

Les mathématiques embarquées dans la mission Herschel.

Yves Meyer

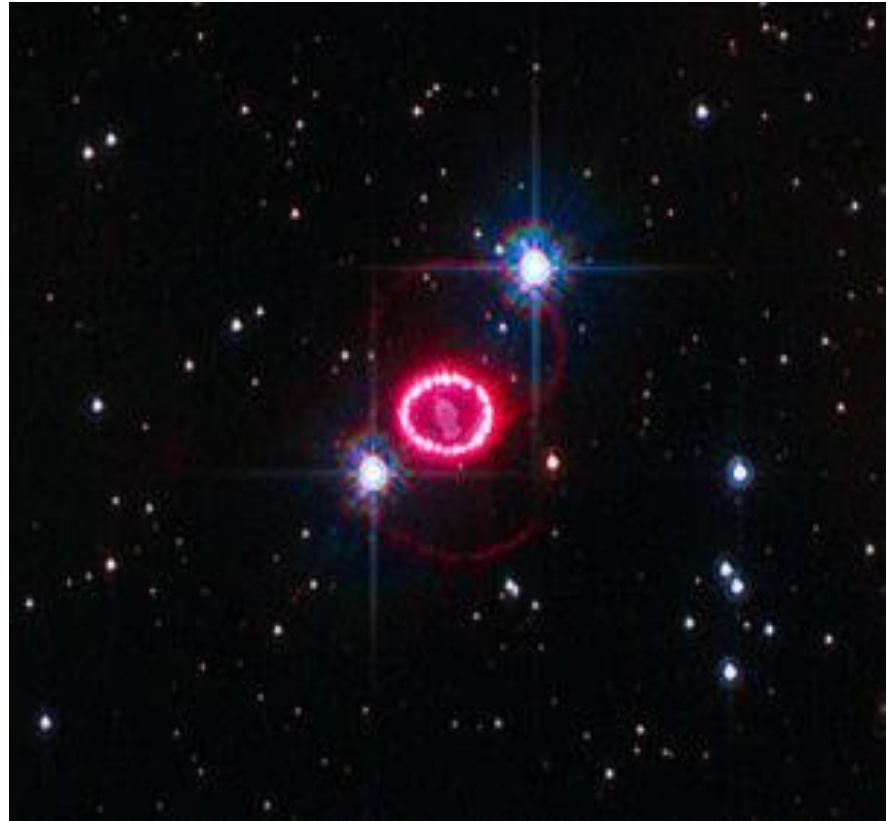
Grenoble, le 25 octobre 2011

**The Nobel Prize in Physics 2011 : Saul Perlmutter,
Brian P. Schmidt and Adam G. Riess**



"for the discovery of the accelerating expansion of the Universe through observations of distant supernovae".

SN1987A



SN 1987A was a supernova in the outskirts of the Tarantula Nebula in the Large Magellanic Cloud, a nearby dwarf galaxy.

It occurred approximately 51.4 kiloparsecs from Earth, approximately 179,000 light-years, close enough that it was visible to the naked eye. It could be seen from the Southern Hemisphere.

It was the closest observed supernova since SN 1604, which occurred in the Milky Way itself. The light from the supernova reached Earth on February 23, 1987. As it was the first supernova discovered in 1987, it was labeled "1987A".

It was discovered by Ian Shelton and Oscar Duhalde at the Las Campanas Observatory in Chile on February 24, 1987.

Plan de l'exposé

- (1) Les missions du télescope spatial Herschel.
- (2) Les plaques CCD (Charge-coupled Device)
- (3) La géométrie en très grande dimension.
- (4) Le compressed sensing.

1. LE TÉLESCOPE SPATIAL HERSCHEL

Le télescope spatial Herschel a été lancé de Kourou le 14 mai 2009 par une fusée Ariane 5. Herschel est désormais en position d'équilibre stable au point L2 de Lagrange à 1 million cinq cents mille kilomètres de la Terre.



Herschel possède le plus grand miroir construit à ce jour pour l'astronomie spatiale. Le télescope spatial Herschel est composé d'un miroir primaire de 3,5 m de diamètre et d'un miroir secondaire. Herschel, en tant que satellite, mesure 7,5 m de haut pour une section de 4x4 m. Il pèse 3,3 tonnes.

Trois instruments sont embarqués à son bord: (a) HIFI, un spectromètre à haute résolution dédié à l'étude de la chimie de l'Univers, (b) PACS, une caméra de bolomètres pour cartographier l'émission infrarouge des grains de poussière et (c) SPIRE, qui remplit les mêmes fonctions que PACS mais à de plus grandes longueurs d'onde, dans **l'infrarouge submillimétrique**.





Comment les galaxies se forment-elles et évoluent-elles?
Comment les étoiles se forment-elles et évoluent-elles en relation avec le milieu interstellaire des galaxies?

Dans les phases précoces de leur formation et de leur vie, les étoiles et les planètes sont dissimulées à l'intérieur ou derrière des nuages de gaz et de poussière qui sont à l'origine de leur constitution. Observer dans l'infrarouge permettra de sonder ces nuages de poussière. L'astronomie dans l'infrarouge lointain est une science encore jeune. Un télescope infrarouge situé dans l'espace permet l'étude des systèmes solaires en cours de formation ou des processus physiques conduisant à la naissance des étoiles et des galaxies et ceci sans les restrictions imposées par l'atmosphère terrestre. C'est pourquoi l'Agence Spatiale Européenne a décidé de construire l'observatoire spatial Herschel. Herschel étudiera également la chimie du milieu interstellaire des galaxies, des comètes et du système solaire.

Herschel a ouvert une nouvelle fenêtre d'observation sur l'univers. Il va désormais s'attaquer aux mystères de la naissance des étoiles et de l'évolution de la vie des galaxies.



Cette région sombre et froide est parsemée d'étoiles en formation comme des perles sur un fil cosmique.

2. LES PLAQUES CCD.

Les plaques photographiques “argentiques” n’enregistraient que de deux à trois pour cent de la lumière. Les plaques CCD enregistrent plus de 80% de la lumière les atteignant. L’utilisation de capteurs numériques (plaques CCD) a révolutionné l’astronomie.

En 1905, Einstein a démontré que lorsque la lumière frappe certains matériaux, des électrons sont produits. Plus il y a de lumière, plus nombreux sont les électrons produits. C’est l’effet photoélectrique d’Einstein.

Les plaques CCD utilisent cette découverte d’Einstein. Elles ont été inventées en 1969 par Willard Boyle et George E. Smith aux laboratoires Bell AT&T. Imaginez la surface d’une plaquette CCD comme une feuille de papier quadrillé. Chaque carré est sensible à la lumière, un peu comme le photomètre d’une caméra. Si l’on compte le nombre d’électrons produits dans chaque carré, on peut en déduire le nombre de photons qui sont tombés sur chaque carré.

La nature numérique des CCD ouvre la voie au traitement numérique de l’image. En 2009 le prix Nobel de physique récompensa Willard Boyle et George E. Smith pour l’invention des plaques CCD.

Jean-Luc Starck, (Astrophysique, CEA), élève d'Albert Bijaoui, est l'un des pionniers de l'utilisation des représentations creuses en Astronomie et en Astrophysique. E. Candès, D. Donoho et J-L. Starck ont développé et amélioré les algorithmes de compressed sensing.

Le compressed sensing permet d'adapter le nombre de mesures à faire à la complexité du signal ou de l'image. Nous reviendrons sur la définition de la complexité.

Le compressed sensing est l'une des deux méthodes utilisées pour transmettre les images enregistrées par Herschel et c'est la méthode la plus économique. Herschel est situé (en orbite stationnaire sur le second point de Lagrange du système Terre-Soleil) à 1 million cinq cents mille kilomètres de la Terre et la transmission de chaque bit d'information sur une telle distance consomme beaucoup d'énergie.

Il ne faut pas gaspiller les bits. Il faut transmettre l'information en envoyant des "messages les plus concis possible, compte tenu de la précision exigée".

3. LA GÉOMÉTRIE EN TRÈS GRANDE DIMENSION.

Une image numérique est une gigantesque matrice dont les coefficients sont les pixels. Votre caméra digitale vous propose des images comportant $N =$ plusieurs millions de pixels. Une image est donc un vecteur $X \in \mathbb{R}^N$. Les coordonnées $x(j)$ de l'image X sont l'éclairement au pixel j . Dans une image en couleurs, $x(j)$ est un vecteur à trois coordonnées qui correspondent aux "couleurs primaires" (rouge, vert et bleu). Une telle image numérique contient autant de bits que les 1118 pages des Essais de Montaigne. Or il nous suffit d'un dixième de seconde pour "voir" une image tandis que plusieurs mois sont nécessaires pour lire les "Essais". Quel est le truc? Le truc vient de ce que les images ont une structure beaucoup plus simple qu'il ne paraît et que notre système visuel est "programmé" pour reconnaître ces structures en un temps extrêmement court.

L'objet du "compressed sensing" est d'enregistrer une image numérique à la même vitesse que l'œil met à la voir. Ceci n'est possible que si l'image possède une structure interne, mais il n'est pas nécessaire pour appliquer le compressed sensing de connaître cette structure dans tous ses détails.

Dans les travaux de J-L. Starck, la structure interne des images de l'astronomie implique la parcimonie de leur décomposition dans la base des ondelettes.

Toute une partie des mathématiques contemporaines consiste à comprendre la géométrie en grande dimension finie. Ce nouveau chapitre des mathématiques s'appelle “concentration de la mesure” et a été initié par Vitali Milman. Voici un premier énoncé pour vous faire rêver. Considérons la boule unité $B_N \subset \mathbb{R}^N$ en très grande dimension. Alors l'intérieur de B_N est essentiellement vide. La masse a tendance à être concentrée près de la sphère unité S^{N-1} . De même presque toute la surface de cette sphère S^{N-1} est concentrée près de l'équateur. Plus généralement une fonction 1-lipschitzienne sur S^{N-1} est essentiellement une fonction constante. Un second énoncé est le théorème de Johnson et Lindenstrauss (1984).

Théorème 3.1. *Soit un ensemble fini V composé de N points situés dans un espace vectoriel euclidien E . Pour tout $\varepsilon \in (0, 1)$, il existe un ensemble W situé dans un espace de “petite” dimension $n = C(\varepsilon) \log N$ et une application bijective $f : V \mapsto W$ qui est ε -proche d'une isométrie entre V et W .*

La dimension de E ne joue aucun rôle et E peut aussi bien être un espace de Hilbert.

Il existe évidemment un plongement isométrique de V dans l'espace affine de dimension $N - 1$ engendré par V . Le théorème de Johnson et Lindenstrauss permet donc de réduire $N - 1$ à $C(\varepsilon) \log N$.

$$\text{On a } C(\varepsilon) \leq \frac{4}{(\varepsilon^2/2 - \varepsilon^3/3)}.$$

La condition sur f s'écrit

$$(1 - \varepsilon)|x - y|^2 \leq |f(x) - f(y)|^2 \leq (1 + \varepsilon)|x - y|^2$$

pour tout $x \in V$ et tout $y \in V$.

Le théorème de Johnson et Lindenstrauss n'est intéressant que si

$$N \geq C\varepsilon^{-2} \log(1/\varepsilon)$$

Ce théorème nous apprend que la dimension effective d'un ensemble de N points est $C \log N$ ce qui est tout-à-fait surprenant.

Prenons, par exemple, $V = \{0, e_1, \dots, e_N\}$ où e_j , $1 \leq j \leq N$, est la base canonique de \mathbb{R}^N .

Les distances entre deux points de V valent 1 ou $\sqrt{2}$. En appliquant le théorème de Johnson et Lindenstrauss on peut transformer V de façon presque isométrique en un ensemble W situé dans un espace de dimension $n = C(\varepsilon) \log N$. Les distances entre les points $f_j = f(e_j)$ seront alors comprises entre $(1 - \varepsilon)\sqrt{2}$ et $(1 + \varepsilon)\sqrt{2}$, tandis que les distances de ces mêmes points à $f(0)$ seront comprises entre $1 - \varepsilon$ et $1 + \varepsilon$. On en déduit en remplaçant $f(0)$ par 0 et en renormalisant les distances entre les points f_j et 0, l'existence, pour tout $\varepsilon > 0$, de $N = \exp(\gamma(\varepsilon)n)$ points z_j , $1 \leq j \leq N$, situés sur la sphère unité $S^n \subset \mathbb{R}^n$ et dont les distances mutuelles $|z_j - z_k|$ sont comprises entre $(1 - \varepsilon)\sqrt{2}$ et $(1 + \varepsilon)\sqrt{2}$.

Nous avons pour $1 \leq j, k \leq N$,

$$|z_j \cdot z_k| \leq 2\varepsilon + \varepsilon^2$$

Les vecteurs unitaires z_j sont donc presque orthogonaux bien que leur nombre N soit fabuleusement grand:

$$N \sim \exp(\gamma(\varepsilon)n), \quad \gamma(\varepsilon) = \frac{(\varepsilon^2/2 - \varepsilon^3/3)}{4}.$$

Ce résultat se situe dans une asymptotique où ε est fixé ce qui permet à N d'être exponentiellement grand devant n . En effet si ε est suffisamment petit pour que $2\varepsilon + \varepsilon^2 < 1/(N - 1)$, tout s'écroule, car les z_j , $1 \leq j \leq N$, sont nécessairement indépendants et l'on a donc $N \leq n$.

On peut essayer de prouver “à la main” ce corollaire du théorème de Johnson et Lindenstrauss. On considère les 2^n vecteurs $z_j = n^{-1/2}(\pm 1, \dots, \pm 1)$ de la sphère S^n . Mais ces vecteurs ne sont pas presque orthogonaux puisque beaucoup sont presque colinéaires: en ne modifiant, par un changement de signe, qu'une coordonnée de z_j , on obtient un vecteur z'_j qui fait avec z_j un angle $\theta = 2 \operatorname{Arcsin}(1/\sqrt{n})$. Il convient donc de trier les z_j (en en jetant beaucoup à la poubelle) pour en extraire un bouquet presque orthogonal. C'est ce triage subtil qui est au cœur du théorème de Johnson et Lindenstrauss.

Cette construction paradoxale ouvre la voie au “compressed sensing”. Le compressed sensing repose sur la construction d'un nombre exponentiel de points z_1, \dots, z_N encore plus régulièrement disposés sur la sphère S^n . Dans l'exemple précédent on demandait simplement que les distances entre deux quelconques des N points z_j soient toutes égales à $\sqrt{2}$ (à ε près). Les vecteurs z_j sont alors presque orthogonaux. On demande maintenant une propriété beaucoup plus précise, dite *RIP*.

Définition 3.1. Soient $\varepsilon \in (0, 1)$ et $1 \leq k \leq N$. On dit que N vecteurs z_j vérifient la propriété $RIP(\varepsilon, k)$ si l'on a

$$(1) \quad (1 - \varepsilon) \sum_1^N |c_j|^2 \leq \left| \sum_1^N c_j z_j \right|^2 \leq (1 + \varepsilon) \sum_1^N |c_j|^2$$

pour tous les choix des coefficients c_j , $1 \leq j \leq N$, sous la seule condition que seulement k de ces coefficients soient non nuls.

Dans la construction précédente on se limitait à $k = 2$ et aux choix de suites c_k telles que $c_k = 0$ sauf si $k \in \{l, m\}$ avec dans ce cas $c_l = 1$, $c_m = -1$. Si l'on suppose que $k \leq c(\varepsilon)n / \log(N/n)$ cette construction est possible. La matrice Φ dont les colonnes sont les z_j , $1 \leq j \leq N$, est alors la matrice servant à faire les n mesures sur les vecteurs (ou signaux) k -parcimonieux de longueur N .

Définition 3.2. Supposons $0 \leq k \leq N$. Un signal $Y = (y_1, \dots, y_N)$ est k -parcimonieux dans une base \mathcal{B} de $l^2\{1, \dots, N\}$ lorsque seulement k des coordonnées α_n , $1 \leq n \leq N$, de Y dans \mathcal{B} sont non nulles. On ne cherche pas à savoir quels sont les indices des coordonnées non nulles.

La parcimonie stricte est rare et elle est le plus souvent remplacée par une condition portant sur le nombre $N(\varepsilon)$ des indices des α_n tels que $|\alpha_n| \geq \varepsilon$.

Soit \mathcal{F} une famille de signaux ou d'images. La décomposition de Karhunen-Loève est la recherche d'une base \mathcal{B} dans laquelle l'approximation des vecteurs $X \in \mathcal{F}$ par les k premières coordonnées soit la plus efficace possible en moyenne sur \mathcal{F} . Le point de vue de la parcimonie consiste à rechercher une base \mathcal{B} telle que tout vecteur $X \in \mathcal{F}$ ait une représentation parcimonieuse dans \mathcal{B} , les indices des coordonnées activées dans cette décomposition parcimonieuse dépendant de X .

La décomposition de Karhunen-Loève appliquée à un corpus d'images naturelles conduit inmanquablement à l'utilisation des séries de Fourier tandis que le point de vue de la parcimonie conduit à l'analyse par ondelettes. Les coordonnées actives sont alors associées aux bords des images et dépendent donc de l'image étudiée.

Le théorème du “compressed sensing” nous dit la chose suivante:

Théorème 3.2. *Supposons que la matrice Φ ait la propriété RIP avec $0 < \varepsilon < \sqrt{2} - 1$. Alors tout signal Y qui est k -parcimonieux est entièrement déterminé par la donnée de $X = \Phi(Y)$. En outre Y peut être reconstruit en résolvant le problème d'optimisation convexe*

$$\inf \left\{ \sum_1^N |z_j| ; X = \Phi(Z) \right\}$$

où, cette fois, on ne fait plus aucune hypothèse de parcimonie sur les vecteurs $Z = (z_1, \dots, z_N)$.

Appliquer ce théorème à une classe F de signaux ou d'images demande trois choses

- (a) Disposer de bases dans lesquelles ces signaux ou ces images $f \in F$ soient comprimés (parcimonieux).
- (b) Généraliser le théorème 3.2 au cas où la parcimonie au sens strict est remplacée par une propriété de décroissance rapide des coefficients.
- (c) Disposer de matrices Φ qui aient la propriété RIP et dont le stockage ne demande pas une charge en mémoire prohibitive.

Un exemple de matrices $n \times N$ convenant à la troisième contrainte est fourni par la construction suivante, due à Emmanuel Candès. On suppose que $k \leq c(\varepsilon)n / \log(N/n)$ et l'on considère la matrice aléatoire $\Phi(\omega)$ dont les coefficients sont $\pm 1/\sqrt{n}$. Alors avec une probabilité écrasante (supérieure à $1 - \epsilon_n$ où $\epsilon_n = \exp[-\gamma(\varepsilon)n]$) la matrice $\Phi(\omega)$ vérifie (1).

4. RÉFÉRENCES

- (1) J. Bobin, J-L. Starck, and R. Ottensamer, Compressed sensing in Astronomy, IEEE (Selected topics in signal processing), Vol. 2, No. 5, October 2008.
- (2) R. Coifman, F. Geshwind, and Y. Meyer, Noiselets, Applied and Computational Harmonic Analysis, 10 (2001), pp. 2744.
- (3) E. Candes and J. Romberg, Sparsity and incoherence in compressive sampling, Applied and Computational Harmonic Analysis, 23 (2007), pp. 969-985.
- (4) H. Nickisch and M. Steeger, Bayesian experimental design for compressed sensing (2008).