

TRAVAUX PRATIQUES N°1

Débruitage de signaux 1D

Objectif : Dans ce TP, on introduit le concept du débruitage de signal pour proposer une première application de la méthode du gradient explicite.

1 Du bruit dans les signaux

Soit $n \in \mathbb{N}$. Soit $u \in \mathbb{R}^n$ un signal (ici un vecteur) qui nous intéresse. Malheureusement, nous ne sommes capables que d'en mesurer une version *bruitée* (c'est-à-dire corrompue par des erreurs et / ou des imprécisions de mesure) $v \in \mathbb{R}^n$. On peut décomposer le signal *mesuré* v de la manière suivante :

$$v = u + b$$

Le vecteur b est appelé *bruit* (en anglais, *noise*) ; dans cette écriture, on considère un modèle de bruit *additif*. Ses valeurs doivent être petits comparés à celles du vecteur u : cela traduit le fait que, globalement, le signal u domine le bruit. Dans le cas contraire (signal noyé dans le bruit), il ne serait pas raisonnable d'espérer pouvoir séparer le signal du bruit.

Lorsque les sources de bruit ne sont pas identifiables, mais qu'elles résultent d'un phénomène aléatoire qui n'a pas de biais particulier (par exemple, s'il ne s'agit pas d'un arrondi systématique à l'entier supérieur, qui aurait en moyenne tendance à surévaluer les valeurs) et qui affectent les données de manière indépendante (sur un vecteur par exemple, les erreurs sur une composante n'influent pas sur celles des autres composantes), alors une bonne façon de modéliser ces erreurs est de considérer que les composantes du vecteur de bruit b sont des réalisations **indépendantes** d'une loi normale centrée (de moyenne nulle) et de variance σ constante (identique pour toutes les composantes). Autrement dit,

$$b \sim \mathcal{N}(0, \sigma)^n$$

Suivant les problèmes, d'autres modèles de bruit peuvent être plus adaptés : bruit de POISSON, bruit multiplicatif...

Le principe du débruitage est d'estimer, le plus précisément possible, le vecteur u non bruité à partir de sa version bruitée v .

Le débruitage fait partie d'une classe de problèmes plus large, appelés *problèmes inverses*.

De nombreuses techniques de débruitage existent ; dans ce TP (et ceux affiliés), on s'intéresse à une famille de méthodes, appelées *méthodes variationnelles*. Le principe est le suivant : partant d'un modèle décrivant la formation de v (donc, en particulier, celle de u), on introduit une métrique $J : \mathbb{R}^n \rightarrow [0; +\infty[$ qui permet de mesurer la distance entre tout vecteur $x \in \mathbb{R}^n$ et ce modèle de formation. Autrement dit, plus $J(x)$ est faible, plus x satisfait les hypothèses du modèle. La suite est alors évidente : pour débruiter, on va chercher à trouver un minimiseur de la fonction objectif J .

Ici, pour simplifier, on s'est placé dans le cas d'un signal discret 1D de longueur finie (un vecteur de \mathbb{R}^n). Mais il est tout-à-fait possible d'étendre cette démarche dans le cas de signaux multi-dimensionnés (comme les images) et / ou non discret / non fini. Il est ainsi possible de rechercher, au lieu d'un vecteur, une fonction; c'est la raison pour laquelle J est souvent appelée *fonctionnelle* (ou même *fonctionnelle d'énergie*).

Le choix de la fonction objectif J est déterminant dans la réussite de cette approche. Tout d'abord, il faut que J (et le modèle sous-jacent) soit choisi de manière adéquate par rapport au problème considéré. Autrement dit, il faut que J admette au moins un minimiseur et que les minimiseurs de J soient proches du signal recherché u (idéalement, il faudrait que u soit l'unique minimiseur de J). On voit pour cela que le modèle de formation doit être bien construit et que la métrique J (qui n'est pas forcément unique) permette bien de mesurer la distance à ce modèle. Un autre critère à prendre en considération est la capacité à calculer en pratique un minimiseur pour J , et, suivant les applications, de manière suffisamment efficace.

Les fonctionnelles J comprennent généralement deux types de termes. Tout d'abord un terme d'attache aux données (ou encore de fidélité), qui traduit le fait que le signal reconstruit doit rester proche du signal observé v (puisque l'on suppose qu'il reste dans le signal corrompu suffisamment d'informations pour pouvoir estimer le signal propre). Souvent, la proximité est définie à l'aide de la norme quadratique, car le bruit est (supposé) de nature gaussienne. Le second type de terme est celui des termes de régularisation. Ils servent à encoder des propriétés connues de l'objet à estimer (en général, la régularité du signal, les valeurs extrémales qu'il peut prendre, des conditions sur sa norme).

Enfin, il existe généralement un paramètre de pondération entre ses deux termes pour ajuster le poids de ces deux hypothèses.

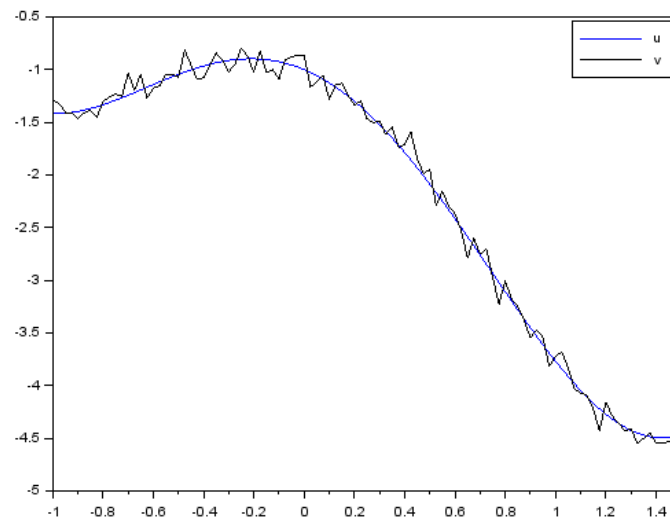


FIGURE 1 – Signal u et exemple de signal v . Attention : ici, le signal u est supposé lisse (ou en pratique, une discrétisation d'un signal lisse) et le signal v est bien discret, mais représenté à l'aide d'une interpolation affine.

2 Position du problème

Dans ce TP, on fait les deux hypothèses suivantes :

1. le bruit est additif et de nature gaussienne, et il est petit devant le signal ;
2. le signal à estimer u est une discrétisation d'un signal (très) lisse.

Autrement dit, le signal observé vérifie

$$v = u + b \quad \text{avec} \quad b \sim \mathcal{N}(0, \sigma)^n$$

et les variations du signal u sont faibles, dans le sens où les différences $u_{(i+1)} - u_{(i)}$ sont petites.

Soit $n \in \mathbb{N}$ et $v \in \mathbb{R}^n$. On considère donc le problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} J(x) = \frac{\lambda}{2} \|x - v\|_2^2 + \frac{1}{2} \|D(x)\|_2^2$$

où $D : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$ est une application linéaire, définie pour tout $x = (x_{(i)})_{1 \leq i \leq n}$ par les différences finies

$$\forall i \in \llbracket 1; n \rrbracket, \quad (D(x))_{(i)} = \begin{cases} x_{(i+1)} - x_{(i)} & \text{si } i \neq n \\ 0 & \text{sinon} \end{cases}$$

Cette régularisation est connue sous le nom de *régularisation de TIKHONOV*.

Le paramètre λ est un réel strictement positif, qui permet d'ajuster la régularité du signal débruité. Plus il est important, plus le terme de fidélité va être décisif dans la minimisation, et plus les minimiseurs de J vont être proches de v . *A contrario*, s'il est faible, les minimiseurs de J vont avoir des termes de régularité associées faibles, donc être très réguliers.

Exercice 1 – Optimisation de J

- (a) Montrer que J est différentiable et que

$$\forall x \in \mathbb{R}^n, \quad \nabla J(x) = \lambda(x - v) + D^*(D(x))$$

- (b) Vérifier que

$$\forall x \in \mathbb{R}^n, \quad D(x) = \underbrace{\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -1 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}}_{=D} \begin{pmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{pmatrix}$$

- (c) En déduire que

$$\forall x \in \mathbb{R}^n, \quad \nabla J(x) = (\lambda I_n + D^\top D) x - v$$

- (d) Montrer que J est convexe et fortement convexe de module λ .

- (e) En déduire que J admet un unique minimiseur x^* , donné par

$$x^* = (\lambda I_n + D^\top D)^{-1} v$$

Pour les expériences numériques, on va générer des données v satisfaisant les hypothèses énoncées en début de section. On va donc ajouter un bruit additif gaussien à un signal très lisse u . On peut choisir par exemple

$$\forall i \in \llbracket 1; n \rrbracket, \quad u_{(i)} = -\exp(t_i) - 6 t_i \sin t_i + 4 t_i^2;$$

avec t_i des valeurs régulièrement espacées dans \mathbb{R} .

Exercice 2 – Optimum

Afficher sur un même graphique les vecteurs v , x^* et u . Tester avec différentes valeurs de λ (par exemple $\lambda \in \{1, 0.1, 0.001\}$). Commenter quant à l'influence du paramètre λ sur la reconstitution du signal. Afficher l'erreur de reconstruction $\|x^* - u\|$ pour chaque valeur de λ testée.

3 Méthode de gradient explicite

Dans le problème considéré dans la section précédente, on a vu qu'il était possible de calculer explicitement le minimiseur de la fonctionnelle J . Il n'est donc pas nécessaire de recourir à un algorithme d'optimisation pour la minimiser. Cependant, dans cette section, on va le faire à titre d'exercice.

On rappelle les itérations de la méthode du gradient explicite à pas fixe pour la minimisation de la fonction objectif J :

$$x_0 \in \mathbb{R}^n \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si la fonction J est convexe et à gradient lipschitzien de constante de LIPSCHITZ L , alors la convergence de la méthode du gradient est assurée dès que le pas de temps τ est choisi strictement inférieur à $2/L$.

Exercice 3 – Régularité de J

(a) Montrer que
$$\|D\| = \sup_{x \neq 0} \frac{\|D(x)\|_2}{\|x\|_2} \leq 2.$$

On pourra majorer la quantité $\|D(x)\|_2^2$ à l'aide de l'inégalité $(a-b)^2 \leq 2(a^2+b^2)$ valable pour tout couple $(a, b) \in \mathbb{R}^2$.

(b) Montrer que ∇J est lipschitzien, de constante de LIPSCHITZ

$$L = \lambda + 4$$

Exercice 4 – Méthode du gradient à pas fixe

Mettre en œuvre la méthode du gradient à pas fixe. Pour l'initialisation de la méthode, on pourra choisir $x^0 = v$. Comme critère d'arrêt, on pourra fixer un seuil sur la norme du gradient de J au point courant x_k . *On pensera à afficher les graphiques correspondant à l'évolution de $J(x_k)$, $\|\nabla J(x_k)\|$ et de $\|x_k - x^*\|$ au cours du temps. Vers quoi ces suites tendent-elles ? Qu'en est-il de la suite des $\|x_k - u\|$?*

4 Compléments

Outre le pas fixe dans le cas régulier, il existe différentes façons de choisir le pas de temps dans la méthode du gradient explicite. Ces choix, qui conduisent à des pas variables, visent à adapter le pas en fonction de la topologie locale de la fonction objectif J , et *in fine*, à accélérer la convergence. En général, de tels procédés ont démontré une efficacité empiriques, mais non démontrée théoriquement.

Dans la méthode du gradient à pas optimal, on choisit pour tout $k \in \mathbb{N}$

$$\tau_k \in \arg \min_{\tau \in \mathbb{R}} J(x_k + \tau \nabla J(x_k))$$

L'idée est d'aller le plus loin possible (dont, au-delà de la limite imposée par la constante de LIPSCHITZ dans le cas des pas fixes) dans la direction courante.

Exercice 5 – Pas optimal

- (a) Soit $d \in \mathbb{R}^n$. Justifier que la fonction réelle

$$\tau \mapsto J(x + \tau d)$$

est une fonction convexe infinie à l'infini

- (b) En déduire que son minimiseur τ^* existe et est unique, et est caractérisé par la condition nécessaire et suffisante d'optimalité du premier ordre, qui s'écrit dans ce cas

$$\langle \nabla J(x + \tau^* d), d \rangle = 0$$

Montrer que

$$\tau^* = - \frac{\langle \nabla J(x), d \rangle}{\lambda \|d\|_2^2 + \|D d\|_2^2}$$

Dans le cas quadratique (attache aux données quadratique et régularisation de TIKHOV), le pas optimal est calculable. Il faut noter qu'en général, ce n'est pas le cas (et dans ce cas, une manière de procéder est de calculer le pas optimal non pour la fonction J , mais pour une approximation quadratique de J autour du point courant). Notons par ailleurs que le calcul du pas optimal peut être coûteux (puisqu'il est recalculé à chaque itération, contrairement au pas fixe).

Exercice 6 – Méthode du gradient à pas optimal

Mettre en œuvre la méthode du gradient à pas optimal. On utilisera la même initialisation et le même critère d'arrêt que dans la méthode du gradient à pas fixe. Comparer la performance des deux algorithmes testés (nombre d'itérations, temps de calculs, erreur finale).

D'autres choix de pas de temps sont possibles. Ainsi, pour la méthode de BARZILAI-BORWEIN, on choisit pour tout $k \in \mathbb{N}$

$$\tau_k = \frac{\langle \Delta g_k, \Delta x_k \rangle}{\|\Delta g_k\|^2}$$

avec $\Delta x_k = x_k - x_{k-1}$ et $\Delta g_k = \nabla J(x_k) - \nabla J(x_{k-1})$.

REMARQUE : Attention : on voit que, pour calculer τ_k , on doit avoir déjà calculé deux points x_k et x_{k-1} . En pratique, cela signifie qu'il faut non seulement initialiser l'algorithme avec

un point x_0 , mais également avec un point précédent x_{-1} différent de x_0 . On pourra choisir par exemple $x_{-1} = -x_0$ si $x_0 \neq 0$.

Exercice 7 – Méthode de BARZILAI–BORWEIN

Mettre en œuvre la méthode de BARZILAI–BORWEIN. On utilisera la même initialisation et le même critère d'arrêt que dans la méthode du gradient à pas fixe. Comparer la performance des trois algorithmes testés (nombre d'itérations, temps de calculs, erreur finale).