Présentation de ...
One-bit Compressed Sensing by Linear Programming
Yaniv Plan & Roman Vershynin (CPAM, 2013)

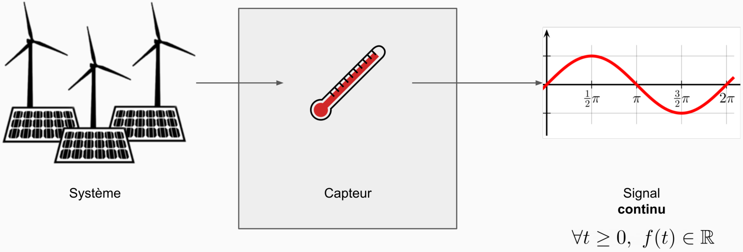
Théodor Lemerle & Théophile Cantelobre
Master M2A – Sorbonne Université

11 mars 2021

Sommaire

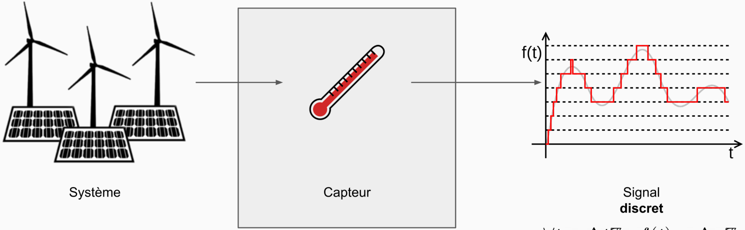
Motivation

Motivation : traitement du signal & CS idéal



- CS permet de reconstruire complètement le signal à partir de mesures... sans atteindre la fréquence de Nyquist-Shannon.
- Hypothèse : $f(t) \in \mathbb{R}$... précision infinie !

Motivation : traitement du signal & CS avec quantization



$$\forall t \in \Delta t\mathbb{Z},\ \ f(t) \in \Delta x\mathbb{Z}$$

- Quantization comme bruit : CS permet de le corriger !
- Deux approches pour traiter le bruit de discrétisation : même nombre de bits !
  - Augmenter le nombre de bits par mesure : hardware lent.
  - Augmenter le nombre de mesures : hardware rapide.

Quantization extrême : 1-bit compressed sensing

Poussé à l'extrême : **mesures avec 1-bit** (signes).

Perte de l'amplitude

$$\forall \lambda > 0, \ \operatorname{sign}(A\lambda x) = \operatorname{sign}(Ax)$$

Problème de dégénérescence $(\mathcal{BP})$

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \tag{1}$$

$$\text{s.t.} \ \operatorname{sign}(Ax) = y \tag{2}$$

...decrease objective by multiplying admissible iterate by $\lambda \to 0^+$.

Garanties théoriques

Principal résultat

Theorem (Théorème 1.1)
Soient n, m, et $s > 0$. Soit A une ***matrice gaussienne*** *de taille $m \times n$. Soit $\delta > 0$ dépendant de s, m et n. Alors, avec une probabilité au moins $1 - C\exp(-c\delta m)$, on a* ***uniformément*** *sur tout $x \in \mathbb{R}^n$ qui sont* ***essentiellement parcimonieux*** *d'ordre s : si $y = \operatorname{sign}(Ax)$. Alors, la solution $\hat{x}$ du problème $(\mathcal{P}_{1-bit})$ vérifie :*

$$\left\| \frac{\hat{x}}{\|\hat{x}\|_2} - \frac{x}{\|x\|_2} \right\|_2 \leq \delta.$$

Situation par rapport au cours

Points communs	Différences
Garanties uniformes.	Essentiellement parcimonieux.
Matrice de mesure gaussienne.	Reconstruction *inexacte*.

Remarque : essentiellement parcimonieux

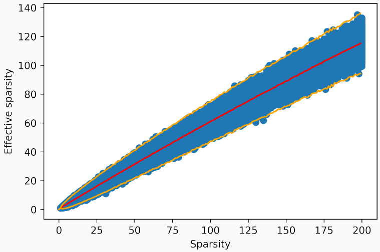
Essentiellement parcimonieux

$$ES(x) = \frac{\|x\|_1^2}{\|x\|_2^2} \leq s$$

Expérience numérique

Pour des vecteurs s-parcimonieux aléatoires, $ES(x)$ estime bien $\|x\|_0$:

$$0.6s \lessapprox ES(x) \lessapprox 0.75s$$



Stratégie de la preuve

1. Géométrie de la reconstruction à partir du signe.

Outils principaux : random tesselations & ε-nets.

Résultat objectif : pour m bien choisi, w.h.p. si $\mathrm{sign}(Ax) = \mathrm{sign}(Ax')$ alors $\|x - x'\|_2 \leq \delta$.

2. Vecteurs essentiellement parcimonieux

Outils principaux : concentration de la mesure.

Résultat objectif : Une solution de notre problème d'optimisation préserve la parcimonie essentielle du signal d'intérêt :

$$ES(\hat{x}) \leq ES(x) \cdot C\sqrt{\log(2n/m + 2m/n)}$$

$\rightarrow$ **Mise en commun**

$\hat{x}$ vérifie 2, donc 1 tient pour x et $\hat{x}$ en particulier.

Esquisse de la preuve : partie 1/2

Idée générale

Quantifier combien de mesures gaussiennes de 1-bit il faut pour séparer tous les signaux possibles :
avec h.p. uniformément sur les x, y, si $\|x - y\|_2 > \delta$, il existe $\Omega(m)$ hyperplans qui séparent x et y.

1. Contrôles des têtes et des queues

1. Séparer x (et y) en : $x = \underbrace{x_0}_{\in \text{ net}} + \varepsilon x'$.
2. S'assurer que les x_0 et y_0 sont séparés par $\Omega(m)$ hyperplans.
3. S'assurer que les x' et y' le sont par $\Omega(m) - o(m)$ hyperplans.
4. Tout mettre ensemble.

2. Plus petit ε-net possible.

$$\log N(K_{n,s}, \varepsilon) \leq \frac{Cs}{\varepsilon^2} \log\left(\frac{2n}{s}\right)$$

Esquisse de la preuve : partie 2/2

On veut montrer : $\frac{\|\hat{x}\|_2^2}{\|\hat{x}\|_1^2} \leq \frac{\|x\|_2^2}{\|x\|_1^2} \cdot C\sqrt{\log(2n/m + 2m/n)}$

Un résultat de concentration uniforme sur la déviation moyenne.

$$\frac{1}{m}\|Ax\|_1 \geq \frac{\|x\|_2}{2} \text{avec proba. au moins } 1 - C\exp(-cm)$$

La preuve reprend les grandes idées de la preuve de la propriété RIP pour les matrices à entrées gaussiennes.

Une borne inférieure sur $\|\hat{x}\|_2$

$$\|\hat{x}\|_2 \geq c/\sqrt{\log(2n/m + 2m/n)}$$

Mis bout à bout :

$$\frac{\|\hat{x}\|_1}{\|\hat{x}\|_2} \leq \frac{\|x\|_1}{\|Ax\|_1\|\hat{x}\|_2} \leq \frac{2\|x\|_1}{\|\hat{x}\|_2\|x\|_2} \leq \frac{\|x\|_2^2}{\|x\|_1^2} \cdot C\sqrt{\log(2n/m + 2m/n)}$$

Résumé : un programme linéaire (LP) à résoudre

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \ & \|x\|_1 \\
\text{s.t.} \ & \operatorname{sign}(Ax) = y \\
& \|Ax\|_1 \geq m
\end{aligned}
$$

Résumé : un programme linéaire (LP) à résoudre

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & \|x\|_1 \\
\text{s.t.} \quad & \operatorname{sign}(Ax) = y \\
& \|Ax\|_1 \geq m
\end{aligned}
$$

$$
\begin{aligned}
\min_{u \in \mathbb{R}^n} \quad & \mathbf{1}^T u \\
\text{s.t.} \quad & \forall i,\ -u_i \leq y_i \leq u_i \\
& \forall i,\ u_i \geq 0 \\
& \forall i,\ y_i \langle A_i, x \rangle = 1 \\
& \langle y, Ax \rangle \geq m
\end{aligned}
$$

Expériences

Méthodologie expérimentale

Protocole expérimental

1. Générer un signal parcimonieux (dans une certaine base) : x.
2. Tirer une matrice $A \sim \mathcal{N}(0, 1)$.
3. Observer $y = \mathrm{sign}(Ax)$.
4. Résoudre $(\mathcal{P}_{1-bit})$: $\hat{x}$.

... répéter N_{trials} fois.

Méthodologie expérimentale

Protocole expérimental

1. Générer un signal parcimonieux (dans une certaine base) : x.
2. Tirer une matrice $A \sim \mathcal{N}(0, 1)$.
3. Observer $y = \mathrm{sign}(Ax)$.
4. Résoudre $(\mathcal{P}_{1-bit})$: $\hat{x}$.

... répéter N_{trials} fois.

Métriques : comment évaluer une reconstruction ? voir [18]

- Erreur angulaire : $\epsilon(x, \hat{x}) = \frac{1}{\pi} \arccos\langle x, \hat{x}\rangle$
- Signal-to-Noise ratio : $\mathrm{SNR}(x, \hat{x}) = 20 \log_{10} \frac{\|x\|_2}{\|x-\hat{x}\|_2}$
- (Distance de Hamming : $d_H(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{XOR}(\mathrm{sign}(Ay), \mathrm{sign}(A\hat{y}))$)

Méthodes comparées : Basis pursuit, BIHT... and 1-bit LP

Basis pursuit (avec seuillage)

Programmation linéaire puis seuillage :

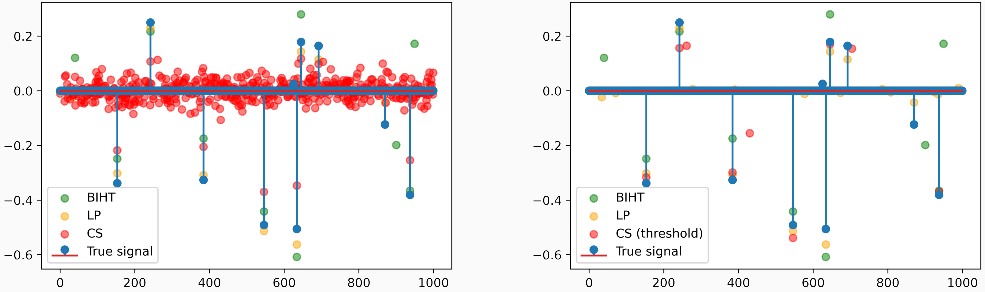
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \|x\|_1 \\ \text{s.t.} \ & Ax = y \end{aligned}$$

Binary Iterative Hard Thresholding [18]

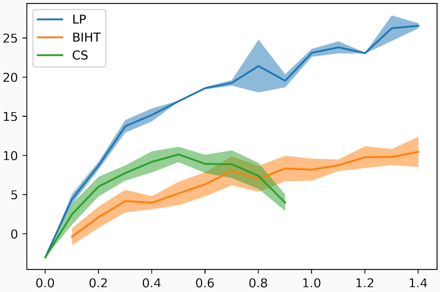
Descente de gradient projetée (pour $\|\cdot\|_0$ = Hard Thresholding) :

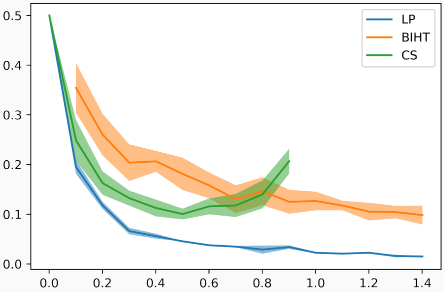
$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} \|y - Ax\|_2^2 \\ \text{s.t.} \ & \|x\|_0 = s \end{aligned}$$

Reconstruction de signaux parcimonieux synthétiques (1/2)



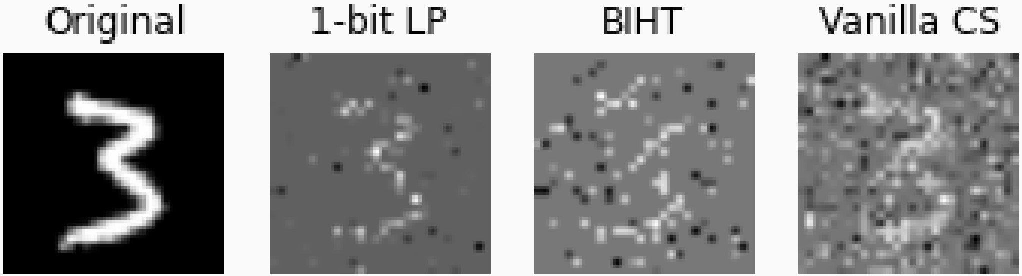
Reconstruction de signaux parcimonieux synthétiques (2/2)





Performance moyenne (10 essais) avec écart type.
À gauche : SNR. À droite : erreur angulaire. En abscisse : m/n.

Eye candy : MNIST



Résumé & Conclusion

Résumé & Conclusion

1-bit compressed sensing with LP

- Reconstruction approximative du signal d'intérêt sur $\mathbb{S}^{n-1}$.
- Garanties uniformes sur les signaux essentiellement parcimonieux ($\Sigma_s \subset K_{n,s}$).
- Programme linéaire.

CS indissociable du mode d'acquisition

- Problématique hardware : vitesse d'échantillonnage vs qualité.
- Voir aussi choix du BOS, tomographie, ...

Des questions ?

Bibliographie essentielle

[0] : *One-bit Compressed Sensing by Linear Programming*, Plan & Vershynin (2013)

[6] : *1-bit compressive sensing*, Boufounous & Baraniuk (2008)

[18] : *Robust 1-bit compressive sensing via binary stable embeddings of sparse embeddings*, Jacques et al. (2011)