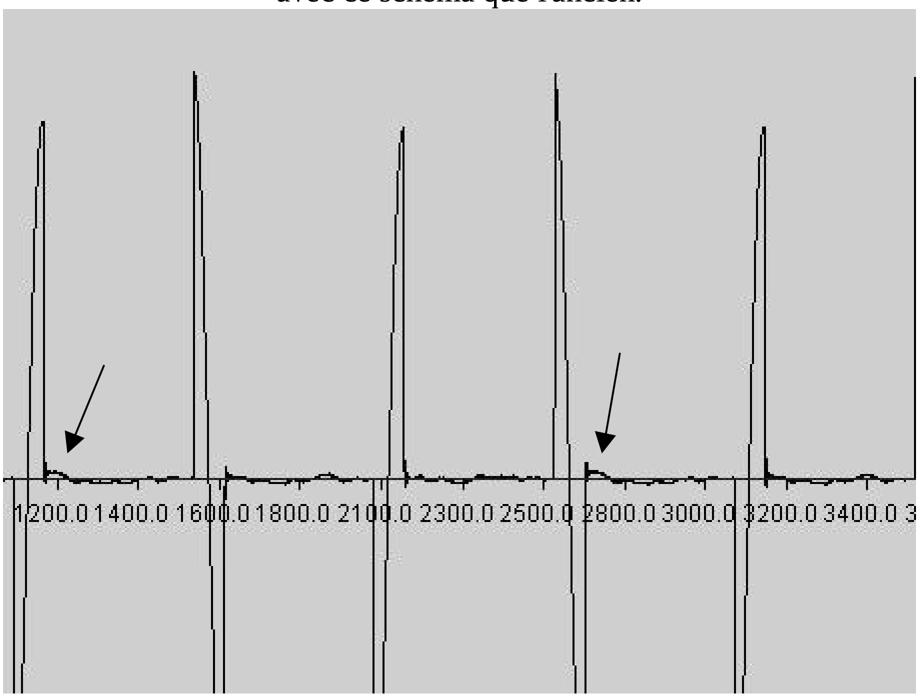


Figure 2. Même signal traité avec SIM3. Les flèches indiquent des endroits où le signal semble meilleur avec ce schéma que l'ancien.



Conclusion

Ce nouveau schéma devrait mener à une stabilité plus grande et une meilleure qualité de signal. Elle pourra diminuer l'importance du filtrage passe-haut et de la superposition (stacking) dans système de THEM Geophysics.

L'intégration au logiciel de THEM Geophysics (via un module en C) ne pose aucune difficulté, mais il convient au préalable de voir s'il est nécessaire de faire certains ajustements. En particulier, il est important d'évaluer avec soin ce nouveau schéma et pour ce faire, une visualisateur écrit en Java permettant de comparer les deux lignes de base est fourni.

Figure 3. Logiciel en Java pour comparer les algorithmes.