Université Paris Cité
UFR de Mathématiques et d'Informatique

Année 23–24

Master 2 - MMA
Optimisation

TP 1 - **Compressed sensing et algorithmes proximaux**

L'objectif de ce TP est de trouver une estimation $\hat{u}$ d'une inconnue u à partir de la donnée

$$d = Au + b, \quad \text{où } A \text{ est une application linéaire connue et } b \text{ est un bruit blanc gaussien.}$$

1. *Cas d'une bijection, sans bruit.* On suppose que

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Calculer la donnée d et lui appliquer la matrice inverse A^{-1} pour retrouver l'inconnue u.

2. *Cas d'une bijection, avec bruit.* On suppose maintenant que $b = (0.1, -0.1, 0.1, -0.1)^T$.

   (a) Calculer la donnée d et lui appliquer la matrice inverse A^{-1}. Commenter le résultat obtenu.

   (b) Afin d'éviter l'explosion du bruit, on se propose d'utiliser une régularisation de Tikhonov ou, en d'autres termes, de prendre comme solution du problème le minimiseur de la fonction
   $$f \;:\; x \mapsto \frac{1}{2}\|Ax - d\|_2^2 + \frac{\lambda}{2}\|x\|_2^2.$$
   Montrer que le minimiseur $\hat{u}_\lambda$ dépend linéairement de d et déterminer numériquement $\hat{u}_\lambda$ pour plusieurs valeurs de λ. Commenter le résultat obtenu.

   (c) On sais cependant que l'inconnue u possède plusieurs composantes nulles. Il conviendrait donc d'utiliser une régularisation qui favorise les solutions parcimonieuse comme la norme $\|\cdot\|_1$.
   $$f \;:\; x \mapsto \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_1.$$
   Afin de faciliter la minimisation de f, on va approcher $\|\cdot\|_1$ par sa régularisé de Moreau-Yosida qui est différentiable.
   $$f_\epsilon \;:\; x \mapsto \frac{1}{2}\|Ax - d\|_2^2 + \lambda\sum_{k=1}^{4} g_\epsilon(x_k), \quad \text{où} \quad g_\epsilon(t) = \begin{cases} \frac{t^2}{2\epsilon}, & \text{si } |t| \le \epsilon, \\ t - \frac{\epsilon}{2}, & \text{si } t > \epsilon, \\ t + \frac{\epsilon}{2}, & \text{si } t < -\epsilon. \end{cases}$$
   Calculer le gradient de f_ϵ et minimiser la fonction à l'aide d'une descente de gradient à pas fixe, pour plusieurs valeurs des paramètres λ et ϵ.

   (d) Utiliser l'algorithme accélé de Nesterov.

(e) Calculer le sous-différentiel de la fonction f non-régularisée;

(f) Minimiser f par une descente de sous-gradient.

(g) Calculer l'opérateur proximal de la régularisation $\lambda\|\cdot\|_1$. Pourquoi un algorithme explicite-implicite est-il bien adapté pour minimiser f ?

(h) Minimiser f par l'algorithme explicite-implicite.

(i) Minimiser f par l'algorithme FISTA.

(j) Pour un paramètre $\tau > 0$ suffisamment petit, déterminer l'opérateur proximal de la fonction $(x \mapsto \frac{\tau}{2}\|Ax - d\|^2)$. Pourquoi l'algorithme de Douglas-Rachford est-il lui-aussi bien adapté pour minimiser f ?

(k) Minimiser f par l'algorithme de Douglas-Rachford.

(l) Comparer la vitesse de convergence des quatre algorithmes.

3. *Cas d'une application non injective (acquisition comprimée), sans bruit.* On suppose maintenant que A est la réalisation d'une matrice aléatoire gaussienne 100×200, centrée, de variance 1, et $b = 0$. Pour construire u, on prend la réalisation d'un vecteur aléatoire gaussien 200×1, centré, de variance 1, puis on met à zéro tous les coefficients de valeur absolue inférieure à $1,5$.

   (a) Générer la matrice A, l'inconnue u et construire la donnée d.

   (b) Trouver une estimation $\hat{u}_\lambda$ de u à partir de d en minimisant la fonction f (régler le paramètre λ afin d'obtenir un signal $\hat{u}_\lambda$ parcimonieux).

4. *Cas d'une application non injective, avec bruit.* On suppose maintenant que b est la réalisation d'un vecteur aléatoire gaussien, centré, de variance 4.

   (a) Générer le vecteur b et construire la donnée d.

   (b) Trouver une estimation $\hat{u}_\lambda$ de u à partir de d en minimisant la fonction f. Que peut-on observer lorsque λ est choisi suffisamment grand pour obtenir un signal parcimonieux ?

   (c) Dans l'algorithme explicite-implicite, changer le seuillage doux en un seuillage dur et relancer l'algorithme en l'initialisant avec l'estimation $\hat{u}$ obtenu précédemment. Interpréter le résultat obtenu.

5. *Cas d'un bruit de basse-fréquence.* La matrice A est l'identité et on suppose que seuls les 25 coefficients de Fourier de plus basse fréquence sont bruités.

   (a) On reprend le même signal original u qu'à la question précédente. Pour le bruit, on prend la réalisation d'un vecteur aléatoire gaussien, centré, de variance 4, puis on met à 0 ses coefficients de Fourier haute fréquence. Construire la donnée $d = u + b$.

   (b) Comme le bruit est de basse-fréquence, on se propose de prendre, comme estimation de u,
   $$\hat{u} = \arg\min_{x \in B} \|x\|_1$$
   où B est l'espace des signaux dont les coefficients de Fourier sont identiques à ceux de d à l'exception des 25 coefficients de plus basse fréquence. Pourquoi l'algorithme de Douglas-Rachford est-il bien adapté à ce problème d'optimisation ?

   (c) Calculer l'opérateur de projection sur B.

   (d) Construire une estimation de u.