

# APPLICATION DES MODÈLES GÉNÉRATIFS AUX PROBLÈMES DE COMPRESSED SENSING

Gabriel Kasmi, Hugo Thimonier

Mars 2020

# Plan

- 1 Objectifs et présentation générale
- 2 Contributions et résultats théoriques
- 3 Expérimentations et répliques
- 4 Insertion dans la littérature

# Objectifs et présentation générale

## Problématique

Objectif en compressed sensing : reconstruction de signal  $x \in \mathbb{R}^n$  à partir d'un nombre de mesures  $m \ll n$ . Le problème est le suivant (dans sa version bruitée) :

$$y = Ax^* + \eta \quad (1)$$

Le problème est formulé comme suit :

$$\hat{x}_0 \in \arg \min_{t \in \mathbb{R}^n, At=y} \|t\|_1 \quad (2)$$

# Objectifs et présentation générale

## Problématique

Hypothèse cruciale sur la structure de  $x$  :  $x \in \Sigma_s$ . L'idée de l'approche générative est de formuler une autre hypothèse de structure, à savoir que l'on suppose que  $x \in \{G(z), z \in \mathbb{R}^k\}$ , où  $G$  est un générateur (par exemple un GAN, un VAE, etc).

Le problème se réécrit comme suit :

$$y = AG(z^*) + \eta \quad (3)$$

Le problème de reconstruction du signal se réécrit quant à lui :

$$\min_z \|AG(z) - y\|_2^2 \quad (4)$$

# Objectifs et présentation générale

## Motivation

Pourquoi formuler une telle hypothèse de structure ?

- Les avancées rendues possibles par les GAN permettent d'approximer des densités de probabilité complexes
- *l'a priori* que l'on formule sur la distribution des données est plus informatif : les performances avec peu de mesures
- Généralise l'approche sparse

# Plan

- 1 Objectifs et présentation générale
- 2 Contributions et résultats théoriques**
- 3 Expérimentations et répliques
- 4 Insertion dans la littérature

Bora et al. (2017) : premiers à appliquer les méthodes génératives aux problèmes de compressed sensing.

Contributions :

- Proposer une généralisation des conditions de design sur les matrices de reconstruction
- Etendre les théorèmes garantissant une "bonne" reconstruction à des larges classes de générateurs
- Expérimentations qui montrent que la méthode fonctionne

## Théorème 1

Soient  $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$   $L$ -Lipschitz et  $A \in \mathbb{R}^{m \times n}$  une matrice aléatoire Gaussienne pour  $m = O\left(k \log \frac{Lr}{\delta}\right)$  telle que ses entrées sont

$a_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{m}\right)$ . Pour tout  $x^* \in \mathbb{R}^n$  et toute observation  $y = Ax^* + \eta$ , soit  $\hat{z} = \arg \min_z (\|y - AG(z)\|_2 + \varepsilon)$  tel que  $\|\hat{z}\|_2 \leq r$ . Alors, avec probabilité  $1 - \exp(-\Omega(m))$ ,

$$\|G(\hat{z}) - x^*\|_2 \leq 6 \min_{z^* \in \mathbb{R}^k, \|z^*\|_2 \leq r} \|G(z^*) - x^*\|_2 + 3 \|\eta\|_2 + 2\varepsilon + 2\delta$$



## Théorème 2

Soit  $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$  un modèle génératif issu d'un réseau de neurones à  $d$  couches utilisant des fonctions d'activation ReLU. Soit  $A \in \mathbb{R}^{m \times n}$  une matrice aléatoire Gaussienne pour  $m = O(kd \log n)$  telle que ses entrées sont  $a_{ij} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{1}{m}\right)$ . Alors pour tout  $x^* \in \mathbb{R}^n$  et toute observation  $y = Ax^* + \eta$ , soit  $\hat{z} = \arg \min_z (\|y - AG(z)\|_2 + \varepsilon)$ , avec probabilité  $1 - \exp(-\Omega(m))$ ,

$$\|G(\hat{z}) - x^*\|_2 \leq 6 \min_{z^* \in \mathbb{R}^k} \|G(z^*) - x^*\|_2 + 3 \|\eta\|_2 + 2\varepsilon$$

# Contributions théoriques

Dhar et al : généralisation

Dhar et al. (2017) : palient les faiblesses de Bora et al et proposent un lien entre l'approche sparse et l'approche générative.

- Les garanties théoriques sont unifiées (on a des garanties pour le régime sparse et le régime génératif)
- Posent les jalons pour des approches appliquées prometteuses (Transfer CS)

## Théorème 1

Soit  $G : B^k(r) \rightarrow \mathbb{R}^n$   $L$ -Lipschitz. Soient  $\alpha \in (0, 1)$ ,  $s > 0$  et  $A \in \mathbb{R}^{n \times m}$  une matrice aléatoire Gaussienne telle que

$$m = O\left(\frac{1}{\alpha^2} \left(k \log \frac{Lr}{\delta} + s \log \frac{n}{s}\right)\right)$$

satisfaisant la condition  $S - REC(S_{1,5s,G}, 1 - \alpha, \delta)$  et  $RIP(2s, \alpha)$  (condition garantie par le lemme ??.) Soit  $\Delta$  un décodeur satisfaisant le lemme ??. Alors, pour tout  $x \in \mathbb{R}^n$  avec probabilité  $1 - \exp(-\Omega(\alpha^2 m))$ , on a

$$\|x - \Delta(Ax + \varepsilon)\|_2 \leq \frac{1}{\sqrt{2s}} C_0 \sigma_{s,G}(x) + C_1 \eta_{\max} + \delta'$$

$$\text{où } C_0 = 2 \left( \frac{1 + \alpha}{1 - \alpha} + 1 \right), \quad C_1 = \frac{2}{1 - \alpha} + 1 \text{ et } \delta' = \frac{\delta}{1 - \alpha}$$

## Théorème 2

Soit  $G : \mathbb{R}^k \rightarrow \mathbb{R}^n$  un réseau de neurones avec  $d$  couches, au plus  $c$  noeuds par couche et des activations ReLU. Soient  $\alpha \in (0, 1)$ ,  $s > 0$  et  $A \in \mathbb{R}^{n \times m}$  une matrice aléatoire Gaussienne telle que

$$m = O\left(\frac{1}{\alpha^2} \left((k + s)d \log c + (k + s) \log \frac{n}{s}\right)\right)$$

satisfaisant la condition  $S - REC(S_{1,5s,G}, 1 - \alpha, \delta)$  et  $RIP(2s, \alpha)$  Soit  $\Delta$  un décodeur satisfaisant le lemme 1 de Dhar et. al.. Alors, pour tout  $x \in \mathbb{R}^n$  avec probabilité  $1 - \exp(-\Omega(\alpha^2 m))$ , on a

$$\|x - \Delta(Ax + \varepsilon)\|_2 \leq \frac{1}{\sqrt{2s}} C_0 \sigma_{s,G}(x) + C_1 \eta_{\max} + \delta'$$

où  $C_0 = 2 \left( \frac{1 + \alpha}{1 - \alpha} + 1 \right)$ ,  $C_1 = \frac{2}{1 - \alpha} + 1$  et  $\delta' = \frac{\delta}{1 - \alpha}$

# Plan

- 1 Objectifs et présentation générale
- 2 Contributions et résultats théoriques
- 3 Expérimentations et répliques**
- 4 Insertion dans la littérature

Cadre expérimental (commun aux deux articles) :

- Bases de données MNIST (et Omniglot) et celebA
- Le LASSO est proposé en point de référence pour la comparaison
- Le bruit d'observation est supposé Gaussien
- Les modèles utilisés sont des GAN ou des VAE
- La métrique retenue est l'erreur  $\ell_p$  de reconstruction

Concernant l'approche générative de Bora et al,

- A faible nombre de mesures, l'approche surpasse le LASSO en termes de précision
- Plafonnement lorsque le nombre de mesures dépasse la dimension de l'espace latent
- La principale source d'erreur est une erreur dite de *représentation*

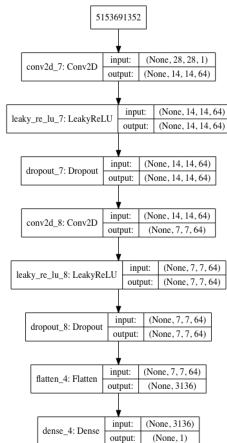
Expérimentalement, les résultats sur l'approche sparse-gen montrent que:

- L'approche sparse-gen allie le "meilleur" des deux approches
- Elle permet de juguler l'erreur de représentation en autorisant les déviations sparses
- Du point de vue pratique, le Transfer CS est une technique prometteuse.

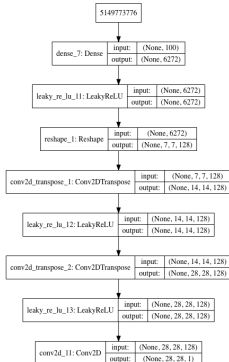
Nous essayons de répliquer les résultats de Dhar et al. avec notre propre GAN entraîné sur des données MNIST et notre propre implémentation du basis pursuist en guise de point de comparaison.



Figure: Architecture du générateur et du discriminant



(a) Discriminant



(b) Générateur

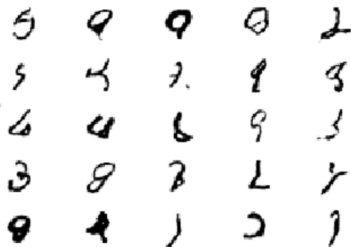
# Répliques

## Entrainement du modèle

Figure: Evolution des performances du générateur



(a) 10 epochs



(b) 100 epochs

# Réplications

## Reconstruction LASSO

Top: Original digits  
2nd line: 25 measurements  
2nd line: 100 measurements  
Bottom: 500 measurements

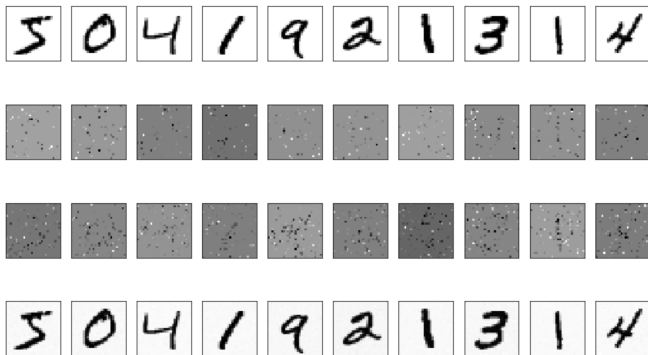


Figure: Reconstruction du signal par LASSO

# Réplications

## Reconstruction LASSO

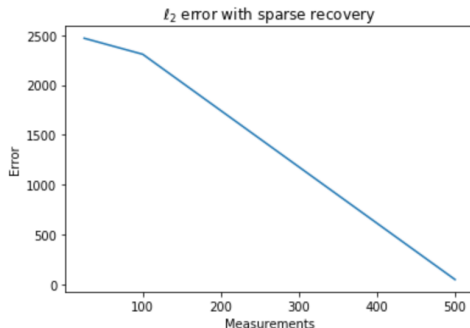


Figure: Erreur  $\ell_2$  du LASSO en fonction du nombre d'observations

La courbe a la même forme que la figure 1 (a) de Bora et al. (pour la courbe du LASSO)

# Répliques

## Tentatives ultérieures

Nous ne sommes malheureusement pas parvenus à répliquer l'approche générative.

# Plan

- 1 Objectifs et présentation générale
- 2 Contributions et résultats théoriques
- 3 Expérimentations et répliques
- 4 Insertion dans la littérature**

Les résultats sont essentiellement une *extension* du cadre sparse.

- Le paramètre de sparsité  $s$  est remplacé dans les ordres de grandeur du nombre de mesure par la dimension de l'espace latent  $k$ ,
- Dans le cas lipschitzien, le terme logarithmique ne dépend pas de la dimension de l'espace du signal  $n$
- Le nombre minimal de mesures dans l'approche sparse-gen est plus important que dans le cas génératif ; ainsi dans le cas lipschitzien le terme  $s \log \frac{n}{s}$  réapparaît.
- Dans le cas sparse-gen, comme pour la borne d'oracle du théorème 6.1 de Bickel et al, la borne dépend de la sparsité.

**Deep CS** L'idée est d'entraîner le générateur en même temps que l'on cherche à reconstruire le signal. En effet il n'est pas forcément possible d'avoir un générateur pré-entraîné à disposition pour l'application concernée.

On procède en deux étapes : reconstruire un signal à partir du générateur puis réentraîner celui-ci, puis à nouveau reconstruire le signal etc. Cela permet notamment "d'ajuster" l'ensemble image de  $G$  et ainsi de limiter l'erreur de représentation. Cela suppose cependant d'être dans un domaine *data rich*



**Bayesian CS** En Compressed sensing, l'approche bayésienne permet de spécifier une distribution sur la sparsité. On dispose ainsi de davantage d'informations puisqu'en plus de l'estimation ponctuelle on a une distribution.

Pour les GANs, l'approche bayésienne permet de disposer d'une distribution sur  $\theta_G$  les paramètres du générateurs. Ainsi mêler les deux approches pourrait permettre de trouver la distribution  $G(\theta_G, \bullet)$  la plus adaptée aux données que l'on cherche à reconstruire.

C'est une approche alternative au deep CS où plutôt que de réentraîner le modèle on explore l'espace des paramètres de ce dernier.

A notre connaissance, aucun papier n'a encore essayé de proposer une approche bayésienne dans le cadre du generative CS.

Merci pour votre attention