

Traitement du signal

Yves Meyer

Les découvertes faites par Emmanuel Candès et Terence Tao en traitement du signal et de l'image s'inscrivent dans la problématique des problèmes inverses mal posés. Il s'agit de reconstruire un objet inconnu alors qu'on ne dispose à son sujet que de quelques informations fragmentaires. Fournir dans ces conditions un résultat exact, comme si l'on disposait d'une information complète, est impossible et l'on doit pallier l'information manquante par une hypothèse a priori. Pour que la démarche proposée soit intéressante, il convient que cette connaissance a priori soit suffisamment souple et n'exclue aucun des signaux ou des images que l'on cherche à reconstruire.

Désignons par $X \in \mathbb{R}^n$ le signal (ou l'image) inconnu et supposons que l'on n'ait réalisé que quelques mesures $y_j = \langle A_j, X \rangle$, $1 \leq j \leq p$, sur ce vecteur inconnu X . Les vecteurs $A_j \in \mathbb{R}^n$ sont donnés. Dans les applications la dimension n est souvent supérieure à 10^6 (5 mégapixels, par exemple) et p est beaucoup plus petit que n . La connaissance a priori sur la solution X est modélisée par un certain ensemble (en général non convexe) Γ et s'écrit $X \in \Gamma$. Alors on cherche le vecteur inconnu X en essayant de résoudre

$$y_j = \langle A_j, X \rangle, \quad j = 1, \dots, p, \quad X \in \Gamma \quad (1)$$

L'objet des travaux d'Emmanuel Candès et de Terence Tao est double. Pour une classe importante d'ensembles Γ , Candès et Tao démontrent que si (1) a une solution, alors cette solution est unique et, dans un second temps, ils proposent un algorithme pour la déterminer.

La définition de la classe Γ est dictée par les travaux sur les *représentations creuses*. Le point de départ concerne l'analyse par ondelettes des images. Les images que l'on peut modéliser par des fonctions appartenant à l'espace BV des fonctions à variation bornée ont une représentation creuse dans une base d'ondelettes (Albert Cohen) et cette découverte mathématique est en plein accord avec les acquis de la neurophysiologie (travaux de David Hubel et Torsten Wiesel sur la détection des contours par les neurones de l'aire V_1 du cortex visuel primaire). Une représentation creuse signifie que l'ensemble T des indices des coefficients d'ondelettes significatifs a une petite cardinalité $|T|$, compte tenu de la taille du signal. Mais ce résultat ne fournit aucune information sur la localisation de cet ensemble T ; cette localisation est déterminée par les positions des bords de l'image. C'est ce qui amène Candès et Tao à considérer pour tout entier $q \leq n$, l'ensemble Γ_q des signaux $X = (x_1, \dots, x_n)$ tels que $|T| \leq q$ où $|T|$ est le cardinal du support $T = \{j; x_j \neq 0\}$ de X . Cet ensemble Γ_q n'est évidemment pas un espace vectoriel.

Le théorème suivant de Candès et Tao éclairera notre discussion. Soit p un nombre premier et soit \mathbb{F}_p le corps fini à p éléments. La transformée de Fourier d'une fonction f à valeurs réelles ou complexes, définie sur \mathbb{F}_p est, elle-même, définie sur \mathbb{F}_p et est notée \hat{f} .

Théorème 1. Soit Ω une partie de \mathbb{F}_p . Alors toute fonction f définie sur \mathbb{F}_p est déterminée de façon unique par la restriction de sa transformée de Fourier à Ω lorsque le cardinal du support T de f vérifie la condition $|T| \leq \frac{1}{2}|\Omega|$. Inversement, si Ω n'est pas \mathbb{F}_p tout entier, il existe deux fonctions différentes f et g telles que (a) le cardinal du support de f ne dépasse pas $\frac{1}{2}|\Omega| + 1$, (b) la même condition soit satisfaite pour g , mais que (c) $\hat{f}(\omega) = \hat{g}(\omega)$, $\omega \in \Omega$.

Ce remarquable énoncé est faux lorsque \mathbb{F}_p est remplacé par $\mathbb{Z}/n\mathbb{Z}$ et que n n'est pas un nombre premier. Néanmoins il redevient vrai si Ω est un ensemble ayant des propriétés arithmétiques particulières, comme nous allons le voir dans un instant. Le théorème 1 apparaît aujourd'hui comme le point de départ des travaux ultérieurs de Candès et Tao. Mais l'histoire est différente. Le vrai point de départ fut un problème posé par la tomographie. Il s'agissait de reconstruire une image simple f à partir d'une information partielle sur sa transformée de Fourier. La seule donnée dont on dispose est la connaissance de $c(\omega) = \hat{f}(\omega)$, $\omega \in \Omega$. L'ensemble Ω se compose d'une centaine de lignes passant par l'origine. L'hypothèse a priori est que f est constante par morceaux, c'est-à-dire que f est égale à une constante c_j à l'intérieur d'un domaine D_j et que ces domaines sont en nombre fini. Les D_j ont des bords réguliers et f est discontinue sur ces bords. Alors Candès et Justin Romberg découvrirent expérimentalement que f est la solution du problème

$$\inf \|g\|_{BV}; \hat{g}(\omega) = c(\omega), \omega \in \Omega \tag{2}$$

Le gradient ∇f de f est porté par un ensemble fini de lignes. Il a donc un petit support et nous nous retrouvons dans un cadre qui ressemble à celui du théorème 1. On observera que $\|g\|_{BV} = \|\nabla g\|_1$ et l'on pourra alors considérer (2) comme un cas particulier du théorème 2 ci-dessous. Mais le théorème 1 ne disait rien sur l'algorithme permettant de calculer f et n'explique donc pas ce phénomène. En outre l'hypothèse que p est premier n'est pas satisfaite dans la version discrétisée du problème de tomographie. Voici un théorème plus général obtenu par Candès et Tao. Commençons par une définition. Considérons $q \leq p$ vecteurs Z_j , $1 \leq j \leq q$, appartenant à l'espace euclidien \mathbb{R}^p . Pour un $0 < \varepsilon < 1$, on dira que ces vecteurs sont ε -orthogonaux si, pour tout choix des coefficients c_j , on a

$$(1 - \varepsilon) \sum_1^q |c_j|^2 \leq \left\| \sum_1^q c_j Z_j \right\|^2 \leq (1 + \varepsilon) \sum_1^q |c_j|^2 \tag{3}$$

Une matrice A est dite (q, ε) incohérente si la condition (3) est vérifiée pour tout ensemble de q vecteurs distincts pris parmi les vecteurs colonnes de A . Pour A et q fixés, on désigne par ε_q la borne inférieure de ces valeurs de ε . On revient alors à (1) où $\Gamma = \Gamma_q$ est défini comme ci-dessus (le cardinal du support de $X \in \Gamma_q$ ne dépasse pas q) et où A est la matrice dont les lignes sont les vecteurs A_j , $1 \leq j \leq p$. On pose enfin $\|X\|_1 = |x_1| + \dots + |x_n|$, $X \in \mathbb{R}^n$.

Théorème 2. Avec les notations précédentes, on suppose que

$$\varepsilon_q + \varepsilon_{2q} + \varepsilon_{3q} < 1 \tag{4}$$

Alors le système (1) où $\Gamma = \Gamma_q a$, au plus, une solution. Si elle existe, cette solution s'obtient en résolvant le problème d'optimisation convexe

$$\inf \|X\|_1; y_j = \langle A_j, X \rangle, j = 1, \dots, p \quad (5)$$

On notera que la contrainte $X \in \Gamma_q$ n'apparaît plus dans (5). La solution de (5) vérifie automatiquement cette condition lorsqu'elle existe, c'est-à-dire lorsque $y_j = \langle A_j, Y \rangle$ pour $1 \leq j \leq p$ et pour un certain $Y \in \Gamma_q$. On pourrait penser que l'algorithme fourni par le théorème est instable. Il n'en est rien. Candès et Tao démontrent un remarquable théorème de robustesse en prouvant que le théorème 2 s'applique à des signaux bruités et fournit alors un algorithme de débruitage.

La principale difficulté est de construire des matrices A d'usage commode vérifiant (4). Les valeurs de q et de n nous sont imposées et l'on souhaite que p soit, à un terme logarithmique près, du même ordre de grandeur que q . Candès et Tao, d'une part, Alain Pajor d'autre part, montrent que des matrices aléatoires (dont les coefficients sont des gaussiennes indépendantes, centrées et de même loi) conviennent. Il est indispensable d'en savoir plus et ce domaine de recherche touche à la physique statistique et à la théorie des nombres. David Donoho et Alain Pajor ont relié les travaux de Candès et Tao à la géométrie des ensembles convexes en grande dimension. Le site web de Candès est <http://www.acm.caltech.edu/~emmanuel/>, celui de David Donoho est <http://www-stat.stanford.edu/~donoho/> et celui d'Alain Pajor est <http://perso-math.univ-mlv.fr/users/pajor.alain/>.

Références

- [1] E.J. Candès, J. Romberg and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59, (2005) 1207-1223.
- [2] E.J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51, (2004) 4203-4215.
- [3] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections : universal encoding strategies. *IEEE Trans. Inform. Theory*, 52, (2004) 5406-5425.
- [4] E.J. Candès, J. Romberg and T. Tao. Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52, (2004) 489-509.