1 Représentation parcimonieuse dans un dico. orthonormal

Nous avons déterminé la représentation parcimonieuse d'un signal dans la base de la transformation en cosinus discret DCT et dans celle de Fourier discrète DFT.
Pour des signaux de taille N, la matrice C de la DCT est définie, pour $0 \leqslant i, j \leqslant N-1$ par :

$$\begin{cases} C_{o,j} = \frac{1}{\sqrt{N}} \\ C_{i,j} = \sqrt{\frac{2}{N}} \cos\left(\frac{2j+1}{2N} i\pi\right) \qquad \text{pour } i \geqslant 1 \end{cases}$$

Pour la DFT, la matrice F est complexe, définie, pour $0 \leqslant p, q \leqslant N-1$, par :

$$F_{p,q} = \frac{1}{\sqrt{N}} \exp\left(\frac{-2pq}{N} i\pi\right)$$

Ici, i représente le nombre complexe de carré égal à -1.
Dans ces deux cas nous avons vu que l'obtention de cette représentation se faisait de manière directe sans besoin d'algorithme d'approximations successives.
Questions :

1. Démontrer que la matrice F de la transformation de Fourier discrète est bien une base orthonormée.
2. Générer un signal de taille $N = 500$, sur le modèle de celui du TD1, puis déterminer sa représentation parcimonieuse directement dans la base C puis dans la base F.
3. Déterminer la représentation parcimonieuse de ce même signal dans ces mêmes bases mais à l'aide de l'algorithme OMP (Orthogonal Matching Pursuit). Vous prendrez 100 comme nombre maximal d'itérations. Comparer les résultats des deux méthodes en mesurant la norme de la différence des deux représentations.
4. Recommencer les 2 procédés avec un signal généré de taille $N = 100$, et commenter les résultats.
5. Expliquer enfin pourquoi les algorithmes de type OMP ne sont pas adaptés aux dictionnaires qui forment une base comme c'est le cas ici.

2 Un autre algorithme de représentation parcimonieuse(IRLS)

Pour déterminer une représentation parcimonieuse d'un signal x dans un dictionnaire D, nous allons rempalcer le problème difficile :

$$(\mathcal{P}_0) \;:\qquad min\ \|\alpha\|_0 \text{ sous contrainte } D\alpha = x$$

par une version utilisant une autre pseudo-norme $\|x\|_p = \left(\sum_{i=1}^{k} |x_i|^p\right)^{\frac{1}{p}}$ avec $0 < p < 1$. Le nouveau problème :

$$(\mathcal{P}_p) \;:\qquad min\ \|\alpha\|_p \text{ sous contrainte } D\alpha = x$$

est ensuite transformé en problème des moindres carrés sous la forme suivante :

$$(\mathcal{P}_2) \;:\qquad min\ \|W\alpha\|_2 \text{ sous contrainte } D\alpha = x$$

Avec W matrice diagonale. Pour ce faire, on réécrit :

$$\sum_{i=1}^{k} |\alpha_i|^p = \sum_{i=1}^{k} w_i {\alpha_i}^2$$

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

Questions :

1. Monter l'équivalence du problème $(\mathcal{P}_p)$ avec le problème des moindres carrés $(\mathcal{P}_2)$, en précisant les coefficients de la matrice W.
2. Montrer que la solution de $(\mathcal{P}_2)$ peut s'écrire :

   $\alpha = QD^T \left(DQD^T\right)^{-1} x$

   où Q est une matrice diagonale liée à W et qu'il faut préciser.
3. L'algorithme IRLS (Iterated Reweighted Least-Squares) détermine α de manière itérative en fixant les poids w_i en fonction de l'estimation précédente $\alpha^{(n-1)}$ et affectant la solution à la nouvelle estimation $\alpha^{(n)}$.

   Afin d'éviter d'éventuelles divisions par 0 à cause de composantes nulles de $\alpha^{(n-1)}$, un coefficient de régularisation ε est introduit permettant de définir : $w_i = \left(|\alpha_i^{(n-1)}|^2 + \varepsilon\right)^{\frac{p}{2}-1}$.

   Ce coefficient de régularisation est diminué au fur et à mesure et sert également de critère d'arrêt. Voici le détail de cet algorithme :

   (a) On initialise le compteur d'itération $k = 0$, le nombre maximum d'itérations $kmax$, le coeffcient de régularisation à $\varepsilon = 0.1$, et la solution de départ :

   $$\alpha^0 = D^T(DD^T)^{-1}x.$$

   (b) On calcule les poids $w_i = \left(|\alpha_i^{(k-1)}|^2 + \varepsilon\right)^{\frac{p}{2}-1}$.

   (c) On calcule l'itération suivante :

   $$\alpha^{(k)} = QD^T \left(DQD^T\right)^{-1} x.$$

   i. Si $\left|\|\alpha^{(k)}\|_2 - \|\alpha^{(k-1)}\|_2\right| > \frac{\sqrt{\varepsilon}}{100}$ et $k < kmax$ on retourne à l'étape (b) après avoir incrémenté le compteur k.

   ii. Si $\left|\|\alpha^{(k)}\|_2 - \|\alpha^{(k-1)}\|_2\right| < \frac{\sqrt{\varepsilon}}{100}$ et $\varepsilon > 10^{-8}$, alors on modifie ε en $\frac{\varepsilon}{10}$.

   A. Si $k < kmax$, on incrémente k et on retourne à l'étape (b).

   B. Sinon, on termine.
4. Implémenter l'algorithme IRLS à côté des algorithmes MP et OMP.

3 Apprentissage d'un dictionnaire (k-SVD)

1. Implémenter l'algorithme d'apprentissage de dictionnaire k-SVD.
2. Appliquer le pour construire un dictionnaire de 100 atomes adapté aux 108 signaux du fichier "DonneesProjet.xlsx", en partant des 100 premières colonnes.
3. Comparer l'efficacité des 3 algorithmes de représentation parcimonieuse (MP, OMP et IRLS) en les appliquant aux 3 signaux test avec le dictionnaire obtenu par k-SVD.

4 Matrices de mesure

Nous allons utiliser le dictionnaire obtenu dans la partie précédente pour reconstruire un signal à partir d'une acquisition comprimée grâce à l'une des matrices de mesure vues en cours.

1. Choisir une matrice de mesure parmi les cinq vues en cours en justifiant votre choix.
2. Appliquer la procédure de Compressed Sensing aux 3 signaux test pour les reconstruire et comparer les signaux reconstruits avec les signaux d'origine.

5 Modalités de rendu

5.1 Date

- Vous devez adresser votre travail par mail à votre prof. avant le17avril 2022 à 23h59.
- Tout retard entraînera une note 0.

5.2 Composition des groupes

- Chaque groupe sera formé de 3 étudiants au maximum.
- Les noms des membres de chaque groupe doivent être transmis à votre prof avant le 18 mars 23h59.
- Cette composition ne peut être modifiée après cette date.

5.3 Format du rendu

- Le rendu se fera impérativement via une archive .zip nommée **MiniProjet-CS-Nom1-Nom2-Nom3.zip**.
- Cette archive doit contenir au moins deux fichier : un rapport pdf et un (ou plusieurs) fichier(s) comportant vos codes python, scilab, matlab.
- Le rapport pdf doit avoir une qualité de travail scientifique, les formules mathématiques doivent être propres et lisibles (latex conseillé).
- Aucun code informatique ne doit y être inséré, mais des graphiques et des tableaux comparatifs ainsi qu'une analyse critique de vos résultats sont attendus.
- Sur la première page de votre rapport vous indiquerez les noms et prénoms des membres du groupe et le titre du TP.