

## 1 | Déroulement du mini projet

Le travail s'organisera autour de quatre phases :

1. Apprendre un dictionnaire à partir de signaux d'apprentissage.
2. La seconde phase de ce travail consistera à développer des méthodes de codage parcimonieux.
3. La troisième Générer des matrices de mesures et valider vos choix de nombre de mesure.
4. Enfin, appliquer le procédé du compressive sensing afin de compléter des données manquantes.

### Rendu

Un rapport reprenant le travail d'analyse initial augmenté des aspects de tests sera fourni (vous devez répondre aux questions sous forme d'un rapport et non d'une copie d'examen). Des **graphiques et des tableaux comparatifs** permettront de mettre en évidence les résultats des simulations pour la reconstruction parcimonieuse et le procédé du compressive sensing et **avoir un avis critique sur ces résultats**.

### Modalités

Ce mini projet doit être réalisé par groupes de 3 élèves. La qualité du rapport, du code et le bon fonctionnement de celui-ci contribueront à la note finale de votre mini projet.

## 2 | Objectifs

Dans le procédé du compressive Sensing, les méthodes de reconstruction permettent de retrouver le signal à partir de sa forme mesurée car l'information portée par le signal se retrouve intégralement dans la forme compressée (mesurée).

Nous nous intéresserons dans ce travail à la validation des méthodes de codage parcimonieux et la possibilité d'exploitation du dictionnaire appris pour reconstruire un signal dont des données sont manquantes en utilisant la méthode (de codage parcimonieux) validée.

Les données qu'on va utiliser seront les mêmes représentées dans le fichier "data" et qui correspondent à la production des puits en gaz, la production d'huile et la production en eau. Ces données seront considérées comme une matrice. Cette matrice est de taille  $\mathbf{N} \times \ell$ ,  $\mathbf{N} = 98$  et  $\ell = 108$ .

Les colonnes  $\mathbf{X}_i$ ,  $i = 1, \dots, \ell$  représentent des vecteurs d'apprentissage et nous permettrons d'apprendre un dictionnaire  $\mathbf{D} \in \mathcal{M}_{\mathbf{N} \times \mathbf{K}}(\mathbb{R})$  dans lequel les signaux  $\mathbf{X}_i$  sont parcimonieux.

Le dictionnaire appris par la méthode K-SVD utilise la méthode de l'OMP est une matrice  $\mathbf{D} \in \mathcal{M}_{\mathbf{N} \times \mathbf{K}}(\mathbb{R})$  où  $\mathbf{N} = 98$ ,  $\mathbf{K} = 100$ ,  $\mathbf{L} = 10$  et  $\epsilon = 10^{-6}$ .

On notera  $\mathbf{z}$  le vecteur dont des données sont manquantes à cause d'une défaillance des capteurs de mesure des données dans un puits pétrolier.

### 3 | Apprentissage d'un dictionnaire

En utilisant la méthode d'apprentissage du KSVD, apprendre un dictionnaire  $\mathbf{D}$  sur les vecteurs d'apprentissage.

#### Question 1.

1. Décrire l'algorithme de l'apprentissage d'un dictionnaire utilisant la méthode KSVD (l'algorithme KSVD doit être décrit).
2. Implémenter l'algorithme pour en déduire le dictionnaire  $\mathbf{D}$  adapté.

### 4 | Codage parcimonieux

La résolution du problème de minimisation ( $\ell_0$ )

$$\min_{\tilde{\alpha}} \|\alpha\|_0 \text{ sous la contrainte } \mathbf{D}\tilde{\alpha} = x \quad (1)$$

nécessite une recherche exhaustive de la solution la plus parcimonieuse  $\tilde{\alpha}$  et est très complexe à mettre en œuvre. Plusieurs méthodes ont été proposées dans la littérature pour contourner ce problème.

#### 4.1 Méthodes gloutones

La famille des algorithmes gloutons est large et elle repose sur des sélections d'atomes qui contribuent le plus au résiduel pris lorsqu'on ôte du signal original la contribution des atomes sélectionnés. Il existe des algorithmes permettant

- ♣ la sélection d'un seul atome par itération (Matching pursuit et Orthogonal Matching Pursuit),
- ♣ la sélection de plusieurs atomes par itération (Stagwise Orthogonal Matching Pursuit),
- ♣ la sélection et le rejet d'atomes par itération (Compressive Sampling du Matching Pursuit).

#### Description de la méthode CoSaMP

L'algorithme du Compressive Sampling du Matching Pursuit (CoSaMP) consiste à la **sélection** des atomes et au **rejet** de quelques atomes déjà sélectionnés en une itération. Le CoSaMP s'initialise comme suit :

$\alpha = 0_{\mathbf{K},1}$ ,  $\mathbf{R} = \mathbf{x}$ ;       $\mathbf{R}$  : le résiduel  
 $\text{Supp} = \emptyset$        $\text{Supp}$  : le support de la solution parcimonieuse

Le CoSaMP suit cinq étapes principales et repose sur la connaissance a priori de l'ordre  $\mathbf{s}$  de parcimonie de la solution :

Tant que le critère d'arrêt n'est pas satisfait faire,  $k \geq 1$ ,

1. La sélection : Déterminer les  $2s$  atomes de plus grande contribution au résiduel et en déduire le vecteur des positions de ces atomes, on le notera le support de la sélection **supp1**.
2. Mise à jour du support : Fusionner le support de la sélection **supp1** avec le support de la solution de l'itération précédente **supp** :

$$\mathbf{supp} = \mathbf{supp} \cup \mathbf{supp1}$$

On note  $\mathbf{AS} = \mathbf{D}(:, \mathbf{supp})$  la matrice des atomes actifs sélectionnés.

3. Estimation : Utiliser la méthode des moindres carrés pour estimer les coefficients  $\mathbf{z}_i$  de la solution parcimonieuse  $\alpha$  correspondant au support mis à jour.
4. Rejet : Considérer les  $s$  plus grands coefficients de  $\mathbf{z}$ .
5. Mise à jour du résiduel :  $\mathbf{R} = \mathbf{x} - \mathbf{D}\alpha$ ;

### Question 2.

1. En se basant sur vos résultats de l'apprentissage du dictionnaire par KSVD, déduire une valeur pour l'ordre de parcimonie des vecteurs d'apprentissage.
2. Implémenter la méthode CoSaMP.

## 4.2 Relaxation $\ell_p$ et optimisation convexe

La résolution du problème d'optimisation  $\ell_0$  au sens de la pseudo norme  $\ell_0$  est très difficile, pour palier ce problème, on peut relâcher cette contrainte en la remplaçant par une autre pseudo norme  $\ell_p$  avec  $0 < p < 1$ .

### Définition 1.

Soit un vecteur  $\alpha \in \mathbb{R}^{\mathbf{K}}$  et  $p > 0$ , la pseudo norme  $\ell_p$  se définit par

$$\ell_p(\alpha) = \|\alpha\|_p = \left( \sum_{i=1}^{\mathbf{K}} |\alpha_i|^p \right)^{\frac{1}{p}}.$$

La pseudo norme  $\ell_p$  avec  $0 < p < 1$  est un bon indicateur de la parcimonie d'un vecteur. On s'intéresse donc à la possibilité de résoudre le problème  $(P_p)$  suivant :

$$(\mathcal{P}_p) : \quad \min \|\alpha\|_p, \quad \text{tel que } \mathbf{x} = \mathbf{D}\alpha$$

**Question 3.** Montrer que le problème  $(\mathcal{P}_p)$  s'écrit sous forme d'un problème de moindres carrés pondérés :

$$(\mathcal{P}_2) : \quad \min \|W\alpha\|_2, \quad \text{tel que } \mathbf{x} = \mathbf{D}\alpha$$

où  $W$  est une matrice qui dépend de  $\alpha$ .

Il s'agit d'un problème non linéaire. On peut alors procéder de manière itérative, en fixant la valeur de  $W$  à partir de la solution estimée  $\hat{\alpha}$  précédente, ce qui permet de calculer une nouvelle estimation de  $\alpha$  et ainsi de suite.

L'algorithme IRLS, pour Iteratively Reweighted Least-Squares, cherche la solution de :

$$(\mathcal{P}_2) : \quad \min \sum_{i=1}^K w_i^2 \alpha_i^2, \quad \text{tel que } \mathbf{x} = \mathbf{D}\alpha$$

où les poids sont tels que  $w_i = |\alpha_i|^{\frac{p}{2}-1}$ .

**Question 4.** Etablir les conditions d'existence de solutions du problème  $(\mathcal{P}_2)$

On cherche à obtenir la solution de  $(\mathcal{P}_2)$  par une méthode itérative. On note le poids à l'itération  $n$ ,  $w_i^{(n)} = |\alpha_i^{(n-1)}|^{\frac{p}{2}-1}$  où  $\alpha_i^{(n-1)}$  est l'estimée du signal à l'itération précédente  $n-1$ .

**Question 5.** Supposons l'itéré  $\alpha^{(k-1)}$  calculé, montrer que l'itéré courant peut s'obtenir selon la formule suivante

$$\alpha^{(k)} \leftarrow Q\mathbf{D}' \left( \mathbf{D}Q\mathbf{D}' \right)^{-1} \mathbf{x}.$$

où  $Q$  est une matrice diagonale.

Comme on s'intéresse au cas où  $0 < p < 1$ , il faut prendre en considération le fait que les poids  $w_i$  ne sont plus définis si à un moment donné,  $\exists \alpha_i = 0$ . On propose alors une régularisation des poids, on ajoute un  $\varepsilon > 0$  dans le poids :

$$w_i^{(n)} = \left( |\alpha_i^{(n-1)}|^2 + \varepsilon \right)^{\frac{p}{2}-1}.$$

En procédant ainsi, on peut garder les termes nuls de  $\alpha$  dans la matrice  $Q$  (lorsque  $w_i$  est nul, il reste le terme en  $\varepsilon$ ). Ce terme de régularisation est variable : en effet, au début des itérations, on le choisit relativement élevé pour que son influence soit minime. Ensuite, ce terme est diminué lorsque l'évolution entre deux itérations successives n'est plus suffisante, car on suppose que le signal devient parcimonieux et le recours à la régularisation devient nécessaire. On fixe un seuil minimal à ce paramètre afin de définir un critère d'arrêt pour l'algorithme. On remarque que ce seuil minimal conditionne en partie l'erreur de reconstruction.

**Algorithme IRLS**

1. On initialise le compteur d'itération  $k = 0$ , le nombre maximum d'itérations  $k_{\max}$ , le coefficient de régularisation à  $\varepsilon = 0, 1$ , et la solution de départ :

$$\alpha^{(0)} = \mathbf{D}' (\mathbf{D}\mathbf{D}')^{-1} \mathbf{x}.$$

2. On calcule les poids  $w_i = \left( \left| \alpha_i^{(k-1)} \right|^2 + \varepsilon \right)^{\frac{p}{2}-1}$  ;
3. On calcule la prochaine itération  $\alpha^{(k)} = \mathbf{Q}\mathbf{D}' (\mathbf{D}\mathbf{Q}\mathbf{D}')^{-1} \mathbf{x}$ .
  - (a) Si  $\left| \left\| \alpha^{(k)} \right\|_2 - \left\| \alpha^{(k-1)} \right\|_2 \right| > \frac{\sqrt{\varepsilon}}{100}$  et  $k < k_{\max}$  on retourne à l'étape 2 après avoir incrémenté le compteur  $k$ .
  - (b) Si  $\left| \left\| \alpha^{(k)} \right\|_2 - \left\| \alpha^{(k-1)} \right\|_2 \right| < \frac{\sqrt{\varepsilon}}{100}$  et  $\varepsilon > 10^{-8}$  alors on modifie  $\varepsilon = \frac{\varepsilon}{10}$  :
    - i. Si  $k < k_{\max}$ , on incrémente  $k$  et on retourne à l'étape 2,
    - ii. sinon, on termine.

**Question 6.** Implémenter l'algorithme IRLS.

## 5 | Procédé du compressive sensing

On considère les matrices de mesures proposées dans le cours (chapitre 3) :  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$  et  $\Phi_5$ . On rappelle que le nombre  $\mathbf{M}$  de mesures considérées nous fournit un ratio de mesures

$$r = \frac{\mathbf{M}}{\mathbf{N}}.$$

On considère le pourcentage  $\mathbf{P}$  de mesures souhaitées par rapport à la taille  $\mathbf{N}$  du vecteur d'origine,  $\mathbf{P} = r \times 100$  et le nombre de mesures  $\mathbf{M}$  comme un arrondi supérieur de  $\frac{\mathbf{P} \times \mathbf{N}}{100}$ .

**Question 7.**

1. On considère le Pourcentage de mesures  $\mathbf{P}$  égal à 15, 20, 25, 30, 50 et 75. Déterminer les vecteurs de mesures  $\mathbf{y}_i, i = 1, \dots, 5$  en utilisant les 5 matrices de mesure.
2. Calculer la cohérence mutuelle entre le dictionnaire  $\mathbf{D}$  et les différentes matrices de mesure pour chacun des pourcentages en remplissant le tableau suivant :

Cohérence mutuelle	$P = 15$	$P = 20$	$P = 25$	$P = 30$	$P = 50$	$P = 75$
$\Phi_1$						
$\Phi_2$						
$\Phi_3$						
$\Phi_4$						
$\Phi_5$						

**Question 8.** Commenter vos résultats.

On considère 3 vecteurs pour validation des résultats (fichier : vecteurs pour valider).

**Question 9.**

1. Utiliser le procédé du compressive pour reconstruire les vecteurs d'origine à l'aide des méthodes de de codage parcimonieux : MP, OMP, StOMP, CoSaMP et IRLS.
2. Calculer l'erreur relative de reconstruction en remplissant le tableau suivant pour chaque vecteur  $\mathbf{X}_j$ ,  $j = 1, 2, 3$  :

Erreurs relatives	$P = 15$	$P = 20$	$P = 25$	$P = 30$	$P = 50$	$P = 75$
MP						
OMP						
StOMP						
CoSaMP						
IRLS						

**Question 10.** Commenter vos résultats.

## 6 | Données manquantes

Pour des raisons de défaillance sur les capteurs de mesure sur les puits de forages, il arrive que des vecteurs ne sont pas entièrement complétés, mais les connaissances a priori du comportement des puits peut s'avérer utile pour compléter ces vecteurs afin de compléter l'information et en faire l'usage nécessaire.

On se propose dans cette partie de compléter le vecteur  $\mathbf{z}$ . Ce vecteur vous sera proposé avec des échantillons vides lorsque l'information est absente.

**Question 11.**

1. Considérer le vecteur (réduit)  $\mathbf{z}_r$  sans tenir compte des échantillons absents puis utiliser la (les) matrice(s) de mesure  $\Phi$  que vous avez validé dans les exercices précédents pour acquérir des mesures sur  $\mathbf{z}_r$ . Reconstruire le vecteur  $\mathbf{z}$ .

2. Considérer le vecteur  $\mathbf{z}_0$  en considérant les échantillons absents comme des valeurs nulles, puis utiliser la (les) matrice(s) de mesure  $\Phi$  que vous avez validé dans les exercices précédents pour acquérir des mesures sur  $\mathbf{z}_r$ . Reconstruire le vecteur  $\mathbf{z}$ .