

Estimation de corrélation sur signaux compressés

A. G. Zebadua, P.O. Amblard, O.J.J. Michel et E. Moisan

GIPSA-lab, DIS(UMR CNRS 5216), ENSE3-BP 46 38402 Saint Martin d'Hères Cedex

{pierre-olivier.amblard,olivier.michel,eric.moisan}@gipsa-lab.grenoble-inp.fr,

Résumé – Ce papier compare l'estimation usuelle de fonctions de corrélation au calcul direct sur des signaux compressés (obtenus par des projections aléatoires). A nombre de mesures fixé, les résultats théoriques montrent que la compression permet de construire un estimateur plus performant que le classique, pour peu que les signaux traités soient suffisamment corrélés. Cette condition peut être considérée comme une condition de parcimonie. Nous établissons ce résultat et l'illustrons en étudiant un processus à corrélation exponentielle, et en l'appliquant à l'estimation de temps de retard.

Abstract – This paper compares the usual empirical correlation function estimator and the estimator built on compressed measurements (obtained by random projections). When the number of measurements is fixed, theoretical results show that compressive measurements allow to have a more efficient estimator if the processes studied have a sufficiently low decreasing correlation function. This condition can be interpreted as a sparseness condition. The results are established and illustrated by studying an exponentially correlated process. They are further applied to the estimation of time delay.

1 Motivations

Le contexte de cette étude est la surveillance passive d'un milieu, à l'aide d'un réseau de capteurs disséminés au hasard. Les variations du milieu se reflètent dans la géométrie du réseau. Estimer cette géométrie permet alors de collecter des informations sur le milieu et, partant, de le surveiller. Connaître la géométrie du réseau nécessite de déterminer les distances inter-capteurs. Celles-ci sont évaluées grâce à l'identité de Ward, qui relie la fonction de Green du système étudié entre les capteurs considérés, à la fonction d'inter-corrélation des signaux mesurés en chaque capteur ([3, 4]). L'aspect passif entraîne que les sources apparaissent aléatoirement, à des positions inconnues et présentent des contenus spectraux non contrôlés.

Dans cette application, les capteurs sont contraints à n'utiliser que très peu d'énergie. Ils ne peuvent effectuer que peu de calculs localement, mais surtout, ils ne peuvent communiquer que très peu avec leurs voisins ou avec une station de base. Nous nous plaçons donc dans un contexte d'estimation distribuée avec des contraintes fortes de communication. Il est par exemple exclu qu'un capteur communique en permanence avec un voisin, pour lui transmettre toutes les perturbations qu'il mesure, afin que ce voisin corrèle ces perturbations avec les siennes pour estimer un temps de retard. Notre objectif consiste à réduire les communications dans le réseau, afin de minimiser l'énergie nécessaire au fonctionnement de chaque capteur.

Dans ce papier nous considérons l'estimation de la fonction de corrélation entre deux signaux aléatoires après une acquisition compressée. Evidemment, l'intérêt est d'éviter le passage par les étapes de reconstruction classiquement étudiées dans la littérature sur l'acquisition compressée [7]. En effet, cette phase de reconstruction s'effectue à l'aide d'algorithmes d'optimi-

sation que l'on ne peut envisager d'implanter localement (au niveau capteur). De même, la communication d'échantillons compressés à une station de base opérant la reconstruction requerrait trop d'énergie. La stratégie envisagée est de corrélérer puis estimer localement pour ne transmettre que les informations nécessaires, entre voisins proches.

A l'instar des idées présentées dans [6], il est possible de traiter les signaux acquis dans le domaine compressé. Dans le même esprit, nous avons déjà obtenu des résultats encourageants lors d'un problème de classification de signaux neuronaux compressés [5]. Dans ce papier, nous estimons la fonction de corrélation entre deux signaux aléatoires à partir d'échantillons compressés. Utiliser les idées de l'acquisition compressée pour l'estimation de matrices de corrélation a été envisagé par exemple dans [8] (et les références incluses). Dans ces études, le principe est de représenter la matrice de corrélation comme combinaison linéaire de matrices de corrélation connues, et d'estimer les coefficients de la combinaison à partir d'échantillons compressés (par des techniques de « correlation matching »). Ici, nous estimons directement la corrélation et montrons que, dans certaines situations, l'estimée effectuée à partir des échantillons compressés est proche de l'estimation sans compression.

La section suivante présente les estimateurs et leur étude à l'ordre 2. Pour illustrer et interpréter les résultats, nous considérons le cas d'une corrélation exponentielle. Nous discutons également le choix des matrices de mesure, et montrons que des matrices très creuses et permettant des calculs simples conduisent à des estimations très acceptables. Nous montrons par exemple que l'estimation de temps de retard à l'aide de ces fonctions compressées reste très honorable.

2 Estimateurs

Les divers signaux qui parviennent aux capteurs sont supposés aléatoires, de moyenne nulle, conjointement stationnaires et ergodiques. Comme cela a été souligné dans l'introduction, l'identification de la distance inter-capteurs, à partir de perturbations non contrôlées, requiert l'estimation de la fonction d'inter-corrélation des signaux mesurés en chacun des capteurs. Soient $x(t)$ et $y(t)$ les signaux mesurés par deux capteurs voisins. On note $\Gamma_{xy}(\tau) = E[x(t)y(t+\tau)]$ leur fonction d'intercorrélacion pour le retard τ .

Sur chaque capteur, les signaux sont échantillonnés à une même cadence, respectant la condition de Shannon, de sorte que les quantités manipulées s'écrivent comme des vecteurs de taille N , nombre d'échantillons successifs stockés sur chaque capteur : $\mathbf{x} = (x_t, \dots, x_{t-N+1})^\top$ et $\mathbf{y} = (y_{t+\tau}, \dots, y_{t+\tau-N+1})^\top$.

L'estimateur usuel de l'intercorrélacion est $\hat{\Gamma}_{N,xy}(\tau) = N^{-1}\mathbf{x}^\top\mathbf{y}$. Notre objectif est de le comparer à l'estimateur qui découle du traitement des signaux compressés, $\Phi\mathbf{x}$ et $\Phi\mathbf{y}$, où Φ désigne la matrice $M \times N$ de compression ($M < N$). Cette matrice aléatoire est constituée d'éléments φ_{ij} , indépendants et identiquement distribués. Nous précisons ultérieurement la loi de probabilité f_φ de ces coefficients, mais ils sont toujours supposés de moyenne nulle et indépendants des signaux eux-mêmes. Dans ce cas, l'intercorrélacion s'estime selon

$$\hat{C}_{N,xy}(\tau) = (\Phi\mathbf{x})^\top(\Phi\mathbf{y}) = \mathbf{x}^\top\Phi^\top\Phi\mathbf{y} \quad (1)$$

Les propriétés de l'estimateur classique $\hat{\Gamma}_{N,xy}(\tau)$ sont bien connues ([1, 2]). Il est non biaisé et sa variance est donnée par

$$\text{Var}[\hat{\Gamma}_{N,xy}(\tau)] = \frac{1}{N^2} \sum_{k=-N}^N (N - |k|) f_{xy}^\tau(k) \quad (2)$$

où $f_{xy}^\tau(k) = \text{Cum}[x(t), y(t+\tau), x(t+k), y(t+k+\tau)] + \Gamma_{xy}(\tau+k)\Gamma_{xy}(\tau-k) + \Gamma_{xx}(k)\Gamma_{yy}(k)$ (Cum désigne le cumuland des variables en argument). L'estimateur converge en $1/N$ sous réserve de sommabilité des fonctions de corrélation d'ordre 2 et 4 apparaissant dans $f_{xy}^\tau(k)$. Pour la version compressée, on obtient pour la moyenne

$$E[\hat{C}_{N,xy}(\tau)] = MN\Gamma_{xy}(\tau) \text{Var}[\varphi], \quad (3)$$

qui implique d'imposer aux coefficients de la matrice Φ de satisfaire $\text{Var}[\varphi] = (MN)^{-1}$ comme contrainte de non biais.

Le calcul est plus fastidieux à l'ordre deux. Si on utilise $\text{Var}[\hat{C}_N] = E_{xy}[\text{Var}[\hat{C}_N|x, y]] + \text{Var}_{xy}[E[\hat{C}_N|x, y]]$ et la contrainte de non biais, on aboutit à

$$\text{Var}[\hat{C}_{N,xy}(\tau)] = (1 + \frac{1}{M})\text{Var}[\hat{\Gamma}_{N,xy}(\tau)] +$$

$$\frac{1}{M}(\Gamma_{xx}(0)\Gamma_{yy}(0) + \Gamma_{xy}(\tau)^2 + \frac{1}{N^2} \sum_{k=-N}^N (N - |k|) g_{xy}^\tau(k))$$

$$+ MN\text{Cum}_4[\varphi] (g_{xy}^\tau(0) + \Gamma_{xx}(0)\Gamma_{yy}(0))$$

La correction à la variance de $\hat{\Gamma}_{N,xy}(\tau)$ est d'ordre $1/M$ évidemment. Toutefois, il est intéressant de comparer ce comportement en $1/M$ à la variance que l'on obtient avec l'estimateur empirique usuel sur M échantillons, à savoir $\hat{\Gamma}_{M,xy}(\tau)$.

Notons enfin que la matrice de compression intervient explicitement dans le résultat à travers le cumuland d'ordre 4 de ses éléments.

Pour comparer ces résultats aux précédents, il devient nécessaire de définir la loi des processus x et y , afin d'expliciter les moments d'ordre 4 qui apparaissent dans ces équations, ainsi que la loi des coefficients de la matrice de compression.

2.1 Processus gaussien AR(1)

Limitons nous à l'estimation de l'autocorrélacion d'un processus autorégressif $x_t = ax_{t-1} + \sqrt{1-a^2}\varepsilon_t$, où ε_t désigne une séquence i.i.d. de variables aléatoires gaussiennes centrées réduites, tandis que le coefficient a appartient à l'intervalle $] -1, 1[$. La fonction d'autocorrélacion de x_t est donnée par $\forall \tau \in \mathbb{Z} \quad \Gamma_{xx}(\tau) = a^{|\tau|}$.

On ne considère pour comparer les estimateurs que l'estimation de $\Gamma_{xx}(0)$. La variance de l'estimateur usuel, exploitant M échantillons successifs des signaux vaut

$$\text{Var}[\hat{\Gamma}_{M,xy}(0)] = \frac{2}{M} + \frac{4a^2}{M(1-a^2)} \left(1 - \frac{1-a^{2M}}{M(1-a^2)}\right) \quad (4)$$

alors que celle de l'estimateur compressé devient

$$\begin{aligned} \text{Var}[\hat{C}_{N,xy}(0)] &= \frac{2}{M} + \frac{2}{N} + \frac{4}{MN} + \frac{3\text{Cum}_4[\varphi]}{MN} \\ &+ \frac{4(M+2)a^2}{MN(1-a^2)} \left(1 - \frac{1-a^{2N}}{N(1-a^2)}\right) \end{aligned} \quad (5)$$

On étudie alors la différence entre les deux variances $\delta_M := \text{Var}[\hat{C}_{N,xy}(0)] - \text{Var}[\hat{\Gamma}_{M,xy}(0)]$. Nous discutons dans la suite le choix de la matrice de compression et l'influence du coefficient de régression a sur δ_M .

2.2 Choix d'une matrice de compression

Jusqu'à présent, nous avons supposé les coefficients φ_{ij} i.i.d. et centrés, de sorte que leurs moments d'ordre impair sont nuls. De plus, leur variance est prise égale à $1/MN$, pour que l'estimateur de la corrélation soit non biaisé. Il reste un degré de liberté quant au moment d'ordre quatre. Examinons quelques possibilités dont la mise en œuvre est aisée.

Matrice gaussienne $\varphi \sim \mathcal{N}(0, 1/MN)$: ce choix relativement fréquent dans la littérature a pour avantage d'annuler le terme $\text{Cum}_4[\varphi]$ intervenant dans le calcul de la variance (équ. 5). Un inconvénient majeur est que, localement, le calcul des compressions est alourdi par le fait que tous les coefficients diffèrent les uns des autres.

Matrice de Bernoulli $\varphi = \pm 1/\sqrt{MN}$, avec la même probabilité $1/2$: le moment d'ordre quatre vaut $(MN)^{-2}$, si bien que le cumuland d'ordre 4 $\text{Cum}_4[\varphi] = -2(MN)^{-2}$ est négatif. Cette correction avantageuse est toutefois d'un ordre de grandeur inférieur au comportement général. L'avantage de cette matrice réside donc plutôt dans une implantation plus aisée que la matrice gaussienne, mais requérant toutefois également beaucoup de calculs.

Matrice ternaire $\varphi = \pm(2N)^{-1/2}$ avec la même probabilité M^{-1} et $\varphi = 0$ avec la probabilité complémentaire $1 - 2M^{-1}$: la matrice résultante est creuse, ce qui allège considérablement le calcul de la compression (il n'y a en moyenne que $2N/M$ termes non nuls par ligne). Par contre, le cumuland d'ordre 4 vaut $\text{Cum}_4[\varphi] = (MN)^{-2}(M/2 - 3)$, qui varie en $1/M$. Utiliser cette matrice creuse se paie donc au prix d'une perte de performance, qui reste toutefois minime comme illustré dans la suite.

Nous pouvons conclure cette discussion en notant que le choix de la matrice de compression n'est pas majeur quant à la différence des variances des estimateurs. Dans la suite, nous considérons alors le cas d'une matrice de compression gaussienne, afin de négliger le terme dépendant du cumuland d'ordre 4 des φ dans les expressions des variances.

2.3 Comparaison en fonction de a

Le modèle simple du processus autorégressif permet d'examiner l'influence de la compression en fonction de la nature spectrale des signaux. Pour $a = 0$, le signal est un bruit blanc, l'archétype d'un signal non compressible. Notons que nous n'employons pas le terme compressible ici comme dans la littérature usuelle de l'acquisition compressée. Usuellement, la compressibilité est reliée à l'amplitude des signaux (ou des coefficients d'une représentation). La compressibilité ici est liée à la redondance apportée par la corrélation entre échantillons du signal. Lorsque $|a|$ s'éloigne de 0, le signal devient de plus en plus compressible, et la limite $|a| = 1$ en est le comble puisque dans ce cas le modèle dégénère en un signal constant (rappelons que pour garantir la stationarité, la variance de l'innovation est fixée à $1 - a^2$).

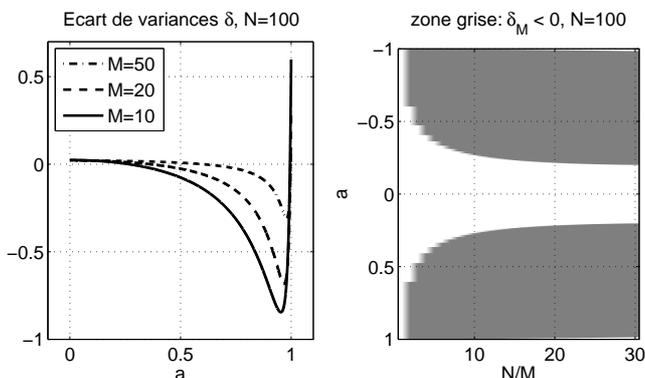


FIGURE 1 – Gauche : Différence des variances des estimateurs, compressé - classique, en fonction de la valeur absolue du coefficient de régression du processus x_t . Pour un couple (M, N) donné, il existe une plage dans laquelle l'estimateur compressé est meilleur que le classique. Droite : en gris, zone du plan $(a, N/M)$ dans laquelle l'estimateur compressé est meilleur que l'estimateur utilisant M échantillons consécutifs. Pour les deux figures, $N = 100$.

La différence de variances δ_M est tracée sur la figure (1)

gauche, pour $N = 100$ et trois valeurs de M (10, 20 et 50).

Pour $a = 0$, $\delta_M = 2N^{-1}(1 + 2M^{-1})$ est positif. Mais δ_M devient négatif pour des coefficients de régression dont le module augmente ! L'effet perdu pour des valeurs de $|a|$ proches de 1. Pour a très proche de 1, la tendance s'inverse à nouveau, et δ_M redevient positif.

Lorsque le signal est très corrélé, typiquement $a = 0.9$, M échantillons successifs apportent moins d'information que M compressions issues d'un horizon de N échantillons, donc plus large. Par contre, lorsque a est proche de zéro, considérer M échantillons aléatoirement construits à partir de N successifs est équivalent à considérer M échantillons successifs. Ainsi, plus le signal est compressible, plus la compression améliore l'estimation vis à vis de M échantillons successifs.

Il est naturel que δ_M redevienne positif pour a très proche de 1. Si $a=1$, le processus est dégénéré et $x_t = x_0, \forall t$ où x_0 est la variable aléatoire initiale gaussienne, centrée réduite. L'estimateur de la puissance est constant $\hat{\Gamma}_{M,xx}(0) = x_0^2$, ainsi que l'estimateur compressé $\hat{C}_{N,xx}(0) = x_0^2 \sum_{k,i,j} \varphi_{ki} \varphi_{kj}$. Dans le cas d'une matrice gaussienne, l'écart de variance vaut alors $6/M$. Comme l'observation est constante dans ce cas, il est naturellement moins bon de faire des combinaisons aléatoires de cette constante que de la considérer seule.

D'autre part, nous étudions le comportement de l'estimation en fonction du rapport $N/M > 1$ qui définit le taux de compression. La figure (1) gauche montre que le comportement décrit avant est stable pour trois rapports égaux à 2, 5 et 10. Toutefois, la zone des a pour lesquels l'amélioration existe dépend du taux de compression. La figure (1) droite représente en grisé la zone $\delta_M < 0$ d'amélioration, en fonction de a et du taux de compression.

Pour illustrer visuellement l'effet décrit, la figure (2) montre la variance expérimentale des trois estimateurs $\hat{\Gamma}_{N,xx}(\tau)$, $\hat{\Gamma}_{M,xx}(\tau)$ et $\hat{C}_{N,xx}(\tau)$ en fonction du retard τ . Les fonctions tracées sont multipliées par N , et nous avons choisi pour cet exemple $M = N/10 = 100$. Le coefficient de régression a prend trois valeurs (0, 0.1 et 0.7).

Pour $a = 0$ on retrouve la correction en $1/M$ de $\hat{C}_{N,xx}(\tau)$ par rapport à $\hat{\Gamma}_{M,xx}(\tau)$. Passer de $a = 0$ à $a = 0.7$ montre clairement que l'estimateur compressé, en étalant ses M mesures utiles sur un horizon de taille N , lutte plus efficacement contre la corrélation que l'estimateur empirique usuel. D'autre part, cet exemple montre que la perte en terme de variance de l'estimateur compressé par rapport à l'empirique sur N échantillons est d'autant plus faible que la corrélation décroît lentement.

2.4 Estimation de temps de retard

Pour replacer cette étude dans le contexte de l'introduction, on applique ici les estimateurs précédents pour estimer un temps de retard. On considère un processus aléatoire x_t généré comme sortie du filtre de réponse impulsionnelle causale $h_t = \cos(2\pi 5t)e^{-8t} + 0.4 \cos(2\pi 26t)e^{-t}, t \geq 0$ attaqué par un bruit blanc gaussien centré de puissance unité. Le signal y_t est une version retardée bruitée $y_t = x_{t-\tau} + \sigma b_t$, où b_t est un bruit blanc gaussien centré normé indépendant des

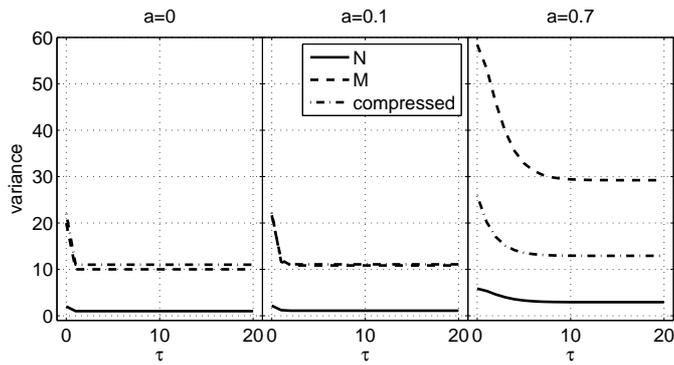


FIGURE 2 – Variance des estimateurs de la fonction de corrélation d'un processus autorégressif d'ordre 1, pour différentes valeurs du coefficient de régression. A nombre de points constant, l'estimateur construit sur les signaux compressés peut s'avérer meilleur que l'estimateur usuel, pour peu que le processus soit suffisamment corrélé.

signaux d'intérêt. τ est choisi à $0.297s$ alors que la période d'échantillonnage est $0.01s$. Le temps de retard n'est donc pas sur la grille d'échantillonnage. Nous faisons varier σ de 0 à 1, simulons pour chaque valeur 1000 réalisations des signaux de façon à estimer la position du maximum dans les fonctions de corrélation. Nous représentons figure (3) gauche la probabilité empirique que le maximum soit à la position 31 dans les fonctions de corrélation estimées. Dans cet exemple, $N = 1000$, et nous utilisons trois taux de compression $N/M = 5, 10$ et 20 . La figure droite montre des fonctions de corrélation estimées sur la même réalisation pour les trois estimateurs $\hat{\Gamma}_{1000}$, $\hat{\Gamma}_{50}$ et \hat{C}_{50} . Il est remarquable de constater que pour un rapport de compression aussi fort que 20, le fait d'utiliser des échantillons compressés permet de récupérer environ 5% de probabilité de bonne estimation. On remarque également que, pour de faibles taux de compression, l'estimation compressée supporte moins bien les bruits d'observation que l'estimation usuelle. Pour terminer, notons que la matrice de compression choisie ici est la matrice ternaire. Ces résultats suggèrent ainsi que des capteurs utilisant peu de mesures et peu d'échanges avec leurs voisins peuvent estimer localement des temps de vol entre leurs voisins et eux.

3 Conclusion

La décroissance des fonctions de corrélation peut être interprétée en terme de compressibilité d'un signal aléatoire. De manière évidente, pour des signaux présentant des compressibilités fortes, il est avantageux pour l'estimation de corrélation de travailler sur des échantillons compressés de taille M choisis sur un horizon de taille $N > M$, plutôt que considérer M échantillons successifs. Ce fait n'est pas très intéressant si les moyens de calculs disposent d'énergie et de possibilité de stockage en quantité, mais devient remarquable dans des contextes très contraints comme les réseaux de capteurs autonomes dispersés qui sont de plus en plus utilisés pour des tâches de sur-

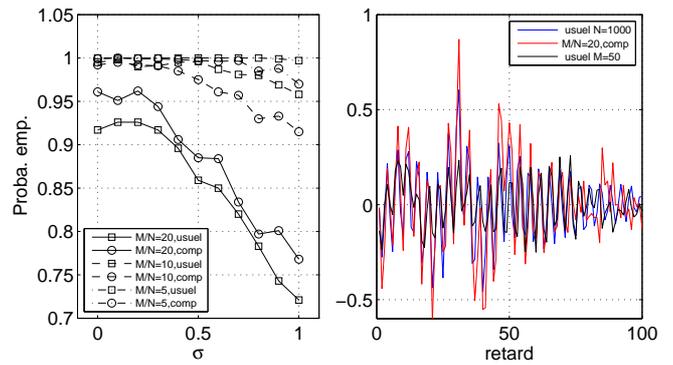


FIGURE 3 – Gauche : Probabilité de bonne localisation du maximum de la fonction d'intercorrélation en fonction de l'écart type du bruit d'observation dans le problème d'estimation de temps de retard étudié dans le texte. Droite : Illustration des fonctions de corrélation estimées, pour une puissance de bruit d'observation de 1, $N = 1000$ et $M = 50$. L'estimateur empirique usuel sur M échantillons est en noir, sur N en bleu, et le compressé en rouge.

veillance.

Nous travaillons à la poursuite de l'étude esquissée ici dans deux directions. La première consiste en l'analyse asymptotique à taux de compression fixé. La deuxième vise à coupler la compression temporelle à de la compression spatiale. Nous observons en effet que les estimations présentées ici résistent particulièrement bien à des quantifications fortes des signaux compressés.

Références

- [1] P. J. Brockwell and R. A. Davis. *Time Series : Theory and Methods*. Springer, 1991.
- [2] D. R. Brillinger. *Time Series. Data analysis and theory*. SIAM, Philadelphia, 2001.
- [3] M. Carmona, O.J.J.Michel,J.L. Lacoume, N.Sprinsky, B. Nicolas. *Ward identities for visco-acoustic and visco elastic propagation media* Wave motion, vol.49-4, pp. 484-489, june 2012.
- [4] M.Carmona, O.Michel, J.L.Lacoume, B. Nicolas, R. Vincent. *On Green's correlation of Stokes equation* Wave motion, fev. 2015, dx.doi.org/10.1016/j.wavemoti.2015.02.011.
- [5] Coppa, B. and Héliot, R. and Michel, O. and Moisan, E. and David, D. *Low-Cost Intracortical Spiking Recordings Compression with Classification Abilities for Implanted BMI Devices*. IEEE Engineering in Medicine and Biology Society IEEE, EMBC 2012.
- [6] M.A. Davenport, P.T. Boufounos, M.B. Wakin, R.G. Baraniuk. *Signal processing with compressive measurements* IEEE Journal of selected topics in signal processing, vol. 4 n. 2, pp.445-460, 2010.
- [7] Y. C. Eldar and G. Kutyniok, editors. *Compressed sensing. Theory and applications*. Cambridge University Press, 2012.
- [8] D. Romero, R. López-Valcarce, and G. Leus. Compression limits for random vectors with linearly parametrized second-order statistics. *IEEE Trans. Inform. Theory*, 61(3) :1410–1425, 2015.