
RACONTE-MOI ... LE *COMPRESSIVE SENSING*

par

Simon Foucart

Aux alentours de 2005, la communauté scientifique et technologique s'enflamma pour un nouveau sujet appelé *compressive sensing*⁽¹⁾, terme interchangeable avec *compressed sensing* et *compressive sampling*. En ce début de 2021, une recherche combinée de ces trois termes sur Google Scholar produit à peu près dix-sept mille résultats. Si cette profusion s'explique par un intérêt venu de nombreuses disciplines, il est bon de noter que le sujet possède bel et bien une origine mathématique dans les travaux fondateurs de Candès, Romberg, et Tao (par exemple [CRT06]) et de Donoho (par exemple [Don06]). Une quinzaine d'années après ces travaux, il est opportun de faire un point sur la théorie développée depuis. De nombreux lecteurs ont entendu vaguement parler du *compressive sensing*, y associant sans doute les mots-clés de parcimonie, minimisation ℓ_1 , et matrices aléatoires, peut-être réduisant le sujet à (un sous-ensemble de) ces trois notions. Je vais m'atteler à en donner une description plus large, mais forcément incomplète. Les lecteurs souhaitant approfondir trouveront plus de détails dans les ouvrages [CGLP12, EK12, FR13] ou dans l'article [Fou16].

Introduction

Dans cette section, nous présentons de manière informelle le problème au centre de la théorie du *compressive sensing*.

Motivation. — Le principe de base du *compressive sensing* consiste à effectuer la compression des données en même tant que leur acquisition (d'où le nom). Prenons l'exemple d'un appareil photo numérique, capturant tous les pixels au déclenchement pour immédiatement compresser l'image. C'est une forme de gâchis : la théorie du *compressive sensing* justifie que l'on puisse acquérir juste les données qu'il faut. Si le principe n'est pas implémenté industriellement dans ce cas précis, il l'a été en imagerie par résonance magnétique, brandie comme la *success story* du *compressive sensing*.

⁽¹⁾J'utiliserai ici le terme anglophone car une traduction acceptée de *compressive sensing* en Français ne semble pas exister, bien qu'"acquisition compressive" soit fidèle aux principes de base.

Le processus d'acquisition étant fastidieux (le patient ne doit pas bouger) et potentiellement dangereux, il est important de limiter le temps d'observation. Certains appareils, actuellement en service, réduisent ce temps par un facteur huit au moins. Cette réussite vient, encore une fois, du fait qu'une image est compressible, c'est-à-dire qu'elle peut être décrite de manière satisfaisante avec seulement une fraction de son contenu d'information. Par exemple, une image JPEG 2000, représentée par ses coefficients en ondelettes, est compressée en éliminant une grande partie de ces coefficients pour ne garder qu'un vecteur peuplé principalement de zéros—un vecteur parcimonieux.

Le problème standard. — Dans le problème ci-dessus comme dans d'autres, on s'intéresse à des objets qui peuvent être modélisés par des vecteurs $x \in \mathbb{K}^N$ de grande dimension, \mathbb{K} désignant \mathbb{R} ou \mathbb{C} . Les données sont acquises au travers d'observations linéaires, disons

$$(1) \quad y_i = \langle a_i, x \rangle, \quad i \in \{1, \dots, m\}.$$

Ces observations peuvent s'écrire sous la forme $y = Ax$, où a_1, \dots, a_m constituent les lignes d'une matrice $A \in \mathbb{K}^{m \times N}$ représentant le processus d'acquisition. Le nombre m d'observations est restreint à cause, par exemple, de leur coût élevé. Est-il possible de déduire les vecteurs $x \in \mathbb{K}^N$ à partir de leurs vecteurs $y = Ax \in \mathbb{K}^m$ d'observations quand $m \ll N$? Non, bien sûr : le système linéaire $Ax = y$ étant sous-déterminé, il possède une infinité de solutions. Mais si nous savons a priori que les vecteurs $x \in \mathbb{K}^N$ sont parcimonieux de niveau s , c'est-à-dire que

$$(2) \quad \|x\|_0 = |\text{support}(x)| \leq s, \quad \text{où} \quad \text{support}(x) = \{j \in \{1, \dots, N\} : x_j \neq 0\},$$

cette information va nous permettre d'accomplir une tâche qui semblait impossible. Bref, le problème considéré dans cet article est un problème d'algèbre (presque) linéaire qui devrait trouver sa place dans une formation scientifique moderne, à savoir :

$$(3) \quad \text{Résoudre le système linéaire sous-déterminé } y = Ax \text{ sachant que } \|x\|_0 \leq s.$$

Cela requiert évidemment $m \geq s$, même si le support de x était connu d'avance, ce qui n'est pas le cas. En théorie, le problème (3) peut être résolu en parcourant tous les $\binom{N}{s}$ supports de taille s , mais cette approche naïve est loin d'être pratique. Du point de vue de la complexité algorithmique, le problème (3) est même NP-difficile dans toute sa généralité. Cependant, le *compressive sensing* n'a pas pour ambition de résoudre ce problème dans toute sa généralité. En effet, un des piliers de la théorie—la possibilité de choisir des processus d'observation favorables—rend le problème abordable. En réalité, nous avons devant nous deux sous-problèmes intimement liés :

- (i) Comment choisir une matrice $A \in \mathbb{K}^{m \times N}$ de sorte que l'application $x \mapsto Ax$ soit, au moins, injective sur l'ensemble

$$\Sigma_s^N := \{x \in \mathbb{K}^N : \|x\|_0 \leq s\}$$

des vecteurs parcimonieux de niveau s ?

- (ii) Etant donné une telle matrice A , comment reconstruire de manière efficace les vecteurs $x \in \Sigma_s^N$ à partir des observations $y = Ax \in \mathbb{K}^m$?

Le processus d'observation

Dans cette section, nous débutons une discussion détaillée à propos du sous-problème (i), quelque peu étendu.

Nombre minimal d'observations. — En fixant les paramètres $s \ll N$, nous nous interrogeons sur le nombre minimal m_{inj} d'observations qui permet de trouver une matrice $A \in \mathbb{K}^{m \times N}$ injective sur Σ_s^N . Il n'est pas difficile de voir que cette injectivité est équivalente au fait que $\Sigma_{2s}^N \cap \ker(A) = \{0\}$, ou encore au fait que $2s$ colonnes de A soient toujours linéairement indépendantes. Cette dernière condition implique directement que $m \geq 2s$. De plus, on peut mettre en évidence des matrices A de taille $2s \times N$ vérifiant cette condition, par exemple en sélectionnant $2s$ lignes d'une matrice totalement positive de taille $N \times N$. De ce fait, nous avons

$$(4) \quad m_{\text{inj}} = 2s.$$

C'est assez intuitif : localiser le support nécessite s unités d'information et déterminer les valeurs non nulles en nécessite s autres. Par ailleurs, si les observations sont prises comme les $m_{\text{inj}} = 2s$ premiers coefficients de Fourier discrets, nous avons même à notre disposition un algorithme de reconstruction, inspiré de la méthode bi-séculaire de Prony [Pro95], qui paraît tout à fait convenable au premier abord. Pour autant, le problème standard n'est pas complètement résolu...

Stabilité et robustesse. — La méthode de Prony est numériquement réalisable, mais elle dégénère quand de petites perturbations sont introduites. Nous voulons des processus d'observation permettant une reconstruction qui soit à la fois stable quand les vecteurs $x \in \mathbb{K}^N$ ne sont pas exactement parcimonieux et robuste quand les observations $y = Ax + e \in \mathbb{K}^m$ contiennent des erreurs $e \in \mathbb{K}^m$ non nulles. Pour simplifier, nous ne discuterons ci-après que de la stabilité. Dans cette optique, nous sommes à la recherche de matrices $A \in \mathbb{K}^{m \times N}$ pour lesquelles il existe des processus de reconstruction $\Delta : \mathbb{K}^m \rightarrow \mathbb{K}^N$ tels que, pour tout $x \in \mathbb{K}^N$,

$$(5) \quad \|x - \Delta(Ax)\|_p \leq \frac{C}{s^{1-1/p}} \min_{z \in \Sigma_s^N} \|x - z\|_1.$$

L'indice p se trouve dans $[1, 2]$ et nous n'élucidons pas la présence de $s^{1-1/p}$ ni le choix de la norme ℓ_1 dans la partie droite de (5). Notons simplement que si x est parcimonieux de niveau s , ce terme est nul, donc $\Delta(Ax) = x$, ce qui signifie que x est parfaitement reconstruit à partir de Ax . La condition d'existence de $\Delta : \mathbb{K}^m \rightarrow \mathbb{K}^N$ permettant (5) ne dépend en fait que du noyau de la matrice A . Elle est en effet équivalente à

$$(6) \quad \forall v \in \ker(A), \forall S \subseteq \{1, \dots, N\} \text{ avec } |S| \leq 2s, \quad \|v_S\|_p \leq \frac{C'}{s^{1-1/p}} \|v_{S^c}\|_1,$$

où la notation $v_T \in \mathbb{K}^N$ est utilisée pour le vecteur défini par $(v_T)_j = v_j$ si $j \in T$ et $(v_T)_j = 0$ si $j \in T^c$. Partant de (6), nous pouvons ensuite établir que le nombre

minimal d'observations rendant possible une acquisition/reconstruction stable au sens de (5) est quasi linéaire en s , c'à d

$$(7) \quad m_{\text{sta}} \asymp s \ln(N/s).$$

Le premier argument montrant que $m_{\text{sta}} \geq C'' s \ln(N/s)$ reposait sur l'estimation de l'épaisseur de Gelfand de la boule ℓ_1 dans ℓ_p^N quand $p > 1$. Un argument ultérieur couvrant aussi le cas $p = 1$ évite ce détour, tout en offrant une explication plus intuitive, basée sur le *compressive sensing*, de l'estimation de cette épaisseur de Gelfand. La justification de $m_{\text{sta}} \leq C''' s \ln(N/s)$ découle de l'existence de matrices $A \in \mathbb{K}^{m \times N}$ ayant une propriété d'isométrie restreinte (voire ci-après) dans le régime de paramètres $m \asymp s \ln(N/s)$ et de la stabilité (et robustesse) des algorithmes de reconstruction que nous allons maintenant discuter. Juste avant ça, mentionnons que les observations (1) considérées jusqu'ici sont non adaptatives, dans le sens où le choix de a_i ne dépend pas de $a_1, y_1, \dots, a_{i-1}, y_{i-1}$. Autoriser des observations adaptatives ne réduirait toutefois pas l'estimation (7) du nombre d'observations garantissant la stabilité.

Le processus de reconstruction

Dans cette section, nous focalisons sur le sous-problème (ii) : en fixant la matrice $A \in \mathbb{K}^{m \times N}$, nous exposons l'intuition derrière quelques algorithmes de reconstruction fréquemment utilisés.

Minimisation ℓ_1 . — Formellement, le problème (3) se résout en minimisant le niveau de parcimonie $\|z\|_0$ d'une variable vectorielle $z \in \mathbb{K}^N$ qui doit satisfaire $Az = y$. Comme ce problème d'optimisation n'est pas abordable en pratique, on peut penser à remplacer $\|z\|_0$ par une (quasi)-norme $\|z\|_q$ pour $q > 0$, puisque $\|z\|_q^q$ tend vers $\|z\|_0$ quand q tend vers zéro. Un tel problème étant non convexe tant que $q < 1$, nous sommes amenés à minimiser la norme ℓ_1 de z , laquelle peut être interprétée comme la relaxation convexe de $\|z\|_0$. Plus précisément, la très populaire minimisation ℓ_1 , aussi appelée *basis pursuit*, consiste à associer à $y \in \mathbb{K}^m$ une solution du problème

$$(8) \quad \underset{z \in \mathbb{K}^N}{\text{minimiser}} \sum_{j=1}^N |z_j| \quad \text{sous la contrainte } Az = y.$$

Le succès de cette stratégie se schématise souvent par la Figure 1. Cette figure est quelque peu trompeuse, car tous les $x \in \mathbb{K}^N$ parcimonieux de niveau 1 ne peuvent pas être reconstruits de la sorte à partir de $y = Ax$ —il faudrait que l'espace affine soit une droite plutôt qu'un plan. Néanmoins, l'interprétation géométrique est valide en dimension supérieure, la boule ℓ_1 devenant de plus en plus pointue. Pour $\mathbb{K} = \mathbb{R}$, une autre explication intuitive du succès de la minimisation ℓ_1 provient du fait que le problème (8) ait toujours une solution parcimonieuse de niveau au plus m . En ayant l'algorithme du simplexe en tête, cela peut se justifier en remarquant que (8) est équivalent au problème d'optimisation linéaire écrit sous la forme standard

$$(9) \quad \underset{z^+, z^- \in \mathbb{R}^N}{\text{minimiser}} \sum_{j=1}^N (z_j^+ + z_j^-) \quad \text{sous les contraintes } A(z^+ - z^-) = y, \quad z^+ \geq 0, \quad z^- \geq 0.$$

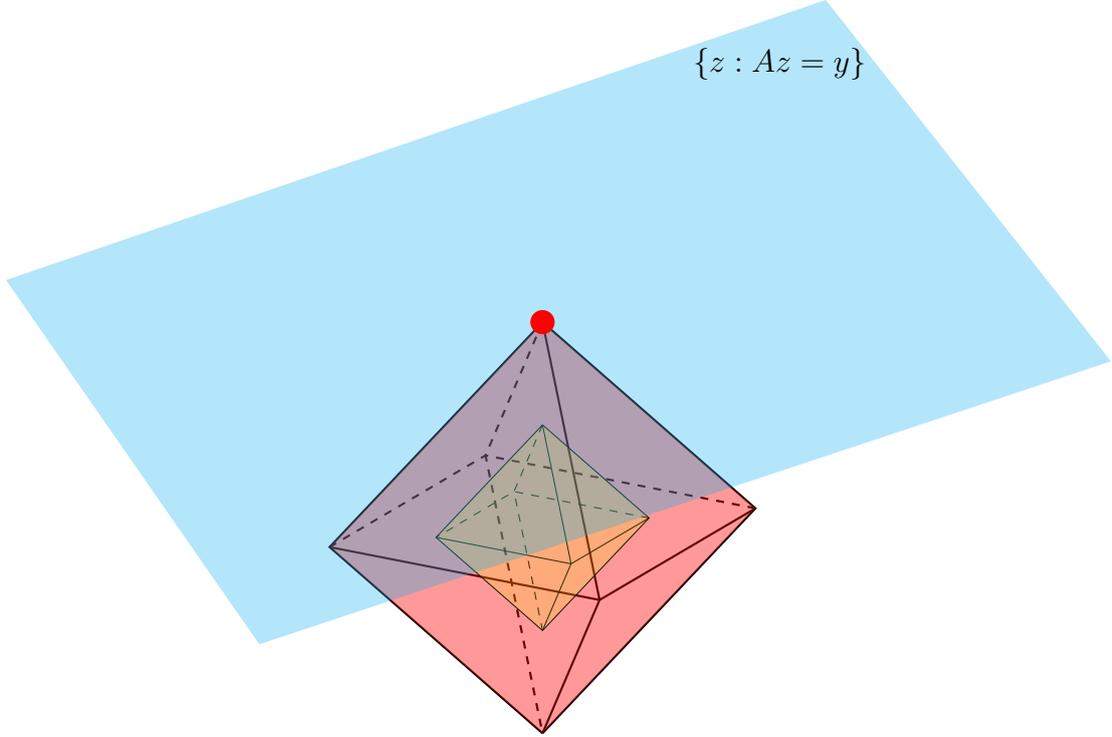


FIGURE 1. La boule ℓ_1 de plus petit rayon qui intersecte l'espace affine $\{z \in \mathbb{K}^N : Az = y\}$ la touche en un vecteur parcimonieux, ici de niveau 1.

Il y a une condition nécessaire et suffisante pour que la minimisation ℓ_1 puisse reconstruire tous les vecteurs parcimonieux $x \in \mathbb{K}^N$ de niveau au plus s à partir de $y = Ax$. Cette condition dépend encore une fois uniquement du noyau de A et s'appelle d'ailleurs la *null space property* d'ordre s . Elle s'énonce comme suit :

$$(10) \quad \forall v \in \ker(A), \forall S \subseteq \{1, \dots, N\} \text{ avec } |S| \leq s, \quad \|v_S\|_1 \leq \|v_{S^c}\|_1.$$

Cette condition est clairement nécessaire, puisque $Av = 0$ donne $A(v_S) = A(-v_{S^c})$, donc $\|v_S\|_1 \leq \|v_{S^c}\|_1$ si un vecteur parcimonieux possède toujours la norme ℓ_1 la plus petite parmi les vecteurs partageant les mêmes observations. Il n'est pas trop difficile de voir qu'elle est aussi suffisante. En fait, un renforcement de la condition (10), à savoir

$$(11) \quad \forall v \in \ker(A), \forall S \subseteq \{1, \dots, N\} \text{ avec } |S| \leq s, \quad \|v_S\|_p \leq \frac{\rho}{s^{1-1/p}} \|v_{S^c}\|_1$$

avec une constante $\rho < 1$ est suffisante pour garantir la stabilité de la minimisation ℓ_1 au sens de (5). L'analogie entre (11) et (6)—il n'y a que la taille de S et les constantes qui changent—suggère une certaine universalité de la minimisation ℓ_1 . Précisément, s'il existe un processus de reconstruction parcimonieuse stable à un niveau s , alors la minimisation ℓ_1 sera également stable à un niveau cs .

Algorithmes itératifs. — Si le côté boîte-noire de la stratégie de minimisation ℓ_1 déplaît, on peut se tourner vers des algorithmes de reconstruction plus explicites et pouvant présenter certains avantages, notamment en terme de vitesse. Un premier exemple est l'*Orthogonal Matching Pursuit*. Cet algorithme cherche à reconstruire le support du vecteur parcimonieux en créant un nouvel indice à chaque itération et en produisant, étant donné le support courant, le vecteur qui s'accorde le plus avec les observations disponibles. Commencant avec $S^0 = \emptyset$ et $x^0 = 0$, l'algorithme s'écrit

$$\begin{aligned} S^k &= S^{k-1} \cup \{j^k\}, & j^k &= \operatorname{argmax}_{j \in \{1, \dots, N\}} |(A^*(y - Ax^{k-1}))_j|, \\ x^k &= \operatorname{argmin}_{z \in \mathbb{K}^N} \|y - Az\|_2 & & \text{sous la contrainte } \operatorname{support}(z) \subseteq S^k. \end{aligned}$$

Dans la première étape, choisir l'indice j^k correspondant à la plus grande composante (en valeur absolue) de la matrice adjointe $A^* = \bar{A}^\top$ multipliant le résidu permet une réduction presque optimale de $\|y - Ax^k\|_2^2$ à chaque itération; dans la seconde étape, le problème des moindres carrés se résume à la résolution d'un système linéaire de taille $k \times k$.

Un deuxième exemple est l'*Iterative Hard Thresholding*. Il imite une méthode itérative visant à résoudre le système linéaire carré $A^*Ax = A^*y$, mais incorpore une étape de seuillage parcimonieux à chaque itération. Partant par exemple de $x^0 = 0$, l'algorithme s'écrit

$$(12) \quad x^{k+1} = H_s(x^k + A^*(y - Ax^k)),$$

où l'opérateur H_s conserve les s plus grandes composantes (en valeur absolue) et remplace les autres par des zéros. Le fait que l'algorithme n'implique aucune résolution de système linéaire est assurément un avantage. Mais si de telles résolutions ne sont pas vues comme des obstacles, comme dans *Orthogonal Matching Pursuit*, d'autres avantages apparaissent en considérant une variante appelée *Hard Thresholding Pursuit*. Commencant avec $S^0 = \emptyset$ et $x^0 = 0$, cette variante s'écrit

$$\begin{aligned} S^k &= \{\text{indices des } s \text{ plus grandes composantes de } |A^*(y - Ax^{k-1})|\}, \\ x^k &= \operatorname{argmin}_{z \in \mathbb{K}^N} \|y - Az\|_2 \quad \text{sous la contrainte } \operatorname{support}(z) \subseteq S^k. \end{aligned}$$

Succès des algorithmes de reconstruction

Dans cette section, nous retournons au sous-problème (i) et nous introduisons des conditions sur la matrice $A \in \mathbb{K}^{m \times N}$ qui garantissent le bon fonctionnement des stratégies qui viennent d'être présentées.

Cohérence et propriété d'isométrie restreinte. — La notion de cohérence $\mu(A)$ d'une matrice $A \in \mathbb{K}^{m \times N}$ formalise le fait que ses colonnes sont deux à deux presque orthogonales. Pour tous les algorithmes ci-dessus, on peut montrer qu'une cohérence $\mu(A) < c/s$ garantit le succès de la reconstruction parcimonieuse de niveau s . Cependant, la cohérence ne peut pas être aussi petite que désirée : la borne de Welch stipule que $\mu(A) \geq c'/\sqrt{m}$ quand $N \geq Cm$. Dès lors, les conditions basées sur la cohérence

ne s'appliquent que dans le régime $m \geq c''s^2$, bien loin du régime recherché où m est quasi linéaire en s .

La percée est venue de l'introduction dans [CT05] de la notion de propriété d'isométrie restreinte. Celle-ci formalise le fait que tout jeu de s colonnes de A forme un système presque orthogonal. Plus rigoureusement, on dit que la matrice A satisfait la propriété d'isométrie restreinte d'ordre s avec constante $\delta > 0$ si

$$(13) \quad (1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta)\|z\|_2^2 \quad \text{pour tout } z \in \Sigma_s^N,$$

et on écrit $\delta_s(A)$ pour le plus petit $\delta > 0$ telle que (13) soit valide. Pour tous les algorithmes ci-dessus, on peut montrer qu'une condition du type $\delta_{\kappa s}(A) < \delta_*$ garantit le succès de la reconstruction parcimonieuse de niveau s . Pour *Iterative Hard Thresholding* et *Hard Thresholding Pursuit*, $\kappa = 3$ et $\delta_* = 1/\sqrt{3}$ conviennent. Pour *Orthogonal Matching Pursuit*, $\kappa = 13$ et $\delta_* = 1/6$ conviennent (ce qui est sans doute améliorable). Étonnamment, il faut effectuer plus de s itérations afin de reconstruire les vecteurs parcimonieux de niveau s . Néanmoins, le nombre d'itérations est au plus proportionnel à s , ce qui est aussi le cas pour *Hard Thresholding Pursuit*. Pour *Basis Pursuit*, $\kappa = 2$ et $\delta_* = 1/\sqrt{2}$ conviennent, ce dernier choix ne pouvant pas être amélioré. D'autres choix de κ et δ_* permettent de déduire la version stable (11) de la *null space property* d'ordre s avec un argument assez élémentaire, mais juste un peu trop long pour être inclus ici. L'argument aurait cependant été instructif pour s'apercevoir que la norme choisie sur Az dans (13) ne joue presque aucun rôle.

Matrices avec la propriété d'isométrie restreinte. — Pour boucler la boucle, il nous faut mettre en avant des matrices $A \in \mathbb{K}^{m \times N}$ qui satisfont la propriété d'isométrie restreinte d'ordre s avec $m \asymp s \ln(N/s)$. Ceci se fait via des arguments probabilistes : considérant un certain type de matrices aléatoires $A \in \mathbb{K}^{m \times N}$, on veut montrer que (13) est valide avec une faible probabilité d'échec dès que m dépasse un certain seuil. Par exemple, si A est peuplée de variables indépendantes gaussiennes de moyenne nulle et de variance $1/m$, alors la probabilité d'échec est $\exp(-cm)$ et le seuil est $s \ln(N/s)$, comme espéré. Le résultat s'étend aux variables sous-gaussiennes, dont les variables de Rademacher. Pour les variables sous-exponentielles, dont les variables de Laplace, la propriété d'isométrie restreinte (13) n'est pas valide dans le régime $m \asymp s \ln(N/s)$, mais la *null space property* l'est néanmoins, même dans sa version stable (11) avec $p \in [1, 2]$.

Les arguments sont plus subtils si l'on considère des observations de Fourier, que l'on rencontre notamment en imagerie par résonance magnétique. Les sous-matrices aléatoires de taille $m \times N$ d'une matrice de Fourier discrète, construites en sélectionnant m lignes au hasard, satisfont bien la propriété d'isométrie restreinte d'ordre s , mais la probabilité d'échec est N^{-c} et le seuil est $s \ln^3(N)$. Déterminer s'il peut être réduit à $s \ln(N)$ est un problème ouvert.

Malheureusement, bien que presque toutes les matrices de taille $m \times N$ satisfassent la propriété d'isométrie restreinte d'ordre s dans le régime optimal $m \asymp s \ln(N/s)$, personne n'est capable à l'heure actuelle d'en construire une de façon déterministe. Découvrir une telle construction est le grand défi auquel la théorie du *compressive sensing* doit faire face. Même le régime $m \asymp s^\alpha$ avec $\alpha < 2$ semble pour l'instant

hors de portée—à l’exception de [BDF⁺11], qui donne un exposant α extrêmement proche de 2 pour certaines sous-matrices de Fourier presque carrées. Si l’on est uniquement intéressé par la reconstruction parcimonieuse de niveau s , il y a cependant de meilleures nouvelles : pour tout $\alpha > 1$, les graphes bipartis expandeurs donnés explicitement dans [GUV09] fournissent des matrices d’adjacence $A \in \{0, 1\}^{m \times N}$ qui satisfont la version stable (11) de la *null space property* dans le régime $m \asymp s^\alpha$. L’inconvénient est de devoir imposer $p = 1$ sans parvenir à couvrir le domaine $p \in [1, 2]$ en entier.

Quelques extensions

En guise de conclusion, nous évoquons brièvement quatre variations du problème standard, variations toujours conformes au principe de base consistant à effectuer l’acquisition et la compression des données simultanément.

Représentations parcimonieuses par rapport à un dictionnaire. — Il n’est pas rare que l’étape de modélisation nous amène à considérer des vecteurs de la forme $f = Dx \in \mathbb{K}^n$ pour un dictionnaire $D \in \mathbb{K}^{n \times N}$, $n < N$. En supposant sans perdre beaucoup de généralité que $DD^* = I_n$, la parcimonie de f par rapport à D peut être comprise de deux façons : soit un des vecteurs $x \in \mathbb{K}^N$ vérifiant $Dx = f$ est parcimonieux, soit le vecteur $D^*f \in \mathbb{K}^N$ (qui vérifie $D(D^*f) = f$) est parcimonieux. Dans le premier cas, on parle de parcimonie en synthèse, et dans le deuxième cas de parcimonie en analyse. La théorie n’est pas complètement satisfaisante pour des dictionnaires arbitraires car chacun de ces deux scénarios présente des inconvénients à surmonter.

Acquisition compressive de matrices de rang faible. — Hormis le fait que les vecteurs $x \in \mathbb{K}^N$ de niveau de parcimonie s sont supplantés par des matrices $X \in \mathbb{K}^{N \times N}$ de rang r , la situation est extrêmement similaire au problème standard. L’analogie découle du fait que le rang de X n’est autre que $\|\sigma(X)\|_0$, où $\sigma(X)$ dénote le vecteur des valeurs singulières de X . En particulier, la minimisation de $\|z\|_1$ peut être remplacée par la minimisation de $\|Z\|_* = \sum_{j=1}^N \sigma_j(Z)$, qu’on appelle la norme nucléaire. En pratique, cette minimisation s’effectue via programmation semi-définie. La théorie, qui peut également se construire autour d’une propriété d’isométrie restreinte, met en évidence que le nombre minimal d’observations pour la reconstruction stable de matrices $X \in \mathbb{K}^{N \times N}$ de rang au plus r est $m_{\text{sta}} \asymp rN$. Il s’agit essentiellement du nombre de degrés de liberté d’une matrice de taille $N \times N$ et de rang r , sans facteur logarithmique supplémentaire.

Acquisition sans-phase compressive. — Nous retournons dans le monde vectoriel et supposons que les vecteurs $x \in \mathbb{C}^N$ sont acquis au travers d’observations non linéaires de la forme $y_i = |\langle a_i, x \rangle|$, $i \in \{1, \dots, m\}$, càd que l’amplitude des $\langle a_i, x \rangle$ est conservée mais pas leur phase. Si ces vecteurs sont parcimonieux de niveau s , il est encore une fois possible de les reconstruire de manière stable en utilisant $m \asymp s \ln(N/s)$ observations. Sans parcimonie, le problème peut être traité en remarquant que $y_i^2 = \text{tr}(a_i a_i^* X)$, $X = xx^*$, ce qui permet de reconstruire la matrice $X \in \mathbb{C}^{N \times N}$ de rang 1, puis d’en déduire $x \in \mathbb{C}^N$.

One-bit compressive sensing. — Ce problème est en quelque sorte complémentaire du précédent : au lieu de conserver l’amplitude d’observations linéaires $\langle a_i, x \rangle$, celle-ci est perdue. En nous plaçant dans le cas réel, les vecteurs $x \in \mathbb{R}^N$ parcimonieux de niveau s sont donc acquis au travers de $y_i = \text{sgn}(\langle a_i, x \rangle) \in \{-1, +1\}$, $i \in \{1, \dots, m\}$. Ceci représente un cas extrême et un peu artificiel de quantification. Puisque les observations binaires y_1, \dots, y_m ne dépendent que de la direction de x , nous supposons à présent que $\|x\|_2 = 1$. La reconstruction exacte de x n’étant bien entendu plus envisageable, nous souhaitons être capables d’utiliser $y \in \{-1, +1\}^m$ afin de produire algorithmiquement des vecteurs $\hat{x} \in \mathbb{R}^N$ pour lesquels l’erreur de reconstruction ait la forme $\|x - \hat{x}\|_2 \leq C[(s \ln(N/s))/m]^\gamma$. Il est possible d’atteindre l’exposant $\gamma = 1/2$, tandis que des arguments plus simples présentés dans [Fou16] permettent d’obtenir $\gamma = 1/4$. Ces arguments s’appuient sur une variation de la propriété d’isométrie restreinte (13) dans laquelle la norme sur Az est une norme ℓ_1 .

Références

- [BDF⁺11] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Mathematical Journal*, 159(1):145–185, 2011.
- [CGLP12] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing, random matrices, and high dimensional geometry*. Société Mathématique de France, 2012.
- [CRT06] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [CT05] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.
- [Don06] D. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.
- [EK12] Y. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge University Press, 2012.
- [Fou16] S. Foucart. Flavors of compressive sensing. In *Approximation Theory XV: San Antonio 2016*, volume 201, pages 61–104. Springer Proceedings in Mathematics & Statistics, 2016.
- [FR13] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Birkhäuser, 2013.
- [GUV09] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM*, 56(4):1–34, 2009.
- [Pro95] R. Prony. Essai expérimental et analytique sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l’eau et de la vapeur de l’alkool, à différentes températures. *J. École Polytechnique*, 1:24–76, 1795.