

Acquisition Comprimée et Algorithmes dits Approximate Message Passing

Raphaël Berthier

16 octobre 2017

Résumé

L'acquisition comprimée s'intéresse à la reconstruction d'un signal structuré à partir d'un petit nombre de mesures linéaires. Parmi les méthodes récentes pour résoudre ce problème, les algorithmes Approximate Message Passing (AMP) proposent de bonnes performances, un faible coût algorithmique et une bonne compréhension de son comportement grâce à une équation unidimensionnelle qui décrit son erreur, appelée l'équation State Evolution (SE). On présente ici comment, à partir d'une fonction de débruitage pour une classe de signaux, on construit un algorithme AMP d'acquisition comprimée pour cette même classe de signaux.

1 Acquisition comprimée

1.1 Cadre du problème

L'acquisition comprimée s'intéresse à la mesure d'un signal $\boldsymbol{\theta}$, que nous représenterons sous la forme d'un vecteur $\boldsymbol{\theta} \in \mathbb{R}^n$ de dimension n . Ce signal pourrait être une image, un son, ou une vidéo par exemple. Une mesure du signal est un réel y_i dépendant du signal, et l'acquisition comprimée se concentre sur des mesures *linéaires*, c'est-à-dire de la forme $y_i = \langle \mathbf{a}_i, \boldsymbol{\theta} \rangle$ pour un vecteur $\mathbf{a}_i \in \mathbb{R}^n$. On prend m telles mesures de sorte que le processus d'acquisition peut s'écrire de forme synthétique :

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}, \quad (1)$$

où \mathbf{A} est la matrice de taille $m \times n$ dont les lignes sont $\mathbf{a}_1, \dots, \mathbf{a}_m$.

Le but de l'acquisition comprimée est de reconstruire le signal $\boldsymbol{\theta} \in \mathbb{R}^n$ à partir des mesures $\mathbf{y} \in \mathbb{R}^m$ et du processus d'acquisition $\mathbf{A} \in \mathbb{R}^{m \times n}$. Dans le cas simple où \mathbf{A} est de rang n , la reconstruction est facile, sous réserve qu'il soit facile d'inverser une sous-matrice de \mathbf{A} de taille $n \times n$.

Cependant, cela implique que le nombre de mesures m soit plus grand que la dimension du signal n . Au contraire, l'acquisition comprimée s'intéresse au cas où le signal $\boldsymbol{\theta}$ est de grande dimension, et l'on souhaite que le nombre de mesures m soit le plus petit possible. Par exemple, l'acquisition comprimée est utilisée en imagerie par résonance magnétique (IRM)[LDSP08]. Lors d'une IRM, le signal est l'image du corps du patient et des mesures sont prises de manière séquentielle. Prendre un grand nombre de mesures nécessite un temps d'acquisition plus long, pendant lequel il peut être difficile de garder le corps immobile. On ne peut pas demander à un patient malade de garder sa respiration pendant plusieurs minutes ! D'autre part, on peut vouloir prendre plusieurs images

rapidement pour observer l'évolution temporelle (du cerveau par exemple, on parle alors d'IRM fonctionnelle).

Pour ces raisons, on se concentrera sur le cas où le ratio de sous-échantillonnage $\delta = m/n$ est le plus petit possible (et bien sûr plus petit que 1). Dans ce cas, le noyau de la matrice \mathbf{A} n'est pas réduit à $\{0\}$, donc le signal n'est pas déterminé de manière unique à partir des mesures. Pour espérer pouvoir reconstruire le signal, il faut utiliser la *structure* du signal, c'est-à-dire de l'information supplémentaire sur ce dernier. Cette structure peut prendre différentes formulations selon l'application ou le point de vue du statisticien. Cela peut être, par exemple, un ensemble $C \subset \mathbb{R}^n$ dont on sait qu'il contient $\boldsymbol{\theta}$; ou bien une distribution a priori P sur \mathbb{R}^n pour notre signal. Le défi de l'acquisition comprimée est alors d'utiliser la structure du signal et les mesures \mathbf{y} pour reconstruire le signal $\boldsymbol{\theta}$. Le problème est d'autant plus difficile que les signaux sont de grande dimension et donc qu'il faut être vigilant à la complexité des algorithmes proposés.

Pour résumer, une application étant donnée, l'acquisition comprimée essaie de trouver le bon processus d'acquisition (i.e. bien choisir les vecteurs \mathbf{a}_i , mais ils peuvent avoir des contraintes pratiques), la bonne structure pour le signal et le bon algorithme de reconstruction de sorte que le signal $\boldsymbol{\theta}$ puisse être reconstruit avec le moins de mesures possible. La beauté de l'acquisition comprimée est que des réponses abstraites ont été apportées, traitant de manière unifiée les cas de l'IRM, des sons ou encore des matrices de rang faible. En fait, nous verrons que les algorithmes dits *Approximate Message Passing* transforment n'importe quelle fonction de débruitage de signaux en méthode d'acquisition pour la même classe de signaux.

1.2 Parcimonie et reconstruction l_1

En acquisition comprimée, on suppose souvent que le signal étudié est *parcimonieux*. On dit qu'un signal est parcimonieux dans une base \mathcal{B} si peu de ses coefficients dans la base \mathcal{B} sont non nuls. Cette hypothèse est vérifiée pour beaucoup de signaux étudiés en pratique : par exemple, les bases d'ondelettes ont été découvertes, dans lesquelles les sons et les images naturelles sont (approximativement) parcimonieux. Ces bases d'ondelettes sont associées à des transformations rapides, qui permettent de décomposer facilement un signal dans la base d'ondelette à partir de la représentation dans la base canonique, ainsi que calculer la transformation inverse. Ainsi, on pourra supposer dans la suite, sans perte de généralité, que le signal $\boldsymbol{\theta}$ est parcimonieux dans la base canonique.

Plus précisément, on dira que le signal $\boldsymbol{\theta} \in \mathbb{R}^n$ est s -parcimonieux si $\|\boldsymbol{\theta}\|_0 = \#\{1 \leq i \leq n \mid \theta_i \neq 0\} \leq s$, et l'on notera $\Sigma_{s,n}$ l'ensemble des vecteurs s -parcimonieux de \mathbb{R}^n . Si le signal $\boldsymbol{\theta} \in \mathbb{R}^n$ est s -parcimonieux, il peut être reconstruit avec strictement moins de n mesures. En effet, pour reconstruire un signal $\boldsymbol{\theta} \in \Sigma_{s,n}$ à partir de m mesures $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} \in \mathbb{R}^m$, de la matrice \mathbf{A} et de l'hypothèse de s -parcimonie, on cherche notre signal dans l'ensemble $E = \{\boldsymbol{\theta}' \in \Sigma_{s,n} \mid \mathbf{y} = \mathbf{A}\boldsymbol{\theta}'\}$. La reconstruction sera possible si $E = \{\boldsymbol{\theta}\}$.

Étudions l'ensemble E lorsque la matrice \mathbf{A} est une matrice Gaussienne, c'est-à-dire que $\mathbf{A}_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq n$ sont des variables aléatoires i.i.d. Gaussiennes. Alors le noyau de \mathbf{A} est presque sûrement un sous-ensemble de \mathbb{R}^n de dimension $n - m$, en position "générique"¹. Prenons $m = s$. Alors l'intersection entre $\boldsymbol{\theta} + \ker \mathbf{A}$ et n'importe quel sous-espace de dimension s fixe est presque sûrement réduit à un point. Maintenant, constatons que $\Sigma_{s,n}$ est l'union de $\binom{n}{s}$ sous-espaces de dimensions s . Par conséquent, E est p.s. fini, de cardinal $\binom{n}{s}$. En prenant une mesure de plus,

1. On entend par là que sa loi est invariante par rotation.

E devient alors réduit à un singleton.

Par conséquent, en prenant $m = s + 1$ et une matrice \mathbf{A} Gaussienne, on obtient $\{\boldsymbol{\theta}' \in \Sigma_{s,n} | \mathbf{y} = \mathbf{A}\boldsymbol{\theta}'\} = \{\boldsymbol{\theta}\}$. Cela implique

$$\boldsymbol{\theta} = \underset{\boldsymbol{\theta}' : \mathbf{y} = \mathbf{A}\boldsymbol{\theta}'}{\operatorname{argmin}} \|\boldsymbol{\theta}'\|_0. \quad (2)$$

Ainsi, $s + 1$ mesures semblent suffire pour reconstruire un signal s -parcimonieux. s étant beaucoup plus petit que n dans les applications, cette amélioration est significative. Cependant, le problème n'est pas si simple car la minimisation (2) est NP-difficile [DMA97].

Une solution proposée est d'utiliser la relaxation convexe suivante

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}' : \mathbf{y} = \mathbf{A}\boldsymbol{\theta}'}{\operatorname{argmin}} \|\boldsymbol{\theta}'\|_1, \quad (3)$$

où $\|\boldsymbol{\theta}'\|_1 = |\boldsymbol{\theta}'_1| + \dots + |\boldsymbol{\theta}'_n|$. Ce problème étant la minimisation d'une fonction convexe sur un sous-espace affine, il est beaucoup plus simple à résoudre que la minimisation l_0 . Cependant, cette méthode a besoin de plus de mesures pour réussir.

Théorème 1. Soient $\epsilon \in (0, 1)$ et $s \leq n/2$. On rappelle que \mathbf{A} est une matrice Gaussienne. Il existe une constante universelle C telle que, si

$$m \geq C \left(s \log \frac{n}{s} + \log \frac{1}{\epsilon} \right), \quad (4)$$

alors avec probabilité $1 - \epsilon$, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ où $\hat{\boldsymbol{\theta}}$ est l'estimation obtenue par la minimisation l_1 (3).

Ce résultat est prouvé dans [Rau10, Théorèmes 2.3, 2.6 et 2.12]. Notons que le nombre de mesures nécessaires est linéaire en s , à un facteur logarithmique près. Ce théorème signifie donc qu'on peut reconstruire un signal parcimonieux avec un nombre de mesures significativement plus petit que n .

Considérer des matrices \mathbf{A} aléatoires est une manière astucieuse et classique de générer de "bonnes" manières de mesurer, de manière probabiliste. Le cas \mathbf{A} Gaussienne semble être le plus facile à étudier et permet d'obtenir de bons résultats théoriques. Cependant, dans les applications, il y a des contraintes physiques sur le processus de mesures. Par exemple, en IRM, on ne peut que échantillonner la transformée de Fourier de l'image mesurée. Dans ce cas, prendre \mathbf{A} Gaussienne n'est pas possible, et on compose \mathbf{A} en sous-échantillonnant aléatoirement les lignes de la matrice correspondant à la transformée de Fourier. Plus généralement, [Rau10] étudie plusieurs catégories de matrices aléatoires structurées. Les matrices aléatoires structurées mêlent la structure nécessaire pour en faire des matrices de mesures réalistes et l'aléa nécessaire pour leur donner de bonnes propriétés pour la reconstruction.

2 Approximate Message Passing pour la reconstruction l_1

2.1 Description de l'algorithme

Beaucoup d'algorithmes ont été proposés pour résoudre la minimisation (3). Ici, on en introduit un, nommé Approximate Message Passing. On peut le motiver comme perturbation d'une méthode plus simple, nommée Iterative Soft Thresholding (IST, ou seuillage doux itéré en français). L'IST

part d'une estimation quelconque du signal - disons $\hat{\boldsymbol{\theta}}^0 = 0$, et calcule récursivement la suite d'estimateurs

$$\hat{\boldsymbol{\theta}}^{t+1} = \eta_t \left(\hat{\boldsymbol{\theta}}^t + \mathbf{A}^T \mathbf{r}^t \right), \quad (5)$$

$$\mathbf{r}^t = \mathbf{y} - \mathbf{A} \hat{\boldsymbol{\theta}}^t. \quad (6)$$

Ici, η_t est l'opérateur de seuillage doux de paramètre λ_t . Il applique à chaque coordonnée la fonction réelle $\eta(x, \lambda_t) = \text{sign}(x)(|x| - \lambda_t)_+$.

On peut justifier cet algorithme avec l'heuristique suivante : calculer $\hat{\boldsymbol{\theta}}^t + \mathbf{A}^T \mathbf{r}^t$ correspond à faire un pas de la descente de gradient de $\|\mathbf{y} - \mathbf{A} \hat{\boldsymbol{\theta}}^t\|_2^2$. Ainsi, cette étape assure que nos estimations correspondent aux mesures. Ensuite, l'étape de seuillage assure que nos estimateurs sont parcimonieux. En effet, le seuillage doux est l'opérateur proximal de la norme l_1 . Cependant, cette intuition n'indique pas comment choisir les paramètres λ_t .

Dans [DMM09], une alternative à l'IST est proposée, sous le nom d'"Approximate Message Passing" (AMP) :

$$\hat{\boldsymbol{\theta}}^{t+1} = \eta_t \left(\hat{\boldsymbol{\theta}}^t + \mathbf{A}^T \mathbf{r}^t \right), \quad (7)$$

$$\mathbf{r}^t = \mathbf{y} - \mathbf{A} \hat{\boldsymbol{\theta}}^t + \mathbf{b}_t \mathbf{r}^{t-1}, \quad (8)$$

$$\mathbf{b}_t = \frac{1}{\delta} \left\langle \eta'_{t-1} \left(\hat{\boldsymbol{\theta}}^{t-1} + \mathbf{A}^T \mathbf{r}^t \right) \right\rangle. \quad (9)$$

Ici, η'_{t-1} est la fonction qui applique à chaque coordonnée la dérivée de la fonction de seuillage doux, qui est $\eta'(x, \lambda_{t-1}) = \mathbf{1}_{|x| > \lambda_{t-1}}$. De plus, on a noté $\langle \boldsymbol{\beta} \rangle = (\beta_1 + \dots + \beta_n) / n$ la valeur moyenne des coordonnées d'un vecteur.

L'AMP est très similaire avec l'IST, avec toutefois un terme inertiel, appelé le *terme Onsager*, dont l'interprétation n'est pas évidente. Pour obtenir cet algorithme, les auteurs de [DMM09] ont été inspirés d'un algorithme avec passage de messages (Message Passing Algorithm), dont ils ont fait des approximations jusqu'à obtenir la forme (7)-(9). Ceci justifie le nom "Approximate Message Passing".

L'approximation faite dans la genèse de l'AMP est valide dans la limite $n, m \rightarrow \infty$, $\delta = n/m$ restant fixé, pour certaines classes de matrices aléatoires \mathbf{A} , par exemple si les entrées $(A_{ij})_{i \leq m, j \leq n}$ sont i.i.d. selon $\mathcal{N}(0, 1/m)$. On supposera toujours cette hypothèse pour \mathbf{A} par la suite. Nous verrons empiriquement que l'AMP fonctionne bien pour des applications de taille $n \gtrsim 10^4$ - ce qui est vérifié dans les applications. Pour de tels signaux, l'AMP est nettement plus efficace que l'IST (voir [DMM09]).

2.2 Équation dite State Evolution

Un autre avantage majeur de l'AMP sur l'IST est que son comportement asymptotique² est décrit par une équation unidimensionnelle, appelée *state evolution* (SE) :

$$\tau_0^2 = \frac{1}{\delta n} \|\boldsymbol{\theta}\|_2^2, \quad (10)$$

$$\tau_{t+1}^2 = \frac{1}{\delta} \mathbb{E} \left[\frac{1}{n} \|\eta_t(\boldsymbol{\theta} + \tau_t \mathbf{Z}) - \boldsymbol{\theta}\|_2^2 \right], \quad (11)$$

2. Dans tout ce mémoire, "asymptotique" fera référence au régime $n, m \rightarrow \infty$, $\delta = m/n$ restant fixé.

où $\mathbf{Z} \sim \mathbf{N}(0, \mathbf{I}_n)$.

La suite τ_0, τ_1, \dots décrit le comportement de l'AMP selon l'heuristique suivante : asymptotiquement, on a

$$\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t \sim \boldsymbol{\theta} + \tau_t \mathbf{N}(0, \mathbf{I}_n), \quad (12)$$

$$\mathbf{r}^t \sim \tau_t \mathbf{N}(0, \mathbf{I}_m). \quad (13)$$

Commençons par interpréter cette heuristique avant de lui donner une formulation rigoureuse. L'équation (12) indique que $\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t$ est une version bruitée $\boldsymbol{\theta}$, et que l'écart-type de bruit est donné par τ_t , c'est-à-dire par les équations SE.

Pour donner un sens formel à cette heuristique, il faut considérer une suite de problèmes d'acquisition comprimée indexée par la dimension du signal n :

$$\mathbf{y}(n) = \mathbf{A}(n)\mathbf{x}(n), \quad (14)$$

où $m = m(n) = \delta n$. Toutes les quantités qui découlent du problème, tels que les estimateurs de l'AMP $\hat{\boldsymbol{\theta}}^0, \hat{\boldsymbol{\theta}}^1, \dots$, dépendent donc de n ; cependant on omettra cette dépendance à l'avenir pour alléger les notations.

On dira qu'une fonction $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ est pseudo-Lipschitz d'ordre k s'il existe une constante $L \geq 0$ telle que

$$\forall x, y \in \mathbb{R}^2, |\psi(x) - \psi(y)| \leq L(1 + \|x\|_2 + \|y\|_2) \|x - y\|_2. \quad (15)$$

Théorème 2. *On suppose toujours que \mathbf{A} est une matrice Gaussienne dont les entrées sont i.i.d. selon $\mathbf{N}(0, 1/m)$. On suppose de plus que lorsque $m, n \rightarrow \infty$, la distribution empirique des coordonnées de $\boldsymbol{\theta}$ converge vers une loi P_Θ sur \mathbb{R} , et que pour un certain $k \geq 2$, $\mathbb{E}_{P_\Theta}[\Theta^{2k-2}] < \infty$. Enfin, on suppose que $\frac{1}{n} \sum_{i=1}^n \theta_i^{2k-2} \xrightarrow[n, m \rightarrow \infty]{} \mathbb{E}_{P_\Theta}[\Theta^{2k-2}]$. Alors pour toute fonction pseudo-Lipschitz $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ d'ordre k , p.s.,*

$$\frac{1}{n} \sum_{i=1}^n \psi(\hat{\boldsymbol{\theta}}_i^{t+1}, \boldsymbol{\theta}_i) \xrightarrow[n, m \rightarrow \infty]{} \mathbb{E}[\psi(\eta_t(\Theta + \tau_t Z), \Theta)], \quad (16)$$

où $\Theta \sim P_\Theta$ et $Z \sim \mathbf{N}(0, 1)$ sont des variables indépendantes.

Ce théorème est prouvé dans [BM11]. Plus généralement, les auteurs démontrent le même résultat pour n'importe quelle suite de fonctions η_t , $t \geq 0$ séparables : on dit qu'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est séparable si elle est de la forme $f(x_1, \dots, x_n) = (g(x_1), \dots, g(x_n))$.

L'équation SE, que l'on rappelle ici,

$$\tau_{t+1}^2 = \frac{1}{\delta} \mathbb{E} \left[\frac{1}{n} \|\eta_t(\boldsymbol{\theta} + \tau_t \mathbf{Z}) - \boldsymbol{\theta}\|_2^2 \right], \quad (17)$$

s'interprète ainsi : τ_{t+1} est, à un facteur multiplicatif près, l'erreur moyenne quadratique de la fonction η_t dans le débruitage de $\boldsymbol{\theta}$, avec bruit Gaussien, d'écart-type τ_t . Puisque τ_t , $t \geq 0$, dicte la précision de l'algorithme AMP (voir éq. (12)), ce dernier sera d'autant plus performant que la fonction de seuillage doux débruite bien $\boldsymbol{\theta}$.

Lorsque $\eta = \eta_t$ ne dépend plus du temps, l'équation SE devient l'itération d'une même fonction, ce que l'on représente schématiquement dans la Figure 1. On voit que les écarts-types τ_t convergent vers une limite τ_* , dont la valeur est d'autant plus petite que la fonction que l'on itère prend de faibles valeurs. Ainsi, pour avoir une bonne reconstruction, il faut soit que δ soit grand, soit que η débruite bien $\boldsymbol{\theta}$.

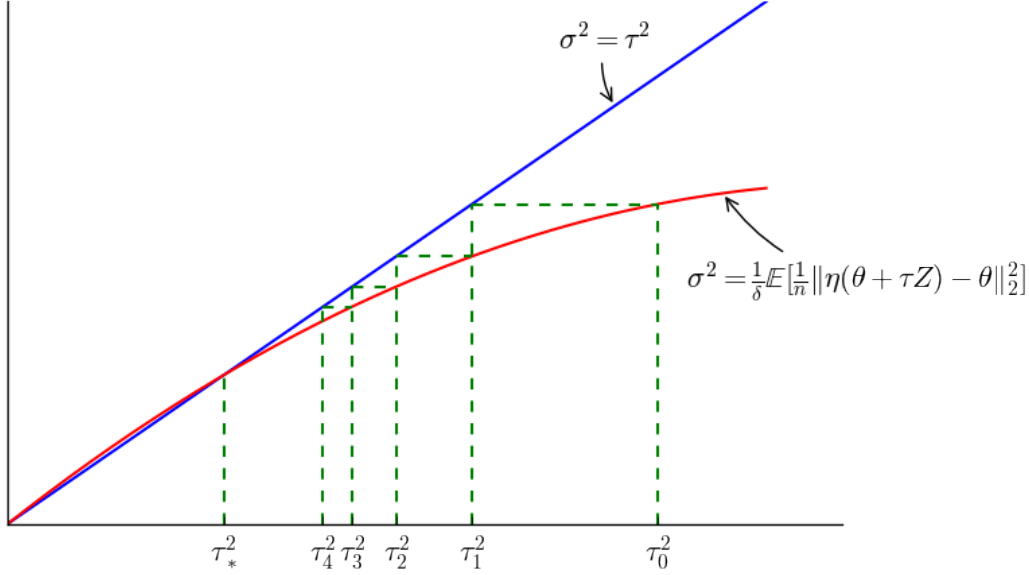


FIGURE 1 – Schéma de l'évolution de τ_t lorsque $\eta = \eta_t$ ne dépend pas du temps.

3 Approximate Message Passing à partir d'une fonction de débruitage

Ce parallèle entre le débruitage et l'AMP, fourni par l'équation SE, suggère de changer η_t . Au lieu de considérer la fonction de seuillage doux, on considèrera n'importe quelle fonction qui débruite θ [MMB16]. L'algorithme et l'équation SE restent quasi-inchangés, au détail près que \mathbf{b}_t doit être redéfini plus généralement comme

$$\mathbf{b}_t = \frac{1}{\delta n} \text{div} \eta_{t-1} \left(\hat{\theta}^{t-1} + \mathbf{A}^\top \mathbf{r}^t \right). \quad (18)$$

L'équation SE décrit toujours asymptotiquement le comportement de l'AMP pour une large classe de fonctions η_t (en particulier Lipschitz, voir [BMN17]).

Pour suggérer la diversité des possibilités offertes en changeant la fonction η_t , on donne deux exemples originaux ici. Par souci de concision, nous omettrons les aspects techniques des simulations, tels que le calcul de $\text{div} \eta_{t-1}$, où η_{t-1} sera une fonction non-triviale.

3.1 Acquisition comprimée de matrice de rang faible avec l'AMP

Un algorithme classique pour débruiteur une matrice est le seuillage doux des valeurs propres (ou singular value thresholding, SVT) [DG14]. Pour une matrice $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ dont la décomposition de valeurs singulières s'écrit

$$\mathbf{X} = \sum_{i=1}^{n_1 \wedge n_2} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top, \quad (19)$$

l'opérateur SVT de paramètre λ produit

$$\mathbf{S}(\mathbf{X}; \lambda) = \sum_{i=1}^{n_1 \wedge n_2} (\sigma_i - \lambda)_+ \mathbf{u}_i \mathbf{v}_i^\top. \quad (20)$$

Cette fonction de débruitage étant efficace pour des matrices de rang faible, il est naturel de l'utiliser dans l'AMP pour faire de l'acquisition comprimée de matrice de rang faible. Puisqu'ici le signal mesuré n'est pas un vecteur mais une matrice, le formalisme est traditionnellement écrit d'une manière différente des parties précédentes. Cependant les idées sous-jacentes restent les mêmes.

On veut reconstruire une matrice $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$ à partir de mesures linéaires $\mathbf{y} \in \mathbb{R}^m$, où

$$\mathbf{y} = \mathcal{A}(\mathbf{X}_0). \quad (21)$$

Ici $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ est un opérateur Gaussien linéaire, ou plus concrètement $y_i = \langle \mathbf{A}_i, \mathbf{X}_0 \rangle$ où $\mathbf{A}_i \in \mathbb{R}^{n_1 \times n_2}$ sont des matrices i.i.d. avec des entrées indépendantes de loi $(\mathbf{A}_i)_{r,c} \sim \mathcal{N}(0, 1/m)$.

L'algorithme AMP suivant permet de reconstruire \mathbf{X}_0 à partir des observations \mathbf{y} :

$$\hat{\mathbf{X}}^{t+1} = \mathbf{S}(\hat{\mathbf{X}}^t + \mathcal{A}^\top \mathbf{r}^t; \lambda_t), \quad (22)$$

$$\mathbf{r}^t = \mathbf{y} - \mathcal{A}(\hat{\mathbf{X}}^t) + \mathbf{b}_t \mathbf{r}^{t-1}, \quad (23)$$

où \mathcal{A}^\top est l'ajoint de l'opérateur \mathcal{A} , et

$$\mathbf{b}_t = \frac{1}{m} \operatorname{div} \mathbf{S}(\hat{\mathbf{X}}^{t-1} + \mathcal{A}^\top \mathbf{r}^{t-1}; \lambda_{t-1}). \quad (24)$$

L'équation SE

$$\tau_0^2 = \frac{1}{\delta n} \|\mathbf{X}_0\|_F^2, \quad (25)$$

$$\tau_{t+1}^2 = \frac{1}{\delta} \mathbb{E} \left[\frac{1}{n} \|\mathbf{S}(\mathbf{X}_0 + \tau_t \mathbf{Z}, \lambda_t) - \mathbf{X}_0\|_F^2 \right], \quad (26)$$

où $\|\cdot\|_F$ désigne la norme de Frobenius, estime que

$$\hat{\mathbf{X}}^t + \mathcal{A}^\top \mathbf{r}^t \sim \mathbf{X}_0 + \tau_t \mathbf{N}(0, \mathbf{I}_{n_1 n_2}), \quad (27)$$

$$\mathbf{r}^t \sim \tau_t \mathbf{N}(0, \mathbf{I}_m). \quad (28)$$

Il est remarquable que cette prédiction nous aide à calibrer les paramètres λ_t . En effet, puisque $n_1 n_2$ est grand dans les applications, on peut invoquer la loi des grands nombres pour justifier que, par (28), $\hat{\tau}_t = \|\mathbf{r}^t\|_2 / \sqrt{m}$ est un bon estimateur du niveau de bruit τ_t . Une règle empirique est alors de choisir $\lambda_t = 2\sqrt{n_1} \hat{\tau}_t \approx 2\sqrt{n_1} \tau_t$ (voir [GD13]). Ce choix de paramètre guidé par l'équation SE, contraste avec le cas de l'IST où il n'était pas évident de choisir les paramètre λ_t .

Des résultats de simulations sont présentés dans la figure 2, où l'on a choisi $m = 0.65 \cdot n_1 n_2$. On affiche l'erreur moyenne quadratique normalisée

$$\text{NMSE}(t) = \frac{\|\hat{\mathbf{X}}^t - \mathbf{X}_0\|_F^2}{\|\mathbf{X}_0\|_F^2}. \quad (29)$$

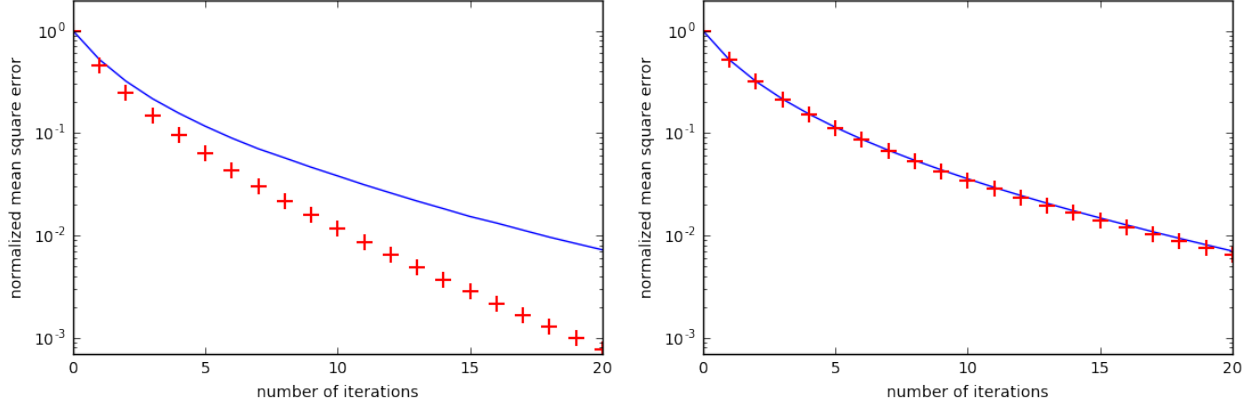


FIGURE 2 – Acquisition comprimée de matrice avec l’AMP : erreur moyenne quadratique normalisée en fonction du nombre d’itération. À gauche : matrice de taille 30×30 et de rang 3. À droite : matrice de taille 170×170 et de rang 17. Plus rouges (+) : simulations. Ligne bleue : prédiction de l’équation SE.

en fonction du nombre d’itération, et on compare avec la prédiction de l’équation SE :

$$\text{NMSE}(t) = \frac{\|\hat{\mathbf{X}}^t - \mathbf{X}_0\|_F^2}{\|\mathbf{X}_0\|_F^2} = \frac{\|\mathbf{S}(\hat{\mathbf{X}}^{t-1} + \mathcal{A}^\top \mathbf{r}^{t-1}, \lambda_{t-1}) - \mathbf{X}_0\|_F^2}{\|\mathbf{X}_0\|_F^2} \quad (30)$$

$$\approx \frac{\|\mathbf{S}(\mathbf{X}_0 + \tau_{t-1} \mathbf{Z}, \lambda_{t-1}) - \mathbf{X}_0\|_F^2}{\|\mathbf{X}_0\|_F^2} \quad \text{où } \mathbf{Z} \sim \mathbf{N}(0, \mathbf{I}_{n_1 n_2}), \text{ en utilisant (28)} \quad (31)$$

$$\approx \frac{\mathbb{E}[\|\mathbf{S}(\mathbf{X}_0 + \tau_{t-1} \mathbf{Z}, \lambda_{t-1}) - \mathbf{X}_0\|_F^2]}{\|\mathbf{X}_0\|_F^2} \approx \frac{n \delta \tau_t}{\|\mathbf{X}_0\|_F^2}. \quad (32)$$

Cette dernière étape est motivée par un argument de concentration de la mesure : la variable aléatoire $\|\mathbf{S}(\mathbf{X}_0 + \tau_{t-1} \mathbf{Z}, \lambda_{t-1}) - \mathbf{X}_0\|_F^2$ est fonction d’un vecteur Gaussien en grande dimension, et donc ne dévie pas trop de son espérance.

La figure 2 montre que la prédiction de l’équation SE est d’autant plus précise que la dimension ambiante est grande. Ceci insiste sur le fait que l’équation SE est avant tout une description *asymptotique*.

3.2 Acquisition comprimée d’images avec l’AMP

On s’intéresse désormais à l’acquisition comprimée d’images. On représente une image comme un tableau de réels en dimension 2, $\boldsymbol{\theta} = (\theta_{i,j})_{i \leq n_1, j \leq n_2}$, que l’on identifie à sa forme vectorielle $\text{vec}(\boldsymbol{\theta}) \in \mathbb{R}^n$, $n = n_1 n_2$. Avec cette dernière représentation, on peut reprendre les notations de la première partie sur l’acquisition comprimée. On obtient un petit nombre de mesures $\mathbf{y} \in \mathbb{R}^m$ par l’équation

$$\mathbf{y} = \mathbf{A} \boldsymbol{\theta}, \quad (33)$$

où $\mathbf{A} \in \mathbb{R}^{m \times n}$ est une matrice Gaussienne.

On rappelle que notre algorithme AMP s'écrit :

$$\hat{\boldsymbol{\theta}}^{t+1} = \eta_t(\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t), \quad (34)$$

$$\mathbf{r}^t = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}^t + \mathbf{b}_t \mathbf{r}^{t-1}, \quad (35)$$

$$\mathbf{b}_t = \frac{1}{m} \operatorname{div} \eta_{t-1}(\mathbf{x}^{t-1} + \mathbf{A}^\top \mathbf{r}^{t-1}), \quad (36)$$

où $\eta_t : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$ est une suite de fonction de débruitage d'images. Comme on l'a expliqué dans la partie 2.2, le choix de la suite de fonctions η_t est cruciale car plus les fonctions de débruitage η_t sont performantes, plus l'on peut compresser nos images.

Un manière traditionnelle de faire de l'acquisition comprimée d'images est de décomposer l'image dans une base d'ondelette où elle est (approximativement) parcimonieuse, puis de résoudre le problème de minimisation (3). Cela revient à utiliser l'algorithme AMP (34)-(36) avec pour fonction de débruitage η_t le seuillage doux dans la base d'ondelette. Ici, on propose d'utiliser une méthode plus originale et efficace, appelée le débruitage par patch (Non-Local Means, NLM). Étant donnée une image bruitée $\boldsymbol{\theta}$, NLM estime le pixel de coordonnées (i, j) comme une moyenne des pixels de $\boldsymbol{\theta}$:

$$\eta(\boldsymbol{\theta})_{i,j} = \frac{\sum_{(k,l)} W_{(k,l),(i,j)}(\boldsymbol{\theta}) \boldsymbol{\theta}_{k,l}}{\sum_{(k,l)} W_{(k,l),(i,j)}(\boldsymbol{\theta})}. \quad (37)$$

Les poids $W_{(k,l),(i,j)}(\boldsymbol{\theta})$ dépendent de la similarité entre les patchs dans $\boldsymbol{\theta}$ centrés autour de (k, l) et (i, j) respectivement, ainsi que de la distance entre les deux pixels. Plus précisément, on choisit une taille de patch $L \in \mathbb{N}_{>0}$, une portée $R > 0$, et un paramètre de précision $h > 0$. Pour une position (k, l) dans l'image, on note $P_{(k,l)}(\boldsymbol{\theta})$ la sous-image (ou patch) de $\boldsymbol{\theta}$, centré en (k, l) , de taille $L \times L$. Alors

$$W_{(k,l),(i,j)}(\boldsymbol{\theta}) = \mathbf{1}_{\|(i,j)-(k,l)\| \leq R} \exp\left(-\frac{\|P_{(k,l)}(\boldsymbol{\theta}) - P_{(i,j)}(\boldsymbol{\theta})\|_2^2}{L^2 h^2}\right). \quad (38)$$

Ainsi, la méthode NLM moyenne les patchs qui sont similaires les uns aux autres.

Dans la figure 3, on montre la reconstruction d'une image de taille 170×170 à l'aide de $m = 0.5 \times 170 \times 170$ mesures. Pour chaque itération $t \in \{0, 1, 2, 3, 4\}$, on montre les estimateurs $\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t$ (colonne de gauche) ainsi que leur version débruitée $\hat{\boldsymbol{\theta}}^{t+1} = \eta_t(\mathbf{x}^t + \mathbf{A}^\top \mathbf{r}^t)$ (colonne de droite). Dans la figure 4, on montre l'évolution de l'erreur moyenne quadratique renormalisée $\text{NMSE}(t) = \|\hat{\boldsymbol{\theta}}^t - \boldsymbol{\theta}_0\|_2^2 / \|\boldsymbol{\theta}_0\|_2^2$ en fonction du nombre d'itération. Encore une fois, l'équation SE semble bien décrire les résultats de la simulation.

Le papier [MMB16] étudie plus d'exemples de reconstruction d'images pour montrer l'efficacité de l'algorithme AMP.

4 Conclusion et perspectives de recherche

L'algorithme AMP permet donc de ramener un problème d'acquisition comprimée à un problème de débruitage avec bruit Gaussien, auquel souvent il existe déjà une solution dans la littérature. Cela permet d'aller au-delà de la reconstruction de signaux parcimonieux.



FIGURE 3 – Reconstruction de l'image Lena en utilisant NLM-AMP, avec un ratio de sous-échantillonnage $m/n = 0.5$: $\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t$ (colonne de gauche) et $\hat{\boldsymbol{\theta}}^{t+1} = \eta_t(\hat{\boldsymbol{\theta}}^t + \mathbf{A}^\top \mathbf{r}^t)$ (colonne de droite) pour $t \in \{0, 1, 2, 3, 4\}$.

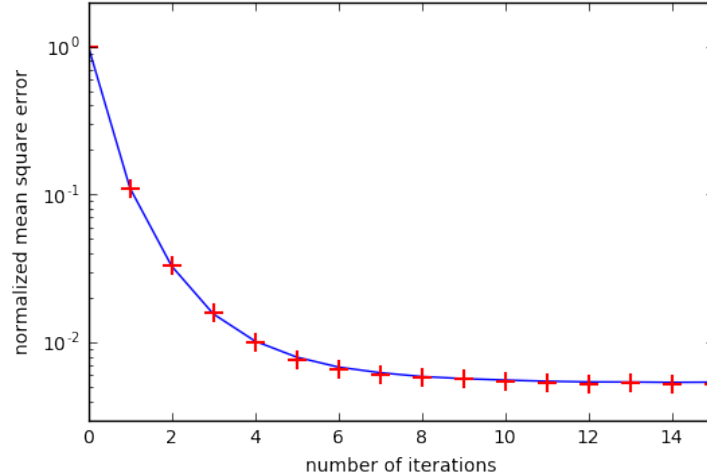


FIGURE 4 – Reconstruction de l’image Lena en utilisant NLM-AMP. Plus rouges (+) : évolution de l’erreur quadratique renormalisée. Ligne bleue : prédiction de l’équation SE.

Nous avons vu que l’équation SE était essentielle pour fournir des bornes théoriques sur la performance de l’algorithme, ainsi que pour choisir les paramètres en pratique. Cependant, elle n’est pas justifiée dans le cadre où \mathbf{A} n’est pas Gaussienne. Par exemple, [BLM⁺15] donne un sens à l’équation SE lorsque \mathbf{A} a des entrées sous-Gaussiennes, et que η_t est un polynôme séparable. Cependant, beaucoup de généralisations restent à faire dans cette direction.

Une autre direction de recherche est l’obtention de résultats non-asymptotiques pour l’équation SE, grâce à des arguments de convergence de la mesure. [RV16] couvre le cas où η_t est séparable.

Références

- [BLM⁺15] Mohsen Bayati, Marc Lelarge, Andrea Montanari, et al., *Universality in polytope phase transitions and message passing algorithms*, The Annals of Applied Probability **25** (2015), no. 2, 753–822.
- [BM11] Mohsen Bayati and Andrea Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Transactions on Information Theory **57** (2011), no. 2, 764–785.
- [BMN17] R. Berthier, A. Montanari, and P.-M. Nguyen, *State Evolution for Approximate Message Passing with Non-Separable Functions*, ArXiv e-prints (2017).
- [DG14] David Donoho and Matan Gavish, *Minimax risk of matrix denoising by singular value thresholding*, Ann. Statist. **42** (2014), no. 6, 2413–2440.
- [DMA97] G. Davis, S. Mallat, and M. Avellaneda, *Adaptive greedy approximations*, Constructive Approximation **13** (1997), no. 1, 57–98.
- [DMM09] David L Donoho, Arian Maleki, and Andrea Montanari, *Message-passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences **106** (2009), no. 45, 18914–18919.

- [GD13] M. Gavish and D. L. Donoho, *The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$* , ArXiv e-prints (2013).
- [LDSP08] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly, *Compressed sensing mri*, IEEE signal processing magazine **25** (2008), no. 2, 72–82.
- [MMB16] Christopher A Metzler, Arian Maleki, and Richard G Baraniuk, *From denoising to compressed sensing*, IEEE Transactions on Information Theory **62** (2016), no. 9, 5117–5144.
- [Rau10] Holger Rauhut, *Compressive sensing and structured random matrices*, Theoretical foundations and numerical methods for sparse recovery **9** (2010), 1–92.
- [RV16] Cynthia Rush and Ramji Venkataramanan, *Finite-sample analysis of approximate message passing*, Information Theory (ISIT), 2016 IEEE International Symposium on, IEEE, 2016, pp. 755–759.