Université Bordeaux 1

Matching Pursuit et Minimisation ℓ_1

Introduction

On considère une matrice ayant n lignes et N colonnes avec $N > n$, $x_0 \in \mathbb{R}^N$ et $y = Ax_0$. On cherche à retrouver ou à approcher x_0 à partir de y_0 en utilisant l'hypothèse que x_0 est parcimonieux c'est à dire que son nombre de composantes non nulles est faible.
Pour une matrice A donnée et un y quelconque il n'existe pas d'algorithme rapide permettant de chercher l'antécédent de y par A qui a la norme ℓ_0 la plus faible. En revanche, il existe des algorithmes qui sous certaines conditions assure la convergence vers le vecteur souhaité. Le but de cette feuille est de démontrer certains critères de convergence du Matching Pursuit et de la minimisation ℓ_1.

Notations et définitions

On note dans la suite $I(x_0) = \{i \leqslant N, x_0(i) \neq 0\}$ le support de x_0 et $\|x_0\|_0 = |I(x_0)|$ le cardinal de $I(x_0)$ c'est à dire le nombre de composantes non nulles de x_0. On appelle aussi norme ℓ_0 cette quantité même si $\|\cdot\|_0$ n'est pas une norme mais une pseudo-norme.
Pour un ensemble d'indices I, on note A_I la matrice extraite de I formée des colonnes de A indicées par I. La matrice A_I possède ainsi n lignes et $|I|$ colonnes.
On note $\langle x, y\rangle$ ou $x^t y$ le produit scalaire de x et de y. On utilisera les deux notations. Si la matrice $(A_I^t A_I)$ est inversible on définit l'*Exact Recovery Coefficient* de I par

$$ERC(I) = \max_{j \notin I} \|(A_I^t A_I)^{-1} A_I^t a_j\|_1 \tag{ERC}$$

On définit la cohérence de A par

$$C = C(A) = \max_{i \neq j} \frac{|\langle a_i, a_j\rangle|}{\|a_i\|_2 \|a_j\|_2} = \max_{i \neq j} \frac{|a_i^t a_j|}{\|a_i\|_2 \|a_j\|_2} \tag{1}$$

La cohérence d'une matrice correspond au cosinus de l'angle minimal entre deux vecteurs colonnes de A. Plus la cohérence est grande plus la reconstruction des vecteurs parcimonieux sera difficile.
On note V_I l'espace vectoriel engendré par les vecteurs $(a_i)_{i \in I}$.

1 Condition de reconstruction par *Matching Pursuit*

Dans cette partie nous allons proposer des conditions qui assurent la convergence de l'algorithme du MP et de l'OMP. C'est à dire des conditions sur x_0 qui assure que le Matching Pursuit converge vers x_0 si on prend comme donnée initiale $y_0 = Ax_0$.

1.1 MP et ERC

1. Soit x à support dans I justifier que $z = Ax \in V_I$.
2. Soit $z \in V_I$ justifier que $z = A_I (A_I^t A_I)^{-1} A_I^t z$.
3. Montrer que
$$\forall j \notin I,\ |\langle a_j, z\rangle| = |a_j^t z| \leqslant ERC(I) \|A_I^t z\|_\infty.$$

4. Si on suppose que $ERC(I) < 1$, montrer que $\max_{i\in I}|\langle z, a_i\rangle| > \max_{j\notin I}|\langle z, a_j\rangle|$.
5. En déduire que si $y_0 = Ax_0$ et $ERC(I(x_0)) < 1$, l'indice du premier atome sélectionné par le MP appartient à I et en déduire que le résidu r obtenu à l'issu e de cette première étape est un élément de V_I.
6. En déduire qu'à chaque étape le MP sélectionne un atome indicé sur I et qu'ainsi si $y_0 = Ax_0$ et $ERC(I(x_0)) < 1$, le MP converge vers x_0.
7. Qu'en est il de l'OMP ?
8. Peut on borner le nombre d'étapes nécessaire à la convergence pour le MP et pour l'OMP dans ce cas ?

2 ERC et parcimonie.

On suppose dans la suite que les colonnes $(a_i)_{i\leqslant N}$ de A sont normées (c'est à dire que $\forall i \leqslant N$, $\|a_i\|_2 = 1$). Soit x_0 un vecteur de $\mathbb{R}^N$, on note I son support et $ERC(I)$ l'ERC du support I.

9. Montrer que la matrice $(A_I^t A_I)$ est inversible si seulement si les vecteurs colonnes $(a_i)_{i\in I}$ forment une famille libre.
10. Si $B = (b_{ij})_{i,j}$ est une matrice on note
$$\|B\|_L = \max_j \sum_k |b_{kj}| \tag{2}$$
11. Montrer que $\|\cdot\|_L$ est une norme d'algèbre.
12. Montrer que pour toute matrice B et pour tout vecteur X, $\|BX\|_1 \leqslant \|B\|_L\|X\|_1$.
13. On note $H = I - (A_I^t A_I)$. Montrer que $\|H\|_L \leqslant C(|I| - 1)$.
14. On suppose que $|I| < \frac{1}{2}(1 + \frac{1}{C})$, montrer que la série $\sum_n H^n$ converge dans $\mathscr{M}_{|I|,|I|}(\mathbb{R})$ et que
$$\|(A_I^t A_I)^{-1}\|_L \leqslant \frac{1}{1 - C(|I| - 1)}. \tag{3}$$
15. Montrer que $\|A_I^t a_j\|_1 \leqslant C|I|$.
16. En déduire que si $|I| < \frac{1}{2}(1 + \frac{1}{C})$ alors $ERC(I) < 1$.
17. En déduire que tout vecteur x_0 tel que $|I| < \frac{1}{2}(1 + \frac{1}{C})$ peut être reconstruit par MP.
18. Soit x_0 tel que $|I| < \frac{1}{2}(1 + \frac{1}{C})$, existe t il un vecteur $x_1 \neq x_0$ tel que $Ax_1 = Ax_0$ et $\|x_1\|_0 \leqslant \|x_0\|_0$?
19. En déduire que si $y_0 = Ax_0$ et $|I| < \frac{1}{2}(1 + \frac{1}{C})$ le MP retrouve l'antécédent de y_0 le plus parcimonieux. **Il est important de noter que cette propriété n'est vrai que sous l'hypothèse où les colonnes de A sont normées.**

2.1 Les colonnes de A ne sont pas normées

Supposons maintenant que les colonnes de A ne sont pas nécessairement normées (mais on suppose quand même qu'elles ne sont pas nulles non plus).
Soit x_0 un vecteur tel que $\|x_0\|_0 < \frac{1}{2}(1 + \frac{1}{C})$ et x_1 un vecteur tel que $Ax_1 = Ax_0$. Nous allons montrer que nécessairement $\|x_1\|_0 > \|x_0\|_0$.
Ainsi il n'existe dans ce cas pas de vecteur plus parcimonieux que x_0 ayant la même image même si les colonnes de sont pas normées.

20. Soit A une matrice dont les colonnes sont normées telle que $C < \frac{1}{3}$ et x_0 un vecteur 1 parcimonieux tel que $x(1) = 1$. Justifier que x_0 est identifiable par MP (c'est à dire que x_0 peut être reconstruit à partir de $y_0 = Ax_0$ en utilisant le MP).
21. Montrer qu'en multipliant la première colonne de A par un facteur α suffisamment petit, la première étape du MP selectionne un mauvais atome.
    Dans ce cas rien ne garantit que le MP sera capable de le retrouver. Si on veut utiliser le MP il faudra donc normaliser les colonnes avant de commencer l'algorithme.
    On note D la matrice diagonale dont les coefficients diagonaux d_{ii} sont égaux à $\frac{1}{\|a_i\|_2}$.
22. Montrer que les colonnes de $B = AD$ sont normées et que $C(B) = C(A)$.
23. On note $\tilde{x}_1 = D^{-1}x_1$ et $\tilde{x}_0 = D^{-1}x_0$. Justifier que $B\tilde{x}_1 = B\tilde{x}_0$ et que $\|\tilde{x}_0\|_0 < \frac{1}{2}(1 + \frac{1}{C(B)})$.
24. En déduire que $\|\tilde{x}_1\|_0 > \|\tilde{x}_0\|_0$.
25. En déduire que $\|x_1\|_0 > \|x_0\|_0$.

3 Minimisation ℓ_1.

Une autre manière d'essayer de retrouver x_0 à partir de $y_0 = Ax_0$ est de résoudre le problème suivant :

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ sous la contrainte } Ax = y_0. \tag{4}$$

Si on veut seulement approcher y_0 par un vecteur parcimonieux on peut aussi résoudre :

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ sous la contrainte } \|Ax - y_0\|_2 \leqslant \varepsilon.$$

Nous allons montrer dans cette partie que sous certaines hypothèses sur x_0 et A, on peut reconstruire x_0 par minimisation ℓ_1. En effet on peut donner des conditions relativement simples qui assurent que x_0 est l'unique solution de (4). Dans un premier temps nous ne nous occuperons pas du caractère unique de la solution de (4) et nous dirons qu'un vecteur x_0 est identifiable, ou identifiable par minimisation ℓ_1 s'il est solution de (4).
L'avantage de (4) par rapport à la minimisation ℓ_0 c'est que c'est un problème convexe que l'on sait résoudre numériquement.

26. Soit A la matrice définie par
    $$\begin{pmatrix} 0 & -3 & 4 \\ 4 & -3 & 1 \end{pmatrix}$$
    Dessiner l'image de la boule ℓ_1 notée $\mathscr{B}_{\ell_1}$ par A.
    D'une manière générale l'image de la boule $\mathscr{B}_{\ell_1}$ par une application linéaire A est le polytope $\mathscr{P}$ qui est l'enveloppe convexe des $(\pm a_i)_{i \leqslant N}$ où a_i désigne la ième colonne de la matrice A.
27. Placer sur le dessin l'image du vecteur $x_1 = (\frac{1}{2}, -\frac{1}{2}, 0)$ par A.
28. Placer l'image du point $x_2 = (\frac{1}{3}, 0, \frac{2}{3})$.
29. Pour chacun de ses deux vecteurs existe il un vecteur de la boule $\mathscr{B}_{\ell_1}$ dont l'image soit de la forme αAx_i avec $\alpha > 1$?
    Dans les questions suivantes nous allons montrer le résultat suivant :

    Proposition 1. *Un vecteur $x_0 \in \mathbb{R}^N$ est identifiable si et seulement si $\frac{Ax_0}{\|x_0\|_1} \in \partial\mathscr{P}$ frontière du polytope $\mathscr{P}$.*

Pour la démonstration on traite le cas où $\|x_0\|_1 = 1$ et on a le cas général en multipliant par la norme ℓ_1 de x_0.
On démontre cette proposition en 2 temps :

30. (a) En considérant une matrice A_J extraite de A adéquate, justifier l'existence d'un réel m qui dépend de A tel que pour tout $z \in Im(A)$, il existe $x \in \mathbb{R}^N$ tel que $z = Ax$ et $\|x\|_1 \leqslant m\|z\|_2$.
    (b) Supposons qu'il existe $x_1 \in \mathbb{R}^N$ tel que $Ax_1 = Ax_0$ et $\|x_1\|_1 < \|x_0\|_1$. Justifier qu'il existe $\varepsilon > 0$ tel que $\|x_1\|_1 < \|x_0\|_1 - \varepsilon$.
    (c) Justifier que la boule ℓ_1 centrée en Ax_0 et de rayon $r = \frac{\varepsilon}{m}$ est inclu dans l'intérieur de $\mathscr{P}$
    (d) En déduire que si $Ax_0 \in \partial\mathscr{P}$ alors x_0 est identifiable.
31. Réciproquement si on suppose que Ax_0 appartient à l'intérieur de $\mathscr{P}$
    (a) Jusifier qu'il existe $\lambda > 1$ tel que $\lambda Ax_0 \in \mathscr{P}$.
    (b) En déduire qu'il existe x_1 tel que $Ax_1 = Ax_0$ et tel que $\|x_1\|_1 \leqslant \frac{1}{\lambda}$.
32. Conclure.

3.1 Un critère sur le support et sur le signe de x_0.

On a démontré qu'un vecteur x_0 était identifiable si et seulement si son image appartenait à la frontière de l'image de la Boule ℓ_1 de rayon $\|x_0\|_1$. Nous allons maintenant proposer une condition d'identifiabilité analytique suffisante.

33. Montrer que s'il existe $u \in \mathbb{R}^N$ tel que

$$\forall z \text{ tel que } \|z\|_1 \leqslant \|x_0\|_1 \text{ on ait } \langle Az, u\rangle \leqslant \langle Ax_0, u\rangle$$

alors x_0 est identifiable. (Indication : Que peut on dire de l'hyperplan d'équation $\langle \cdot - Ax_0, u\rangle = 0$?)
Pour un vecteur x de support I dont la matrice active associée A_I est de rang $|I|$ on peut définir le vecteur de $\mathbb{R}^n$ suivant $d = d(x) = A_I(A_I^t A_I)^{-1}$ sign x_I et la quantité suivante :

$$F(x) = \max_{j \notin I} |\langle a_j, d(x)\rangle| \tag{5}$$

34. Montrer que si d est défini $\langle Ax_0, d(x_0), \rangle = \|x_0\|_1$.
35. Montrer que si $d(x_0)$ est défini et si $F(x_0) \leqslant 1$, alors x_0 est identifiable.
36. Montrer que si $ERC(I) \leqslant 1$ alors $F(x) \leqslant 1$. En déduire que si $ERC(I) \leqslant 1$, alors tout vecteur de support I est identifiable.
Unicité
37. Soit x_0 tel que $F(x_0)$ est défini. Montrer que

$$\langle Az, d(x_0)\rangle = \sum_{k=1}^{N} z(k)\langle a_k, d(x_0)\rangle.$$

On suppose maintenant de plus que $F(x_0) < 1$, que $\|z\|_1 \leqslant \|x_0\|_1$. et qu'il existe $k \notin I$ tel que $z(k) \neq 0$.

38. Montrer que

$$\langle Az, d(x_0)\rangle < \|z\|_1.$$

En déduire que $Az \neq Ax_0$ et donc que x_0 est l'unique minimiseur de (4).

39. En déduire que si $ERC(I) < 1$, tout vecteur x_0 de support I est l'unique minimiseur de (4) pour $y_0 = Ax_0$.